Editorial Preface

From the Desk of Managing Editor...

It may be difficult to imagine that almost half a century ago we used computers far less sophisticated than current home desktop computers to put a man on the moon. In that 50 year span, the field of computer science has exploded.

Computer science has opened new avenues for thought and experimentation. What began as a way to simplify the calculation process has given birth to technology once only imagined by the human mind. The ability to communicate and share ideas even though collaborators are half a world away and exploration of not just the stars above but the internal workings of the human genome are some of the ways that this field has moved at an exponential pace.

At the International Journal of Advanced Computer Science and Applications it is our mission to provide an outlet for quality research. We want to promote universal access and opportunities for the international scientific community to share and disseminate scientific and technical information.

We believe in spreading knowledge of computer science and its applications to all classes of audiences. That is why we deliver up-to-date, authoritative coverage and offer open access of all our articles. Our archives have served as a place to provoke philosophical, theoretical, and empirical ideas from some of the finest minds in the field.

We utilize the talents and experience of editor and reviewers working at Universities and Institutions from around the world. We would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations. Our high standards are maintained through a double blind review process.

We hope that this edition of IJACSA inspires and entices you to submit your own contributions in upcoming issues. Thank you for sharing wisdom.

Thank you for Sharing Wisdom!

Editorial Board

CONTENTS

(ix)

Investigating the User Experience of Mind Map Software: A Comparative Study based on Eye Tracking

Junfeng Wang, Xi Wang, Jingjing Lu, Zhiyu Xu

College of Design and Innovation, Shenzhen Technology University, Shenzhen, China

Abstract—Software for creating mind maps is currently prevalent, and it should have strong usability and create a good user experience. Usability testing can help to uncover flaws in software's usability and support its optimization. This paper took the mind map software "Xmind" and "MindMaster" as study cases and conducted comparative research on three aspects: effectiveness, efficiency, and satisfaction. The research investigated 20 participants' interactions with the two software. Task completion rate, number of errors, and number of requests for help were collected to evaluate the effectiveness. Eye tracking data and task completion time are collected to evaluate efficiency. System usability, interface quality, and emotional dimensions were collected with subjective scales to assess the software's user satisfaction. The data together led to a conclusion: each software has a few usability issues. The use of jargon to explain functions was costly to learn and quickly undermined users' confidence in using the software; the interface's simplicity impacted satisfaction, although users tended to evaluate utility tools in terms of their ease of use and ease of learning. These findings could be used to optimize utility software.

Keywords—Usability; mind map software; comparative research; eye tracking; user experience

I. INTRODUCTION

The Mind Map was created in the twentieth century in England. Tony Buzan discovered a prototype of a form of representation concerning the radioactive mind and its images (i.e., mind map) while studying the neurophysiological science of the brain[1]. As a practical tool to visualize the thinking process, it has been widely used in teaching, business, personal, and team information management and is still penetrating other areas. Because of the continuous development of mobile network technology and application devices, software for drawing mind maps increased gradually and variously. Current research primarily focuses on applying mind map software in specific fields, with little concern about usability and user experience. As a utility tool, mind map software requires good usability and user experience to help users draw mind maps much simpler and increase efficiency and pleasure. In-depth usability research in this paper can help software designers understand better how the interaction logic, information quality, and interface layout influence the user experience together.

This study examines the usability of mind map software from three aspects: effectiveness, efficiency, and satisfaction. 20 subjects were invited to complete the given tasks with mind map software, and data were collected through eye tracking, user interview, and usability questionnaire. The usability problems in the interaction and visual aspects of the mind map software were explored.

The paper is organized as follows. The literature review is explained in Section II. Next, the methods adopted for this research is described in Section III, followed by the experimental tasks and procedure in Section IV, and evaluation metrics in Section V. The research results of usability and user experience are analyzed in Section VI. And the discussion in Section VII. Finally, the conclusion and future works are given in Section VIII.

II. LITERATURE REVIEW

For mind map software, good usability helps users to record and disperse their thoughts efficiently. User experience assessment can identify software problems regarding effectiveness, efficiency, and user satisfaction.

A. Implementation of Mind Map Software

Mind mapping is the process of constructing concepts graphically[2]. The core concept acts as the center, and related and subordinate concepts are grouped into various branches. A new concept is derived from an additional point of an existing concept, forming a tree structure in the diameter direction[3]. With eye-catching colors, simple graphics, and logical and precise lines, mind maps help increase learning and job productivity by making words and theories systematic and neatly structured[4].

In recent years, mind maps have become increasingly significant in education and teaching reform at all levels, from elementary school to university. Many studies use mind mapping to improve teaching and learning. For example, Fung D[5] investigated the effectiveness of mind maps in in Hong Kong Primary Science Classrooms; YANG A et al.[6] used mind maps in the teaching practice of elementary school composition; WEI C et al. [7] used mind maps in the teaching practice of human parasitology. These studies show how visual tools can improve teaching and learning and promote more profound knowledge. Olga Maksimenkova et al. [8] created an automatic grader for educational mind maps (AGEMM) based on the quantitative properties (node count, associations count, image count, and branching levels), which acts like a teacher's assistant.

In several other fields, including Finance [9], Medicine [10], Marketing [11], Economics [12], Engineering [13], Management [14], Computer [15], Fine Art and Design [16], Advertising [17] and Public Relations [18], etc., mind maps are also used to facilitate the collaboration in the workplace. Mind mapping contributes to collaborative work, especially during the COVID-19 pandemic today. Dilshod Kuryazov et al. [19] introduced the tool CMCM (Collaborative Mind-and Concept-Mapping) to demonstrate the viability of tool cooperation made possible by model differencing used for the collaborative modeling of Mind Maps.

However, little research has been done on drawing mind maps. According to Joeran Beel et al.[20], users generally construct "typical" mind maps that are relatively short, usually only a few hundred branches, seldom use annotations, linkages, and other features, and are typically modified repeatedly within a day or two. This indicates that people care more about mind maps' timeliness than the amount of information, which implies that more attention should be paid to the software's efficiency and efficacy while drawing.

However, being efficient and effective does not equate to a positive User Experience (UX). Many scholars have done a significant amount of usability research on digital libraries, online advertising, and systems engineering. Sudatta Chowdhury et al. [21] divided the concept of usability into two main lines (library intelligence and human-computer interaction). Hansen P [22] considered that usability in human-computer interaction systems refers to a particular interface's effectiveness, efficiency, and satisfaction. Usability is defined by the International Organization for Standardization (ISO) as the "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [23]. The specific interpretations are as follows:

- Effectiveness: the user's precision and completeness in achieving the specified goal.

- Efficiency: the resources consumed by the user to precisely achieve the goal.

- Satisfaction: the user's comfort and acceptability of use.

However, ISO 9241 focuses too much on established tasks and goals, emphasizes efficiency and effectiveness, and ignores what is challenging to grasp in user experience. Furthermore, it does not apply to all scenarios, and satisfaction is also unsuitable as a usability indicator in many cases[24]. Even though ISO 9241 has been criticized for being inadequate as a criterion for evaluating usability, it is still extensively used.

Usability is a highly abstract term difficult to quantify and cannot be precisely tested. Many researchers have decided to build an assessment index system based on each index to assess the system's overall usability. According to Nielsen[25], usability can be divided into "efficiency of use," "ease of learning," "memory retention," "use reliability," and "user satisfaction." "response time," "learning time," "long retention time," "user error rate," and "subjective satisfaction" were chosen by Schneiderman[26] as indicators of user interface usability. Preece et al.[27] chose "throughput," "ease of learning," and "attitude" to measure the usability of HCI systems; Shackel[28] combined "effectiveness (speed)," "ease of learning (learning time)," "ease of learning (memory retention)," "effectiveness (error rate)," and "attitude" to define as usability evaluation index systems. In addition, content metrics and user-centered metrics can be categorized under the usability metrics system [29].

Current research has clarified the metrics for usability evaluation and established many theoretical foundations. However, there are specific variances between different genres, and the mind map software evaluation system should be updated to consider the peculiarities of mind mapping and user habits. This study employs a user-centered approach to usability evaluation, accounting for the efficacy, efficiency, learnability, and usability of mind-mapping software, user interface, and user sentiment metrics.

B. Usability Methods based on Eye Tracking

Typical usability assessment techniques include questionnaires, heuristic evaluation, focus groups, think-aloud, and card sorting[30]. However, the main limitations in implementation, particularly in user interface evaluation, are as follows[31]:

- The host will unavoidably affect the subjects and their behavior during the test through words, deeds, expressions, and expectations. This is especially true when there are several staff members.

- The test result's objectivity is not very good. The test mainly gathers qualitative information about individuals' preferences and inclinations.

- The experimental control is not strict enough during the test, and the reproducibility of the test results is low[32].

- The internal processing of the user is complicated for the test data to reflect directly, and the interpretation of the phenomenon is based chiefly on the tester's knowledge and experience, which is subjective.

- The test is costly, the sample size is limited, and there are issues with the statistical significance and representativeness of the test results.

Eye movement is crucial for visual processing information when using a computer interface. Eye tracking was initially recognized in reading studies in the late 19th century, and the first eye tracking apparatus was created in the early 20th century. As eye tracking technology advances, experts and academicians gradually discover the relationship between eye movements and human cognitive processes. Paul[33] and his colleagues were the earliest researchers who employed eye tracking in user experience research in 1947. This method was not frequently used until the late twentieth and early twenty-first centuries. According to Aga Bojko [34], objective data gained from eye tracking aids in better understanding how

different designs affect user experience and the demands, preferences, motivations, and processes of consumers or potential product users. According to Ivory and Hearst, visual recordings can be used for usability testing[35].

Eye tracking is a technology that records an individual's eye movements and can be used for usability assessment, especially for assessing visual user interfaces. This could be achieved by analyzing the data generated using this technology[35], including Fixation, Saccade, Behavior metrics, and data related to areas of interest (AOI) that dominate the data presented here. Data can also be visualized as gaze maps, scan paths, gaze videos, swarm maps, heat maps, focus maps, etc. According to Sandra Milena et al.[36], the aspects of usability evaluation that can be performed in digital information environments by eye tracking metrics primarily include search efficiency, interface quality, information visibility, and expectation flow.

Few studies have used eye tracking technology to evaluate dynamic processes and overall usability, while data from eye-tracking devices alone cannot shed additional light on user experience. This study attempts to assess mind map software's effectiveness, efficiency, and user satisfaction by combining subjective scale scoring and objective eye-movement data. The eye movement heat map was used to investigate the factors influencing the software's overall usability. We also combine interviews, and user behavior observations to more thoroughly argue the deeper reasons for user behavior. As a result, more reliable usability conclusions can be reached.

III. METHODOLOGY

The technique for the experiment's usability testing is described in the following section. The source of the sample software is discussed, along with how the task was created so that the participants could understand how to use the software and incorporate eye tracking into the experiment.

A. Study Design

An assessment test and a comparison test were both employed in this experiment. As much as possible, we gathered qualitative tendency data and quantitative operational data using simulated tasks, interviews, observations, and eye tracking [37]. We can evaluate which alternative is easier to use and learn by contrasting the various design possibilities of two software examples. Additionally, it aids in our comprehension of the advantages and disadvantages of multiple designs.

Due to variances in the mind-mapping process among the individuals, the functions and time spent generating distinct graphs varied. The task scenario was set to "You need to use the software which we offered to draw a mind map for work reporting as shown in the case diagrams" to control the variables. The individuals were instructed to execute the tasks using the sample software, and we gathered usability-related data.

B. Sample Software

A web-based questionnaire was used to conduct a study on the use of vector drawing software before the start of the experiment. Ninety-four valid questionnaires were collected,

with 64 people (68.09%) having previous vector image drawing experience. Mind maps were drawn by 82.81% of those with vector drawing experience in the sample. Xmind was chosen by 38 people, accounting for 71.7% of those who used related software. Visio (11 people, accounting for 20.8%), EDraw (8 people, accounting for 15.1%), MindMaster (6 people, accounting for 11.3%), and ProcessOn (6 people, accounting for 11.3%) were the others in that order. MindMaster is more focused on mind mapping because EDraw and MindMaster are owned by the same company and have similar product designs. According to market share, this study chooses Xmind and MindMaster as the study cases.

In the beginning, the similarities between Xmind and MindMaster as mind map software are primarily centered on the functional level, as both are designed to suit the needs of a daily mind map. Both provide a range of templates when producing a new drawing; the basic features of the drawing are essentially the same. Both provide labels, pictures, and other elements, and both may export standard image formats after the drawing is completed.

In addition to the preceding similarities, the two software differ significantly in information hierarchy, page layout, and visual style. MindMaster externalizes more information than Xmind in terms of information hierarchy, stacking as many functions as possible on the top toolbar. In contrast, Xmind hides more functions in the secondary directory, which must often be retrieved via the top menu bar or right-click. Table I displays the exact differences.

TABLE I.   DIFFERENCES IN DESIGN BETWEEN THE TWO SOFTWARE

Differences	*Xmind*	*MindMaster*
Top Toolbar	Other functions in the first level of the menu bar drop-down menu are only for frequently used parts in the form of text with icons in the second level.	Using tabbed management, huge categories beneath the text and icons to indicate specific available options, and the default state to display the most often used tools.
Free Themes	Different levels from "Topics" and "Subtopics" are placed at the top of the page under "Insert."	In the top toolbar category, "Topics" and "Subtopics" are on the same level.
Side Format Panel	Default Hidden	Default Expand
Presentation of Auxiliary Functions	Preference for textual descriptions.	Pure icons are preferred, with text descriptions loading after the mouse hovers over.
Visual Style	The Minimalist style is dominated by grey and white tones.	Simple yet not excessively energetic green as the theme color
Online Features	None	Cloud files, online sharing, etc.

C. Participants

Five people can discover around 77%-85% of the detection system's usability difficulties, according to Nielsen's fitted regression curve $n=N\ 1-(1-L)^n$ [38] where n is the number of users, N is the total number of usability problems in the design and L is the proportion of usability problems discovered while testing a single user. However, because substantial variations between participant groups from a smaller number of participants are difficult to uncover, Spyridakis [39] stated that

10-12 people would better properly investigate the product's usability concerns.

Based on their familiarity with the software, time spent using it, and frequency of use, inexperienced and experienced participants were purposefully picked for the experiment. Twenty participants were chosen from a Chinese university with various specialties. They were split into two groups, A and B; each group had five novice participants and five veterans.

D. Materials

1) Experimental equipment: The eye movement data were recorded by the Tobii-Glasses2 RU-type oculomotor, and the data were analyzed using the Tobii Pro Lab software. With a weight of only 45 grams, this eye-tracking head module model provides an ultra-lightweight and sturdy non-intrusive head-tracking module that ensures the interviewee's comfort and freedom of movement. Lenses for vision correction are also appropriate for a wider variety of people (independent of age, gender, or myopia)[40].

Under Windows 10, Xmind version 11.2.2 and MindMaster version 9.0.9 were used, as they were the newest versions available in China at the time of the experiment. For post-test interviewing, EV recording software was utilized to record during the task. All software is available in Chinese.

2) Test sample images: According to Beel J's[20] research and pre-test interviews with the participants, users strive to complete the mind map in a timely and efficient manner by adding topics, changing styles, and finally completing the drawing and exporting the required file format. In contrast, mind map software features such as annotations and hyperlinks are used less frequently. A standard tiny mind map is defined as the target, which involves using high-frequency functionalities such as topics, subtopics, floating topics, relationship lines, branches, and styles. When users completed the drawing work under the usage scenario, the collected data was used to examine the usability issues. Fig. 1 shows an example of each software provided to participants as reference.

(a) Xmind Test Sample Images   (b) MindMaster Test Sample Images
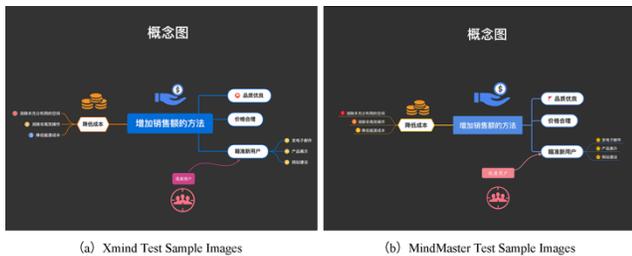
Fig. 1.   Test Sample Images (Chinese Version).

IV. EXPERIMENTAL TASKS AND PROCEDURE

By having the participants create an identical mind map using each software, the experiment aims to compare the many interaction forms of the two kinds of software in-depth and assess the usability of the entire system. The drawing process is divided into ten key operations (see Table II) to collect the critical nodes in the software more effectively and identify the essential components affecting the user experience. The entire task is finished once all ten tasks have been completed.

TABLE II.   EXPERIMENTAL TASKS

Task Number	Task content	Key Operation
Task 1	Launch the software to start a new [mind map].	Find the New entry.
Task 2	Enter [Central Topic] and add [Main Topic].	Find the add Topic entry and its interaction.
Task 3	Add [Subtopic].	Find the add Subtopic entry and its interaction.
Task 4	Add [Floating Topic] and [Relationship].	Free Topic entry and insertion of a Relationship form.
Task 5	Modify the style of [Central Topic], [Main Topic], [Subtopic] and [Branch].	Efficiency and degree of comprehension of the style adjustment area.
Task 6	Add [Mark].	Find the Mark entry.
Task 7	Add [Clipart] from the materials folder.	Clipart add form.
Task 8	Modify [Background Color].	Find the Background Color entry.
Task 9	Modify [Text] size, style, color.	Text adjustment area positioning and understanding efficiency.
Task 10	[Save] file to desktop/personal cloud and [Export] PNG.	Save and Export entry.

Within-subjects design was used in this experiment. The mind maps were created by each participant in the experiment using Xmind and MindMaster, respectively. To prevent the learning effect brought on by order of use. Group A was assigned to use Xmind first, then MindMaster, and group B vice versa.

Fig. 2 depicts the experimental procedure. The participants were briefed on the purpose and details of the experiment before it began. They were also asked to sign an experimental consent form. Pre-test interviews were done to assuage any potential uneasiness in the participants and to learn how they think about the mind map software and how it was used in their ordinary work. Then, the participants were assisted in donning the apparatus, and the machine was calibrated to guarantee the validity of the results. Following the commencement of the experiment, the participants utilized the first software and finished tasks 1 through 10 concerning the example. The participants were not required to speak aloud during the procedure to guarantee the correctness of the task completion time. The subject completed the subjective questionnaires and scales when the task was finished. Then repeat the process above using the second software after a five-minute break and recording every action on screen. After completing all tasks with the two software, participants watched the recorded video while giving a retrospective oral report to help them remember particular usability difficulties. Post-test interviews were conducted afterward to delve deeper into the underlying causes of the users' behavior.
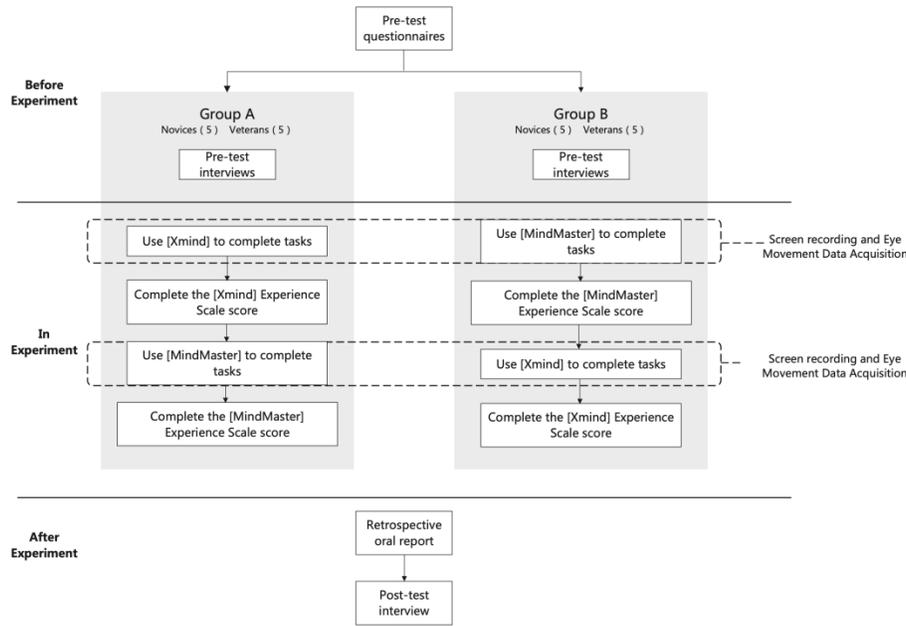
Fig. 2.   Experimental.

V.   EVALUATION METRICS

This study divides the usability index of mind mapping software into three categories based on the ISO definition: effectiveness, efficiency, and satisfaction. Data were gathered by eye tracking and subjective scales corresponding to the subdivided indicators for each category after it had been divided into smaller sub-dimensions. Table III can provide more information. Throughout the test, user behavior and interviews are recorded.

A.   Effectiveness

By counting the task completion rate of 20 participants, the software's effectiveness in attaining the stated objectives can be evaluated. However, the quantity of errors reflects how accurate it is. Additionally, we allowed the participants to seek assistance during the experiment and record the results to make the test procedure more thorough. "Task completion rate," "Number of errors," and "Number of requests for help" together provide for a more comprehensive assessment of the software's effectiveness.

B.   Efficiency

Time is an essential resource for achieving goals while using mind map software. The efficiency of both software can be measured by comparing the "Task Length." Eye movement analysis can be used to measure efficiency in addition to task completion time. The heat distribution before entering the AOI and the time spent before entering the AOI can infer the area where participants are used to finding the target and can reflect the target visibility. Scan routes can be used to compare user interfaces. The search behavior is less effective the more extended the scan path[31]. The duration of the first glance at the AOI and the number of visits can be used to assess the participant's comprehension of the specific area of interest and the ease of operation.

TABLE III.   METRICS FOR EVALUATING USABILITY AND DATA-GATHERING TECHNIQUES

	Evaluation Metrics	Date gathering techniques	
Effectiveness	Task completion rate	Percentage of completers	Experimental records
	Number of errors	Overall errors	
	Number of requests for help	Overall requests for help	
Efficiency	Task Length	Timing	Eye-movement data
	Target Visibility	Heat Map	
		Time to first fixation in AOI	
	Target identifiability	Scan Path Map	
		Duration of first look at AOI	
		Number of AOI visits	
Satisfaction	System Usability	Ease of use	SUS Scale
		Easy to learn	
	Interface Quality	Interface comfort	Subjective Likert scale
		Preference for the interface	
		Richness of interface functions	
		Simplicity of interface	
		Attractiveness of the interface	
	Perceptual Emotional Experience	Level of comfort	
		Level of enjoyment	
		Level of excitement	
		Level of clumsiness (inverse)	
		Level of frustration (inverse)	

C. Satisfaction

In this study, satisfaction was primarily measured using the SUS system usability scale[41], and Bangor[42] determined its reliability coefficient as 0.91 based on a significant sample of trials. The interface quality and user-perceived emotional experience were two additional variables that this experiment added to the user experience[43]. A total of 20 items on a 7-point Likert scale made up the questionnaire, which vehemently opposed is "1," whereas firmly in favor is "7."

D. Credibility Analysis

Measures of scale assessment include validity and reliability. The term "reliability" refers to the consistency and dependability of the results; if these characteristics are weak, the results are unstable and subject to the effects of place and time. The capacity to measure accurately is implied by validity or accuracy. The dependability quality level of the 20 question items in this experiment was measured using Cronbach's α[44]. The result of Cronbach's α was 0.919, which suggests good internal consistency and high-scale reliability. The sample size was appropriate, the KMO values for the three satisfaction-related features were 0.855, 0.739, and 0.635, all of which were higher than 0.6, and Bartlett's sphericity test, $p=0 < 0.05$, was compatible with the sphericity test.

VI. RESEARCH RESULTS

A. Effectiveness

As shown in Fig. 3 and Fig. 4, the fact that MindMaster outperforms Xmind in task completion and has much fewer errors and help requests per task indicates that it is more efficient. It can be noticed that each software has some challenges for inexperienced users by contrasting the two categories of users. Task V had the most mistakes made by the participants, with 72 made using Xmind and 26 using MindMaster. The task also included the majority of requests for assistance. This suggests that there are issues with both software when it comes to changing the drawing style, which is a crucial factor in the software's effectiveness.

B. Efficiency

The difference in efficiency between the two software can be ascertained by contrasting the two software's total and individual task lengths. Table IV shows that the overall task time of MindMaster is lower than Xmind and that this difference is significant, demonstrating that MindMaster is much more efficient than Xmind. The tasks that revealed notable variations were tasks 2, 5, and 10. For tasks 2 and 5, MindMaster performs better than Xmind. Xmind spends less time on task 10. Eye movement data for the three tasks mentioned above was examined to investigate the apparent disparities.
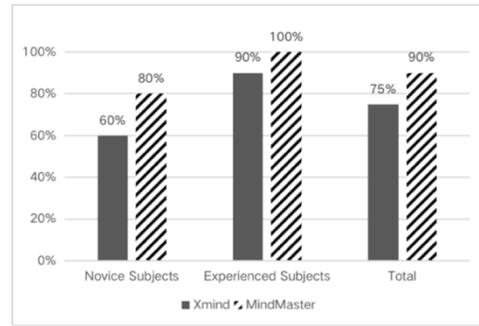
Fig. 3. Task Completion Rate of the Two Software.
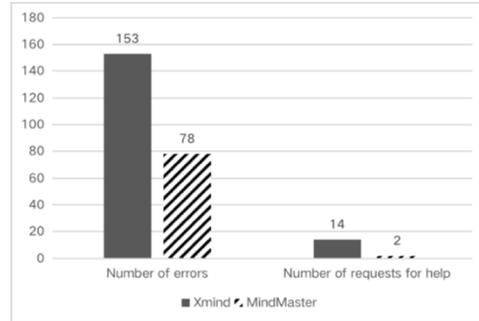
Fig. 4. The Number of Errors and Helps of the Two Software.

TABLE IV. TASK LENGTH AND THE RESULT OF PAIRED t-TEST

Items	Paired(M±SD)		Mean difference	t	Sig.
	Xmind(s)	MindMaster(s)			
Task 1	12.20±5.79	14.10±8.10	-1.89	-0.778	0.450
Task 2	112.19±37.41	88.66±45.11	23.53	2.425	0.029*
Task 3	119.87±34.12	110.15±51.27	9.71	1.000	0.334
Task 4	56.34±23.47	69.26±58.34	-12.92	-0.844	0.413
Task 5	591.76±247.08	246.51±110.79	345.25	5.902	0.000**
Task 6	68.80±36.81	55.75±21.29	13.05	1.333	0.204
Task 7	111.06±60.01	92.11±60.83	18.95	0.761	0.459
Task 8	74.24±35.22	90.45±54.71	-16.20	-1.026	0.322
Task 9	65.79±34.83	94.58±32.95	-28.79	-1.975	0.068
Task 10	59.35±20.11	94.38±31.38	-35.03	-3.700	0.002**
Total Task Length	1271.60±325.95	955.94±301.47	315.66	3.866	0.002**

*Sig.<0.05 **Sig.<0.01

1) Task 2 – add [main topic]: According to the Heat Map in Fig. 5, the participants could concentrate on the "Add Main Topic /Subtopic" choice quite well. And the majority of users also add subjects by selecting "Subtopic." As for the time to the first fixation in " Main Topic, "using Xmind takes more time than using MindMaster (according to Table V). However, it takes 14.316 s to view the "Subtopic" for the first time using MindMaster, compared to 2.283 s while using Xmind. Consequently, in this work, the target visibility of the two programs is quite close.

TABLE V. TIME TO THE FIRST FIXATION IN AOI OF TASK2

	Software	Mean	Std. Dev.	Sig.
Time to first fixation in " Main Topic "	Xmind	18.534	9.781	0.013*
	MindMaster	10.050	9.239	0.018*
Time to first fixation in " Subtopic "	Xmind	2.283	1.890	0.172
	MindMaster	14.316	6.394	0.000**

*Sig.<0.05 **Sig.<0.01

The Scan Path Maps show a few users' saccades when looking at the menu bar at the top of MindMaster and the "Adjust Style" toolbar on the right side. Although this happened just a few times, it made it harder for the subject to find the target and decreased the effectiveness of use.

(a) Xmind-Heat Map (b) Mindmaster-Heat Map
(c) Xmind-Scan Path Map (d) Mindmaster-Scan Path Map
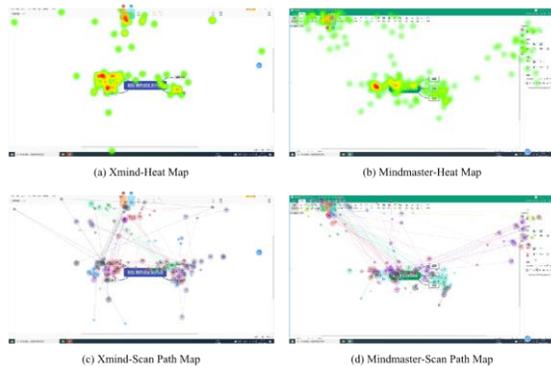
Fig. 5. Heat Map and Scan Path Map of Task2.

The "Subtopic" of Xmind was organized in the top-center toolbar, where there were fewer function options, and participants could locate the desired option more quickly. During this exercise, it was seen that the more skilled participants used shortcut keys more frequently, and these shortcut keys were consistent across both software. As a result, there wasn't much difference between how the experienced participants used the two kinds of software in this assignment. And when it comes to installing the "Subtopic," novice users typically find Xmind to be more user-friendly.

2) Task 5 – modify the style: Task 5 measures how well the two software performs regarding branching, adjusting graphics, and other features. According to Table VI, MindMaster performs better in target visibility since it takes less time for its first fixation in the AOI than Xmind. However, when using MindMaster to complete Task 5, the first stare at AOI is too long. Presumably, participants had

trouble identifying "shapes" and "branches." Division and literal description of function play a significant role. MindMaster uses only icons—no text—that are both small and strikingly identical in appearance, making it more challenging for participants to locate and recognize the target icon.

TABLE VI. TIME TO THE FIRST FIXATION IN AOI OF TASK5

	Software	Mean	Std. Dev.	Sig.
Time to first fixation in "shape"	Xmind	47.298	76.678	0.191
	MindMaster	14.476	31.291	0.121
Time to first fixation in "branch"	Xmind	21.597	29.904	0.023*
	MindMaster	18.003	13.279	0.021*
Duration of the first gaze at "shape."	Xmind	7.540	5.203	0.016*
	MindMaster	13.959	13.146	0.002**
Duration of the first gaze at "branch"	Xmind	7.910	7.633	0.034*
	MindMaster	18.366	13.107	0.000**
Number of visits to "Shape"	Xmind	10.333	6.501	0.011*
	MindMaster	14.462	15.967	0.007**
Number of "branch" visits	Xmind	20.286	16.226	0.000**
	MindMaster	9.857	8.153	0.019*

*Sig.<0.05 **Sig.<0.01

The fact that the side formatting panel was default hidden and had to be opened by selecting the "Panel" button in the top right corner was one of the primary causes of the increased time spent on Xmind. Some participants (mostly the less skilled ones) complained that they could not comprehend what the "panel" meant and that the placement of the alternatives did not correspond to their cognitive tendencies, making them challenging to identify the target.

The Heat Map (Fig. 6) also demonstrates that the participants' perspectives were primarily dispersed in the panel editing area and the right toolbar region when completing task 5 using MindMaster. However, when using Xmind, its viewpoints increasingly straddle other sections. The main challenge is finding the operation target. The mind map's branches are not free by default; you have to check the "Branch Free Layout" box in "Layout," and the user has to move the top tabs in the style panel to discover the "Layout." Users can more easily locate what to do with MindMaster because its control options are more consistent.

(a) Xmind-Heat Map (b) Mindmaster-Heat Map
(c) Xmind-Scan Path Map (d) Mindmaster-Scan Path Map
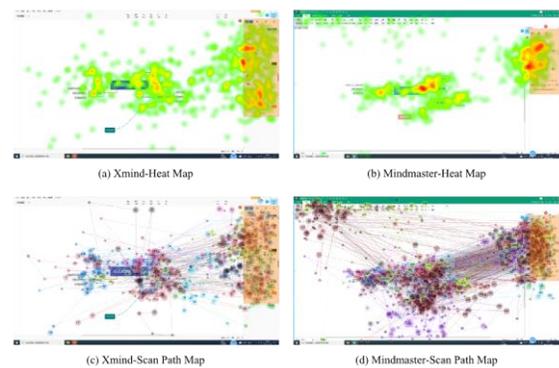
Fig. 6. Heat Map & Scan Path Map of Task5.

While MindMaster's Scan Path Map makes it clear that users will also focus on the menu bar region in the upper left corner of the interface during work, in addition to the right toolbar and the primary drawing interface, Xmind's Scan Path Map is disorganized and erratic. We attempted to analyze a specific topic to investigate the causes of this. According to the findings, novice and skilled users differ significantly, as illustrated in Fig. 7 As you can see, new users of MindMaster are accustomed to looking for functions in the menu bar area in the top left corner of the screen.
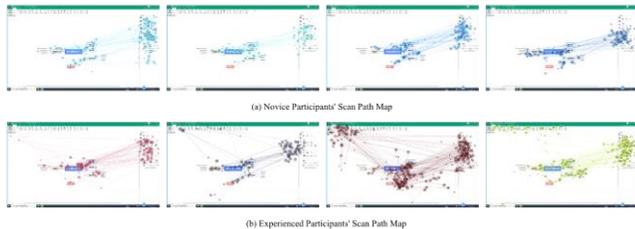
Fig. 7.   Novice & Experienced Participants Scan Path Map.

3) Task 10 – [save] and [export]: Although Xmind's interface has a wide distribution of hotspots, as seen in Fig. 8, but users use Xmind to do Task 10 significantly more quickly than MindMaster. Combined with the post-test interview, it can be found the save & export feature in Xmind uses a conventional drop-down menu is consistent with participants' habits. But it lacks innovation and does not offer features like sharing and cloud storage. In contrast, MindMaster is far more feature-rich and provides greater ease by permitting the export of files in various formats and cloud storage. However, the full-page switching format in MindMaster is pretty abrupt.

Fig. 8.   Task10 Heat Map Comparison.

According to the post-test interview, some participants expressed shock at seeing this switching format, and they initially believed they had done it by accident. Efficiency decreases since more storage alternatives lengthen the user's time to find the target.

C. Satisfaction

The ease of learning and usability scores were multiplied by 12.5 and 3.125, respectively, to match them with the SUS composite score (0～100)[41]. The amount was then multiplied by 20/7 to convert the metrics measuring user affective experience and interface satisfaction results to percentages. The computer analyses of the data from the usability scale produced the statistical findings in Table VII. As can be observed, the two software is significantly different in terms of user satisfaction. In this essay, "S" stands for the overall score, and "M" for the mean value.

TABLE VII.   COMPARISON OF USER SATISFACTION DIFFERENCES

	Variables	Xmind (n=20)	MindMaster (n=20)	t	Sig.
Usability	A1 Intention to use	3.40	3.85	-1.303	0.202
	A2 Complexity	3.55	4.40	-2.628	0.012*
	A3 Ease of operation	3.35	4.30	-2.832	0.007**
	A5 Degree of integration	3.25	4.15	-2.764	0.009**
	A6 Consistency	3.35	3.90	-1.297	0.203
	A7 Universality	3.00	3.95	-2.084	0.044*
	A8 Friendliness	3.25	4.10	-1.732	0.091
	A9 Confidence	3.35	4.60	-2.549	0.015*
	Total Usability	57.81	78.91	-2.457	0.019*
Learnability	A4 Information Support	3.50	4.40	-2.538	0.015*
	A10 Learning Cost	3.60	4.55	-1.736	0.091*
	Total ease of learning	63.75	86.88	-2.324	0.026*
Interface Quality	B1 Interface comfort	4.65	6.05	-2.278	0.028*
	B2 Preference of interface	4.35	6.00	-2.675	0.011*
	B3 Functional richness of the interface	4.45	5.85	-2.004	0.052
	B4 Simplicity of interface	5.65	5.90	-0.373	0.711
	B5 Interface Attractiveness	4.25	5.25	-1.287	0.206
	Total interface quality	66.71	83.00	-1.827	0.076
Emotional Dimensions	C1 Comfort level	4.35	6.25	-2.462	0.018*
	C2 Enjoyment level	4.05	5.95	-2.289	0.028*
	C3 Excitement level	3.80	5.20	-1.544	0.131
	C4 Clumsiness (inverse)	4.40	3.75	0.6	0.552
	C5 Frustration level (inverse)	3.00	3.25	-0.22	0.827
	Total perceived emotional experience	53.71	69.71	-3.762	0.001**
Total user satisfaction		59.81	77.74	-2.761	0.009**

*Sig.<0.05 **Sig.<0.01

The overall satisfaction rating data for the two software revealed a 0.01 level of significance (t=-2.761, Sig.=0.009), and the participants' overall satisfaction with MindMaster was considerably higher than Xmind (S_M=77.74>S_X=59.81). Less experienced participants gave MindMaster higher ratings. Some participants claimed that MindMaster's operating logic and page structure are more similar to Office, which is more recognizable and user-friendly. Moreover, MindMaster has more features that users value most, such as more drawing flexibility and support for online storage. Thus, MindMaster scores higher in these categories.

MindMaster outperformed Xmind in every category of usability, scoring significantly higher than Xmind in A2 complexity, A3 Ease of operation, A5 Degree of integration, A7 Universality, A9 Confidence, and overall usability. All participants stated that both products satisfied their usage requirements and that their willingness to use increased after becoming familiar with the product. However, MindMaster received a higher total score for usability.

The participants evaluate MindMaster's information support and learning costs over Xmind for novice users because the function's descriptions are simpler to understand. The design is similar to traditional office software, making learning easier and less expensive. Participant 18 stated, "MindMaster is simple to use, and the toolbar is identical to the standard Office software," for example.

The evaluation of the two software's interface quality also showed substantial variances. Many participants in post-test interviews praised Xmind's straightforward and consistent visual style. However, it also means that more function entries have been hidden deeper, which makes user difficult to find them. Another group of users was dissatisfied with Xmind's pages because they were too "basic." Even though the green theme color of MindMaster is visually arresting and vibrant, the 13th participant claimed that "the green appearance of the interface has an effect on the user's choice of color when drawing and makes people unconsciously want to choose a color close to the theme color of the software interface." The layout of the functional area's icons and buttons also gives MindMaster a more complicated appearance. The same results were also supported by user ratings, with MindMaster significantly outperforming Xmind in every category except for interface simplicity (M_{XB4}=5.90>M_{MB4}=5.65).

Users gave MindMaster's overall emotional experience a modest edge over Xmind in terms of perceived emotional intensity ($S_{X-sense}$=53.71<$S_{M-sense}$=69.71). The participants experienced greater comfort and enthusiasm after using MindMaster. On the other hand, Xmind made people feel even more awkward and frustrated. For instance, the 3rd participant commented, "Using Xmind was challenging, and I got frustrated." The 12th participant said, "Xmind is not very free, and the experience is not very nice."

VII. DISCUSSION

The study discovered that MindMaster performed better than Xmind in efficacy, efficiency, and satisfaction. The result is based on satisfaction questionnaires, eye-tracking data, and usability testing. This section will detail the causes and provide guidelines for software designers to adhere to.

Effectiveness-wise, a function's cognitive difficulty and location in the user interface determine whether a task can be completed successfully. The inability to modify the branch-free layout under the default "Balanced Layout" is the leading cause of task failure using Xmind. However, the "free-form" and "unconstrained structure" of mind mapping itself are its benefits[45]. This advantage is undermined by how Xmind is designed, which reduces its validity. In MindMaster, the participants can freely drag the branches to change their places, and at the same time, the branches can be intelligently adhered to the next nearest trunk. Intelligent features like this will become a trend. Simultaneously, the panel's entrance is generally difficult to discover. Thus, participants asked for help for a few times. There are two main causes for this. Firstly, the button was difficult for novices to notice in the full-screen mode because it was in the top-right corner of the screen. Second, it was challenging to understand the button's meaning due to ambiguous symbol semantic expressions and overly professional text descriptions. It suggests that designers should carefully take into account both the clear depiction of icons and the visibility of function buttons in the interface layout.

In terms of efficiency, MindMaster performs better. The interface of Xmind is simpler and conceals more functional entrances. Users' search processes become more time-consuming and ineffective due to the low visibility of functional entries. On the contrary, MindMaster directly displays some commonly used functions by default. It reduces the interaction steps for users and improves efficiency. Therefore, designers should consider the usage scenario when software is designed. And optimize the interaction design of the software's fundamental functions to match the user's behavior. The interviews and eye-tracking data show that participants are more drawn to graphical design objects than textual objects. Consumers generally are unwilling to read lengthy passages of text; therefore, using the appropriate graphic elements in the user interface can help users get the content rapidly while keeping their interest. However, it also should be carefully exercised to prevent confusing users with similar icon shapes and layouts. For instance, the MindMaster right toolbar's icons are small and neat. The branch and theme adjustments have similar layouts, identical icons, and repeating issues. As a result, the operation takes longer than it should because users have to take the time to examine the hover text descriptions. However, icons on MindMaster's top-right toolbar have auxiliary textual information and are arranged according to function. They help users to understand and use easier. So, designers can combine the relevance of functions and the level of information for interface design. And suitable auxiliary information is also important.

In terms of user satisfaction. A product's usability and learnability depend on how well it fits into the user's routines and how helpful and amiable it is to them. Xmind's entry position, iconography, and text are inconsistent with users' cognitive habits. And this is why users have low usability ratings for Xmind. On the other hand, MindMaster's top and right sides, which have a multi-entry format and take into

account the preferences of many users, are more inclusive. It makes Mindmaster's usability and learnability scores higher. Adding a multi-entry to functions does improve usability to some extent. At the same time, designers should be careful to use this method to avoid interface quality degradation. For mind map software, users choose designs that are specific and present functions in interface quality. Emotional dimensions are mainly related to the fluency of the operation process. Therefore, the designers must be mindful of the interaction logic's jumps to prevent user operations interruptions.

VIII. CONCLUSION AND FUTURE WORKS

The objective of this paper is to explore the usability and user experience of the existing mind map software. Comparative usability research is conducted to determine the key influential elements. A total of 20 participants participated in the study. According to the experiment results, the interactive logic, level of information, and interface design of mind map software all affect their effectiveness, efficiency, and user satisfaction. The study also suggests that designers must base their decisions on user needs, scenarios, and habits to create a better user experience for users of the mind map software. When evaluating the usability of their design solutions, both software design teams can benefit from the assessment techniques described in this paper.

The limitations of this paper are that the drawing examples used in the tests were slightly more complex and required using multiple functions to complete. It is somewhat different from the actual drawing scenarios of the users. This could be amended by testing as many functions of the software as possible. The future work will expand the number of experiment samples, focus on providing answers and optimize the design by merging the enumerated usability issues, followed by experiments to confirm the optimization's effects.

ACKNOWLEDGMENT

REFERENCES

[1] X. LIU, 'A Research Review on Mind Map in China', J. Sichuan Coll. Educ., vol. 25, no. 05, pp. 109-111+116, 2009.

[2] T. Buzan and D. O'Brien, Mind Map Mastery: the complete guide to learning and using the most powerful thinking tool in the universe. London: Watkins, 2018.

[3] S. Edwards and N. Cooper, 'Mind mapping as a teaching resource', Clin. Teach., vol. 7, no. 4, pp. 236–239, Dec. 2010, doi: 10.1111/j.1743-498X.2010.00395.x.

[4] Z. ZHANG and X. WANG, 'Overview of the application of mind mapping', Heilongjiang Sci, vol. 9, no. 23, pp. 28-29+32, 2018.

[5] D. Fung and T. Liang, 'The Effectiveness of Collaborative Mind Mapping in Hong Kong Primary Science Classrooms', Int. J. Sci. Math. Educ., doi: 10.1007/s10763-022-10279-1.

[6] A. YANG, 'The practice of using thinking maps to optimize elementary school composition teaching', Parents, no. 18, pp. 150–152, 2022.

[7] C. WEI, Z. WANG, Y. YANG, and H. WANG, 'Application of mind mapping in the teaching of human parasitology', Basic Clin. Med., vol. 42, no. 01, pp. 193–196, 2022, doi: 10.16352/j.issn.1001-6325.2022.01.036.

[8] O. Maksimenkova, A. Neznanov, I. Papushina, and A. Parinov, 'On Mind Maps Evaluation: A Case of an Automatic Grader Development', in Teaching and Learning in a Digital World, vol. 716, M. E. Auer, D. Guralnick, and I. Simonics, Eds. Cham: Springer International Publishing, 2018, pp. 210–221. doi: 10.1007/978-3-319-73204-6_25.

[9] 'Show Them the Money: Using Mind Mapping in the Introductory Finance Course by Ernest N. Biktimirov, Ph.D., CFA, Linda Nilson :: SSRN'. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2467958 (accessed Oct. 11, 2022).

[10] D. Thurn and J. Wolstein, 'Measurement of Motives for the Amphetamine-Type Stimulant Use on the Basis of Mind Maps', Suchttherapie, vol. 21, no. 1, pp. 43–50, Feb. 2020, doi: 10.1055/a-0858-1700.

[11] L. T. Eriksson and A. M. Hauer, 'Mind Map Marketing: A Creative Approach in Developing Marketing Skills', J. Mark. Educ., vol. 26, no. 2, pp. 174–187, Aug. 2004, doi: 10.1177/0273475304265634.

[12] S. Mehryar and S. Surminski, 'Investigating flood resilience perceptions and supporting collective decision-making through fuzzy cognitive mapping', Sci. Total Environ., vol. 837, p. 155854, Sep. 2022, doi: 10.1016/j.scitotenv.2022.155854.

[13] P. de Saqui-Sannes, R. A. Vingerhoeds, N. Damouche, E. Razafimahazo, O. Aiello, and M. Cietto, 'Mind Maps Upstream SysML v2 Diagrams', in Syscon 2022: The 16th Annual Ieee International Systems Conference (syscon), New York, 2022. doi: 10.1109/SysCon53536.2022.9773934.

[14] S. Subrahmanyam, K. Sherwani, and N. M. Pereira Ribeiro, 'Mind Mapping - A Critical Gizmo for Corporate Change', Pac. Bus. Rev. Int., vol. 14, no. 8, pp. 35–45, Feb. 2022.

[15] M. Tao and R. Xie, 'Mind Map Based Computer Network Knowledge Graph Visualization Research and Application', in Emerging Technologies for Education, Sete 2021, Cham, 2021, vol. 13089, pp. 3–12. doi: 10.1007/978-3-030-92836-0_1.

[16] G. C. Gokce and S. Aciksoz, 'Evaluation of the Relationship Between Cultural Landscape and Identity of Rural Landscape with Mind Map Method', Fresenius Environ. Bull., vol. 31, no. 7, pp. 6962–6972, 2022.

[17] W. Hang and W. Houlin, 'Produce Memorable TV Commercials based on the Memory Principles', in 2012 International Academic Conference of Art Engineering and Creative Industry (iacae 2012), Newark, 2012, vol. 37, pp. 109–114. Accessed: Oct. 11, 2022. [Online]. Available: https://www.webofscience.com/wos/woscc/full-record/WOS:000396802800025

[18] Y. Liubchenko, P. Miroshnychenko, K. Sirinyok-dolgaryova, and O. Tupakhina, 'Political Communication in the Post-Truth Era: Mind Mapping Values of Ukraine's Volodymyr Zelensky', Commun. Today, vol. 12, no. 2, pp. 146–166, Nov. 2021.

[19] D. Kuryazov, F. Schmalriede, C. Schonberg, K. Meyer, and A. Winter, 'Mind Maps sans Frontières', in 2021 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C), Fukuoka, Japan, Oct. 2021, pp. 8–15. doi: 10.1109/MODELS-C53483.2021.00010.

[20] J. Beel and S. Langer, 'An Exploratory Analysis of Mind Maps', in Doceng 2011: Proceedings of the 2011 Acm Symposium on Document Engineering, New York, 2011, pp. 81–84. Accessed: Oct. 11, 2022. [Online]. Available: https://www.webofscience.com/wos/woscc/full-record/WOS:000303789300014

[21] S. Chowdhury, M. Landoni, and F. Gibb, 'Usability and impact of digital libraries: a review', Online Inf. Rev., vol. 30, no. 6, pp. 656–680, Nov. 2006, doi: 10.1108/14684520610716153.

[22] P. Hansen, 'Evaluation of IR User interface – Implications for User Interface Design', Hum. IT J. Inf. Technol. Stud. Hum. Sci., vol. 2, no. 2, Art. no. 2, 1998, Accessed: Oct. 11, 2022. [Online]. Available: https://humanit.hb.se/article/view/249

[23] 'ISO 9241-11:2018(en), Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts'. https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en (accessed Oct. 11, 2022).

[24] W. Quesenbery, Dimensions of Usability. Lawrence: Erlbaum, 2003.

[25] J. Nielsen, Usability engineering. Boston: Academic Press, 1993.

[26] Schneiderman B, Designing the User Interface：Strategies for Effective Human－Computer Interaction, 2nd Edition. Addison－Wesley, 1992.

[27] J. Preece, Y. Rogers, H. Sharp, D. Benyon, and S. Holland, Human-Computer Interaction. UK: Addison Wesley, 1994.

[28] B. Shackel, Usability－Context, Framework, Definition, Design and Evaluation. UK: Cambridge University Press, 1991.

[29] J. Wang, 'A Literature Review of Progress in Foreign Usability Research', New Technol Libr Inf Serv, no. 9, pp. 7–16, 2009.

[30] J. Jeng, 'What Is Usability in the Context of the Digital Library and How Can It Be Measured', Inf. Technol. Libr., vol. 24, no. 2, p. 47, Jun. 2005, doi: 10.6017/ital.v24i2.3365.

[31] G. ZHANG, M. SHEN, and R. TAO, 'The Application of Eye Tracking in Usability Tests', Chin. Ergon., no. 04, pp. 9-14+70-71, 2001, doi: 10.13837/j.issn.1006-8309.2001.04.003.

[32] J. Leplat, 'Handbook of usability testing: How to plan, design and conduct effective tests - Rubin,J', Trav Hum, vol. 60, no. 2, pp. 223–223, Jun. 1997.

[33] M. F. Paul, E. J. Richard, and L. M. John, 'The Movements of Aircraft Pilots During Instrument-Landing Approaches', Aeronaut. Eng. Rev. 9, 1950.

[34] B. Aga, 'Measuring the Effects of Drug Label Design and Similarity on Pharmacists'Performance', in Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics, Morgan Kaufman, 2008, pp. 271–280.

[35] B. Aga, Eye Tracking the User Experience. Beijing: POSTS & TELECOM PRESS, 2019.

[36] S. M. Roa-Martínez and S. A. B. G. Vidotti, 'Eye tracking y usabilidad en ambientes informacionales digitales: revisión teórica y propuesta de procedimiento de evaluación', Transinformação, vol. 32, p. e190067, 2020, doi: 10.1590/1678-9865202032e190067.

[37] An introduction to human factors engineering | Open University Malaysia Digital Library Portal. Accessed: Oct. 11, 2022. [Online]. Available: https://library.oum.edu.my/oumlib/content/catalog/597121

[38] J. Nielsen and R. L. Mack, Heuristic Evaluation. New York: John Wiley, 1994.

[39] J. H. Spyridakis, 'Conducting Research in Technical Communication: The Application of True Experimental Design', Tech Commun, vol. 39, no. 4, pp. 607–624, 1992.

[40] Y. WANG and F. LV, 'Usability Testing of ATM Machines Interface based on Eye Tracking Data', Chin J Ergon, vol. 23, no. 01, pp. 48–54, 2017, doi: 10/gppx22.

[41] J. Brooke, 'SUS: a retrospective', Usability Stud, vol. 8, no. 2, pp. 29–40, 2013.

[42] Bangor, A., Kortum, P. T., and Miller, J. T, 'Determining what individual SUS scores mean:Adding an adjective rating scale', J. Usability Stud., vol. 4, no. 3, pp. 114–123, 2009.

[43] H. Han and Y. Li, 'The Interaction Design in Mobile Learning Platforms:A Comparative Study on Usability and User Experience', J Mod Inf, vol. 41, no. 04, pp. 55–68, 2021.

[44] S. Wu, SPSS and Statistical Thinking. Beijing: Tsinghua University Press, 2019.

[45] M. Davies, 'Concept mapping, mind mapping and argument mapping: what are the differences and do they matter?', High. Educ., vol. 62, no. 3, pp. 279–301, Sep. 2011, doi: 10.1007/s10734-010-9387-6.

Character Level Segmentation and Recognition using CNN Followed Random Forest Classifier for NPR System

U. Ganesh Naidu[1]

Research Scholar, Department of
Computer Science and Engineering
Annamalai University
Annamalai Nagar, Tamil Nadu, India

Dr. R. Thiruvengatanadhan[2]
Dr. P. Dhanalakshmi[4]

Assistant Professor, Department of
Computer Science and Engineering
Annamalai University, Annamalai
Nagar, Tamil Nadu, India[2, 4]

Dr. S. Narayana[3]

Professor, Department of Computer
Science and Engineering
Gudlavalleru Engineering College
Gudlavalleru, Andhra Pradesh, India

Abstract—**The number plate recognition system must be able to quickly and accurately identify the plate in both low and noisy lighting conditions, as well as within the specified time limit. This study proposes automated authentication, which would minimize security and individual workload while eliminating the requirement for human credential verification. The four processes that follow the acquisition of an image are pre-processing, number plate localization, character segmentation, and character identification. A human error during the affirmation and the enrolling process is a distinct possibility since this is a manual approach. Personnel at the selected location may find it difficult and time-consuming to register and compose information manually. Due to the printed edition design, it is impossible to communicate the information. Character segmentation breaks down the number plate region into individual characters, and character recognition detects the optical characters. Our approach was tested using genuine license plate images under various environmental circumstances and achieved overall recognition accuracy of 91.54% with a single license plate in an average duration of 2.63 seconds.**

Keywords—Character segmentation; convolutional neural networks; bilateral filter; character recognition; SVM classifier

I. INTRODUCTION

An automatic number plate recognition (ANPR) system identifies the vehicle's license plate without a person's involvement. Image acquisition and character retrieval are the two primary operations of ANPR systems. During this step, images of the vehicles are acquired. These approaches are used to process images of number plates and extract their characters. Surveillance cameras are often used in ANPR systems. Due to the fact that they are used for broad purposes, these cameras tend to capture larger-than-average images. The first step is to find the number plate area in the image and extract it. Afterward, the number plate's alphanumeric characters need to be deciphered from the backdrop. Then deep learning-based methods are used to identify the characters. The ANPR system comprises different phases as seen in Fig. 1: (1) image acquisition (2) number plate identification and extraction (3) character segmentation and (4) character recognition [1]. The first two phases identify and capture an image of the vehicle. Next, a number plate's

possible locations in an image are identified. The identified area is used to isolate the characters in the third phase. Finally, the characters are identified in the last stage. This paper focuses on the character isolation phase. Therefore, this study does not include vehicle detection techniques. The contribution of this paper is to improve character recognition accuracy.
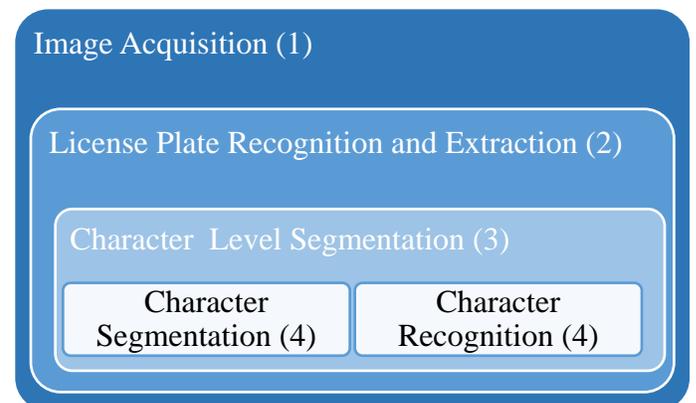
Fig. 1. Different Phases Involved in the ANPR System.

The paper is structured in the following manner. Section II gives an overview of the ANPR systems and techniques. Section III outlines the nature of the issue and the difficulties. The proposed methodology is narrated in Section IV. Section V details the results of the experiments and evaluations of the performance. Section VI concludes the paper.

II. RELATED WORKS

The convolutional neural network (CNN) has a state-of-the-art accuracy in HCR [2]. The performance of the recurrent neural network (RNN) is improved in [3]. In [4], presented a novel technique by combining the conventional RNN and deep CNN. HCR technology has been researched for a while and is used by the industry, but its poor accuracy hinders its usability and overall performance. The current character recognition systems aren't very reliable and require additional research to be used widely.

The YOLOv3 algorithm and background subtraction are suggested in [5] for the identification of vehicles and the localization of license plates. In [6], developed a system for optical character recognition to identify stolen vehicle license plates. CNN is used in [7] to develop an embedded system that can identify and recognize Brazilian license plates. Finally, the system uses a Tiny YOLOv3 architecture and a NN is trained on synthetic images and fine-tuned on actual license plate images to identify the license plate characters.

A. Character Segmentation

Using k-means clustering to discover linked pixel areas and combining appropriate pixels into components to efficiently extract each character, a method known as connected components labelling analysis (CCLA) was described in [8] for segmenting license plate (LP) characters. In [9], created a unique sharpness-based methodology for segmenting the characters in LP images. In the segmentation procedure, the model faced a gradient vector and a lack of accuracy. In [10] used a successful model for LP detection on a variety of lighting systems. Binarization and the superpixel paradigm were used in conjunction with this model to segment the characters in LP. LP recognition is expected to be implemented using edge detection and CNN [11] [12].

Prediction and identification of LP vehicles based on character features were utilized by [13] [14]. Segmenting characters from LP areas was initially the model's first step toward refinement. There were both vertical and horizontal projection elements in the approach for character segmentation. For images with challenging backgrounds, the features given may not work. In [15] developed a technique for character segmentation using a binarized input image without character shapes or presence among the characters improves the performance quite a little. It is thus exceedingly difficult to use a binarization model that distinguishes foreground and background data in images consisting of complicated backgrounds. Finally, it was discovered that a greater variety of ways had been explored in an effort to address the problems caused by the reduced light impacts on the environment. However, it does not include issues such as blur, touch, and challenging backdrops.

B. Character Recognition

Riesz fractional-centric detection and identification of LPs was proposed in [16]. This method is used to explain why LPs are difficult to find and identify. An enhancement in LP images may lead to an increase in recognition outcomes, which is not ideal for real-time environments, according to the experimental results. For low-quality images, [17] used an ensemble of Adaboost cascades for LP recognition. Classification models for LP analysis from images impacted by different variables are used in this model to identify that the texture assigned relies on the LBP job completed by the user. Consequently, this technique's primary job is to learn from and count examples of simulated situations. Text prediction has also been constrained in terms of value, although the established technique still allows for recognition. Due to the fact that the detection procedure does not acquire whole character forms, text detection is simpler than text recognition. In recent years, there is more development of numerous DL techniques for the identification of low-level features, such as LPs. In [18], suggesting a CNN-based approach to LP identification. LP analysis has been performed by using R-CNN. In [19], developed a deep localization and error detection technique for LP detection. In [20], used kernel-based Extreme Learning Machines (ELM) with significant attributes to develop Chinese vehicle LP identification. Once this was established for LP identification, the researchers looked at how CNN and ELM might be combined to improve performance. Deep learning (DL) modules that function effectively in the presence of a large number of predefined samples were found to have these characteristics. However, images impacted by a number of negative circumstances make it difficult to identify a predetermined instance that demonstrates plausible distinctions from LP recognition. Additionally, the DL approach is hampered by inadequacies like as parameter optimization for various databases and the ability to retain the dependability of DNN in the face of these issues [21]. The established criteria show that there is a discrepancy between the previous models and the more current demands. This discovery prompted the development of a new method for LP identification that did not rely on classification models and counted more labelled samples than the prior techniques. Based on the genetic algorithm and an improved neutrosophic set (NS), a new LP recognition model has been developed in [22].

III. CHALLENGES IN ANPR

A. A Lack of Standardized Number Plates

Various nations and areas within the same country have different number plates [23]. Number plates are not standardized. Several factors, such as plate sizes, plate backdrops, character sizes, and plate textures, all contribute to the variety of number plate forms [24]. Intelligent algorithms [25] are necessary for improved number plate recognition. It has been suggested that a number of strategies may be used to enhance the system. As a result, this study is a difficult one that is limited to a small area.

B. Erroneous Character Interpretations

The characters such as "O - 0, I - 1, B - 8, C - G, A - 4, D - 0, D - O, G - 6, and 2 –Z" are close enough that character recognizers may get them mixed up. All of these flaws should be addressed by character recognition algorithms [1, 10, 20]. In spite of the increased emphasis given to character recognition in ANPR systems, the ambiguous character problem remains a major issue. Furthermore, the issue of ambiguous characters has not been examined in depth.

In addition, segmenting the characters, which is the third stage in the ANPR process, is one of the most difficult. There are a variety of techniques used to categorize the characters of various license plates in the literature. Optical character recognition systems for letters and numbers may be employed separately to address the issue posed by ambiguous characters. This will improve the accuracy rate of the character identification phase since it will eliminate the ambiguous character issue.

In this paper, an image processing-based approach is described to address the issues outlined above. In the character recognition stage, we may employ various models specifically built for letters and numbers provided we accurately segregate the areas. The first step is to determine whether a character is a letter or number before we can do anything further. According to our knowledge, no research has looked at identifying the character's kind (letter, number).

IV. METHODOLOGY

Fig. 2 gives the workflow of the proposed license plate recognition, which will be explored in more detail below. This method includes different phases. The four processes that follow the acquisition of an image are pre-processing, number plate extraction, character segmentation, and character identification. Character level segmentation is performed using CNN and character recognition is done by using a Random forest classifier. The proposed method would expedite registration and provide additional advantages such as vehicle security and traffic control.
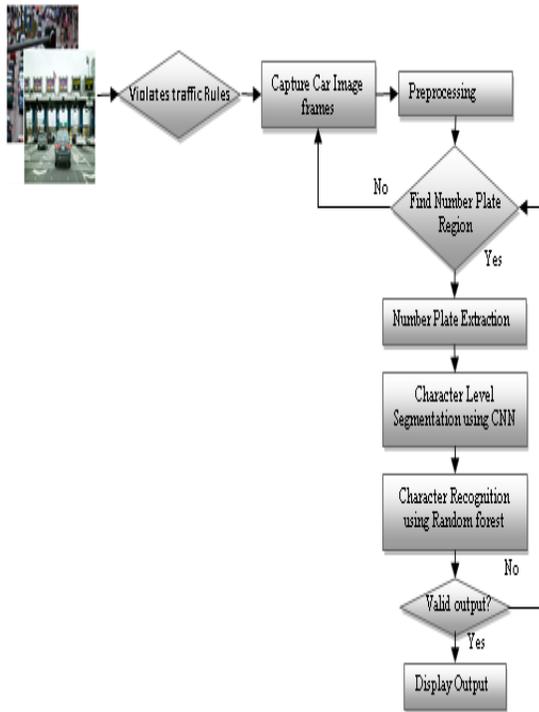
Fig. 2. Flow Diagram of the Proposed License Plate Recognition.

A. Pre-Processing

Input images face a myriad of issues, including interference, deformation, underexposure, and other. These effects may be minimized by doing image preprocessing on the input, which simplifies and speeds up the processing of images. In the pre-processing step, first, the input images are converted into grayscale images. The reason behind this is, standard number plates come in white and black, which are the only two colors available. As a result, it's not necessary to include all of the colors in the image.

1) Bilateral filtering: Noise reduction is made possible with the employment of a bilateral filter. To remove noise while maintaining the sharpness of characters, a kernel size of 3x3 is highly recommended. Filtering is the process of replacing each pixel value in an image with the average weighted sum of nearby pixels. The number of pixels in the weighted average is determined by the kernel size. Because larger kernels take longer to process and smaller kernels are useless at removing noise, a kernel size of three is the most efficient in this paper for decreasing noise.

2) Contrast enhancement: When the lighting is dim, boosting the contrast is a necessary step in number plate recognition. In order to better distinguish between the black text and the white background on the number plate, the contrast between the two colors must be increased. CLASH is used to increase contrast. Using this method, the image is broken up into smaller units known as tiles. Then histogram equalization is performed on each 8x8 tile separately. After contrast enhancement, the image can be converted into binary. When converting an image in to a binary image, the pixel values are reduced to two values [0 and 255].

3) Dilation of image: Dilating a number plate's characters makes them more prominent, which in turn makes their edges more prominent. When an object's area rises, so does the image's overall perimeter (edges). The number plate may be more precisely located if the character's borders are thicker, as edge-detection is used to do so. A number plate frame with no extra edges avoids the false edges from being picked up by the camera.

B. Localization of Number Plate

By recognizing the image's boundaries, edge detection can be utilized to determine the location of a number plate. The number plate region, which is actually the number plate characters themselves, has a lot of edges. This is because most of the number plates have black writing on a white background. As a result, the edges of the characters stand out against the background. Number plates tend to be densely packed with letters and numbers since it is where localized edges are found to be the most numerous. The number plate has the highest concentration of edges in the picture, despite the fact that the edges are visible across the image. Through the use of pre-processing, the contours of these objects become clearer. Because of this, the number plate is shown by a bounding rectangle. The picture is visited to count the number of edges it covers. The number plate is located in the portion of the image that has the most localized number of edges in the enclosing rectangle. Number plate locator software used an ingenious edge-detecting mechanism. Because it is more accurate, Canny edge-detection can pick up on both horizontal and vertical edges. When a car is in motion, the horizontally structured rear windshield usually sits immediately below the license plate just above it. Similarly, the bumper just under the license plate has a structure that is almost horizontal as well.

C. Character Segmentation using CNN_CL4 Network

The characters of the number plate are retrieved using a localized number plate as input. CNN has garnered a lot of attention as one of the most successful deep-learning models

for picture segmentation. It's employed in a slew of new applications, including medical picture categorization, vehicle monitoring, self-driving vehicles, and face recognition, among others. Because of the self-optimization feature of artificial neurons, CNNs are able to learn just like brain neurons. Other algorithms cannot extract characteristics and segment images as accurately as it does because of this self-optimizing nature. The input data has to be processed minimally, yet it produces very accurate and exact outputs. HyperOn the basis of these qualities, CNN attempts to comprehend and distinguish between the images. Low-level information such as edges, gradient direction, or color may be captured by the first few layers of convolutional neural networks (CNNs). However, as the number of convolutional layers increases, it begins to extract characteristics at a higher level. Higher dimensions and convolution lead to exponentially increasing network parameters. As a result, CNN computations are slowed. This is because the number plate images have a lower pixel density than the images that are given to the input layers.

The CNN model for character-level segmentation is given in Fig. 3. Layers of the convolutional network are fed images from the input layer. Because it only contains four convolutional layers, the model is small and quick to run on a computer. Convolutional layer with 3 x 3 kernels and ReLU activation function in the first layer. 2D convolutional layer in the second layer. There are several deep learning algorithms that make heavy use of the ReLU activation function. In contrast to other activation functions, such as tanh, neurons are only partially stimulated when using ReLU. Except for zero, the continuous and differentiable function ReLU is a piecewise linear one. Additionally, it has a lower probability of a vanishing gradient than other methods because of its simplicity and empirical ease. Negative values can be simply returned as zero, which is shown in Fig. 4 by ReLU's basic concept of returning positive input values to output. A max-pooling layer and ReLU-activation function follow the first three 2D convolutional layers. To reduce the size of our input images, a sample-based discretization max-pooling is used. It helps to minimize the network's dimensionality by pooling the maximum value from each feature map patch. Over fitting may be avoided by removing non-essential factors and thereby reducing the overall number of parameters. To avoid overfitting the initial convolutional layer, a 2 x 2 max-pooling layer is added to all except the first layer. A 1D string is formed as a result of the fourth convolutional layer, which is then passed on to the final fully connected layer. The activation units in the next layer are linked to all neurons in the completely connected layer. The first layer's neurons are all linked to the activation unit of the second layer's neurons in a two-layer model. All inputs are then passed to the Soft max activation algorithm, which segments the characters on the number plate. Finally, the segmented characters are given to the SVM classifier to classify the numbers and the alphabet.

D. Character Recognition

The maximal margin classifier and random forest (RF) classifier have been utilized to categorize these segmented characters into respective digits and alphabets, and their performance has been evaluated.

1) Support Vector Machine: SVM is a maximal margin classifier. By constructing hyperplane, it categorizes the data. It works better for categorizing binary classes. It is also used for solving multiclass issues. Figure shows the maximum margin hyper plane of the support vector machine. There are two types of SVM classifications are there name one-vs-one (OvO) and one-vs-rest (OvR).

In this paper, a polynomial kernel of degree three is used. The magnitude of the allowable misclassification is determined by parameter 'C.' Larger values of 'C' will choose a hyperplane with a narrower margin. Smaller values of 'C', on the other hand, compel the classifier to search for a bigger margin, even if the hyperplane misclassifies the points. In general, a high 'C' value is desired, although it may lead to over fitting [2]. The shape of the class dividing hyperplanes affected by the parameter γ. When γ is high, only points near the hyperplane are considered, while a lesser value of γ considers points far away from the hyperoverplane. In general, a low value of γ is preferable, since a higher number might result in overfitting. The low value of kernel width parameter is selected with greater value of optimal cost parameter. Table I lists all of the critical parameters [5].
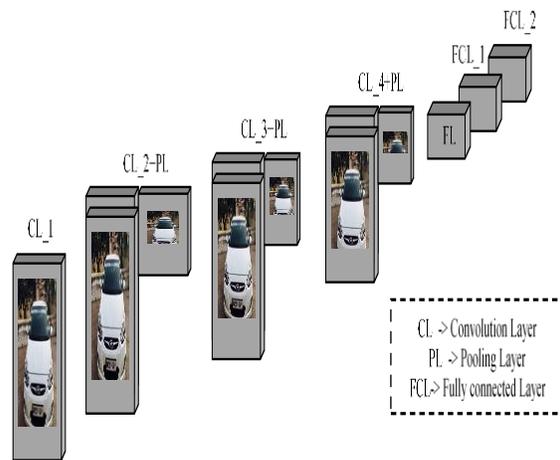
Fig. 3.   CNN Model for Character Level Segmentation.
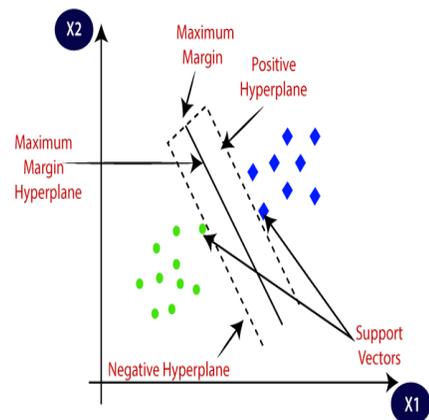
Fig. 4.   Support Vector Machine.

TABLE I.   PARAMETERS OF SUPPORT VECTOR MACHINE

Classifier	SVM
Approach	One-versus-all
Cost parameter (C)	1
Kernel width parameter (γ)	Auto-deprecated
Degree of polynomial kernel	3

2) Random Forest (RF) Classifier: It is based on an ensemble of decision trees. Nodes make up a tree, and each node makes a decision based on a parameter. It has been trained using a bagging technique. Random forest employs a slightly different method to bagging, in which a subset of characteristics is chosen for node split, while bagging utilizes all features. As a consequence, an RF is a collection of trees that lowers the influence of noise in a single tree. As a consequence, bagging improves the overall outcome. Two key parameters must be set: Two variables are the number of features in each split (F_s) and the number of decision trees in the forest (N_t). Although the big value of parameter N_t is maybe superfluous, it has no negative impact on the model. It will almost certainly improve the accuracy of the forecasts, but it may slow down the model. The number of features to take into account while dividing a node is specified by the parameter F_s. It is always a subset of the whole amount of characteristics. The value for N_t was set to 100, and the value for F_s is set to the square root of the model's number of features.

V.   RESULTS AND DISCUSSIONS

The characters from the number plate are extracted using character segmentation, which is subsequently used to train different classification models. All segmented characters for training have been reduced to 20 x 20 pixels. SVM and Random Forest classifiers are used in this study. The performance is also compared with other existing methodologies such as KNN, and Neural networks. The simulation results are given in Fig. 5. Initially, the number plate region is extracted from the car, and the extracted region is given in Fig. 5(a). Then the images are given to preprocessing stage where the input images are preprocessed, enhanced and finally converted to a binary image and it is shown in Fig. 5(b). The bounding box for character segmentation is given in Fig. 5(c). Finally, the recognized number plate is given in Fig. 5(d).The comparative analysis on the recognition accuracy of different methods are tabulated in Table II. The performance of four different classifiers such as Random Forest Classifier, SVM Classifier, KNN and Neural network has been analyzed. From the analysis, we can clearly say that, the random forest classifier gives better recognition accuracy of 91.54% whereas, the lowest recognition accuracy is given by the neural network classifier.

A. Comparative Analysis

The graphical representation on the comparative analysis of different classification methods are given in Fig. 6. The graphical analysis shows that the random forest classifier performs well when compared to other techniques.

| (a) | (b) | (c) | (d) |

Fig. 5.   (a) Original Image (b) Binary Image (c) Boundary Box for Character Segmentation (d) Recognized Number Plate Value.

TABLE II.   COMPARATIVE ANALYSIS

Classifier	Recognition Accuracy
Random Forest Classifier	91.54
SVM Classifier	89.47
Neural Network	86.82
KNN	82.19
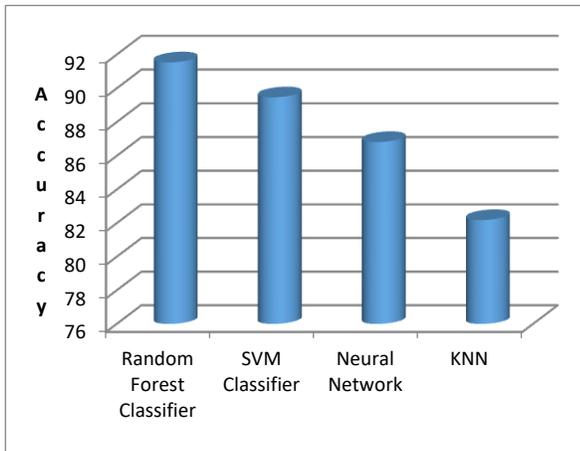
Fig. 6.   The Graphical Representation of Comparative Analysis of Different Methods.

For a variety of reasons, RF outperforms SVM in this scenario. SVM is a binary classifier, but random forests are inherently multiclass classifiers. It is necessary to convert the multiclass issue to a multiple binary class problem for SVM to operate. Despite this, random forest beats SVM since it is a multiclass classifier. The second problem is that pictures, no matter how much preprocessing is done, will always be noisy. In this instance, the noise-resistant classifier is expected to perform better. There is evidence to support this notion since RF outperforms SVM. Consensus results from several trees are used to form RF's overall classification. As a consequence, even if some trees are trained using noisy input, the end outcome should still be as predicted. While SVM is notoriously slow to train, using RF you can train several models at the same time, which is something you can only accomplish with SVM. Neuronal networks can handle noise, thus they are accurate. There is no substitute for a random forest, which is an ensemble of decision trees based on a kind of bagging strategy, in terms of accuracy.

VI.   CONCLUSION

The four stages: Preprocessing, number plate localization, character level segmentation, and character recognition are the processes discussed in this study. Converting the RGB image to grayscale, using the bilateral filter to reduce noise, improving the image's contrast, translating the image to a binary image, and finally distorting the image are all part of the preprocessing stage. Character segmentation begins by segmenting the characters from the number plate that have been accomplished using CNN after removing an unnecessary area of the number plate, The first step in character segmentation is to separate the characters from the number plate, which is done using CNN. Individual characters are recognized using character recognition utilizing random forest classifier and support vector classifier. The proposed method achieved an overall recognition accuracy of 91.54% with a

single license plate in an average duration of 2.63 seconds. This method is also compared to state-of-the-art methods, demonstrating that the suggested method outperforms other existing methods.

REFERENCES

[1] Nazmus Saqib, Khandaker Foysal Haque, Venkata Prasanth Yanambaka and Ahmed Abdelgawad, " Convolutiona-Neural-Network-Based Handwritten Character Recognition: An Approach with Massive Multisource Data, Agorithms, 15,129,pp.1-25, 2022.

[2] Tabik, S.; Alvear-Sandoval, R.F.; Ruiz, M.M.; Sancho-Gómez, J.L.; Figueiras-Vidal, A.R.; Herrera, F. MNIST-NET10: A heterogeneous deep networks fusion based on the degree of certainty to reach 0.1% error rate. ensembles overview and proposal. Inf. Fusion 2020, 62, 73–80.

[3] Pham, V.; Bluche, T.; Kermorvant, C.; Louradour, J. Dropout Improves Recurrent Neural Networks for Handwriting Recognition. In Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), Heraklion, Greece, 1–4 September 2014; pp. 285–290.

[4] Shi, B.; Bai, X.; Yao, C. An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 2298–2304.

[5] Jadhav, Atharva V., Omkar K. Dongre, Tanmay K. Shinde, Deepak S. Patil, And Jaya H. Dewan. "A Study On Approaches For Automatic Number Plate Recognition (ANPR) Systems." In Advances In Data And Information Sciences, Pp. 647-658. Springer, Singapore, 2022.

[6] Luna, A.C., Trajano, C., So, J.P., Pascua, N.J., Magpantay, A. and Ambat, S., 2022. License Plate Recognition for Stolen Vehicles Using Optical Character Recognition. In *ICT Analysis and Applications* (pp. 575-583). Springer, Singapore.

[7] Izidio, D.M.F., Ferreira, A.P.A., Medeiros, H.R. *et al.* An embedded automatic license plate recognition system using deep learning. *Des Autom Embed Syst* 24, 23–43 (2020). https://doi.org/10.1007/s10617-019-09230-5.

[8] Lakshmana Phaneendra Maguluri, Efficient Smart energy Response Systems for Fire Hazards using IOT. International Journal of Advanced Computer Science and Applications, Vol. 9, No. 1, 2018, 314-320.

[9] V. Khare, P. Shivakumara, P. Raveendran, L. K. Meng, and H. H. Woon, ``A new sharpness based approach for character segmentation in license plate images,'' in *Proc. 3rd IAPR Asian Conf. Pattern Recognit. (ACPR)*, Nov. 2015, pp. 544_548.

[10] D. Kim, T. Song, Y. Lee, and H. Ko, ``Effective character segmentation for license plate recognition under illumination changing environment,'' in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2016, pp. 532_533.

[11] P. Dhar, S. Guha, T. Biswas, and M. Z. Abedin, ``A system design for license plate recognition by using edge detection and convolution neural network,'' in *Proc. Int. Conf. Comput., Commun., Chem., Mater. Electron. Eng. (IC4ME2)*, Feb. 2018, pp. 1-4.

[12] Lakshmana Phaneendra Maguluri1 , R. Ragupathy A Cluster based Non-Linear Regression Framework for Periodic Multi-Stock Trend Prediction on Real Time Stock Market Data, International Journal of Advanced Computer Science and Applications, Vol. 11, No. 9, 2020.

[13] S. K. Ingole and S. B. Gundre, ``Characters feature-based Indian vehicle license plate detection and recognition,'' in *Proc. I2C2*, 2017, pp. 1-5.

[14] Haengwoo Lee Namseoul University, South Korea "Analysis on performances of the optimization algorithms in CNN speech noise attenuator" Indonesian Journal of Electrical Engineering and Informatics (IJEEI) Vol. 9, No. 4, December 2021, pp. 827~833 ISSN: 2089-3272, DOI: 10.52549/ijeei.v9i4.3245.

[15] A. Radchenko, R. Zarovsky, and V. Kazymyr, ``Method of segmentation and recognition of Ukrainian license plates,'' in *Proc. IEEE Int. Young Sci. Forum Appl. Phys. Eng. (YSF)*, Oct. 2017, pp. 62-65.

[16] K. S. Raghunandan, P. Shivakumara, H. A. Jalab, R. W. Ibrahim, G. H. Kumar, U. Pal, and T. Lu, ``Riesz fractional based model for enhancing license plate detection and recognition,'' *IEEE Trans. Circuits Syst.*

Video Technol., vol. 28, no. 9, pp. 2276_2288, Sep. 2018.

[17] Maganti Syamala, N.J.Nalini An Efficient Aspect based Sentiment Analysis Model by the Hybrid Fusion of Speech and Text Aspects, Vol. 12, No. 9, 2021, 160-169.

[18] M. Dong, D. He, C. Luo, D. Liu, and W. Zeng, ``A CNN-based approach for automatic license plate recognition in the wild,'' in *Proc. Brit. Mach. Vis. Conf.*, 2017, pp. 1-12.

[19] O. Bulan, V. Kozitsky, P. Ramesh, and M. Shreve, ``Segmentation- and annotation-free license plate recognition with deep localization and failure identi_cation,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2351_2363, Sep. 2017.

[20] Nurnajmin Qasrina Ann1 , Dwi Pebrianti1,4 , Mohammad Fadhil Abas2 , Luhur Bayuaji3,4, Mohammad Syafrullah4 "Parameter Prediction for Lorenz Attractor by using Deep Neural Network" Indonesian Journal of Electrical Engineering and Informatics (IJEEI) Vol. 8, No. 3, September 2020, pp. 532~540 ISSN: 2089-3272, DOI: 10.11591/ijeei.v8i3.1272.

[21] Y. Yang, D. Li, and Z. Duan, ``Chinese vehicle license plate recognition using kernel-based extreme learning machine with deep convolutional features,'' *IET Intell. Transp. Syst.*, vol. 12, no. 3, pp. 213-219, Apr. 2018.

[22] B. B. Yousif, M. M. Ata, N. Fawzy, and M. Obaya, ``Toward an optimized neutrosophic K-means with genetic algorithm for automatic vehicle license plate recognition (ONKM-AVLPR),'' *IEEE Access*, vol. 8, pp. 49285-49312, 2020.

[23] Akhtar, Z., Rashid, A. Automatic number plate recognition using random forest classifier. SN Computer Science 2020, 1.3, 1-9.

[24] Dalarmelina, N. D. V., Teixeira, M. A., Meneguette, R. I. A real-time automatic plate recognition system based on optical character recognition and wireless sensor networks for ITS. Sensors 2020, 20(1), 55.

[25] Mahalakshmi, S., Tejaswini, S. Study of character recognition methods in automatic license plate recognition (ALPR) system. International Research Journal of Engineering and Technology 2017, 4, 1420-1426.

Prediction of Micro Vascular and Macro Vascular Complications in Type-2 Diabetic Patients using Machine Learning Techniques

Bandi Vamsi[1]
Department of Computer Science
Artificial Intelligence and Data Science
Madanapalle Institute of Technology and Science
Andhra Pradesh - 517325, Madanapalle, India

Ali Al Bataineh[2]
Department of Electrical
and Computer Engineering
Norwich University
VT 05663, United States

Bhanu Prakash Doppala[3]
Lead Instructor, Data Analytics
Academy Xi
Sydney NSW 2000, Australia

Abstract—A collection of metabolic conditions known as diabetes mellitus (DM) is defined by hyperglycemia brought on by deficiencies in insulin secretion, action, or both. In terms of mortality rate, type-2 diabetes is 20 times higher when compared with type-1. Based on the earlier research, there is still scope to identify different risk levels of type-2 diabetes complications. To achieve this, we have proposed a T2DC machine learning-based prediction system using a decision tree as a base estimator with random forest to identify the severity of T2-DM complications at an early stage. Our proposed model achieved accuracies of 95.43%, 94.62%, 96.25%, 97.55%, and 97.83% for Nephropathy, Neuropathy, Retinopathy, Cardio Vascular and Peripheral Vascular complications in T2-DM patients. The proposed model has the potential to improve clinical outcomes by promoting the delivery of early and personalized care to T2-DM patients.

Keywords—*Diabetes mellitus; micro vascular; macro vascular; machine learning; type-2 complications*

I. INTRODUCTION

One of the most widespread health problems affecting all age groups is diabetes mellitus (DM). According to WHO (World Health Organization) statistics, nearly 180 million people worldwide have type 2 diabetes mellitus (T2-DM), with 95 percent having DM in this structure [1]. The number of people having this T2-DM is estimated to rise drastically by 2030. As per the highest cases recorded in the world, India ranked in 2nd place with 60 million DM records and is estimated to rise by 109 million people with DM by 2035 [2]. It is a condition that is identified when the pancreas fails to produce insulin in the required amount needed for the body, or due to damage to tissues and cells in the human body. T2-DM is a condition that is strongly connected to both micro vascular (MIV) and macro vascular (MAV) problems, which include nephropathy, retinopathy, neuropathy (MIV), and peripheral vascular disease (MAV), contributing to the effect on internal organs and blood vessel-related complications [3]. DM can be identified in three different forms, namely: type 1 diabetes mellitus (T1-DM) affects the pancreas by producing insulin in a lower amount than needed by the entire human body [4]. To keep the body's insulin level at the right level, it needs supplements from outside the body. T2-DM is a condition characterized by a significant increase in blood sugar levels. In this structure of DM, the insulin levels are disturbed, and the body fails to utilize and produce [5]. It is a type of

hormone developed in the pancreas that aids in maintaining sugar levels in the blood. In particular, this hormone maintains the amount of glucose that flows through the cells. In general, after the consumption of food, blood sugar levels in the blood are identified in the high range [6]. The extra glucose in the blood is transferred into the cells when the pancreas secretes insulin, which diminishes the quantity of glucose in the bloodstream [7]. When this abnormality is not identified at an early stage using proper diagnosis, this T2-DM can lead to severe chronic health disorders. Chronic hyperglycemia, a side effect of diabetes, can damage, weaken, or kill many organs over time, especially the kidneys, eyes, heart, nerves, and blood vessels [8].

A. MIV and MAV Complications

Based on the research on DM, it is evident that the following are the damaging factors for human health:

- latent loss of eyesight with Retinopathy (RET).

- fluid accumulation causes Nephropathy (NEP), which in turn leads to hyperglycemia.

- Neuropathy (NEU) is a condition that affects the supply of blood by causing late healing in the functioning of nerves, shortening the sensation in the feet, and ulcers.

- the obstruction and shortened flow of oxygenated blood to the bladder and kidneys are caused due to cardio vascular disease (CVD).

- the internal layers of large and small arteries are the complications caused due to peripheral vascular diseases (PVD).

B. Role of Machine Learning in Diabetes Detection

Machine Learning (ML) models [9] have been developed in most medical implementations as an encouraging tool to help in taking spontaneous conclusions related to various infections, along with DM, which produces favorable results. With ML algorithms, vast amounts of data are processed by minimizing the effort [10]. This data is used to train models, which then generate the most appropriate results associated
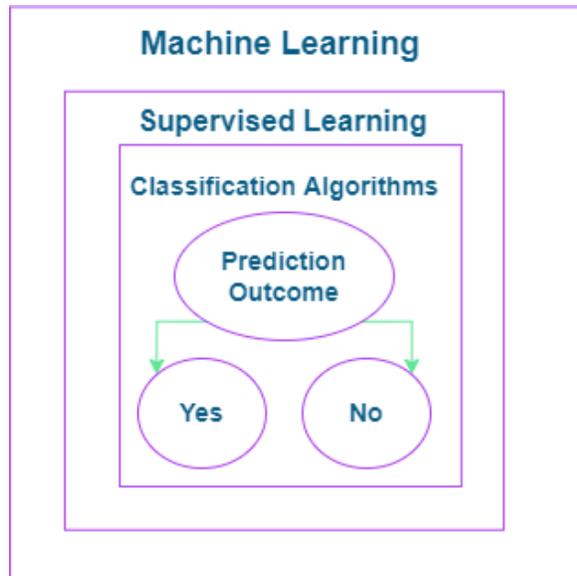
Fig. 1. Machine Learning Classification.

with the input data [11]. It is possible to train the models using any characteristics that are feasible and medically necessary. These parameters differ in accordance with the wide range of symptoms. Some of the learning models used for classification purposes are decision trees, support vector machines, and regression analysis [12]. The approaches are often embedded in statistical analysis to extract useful information from the core data. A combination of these approaches can be used to build a predictive model that identifies the risk complications of T2-DM based on the deciding risk prediction calculator. A scenario of disease prediction with machine learning models is depicted in Fig. 1.

This research study focuses on identifying the risk levels of MIV and MAV complications in T2-DM using different ML algorithms. Accurate prediction of MIV and MAV complications could aid in more targeted measures that would prevent or slow their progression.

The rest of this paper has the following structure. Section II discusses related research and its limitations. Section III presents the proposed methodology. Section IV discusses the experimental results, and Section V concludes the study with potential future work.

II. RELATED WORK

Allen A. et al. [13] trained two ML algorithms to predict DKD severity stages. To assess performance, they compared them to the Centers for Disease Control and Prevention (CDC) risk score. The algorithms were validated using both a hold-out test set and an external dataset obtained from different facilities. In both the hold-out and external datasets, their proposed algorithms outperformed the CDC risk score, achieving an area under the receiver operating characteristic curve (AUROC) of 0.75 on the hold-out set for the prediction of any-stage DKD and an AUROC of over 0.82 for more severe endpoints, compared to the CDC risk score, which had an AUROC of less than 0.70 on all test sets and endpoints. Lu H. et al.

[14] suggested that the perception of acute infection changes and predicting people inculcating the threat of treatment-resistant infection are mainly considered. For chronic disease identification, an ensemble of original patient-channel and ML techniques is proposed. This proposed method is used in networks with health scenarios. T2-DM is identified in a subset of patients for this purpose. This method identified the factors in identifying the acute infection threat using eight ML techniques. The significant observations show that the advanced structure with ML classifiers achieved an AUC ranging from 0.79 to 0.91. Rashid, M. et al. [15] identified a root cause of death among T2-DM patients due to micro vascular problems. Their study aims to examine the use of the entire ML procedure in identifying issues using people's medical, clinical, and statistical examinations. The records of 96 people from Bangladesh were examined with T2-DM. They are examined through a chi-squared examination to demo-graphically represent the major key points in identifying the micro vascular problems like CAN, DPN, and RET. Various ML models like LR, RF, and SVM were used for the exam-ination of micro vascular problems. The exact outcomes are determined through the random forest through hypertension, gender, micro albuminuria, and smoking habit. The authors showed ML represents accurate results in identifying micro vascular problems in T2-DM patients. Based on their records, which aid in controlling these people by later micro vascular problems that lead to early death. Deberneh, H. et al. [16] considered factors like FPG, triglycerides, HbA1c, gamma-GTP, BMI, family history, physical activity, smoking, drinking, gender, age, and uric acid in their study. Then the engaged LR, RF, SVM, XGBoost and ensemble ML procedures relied on these attributes to identify the result as normal, diabetic, or pre-diabetic. Depending on the hypothetical outcomes, the execution of the identified method strives to preferably better in predicting the circumstance of type 2 diabetes. This method also helps doctors and patients with required forecasting data on the probability of occurring type 2 diabetes. Fazakis et al. [17] proposed a worker-centric, IoT-enabled, unobtrusive health, well-being, and functional ability monitoring frame-work with AI tools for the early detection of T2-DM. Their diabetes risk prediction system used several ML models to ap-ply, evaluate, and incorporate KDD components. The ensemble WeightedVotingLRRFs ML model's AUC of 0.884 improves diabetes prediction. Jian, Y. et al. [18] proposed multiple ML algorithms to predict and classify eight diabetes complications. Metabolic syndrome, dyslipidemia, neuropathy, nephropathy, diabetic foot, hypertension, obesity, and retinopathy are among the complications. The authors used a dataset with 884 cases and 79 features. The models' performance was evaluated using accuracy and F1-score metrics, which reached a maximum of 97.8% and 97.7%, respectively. Neha Prerna et al. [19] proposed a research study that uses different ML algorithms to predict the risk of type 2 diabetes among individuals based on their lifestyle and family history. The experiment was carried out with 952 instances collected via an online and offline ques-tionnaire, which included 18 questions about health, lifestyle, and family background. The proposed ML algorithms were also tested on the "Pima Indian Diabetes database." Their experimental results revealed that the random forest classifier performed the best in terms of accuracy. Jung, L. C et al. [20] described a method for identifying subtle effects of genetic variants using whole genome sequencing data and improving

the prediction accuracy of T2-DM at the population level. The method entailed first performing sparse principal component analysis to genotype data to obtain orthogonal features, then creating a new classifier with single nucleotide polymorphism (SNP)-specific regularization parameters to reduce the false positive rate of feature selection, and finally verifying feature relevance with penalized logistic regression. The researchers used a dataset containing 625597 SNPs and 23 environmental variables from 3326 people. The method identified 271 genetic variants with minor effects on T2DM prediction. It is also more than 15 times faster than random forest and extreme gradient boosting (XGBoost) classifiers. Hasan, M.K. et al. [21] proposed a method for DM identification for the metadata, involving 768 women, 268 of who had high blood sugar levels and 500 of whom were normal. In their study, initialization is critical for maintaining cutting-edge results. This includes edge elimination, replacement with the average for lost numbers, data stabilization, factor alternative, and five-fold validation. K-Nearest Neighbor, Decision Trees, Random Forest, AdaBoost, and other Different ML models were used. The authors also proposed a weighted ensembling of different ML models. AUC was chosen as the performance metric. The experiments demonstrated that the ensembling classifier outperforms all others, with sensitivity, specificity, false omission rate, diagnostic odds ratio, and AUC values of 0.789, 0.934, 0.092, 66.234, and 0.950, respectively. Islam, M.S. et al. [22] proposed a study to predict the Hemoglobin Alc (HbAlc) levels in advance using ML methods in order to enable early diagnosis and prevent diabetes complications. The fractional derivative, glucose variability, time in range, and wavelet decomposition methods were used to extract features from continuous glucose monitoring (CGM) data. The CGM data from the Diabetes Research in Children Network (DirecNet) was utilized. According to the results, the ensembling of the random forest and extreme gradient boosting algorithms, combined with the feature fusion, produced the best performance with a low mean absolute error (MAE) of 3.39 mmol/mol and a high coefficient of determination (R-squared) score of 0.81. Kopitar, L. et al. [23] compared ML models for the prediction of T2-DM by using various regression techniques on undiagnosed patient data. Fasting plasma samples are measured over a six-month period using 100 computational iterations. These iterations were examined using various data subsets. According to the study analysis, the linear regression model had the lowest average RMSE of 0.838, followed by random forest with 0.842 and Xgboost with 0.881. Dagliati et al. [24] proposed a data mining pipeline to derive a set of predictive models of T2DM complications from nearly 1000 patients' electronic health record data. Clinical center profiling, predictive model targeting, predictive model construction, and model validation are all part of the pipeline. The logistic regression-based method with stepwise feature selection was used to predict the onset of retinopathy, neuropathy, or nephropathy at three different time intervals: three, five, and seven years after the initial visit to the Hospital Center for Diabetes. Gender, age, time since diagnosis, BMI, hypertension, HbA1c, and smoking habit were all factors considered in the study. The final models, customized for the complications, had an accuracy of up to 0.838. For each complication and time scenario, different attributes were chosen. This led to specialized models that are easy to use in clinical practice. Wei, S. et al. [25] used a variety of machine learning techniques, i.e., Neural Net-

works, Support Vector Machines, and Decision Trees to detect diabetes. The best accuracy acquired was 77.86% through the 10-fold cross-validation approach. Fan Yuting et al. [26] proposed an effective ML model for identifying the problem of blood sugar levels in non-adherent T2-DM. The authors looked at people who had not had glycosylated hemoglobin in the previous month to identify the risk in blood sugar levels. Seven different ML procedures are utilized to implement eighteen identification methods. Identification achievement is majorly analyzed through the AUC of the examining group. Based on 800 patients' data, 20.6% could meet the insertion range, of which 78.2% had poor glycemic guide. The greater AUC of the analysis set for nephropathy, peripheral neuropathy, angiopathy, retinopathy, and glycosylated hemoglobin is determined as 0.902 ± 0.040, 0.859 ± 0.050, 0.889 ± 0.059, 0.832 ± 0.086, and 0.825 ± 0.092 accordingly. Both the ML models and univariate testing attained an equal outcome. Fiarni et al. [27] proposed Naive Bayes and C4.5 classification approaches, as well as k-means clustering, for identifying the risk complications of T2-DM patients. The authors analyzed each technique's reliability and identified the associated elements and sub-features as clinical contributing factors. As a result, the most major risk factor for retinopathy is a female patient who is now experiencing a hypertensive problem. In terms of nephrotic syndrome, the most major risk factor is a history of diabetes lasting over four years. Furthermore, it was more prevalent in female patients with a BMI over 25. There is no clear association between the duration of diabetes and certain complications. The overall accuracy of the suggested model is 68%, which means that it could be used as an alternative way to find diabetes complications early. Sudharsan, B. et al. [28] proposed a reducing phenomenon of blood sugars in people having T2-DM. The authors analyzed the risk of complications through different data sets. The quantity of self-regulation of blood sugar levels required by the method is nearly ten per week. The vulnerability of the method for identifying blood sugar levels in the coming 24 hours is 92%, and the selectivity is 70%. In their work, the identification group was four hours of blood sugar, and the selectivity advanced to 90%. The advanced ML methods can identify blood sugar levels with an accurate level of vulnerability and selectivity.

After taking earlier research work into consideration, we observed that the majority of the work concentrated on identifying the T2-DM with limited complications. To overcome this limitation and to identify the severity levels, in this proposed work, we concentrated on predicting the risk levels at an early stage, falling under low, medium, and high categories. These levels are discriminated against among the T2-DM patients with MIV and MAV complications such as Nephropathy, Neuropathy, Retinopathy, cardio vascular and Peripheral vascular.

III. METHODOLOGY

This section goes into more detail about the statistical analysis of the dataset used for this study, the T2-DM patient network, data preprocessing approaches, ML methods, and the proposed model.

A. Statistical Analysis of the Dataset

For this study, 3068 records are considered, particularly with subjects between 30 and 80 years of age, of which 1565

are males, and 1503 are females. The mean, standard deviation, error rate, and P-value of primary attributes that are available in the dataset are displayed in Table I. These attributes are mainly considered in identifying the risk levels of MIV and MAV complications in T2-DM patients collected from various multi-specialty hospitals in India [29].

The distribution of samples depends on the parameters in a dataset and can be represented using frequency tables. These are useful in making decisions that appear more or less within the dataset. Every parameter and its selection range can be easily identified. From this, basic charts like histograms and bar charts can be generated for data visualization. The representation of some primary attributes in the dataset is depicted using a histogram, shown in Fig. 2.

B. T2-DM Complications Network

In this work, the graph theory concepts are used to construct the T2-DM complications network. This network represents the common complications among different patients. For this, four different patients with MIV and MAV complications are considered from the dataset. The relation among patients and various complications are determined by Eq. (1).

$$G = (V, E) \tag{1}$$

where 'V' represents the nodes that indicate the patient's complications. The common complications for these patients are connected through 'edges' (E). Hence, these sets of nodes and edges are combined together to form a complication network (G).

Fig. 3. represents the complication network among T2-DM patients to identify the relationship between patients and their common health complications. The left-hand side of this figure shows three male and one female patients' data along with their disease complications, respectively. The middle part represents the edge relation between patients connected through nodes. For instance, patients p1, p2, and p4 all have T2-DM in common. Due to this, "neuropathy" is a common complication. Also, patients p2 and p4 show a common complication of "nephropathy". Hence, by considering this scenario, the patient's p1, p2, and p4 are connected through nodes, and their common health comorbidities are interlinked with edges, thereby forming a subgraph to construct a complications network. By constructing this network, the classification between T2-DM and non-T2-DM becomes easier through ML models.

C. Data Preprocessing

When developing ML models, data preprocessing is the first step in the process. Real-world data is tainted and contaminated by inconsistencies, noise, incomplete information, missing values, inaccurate (containing errors or outliers), and lack of specific attribute values [31]. This is where data preprocessing comes into play; it helps to clean, format, and organize raw data, preparing it to build ML models. Simply put, data preprocessing helps improve data quality and promotes the extraction of meaningful insights from data to train more accurate prediction models. The data preprocessing procedure in ML includes the following steps [32].

- **Missing Values:** This step involves identifying and appropriately handling missing values; failing to do so could lead to inaccurate and erroneous conclusions and inferences drawn from the data. There are two methods for dealing with missing data: (a) deleting a specific row, in which we remove a particular row that contains a null value for a feature or a particular column where more than 70% of the values are missing; and (b) calculating the mean, median, or mode of a specific feature, column, or row that contains a missing value and replacing the outcome with the missing value. This method is helpful for features with numeric data, such as salary, year, and so on, and it can add variance to the dataset while efficiently negating any data loss. As a result, it produces better results than the first method.

- **Removing Duplicates:** This step includes deleting duplicate entries. During model training, an entry that appears more than once is given disproportionate weight. Where identical entries are not all in the same set, duplicate entries can ruin the split between training, validation, and test sets. This can lead to biased estimates of how well the model will perform, which can cause the model to underperform in production.

- **Removing irrelevant data:** This step involves removing irrelevant entries from the dataset. Data often comes from a variety of sources, and a given set of data is likely to have entries that don't belong.

- **Detecting Outliers:** This step entails detecting outliers by exploring the ranges and possibilities for categorical and numerical data entries. For instance, a negative price for a vehicle is an outlier. Outlier detection or anomaly detection algorithms, such as Isolation Fores or KNN, can also be used to detect and remove outliers automatically.

- **Categorical Data Encoding:** This step includes transforming categorical data (i.e., a patient's gender) into numerical values. ML models are built on mathematical equations that can only work with numbers. As a result, the categorical values of the features must be converted into numerical values, which can then be fed into ML models to learn from and improve performance.

- **Feature Selection:** This step always plays an important role in machine learning, where we will have several features in the dataset and have to select the best ones when building a model. The inclusion of irrelevant features reduces the model's generalization capability and may reduce a classifier's overall accuracy. In addition, the model's overall complexity grows as more features are added. Feature selection methods in machine learning can be broadly classified as Wrapper, Embedded, and Filter. Wrapper methods use a greedy search approach, evaluating all possible feature combinations against the evaluation criterion. Embedded methods are iterative in the sense that they handle each iteration of the model training process and extract those features which contribute the most to the training for that iteration. Filter methods pick

TABLE I. SUMMARY OF PRIMARY ATTRIBUTES OF THE DATASET [29]

Parameter	Description	Range	Population Size (n)	Mean	Standard Deviation	Standard Error Mean	P-value
Age	Life span of patient	30 to 49	1222	54.92	14.494	0.262	0.011
		50 to 59	600				
		60 to 69	598				
		70 to 80	648				
Sex	Identity of patient	Male	1565	0.51	0.500	0.009	0.048
		Female	1503				
BMI	Determines the level of fat	18.5	155	28.823	6.3138	0.1140	0.025
		18.5 to 24.9	779				
		25.0 to 29.9	736				
		30.0	1398				
SBP	Pressure in arteries when heart beats	120	462	139.32	20.008	0.361	0.015
		120 to 139	891				
		140	1715				
DBP	Pressure in arteries when heart rest among the beats	80	822	84.67	11.348	0.205	0.037
		80 to 89	755				
		90	1491				
HbA1C	Blood pressure attached to hemoglobin	5.7%	127	11.1308	2.671	0.4823	0.016
		=5.7% to 6.4%	617				
		=6.5%	2324				
FBS	Blood sugar level after fasting	=100 mg/dL	40	230.19	76.879	1.388	0.006
		100 to =125 mg/dL	288				
		=126 mg/dL	2740				
PPBS	Determine type of sugar	180 mg/dL	447	334.67	123.539	2.230	0.001
		=180 to 250 mg/dL	492				
		=250 mg/dL	2129				
DIA_LIFE	Span of diabetes in months	40	1350	43.14	12.367	0.223	0.038
		=40 to 60	1388				
		=60	330				
Smoking	-	No	847	1.44	1.129	0.020	0.017
		Ex-smoker	752				
		Occasionally	740				
		Current	729				
Medical Usage	Medicine usage	No	1498	0.51	1.141	0.009	0.019
		Yes	1570				
Medical adherence	Medicine usage pertained to time	Low	1083	0.97	0.822	0.015	0.001
		Medium	994				
		High	991				

TABLE II. PERFORMANCE OF MICRO VASCULAR AND MACRO VASCULAR MODULES

Module	Model	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Nephropathy (NEP)	RF with base DT	95.43%	96.57%	94.0%	96.81%	95.27%
	RF with base LR	94.78%	97.80%	91.13%	98.12%	94.35%
	RF with base AB	92.91%	93.85%	90.73%	94.81%	92.26%
Neuropathy (NEU)	RF with base DT	94.62%	96.84%	92.0%	97.13%	94.35%
	RF with base LR	93.05%%	94.57%	91.26%	94.81%	92.88%
	RF with base AB	91.12%	93.52%	88.54%	93.75%	90.96%
Retinopathy (RET)	RF with base DT	96.25%	96.78%	95.85%	96.66%	96.32%
	RF with base LR	93.37%	94.83%	92.25%	94.58%	93.53%
	RF with base AB	90.78%	92.65%	88.94%	92.71%	90.76%
Cardio Vascular (CVD)	RF with base DT	97.55%	96.0%	98.96%	96.28%	97.46%
	RF with base LR	93.91%	93.54%	93.95%	93.88%	93.75%
	RF with base AB	91.04%	92.91%	88.63%	93.39%	90.72%
Peripheral Vascular (PVD)	RF with base DT	97.83%	98.0%	86.88%	98.0%	92.11%
	RF with base LR	97.56%	97.9%	71.42%	98.0%	83.33%
	RF with base AB	97.67%	98.0%	78.94%	98.0%	88.23%

up the intrinsic properties of the features measured using univariate statistics. Information Gain and the Chi-square Test are two of the Filter methods. In this research, we have used the Chi-square Test to select the features related to T2-DM. The Chi-square between each feature and the target is calculated, and the number of features with the best Chi-square scores is selected. The Formula for Chi-square is given in Eq. (2), where c is the degrees of freedom, O is the observed value(s), and E is the expected value(s).

$$\chi_c^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{2}$$

- **Data Split:** This step includes splitting the dataset for the ML model into two or more separate sets. Typically, with a two-part split, one part (training dataset) is used to train the ML model, and the other (test dataset) is used to evaluate or test the model. The testing data set is used following training. Usually, the dataset is split into an 80:20 ratio or 70:30 ratio. This
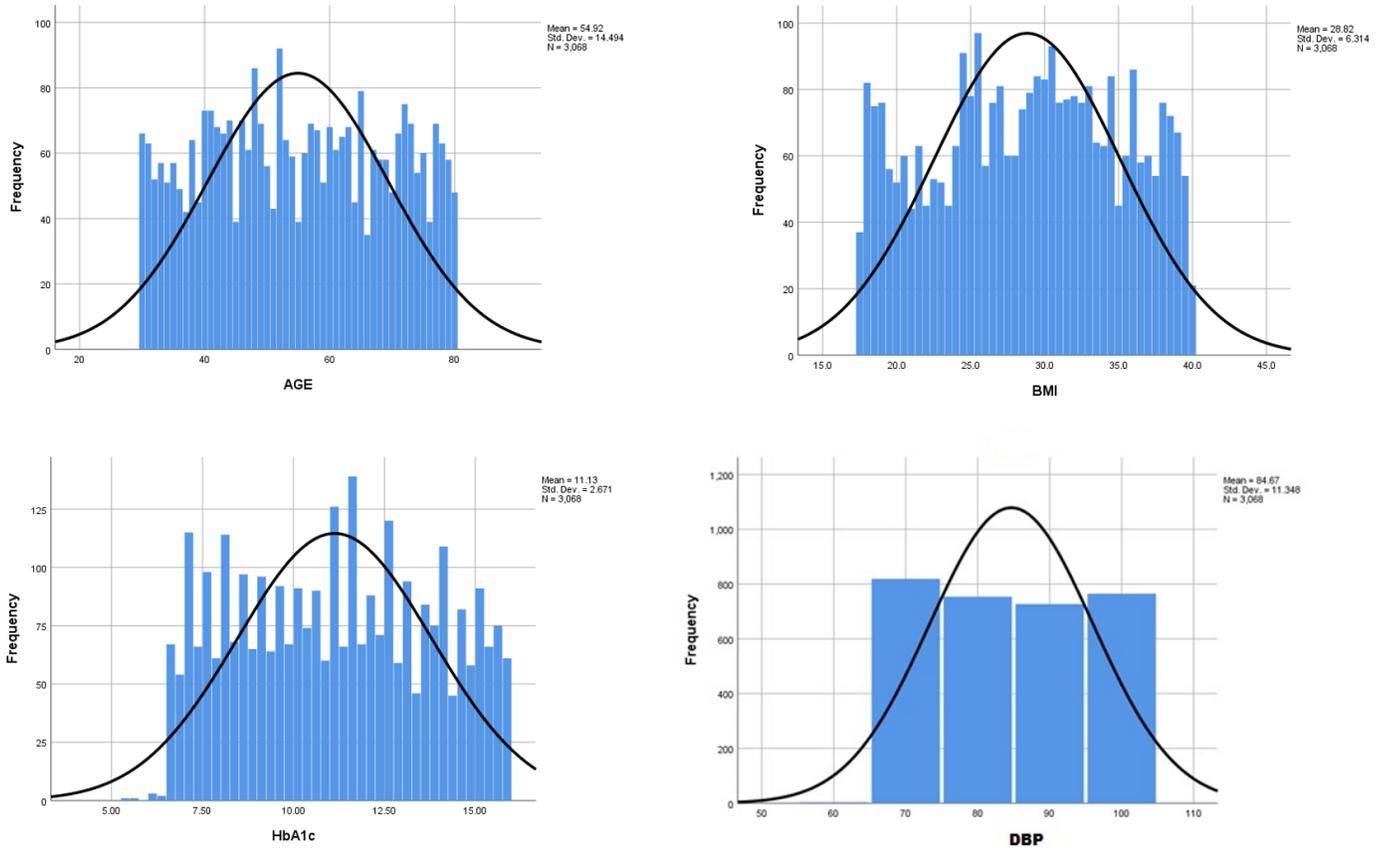
Fig. 2. A Histogram of Some Primary Features in the Dataset.

TABLE III. EXISTING COMPLICATIONS

Patient	Existing Complications				
	NEP	NEU	RET	CVD	PVD
A	0	1	0	0	0
B	1	0	1	0	0
C	0	0	0	0	0
D	1	0	1	0	0
E	1	1	0	0	0
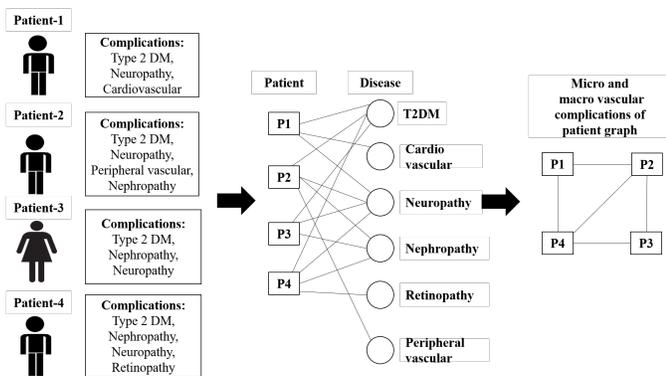
Fig. 3. MIV and MAV Complication Network.
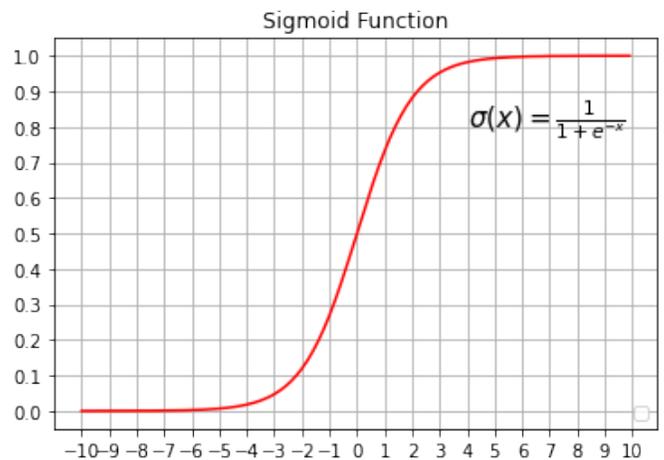
Fig. 4. The Sigmoid Function.

means that 80% or 70% of the data is used for training the model, while the remaining 20% or 30%, is used

Algorithm: Generate_decision_tree. Generate a decision tree from the training tuples of data partition *D*.

Input:

- Data partition, *D*, which is a set of training tuples and their associated class labels;
- *attribute_list*, the set of candidate attributes;
- *Attribute_selection_method*, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes. This criterion consists of a *splitting_attribute* and, possibly, either a *split point* or *splitting subset*.

Output: A decision tree.

Method:

(1) create a node *N*;
(2) **if** tuples in *D* are all of the same class, *C* **then**
(3)   return *N* as a leaf node labeled with the class *C*;
(4) **if** *attribute_list* is empty **then**
(5)   return *N* as a leaf node labeled with the majority class in *D*; // majority voting
(6) apply Attribute_selection_method(*D*, *attribute_list*) to find the "best" *splitting_criterion*;
(7) label node *N* with *splitting_criterion*;
(8) **if** *splitting_attribute* is discrete-valued **and**
        multiway splits allowed **then** // not restricted to binary trees
(9)   *attribute_list* ← *attribute_list* − *splitting_attribute*; // remove *splitting_attribute*
(10) **for each** outcome *j* of *splitting_criterion*
        // partition the tuples and grow subtrees for each partition
(11)   let *D_j* be the set of data tuples in *D* satisfying outcome *j*; // a partition
(12)   **if** *D_j* is empty **then**
(13)   attach a leaf labeled with the majority class in *D* to node *N*;
(14)   **else** attach the node returned by **Generate_decision_tree**(*D_j*, *attribute_list*) to node *N*;
     endfor
(15) return *N*;

Fig. 5. Basic DT Algorithm [30].

TABLE IV. PREDICTED COMPLICATIONS

Patient	Proposed model risk predictions				
	NEP	NEU	RET	CVD	PVD
A	Low	Medium	No	No	No
B	Medium	No	Medium	Low	No
C	Low	No	No	No	No
D	Medium	No	Medium	Low	No
E	Medium	Medium	No	Low	No

for testing. Most often, data is separated into three or more sets. With three sets, the additional set is the validation set, which is used to modify the parameters of the learning process.

- **Data Scaling:** This crucial step concludes the data preprocessing phase in ML. It is a technique for converting all independent features of a dataset to the same scale. This allows for faster learning convergence and more uniform influence across all weights. Normalization and Standardization are two commonly used methods for feature scaling.

  - Normalization: This technique is known as Min-Max scaling, in which all independent feature values are changed between 0 and 1, as defined below:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

where X_{min} and X_{max} are the minimum and the maximum values of the feature, respectively.

  - Standardization: In this technique, the independent feature values are standardized by removing the mean and scaling to unit variance. The standard score of a sample X is calculated as:

$$X' = \frac{X - \mu}{\sigma} \quad (4)$$
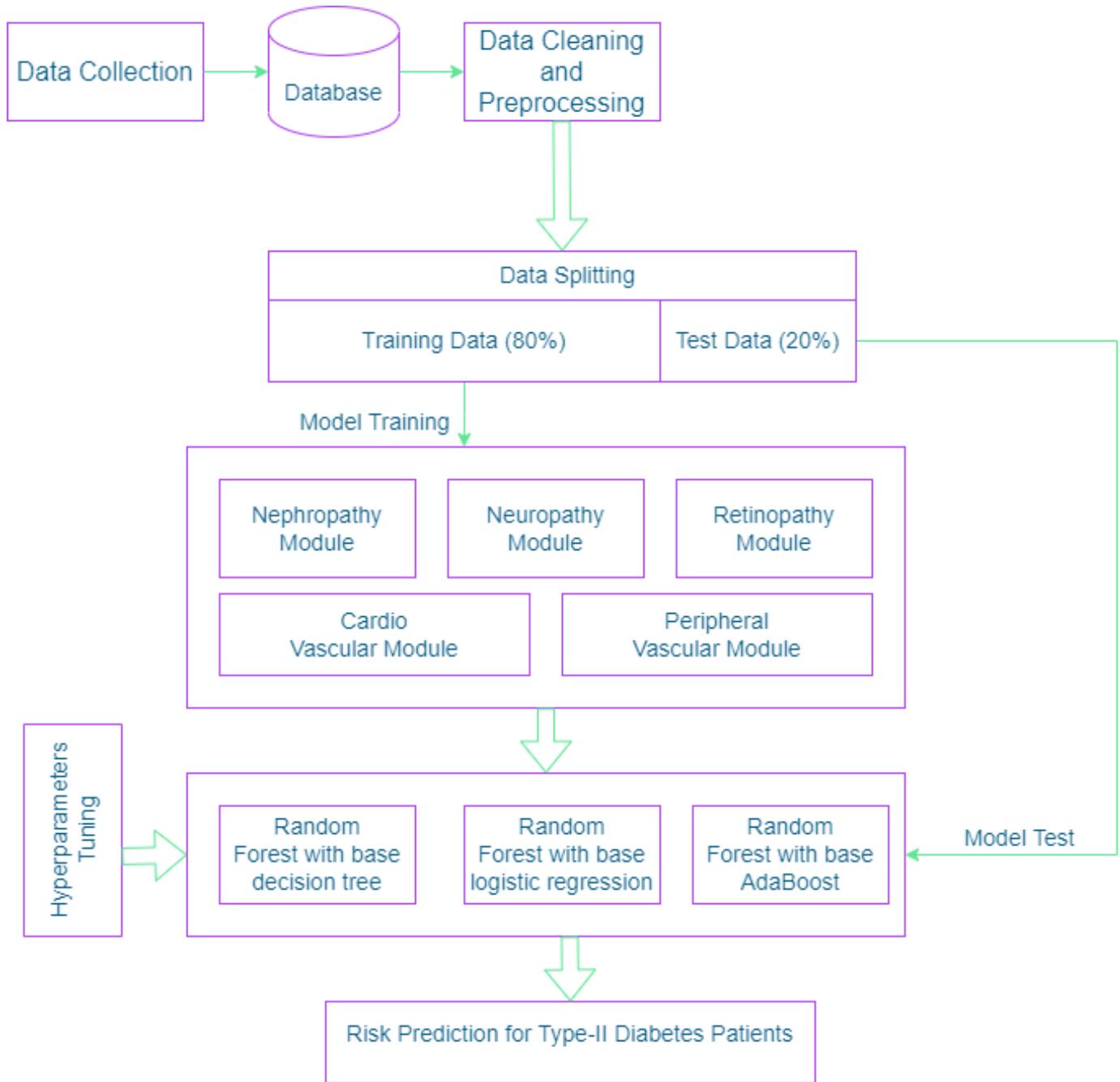
Fig. 6. Proposed Framework.

where μ is the mean of the feature values, and σ is the standard deviation of the feature values.

The data preprocessing procedure can vary slightly according to each dataset, but many of the aforementioned steps are applicable to all situations.

D. ML Models

1) Logistic Regression: Logistic Regression (LR) is a powerful ML algorithm commonly used to solve binary classification problems. It is called after the core function of the method, the logistic function. The logistic function (a.k.a. sigmoid function) has an S-shaped curve that can map any real-valued number to a value between 0 and 1 [33]. The sigmoid function, usually denoted by $\sigma(x)$ is defined as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

where e is Euler's number and x is the actual numerical value to be transformed. Fig. 4 shows a plot of the numbers between -10 and 10 transformed into the range 0 and 1 using the sigmoid function. LR, like linear regression, uses an equation as its representation. To predict an output value (y),

- TP (True Positives): Number of times the model correctly predicts positive samples.
- TN (True Negatives): Number of times the model correctly predicts negative samples.
- FN (False Negatives): Number of times the model incorrectly predicts positive samples as negatives.
- FP (False Positives): Number of times the model incorrectly predicts negative samples as positives.

Fig. 7. Confusion Matrix for a Binary Classifier.
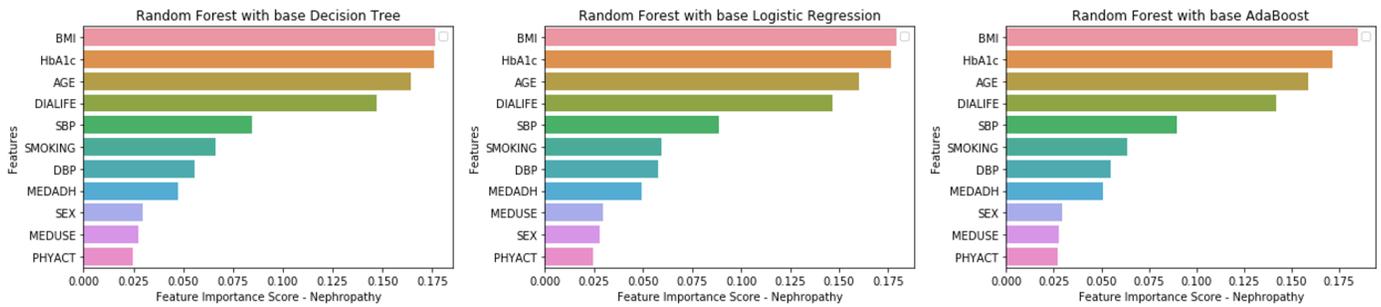
Fig. 8. Feature Importance Scores for Neuropathy.
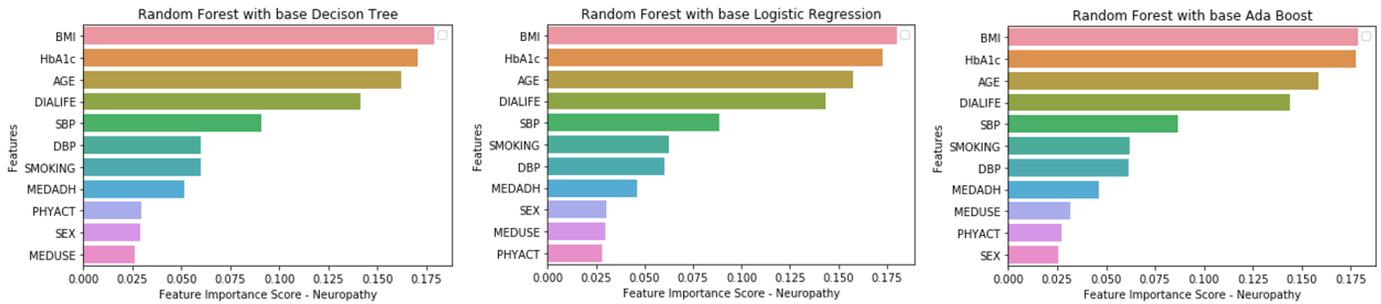
Fig. 9. Feature Importance Scores for Nephropathy.
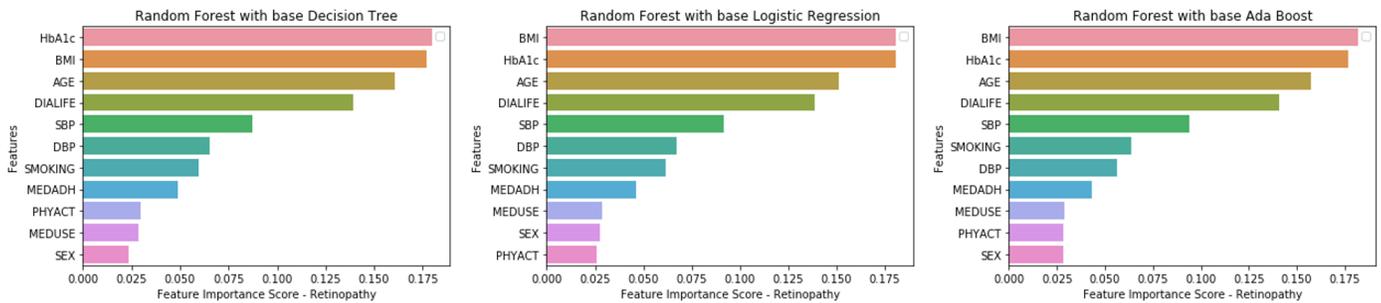
Fig. 10. Feature Importance Scores for Retinopathy.
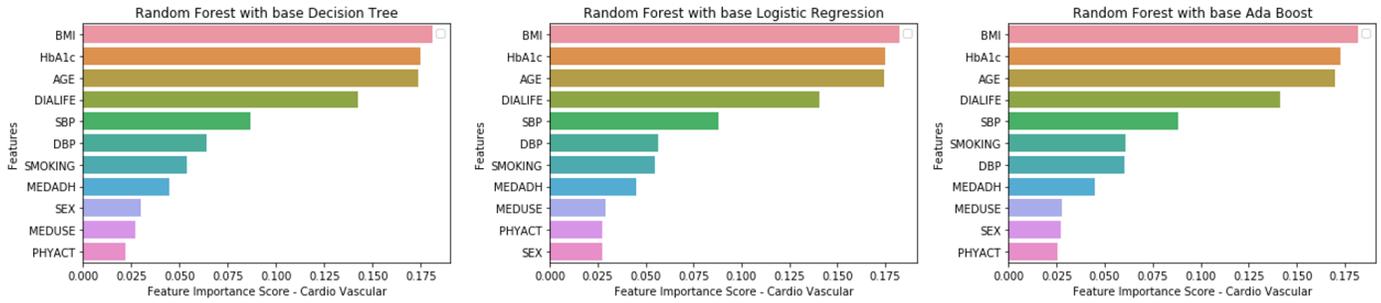
Fig. 11. Feature Importance Scores for Cardio Vascular.
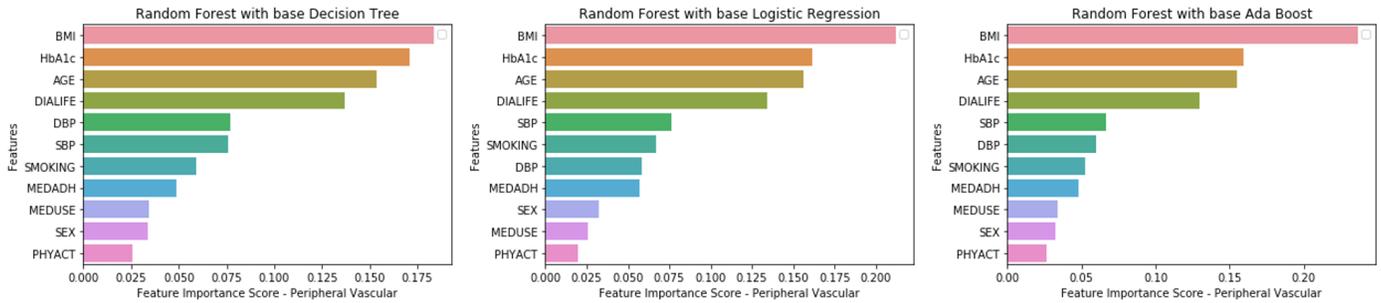
Fig. 12. Feature Importance Scores for Peripheral Vascular.
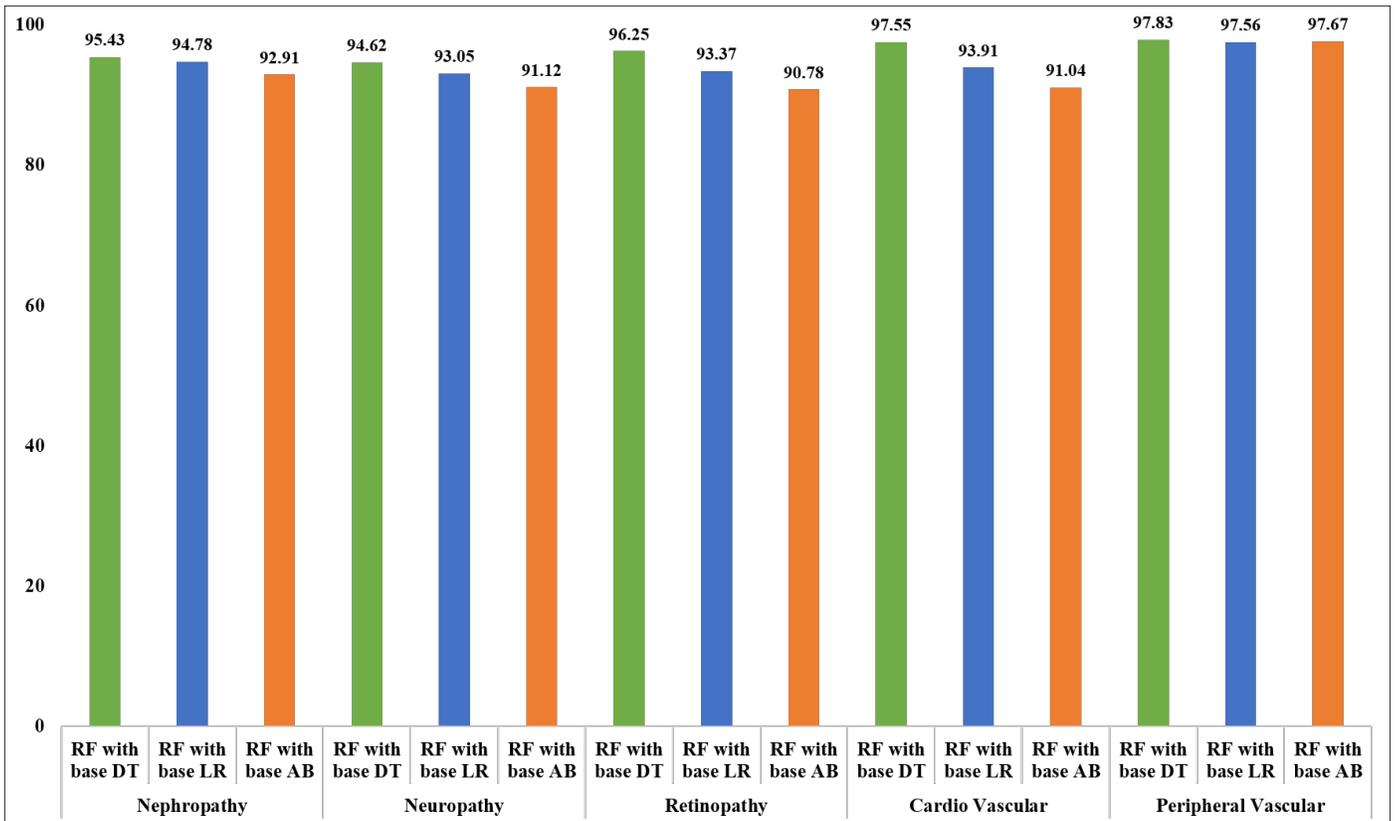
Fig. 13. Accuracy Results for All Complication Modules.

input values (x) are linearly combined using weights values. The output value being modeled is a binary value (0 or 1) rather than a numeric value, which distinguishes it from linear regression. An example LR equation is shown below:

$$y = \frac{1}{1 + e^{-(w_0 + w_1 \times x)}} \quad (6)$$

Where y is the predicted output, w_0 is the bias and w_1 is the weight for the single input value (x). Each column in the dataset has an associated w weight that must be learned from the training dataset. LR predicts the probability of the positive class (e.g., a patient has diabetes). For example, if we're predicting whether or not a patient has diabetes or not based on their age, then the positive class could be diabetic, and the LR model could be written as the probability of being diabetic given a patient's age, or more formally:

$$P(x) = P(y = \text{diabetic} \mid x) \quad (7)$$

To get a clear answer, we can snap the probabilities to a binary class value, such as:

$$
\begin{aligned}
\text{prediction} &= 0 \text{ (non-diabetic)} \text{ IF } p(\text{ diabetic }) < 0.5 \\
\text{prediction} &= 1 \text{ (diabetic)} \text{ IF } p(\text{ diabetic }) \geq 0.5
\end{aligned}
\quad (8)
$$

2) Decision Tree: Decision Tree (DT) algorithm is another popular supervised ML algorithm used for solving both classification and regression problems. The goal of using a DT is to learn simple decision rules from training data to create an efficient model that predicts the class of the target variable [34]. DT is a flowchart-like tree structure in which each leaf node corresponds to a class label, and features are represented on the internal node of the tree. The root node is the topmost node in a tree. The DT can be better understood with the help of the algorithm summarized in Fig. 5.

By adopting the above algorithm to the problem of determining whether an instance belongs to class-0 or class-1, we can construct a decision tree by selecting a root node, internal nodes, and leaf nodes, and then defining the class's splitting criteria. For example, we can select the "glucose" feature to be the root node and based on it and other features such as "systolic BP," "Diastolic BP," "Age," and "BMI," we can construct our tree.

The most challenging aspect of the DT algorithm is selecting The root node or first test attribute based on what we will start splitting the data. It is selected based on statistical measures like Information Gain (IG), Gain Ratio, or Gini Index. In this paper, we used the IG measure. IG helps to measure the reduction of uncertainty of a certain feature. It also helps decide which feature is good as a root node. The calculation of $IG(a)$ shows us the formula for the gain in the general case. Let S denote the dataset to be split by creating a new node. Let's suppose that the attribute a can take m values: $a_1, a_2, \ldots, a_m$, and that p_i is the fraction of the objects in S with $a = a_i$. Then, the information gain of a is:

$$IG(a) = H(S) - \sum_{i=1}^{m} p_i H(S \mid a = a_i) \quad (9)$$

Here $H(s)$ is the Entropy of the dataset, is a function H of probabilities $p_1, p_2, \ldots, p_n$. Entropy can be thought of as the amount of variance in the data. For binary classification problems, the following formula is used to compute Entropy.

$$H(S) = -\sum_{i=1}^{n} p_i \log_2 p_i \quad (10)$$

3) Random Forest: Random Forest (RF) is another popular ML technique used to solve regression and classification problems. RF consists of many decision trees. The 'forest' generated by the random forest algorithm is trained using an ensemble method known as bootstrap aggregation (or bagging for short) [35]. Bagging is a technique that combines the predictions from multiple decision trees to make more accurate predictions than any single model. The model's final output is based on the majority of the predictions if the problem is a classification, or the mean of the predictions if the problem is a regression.

4) AdaBoost: Boosting is an ensemble technique for constructing a strong classifier from a collection of weak classifiers. This is accomplished by first creating a model from the training data, followed by the creation of a second model that attempts to correct the errors in the first model. Models are added until the training dataset is perfectly predicted or until the maximum number of models is reached [36]. AdaBoost, shortened for Adaptive Boosting [37], was the first successful boosting algorithm developed for binary classification problems. Decision trees with one level are the most appropriate and widely used algorithm with AdaBoost because these trees are so short and only have one classification decision. They are commonly known as "decision stumps." Weights are assigned to each instance in the training dataset. The initial weight is set as follows:

$$\text{weight}(x_i) = \frac{1}{M} \quad (11)$$

where x_i is the i'th training instance and M is the number of training instances. To train a single model, a weak classifier is prepared on the training data using the weighted samples. Only binary classification tasks are supported, so each decision stump makes one decision on one input variable and outputs a $+1.0$ or -1.0 value for the first and second class value.

The misclassification rate (E) for the trained model can be calculated as follows:

$$E = \frac{r - M}{M} \quad (12)$$

where r is the number of training instances predicted correctly by the model and M is the total number of the training instances. The opposite of misclassification rate would be accuracy, calculated as:

$$\text{Accuracy} = 1 - E \quad (13)$$

To take into account the weighting of the training instances, the weighted sum of the misclassification rate is computed as:

$$E = \frac{\sum_{i=1}^{n} (w_i \times p_i)}{\sum_{i=1}^{n} w} \quad (14)$$

where w is the weight for training instance i and p_i is the prediction error for training instance i, which is 1

if misclassified and 0 if correctly classified. For the trained model, a stage value is calculated, which provides a weighting for any predictions made by the model. A trained model's stage value s_v is calculated as follows:

$$s_v = \ln\left(\frac{1-E}{E}\right) \qquad (15)$$

The stage weight has the effect of giving more weight or contribution to the final prediction to more accurate models. The training weights are adjusted so that incorrectly predicted instances receive more weight and correctly predicted instances receive less weight. For instance, the weight w of one training instance is updated as follows:

$$w = w \times e^{(s_v \times p)} \qquad (16)$$

E. Proposed T2DC Prediction Model

The proposed framework for predicting the risk levels of MIV and MAV complications among T2-DM patients using the RF-based method is shown in Fig. 6. It is broken down into six major steps: a) gathering data from reliable sources; b) cleaning and preprocessing the data; c) dividing the cleaned data into two sets—a training set (80%) and a testing set (20%); d) training the model with the training set; e) evaluating the model performance of the trained model with various base estimators such as DT, LR, and Adaboost models; and f) tuning the model's hyperparameters to see if its accuracy can be improved.

IV. RESULTS AND DISCUSSIONS

This section presents all of the results obtained by the proposed framework, as well as related discussions.

A. Model Performance Evaluation Metrics

Evaluation of the performance of a classification model is based on the number of test samples that the model correctly and incorrectly predicts. The confusion matrix extracts additional information about the performance of a predictive model. It demonstrates which classes are correctly and incorrectly predicted and what types of errors are made. Fig. 7 shows an illustration of a confusion matrix for a binary classifier. The four classification metrics (TP, FP, FN, TN) are calculated, and the confusion matrix compares the model's predicted value to the actual value. The confusion matrix is not exactly a performance metric, but it is used to calculate important classification metrics like accuracy, precision, recall, specificity, and, most importantly, the f1-score, which are used to evaluate the results.

1) **Accuracy:** It is the fraction of correct predictions made by the model.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \qquad (17)$$

2) **Precision:** It is the proportion of true positives to all positive predictions made by the model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \qquad (18)$$

3) **Recall:** It is the proportion of actual positives correctly identified by the model.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \qquad (19)$$

4) **Specificity:** It is the proportion of actual negatives correctly identified by the model.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\% \qquad (20)$$

5) **F1-score:** It is the proportion of actual positives correctly identified by the model.

$$F1\text{-}score = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \times 100\% \qquad (21)$$

B. Feature Importance Score

Feature importance is a process that involves calculating the score for the input features for a given model — the scores simply represent the "importance" of each feature. A higher score indicates that the specific feature will have a greater impact on the model used to predict a specific class. This can enhance a predictive model's efficiency and effectiveness on the problem. The feature importance scores on T2-DM complications such as nephropathy, neuropathy, retinopathy, cardio vascular, and peripheral vascular modules of the RF model with different base estimators, DT, AdaBoost, and LR, are shown in Fig. 8 to 12, respectively. It is evident from the figures that BMI and HbA1c are the most significant features in the dataset.

C. Model Evaluation Discussion

Table II represents the evaluation of MIV and MAV complications in T2-DM patients. After evaluating the proposed ML models, the decision tree as a base model with RF provides the best performance in all evaluation metrics. The visualization of the accuracy metric for all complication modules is shown in Fig. 13.

D. Risk Prediction Identification

Table III shows the results of T2-DM patients with existing complications, and Table IV shows the predicted complications of the same patients. The existing complications only represent True (1) or False (0) in regard to NEP, NEU, RET, CVD, and PVD. This type of result is not sufficient to assist doctors in identifying further complications. To overcome this, our proposed model can automatically identify the risk levels with respect to low, medium, and high depending upon the severity of T2-DM complications among patients. For instance, T2-DM patient-A has only one NEU complication in existing data. In contrast, the same instance, when evaluated on the proposed risk prediction model, shows a "low" risk in NEP and a "medium" risk in NEU. This helps healthcare providers make accurate treatment plans for T2-DM patients so they can give them good clinical care.

V. Conclusion

AI and its applications, such as machine learning (ML) in medical diagnosis and healthcare, have shown enormous promise in recent years, both in terms of improving care and alleviating the enormous strains on the healthcare system. ML-based solutions are revolutionizing diabetes care and helping the medical community gain ground in the fight against the disease. The number of people with diabetes who go undiagnosed can be lowered with the help of ML algorithms that allow for accurate early diagnosis. In this research, we have identified MIV and MAV risk levels of T2-DM complications by proposing a T2DC ML-based prediction model. We have used a decision tree as a base estimator with random forest and obtained better accuracy when compared to other base models. The proposed model achieves 95.43%, 94.62%, 96.25%, 97.55%, and 97.83% accuracies for NEP, NEU, RET, CVD, and PVD complications, respectively. The model can be used as a viable aid in clinical decision-making for practitioners and diabetes educators to improve the quality of life of patients. In the future, we will study how T2-DM and related complications affect pregnant women.

References

[1] S. Ciardullo and G. Perseghin, "Prevalence of elevated liver stiffness in patients with type 1 and type 2 diabetes: A systematic review and meta-analysis," *Diabetes Research and Clinical Practice*, p. 109981, 2022.

[2] X. Ding, S. Rong, Y. Wang, D. Li, L. Wen, B. Zou, D. Zang, K. Feng, Y. Liang, F. Wang *et al.*, "The association of the prevalence of depression in type 2 diabetes mellitus with visual-related quality of life and social support," *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, vol. 15, p. 535, 2022.

[3] N. Sambyal, P. Saini, and R. Syal, "A review of statistical and machine learning techniques for microvascular complications in type 2 diabetes," *Current Diabetes Reviews*, vol. 17, no. 2, pp. 143–155, 2021.

[4] Q. Xu, L. Wang, and S. S. Sansgiry, "A systematic literature review of predicting diabetic retinopathy, nephropathy and neuropathy in patients with type 1 diabetes using machine learning," *J. Med. Artif. Intell*, vol. 3, no. 6, 2020.

[5] E. Adua, E. A. Kolog, E. Afrifa-Yamoah, B. Amankwah, C. Obiriko-rang, E. O. Anto, E. Acheampong, W. Wang, and A. Y. Tetteh, "Predictive model and feature importance for early detection of type ii diabetes mellitus," *Translational Medicine Communications*, vol. 6, no. 1, pp. 1–15, 2021.

[6] R. Wang, Z. Miao, T. Liu, M. Liu, K. Grdinovac, X. Song, Y. Liang, D. Delen, and W. Paiva, "Derivation and validation of essential predictors and risk index for early detection of diabetic retinopathy using electronic health records," *Journal of Clinical Medicine*, vol. 10, no. 7, p. 1473, 2021.

[7] M. Maniruzzaman, M. M. Islam, M. J. Rahman, M. A. M. Hasan, and J. Shin, "Risk prediction of diabetic nephropathy using machine learning techniques: A pilot study with secondary data," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 15, no. 5, p. 102263, 2021.

[8] L. Muhammad, E. A. Algehyne, and S. S. Usman, "Predictive supervised machine learning models for diabetes mellitus," *SN Computer Science*, vol. 1, no. 5, pp. 1–10, 2020.

[9] A. Al Bataineh and A. Jarrah, "High performance implementation of neural networks learning using swarm optimization algorithms for eeg classification based on brain wave data," *International Journal of Applied Metaheuristic Computing (IJAMC)*, vol. 13, no. 1, pp. 1–17, 2022.

[10] A. A. Bataineh, "A comparative analysis of nonlinear machine learning algorithms for breast cancer detection," *International Journal of Machine Learning and Computing*, vol. 9, no. 3, pp. 248–254, 2019.

[11] A. Al Bataineh, D. Kaur, and S. M. J. Jalali, "Multi-layer perceptron training optimization using nature inspired computing," *IEEE Access*, vol. 10, pp. 36 963–36 977, 2022.

[12] H. F. Ahmad, H. Mukhtar, H. Alaqail, M. Seliaman, and A. Alhumam, "Investigating health-related features and their impact on the prediction of diabetes using machine learning," *Applied Sciences*, vol. 11, no. 3, p. 1173, 2021.

[13] A. Allen, Z. Iqbal, A. Green-Saxena, M. Hurtado, J. Hoffman, Q. Mao, and R. Das, "Prediction of diabetic kidney disease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus," *BMJ Open Diabetes Research and Care*, vol. 10, no. 1, p. e002560, 2022.

[14] H. Lu, S. Uddin, F. Hajati, M. A. Moni, and M. Khushi, "A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus," *Applied Intelligence*, vol. 52, no. 3, pp. 2411–2422, 2022.

[15] M. Rashid, M. Alkhodari, A. Mukit, K. I. U. Ahmed, R. Mostafa, S. Parveen, and A. H. Khandoker, "Machine learning for screening microvascular complications in type 2 diabetic patients using demographic, clinical, and laboratory profiles," *Journal of Clinical Medicine*, vol. 11, no. 4, p. 903, 2022.

[16] H. M. Deberneh and I. Kim, "Prediction of type 2 diabetes based on machine learning algorithm," *International journal of environmental research and public health*, vol. 18, no. 6, p. 3317, 2021.

[17] N. Fazakis, O. Kocsis, E. Dritsas, S. Alexiou, N. Fakotakis, and K. Moustakas, "Machine learning tools for long-term type 2 diabetes risk prediction," *IEEE Access*, vol. 9, pp. 103 737–103 757, 2021.

[18] Y. Jian, M. Pasquier, A. Sagahyroon, and F. Aloul, "A machine learning approach to predicting diabetes complications," in *Healthcare*, vol. 9, no. 12. MDPI, 2021, p. 1712.

[19] G. Naveen Kishore, V. Rajesh, A. Vamsi Akki Reddy, K. Sumedh, and T. Rajesh Sai Reddy, "Prediction of diabetes using machine learning classification algorithms," *Int J Sci Technol Res*, vol. 9, no. 01, pp. 1805–1808, 2020.

[20] L. C. Jung, H. Wang, X. Li, and C. Wu, "A machine learning method for selection of genetic variants to increase prediction accuracy of type 2 diabetes mellitus using sequencing data," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 13, no. 3, pp. 261–281, 2020.

[21] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," *IEEE Access*, vol. 8, pp. 76 516–76 531, 2020.

[22] M. S. Islam, M. K. Qaraqe, and S. B. Belhaouari, "Early prediction of hemoglobin a1c: A novel framework for better diabetes management," in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2020, pp. 542–547.

[23] L. Kopitar, P. Kocbek, L. Cilar, A. Sheikh, and G. Stiglic, "Early detection of type 2 diabetes mellitus using machine learning-based prediction models," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.

[24] A. Dagliati, S. Marini, L. Sacchi, G. Cogni, M. Teliti, V. Tibollo, P. De Cata, L. Chiovato, and R. Bellazzi, "Machine learning methods to predict diabetes complications," *Journal of diabetes science and technology*, vol. 12, no. 2, pp. 295–302, 2018.

[25] S. Wei, X. Zhao, and C. Miao, "A comprehensive exploration to the machine learning techniques for diabetes identification," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*. IEEE, 2018, pp. 291–295.

[26] Y. Fan, E. Long, L. Cai, Q. Cao, X. Wu, and R. Tong, "Machine learning approaches to predict risks of diabetic complications and poor glycemic control in nonadherent type 2 diabetes," *Frontiers in Pharmacology*, vol. 12, p. 1485, 2021.

[27] C. Fiarni, E. M. Sipayung, and S. Maemunah, "Analysis and prediction of diabetes complication disease using data mining algorithm," *Procedia computer science*, vol. 161, pp. 449–457, 2019.

[28] B. Sudharsan, M. Peeples, and M. Shomali, "Hypoglycemia prediction using machine learning models for patients with type 2 diabetes," *Journal of diabetes science and technology*, vol. 9, no. 1, pp. 86–90, 2014.

[29] B. Vamsi and D. Bhattacharyya. (2021) Micro and macro

vascular complications in type_ii diabetes. [Online]. Available: https://data.mendeley.com/datasets/dsjcb6pyd8/1

[30] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.

[31] A. Al Bataineh and D. Kaur, "Immunocomputing-based approach for optimizing the topologies of lstm networks," *IEEE Access*, vol. 9, pp. 78 993–79 004, 2021.

[32] A. Al Bataineh and S. Manacek, "Mlp-pso hybrid algorithm for heart disease prediction," *Journal of Personalized Medicine*, vol. 12, no. 8, p. 1208, 2022.

[33] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, vol. 398.

[34] J. R. Quinlan, "Probabilistic decision trees," in *Machine Learning*. Elsevier, 1990, pp. 140–152.

[35] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[36] J. Brownlee, *Machine learning algorithms from scratch with python*. Machine Learning Mastery, 2016.

[37] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

Using Incremental Ensemble Learning Techniques to Design Portable Intrusion Detection for Computationally Constraint Systems

Promise R. Agbedanu[1]
African Centre of Excellence in Internet of Things
University of Rwanda
Kigali, Rwanda

Richard Musabe[2]
Rwanda Polytechnic
Kigali, Rwanda

James Rwigema[3]
African Centre of Excellence in Internet of Things
University of Rwanda
Kigali, Rwanda

Ignace Gatare[4]
University of Rwanda
Kigali, Rwanda

Abstract—Computers have evolved over the years, and as the evolution continues, we have been ushered into an era where high-speed internet has made it possible for devices in our homes, hospital, energy, and industry to communicate with each other. This era is known as the Internet of Things (IoT). IoT has several benefits in a country's economy's health, energy, transportation, and agriculture sectors. These enormous benefits, coupled with the computational constraint of IoT devices, make it challenging to deploy enhanced security protocols on them, making IoT devices a target of cyber-attacks. One approach that has been used in traditional computing over the years to fight cyber-attacks is Intrusion Detection System (IDS). However, it is practically impossible to deploy IDS meant for traditional computers in IoT environments because of the computational constraint of these devices. This study proposes a lightweight IDS for IoT devices using an incremental ensemble learning technique. We used Gaussian Naive Bayes and Hoeffding trees to build our incremental ensemble model. The model was then evaluated on the TON IoT dataset. Our proposed model was compared with other proposed state-of-the-art methods and evaluated using the same dataset. The experimental results show that the proposed model achieved an average accuracy of 99.98%. We also evaluated the memory consumption of our model, which showed that our model achieved a lightweight model status of 650.11KB as the highest memory consumption and 122.38KB as the lowest memory consumption.

Keywords—*Cyber-security; ensemble machine learning; incremental machine learning; Internet of Things; intrusion detection; online machine learning*

I. INTRODUCTION

As the evolution of computing technology continues, the ability of things, such as fridges, air-conditioners, medical equipments, and meters among others to communicate has become a reality due to fast communication technologies. A paradigm popularly known as the Internet of Things (IoT) has not only become a household term with smart homes, but it also has numerous uses in energy, agriculture, manufacturing, healthcare, and transportation. There is no doubt that the IoT has many benefits, which is why the number of IoT devices is growing exponentially. According to [1], the number of

IoT devices is estimated to reach 30.9 billion by 2025. The numerous benefits of the IoT ecosystem make it attractive to cyber-attacks. An attack statistic presented by SAM Seamless Network shows that over 1 billion IoT-based attacks happened in 2021 [2]. Although methodologies, such as encryption and secured architecture are progressively being deployed to ensure that IoT devices are secured, the computational constraint of these devices makes it challenging to implement these security measures to their fullest potential. Another approach to securing these devices from cyber-attacks is to detect these attacks before an attacker exploits them. Intrusion Detection Systems have been around for more than four decades, with the development of these IDSs focused on traditional computing systems [3]. They have been among the primary methodologies used to protect computer networks. Vacca [4] defines intrusion detection as the process of detecting activities perpetrated against computer systems by intruders. Over the past 40 years, many breakthroughs have been made in intrusion detection. One of the most significant breakthroughs in this area is the use of machine learning in detecting intrusions. However, with all these breakthroughs, it is practically impossible to deploy traditional computing-based IDS methods in the IoT. The impossibility of deploying these IDSs in IoT systems has been created because of the computational constraints of IoT devices. The constraints have led to several studies being carried out to design IDSs that can be deployed in IoT systems without significantly affecting the computational resources of these devices. Several approaches have been proposed in designing lightweight IDSs for IoT environments. However, these studies fail to either report how these lightweight IDSs are achieved or how much computational resources these proposed approaches consume. For example, [5]–[9] proposed various techniques that are supposed to translate into lightweight IDSs. However, these works either failed to report how these methods translate into lightweight IDS or how much computational resources these proposed methods consume. This study proposes a novel lightweight intrusion detection system using an incremental machine learning approach. The main contributions of this study are as follows:

- Using an incremental machine learning approach to design a lightweight IoT intrusion detection system.

- Measuring the memory consumption of our proposed model.

- The study uses an incremental ensemble approach to achieve improved accuracy.

- The study evaluates the proposed IDS model on an IoT dataset.

The remainder of the paper is structured as follows. Section II discusses the study's background. Section III focuses on works relevant to our study, whereas Sections IV, V and VI focus on the proposed model, experimental evaluation, and conclusion, respectively.

II. BACKGROUND

A. Intrusion Detection System

An intrusion detection system (IDS) is a security device that detects illegal access to data within a networked or computer-based environment in order to threaten the integrity, availability, or confidentiality of the computing device [10], [11]. An IDS's objective is to continuously monitor network traffic and flag any activity that violates the normal usage of the system [12]. According to [13], typically, an IDS consists of sensors, an analysis engine, and some reporting system. Intrusion detection systems can be classified either on how they are deployed or detect illegal activities. From a deployment perspective, an IDS can be classified as distributed, centralized, or hybrid. On the other hand, an IDS can be classified as signature-based, anomaly-based, specification-based, or hybrid. According to [14], signature-based detection is the set of pre-defined rules, such as the sequence of bytes in network traffic that are pre-loaded to trigger an alert when a matched sequence is detected. On the other hand, anomaly detection records the normal behavior of a network and then compares them with the system's current behavior. The authors also explained the specification-based detection method as an approach that uses input specifications designed manually. Finally, hybrid detection methods deploy a combination of signature-based, anomaly-based, and specification-based detection methods to improve accuracy and reduce false positive rates.

1) Ensemble Learning: According to [15], ensemble learning is a machine learning technique that combines the strengths of different machine learning algorithms into a single algorithm. The primary goal of ensemble learning is to improve accuracy by leveraging the strengths of the ensemble learners [16]. There are instances where traditional machine learning models do not achieve high accuracy [17]. Several ensemble-based techniques have been developed over time, but the most popular are bagging, boosting, stacking generalization, and expert mixture [16].

In the preceding paragraphs, we briefly explain the three categories of ensemble learning.

2) Bagging-based Learning: Bagging, short for bootstrap aggregation, is an algorithm that is best suited for problems with a small training dataset. Given a training set S with a cardinality n, the bagging algorithm trains several independent classifiers T. Each of these classifiers are trained using a percentage of N [16] sampling. Linear classifiers such as linear SVM, decision stumps, and single-layer perceptrons are excellent candidates for bagging [16]. Classifiers are trained and then combined using simple majority voting in bagging. Bagging, an abbreviation for bootstrap aggregation, is a method that works well with issues that have a limited training dataset. The bagging algorithm learns several independent classifiers T given a training set S with a cardinality of n. Each of these classifiers is trained using a proportion of N sampling. Linear classifiers like linear SVM, decision stumps, and single-layer perceptrons are great candidates for bagging [16]. In bagging, classifiers are trained and then concatenated using simple majority voting.

3) Boosting-based Learning: An iterative approach can be used to generate a robust classifier from a set of weak classifiers. Although boosting also combines a large number of weak learners through simple majority voting, there is one significant difference between boosting and bagging. Every instance in bagging has an equal chance of being in each dataset used in training. In boosting, on the other hand, the dataset used to train each subsequent model focuses on instances misclassified by the previous model. At any given time, a boosting designed for a binary class problem generates a set of three weak classifiers. The first learning classifier is trained on a random subset of the training data available. A different subset of the original training dataset is used to train the second learning classifier [18].

4) Stack Generalization: Non-trainable combiners are used in bagging and boosting methods. The combination weights in non-trainable combiners are determined after the classifiers have been trained. The combination rule used in non-trainable combiners does not allow determining which member classifier learned from which partition of the feature space [16]. Trainable combiners can be used to solve this problem, and individual ensemble members can be combined using a separate classifier in stacked generalization.

5) Mixture of Experts: A sampling technique is used to train an ensemble of classifiers in a mixture of experts. The classifiers are then combined using a weighted combination rule [19]. Furthermore, a mixture of experts can encompass the selection of algorithms, with each classifier trained to become an expert in a different aspect of the feature space. Individual classifiers are usually not weak since they are trained to become experts.

B. Online/Incremental Machine Learning

Online machine learning, also known as incremental machine learning, is increasingly becoming popular in real-time data streams. According to [20], online algorithms instantaneously build machine learning models after seeing a small portion of the data. This characteristic leads to the inability to undo less optimal decisions made earlier because the data will no longer be available for the algorithm. The concept of machine learning models acquiring knowledge from continuous data without accessing the original data has been applied to domains like intelligent robots, auto-driving, and unmanned aerial vehicles [21]–[23]. According to [24], as reported by [20],, in data stream models infinite stream of data arrives

continuously, and these streams of data are to be processed by systems with resource constraints. The main restriction of data stream models is that memory of these systems are usually small and can only hold a minimal portion of the data stream. Regarding data stream models, only a minimal subset of the data can be kept for instant data analysis [25]. Fig. 1 shows an online machine-learning model using an offline dataset, while Fig. 2 shows the same model using streams of network traffic.

III. Related Work

Throughout this section, we will look at some studies that are relevant to our work. Yang et al. [26] proposed an ensemble framework for intrusion detection systems (IDSs) in IoT environments, focusing primarily on idea drift adaptability. The suggested framework employs a performance-weighted probability averaging ensemble to manage concept drift in IoT anomaly detection. When compared to other cutting-edge approaches, the suggested framework performed better. Even though the author's proposed method took less time to run than the other methods they looked at for their study, they did not look into how the proposed model affected other computing parameters, such as memory.

Jan et al. [5] used a supervised Support Vector Machine (SVM) to detect IoT adversarial attacks. The authors used only the sensor node's packet arrival rate to design the proposed IDS. The accuracy of the proposed IDS showed better performance compared to other models like neural networks, KNN, and decision trees. One of the drawbacks of this study is that it only considered DDoS attacks. Additionally, the authors failed to report how the proposed approach leads to a lightweight IDS. Parameters such as memory consumption and model running time were not reported.

In a similar study, [27] proposed a lightweight IDS for IoT ecosystems using a Deep Belief Network and Genetic Algorithm. According to the study, the proposed system was more accurate than other methods that were looked at for the study. However, the study failed to report how the proposed approach translates to a lightweight model. Moreover, the dataset used for the experimental validation is not an IoT-based dataset. Like in other studies, parameters that are supposed to prove the lightweight status of the proposed method were not considered in the study.

Roy et al. [7] also designed a lightweight IDS for IoT systems using a set of optimization techniques. They used multicollinearity, sampling, and dimensionality reduction to reduce the training data, which resulted in a shorter training time. Like other earlier related works considered in this section, although their proposed approach reduces the training time of the model, the study did not report how much memory the model consumes.

Zhao et al. [8] suggested a network intrusion detection method for IoT devices utilizing a lightweight neural network. To minimize the dimensionality of features, the authors employed a principal component analysis approach. The proposed method was tested using the UNSW-NB15 and Bot-IoT datasets. Although the authors determined that both the ultralight feature extraction network and principal component analysis contributed to the suggested model's lightweight per-

formance, they did not report on the computational complexity of their proposed method.

In order to make a lightweight IDS for IoT systems, [9] said that they used a mix of feature selection techniques on different datasets to make a lightweight IDS algorithm for IoT traffic. However, two datasets used to evaluate their proposed lightweight IDS were non-IoT related. Also, the authors did not talk about how their proposed model would affect the computing power in their experimental environment.

Latif et al. [6] reported using a Dense Random Neural Network to develop a lightweight intrusion detection system for IoT environments. The proposed model was evaluated on the ToN-IoT dataset, and the results show a detection accuracy of 99.14% for binary class classification and 99.05% for multiclass classifications. However, Latif et al. did not report on the computational complexity of their proposed model. A parameter is required to measure the lightness of the proposed model.

Pan et al. [28] also suggested a lightweight, intelligent intrusion detection system (IDS) architecture for wireless sensor networks. The authors used KNN and the sine cosine technique to create their model. The authors reported that combining the above techniques improves classification accuracy and reduces false alarms. However, the authors failed to report how the lightweight model was achieved or what parameters were used to determine the lightweight status of the model.

Reis et al. [29] created an IDS for cyber-physical systems using incremental support vector machines. In their study, a one-class support vector machine was applied to each sensor to retrieve abnormal behaviors. These anomalies are orchestrated as an output of the proposed incremental machine learning model. Although the model proposed by Reis et al. achieved an accuracy higher than 95%, the study did not detail how the proposed method would affect the computational resources of cyber-physical systems.

To reduce the computation overhead, [30], introduced a privacy-preserving pipeline-based intrusion detection for distributed incremental learning that selects unique features using an innovative extraction technique. Current incremental learning techniques are computationally expensive, and the distributed intrusion detection method is used to distribute the load across IoT and edge devices. Theoretical analysis and experiments show that state-of-the-art techniques require less space and time. The study, however, reported on time complexity but not space complexity. Furthermore, the experimental validation dataset is not an IoT-based dataset.

IV. Proposed Model

This section details the design and conceptual implementation of our suggested approach. The suggested model is based on a machine learning technique, ensuring the creation of a lightweight IDS model suitable for the IoT environment. The proposed model employs incremental machine learning and data streaming ensemble learning approaches to create a lightweight intrusion detection system for the IoT environment. The proposed model processes network data generated in IoT environments as data streams and train each data stream. After each iteration, the model is updated. Fig. 3 depicts our proposed model.
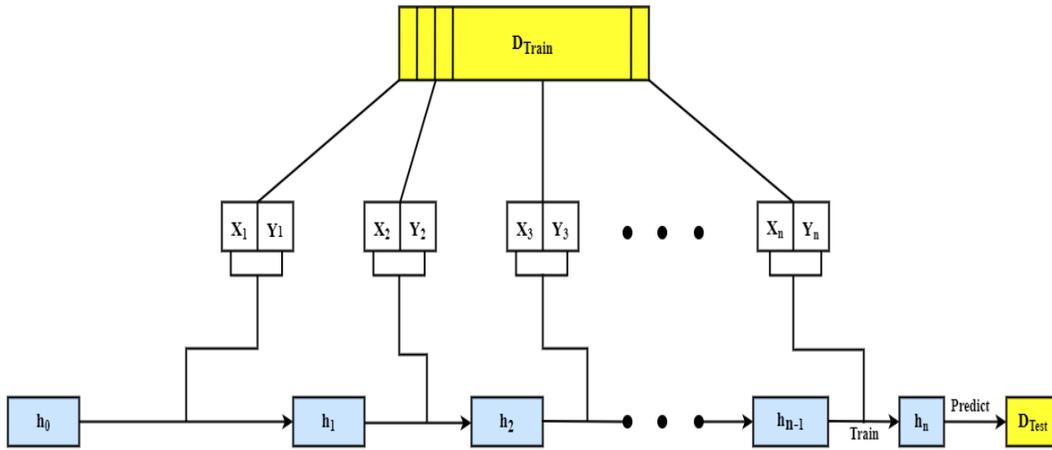
Fig. 1. Online Machine Learning using Offline Dataset.
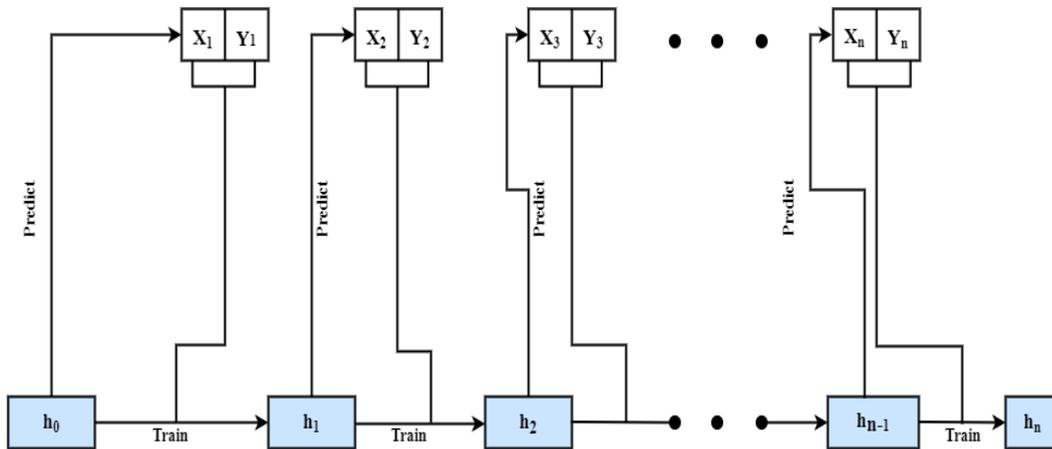
Fig. 2. Online Machine Learning using Data Stream.

1) Pre-Processing: The dataset used to train our proposed model is cleaned at this stage. The pre-processing data approach used in this study includes imputing missing data values and transforming and selecting important features to train our machine learning model. We employed one-hot encoding as one of the techniques to pre-process our data. A single hot encoding transformer will encode all the features provided to it. If a list or set is supplied, this transformer will encode each item in the list or set by composing it with compose.Select command in River; the encoding can apply to a subset of features.

2) Model Training: This study proposes a novel online stacking ensemble machine learning technique using Gaussian Naive Bayes and Hoeffding Tree Classifier. We chose these two machine learning models to build our ensemble learning because we want to achieve the following three objectives

- Design a model that consumes a minimal computational resource (lightweight).
- Building a fast model.
- A model that achieves a high accuracy.

Gaussian Naive Bayes and Hoeffding Tree Classifier are used as the base classifiers of our proposed model; whiles Hoeffding Tree is used as the meta classifier of the proposed model. Each observation of the dataset is read as a stream and is then used to train the base and meta classifier. Each base classifier predicts each stream of data, that is, $X_i Y_i$ which becomes feature input to the meta classifier. The meta classifier (HT) then uses the outputs of the base classifiers to make a better prediction. We chose Hoeffding trees because they learn patterns in data without continuously storing data samples for future reprocessing, and this makes them particularly suitable for use on embedded devices. Similarly, Gaussian NB is quick and flexible and produces highly reliable results. It works well with large amounts of data and requires little training time, and it also improves grading performance by removing insignificant specifications [31], [32].

3) Model Evaluation: The final stage of the model is the model evaluation stage. The proposed model's accuracy, precision, recall, F1, model training time, and memory consumption are all evaluated.

We chose incremental and ensemble machine learning to develop our framework in this work because of the the following

benefits.

1) Network traffic is generated in blocks as a data stream. By using incremental learning on network traffic, models can predict the nature of traffic without having to be trained on large datasets.
2) The computational constraints of IoT devices make loading an entire training dataset into main memory difficult and impractical. Even if the entire training data can fit into the main memory of an IoT device, the device's computational power will be drastically reduced.
3) New data is constantly available because of the sophisticated nature of cyber-attacks. Retraining the model on the entire dataset will be time-consuming and computationally expensive. Because models in online machine learning are trained with data streams, they can quickly learn from new data examples without consuming much computational power.
4) Real-time network traffic is generated, which must be analyzed in real-time to prevent intruders from gaining unauthorized access to devices. Online machine learning has proven to be an effective learning method in real-time environments.
5) Traffic flow is dynamic and constantly changing. Changes in network traffic can impact the predictive performance of machine learning models, referred to as concept drift in machine learning. Models should be able to self-adapt to changes in the relationship between input and output data to handle concept drifts.
6) Most of the time, ensemble methods have produced higher accuracy than the individual models that were used to make them.

Additionally, the use of the above method poses the following limitation.

1) Getting a good tradeoff between accuracy, speed, and minimal resource consumption is going to be a challenge because combining models increases the computational consumption of the final output.

A. Gaussian Naive Bayes

According to [33], the Naive Bayes algorithm is a typical illustration of how generative hypotheses and parameter guesses can facilitate learning. Consider the problem of predicting a label y $\in \{0,1\}$ from a vector of characteristics $\mathbf{X} = (x_1,..., x_d)$, where each x_i is in the range of $\{0,1\}$. The optimal classifier of Bayes is given below

$$h_{Bayes}(\mathbf{X}) = argmaxP[Y = y|X = x], y \in \{0,1\} \quad (1)$$

We need 2^d parameters to define the probability function P[Y = y—X = x], each of which relates to P[Y = 1—X = x] for a given value of y $\in \{0,1\}^d$. This means that when the number of features increases, so does the number of instances necessary. In the Naive Bayes technique, we make the generative assumption that, given the label, the features are independent of one another. To put it another way,

$$P[Y = y|X = x] = \prod_{i=1}^{d} P[Y = y|X = x] \quad (2)$$

The Bayes optimum classifier can be reduced further using this assumption and the Bayes rule:

$$h_{Bayes}(\mathbf{X}) = argmaxP[Y = y] \prod_{i=1}^{d} P[X_i = x_i|Y = y] \quad (3)$$

That is, the set of parameters to estimate has been reduced to 2d +1. In this situation, the generative assumption we made considerably decreased the number of parameters we needed to learn. When the maximum likelihood principle is used to determine out the parameters, the resulting classification model is called the Naive Bayes classifier.

One typical technique to handle continuous attributes in Naive Bayes classification is to use Gaussian distributions to express the probabilities of the features based on the classes. As a result, every attribute is represented as Xi N(, 2) by a Gaussian probability density function (PDF), [34] as reported by [35].

$$X_i \sim N(\mu, \sigma^2) \quad (4)$$

The Gaussian PDF is shaped like a bell and is defined by the equation below where μ is the mean and σ^2 is the variance.

$$N(\mu, \sigma^2)(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2a^2}} \quad (5)$$

B. Hoeffding Tree (HT)

Hulten et al [36] are the first to propose Hoeffding trees. The Hoeffding tree algorithm is a fundamental algorithm for stream data classification. It is an induction of a decision tree algorithm that could learn from enormous data streams if the distributed generating examples remain constant over time. It creates decision trees that are similar to the standard batch learning method. Asymptotically, Hoeffding trees as well as decision trees are connected. The HT technique is based on the basic premise that a modest sample size can frequently be sufficient to identify an optimal splitting feature. The key point to understand here is that classic batch learning algorithms produce decision trees based on attribute splitting. The HT method is mathematically verified to use the Hoeffding bound. To comprehend the significance of the Hoeffding bound, a few assumptions must be made. Let's say we get *N* separate samples of a random variable *r* with a range of *R*, where *r* is a measure of attribute selection. In the case of Hoeffding trees, *r* is information gain, and if we calculate the mean value of r_{mean} for this sample, the Hoeffding limit indicates that the true mean of *r* is at least 1-δ. The primary benefits of the HT algorithm are as follows:

1) it is incremental in nature
2) it achieves high accuracy with small sample size.
3) scans on the same data are never performed.
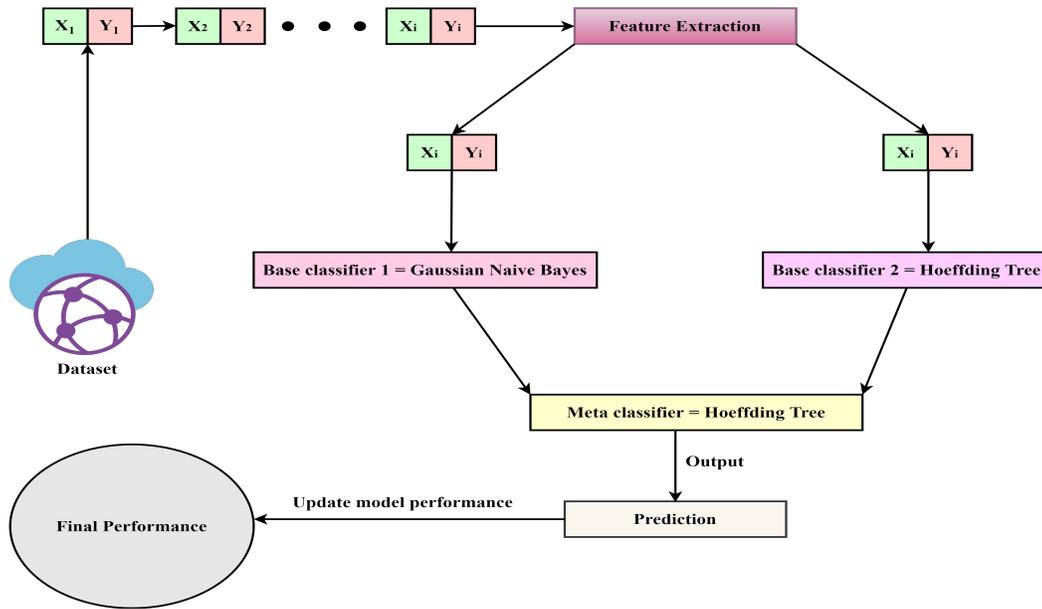
Fig. 3. Our Proposed Model.

However, Hoeffding Tree has a few disadvantages. The main disadvantage is that HT cannot handle concept drift because the node cannot be changed once it is created. Wang et al [37] described how to deal with concept drift using classifiers. The algorithm devotes a significant amount of time to attributes with nearly identical splitting quality. Furthermore, memory utilization can be further optimized.

$$[!h] \in = \sqrt{\frac{R^2 1 n \frac{1}{\delta}}{2n}} \qquad (6)$$

The algorithm for Hoeffding tree is shown in Algorithm 1.

V. EXPERIMENTAL EVALUATION

A. Experimental Environment

The proposed method was implemented using Python 3.8 with River as our framework for online machine learning. The proposed method was implemented on a MacBook Pro with an M1 chip with 16 GB of RAM. The TON-IoT dataset was used to evaluate the proposed framework. There are several incremental or online streaming libraries that provide machine functionalities. Some of these libraries are Creme, scikit-multiflow, and River. In this study, we choose to build our incremental learning models using River. According to [38], River is a merger of creme and Scikit-multiflow. River is a library that allows continual learning by handling dynamic data streams. We chose River because it includes data transformation methods, learning algorithms, and optimization algorithms. Its distinct data structure lends itself well to streaming data and web application settings.

B. Dataset

According to [39], the TON-IoT dataset was built by the Cyber Range and IoT Las at the University of South Wales.

The dataset has nine (9) types of cyber-attacks. These are Denial of Service (DoS), Distributed Denial of Service (DDoS), ransomware, backdoor, data injection, scanning, Cross-site Scripting (XSS), password cracking, and Man-in-The-Middle (MiTM). The generated data were from seven IoT and IIoT devices: fridge, motion light, garage door, GPS tracker, thermostat, and weather. The fridge dataset has a total of 587076 records; the motion light dataset has 452262 records; the garage door has 591446 records; the GPS tracker produced 595686 records, whiles the thermostat and weather produced 442228 and 650242 records, respectively. The statistics of the dataset used are shown in Tables I and II below.

C. Evaluation Metrics

True positive (V_P): Positive intrusion that is both expected and confirmed.

False positive (U_P): An intrusion that was expected to be positive but ended up turning out to be negative.

True negative (V_N): The intrusion is expected to be negative and confirmed to be negative.

False negative (U_N): The intrusion was expected to be negative, but it turned out to be positive.

1) Accuracy: A model's overall accuracy can be measured by the number of correctly predicted events made by the given model. The formula below computes the total accuracy of the model.

$$Accuracy = \frac{V_P + V_N}{V_P + V_N + U_P + U_N}$$

2) Precision: Precision is found by dividing the total number of positive detections by the number of positive detections that were correctly identified as positive.

Algorithm 1: Hoeffding Tree Algorithm [36]

Input: S is a sequence of examples,
    X is a set of discrete attributes,
    G(.) is a split evaluation function,
    δ is one minus the desired probability of
    choosing the correct attribute at any given
    node

Output: *HT* is a decision tree

  procedure HOEFFDINGTREE(S, X, G, δ)

    Let *HT* be a tree with a single leaf l_1 (the root)

    Let $X_1 = X \cup \{X_\theta\}$

    Let $\bar{G}_1(X_\theta)$ be the $\bar{G}$ obtained by predicting the most frequent class in S

    for *each class Y_k* **do**

      for *each value X_{ij} of each attribute $X_i \in \mathbf{X}$*

    do

        Let $n_{ijk}(l_1) = 0$

      end

    end

    for *each example (X, Y_k) in S* **do**

      Sort (x, y) into a leaf l using HT

      for *each X_{ij} in X ssuch that $X_i \in \mathbf{X}$* **do**

        Increment $n_{ijk}(l)$

      end

    end

    Label *l* with the majority class among the examples seen so far at *l*

    if *the examples seen so far at l are not all of the same class* **then**

    end

    Compute $\bar{G}_l(X_i)$ for each attribute $X_i \in X_l - X_\theta$ using the counts $n_{ijk}(l)$

    Let X_a be the attribute with highest $\bar{G}_l$.

    Let X_b be the attribute with second-highest $\bar{G}_l$.

    Compute $\in$ using Equation 1

    if $\bar{G}_l(X_a) - \bar{G}_l(X_b) > \in$ *and* $X_a \neq X_\theta$, **then**

    end

    Replace *l* by an internal node that splits on X_a.

    For each branch of the split

    Add a new leaf l_m and let $X_m = X - \{X_a\}$.

    Let $G_m(X_\theta)$ be the G obtained by predicting the most frequent class at *l*

    for *each class Y_k and each value X_{ij} of each attribute $X_i \in X_m - \{X_\theta\}$* **do**

    Let $n_{ijk}(lm) = 0$

    end

    return *HT*

  end procedure

TABLE I. STATISTICS OF TON IoT DATASET [39]

Fridge IoT dataset	
Type of attack	**No of rows**
Backdoor	35568
DDoS	10233
Injection	7079
Normal	500827
Password	28425
Ransomware	2902
XSS	2042
GPS tracker IoT dataset	
Backdoor	35571
DDoS	10226
Injection	6904
Normal	513849
Password	513849
Ransomware	2833
Scanning	550
XSS	577
Motion light IoT dataset	
Backdoor	28209
DDoS	8121
Injection	5595
Normal	388328
Password	17521
Ransomware	2264
Scanning	1775
XSS	449
Weather IoT dataset	
Backdoor	35641
DDoS	15182
Injection	9726
Normal	559718
Password	25715
Ransomware	2865
Scanning	529
XSS	866
Garage IoT dataset	
Backdoor	35568
DDoS	10230
Injection	6331
Normal	515443
Password	19287
Ransomware	2902
Scanning	529
XSS	1156

TABLE II. STATISTICS OF TON IoT DATASET CONTINUATION [39]

Modus IoT dataset	
Backdoor	40035
Injection	7079
Normal	405904
Password	24269
Scanning	529
XSS	577
Thermostat IoT dataset	
Backdoor	35568
DDoS	10230
Injection	6331
Normal	515443
Password	19287
Ransomware	2902
Scanning	529
XSS	1156

$$Precision = \frac{V_P}{V_P + U_P}$$

3) Recall: The recall is defined as the ratio of true positive detections to the number of real abnormal samples.

$$Recall = \frac{V_P}{V_P + U_N}$$

4) F1 Score: The F1 score is the average of precision and recall. The F1-score is determined as the weighted average of precision and recall, taking both the U_P and U_N into consideration.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * V_P}{2 * V_P + U_P + U_N}$$

5) Memory: The computational constraint of IoT devices makes it difficult and sometimes impossible to run IDS meant for traditional computers on these IoT devices. This calls for developing models that consume minimal memory (lightweight models).

6) Model Running Time: In this evaluation metric, we measure the total time it takes for the proposed model to run.

D. Results

In this section, we present the proposed results and compare them with other state-of-the-art techniques. To begin with, this study compares the results of the proposed model with other state-of-the-art IDS proposed and evaluated with the TON IoT dataset. We decided to limit the state-of-the-art studies that used ToN IoT dataset because we wanted to eliminate biases. In comparing these studies, we considered the category of the TON IoT dataset used in each study, the method proposed by each of the works under consideration, the highest accuracy recorded, and whether the study records the time used to build the model as well as the amount of memory the model consumes. The comparison of our approach with other state-of-the-art IDS for IoT systems is presented in table III. When the authors fail to report a parameter, we indicate it as non-available (N/A). Table III shows that out of the five state-of-the-art IDSs considered in the study, none of them reports on the model or the memory consumption of their proposed technique. Although [40], [41] both report 100% accuracy, our proposed model outperforms the methods proposed in those studies for the following reasons:

1) The accuracy reported in our study is the average accuracy of our proposed model whereas [40], [41] report total accuracy.
2) Our study focused on multi-class classification, whereas [40], [41] focused on binary-class classification.

Table IV compares the accuracy, time, and memory consumption of the models used to build our incremental ensemble technique with our proposed model. Although the time and memory consumption of the individual models are lower than our proposed model, we wanted to propose a model that achieves a trade-off between accuracy, time, and memory consumption. Our proposed model achieved a higher accuracy without significantly increasing the time and memory consumption. The time and memory consumption of the proposed model shows it can run on computationally constrained devices without negatively impacting the computational resources of these devices.

Fig. 4 below shows the output of our proposed model in terms of the time taken to build the model. The results show that our proposed model recorded the least training time of 59 seconds on the thermostat dataset and the highest training time of 114 seconds on the weather IoT dataset. The concept of incremental learning allows our model to learn one stream of data at a time. Therefore, the time used to train the model on a stream of data will be the total observations in the dataset divided by the total model training time. This makes our model very fast irrespective of the size of the dataset.
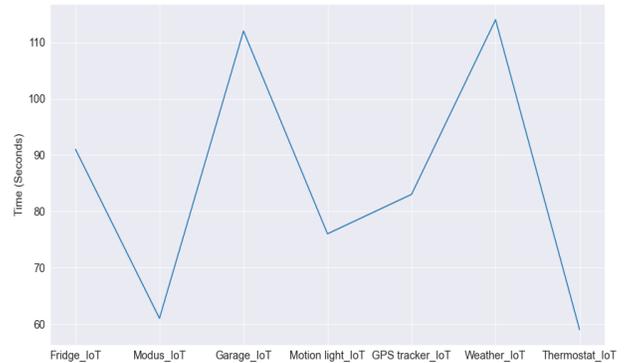
Fig. 4. Model Training Time of our Model using the TON IoT Dataset.

When tested on the modus dataset, our proposed model achieved a superior average accuracy of 96.81%, with precision, recall, and F1 scores of 97.23%, 96.81%, and 96.92%, respectively. The Hoeffding tree had an average accuracy of 92.96%, precision, recall, and F1 scores of 92.36%, 92.36%, and 92.36%, respectively. Using the GPS IoT dataset, the Gaussian NB had an average accuracy of 77.60% and precision, recall, and F1 scores of 60.21%, 77.60%, and 67.81%, respectively. Fig. 5 shows the results of our model when evaluated using the modus IoT dataset.
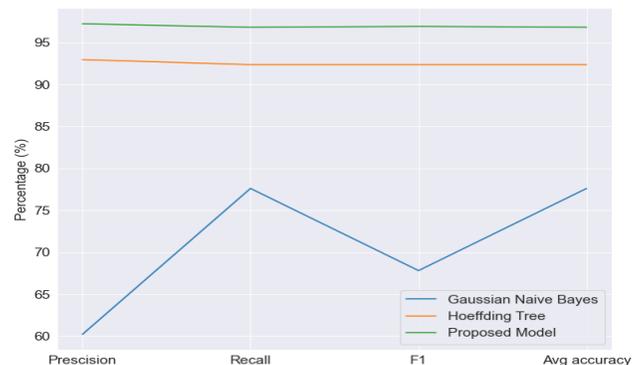
Fig. 5. Accuracy of Gaussian NB, HT and Proposed Model on Modus Dataset.

Similarly, our proposed model has a superior average accuracy of 99.98% when it was evaluated using the fridge IoT dataset. The proposed model also recorded the same precision, recall, and F1 score value using the same dataset. Hoeffding tree algorithm recorded 98.63%, 98.68%, 98.52%, and 98.52% for precision, recall, F1 score, and average accuracy, respectively. Gaussian NB recorded the lowest average accuracy, an average accuracy of 85.31%. Gaussian NB also recorded the least values for precision, recall, and F1 scores, with 72.8%, 85.31%, and 78.55%, respectively. Fig. 6 illustrates the outcomes of our model when tested against the fridge IoT dataset.

TABLE III. USING THE TON IOT DATASET, WE COMPARED OUR PROPOSED MODEL TO STATE-OF-THE-ART MODELS THAT HAD BEEN TESTED USING THE SAME DATASET

Study	Year of the study	Method used	Highest accuracy	Model training time (S)	Memory consumption (KB)
[6]	2021	Dense Random Neural Network	99.14%	N/A	N/A
[40]	2022	Optimized decision tree	100	N/A	N/A
[41]	2022	Ensemble based voting	100	N/A	N/A
[42]	2022	Graph Neural Network	97.87	N/A	N/A
[43]	2021	Synthetic minority oversampling technique	99.0	N/A	N/A
Our proposed model	2022	Stack-based Incremental ensemble (HT and Gaussian NB)	99.98	71	122.38

TABLE IV. COMPARING THE ACCURACY (ACC), MODEL TIME CONSUMPTION (TIME) AND MEMORY USAGE OF THE BASE CLASSIFIERS AGAINST OUR MODEL ON THE DIFFERENT DATASETS

Fridge IoT dataset								
Gaussian NB			Hoeffding Tree			Our proposed model		
Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)
85.31%	29.9	15.85	98.68%	20.7	532.71	99.98%	104	650.11
Modus IoT dataset								
Gaussian NB			Hoeffding Tree			Our proposed model		
Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)
77.60%	19.2	19.2	92.36%	14	124.58	96.81%	75	495.25
Garage IoT dataset								
Gaussian NB			Hoeffding Tree			Our proposed model		
Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)
85.96%	37	27.05	95.70%	29.5	75.85	99.96%	131	394.95
Motion light IoT dataset								
Gaussian NB			Hoeffding Tree			Our proposed model		
Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)
85.86%	22.9	20.6	92.06%	15.5	33.56	99.98%	79	219.58
GPS Tracker IoT dataset								
Gaussian NB			Hoeffding Tree			Our proposed model		
Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)
85.20%	28.3	10.56	98.29%	18.7	120.36	99.97%	97	281.94
Weather IoT dataset								
Gaussian NB			Hoeffding Tree			Our proposed model		
Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)
86.08%	33.4	20.58	98.37%	23.7	314.91	99.93%	116	627.81
Thermostat IoT dataset								
Gaussian NB			Hoeffding Tree			Our proposed model		
Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)	Acc	Time (s)	Memory (KB)
87.27%	19.9	6.85	99.12%	13.8	55.07	99.94%	71	122.38

The experimental findings show that our proposed method performed better when tested using the motion IoT dataset. The proposed ensemble model achieved an average accuracy of 99.98% with precision and recall of the same value whiles recording 99.97% for the F1 score. The same dataset revealed that the Hoeffding tree recorded an average accuracy of 92.06% whiles recording a precision, recall, and F1 score of 88.56%, 92.06%, and 89.64%, respectively. The average accuracy, precision, recall, and F1 score recorded by Gaussian NB is 85.86%, 73.73%, 85.86%, and 79.33%, respectively. Fig. 7 depicts the results of our model when tested against the motion IoT dataset.

When tested on the garage IoT dataset, our proposed ensemble model again had the highest precision, recall, F1 score, and average accuracy. Our proposed model recorded an average accuracy of 99.96%, with the same value recorded for precision, recall, and F1 score. Hoeffding tree, on the other hand, recorded a precision, recall, F1 score, and average accuracy of 95.52%, 95.70%, 95.26% and 95.70% respectively. Gaussian Naive Bayes recorded a precision, recall, F1 score, and average accuracy of 73.88%, 85.96%, 79.46%, and 85.96 respectively when it was evaluated using the garage IoT dataset. Fig. 8 shows the results of our model when tested against the garage IoT dataset.

Evaluating our proposed model on the GPS tracker dataset, our model achieved a superior average accuracy of 99.97% with precision, recall, and F1 score of 99.97% each. Hoeffding tree recorded an average accuracy of 98.29% whiles recording a precision, recall, and F1 score of 98.36%, 98.29% and 98.08%, respectively. Evaluating the Gaussian NB using the GPS IoT dataset revealed an average accuracy of 85.20% with precision, recall, and F1 score of 88.26%, 85.20%, and 82.02%, respectively. Fig. 9 shows the results of our model when evaluated using the GPS tracker IoT dataset.
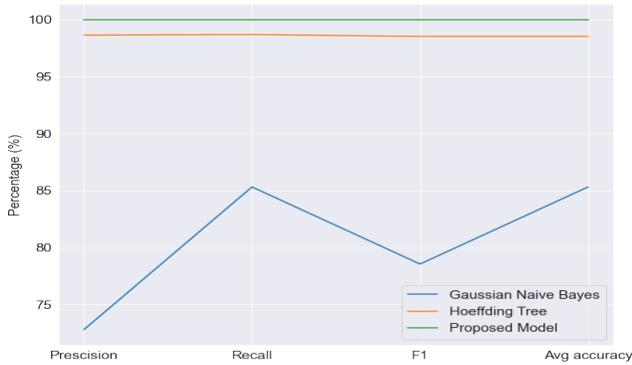
Fig. 6. Accuracy of Gaussian NB, HT and Proposed Model on Fridge Dataset.
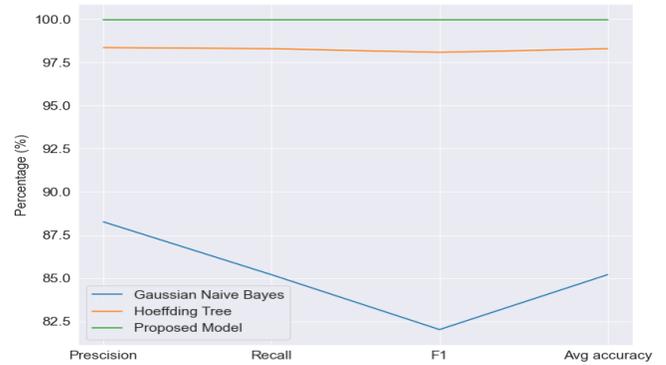
Fig. 9. Accuracy of Gaussian NB, HT and Proposed Model on GPS Tracker Dataset.
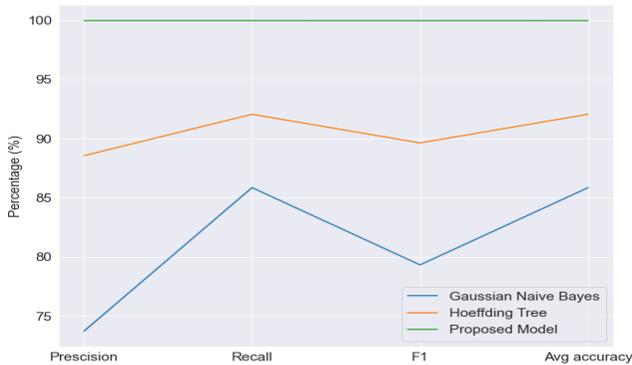
Fig. 7. Accuracy of Gaussian NB, HT and Proposed Model on Motion Dataset.
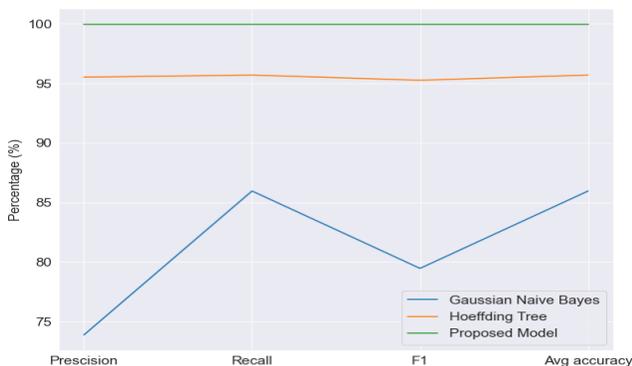
Fig. 8. Accuracy of Gaussian NB, HT and Proposed Model on Garage Dataset.

model achieved an average accuracy of 99.94% with precision, recall, and F1 score of 99.94% for each of them, respectively. Gaussian NB showed an average accuracy of 87.27% whiles recording a precision, recall, and F1 score of 76.17%, 87.27%, and 81.34%, respectively. Hoeffding tree showed an average accuracy of 99.12% with precision, recall, and F1 score of 99.07%, 99.12% and 99.03%, respectively. Fig. 10 shows the results of our model when tested against the thermostat IoT dataset.
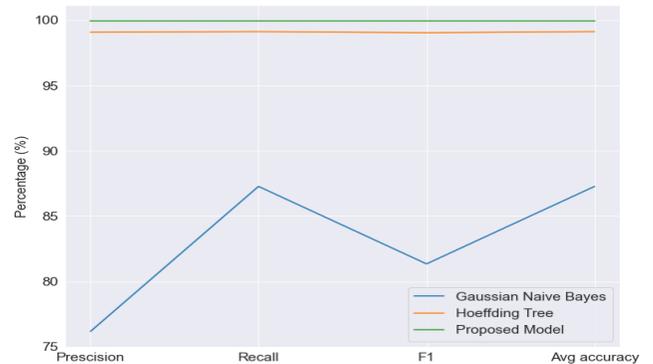
Fig. 10. Accuracy of Gaussian NB, HT and Proposed Model on the Thermostat Dataset.

We also evaluated our model on the weather IoT dataset, one of the datasets found in the TON IoT dataset. The results show that Gaussian NB recorded precision, recall, F1, and average accuracy of 80.02%, 86.08%, 76.65%, and 86.08%, respectively. On the other hand, the Hoeffding tree recorded precision, recall, F1, and average accuracy of 98.47%, 98.37%, 98.30%, and 98.37%, respectively. However, our proposed ensemble technique recorded an average accuracy of 99.93% with precision, recall, and F1 score of 99.93%, respectively. Fig. 11 illustrates the outcomes of our model when tested against the weather IoT dataset.

The experimental result shows that when our proposed model is evaluated using the thermostat IoT dataset, the

The Fridge IoT dataset recorded the highest consumption of

Fig. 11. Accuracy of Gaussian NB, HT and Proposed Model on Weather Dataset.

650.11KB, whiles the weather IoT dataset recorded the lowest memory consumption of 122.38KB. The results of the memory consumption of the model proposed in this study show that even at the highest memory consumption, the proposed IDS achieves a lightweight status and can potentially run on IoT devices without significantly affecting the available memory of these devices. The memory consumption of our proposed model is shown in Fig. 12.

The precision, recall, and F1 score of Gaussian Naive Bayes, Hoeffding tree, and our proposed model on different attack categories are shown in Tables V and VI below.
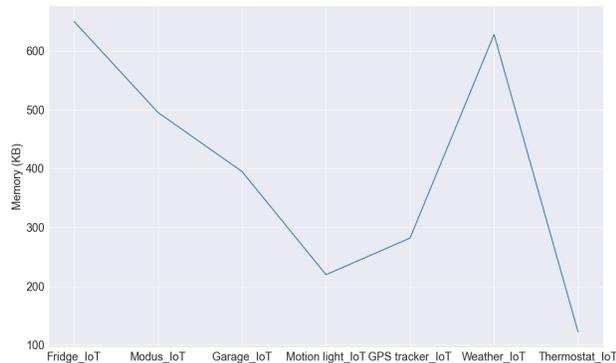
Fig. 12. Memory Consumption of our Proposed Model on Various Sub-Datasets of TON IoT Dataset

VI. LIMITATION OF THE STUDY

The main limitation of the study is how the proposed method can be used to achieve good accuracy, higher detection speed, and lower resource consumption at the same time. One approach that can be used to overcome this limitation is to deploy the proposed model on a resource-constrained device while fine-tuning the model to achieve a good tradeoff among the parameters mentioned above.

VII. CONCLUSION

The security of the IoT ecosystems is increasingly gaining significant importance due to its numerous applications. The security of IoT systems has gone beyond encryption, authentication, and secured architecture. Recently, much security-based research in IoT systems has been focused on detecting attacks and anomalies in network traffic. However, because of the computational constraints of IoT devices, IDS developed for traditional computing systems cannot be deployed in IoT environments. It is, therefore, expedient to design lightweight IDS that can be deployed on IoT devices. In this view, we used the incremental machine learning technique to design a lightweight IDS for IoT systems using incremental ensemble machine learning algorithms. Our proposed model was evaluated using the TON IoT dataset. The results show that our proposed model achieved a high average accuracy rate of 99.98%. The experimental results show the highest memory consumption at 650.11KB and the lowest at 122.38KB. The experimental result shows that our approach has led to the design of an IDS with a high accuracy rate and a lightweight model that can potentially run on IoT devices. In the future, we plan to evaluate our approach to other IoT-based datasets and deploy our model on an IoT device to evaluate parameters such as CPU usage, memory, and energy consumption. Additionally, future work could consider exploring how concept drifts in these datasets could be handled.

REFERENCES

[1] Statista, "• Global IoT and non-IoT connections 2010-2025 — Statista." [Online]. Available: https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/

[2] SAM Seamless Network, "2021 IoT Security Landscape - SAM Seamless Network." [Online]. Available: https://securingsam.com/2021-iot-security-landscape/

[3] J. P. Anderson, "Computer security threat monitoring and surveillance," *Technical Report, James P. Anderson Company*, 1980.

[4] J. R. Vacca, *Computer and information security handbook.* Newnes, 2012.

[5] S. U. Jan, S. Ahmed, V. Shakhov, and I. Koo, "Toward a lightweight intrusion detection system for the internet of things," *IEEE Access*, vol. 7, pp. 42 450–42 471, 2019.

[6] S. Latif, Z. e Huma, S. S. Jamal, F. Ahmed, J. Ahmad, A. Zahid, K. Dashtipour, M. U. Aftab, M. Ahmad, and Q. H. Abbasi, "Intrusion detection framework for the internet of things using a dense random neural network," *IEEE Transactions on Industrial Informatics*, 2021.

[7] S. Roy, J. Li, B.-J. Choi, and Y. Bai, "A lightweight supervised intrusion detection mechanism for iot networks," *Future Generation Computer Systems*, vol. 127, pp. 276–285, 2022.

[8] R. Zhao, G. Gui, Z. Xue, J. Yin, T. Ohtsuki, B. Adebisi, and H. Gacanin, "A novel intrusion detection method based on lightweight neural network for internet of things," *IEEE Internet of Things Journal*, 2021.

[9] T. D. Diwan, S. Choubey, H. Hota, S. Goyal, S. S. Jamal, P. K. Shukla, and B. Tiwari, "Feature entropy estimation (fee) for malicious iot traffic and detection using machine learning," *Mobile Information Systems*, vol. 2021, 2021.

[10] S. Mukkamala, G. Janoski, and A. Sung, "Intrusion detection using neural networks and support vector machines," in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, vol. 2. IEEE, 2002, pp. 1702–1707.

TABLE V. COMPARING THE PRECISION (P), RECALL (R) AND F1 SCORE OF EACH MODEL ON EACH ATTACK CATEGORY

	Gaussian NB			Hoeffding Tree			Our proposed model		
Fridge IoT dataset									
Attack category	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
Backdoor	0.00	0.00	0.00	98.73	99.38	99.06	99.97	99.76	99.86
DDoS	0.00	0.00	0.00	88.65	95.58	91.99	99.98	99.95	99.97
Injection	0.00	0.00	0.00	75.37	83.26	79.12	99.87	99.97	99.92
Normal	85.31	100.00	92.07	99.96	100.00	99.98	99.98	100.00	99.99
Password	0.00	0.00	0.00	90.09	92.99	91.52	99.97	99.97	99.97
Ransomware	0.00	0.00	0.00	43.83	27.29	33.64	99.59	99.69	99.64
XSS	0.00	0.00	0.00	100.00	11.31	20.33	100.00	99.41	99.71
Modus IoT dataset									
Backdoor	0.00	0.00	0.00	83.16	71.65	76.97	97.35	87.40	92.10
Injection	0.00	0.00	0.00	41.83	18.16	25.33	75.70	69.80	72.63
Normal	77.60	100.00	87.39	99.97	100.00	99.99	99.99	99.98	99.99
Password	0.00	0.00	0.00	46.42	68.78	55.43	71.61	89.00	79.36
Scanning	0.00	0.00	0.00	47.65	63.14	54.31	86.60	71.64	79.62
XSS	0.00	0.00	0.00	14.29	0.20	0.40	17.51	25.10	20.63
Garage IoT dataset									
Backdoor	0.00	0.00	0.00	99.55	94.32	96.86	99.67	99.91	99.79
DDoS	0.00	0.00	0.00	35.38	98.51	52.06	99.65	98.99	99.32
Injection	0.00	0.00	0.00	0.00	0.00	0.00	99.81	99.78	99.79
Normal	85.96	100.00	92.45	99.97	100.00	99.99	100.00	100.00	100.00
Password	0.00	0.00	0.00	66.51	47.23	55.23	99.85	99.88	99.87
Ransomware	0.00	0.00	0.00	0.00	0.00	0.00	99.90	99.83	99.86
Scanning	0.00	0.00	0.00	0.00	0.00	0.00	99.81	98.68	99.24
XSS	0.00	0.00	0.00	0.00	0.00	0.00	100.00	99.13	99.57
Motion light IoT dataset									
Backdoor	0.00	0.00	0.00	43.94	99.28	60.92	99.97	99.70	99.83
DDoS	0.00	0.00	0.00	0.00	0.00	0.00	99.96	99.94	99.95
Injection	0.00	0.00	0.00	0.00	0.00	0.00	99.93	99.95	99.94
Normal	85.86	100.00	92.39	99.95	100.00	99.97	99.98	100.00	99.99
Password	0.00	0.00	0.00	0.00	0.00	0.00	99.98	99.98	99.98
Ransomware	0.00	0.00	0.00	0.00	0.00	0.00	99.82	99.82	99.82
Scanning	0.00	0.00	0.00	0.00	0.00	0.00	99.77	99.77	99.77
XSS	0.00	0.00	0.00	0.00	0.00	0.00	100.00	99.11	99.55
GPS Tracker IoT dataset									
Backdoor	95.05	7.50	13.90	97.87	99.47	98.66	99.97	99.73	99.61
DDoS	98.77	8.65	15.91	69.66	92.50	79.47	99.87	99.97	99.92
Injection	16.77	46.00	24.58	79.02	79.78	79.40	99.84	99.88	99.86
Normal	88.42	96.57	92.31	99.96	100.00	99.98	99.98	100.00	99.99
Password	100.00	11.11	20.00	85.27	83.39	84.32	99.96	99.94	99.95
Ransomware	20.08	58.84	29.94	99.54	7.70	14.29	99.12	99.61	99.37
Scanning	100.00	0.18	0.36	100.00	20.36	33.84	100.00	99.64	99.82
XSS	10.92	13.52	12.08	0.00	0.00	0.00	100.00	95.84	97.88

TABLE VI. COMPARING THE PRECISION (P), RECALL (R) AND F1 SCORE OF EACH MODEL ON EACH ATTACK CATEGORY CONTINUATION

	Gaussian NB			Hoeffding Tree			Our proposed model		
Attack category	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
Weather IoT dataset									
Backdoor	100.00	0.05	0.10	96.42	99.48	97.92	99.89	99.98	99.94
DDoS	0.00	0.00	0.00	83.59	93.53	88.28	99.89	99.84	99.87
Injection	0.00	0.00	0.00	69.59	89.18	78.18	97.39	99.88	98.62
Normal	86.08	100.00	92.52	99.97	100.00	99.98	100.00	100.00	100.00
Password	0.00	0.00	0.00	89.57	78.11	83.45	99.93	98.87	99.45
Ransomware	100.00	0.31	0.63	86.95	34.66	49.56	96.71	99.55	98.11
Scanning	0.00	0.00	0.00	100.00	38.56	55.66	95.19	86.01	90.37
XSS	0.00	0.00	0.00	100.00	39.84	56.98	100.00	94.11	96.97
Thermostat IoT dataset									
Backdoor	0.00	0.00	0.00	97.48	99.43	98.44	99.93	99.75	99.84
Injection	0.00	0.00	0.00	88.55	93.47	90.94	99.11	99.80%	99.45
Normal	87.27	100.00	93.20	99.95	100.00	99.97	99.98	100.00	99.99
Password	0.00	0.00	0.00	85.37	84.64	85.00	99.52	98.99	99.26
Ransomware	0.00	0.00	0.00	72.39	44.70	55.27	99.55	98.23	98.89
Scanning	0.00	0.00	0.00	0.00	0.00	0.00	69.33	85.25	76.47
XSS	0.00	0.00	0.00	100.00	1.34	2.64	100.00	94.65	97.25

[11] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *computers & security*, vol. 28, no. 1-2, pp. 18–28, 2009.

[12] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on software engineering*, no. 2, pp. 222–232, 1987.

[13] B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in internet of things," *Journal of Network and Computer Applications*, vol. 84, pp. 25–37, 2017.

[14] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac, and P. Faruki, "Network intrusion detection for iot security based on learning techniques," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2671–2701, 2019.

[15] S. Otoum, B. Kantarci, and H. T. Mouftah, "A novel ensemble method for advanced intrusion detection in wireless sensor networks," in *Icc*

2020-2020 ieee international conference on communications (icc). IEEE, 2020, pp. 1–6.

[16] C. Zhang and Y. Ma, *Ensemble machine learning: methods and applications.* Springer, 2012.

[17] T. T. Khoei, G. Aissou, W. C. Hu, and N. Kaabouch, "Ensemble learning methods for anomaly intrusion detection system in smart grid," in *2021 IEEE International Conference on Electro Information Technology (EIT)*. IEEE, 2021, pp. 129–135.

[18] R. E. Schapire, "The strength of weak learnability," *Machine learning*, vol. 5, no. 2, pp. 197–227, 1990.

[19] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural computation*, vol. 3, no. 1, pp. 79–87, 1991.

[20] A. A. Benczúr, L. Kocsis, and R. Pálovics, "Online machine learning in big data streams," *arXiv preprint arXiv:1802.05872*, 2018.

[21] A. Khannoussi, A.-L. Olteanu, C. Labreuche, P. Narayan, C. Dezan, J.-P. Diguet, J. Petit-Frère, and P. Meyer, "Integrating operators' preferences into decisions of unmanned aerial vehicles: multi-layer decision engine and incremental preference elicitation," in *International Conference on Algorithmic Decision Theory*. Springer, 2019, pp. 49–64.

[22] A. Mozaffari, M. Vajedi, and N. L. Azad, "A robust safety-oriented autonomous cruise control scheme for electric vehicles based on model predictive control and online sequential extreme learning machine with a hyper-level fault tolerance-based supervisor," *Neurocomputing*, vol. 151, pp. 845–856, 2015.

[23] F. Feng, R. H. Chan, X. Shi, Y. Zhang, and Q. She, "Challenges in task incremental learning for assistive robotics," *IEEE Access*, vol. 8, pp. 3434–3441, 2019.

[24] S. Muthukrishnan *et al.*, "Data streams: Algorithms and applications," *Foundations and Trends® in Theoretical Computer Science*, vol. 1, no. 2, pp. 117–236, 2005.

[25] M. R. Henzinger, P. Raghavan, and S. Rajagopalan, "Computing on data streams." *External memory algorithms*, vol. 50, pp. 107–118, 1998.

[26] L. Yang, D. M. Manias, and A. Shami, "Pwpae: An ensemble framework for concept drift adaptation in iot data streams," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 01–06.

[27] V. Shakhov, S. U. Jan, S. Ahmed, and I. Koo, "On lightweight method for intrusions detection in the internet of things," in *2019 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*. IEEE, 2019, pp. 1–5.

[28] J.-S. Pan, F. Fan, S.-C. Chu, H.-Q. Zhao, and G.-Y. Liu, "A lightweight intelligent intrusion detection model for wireless sensor networks," *Security and Communication Networks*, vol. 2021, 2021.

[29] L. H. A. Reis, A. Murillo Piedrahita, S. Rueda, N. C. Fernandes, D. S. Medeiros, M. D. de Amorim, and D. M. Mattos, "Unsupervised and incremental learning orchestration for cyber-physical security,"

Transactions on emerging telecommunications technologies, vol. 31, no. 7, p. e4011, 2020.

[30] A. Tabassum, A. Erbad, A. Mohamed, and M. Guizani, "Privacy-preserving distributed ids using incremental learning for iot health systems," *IEEE Access*, vol. 9, pp. 14 271–14 283, 2021.

[31] S. D. Jadhav and H. Channe, "Comparative study of k-nn, naive bayes and decision tree classification techniques," *International Journal of Science and Research (IJSR)*, vol. 5, no. 1, pp. 1842–1845, 2016.

[32] A. McCallum, K. Nigam *et al.*, "A comparison of event models for naive bayes text classification," in *AAAI-98 workshop on learning for text categorization*, vol. 752, no. 1. Citeseer, 1998, pp. 41–48.

[33] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[34] T. M. Mitchell and T. M. Mitchell, *Machine learning*. McGraw-hill New York, 1997, vol. 1, no. 9.

[35] C. Bustamante, L. Garrido, and R. Soto, "Comparing fuzzy naive bayes and gaussian naive bayes for decision making in robocup 3d," in *Mexican International Conference on Artificial Intelligence*. Springer, 2006, pp. 237–247.

[36] G. Hulten, L. Spencer, and P. Domingos, "Mining time-changing data streams," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 97–106.

[37] H. Wang, W. Fan, P. S. Yu, and J. Han, "Mining concept-drifting data streams using ensemble classifiers," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 226–235.

[38] J. Montiel, M. Halford, S. M. Mastelini, G. Bolmier, R. Sourty, R. Vaysse, A. Zouitine, H. M. Gomes, J. Read, T. Abdessalem *et al.*, "River: machine learning for streaming data in python," 2021.

[39] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and A. Anwar, "Ton_iot telemetry dataset: A new generation dataset of iot and iiot for data-driven intrusion detection systems," *IEEE Access*, vol. 8, pp. 165 130–165 150, 2020.

[40] Q. Abu Al-Haija, A. Al Badawi, and G. R. Bojja, "Boost-defence for resilient iot networks: A head-to-toe approach," *Expert Systems*, p. e12934, 2022.

[41] M. A. Khan, M. A. Khan Khattk, S. Latif, A. A. Shah, M. Ur Rehman, W. Boulila, M. Driss, and J. Ahmad, "Voting classifier-based intrusion detection for iot networks," in *Advances on Smart and Soft Computing*. Springer, 2022, pp. 313–328.

[42] W. W. Lo, S. Layeghy, M. Sarhan, M. Gallagher, and M. Portmann, "E-graphsage: A graph neural network based intrusion detection system for iot," in *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2022, pp. 1–9.

[43] A. R. Gad, A. A. Nashat, and T. M. Barkat, "Intrusion detection system using machine learning for vehicular ad hoc networks based on ton-iot dataset," *IEEE Access*, vol. 9, pp. 142 206–142 217, 2021.

Blockchain based Framework for Efficient Student Performance Tracking (BloSPer)

Aisha Zahid Junejo[1], Anton Dziatkovskii[2], Manzoor Ahmed Hashmani[3], Uladzimir Hryneuski[4], Ekaterina Ovechkina[5]

Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Sri Iskandar, Malaysia[1, 3]
Platinum Software Development Company, Australia[2, 4, 5]

Abstract—**For maintaining sustainable economy, the government of Malaysia is working towards improvising the standards of education in higher education institutes. According to reports, around 32% of enrolled students in Public Universities of Malaysia are unable to graduate on time due to unknown reasons. To ensure more students graduate on time with high quality of education, continuous monitoring of the student is essential. Continual tracking will allow the student as well as the educator to analyze the weak performer at an early stage. Tracking the student performance manually is challenging but with the advancements in information technology, keeping a track of student performance has been relatively easier. Therefore, the fundamental aim of this paper is to present a novel blockchain framework for record keeping and student performance tracking. We name this framework BloSPer (Blockchain Student Performance Tracking System). BloSPer has an edge over the existing systems as current systems face problems of single point of failure and unreliable data. The proposed framework will enable the students and educators to track the performance of the students in a more convenient and transparent manner. Due to this, it will be simpler for them to analyze the reasons of a students' poor performance. Moreover, the data gathered through the system will be more reliable and worthy for data analytics because of tamper resistance provided through blockchain. This will result in much knowledgeable decisions by the institutions regarding improving the performance of each individual candidate.**

Keywords—Blockchain; education; performance tracking; trackability; student data analytics; student monitoring

I. INTRODUCTION

A. Background

For any nation to progress, education plays a significant role and is therefore usually the main focus for most developed countries. This is because the advancement in any sector depends on the level of education. United Nations has listed quality education as the 4th out of 17 Sustainable Development Goals (SDG) for its 2030 agenda. Education has a long-term positive impact on a country's economic growth [1]. Education plays a vital role in Malaysia for building a resilient nation, encouraging the creation of a just society, and maintaining sustainable economic growth [2]. Thus, improvising education sector is one of the major concerns of the Malaysian government nowadays.

According to Malaysian Qualifications Agency (MQA) criteria for higher education has been set through a code of practice that includes, vision, mission and learning outcomes; curriculum design and delivery; student selection and support services; assessment of students; educational resources; continual quality improvement and more. Each institution is evaluated based on the standards to ensure high quality of education. However, the way each institution works towards achievement of those standards varies. Despite of providing high quality of education, demand on predicting student academic performance become more critical to improve the quality of education and assisting the students to achieve a greater performance in their studies [2].

One way that most of the higher education institutions maintain the standards and quality of education is through tracking the student performance in a classroom setting [3][4]. Keeping a record of student performance at regular intervals will enable an interactive learning environment and promote better educational outcomes for the students. Tracking and monitoring the performance of students from time to time throughout the course of an academic semester has a positive impact for both, the students and the educators [5]. This positive impact may include valuable insights to redesign the course for instructors, decisions for improving standards of education, effective intervention of students and collection of valuable data for further analytics processes (if any). The fundamental aim of performance recording and analysis is to ensure better education standards and improved student performance. In order to achieve higher quality education, it is the core responsibility of each individual student and the academic institute to ensure every student achieves expertise in his area of interest. To reach this goal, various tools are needed by the educators to identify students who are at a potential risk of academic failure and adjust their strategies of delivery education meeting the needs of those students. Moreover, it is also very important for students to know the areas that need further focus and hard work. With student performance tracking the educators and students can continually evaluate the effectiveness of teaching and learning environment in the class. The continual monitoring will help them make more informed instructional decisions.

Information technology has made it easy to do and achieve with the recent advancements. The record keeping and management are now shifting to an innovative technology named Blockchain [6]. A blockchain is a distributed ledger i.e. the ledger is spread among all peers in across the network, and each peer holds a copy of the complete ledger [7]. Fundamental features of blockchain include being, i)

distributed, i.e., disseminating the ledger across the network to restrict the tampering of data [8], ii) transparency [9], i.e., authenticity of each transaction is verified by participants via consensus [10] in the network, and iii) immutable [11], i.e., the data cannot be modified/altered. The technology of Blockchain is known to track and trace the data in a more efficient way then the traditional record keeping approaches.

B. Research Gap

Presently, various tools and techniques are being used to monitor [12] the performances of students, and to predict student grades [13]. However, to the best of our knowledge, there's no way of determining the causes of student failure, i.e., if a student is likely to fail, then what reasons are causing this failure. This information is mainly important as it highlights what an individual lacks. In order to ensure each student's academic success, it is important to adjust the learning materials according to their needs. This can be achieved by tracking history of the student. Existing information systems do not have a reliable track of student performance data. Therefore, in this paper, we intend to overcome this gap by using the cutting-edge technology of the blockchain networks [14]. The main contributions of the manuscript, in this regard, are elaborated in upcoming section.

C. Major Contributions

The major contributions of this research article are twofold: 1) the shortcomings of existing student performance tracking systems are discussed and, 2) blockchain based novel framework for student performance tracking system is proposed. This framework will record the performance of the students in a distributed ledger accessible by various stakeholders involved in the process. The transparency and traceability of the blockchain will ensure that the performance indicators are not tampered with and can be tracked as needed. Moreover, since blockchains are immutable, hence the data collected from the proposed framework will be reliable, tamper free and accurate. This data can further be used for data analytics to determine student behaviours in various educational applications.

D. Organization

The organization of rest of the paper is as follows: Section II, describes the state-of-the-art student performance tracking systems available in literature and describes the shortcomings of the existing systems and discusses how blockchain technology can be used to overcome these challenges. Section III enlists the design goals and reasoning behind the proposed design. Section IV entails the proposed framework BloSPer. Section V presents results and discussions of the proposed framework. The paper finally concludes in Section VI, and a few promising future research directions are suggested in Section VII.

II. LITERATURE REVIEW

According to Malaysian Educational Statistics report published by Education Data Sector [15] about 32% of the students enrolled in public universities, fail to graduate on time. Moreover, this report was further analysed, and it was found out that from 2016 and onwards, each year the budget being used on improvising the education sector is substantially increasing. In 2016 actual educational recurrent expenditure was around 40 Million MYR which increased to around 45 Million MYR in 2018. Despite of the substantial increase in the budget being spent on education (public universities only) the literacy rate did not have a considerable improvement. The overall literacy rate in subsequent years only increased by 0.8%. These insights highlight the importance of taking measures to analyze the reasons behind unsatisfactory performance of a student in higher educational institutes.

A. Existing Student Performance Tracking (SPT) Systems

For determining the success and worth of any educational institute, the number of students graduating [16] each year plays an important role. Not only students graduating but, the students graduating with better scores is also significant [17]. Some students, having a higher intellectual ability, are fast learners while others need more time and undivided attention of the instructor. To maintain the quality of education, many institutions use different tools and techniques to track the output of the students. This tracked output helps the educators to accommodate differing needs of students [18]. With the advancements in technology it has been easier for educators to keep a track of student performance using various factors like student attendance, exam, quiz and lab work score, timely submission of assignments [19] and many more. Recording these features can also be used to predict if a student is likely to pass an exam or fail. In [20] a study was conducted to predict the success rate of the candidate in an English exam. Besides that, scores were also predicted with the help of technology. Similar study was presented in [21] where different data mining techniques were used to predict if a student is likely to pass or fail. This kind of analysis is claimed to be useful for higher education institutes as it will help them analyze students at risk to take preventive measures in time. Several other studies [22][23] have also been conducted to try different methods to predict passing or failing of a student in a course. However, these studies fail to track back to the factors causing failure of a student. Although, it is desirable to know the tendency of a student to fail to prevent it in time, but existing systems and technologies are incapable of tracking and finding the flaws each student possess. The identification of these flaws can lead to build a foundation of more useful strategies to enable students cop up with their inabilities or other study related issues.

Keeping the record of student performance is a usual practice in educational institutes all over the world. Tracking plays a significant role in analysing whether the student is learning well or not. Not only it allows monitoring of the students but recording performance and behavioural patterns of students also help data analysts to predict future performance of the students with the help of information technology. This prediction is highly valuable as it helps educational institutes to take measures to improve the performance and learning ability of students [13], thereby uplifting institutional reputation.

B. Challenges in Existing SPT Systems

A lot of work is being done for tracking student performance. However, this work has various challenges and shortcomings in terms of data storage and retrieval that need

to be addressed for more efficient systems and for better learning of students. Some of these challenges are discussed as follows.

1) Lack of traceability: The most common problem with existing record keeping systems is lack of traceability [24][25]. This means that these systems usually depict the overall statistics of the student performance but fail to give a detailed insight on the history of student evaluation as a learner at various levels of study. This history is important as it determines when the student's performance started declining what factors caused it to decline. This information is significant for correct prediction of student future behavior [26]. Moreover, history tracing allows the educators to make more knowledgeable decisions for improving a certain student's learning ability and overall performance.

2) Single point of failure: Existing systems are built on centralized storages [27] [28]. This means that they have a single point of failure [29]. In case of a disaster or loss of data, it will be very difficult and almost impossible to retrieve back the data if backups are not properly kept.

III. CONCEPTUALIZATION OF PROPOSED BLOCKCHAIN BASED STUDENT PERFORMANCE TRACKING SYSTEM (BLOSPER)

A. Design Goals

Based on the problems found in literature with existing student performance tracking systems (as mentioned in previous section), we have devised a few design goals that must be present in the proposed system. These design goals are as follows:

1) The application must be decentralized to avoid any single point of failure.

2) The record of student submissions must be transparent and tamper proof.

3) The performance of each student must be easily trackable.

To accomplish this, we proposed a student performance tracking system based on the novel technology of blockchain networks.

B. Blockchain as a Solution

The idea of blockchain technology was introduced with the advent of bitcoin, popularly known cryptocurrency [14]. The technology eliminates the need of third parties for transactions and has been proven to be more time efficient as compared to the third-party hosted system [15]. Blockchain is a disruptive innovation based on distributed ledger technology. It is responsible for record keeping in a transparent ledger. The transparent distributed ledger makes it very convenient for the user to trackback even the tiniest form of information as all the data is chained into the blocks. Besides ease of tracking data, blockchain also ensures immutability of records. This implies that the once the data has been validated and stored into the blockchain, it cannot be tampered with, hence a blockchain based system is essentially hack proof. Blockchain has a number of benefits that will help the traditional student performance monitoring and tracking systems more efficient and meaningful. Some of these include:

C. Decentralized and Distributed Ledger

Since the data on a blockchain network is distributed among all the participating peers [16] therefore, the need of trusting a designated authority to control and supervise the chain is not needed. This provides a user a complete autonomy over his data. Any student using this system can completely rely on the system for accurate tracking and monitoring without any ambiguities. Moreover, the decentralization of storage [17] ensures that there is a very little to no chance of the data being misplaced. This ensures the integrity and security of the data to any deploying enterprise/institution.

1) Transparency: Another remarkable feature provided by blockchain technology is transparency. Anyone in the network can view the information as per their convenience. This reduces the chances of illegal transactions.

2) Better trackability: The data on blockchains cannot be altered or removed. This signifies that all the data that is present on the blockchain remains there and every new information is linked to the previous corresponding information. This enables the user to trackback the entire history of chained records from the storage. This history can present detailed insights on the performance of students through various events from the beginning of usage of the system. Since all the transactions in the blockchain are timestamped hence the time and data of submission for different assignments and tasks can also be inferred from the records. Such data can be very useful for data analytics and to examine the weak points in student performance that must be targeted for better learning of the student and overall education system.

3) Immutability: Immutability of the data and transactions in blockchain will ensure no student data is tampered with. This enables a greater trust of the stakeholders into the performance tracking system. Moreover, it guarantees that the data used by data analytics for prediction purposes is original and untampered.

Moreover, according to [7] the internet access to educational institutes is getting better with each passing year. This provides a stronger base for blockchains to be implemented without a problem. Therefore, with these wide variety of benefits and ease of internet access, blockchain-based student performance tracking systems can be successfully deployed in educational institutes thereby improving the education standards of the country.

IV. BLOSPER ARCHITECTURE

A. Proposed Framework Components

Continual monitoring and tracking of students' performance is known to increase the efficiency of the students. In this paper, a novel framework for blockchain based student performance tracking (BloSPer) is proposed to keep the record of student performance on a blockchain network. The fundamental aim of this framework is to make

tracking the performance of students easier and more transparent. The traditional student monitoring and tracking systems are unable to trackback and identify the causes of student failure. The proposed framework will trackback to the entire performance history of the students to recognize the causes of performance lag in students. This ensures better strategies are devised to improve student learning ability and overall quality of education of an institution. The overall framework for the proposed BloSPer is shown in Fig. 1. The proposed framework has three major components i.e., User Interface (the front-end), Processing Layer and, Blockchain Layer. The functionality of each component is explained below:

1) User interface: As the name suggests, this component enables the users to interact with the blockchain system for information storage and retrieval. It captures information including login credentials, assignments, and other essential data from end user and sends it to processing layer.

2) Processing layer: This component consists of all the major modules designed for smooth functionality of BloSPer. The modules include i) Login Module, ii) Registration Module, iii) User Management Module, iv) Data Management Module and, v) Data Analytics Module. The data from front-end is received by processing layer and the query is forwarded to the corresponding module. The purpose of each module is elaborated as follows.

a) Login Module (LM): This module is responsible for ensuring if the login credentials entered by a user are authorised to use the system. Once the user is authorised, the module maintains a session so that the user does not need to login for every transaction triggered.

b) Registration Module (RM): The registration module enables the system administrator to register the records of students and instructors. No one can use the system unless administrator registers them using this module.

c) User Management Module (UMM): This module allows the administrator to add and delete the users from the system.

d) Data Management Module (DMM): This module is responsible for management of data, i.e., it enables the users to store the data into the chain and retrieve the authorised information from the blockchain.

e) Data Analytics Module (DAM): The purpose of this module is to download all the data for analytics purposes. These purposes may include predicting student academic performance, analysing instructor behaviour, and so on. However, it should be noted that this module is yet not a part of the implementation, currently it is only being proposed, hence it is colour coded in red to illustrate the difference between this module and the rest of the modules.

3) Blockchain layer: The third component is the blockchain layer. It provides the decentralised support for the proposed framework. Since blockchain itself cannot store

huge multimedia files, hence this layer also provides blockchain compatible, decentralised storage.

B. Stakeholders

There are three kinds of stakeholders involved in the proposed system, i.e., the instructor or educator, the student and the administrative staff as shown in Fig. 2. The role of each stakeholder is described below:

1) Instructor

The main responsibilities of instructor include:

- Self-authorize to be able to use the system.

- Upload scores of exams, assignments, lab work and other tasks submitted by students.

- Track and trace the performance of students.

2) Student

The main responsibilities of students include:

- Self-authorize to be able to use the system.

- Upload assignments, lab work and other tasks assigned by the instructor.

- Track and trace their performance.

3) Administrator

The main responsibilities of students include:

- Self-authorize to be able to use the system.

- Add student data and grant those students the permission to access the system.

C. Features and Functionalities

The various other features the BloSPer offers for end users include the following:

1) Enabling students to upload assignments/tests/lab work and other assessment materials.

2) Enabling instructors/teachers to review the assessment materials uploaded by students and mark those.

3) Enabling instructors/teachers to award reward points to students. Students can further redeem these reward points for accessing paid teaching and learning materials.

4) Enabling institutional administrators to ensure that students and instructors are uploading and reviewing assessment materials on time.

5) Enabling institutional administrators to download reliable data for further data analysis and decision making on improving student performance.

System users need to use DApp to login to the system using their private keys and interact with the data stored on the chain. It can be noted that the framework has a module.
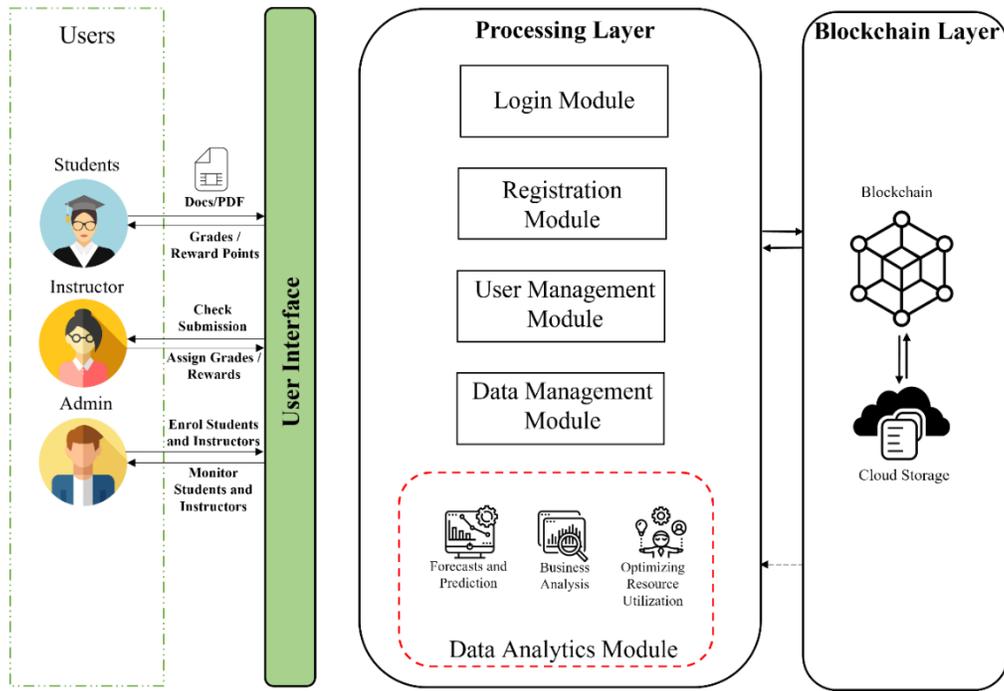
Fig. 1.   Proposed Framework for BloSPer.
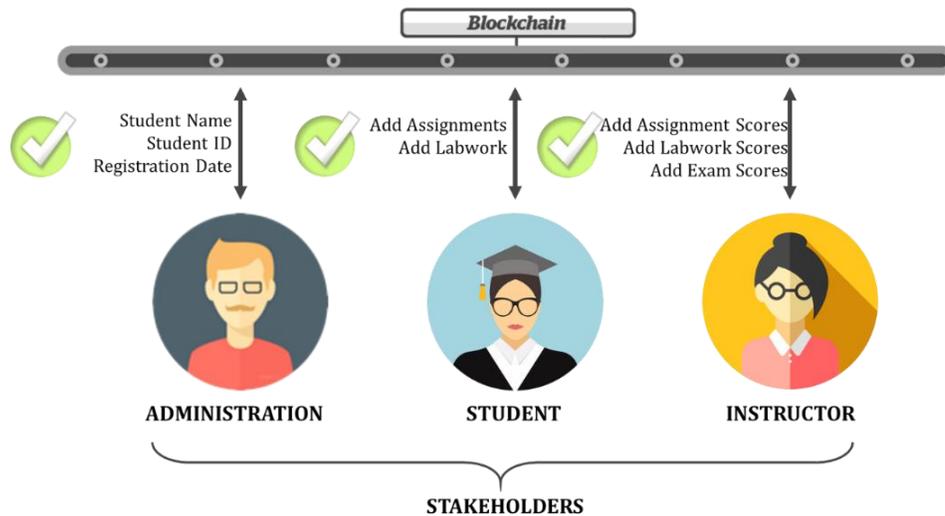
Fig. 2.   Stakeholders in Student Performance Tracking System.

D. Interaction and Transaction Sequence

In this section, we describe how BloSPer transactions are executed and communicated between different stakeholders. A graphical illustration of the system usage is given in Fig. 3. The student will submit his assignment, lab work or other tasks in the form of a transaction in a block. The block must be then validated by all the participating peers in the network. Once the block is validated it is stored into existing blockchain network. Next, the instructor can forward the request for data retrieval. This request will again be validated by participating peers to check if the instructor is authorized to access the data or not. After the request has been validated, the data will be retrieved from the blockchain and sent to the instructor. After reviewing student assignment/lab work, the instructor will insert the relevant grade into the blockchain, which will again be validated by peers and the final data will be stored into the blockchain.

The same procedure will be repeated when student wants to access his grade or when the administrative staff wants to retrieve the data for official purposes.

For using this system, all users are required to identify and authenticate themselves using the private keys received upon registration. Initially, a predefined administrator will login the system. This login request will trigger login () function in LM, from where the LM will authenticate him by matching the user records stored on blockchain. Once the administrator is authorised, he may use the system to register different

participants (i.e., students and teachers/instructors) to use the system, by calling reg(uname, utype) function, where uname refers to the name of the users and utype refers to his role/type (student/teacher). Once the registration module receives the request from admin, it generates public and private keys for the users and returns the same to the admin. Now the student may use these keys to login and submit their assignments. This will trigger the function submit (title, type, desc, teacher)

where title refers to assignment title, type refers to the type of file submitted, desc refers to description of the submission and finally teacher refers to the public ID of the instructor that the assignment is submitted to. Once the student has submitted the assignment, the teacher can login to his portal and view the assignment, mark the assignments and monitor student progress. This entire process is shown in Fig. 4.
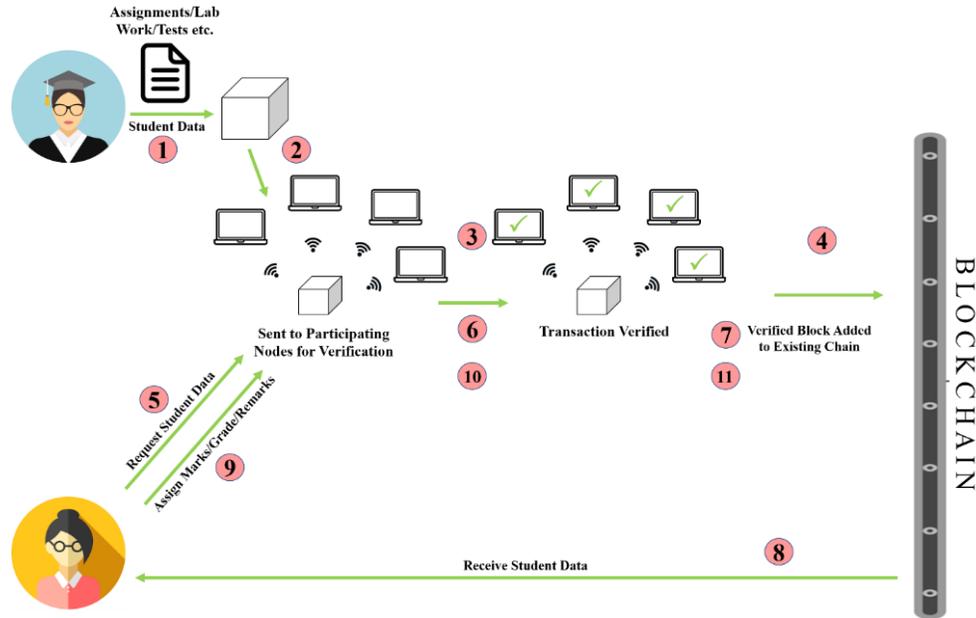
Fig. 3.   Workflow of Students Performance Tracking Systems.
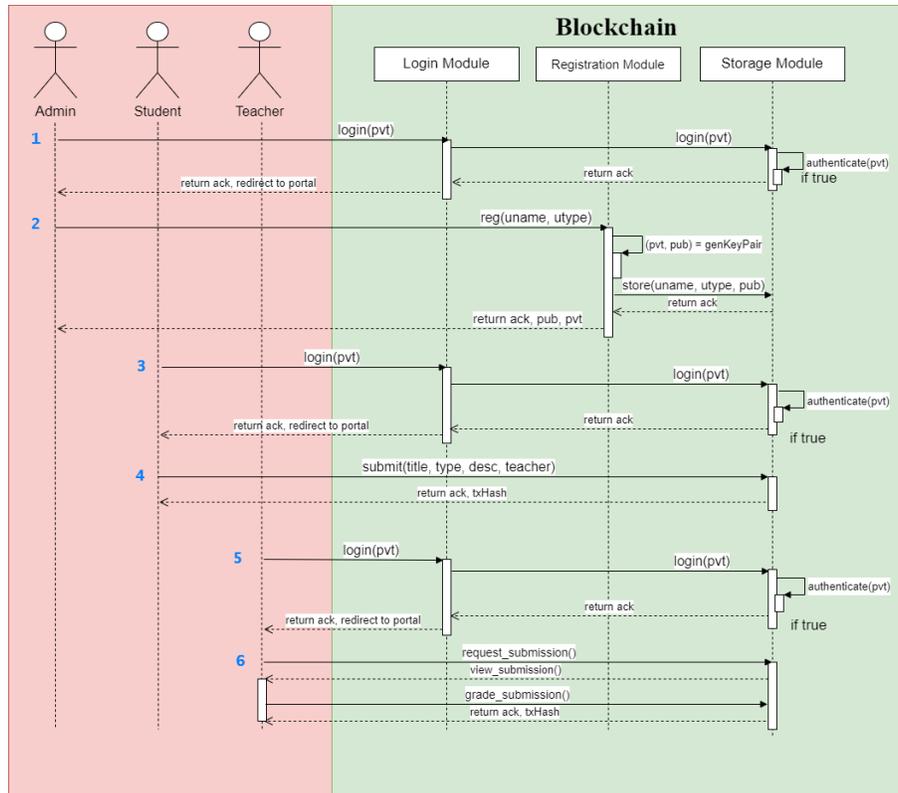
Fig. 4.   Different Types of Transaction Flows in BloSPer.

E. System Architecture

The layered architecture of the proposed system is illustrated in Fig. 5. The proposed framework is divided into four layers: i) Infrastructure Layer, ii) Consensus Layer, iii) Protocol Layer and iv) Application Layer.

The first layer is the infrastructure layer facilitating all the required hardware components for the system including different kinds of nodes (i.e., simple, full, mining), storage and other network facilities. The function of the nodes is to send, receive, store and mine the transactions. Storage in the layer is responsible for storing the updated ledger after each verified transaction is added to the chain.

Followed by infrastructure layer, the proposed architecture consists of consensus layer. A consensus mechanism is an integral part of blockchain networks. It can be defined as a fault-tolerant mechanism that is used in blockchain systems to achieve the necessary agreement on a single data value or a single state of the network among distributed processes or multi-agent systems. Most widely used consensus mechanism include Proof-of-Work (PoW), Proof-of-Stake (PoS), Proof-of-Capacity (PoC), etc. This layer also includes smart contracts which are computer programs consisting of predefined set of rules and agreements among various parties involved in the blockchain network.

The third layer in the proposed architecture is the protocol layer. This layer deals with the protocols that are necessary for a transaction to flow through the network. Whenever a transaction is requested/initiated in a blockchain network, it must first be signed with digital signature of the sender and public key of the receiver. Then the transaction is encrypted using hash functions. This layer is also responsible for handling various protocols required for decentralized data storage.

Finally, there is an application layer. The function of application layer is to provide the user an ease of interacting with the data and communication within blockchain. This is a host layer for network application to make the usage of the system more convenient to the stakeholder.
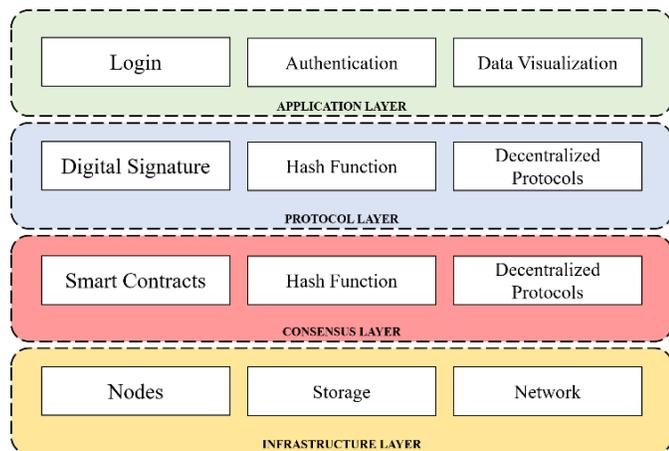
Fig. 5.   Student Performance Tracking System Architecture.

V.   Simulation Results and Discussion

In this section, the simulation results are described and discussed. The experiments are simulated on Intel(R) Core (TM) i7-10510U processor with 8GB RAM and 2304 MHz clock speed. The prototype is implementation using NodeJS programming language and executed using Visual Studio Code. The main screen to use the system is shown in Fig. 6.
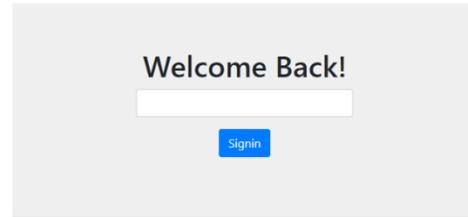
Fig. 6.   Welcome Screen of BloSPer.
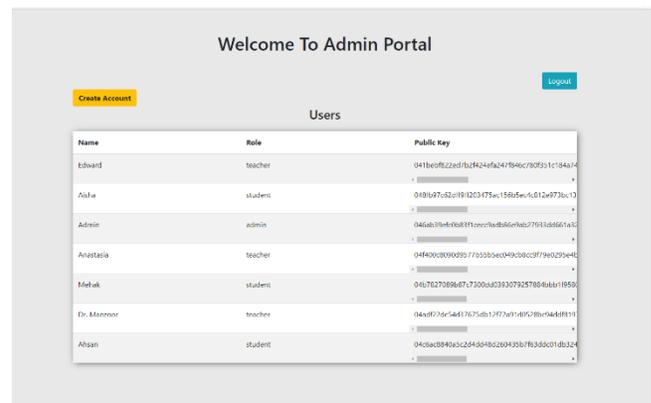
Fig. 7.   Admin Portal.

This screen requires the admin to sign in. The information of the admin has already been added to the system. Once the admin logins using his private key, the admin will be redirected to the admin portal as shown in Fig. 7. On the portal, the admin can see all the users of the system, their roles/types and their public keys. Here, the admin has an option of register more students and teachers to the system using create account button. Moreover, the admin portal has an option named "Block Explorer" which will enable the admin to retrieve information about a user, a block, or a transaction by using their corresponding hashes as shown in Fig. 8.
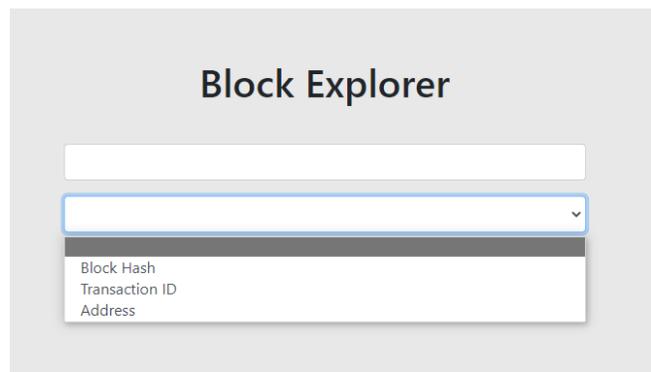
Fig. 8.   Block Explorer – Admin Portal.

As soon as the admin enters the hash and selects the category that the hash belongs to, the entire information about that hash will be visible to the admin for monitoring purposes as shown in Fig. 9.
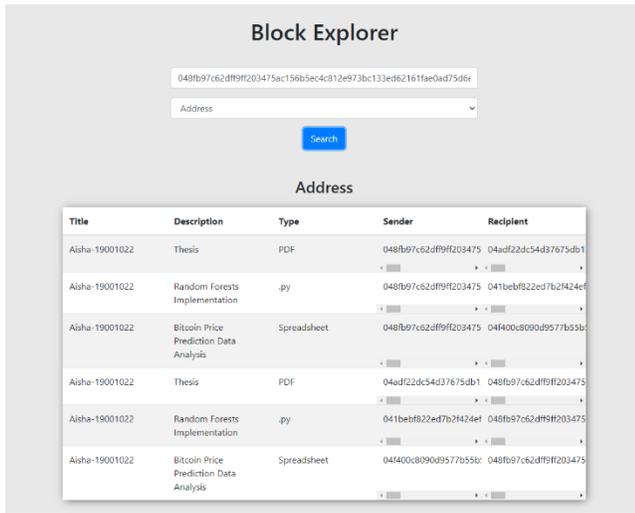
Fig. 9. Retrieving user information – Admin Portal.
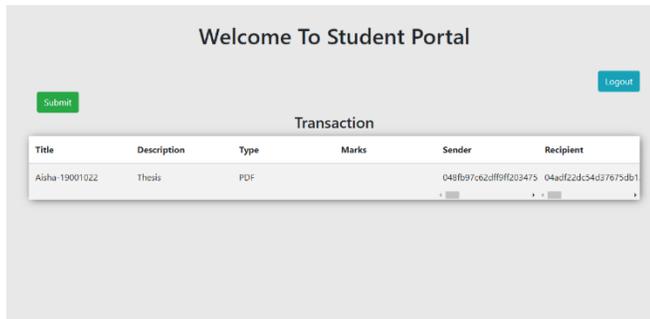
Fig. 10. Student Portal.
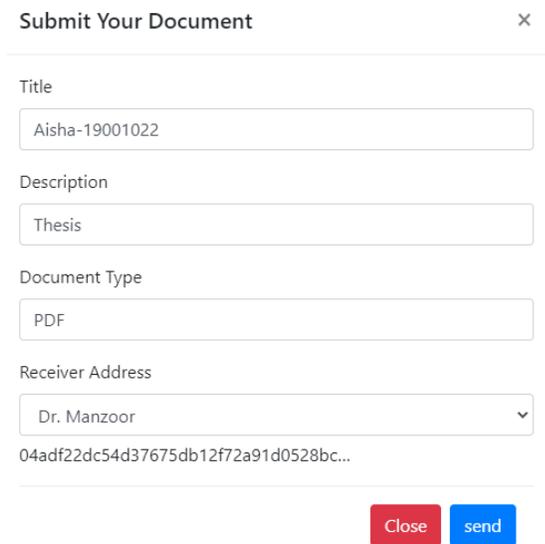
Fig. 11. Task Submission – Student Portal.

The second stakeholder of the proposed BloSPer is the student. As soon as the student logins, he will be redirected to the student portal as shown in Fig. 10. To add a file, the student must click on submit button, which will open a pop-up window for submission as shown in Fig. 11. At the time of submission, there will be no marks assigned as shown in Fig. 10. Once the student has submitted the work, the teachers can view and mark it on their portal as shown in Fig. 12. and Fig. 13. respectively.
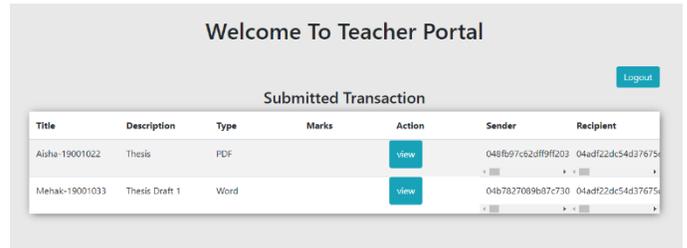
Fig. 12. Viewing Task – Teacher Portal.
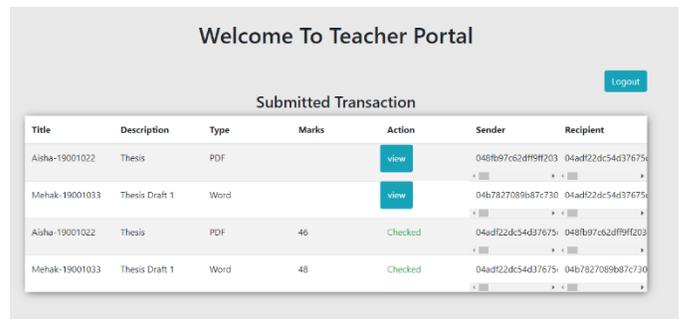
Fig. 13. Marking Task – Teacher Portal.

The data retrieved from the usage of this system will not only provide information regarding student marks but will also describe the student behavior in terms of submission time and other patterns. It will also monitor the behavior of teachers while marking. All this data can further be used to develop insights to improvise student academic performance.

VI. CONCLUSION

High quality education has a great significance for progress and prosperity of any nation. For socio-economic, political and industrial growth and well-being of a country, it is very important for its people to be well learned. Various institutions are striving towards ensuring that their students are well skilled and well trained. To achieve this, these institutions are trying and testing several methods to improve the performance of the students lagging due to unknown reasons. One such widely adopted practice is monitoring and analyzing the performance of students with the help of information technology. Literature shows that there is some remarkable work done in the domain for recording student's past and current behavior and predicting the future performance to improve the quality of education and overall student performance. However, these systems present statistical analysis only. Substantial amount of consideration is not given to trackback to the causes of students' poor performance. Our proposed framework utilizes the features of blockchain to overcome these issues. It provides a platform for

student performance monitoring in a decentralized environment. Due to immutability and trackability of blockchain, the entire history of student performance is stored which cannot be removed or altered. Hence, it provides a clear picture of student's academic standing. Using this information, important actions could be taken in order to optimize and improvise the learning experience for the students.

VII. FUTURE WORK

In this paper, we have presented a novel framework based on innovative technology of blockchain that will not only record student performance but also trace back to the causes of student failure, which is essential to devise better strategies to improve the overall education standard of the institute. However, the given system has some limitations due to time and resource constraints, hence we propose the following future research directions to improvise the study given in this paper:

1) Incentive based learning: Another simple feature could be added to BloSPer where teachers award the students with some reward points based on their good performance. These reward points can later be redeemed to access paid research and study materials, or to apply for certain courses/certifications to further enhance their capabilities.

2) Incorporation of IPFS/CouchDB for multimedia storage: Interplanetary File Storage System or CouchDB are two decentralized storage platforms. These platforms can further be integrated with the proposed system for storing different kinds of multimedia files.

3) Data analytics: The data analytics module in the proposed framework has not been implemented in the tested prototype yet. Hence, the module can be implemented to automatically visualize the data and retrieve insightful information. This is a great promising research direction as it will hugely benefit data analytics industry besides improvising student performance.

ACKNOWLEDGMENT

REFERENCES

[1]   K. Akwei, M. Mutize, and A. L. Alhassan, "Education and economic growth in a developing country," Int. J. Educ. Econ. Dev., vol. 13, no. 2, p. 171, 2022, doi: 10.1504/ijeed.2022.10044287.

[2]   C. Da Wan, M. Sirat, and D. A. Razak, "Economic in Malaysia Towards a Developed Nation," 2018.

[3]   A. Khan and S. K. Ghosh, "Student performance analysis and prediction in classroom learning: A review of educational data mining studies," Educ. Inf. Technol., vol. 26, no. 1, pp. 205–240, 2021, doi: 10.1007/s10639-020-10230-3.

[4]   H. Jarodzka, I. Skuballa, and H. Gruber, "Eye-Tracking in Educational Practice: Investigating Visual Perception Underlying Teaching and Learning in the Classroom," Educ. Psychol. Rev., vol. 33, no. 1, pp. 1–10, 2021, doi: 10.1007/s10648-020-09565-7.

[5]   K. Guill, O. Lüdtke, and O. Köller, "Academic tracking is related to gains in students' intelligence over four years: Evidence from a propensity score matching study," Learn. Instr., vol. 47, pp. 43–52, 2017, doi: 10.1016/j.learninstruc.2016.10.001.

[6]   A. H. Alsaadi and D. M. Bamasoud, "Blockchain-Technology-in-Education-System-A-Survey-Examining-Potential-Uses-of-Blockchain-in-Saudi-Arabia-EducationInternational-Journal-of-Advanced-Computer-Science-and-Applications.pdf," vol. 12, no. 5, pp. 730–739, 2021.

[7]   A. Z. Junejo, M. A. Hashmani, and A. A. Alabdulatif, "A survey on privacy vulnerabilities in permissionless blockchains," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 9, pp. 130–139, 2020, doi: 10.14569/IJACSA.2020.0110915.

[8]   A. Razzaq et al., "Use of Blockchain in Governance: A Systematic Literature Review," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 5, pp. 685–691, 2019, doi: 10.14569/IJACSA.2019.0100585.

[9]   B. Hameed et al., "A review of Blockchain based educational projects," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 10, pp. 491–499, 2019, doi: 10.14569/ijacsa.2019.0101065.

[10]   D. Khan, L. T. Jung, M. Ahmed Hashmani, and A. Waqas, "A Critical Review of Blockchain Consensus Model," 2020 3rd Int. Conf. Comput. Math. Eng. Technol. Idea to Innov. Build. Knowl. Econ. iCoMET 2020, pp. 1–2, 2020, doi: 10.1109/iCoMET48670.2020.9074107.

[11]   A. A. Siyal, A. Z. Junejo, M. Zawish, K. Ahmed, A. Khalil, and G. Soursou, "Applications of Blockchain Technology in Medicine and Healthcare: Challenges and Future Perspectives," Cryptography, vol. 3, no. 1, 2019, doi: 10.3390/cryptography3010003.

[12]   J. B. A. Das, S. K. Mohapatra, and M. N. Mohanty, "Smart Student Performance Monitoring System Using Data Mining Techniques," in Biologically Inspired Techniques in Many Criteria Decision Making, 2022, pp. 337–343.

[13]   B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student' performance prediction using machine learning techniques," Educ. Sci., vol. 11, no. 9, 2021, doi: 10.3390/educsci11090552.

[14]   A. Z. Junejo, M. M. Memon, M. A. Junejo, S. Talpur, and R. M. Memon, "Blockchains Technology Analysis: Applications, Current Trends and Future Directions—An Overview," Lect. Notes Networks Syst., vol. 118, no. May, pp. 411–419, 2020, doi: 10.1007/978-981-15-3284-9_47.

[15]   Ministry of Education, "Quick Facts 2018 Malaysia Educational Statistic," Educ. Data Sect. Educ. Plan. Res. Div. Minist. Educ. Malaysia, pp. 1–48, 2018, [Online]. Available: https://www.moe.gov.my/ penerbitan/1587-quick-facts-2018-malaysia-educational-statistics-1/file.

[16]   Y. Chen, Y. Chen, and A. Oztekin, "A hybrid data envelopment analysis approach to analyse college graduation rate at higher education institutions," INFOR, vol. 55, no. 3, pp. 188–210, 2017, doi: 10.1080/03155986.2016.1262584.

[17]   B. R. Juliana and P. E. Lengkat, "Strategies for Teaching Slow learners in an Inclusive Setup," KIU J. Soc. Sci., vol. 7, no. 1, pp. 225–232, 2021.

[18]   K. Barrington, "The Pros and Cons of Tracking in Schools," 2020. https://www.publicschoolreview.com/blog/the-pros-and-cons-of-tracking-in-schools (accessed Jun. 06, 2022).

[19]   I. Khan, A. Al Sadiri, A. R. Ahmad, and N. Jabeur, "Tracking student performance in introductory programming by means of machine learning," 2019 4th MEC Int. Conf. Big Data Smart City, ICBDSC 2019, pp. 1–6, 2019, doi: 10.1109/ICBDSC.2019.8645608.

[20]   W. Puarungroj, N. Boonsirisumpun, P. Pongpatrakant, and S. Phromkhot, "Application of data mining techniques for predicting student success in English exit exam," ACM Int. Conf. Proceeding Ser., pp. 1–6, 2018, doi: 10.1145/3164541.3164638.

[21]   A. U. Khasanah and Harwati, "A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques," IOP Conf. Ser. Mater. Sci. Eng., vol. 215, no. 1, 2017, doi: 10.1088/1757-899X/215/1/012036.

[22]   M. Adil, F. Tahir, and S. Maqsood, "Predictive Analysis for Student Retention by Using Neuro-Fuzzy Algorithm," 2018 10th Comput. Sci. Electron. Eng. Conf. CEEC 2018 - Proc., pp. 41–45, 2019, doi: 10.1109/CEEC.2018.8674216.

[23] D. M. Krishna*, B. S. B. P. Rani, B. Madhavrao, and S. M. B. Chowdary, "Predicting Student Performance using Classification and Regression Trees Algorithm," Int. J. Innov. Technol. Explor. Eng., vol. 9, no. 3, pp. 3349–3356, 2020, doi: 10.35940/ijitee.c8964.019320.

[24] Y. Wang, R. Wu, Q. Chen, and N. Zheng, "Research on the Construction of Student Learning Effect Traceability System from the Perspective of Blockchain," no. 250, pp. 250–260, 2021.

[25] M. A. Hashmani, A. Z. Junejo, A. A. Alabdulatif, and S. H. Adil, "Blockchain in Education - Track ability and Traceability," 2020 Int. Conf. Comput. Intell. ICCI 2020, no. October, pp. 40–44, 2020, doi: 10.1109/ICCI51257.2020.9247760.

[26] A. Abu, "Educational Data Mining & Students' Performance Prediction," Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 5, pp. 212–220, 2016, doi: 10.14569/ijacsa.2016.070531.

[27] M. K. Yousif and A. Shaout, "Fuzzy logic computational model for performance evaluation of Sudanese Universities and academic staff," J. King Saud Univ. - Comput. Inf. Sci., vol. 30, no. 1, pp. 80–119, 2018, doi: 10.1016/j.jksuci.2016.08.002.

[28] M. Faisal, A. Bourahma, and F. AlShahwan, "Towards a reference model for sensor-supported learning systems," J. King Saud Univ. - Comput. Inf. Sci., vol. 33, no. 9, pp. 1145–1157, 2021, doi: 10.1016/j.jksuci.2019.06.015.

[29] S. Rathore, B. Wook Kwon, and J. H. Park, "BlockSecIoTNet: Blockchain-based decentralized security architecture for IoT network," J. Netw. Comput. Appl., vol. 143, no. June, pp. 167–177, 2019, doi: 10.1016/j.jnca.2019.06.019.

Impact of Mobile Technology Solution on Self-Management in Patients with Hypertension: Advantages and Barriers

Adel Alzahrani[1], Valerie Gay[2]
Faculty of Engineering and Information Technology
University of Technology Sydney
Sydney, Australia

Ryan Alturki[3]
College of Computer and Information Systems
Umm Al-Qura University
Mecca, Saudi Arabia

Abstract—**Hypertension is a major risk factor for cardiovascular morbidity and mortality. It is a condition that increases the high risk of heart, liver, and other diseases. Since hypertension is one of the biggest global public health issues, patients require more interventions to manage their blood pressure. The vast use of mobile phones and applications with medication features has turned a smartphone into a medical device. These tools are helpful for a physician in the treatment of hypertension. Mobile health applications are utilised to manage hypertension at the moment; however, there is a lack of information regarding their efficacy. Smartphones and their applications are evolving quickly hence the rise in the innovation of mobile health applications. Mobile-based applications are helpful in-patient education and reinforce the behaviour through constant reminders, medication, and appointment alarms. The main objective of this study is to determine the impact of mobile health applications on self-management in patients with hypertension and its advantages and disadvantages. We used publications from 2015 and later as a time frame and searched on the first five pages of Google Scholar, JSTOR, Hindawi, PubMed, and ResearchGate. We group all associated terms that might turn up articles on this subject in the search results. The total number of database records that we identified were 213, and the duplicate identified and removed were 117; hence the screened records were 96. The reports excluded based on abstract and title were 31. Articles with full text and have been accessed for final inclusion were 65. The excluded articles were 51, and the studies included in the qualitative analysis were 14.**

Keywords—*Impact; self-management; mHealth; hypertension*

I. INTRODUCTION

Hypertension is a global health challenge in which patients need to be more careful and take more measures to control blood pressure. It is a chronic disease that requires self-management which is the most effective treatment for the patient with hypertension. It is one of the major risks of heart disease, stroke, and renal failure [1]. Despite knowing the fact of the danger of this disease, hypertensive patients do not control their blood pressure [2] . So, it is necessary to encourage them by using smartphone applications.

The arteries in the body are impacted by the prevalent condition of high blood pressure and hypertension. The blood's constant pressure against the artery walls is too high if you have high blood pressure [3]. Worldwide, hypertension is a major contributor to cardiovascular and cerebral events risk and poses a serious threat to public health. With population expansion and aging, it is predicted that more than one in four persons globally have hypertension. A blood pressure reading of 130/80 mm Hg or greater is generally considered hypertension [3]. Hypertension is categorised into four that is normal blood pressure, elevated blood pressure, stage 1 hypertension and stage 2 hypertension [4, 5]. Most people with hypertension do not have symptoms but the few that have experience headaches, nosebleeds and shortness of breath [3].

Mobile application is a type of application software designed to run on a mobile phone, such as a smartphone or tablet PC [6]. Apps are typically small, independent software modules with constrained storage. The mobile applications can be gaming applications, business applications, education applications or travel applications. Mobile applications are developed for Android, IOS and Windows operating systems. Mobile applications are categorised depending on the following factors: the intended users of the app, technology used to develop the app and the expected performance of the application [6].

Mobile health applications have had numerous positive impacts on patients. Mobile health technology, as opposed to a conventional and ineffective telephone connection to doctors and healthcare organisations, allow patients to instantly send secure messages, make appointments, and connect to clinicians 24/7 for telemedicine visits [7]. It is difficult for doctors to keep track of a patient's adherence once they leave the hospital, and with poor medication adherence, the patient's condition can worsen and result in readmission. The mobile health applications will keep track of the patient's medication adherence. The mobile health applications will help in remote monitoring just like home care. The patients will enter data on the mobile applications that will be transmitted to the doctors. This data will include their weight, diet, and blood pressure levels without the need to visit the hospital physically [7]. Mobile health technology means the use of smartphones for health care and self-management [8]. It can integrate everyday life with health care and collects and deliver health services and information in an accessible, convenient, and interactive mode [9]. The use of smartphones and other devices like tablets, and iPad is rapidly increasing. Mobile phones are

important in health as they have become an important platform for using health care applications.

Self-management attempts to give patients with chronic diseases the power to take charge of their care [10]. Mobile health applications have proven to have a positive impact on the level of self-management among patients with chronic conditions. Mobile health applications improve self-management by allowing hypertension patients to record and track their blood pressure levels, diet intake and exercise therapy. Patients with hypertension require self-management on a daily basis because they need to check on their exercise therapy and diet intake [11]. Additionally, several nations have found that increasing self-management is an effective way to lower hypertension [10]. Multiple studies have revealed that remote patient self-monitoring without feedback from healthcare professionals has a negligible impact on lowering blood pressure. The most effective treatment for hypertension is self-management, it concludes the role of patients to manage the symptoms, physical effects, psychosocial, treatment, and changing lifestyle [12]. It is hard to achieve the optimum level of self-management as it needs considerable attention from patients. Therefore, self-management is an important part of healthcare management, as it's important to get patients to

become more involved in managing their blood pressure [4, 13].

II. RESEARCH METHOD

We searched the first five pages of Google Scholar, JSTOR, Hindawi, PubMed, and ResearchGate, with articles published as far back as 2015 as a deadline. We include all relevant terms whose search results can find articles related to this topic. We decided to use an integrated review to determine the impact of mobile applications on patients with hypertension and factors affecting the use of mHealth applications. The review was conducted to establish both the positive and negative impact that mobile health applications have had on the self-management of patients with hypertension. The review was done according to the PRISMA guidelines. The selection and extraction of data were conducted on inclusion and exclusion criteria. The articles that were included were dated from 2015 to 2022 and were concerned with the self-management of hypertension patients.

Out of all the 213 records identified, only 14 studies were included in the qualitative analysis (Fig. 1). Some were excluded basis of title and abstract, and others due to reasons related to no topics of self-management and mobile health applications for hypertension patients.
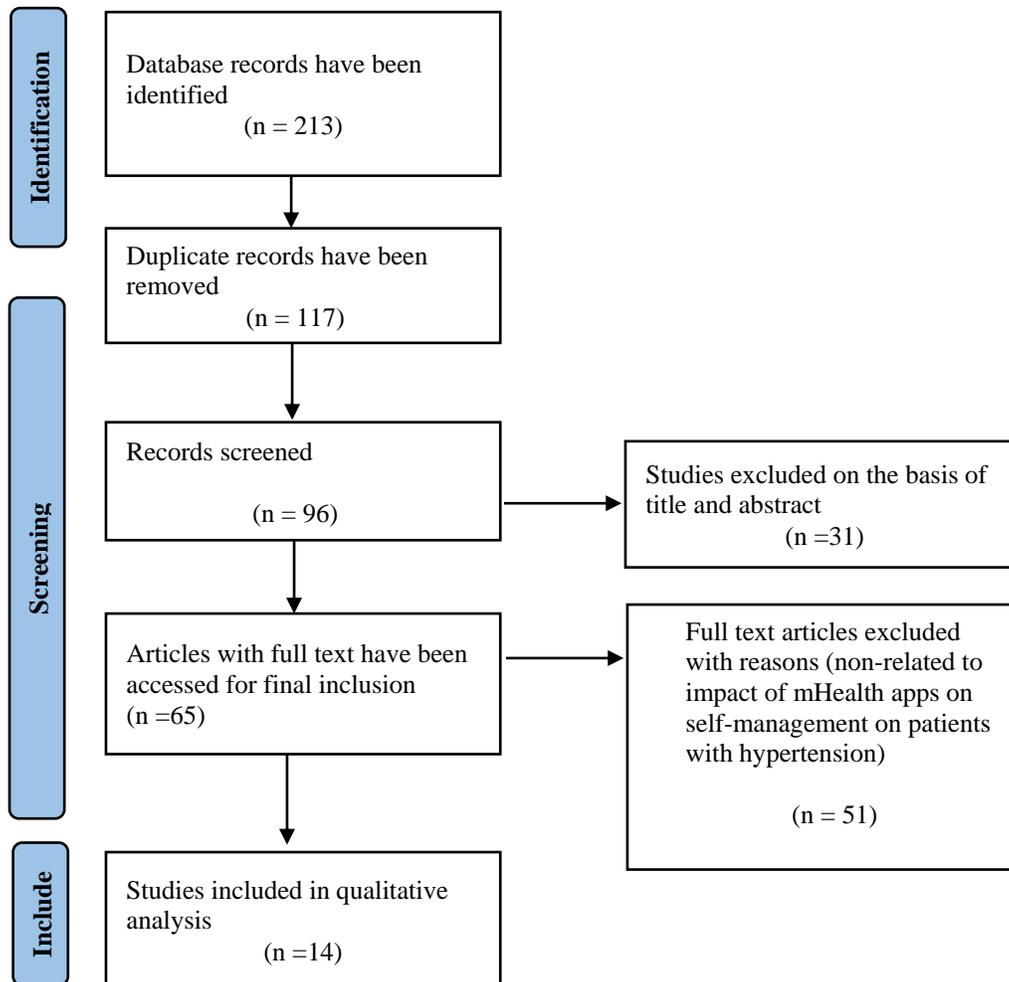
Fig. 1. Prisma Chart for the Investigation.

A. Inclusion Criteria

- The population group in all articles evaluated was 18-79 years.

- The studies were conducted between 2015 and 2022.

B. Exclusion Criteria

- The studies were carried out on less than 18 years and more than 80 years of age.

- The studies were carried out on chronic diseases other than hypertension.

- The studies published in languages other than English were excluded, which likely missed the relevant research.

III. RESULTS AND DISCUSSION

A. Positive Impacts of Mobile Health Applications on Self-Management in Patients with Hypertension

The information monitoring feature for self-care and the app's display, notes and reminders, information access, and information sharing features support the fundamental self-care tactics [3]. Because hypertension patients need to regularly assess their lifestyle, they have found the reminder feature on mobile health applications to be quite beneficial [11, 28]. To remind themselves to take their medications, exercise, check blood pressure, the patients set alarms and reminders. Numerous medications are used to manage hypertension, and not all patients can keep track of when they should take each one. The utilisation of the reminder feature to notify patients when their next appointment is coming is due to the fact that patients occasionally need to see a doctor [28]. Therefore, Reminders have a very positive impact on medication, notification pop up on smartphone screens either to check blood pressure or to take medicines.

Goal setting feature on the mobile application is also helpful because hypertension patients need to check on their weight and food intake daily. By setting a goal of reducing weight and watching their diet on the mobile applications they will be able to achieve their set targets. Moreover, setting goals is a powerful tool for encouraging physical exercise. The goal setting feature allows the patients to set a target and observe their improvements towards achieving their goals. A sedentary lifestyle is closely linked to hypertension. Exercise has been demonstrated to delay the onset of hypertension [29]. Effective blood pressure reduction has been shown to occur with both aerobic and resistance training hence the need for hypertension patients to set a goal of exercising daily. The hypertensive patient can set goals using medical health applications. He can generate a checklist or can use a default list of common goals that have to be achieved or recommended by the physician. Through these applications, hypertensive patients record the achieved goal in progress goals. Mobile health applications can also populate these goals and show overall progress. Achieved or progressed goals can also be linked with calendars and deadlines. Mobile health applications provide tips on one base of these achieved goals [30].

Some applications for mobile health have an educational component that gives more details about the patient's illness and how they may manage it at home. Patients may comprehend the data they input into mobile applications, such as their blood pressure and salt intake, with the aid of the teaching component. The educational function will enable patients to determine whether their condition is improving or declining. There are apps that can give you the contents, including the calories, fat, and sugar, of any specific restaurant item [29]. You may make an informed choice that supports you by rapidly comparing the nutrition facts to your daily objectives. Hypertensive patients also use the educational features of mobile health applications in which applications suggest them food intake, weight management as well as physical activities. These applications are also used to ask a question from a physician or to seek his opinion. The motivational feature of mobile health applications is also very important as it helps people to continue the medication and adherence and finally, they get rid of the disease.

B. Barriers to using Mobile Health Applications on Self-Management for Hypertension Patients

Lack of privacy is one of the major barriers to adoption of mobile health applications. The mobile applications contain patients' crucial information such as their email address, phone number and home address which when a breach occurs the information can be used by cyber criminals for malicious reasons [11, 31].

In addition, usability is a crucial concern for mobile health application systems because it contributes significantly to their success. However, Poor usability is a major obstacle to the widespread use of mobile applications [32]. Therefore, M-health could have unforeseen, detrimental effects if usability is not taken into account, including an increase in medical errors and issues with provider communication [31]. Some applications are very tedious and the data entry process is annoying. Even patients who are familiar with technology found it difficult which is one of the reasons for not using mobile health applications. Numerous mobile health applications are not regulated, which means they may provide consumers inaccurate information and seriously harm their health [29].

Another barrier is the prescription of mobile applications, patients think that only those applications can be beneficial and are recommended by their doctor. Finding an appropriate application is also a barrier to the frequent use of mobile applications. Lack of evidence about the effectiveness of medical health applications is also a barrier, patients demand evidence of whether either application feature is beneficial for hypertension or not [33].

C. Examining the Effects of using Mobile Health Apps for Hypertension Self-Management by Patients in Table I

This study aimed to review the impacts of mobile applications on self-management among patients with hypertension. In this review, 14 studies have been reviewed that were conducted recently between 2015 and 2022. There may be differences in study design however the main crux of these studies was to evaluate the impacts of mobile health applications on self-management in patients with hypertension.

TABLE I.   DETAILS OF ARTICLES INVESTIGATING THE IMPACTS OF MOBILE HEALTH APPLICATIONS ON SELF-MANAGEMENT IN PATIENTS WITH HYPERTENSION BETWEEN 2015 AND 2022

Study (year)	Study title	Aim	Study design
[14]	The Effectiveness of Self-Management of Hypertension in Adults Using Mobile Health.	The goal of this systematic review was to find out how well mHealth helps adults better manage their own hypertension. Blood pressure (BP), blood pressure control, medication management, self-management behaviour, and costs were used to measure the results.	Systematic Review and Meta-Analysis
[15]	Evaluation of the Effectiveness of the Mobile Application on Adherence of Patients with Arterial Hypertension.	The purpose of the study was to evaluate the impact of a mobile application on medication adherence in individuals with documented arterial hypertension.	A multi-centre randomized controlled
[16]	Smartphone Apps to Support Self-Management of Hypertension,	The study's goal is to evaluate and describe all existing applications that aid in self-management of hypertension, as well as to analyses their functionalities.	Review and Content Analysis
[17]	Apps on Google Play Store to assist in self-management of hypertension in Indian context.	The purpose of this research was to examine the analytical and practical features of hypertension self-management apps found in the Google Play Store.	Features analysis study
[18]	Comparing the Acceptance of Mobile Hypertension Apps for Disease Management Among Patients Versus Clinical Use Among Physicians.	The goal of this study is to find out what makes people in German-speaking countries accept mobile hypertension apps for personal use and what makes doctors accept them for clinical use.	Cross-sectional Survey
[19]	Mobile health applications for the management of primary hypertension.	The objective was to evaluate the impact of mobile Health applications on blood pressure control and medication adherence.	A multicenter, randomized, controlled trial
[20]	Chinese Mobile Health APPs for Hypertension Management.	The goals of this study are (a) to provide recommendations for developers in industry and (b) to aid hypertensive patients in selecting suitable APPs by comparing and contrasting the efficacy of hypertension management APPs released in the Chinese market; (c) to gain an understanding of the general circumstances, characteristics, issues, and trends in hypertension management mHealth APPs; and (d) to determine the gaps between products from mainland China and those from outside of mainland China.	A Systematic Evaluation of Usefulness
[21]	Effects of using a mobile health application on the health conditions of patients with arterial hypertension.	To verify the mobile intervention and effects of using a health application on patients with hypertension.	A pilot trial in the context of Brazil's Family Health Strategy
[22]	A content analysis of smartphone-based applications for hypertension management.	This study aims to determine the content of mobile hypertension applications available on Google Play and Apple iTunes.	cross–sectional study
[23]	Impact of a mobile phone app on adherence to treatment regimens among hypertensive patients.	The aim of this study is to investigate the impacts of mobile applications and medication adherence among hypertensive patients.	A randomized clinical trial study
[24]	Effect of an Integrative Mobile Health Intervention in Patients with Hypertension and Diabetes.	The object of this study is to evaluate the use of an integrated mobile application for hypertension, obesity, and blood pressure.	Crossover Study
[25]	The effect of the mobile "blood pressure management application" on hypertension self-management enhancement.	This study assessed the impact of a self-management application on patient adherence to hypertension therapy in context of the increasing use of mobile health in healthcare.	A randomized controlled trial
[26]	The Emerging Role of Mobile-Health Applications in the Management of Hypertension	This study examined the current research on m-health systems and how m-health might impact the management of hypertension.	Study review
[27]	Mobile Health Applications and Medication Adherence of Patients with Hypertension.	The goal of this study is to figure out how health apps impact how people with hypertension take their medicine and what causes those effects.	A systematic Review and Meta-Analysis

These studies have appraised the applications with a combination of functionalities and features, as some were designed for self-monitoring of blood pressure, some to measure blood pressure and other was helpful in log and record updating of a patient. In most studies, participants accepted the applications for blood pressure and self-monitoring. Due to varying places, study designs, and quality of results, it is inconclusive to define the most evident impacts of mobile applications, and which is more effective in lowering blood pressure worldwide. However, it is finalized that applications with more comprehensive functionalities and features were more effective in self-management.

There are different study designs evaluated in this paper, (4/14) studies are randomized control trials, (3/14) are systematic reviews and meta-analysis, (2/14) are cross-sectional surveys and (2/14) are reviews and content analysis whereas other study designs are included in this paper are, features analysis study, pilot trial in the context of Brazil's family health strategy, crossover study. All these studies were on the effectiveness of mobile health applications in patients with hypertension, and many have also assessed the experience and satisfaction with these applications. These are conducted in different environments for different purposes, so there is no standard duration as well as the number of participants.

A systematic review conducted in China to determine the effectiveness of self-management of hypertension patients who use mobile health indicated that mHealth self-management treatments successfully lower blood pressure [14]. Patients with hypertension require self-management regularly. Maintaining blood pressure with medication and avoiding problems are the fundamental therapies for hypertension. Because hypertension is seen as a chronic condition, self-management should be a regular part of the patient's routine. Improvements in self-management behaviour and medication adherence were seen as a result of this review. The mHealth intervention that was the most effective combined personalized messaging with interactive communication and diverse features. With the prevalence of hypertension increasing internationally, mHealth presents a potentially efficient approach for hypertension self-management and control. It is simple to incorporate mHealth into current healthcare systems [14]. Therefore, according to this study, self-management education can increase patients' understanding of hypertension and enable early diagnosis of excessive blood pressure.

A study conducted in the UK to evaluate various mobile applications that support the self-management of hypertension patients and their functions concluded that many mobile health applications but only a few could be considered effective [16]. According to the study, many applications are ineffective because they do not have security measures and do not provide evidence of their effectiveness and usability making it difficult for hypertension patients to determine which mobile health applications are effective [16]. In Brazil, a study was conducted to verify the effects of mHealth applications on arterial hypertension patients. The study indicated that technology use with health information is a development that can benefit hypertension patients' therapeutic plans by promoting better treatment compliance, healthier lifestyle choices, and better health outcomes [21]. mHealth systems help patients manage hypertension in the following ways: setting alarms and reminders for patients to take their medications, linking patients' blood pressure reports to their electronic medical records so that their doctors can check them, providing patients with feedback on their blood pressure tendencies, and acting as point-of-care blood pressure sensors [21]. mHealth apps with alarms and reminders can increase medication compliance, while apps that share ambulatory blood pressure records with patients' doctors can advance patient-health professional communication.

A study conducted in Kazakhstan was aimed at evaluating the effectiveness of mobile health applications on patients diagnosed with arterial hypertension. The study concluded that the mobile health applications did in fact help with patient adherence. However, more research is needed to allow the wider implementation of mobile health applications in healthcare [15]. A study done in India to identify the mobile applications in the Google play store that support self-management for hypertension patients indicated that the applications mainly focused on recording blood pressure and providing statistics and trends and not self-management techniques, hence the need for improvement [17]. Therefore, Developers of hypertension apps should now focus on including additional self-management features.

According to a study conducted in German speaking countries to determine the acceptance of mobile health applications for patients and physicians of hypertension by observing the modes of service and delivery of health services indicated that the expected performance of the applications was the main factor for the acceptance of the mHealth applications [18]. A systematic review done in China to evaluate the effectiveness and usefulness of mobile health applications for hypertension patients concluded that the usefulness of the mobile applications was unsatisfactory hence the need for improvement because there is a rise in the need for mHealth applications for hypertension patients [20].

In Palestine (Gaza Strip), a study was conducted to identify the impact of mobile health applications on adherence to medication on hypertension patients indicated that the applications were indeed effective and reduce cardiovascular mobility and morbidity [23]. A study conducted in South Korea to investigate the effect of using integrated mobile health intervention on patients with hypertension and diabetes indicated that even though the apps body fat and glucose levels they failed to indicate clinical improvement on the patients. The medical information and reminder app was also more appealing to patients than a food diary and fitness monitor [24].

A study conducted in Iran to verify the effect of mobile blood pressure management application on self-management for patients with hypertension indicated that the application is helpful for hypertension patients in developing countries [25]. An analysis is done in France to determine the effectiveness of smartphone-based applications on hypertension management showed that the applications were helpful by recording the patient's blood pressure, heart rate and salt intake. Additionally, Medical app development for HTN, especially medical devices, needs more oversight. [22]. In China a study was conducted to observe the impact of mHealth apps on blood

pressure and medical adherence and the results showed that the mHealth apps helped to lower blood pressure, improve blood pressure control, and medication adherence, all of which can assist to minimise the economic and social burden of hypertension in people [19].

A review conducted to identify the role of mobile health applications on the management of hypertension indicated that improvements in blood pressure outcomes might be achieved through the use of smartphone applications that serve as medication reminders, offer patients with individualised feedback, and provide two-way communication with their healthcare professionals [26]. A study conducted to analyze the impacts of mobile health application on medical adherence of hypertension patients and the underlying factors showed that mobile health applications have a positive impact on medical adherence of hypertension patients and helps improve their conditions [27].

All the above-mentioned studies agreed as a whole that smartphone-based has the potential to manage chronic diseases considerably. The results are in accordance with the impacts of mobile health applications in self-management for patients with hypertension. The results of the application with automatic feedback and recommendation features without involving a physician were more promising in controlling blood pressure. In the same way, the mobile applications in which healthcare products were involved in remotely monitoring the patients with automatic feedback also had a significant impact on blood pressure. It can be assumed that both approaches are effective and beneficiary.

In all the included studies one thing is common hypertensive patients appreciated the features of medical health that have saved their time as compared to traditional ways. They found medical health application valuable if it is simple to use, provide specific and to-the-point recommendations and instructions and share data with their desired individual. Though tracking is the main function, it may distract the patients at the same time.

This study is one of the systematic reviews which explore the impacts of mobile health applications in self-management for patients with hypertension. Lastly, this study might be a roadmap for future studies on this topic.

IV. CONCLUSION

This study determined that mobile health applications are acceptable and assist in controlling and lowering the blood pressure of hypertensive patients. It is also concluded that applications with more comprehensive features and functionalities are more effective. This study also demonstrates its positive impacts and barriers which help the technology giants to design user-friendly applications. In positive impacts, it highlights the most effective impacts, such as reminder, goal setting as well as educational feature of these applications. At the same time, the barrier in patients' perspectives is also highlighted. Implementation of these applications more effectively would help to control the disease which requires self-management and self-monitoring. According to studies, mobile health applications can benefit people with hypertension by keeping track of their diet, salt intake, blood pressure, and heart rate, which can help their condition. The majority of studies showed that mobile health applications had a good effect. To boost the effectiveness of mobile health applications for patients self-managing their hypertension, we advise conducting more research in this area.

REFERENCES

[1] Chobanian, A.V., et al., Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. hypertension, 2003. 42(6): p. 1206-1252.

[2] Chow, C.K., et al., Prevalence, awareness, treatment, and control of hypertension in rural and urban communities in high-, middle-, and low-income countries. Jama, 2013. 310(9): p. 959-968.

[3] Mills, K.T., A. Stefanescu, and J. He, The global epidemiology of hypertension. Nature Reviews Nephrology, 2020. 16(4): p. 223-237.

[4] ALZAHRANI, A., V. GAY, and R. ALTURKI, The Design and Development of a Mobile Technology Solution to Self-Monitor Hypertension (HTN) And Improve Health and Fitness Levels Among Saudi Adults. 2021.

[5] Whelton, P.K., et al., 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Journal of the American College of Cardiology, 2018. 71(19): p. e127-e248.

[6] Dar, H., et al., A systematic study on software requirements elicitation techniques and its challenges in mobile application development. IEEE Access, 2018. 6: p. 63859-63867.

[7] Baxter, C., et al., Assessment of mobile health apps using built-in smartphone sensors for diagnosis and treatment: systematic survey of apps listed in international curated health app libraries. JMIR mHealth and uHealth, 2020. 8(2): p. e16741.

[8] Whittaker, R., Issues in mHealth: findings from key informant interviews. Journal of medical Internet research, 2012. 14(5): p. e1989.

[9] Eren, H. and J.G. Webster, Telemedicine and Electronic Medicine. 2018: CRC Press.

[10] Duan, H., et al., Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR mHealth and uHealth, 2020. 8(2): p. e14466.

[11] Alzahrani, A., V. Gay, and R. Alturki, Exploring Saudi Individuals' Perspectives and Needs to Design a Hypertension Management Mobile Technology Solution: Qualitative Study. International Journal of Environmental Research and Public Health, 2022. 19(19): p. 12956.

[12] Ryan, P. and K.J. Sawin, The individual and family self-management theory: Background and perspectives on context, process, and outcomes. Nursing outlook, 2009. 57(4): p. 217-225. e6.

[13] Alessa, T., et al., Mobile apps to support the self-management of hypertension: systematic review of effectiveness, usability, and user satisfaction. JMIR mHealth and uHealth, 2018. 6(7): p. e10723.

[14] Li, R., et al., The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR mHealth and uHealth, 2020. 8(3): p. e17776.

[15] Nurakysh, S., et al., Evaluation of the Effectiveness of the Mobile Application on Adherence of Patients With Arterial Hypertension. Acta Informatica Medica, 2022. 30(1): p. 18.

[16] Alessa, T., et al., Smartphone apps to support self-management of hypertension: review and content analysis. JMIR mHealth and uHealth, 2019. 7(5): p. e13645.

[17] Kaur, M., et al., Apps on Google Play Store to assist in self-management of hypertension in Indian context: features analysis study. Mhealth, 2022. 8.

[18] Breil, B., C. Salewski, and J. Apolinário-Hagen, Comparing the acceptance of mobile hypertension apps for disease management among patients versus clinical use among physicians: cross-sectional survey. JMIR cardio, 2022. 6(1): p. e31617.

[19] Gong, K., et al., Mobile health applications for the management of primary hypertension: A multicenter, randomized, controlled trial. Medicine, 2020. 99(16).

[20] Liang, J., et al., Chinese mobile health apps for hypertension management: a systematic evaluation of usefulness. Journal of healthcare engineering, 2018. 2018.

[21] Debon, R., et al., Effects of using a mobile health application on the health conditions of patients with arterial hypertension: A pilot trial in the context of Brazil's Family Health Strategy. Scientific Reports, 2020. 10(1): p. 1-10.

[22] Kumar, N., et al., A content analysis of smartphone–based applications for hypertension management. Journal of the American Society of Hypertension, 2015. 9(2): p. 130-136.

[23] Abu-El-Noor, N.I., et al., Impact of a mobile phone app on adherence to treatment regimens among hypertensive patients: A randomised clinical trial study. European Journal of Cardiovascular Nursing, 2021. 20(5): p. 428-435.

[24] Oh, S.W., et al., Effect of an Integrative Mobile Health Intervention in Patients With Hypertension and Diabetes: Crossover Study. JMIR mHealth and uHealth, 2022. 10(1): p. e27192.

[25] Bozorgi, A., et al., The effect of the mobile "blood pressure management application" on hypertension self-management enhancement: a randomized controlled trial. Trials, 2021. 22(1): p. 1-10.

[26] Thangada, N.D., et al., The emerging role of mobile-health applications in the management of hypertension. Current cardiology reports, 2018. 20(9): p. 1-9.

[27] Mikulski, B.S., et al., Mobile health applications and medication adherence of patients with hypertension: a systematic review and meta-analysis. American Journal of Preventive Medicine, 2021.

[28] Qudah, B. and K. Luetsch, The influence of mobile health applications on patient-healthcare provider relationships: a systematic, narrative review. Patient education and counseling, 2019. 102(6): p. 1080-1089.

[29] Pires, I.M., et al., A research on the classification and applicability of the mobile health applications. Journal of personalized medicine, 2020. 10(1): p. 11.

[30] Dicianno, B.E., G. Henderson, and B. Parmanto, Design of mobile health tools to promote goal achievement in self-management tasks. JMIR mHealth and uHealth, 2017. 5(7): p. e7335.

[31] Bol, N., N. Helberger, and J.C. Weert, Differences in mobile health app use: a source of new digital inequalities? The Information Society, 2018. 34(3): p. 183-193.

[32] Alzahrani, A.S., et al., Towards Understanding the Usability Attributes of AI-Enabled eHealth Mobile Applications. Journal of Healthcare Engineering, 2021. 2021.

[33] Byambasuren, O., et al., Barriers to and facilitators of the prescription of mHealth apps in Australian general practice: qualitative study. JMIR mHealth and uHealth, 2020. 8(7): p. e17447.

IRemember: Memorable CAPTCHA Method for Sighted and Visually Impaired Users

Mrim Alnfiai, Sahar Altalhi, Duaa Alawfi
Information Technology Department
College of Computer Science and Information Technology
Taif University, Taif, Saudi Arabia

Abstract—**A CAPTCHA is used to automatically differentiate between human users and automated software to prevent bots from accessing unauthorized websites. Most proposed CAPTCHAs are not accessible to visually impaired users because of the memorability of the CAPTCHA's numerical digits. Recalling six random spoken digits is a difficult task for any human. Visually impaired users must typically play the audio several times to memorize the spoken digits in the correct order. The authors reviewed existing CAPTCHAs for visually impaired users and concluded that the high cognitive load is more susceptible to response errors due to extensive challenge digits intended for visual users. Thus, the authors proposed a novel method that improves current audio CAPTCHA by enhancing the display of the challenge and improving the memorability of its phraseology. The proposed CAPTCHA presents short common phrases, such as "piece of cake." After hearing or seeing the phrases, the users are required to type the first letter of each word from the presented phrases, such as POC for a piece of cake. The study results of 11 visually impaired users concluded that the memorability and success rate for the IRemember CAPTCHA was 82.72%, compared to the audio CAPTCHA at only 48.18%. It has also demonstrated higher memorability and less workload than the traditional audio method. This research indicates that using common knowledge and experience in the design process for a CAPTCHA method for these users can enhance performance and minimize workload and, hence, error rates.**

Keywords—CAPTCHA; blind users; visually impaired users; memorability; accessibility

I. INTRODUCTION

Today, we can consider web services as the backbone of life, especially since the inception of the COVID-19 pandemic. Many services are provided through web-based applications. The Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) is used as an authentication protocol. It has become a necessary step for online services, especially ones designed for public use, free usage cost, or registration to the end user. If you use the internet regularly and deal with a web form, you will typically encounter a CAPTCHA test. Being able to utilize the services on these applications is a right for everyone worldwide. A study by the World Health Organization shows that about 217 million people have some kind of visual impairment, and almost 36 million are totally blind [1]. It is also determined that the rate of people with visual impairments will increase because of the population expansion. CAPTCHAs are annoying for humans to solve and can be difficult or impossible to figure out for those with disabilities. CAPTCHAs have to work for all people, disabled or not, because this tool prevents bots and viruses, but people must be able to use them with ease. Therefore, it is essential to develop new technologies that people with visual impairment can effectively use.

CAPTCHA, initially developed in 1997 by Alta Vista, is a test used to determine whether or not the user is human. It is a significant step when dealing with a web form to prevent automatic bots and harmful invasions. It does this by offering tests that humans can pass, but computer programs cannot. The term CAPTCHA was instituted in 2000 by Luis von Ahn, Nicholas J. Hopper and Manuel Blum of Carnegie Mellon University and John Langford of IBM [2]. Essentially, CAPTCHA blocks robot software from submitting fake or misused online requests. It is also used to protect the integrity of online polls by preventing hackers from sending in repeated false responses using robots.

Currently, CAPTCHA is categorized into five main categories: text-based, image-based, video-based, puzzle-based, and audio-based [20], the text-based type being the most widely deployed. Google, Yahoo, and Microsoft websites have deployed text-based CAPTCHAs for years. Many techniques are used for creating text-based CAPTCHAs, such as Gimpy, which provides an arbitrary number of words and displays them in a distorted manner. The EZ-Gimpy technique uses only one word. The Gimpy-r technique uses random distorted letters with a noise background. Simard's HIP CAPTCHA selects random letters and numbers and then uses arcs and colors to distort them [3]. The problem is that text-based CAPTCHAs are not suitable for individuals with visual impairments since they have a hard time remembering letters and typing them.

Image-based CAPTCHAs require users to select matching images or images that don't fit. A Braille display cannot present an image; therefore, a user who needs to use a Braille display cannot solve the challenge. So, if screen reader software could display the CAPTCHA to a blind user that would defeat the purpose of preventing automatic bots and harmful actions since bots would be able to solve the challenge as well. This is a severe limitation of image-based CAPTCHAs, which cannot be used wherever accessibility has to be guaranteed by governmental institutions.

The third category is video-based CAPTCHAs, which rely on typing information from the video that only humans can notice.

Puzzle-based CAPTCHAs have pictures divided into parts with different techniques and ask the user to find the missing part.

Audio-based CAPTCHA is the last category, and that presents an audio recording with a noisy background of a series of letters or numbers [14]. This CAPTCHA helps visually impaired users, but it is difficult to interpret the results. Many studies and schemes have been done in this area to develop it. One of these studies did try to help blind people by not making them type the word. This mechanism is called HearAct CAPTCHA, and it uses a tap or swipe to allow the user to determine if it is a specific letter in a sound-maker name. Users with comprehensive spelling and English vocabulary can solve this challenge. So, remembering how to spell the word is the gist of this type of CAPTCHA [1]. reCAPGen schemes depend on choosing audio clips from old radio programs, podcasts, and YouTube lectures and adding the amount of noise. Noising minimally affects the human ability to solve the generated audio CAPTCHAs [4].

The Last Two Words (LTW) scheme achieves a success rate of 78 percent with sighted users and 81 percent with visually impaired users with an average response time of 15 seconds [5]. These CAPTCHAs exploit the human effort to generate transcriptions for audio files with high accuracy [6]. This scheme relies on remembering the last two words and knowing how to pronounce them, which is difficult for non-native English speakers.

This paper introduces a new technical solution to address visually impaired individuals' challenges in answering CAPTCHA questions. This solution allows browsing and accessing internet content while maintaining website security. Using an approach to help users recall the response to the CAPTCHA question improves the accessibility. It makes solving the CAPTCHA challenge the idea behind audio, improves accessibility, and allows people with visual impairments to solve the CAPTCHA independently. This approach uses common memorable phrases to enable users to memorize the answer quickly without repeating it several times. In this technique, we will use these sentences in the audio CAPTCHA and ask the user to input the first letter of each word. For example, if we use the common phrase, "How are you?" they are required to input the first letter of each word, in this case, "hay." We will display common phrases to enhance the memorability of the CAPTCHA. The length of the phrases must not exceed four words. Developing this new form of CAPTCHA allows visually impaired users to memorize it.

The objectives of the proposed CAPTCHA method are:

1) To overcome the limitations associated with traditional audio CAPTCHA methods for visually impaired users.

2) To build accessible and secure CAPTCHA challenges for both visually impaired and sighted users.

3) To improve memorability for visually impaired users by reducing the mental workload needed to solve the

CAPTCHA and minimizing their cognitive load needed to recall the CAPTCHA digit.

4) To minimize error rate and completion time to solve the challenge.

The contribution of this research is as follows:

- Creating a novel and memorable CAPTCHA method that uses common phrases in creating CAPTCHA challenges, allowing users to respond with few characters to minimize completion time.

- Developing a CAPTCHA method that is memorable and easy to solve for visually impaired individuals.

- Presenting a user study to investigate the performance of the proposed CAPTCHA and measure the accuracy, usability, and cognitive load required to complete the task. Participants solved a set of CAPTCHA challenges during the experiment.

The paper is organized as follows: a review of the CAPTCHA systems previously designed for visually impaired people, an explanation of the IRemember CAPTCHA design, the methodological approach to testing the proposed CAPTCHA method, results, conclusions, and considerations for future research.

II. RELATED WORK

With the spread of the Internet and its technologies, it has become necessary to propose and develop online verification methods, CAPTCHA, and facilitate it to all society members in proportion to their special needs, such as people with autism spectrum disorders [7], learning disabilities [8], and blinds or visual impairment [9]. For Internet users with a visual disability, our target in this research, the audio CAPTCHA overcomes limitations of other CAPTCHA types because it is based on what they hear, not what they see [1].

Human-Interaction Proof, Universally Usable (HIPUU) with both versions [10] and [11] is an audio CAPTCHA designed to overcome the traditional audio CAPTCHA limitations. The task is based on choosing the appropriate word to describe an image from a drop-down list. In addition, an alternative for the image is provided by playing an audio file corresponding to the same image. In this way, the user can choose the preferred media representing CAPTCHA. In version 1, the list contains 15 different choices and some false decoy answers, which makes it vulnerable to brute force attacks because of the small number of choices. However, HIPUU version 2 increases the security level by requiring solving multiple tasks, each with 35 choices and some false decoy answers. In contrast, version 3.0 requires users to type in the solution using a keyboard instead of choosing from the drop-down menu, which is a time-consuming task for visually impaired users. Similarly, SoundsRight [12] is another audio-based CAPTCHA that asks the user to press the space button each time the users hear a particular sound.

In [13], the CAPTCHA is made of a simple text-mathematical problem converted into speech/audio using text to speech (TTS) system. The user has to listen to the question and then answer by only typing the answer as a number. On

the one hand, this method relies on computer limitations to solve numerical spoken questions, but it is challenging to generate a large number of tests. Thus, it is not a practical solution [14]. Similarly, in [15], the authors propose four scenarios of audio CAPTCHA includes: ask users to calculate a running total, count the occurrence of a character in an alphanumeric series, transcribe the alphanumeric characters they heard, and last which is the categories prototype that asks users to count the number of sounds, in a series, that belonged to a specific category. Three out of four designs have increased vulnerability against random guessing attacks, which makes it limited to situations where high security is not critical. HuMan [16] asks the user a question from his/her favorite field and types the answer in a text box. It is easier than the text or image-based CAPTCHAs for blind users, but still has a high probability of spelling mistakes since it requires a full written answer.

In 2021, Mathai et al. improved the existing audio CAPTCHA as they stated that the traditional audio CAPTCHA is complicated to solve by visually impaired users. In their developed method, users hear a distinctive sound, and they are required to count how many times they hear the intended sound. The distinctive sound is a combination of sound, background noise, and music produced using Generative Adversarial Networks to make it difficult for bots to distinguish the answer. The main advantage of this method is allowing users to solve the challenges using numbers. However, it is too complicated and time-consuming to prepare the sound challenges.

Another distinct study where the CAPTCHA method is developed using an OTP-based QR code [17]. Users can decrypt the OTP using a unique key. It is a secure method, but it requires having another phone device to scan the QR code. Noorjahan designed a CAPTCHA method using fingerprints (2019). This method is accessible for visually impaired users, but it requires external hardware to complete the CAPTCHA solving task, which is a fingerprint scanner. Another drawback of this system is that it is applicable to smartphone devices.

HearAct [18] is an audio-based solution that does not require typing text but reaction. The user listens to a sound of what is called a "sound-maker" and answers a spoken question by tapping if the answer is true. Otherwise, swap left or right. The work achieved a success rate of about 82.05%, which is twice as much as the success rate of the traditional audio CAPTCHA and a faster completion time of 55.35 seconds than 65.64 seconds in the traditional audio CAPTCHA.

A recent and related work [19] proposes Machine Learning (ML) model to classify humans from bot behavior. To solve the CAPTCHA task, the user will be asked by audio to draw a simple shape. Then, the cursor movement will be captured and converted to an image and sent to the backend ML model to decide if it was a human or not.

The Generative Adversarial Network (GAN) algorithm is used in [20] to generate music with unique sound samples and a white noise layer. The CAPTCHA challenge asks the user to count the number of unique sounds made by percussion instruments such as drums. This solution simplifies the task for blind or impaired users by requesting to type the answer as

a number. In addition, auto-generation of music and adding the noise layer make it more secure against speech-to-text bots and APIs.

The work [21] is a simple but efficient solution developed to allow visually impaired users to verify their humanity by walking at least five steps. If this step is done successfully, then the user will proceed to enter a username and password. Otherwise, the verification process will fail, and the user will be prevented from accessing the service that he/she wants.

By reviewing these studies and taking in mind the few numbers of research on audio CAPTCHAs and their accessibility and usability for people with vision impairments, the door is opened for more investigation and encouraged more studies to improve the accessibility, memorability, usability, and speed of solving CAPTCHA tasks for blind Internet users. This motivated us to propose a new solution based on commonly spoken phrases to increase memorability and requires entering only the first letter from each word to speed the task for blind or visual impairment users.

III. IREMEMBER CAPTCHA DESIGN

The proposed method is audio and text-based, where users can see or listen to a short common phrase and then identify the beginning letter of each word in the phrases displayed and spoken. The proposed method is built based on the theoretical concept that people can remember phrases better than random and varied letters [22]. The IRemember CAPTCHA features include:

1) Requiring users to listen to the provided phrase.
2) Recognizing the beginning letter of each word in the phrase.
3) Typing only the first letter of each word in the response field.
4) Ability for user to double tap when updating the CAPTCHA challenge.

For example, users will receive a question like, "Please type the first letter in each word in the following phrase." Then, users are required to listen to the given phrase, "How are you?" and to determine the first letter of each word within the phrase. Once users recognize the beginning letters of the phrase, they must type "hay." If users want to update the challenge, they can double tap anywhere on the screen to receive a new challenge (see Fig. 1).

The proposed CAPTCHA differs from existing methods in two ways: the identification stage and solving stage. In the identification stage, users listen to the phrase and are required to recognize the first letter of each word in the phrase. In the solving phase, users can type 3 to four letters that represent the first letter in each word. Doing so will minimize the time spent to solve the CAPTCHA as users have only to type three digits.

To improve the accessibility of CAPTCHA, two sounds are used. The first speaks the CAPTCHA question, "Type the first letter of each word in the phrase," and the second speaks the phrases. This feature can enhance the method's usability and simplify the differentiation between the question and challenge. To minimize cognitive load, the CAPTCHA

method has one basic CAPTCHA statement, which is "Type the first letter of each word in the CAPTCHA challenge."
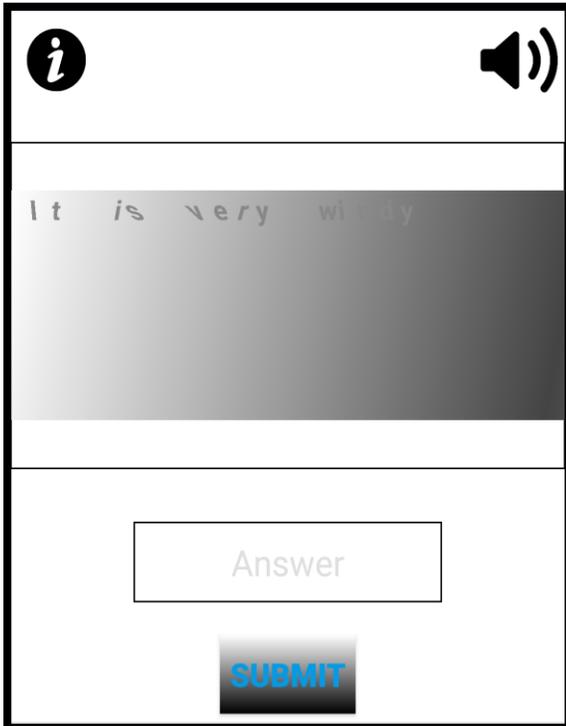
Fig. 1. IRemember CAPTCHA Method Interface.

The proposed approach is different from existing methods in both the identification and solving phases. In the initial phase, the challenges are represented as a phrase, which is easier to recall than presenting different letters. In the second phase, users are required only to type a few letters. To enhance security and reduce confusing recognition software, the CAPTCHA method presents the entire phrase that is distorted using text and audio. With this CAPTCHA style, it is easier for users to identify the first letter of each word of the phrase while simultaneously being difficult for the software to determine the letters.

A more secure and accessible CAPTCHA method is provided by developing the proposed CAPTCHA method. The new method will also benefit a wide range of users as it will simultaneously provide visual and audible challenges. Doing so makes CAPTCHA universal and meets usability and accessibility standards. Another valuable benefit of developing IRemember CAPTCHA is creating an accessible and memorable CAPTCHA method that will enable visually impaired individuals to solve CAPTCHA challenges by themselves without asking for assistance from others to pass this stage. By doing so, the visually impaired will be independent and able to access online materials without support. Another critical benefit of this CAPTCHA is minimizing the time spent to complete the challenge and reducing the cognitive load needed to recall the CAPTCHA digits. In addition, developing an easy-to-recall challenge will minimize the error rate and enhance users' satisfaction with the developed method.

The proposed method will be easier for sighted users to solve as they are only required to type the first letter of each word, which is much easier to recognize and type since its only six digits. From the website security perspective, the proposed method will be more secure as it will use different background noise to prevent software recognition and determine the challenge response. As a result, developing this method will provide websites with a more resistant and secure approach to various attacks. It is more confusing for software to recognize the content of a challenge, which is the entire phrase. Additionally, providing different methods for presenting CAPTCHA challenges will improve accessibility for humans as it considers people's disabilities. In general, building accessible CAPTCHA for visually impaired users and other individuals with disabilities prevents the technical barriers they have faced when browsing web pages and enables them to do their online tasks without complications.

A. Selection of Phrases

The authors gathered a set of common phrases or idioms that individuals typically use in their daily lives that are recognizable by general users without requiring too much concentration and mental effort to memorize the given digits. The used phrases were collected from the EnglishSpeak website [22]. In the implementation phase, the authors used 250 common English phrases to build the first version prototype's library.

B. IRemember CAPTCHA Task Complexity

The primary goal of the IRemember CAPTCHA is to create an accessible and memorable method for individuals with no or low vision. The method presents a phrase that contains at least three words to enhance the complexity of the audio CAPTCHA challenge and, thus, the security. In addition, background noise was added to the text and audio clip to make it more challenging for software to recognize the CAPTCHA content.

C. Implementation of IRemember CAPTCHA

The proposed CAPTCHA is implemented on the Android platform. The authors chose to implement the prototype on the Android platform because it is open source and allows designers to integrate available functions with future iterations. In addition, the proposed CAPTCHA can be implemented on other platforms.

IRemember has one interface that presents the challenges of using textual and audible approaches. At the bottom of the challenge, a large text box appears where users can type the requested response. As a result, it is easier for users to locate the answer box, and they can immediately type the response upon hearing the challenge. For example, the users hear the prompt, "Type the first letter of each word in the phrase" with a particular pitch of a sound. Then, the application will immediately present the phrase "Cup of Joe" on a different pitch. After users hear the phrase, they should recognize the first letter of each word (COJ) and type those letters into the response box. If the response is correct, users pass it onto the website's main page; if it is wrong, users receive a new CAPTCHA challenge.

Fig. 1 shows an example of an IRemember CAPTCHA as described above. Then, if users enter an incorrect answer, the CAPTCHA says, "Try again," and users receive a new challenge. Users follow the same steps to solve the additional challenge.

The key features of the audio CAPTCHA file used to implement IRemember CAPTCHA are:

- Common phrases.
- Phrases ranging between three to four words.
- Background noise.
- Challenge duration differs based on phrase length.
- Randomly presented phrase.

D. IRemember CAPTCHA Security

The IRemember CAPTCHA method provides another level of layer after the authentication process for users. It has been implemented to build an accessible approach for people with no vision to enable them to solve the given challenge in a short time with less effort.

In the implementation process of the proposed CAPTCHA, we considered the essential security requirements that make the CAPTCHA method more secure and eligible [23, 24]. The IRemember CAPTCHA is implemented based on these requirements, including distortion levels, challenge type, randomness, time constraints, and size. Most blind users find it extremely difficult to solve a CAPTCHA following a loud background noise or distortion. Thus, according to the HearAct, CAPTCHA challenge type and size are the predominant design factors to consider and have been used to make the IRemember solution secure and accessible. Users receive the whole phrase and need to recognize the first letter of each word in the phrase; doing this makes it very difficult and costly for an automated bot to attack. The addition of background noise makes it complicated and requires time from software bots to provide the answer. Randomness is another IRemember method presented randomly to increase the hardness of distracting the correct response. In addition, time is another approach to restrict the software bots from recognizing the answer. After a short time, the method will generate a new challenge. Adding a specified time frame for answering the CAPTCHA challenge prevents the bots from having time to identify the right answer. Another factor that enhances the security level of a CAPTCHA method is the size and length of the CAPTCHA challenge. The IRemember CAPTCHA method provides phrases that have at least three or four words.

The steps that are required to break or automatically solve the IRemember CAPTCHA method would include the following procedures:

1) Understand the main question (e.g., "Type the first letter of each word in the phrase?"

2) Separate or remove the random noise from the audio content.

3) Convert the spoken phrase into textual format.

4) Extract concepts from the transcribed text and recognize the first letter of each word from the given phrase.

5) Type the answer by analyzing the challenge and specify the response.

Theoretically, the software bots need to transcribe the audio into a textual format in the traditional audio CAPTCHA. However, the software bots need more than three steps to break the IRemember CAPTCHA: removing the background noise, understanding the main question, and determining the first letter of each word in the phrase. These steps make it more challenging to break the CAPTCHA method and require more time.

E. Advantages of IRemember

The proposed CAPTCHA system is easy to solve and reduces the time users will spend typing by only asking for the first letter of each word. This CAPTCHA is unlike the traditional audio CAPTCHA, where users need to type at least six digits for each challenge [25]. Additionally, this CAPTCHA approach is efficient because once users hear the audio challenge and identify the CAPTCHA, they can immediately solve the challenge without any need to hear the audio clip in its entirety. When users identify the phrase, they can type the response and pass the challenge seamlessly.

The most important advantage of the proposed CAPTCHA is that it is simple to recall the response phrases because they are commonly used in everyday life. So, memorizing a phrase is far easier for users than a string of random digits.

IV. METHODOLOGY

This section aims to investigate the usability and memorability of the IRemember CAPTCHA method for visually impaired users. Our methodology focuses on mixed methods to perform measurements. The quantitative method was used to measure solving time and success rate and to clearly determine user experience through a workload (NASA TLX) Subjective Questionnaire. On the other hand, the qualitative method will be used to collect strengths, limitations, and suggestions through interview questions. The performance measures were used to analyze the efficiency and effectiveness of the examined CAPTCHAs. This section explains the user study procedures and the obtained results in detail.

A. Study Goals and Hypotheses

The primary goal of the proposed IRemember CAPTCHA method is to improve memorability as well as maximize the success rate. The main goal of the user study is to analyze the usability and memorability of the proposed method. In particular, we compared our proposed method with the traditional audio CAPTCHA method taken from [25]. The audio CAPTCHA method is an alternative to text-based CAPTCHA, where it speaks out the given digits, and the users are required to memorize it and type it using the keyboard on the answering box (see Fig. 2).
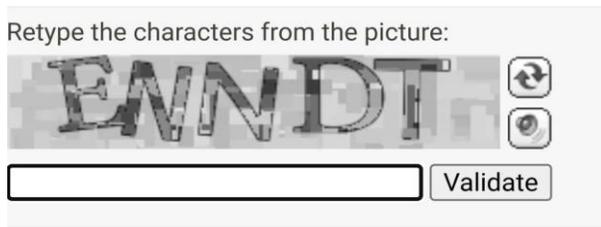
Fig. 2. Common Audio CAPTCHA Method.

The user study hypotheses are:

1) Completing time: The IRemember CAPTCHA can be solved faster than the traditional CAPTCHA due to the easiness of remembering the given challenges and the few numbers of digits that need to bed entered. The IRemember CAPTCHA removes the need to repeat the CAPTCHA challenges several times to memorize the digits limitations of the touchscreen entry method, including locating the location of keys on a touchscreen keyboard.

2) Success rate: The IRemember CAPTCHA will be solved correctly more than the traditional audio CAPTCHA due to the memorability of the given challenges.

3) Workload: The IRemember CAPTCHA will require less workload to complete solving the challenge than the traditional audio CAPTCHA.

B. Participants

To evaluate the proposed CAPTCHA and audio CAPTCHA, eleven participants (six female and five male) were recruited from the Taibh University Accessibility Center and from the Center for Disabled People at Taibh. Participants' ages ranged from 18 to 33 years old and participated in the study voluntarily. All participants were completely blind, have experience using touch screen devices with the support of a screen reader, and were proficient in the English language.

C. Procedure

The study aimed to examine the usability and memorability of two CAPTCHA methods, including IRemember CAPTCHA and the traditional audio CAPTCHA. At the beginning of the experiment, researchers explained the study goals and tasks and described the functionality of both examined CAPTCHAs. They also introduced an overview of the study requirements and study steps and answered participants' queries and provided a clear overview of the study procedures and requirements. After answering participants' questions, they signed a consent form that disclosed that their interaction with the smartphone device while solving CAPTCHA challenges is recorded and their private information is protected. The study contains four main sessions, which are a demographic questionnaire, a training session, a test session, and a Workload Questionnaire.

1) A Demographic questionnaire: To have a clear overview of user experience and performance when using the proposed CAPTCHA method, we collected participants' demographic data, including age, gender, vision level, and touchscreen experience, and English language proficiency.

2) Training session: After collecting participants' demographic data, researchers trained participants on how to solve challenges in both CAPTCHA methods. They were also encouraged to explore both methods and use them for five minutes to ensure they were familiar with the CAPTCHA methods' interfaces. The data from the training session were not included in the study results.

3) Test session: Participants were required to answer 10 CAPTCHA challenges without any assistance or without using other tools to help them record the response to a CAPTCHA question. To avoid order effect, CAPTCHA methods were randomly presented with a 2*2 Latin Square. Challenges for each CAPTCHA method were presented randomly. After listening to a CAPTCHA challenge, participants were requested to speak out the CAPTCHA answer before typing it in order to determine the causes of error when solving a CAPTCHA challenge, whether due to the input method or the difficulty of recalling the given digits. Then, they can type their response. At the end of the test session, participants were required to answer the questionnaires orally.

4) Questionnaires: To measure workload perception [26] for CAPTCHA methods, participants completed the workload questionnaire (NASA TLX) at the end of the study (see Table II). Participants were also asked to identify the strengths and limitations of CAPTCHA methods and recommend any improvements for the memorability aspect. This session took five to ten minutes.

5) Apparatus: We developed a prototype of IRememeber CAPTCHA method on an Android device called Galaxy Nexus phone that has a 4.65-inch-long screen. In the experiment, we asked participants to use the prototype on a portrait mode when solving challenges. In addition, participants were screen recorded when solving CAPTCHA challenges to track the method performance and participants' behavior when interacting with the device. At the end of the experiment, participants were audio recorded to report their answers to the questionnaires.

D. Usability Measurements

Success rate, completing time and the subjective NASA TLX were measured to evaluate the usability and memorability of both CAPTCHA methods.

1) Success rate: The effectiveness (success rate) of the IReamamber CAPTCHA and audio CAPTCHA were evaluated by calculating the challenge completion rate and the number of errors. The total number of challenges completed successfully were divided by the total number of challenges to calculate the CAPTCHA success rate (see Eq. 1).

$$Success\ Rate = \frac{Challenges\ completion\ Successfully}{Total\ number\ of\ Challenges} \qquad (1)$$

2) Completing time: In order to determine the average solving time taken to solve CAPTCHA challenge for each participant, each CAPTCHA challenge was tracked from the start time (the time when Run button was pressed) and and the

completing time (the time when Submit button was pressed). The difference between the completing time and start time determine the solving time, which involves the time spent listening to the challenge and the time spent typing a response.

3) Subjective measure: The Workload NASA TLX questionnaire was used at the end of the study, it contains six factors which are mental demand, physical demand, temporal demand, performance, effort, and frustration (see Table II).

4) Memorability measure: To determine whether participants remember the CAPTCHA content, we asked them to speak out the answer before typing it. In the IRemember CAPTCHA, users cannot repeat listening to the challenge, but in the audio CAPTCHA method, users can repeat several times and then they can speak out their responses.

V. RESULT

The aim of the conducted user study is to evaluate the usability and memorability of the proposed CAPTCHA IRemember. The success rate, speed, user satisfaction, and memorability were computed and discussed in detail in this section.

A. Success Rate

Fig. 3 shows the average success rate of each participant for each CAPTCHA method. The average success rate of the IRemember CAPTCHA method is 81%. Whereas the average success rate for audio CAPTCHA is 48% (see Table I). The results indicate that the success rate for IRemember CAPTCHA is higher and more efficient than the traditional audio CAPTCHA. The results also disclosed that users tend to make more typing errors when using the traditional audio CAPTCHA. Fig. 3 also shows that all participants perform better when solving CAPTCHA challenges using the IRemember CAPTCHA method, which may contribute to the easier recalling of a common phrase than random digits.

One way ANOVA value shows a significant difference between IRemember and the traditional audio CAPTCHA $(F(1, 21) = 78.47; p < 0.000476)$. The t-test value indicates a significant difference between the success rate for IRemember and the traditional audio CAPTCHA $(p= 0.000476)$.
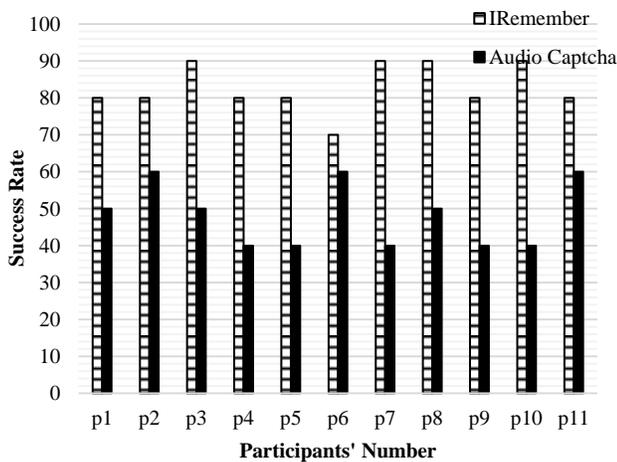
Fig. 3. Average Success Rates for IRemember and Audio CAPTCHA.

TABLE I.   SUMMARY OF THE RESULTS OF BOTH CAPTCHA METHODS FOR VISUALLY IMPAIRED USERS

Methods	Average success rate (standard)	Average solving time (standard)
Audio CAPTCHA (6 digits)	48.18% (8.73)	138.75 s
IRemember CAPTCHA	82.72% (19.20)	54.96 s

B. Completing Time

Fig. 4 shows the average completing time each participant spent on solving IRemember and Audio CAPTCHAs. 54.36 seconds is the average time participants spent Completing IRemember CAPTCHA, and 137.5 seconds for the traditional audio CAPTCHA. The results reveal that all participants spent more time completing the traditional audio CAPTCHA method than IRemember method. That may be due to the need to listen to the CAPTCHA challenge several times to memorize it, unlike the IRemember CAPTCHA, which can recognize the phrase immediately and do not need to listen to it again.

One way ANOVA value shows a significant difference between IRemember and the traditional audio CAPTCHA $(F(1, 21) = 384.13; p < 0.000261)$. The t-test value indicates a significant difference between the success rate for IRemember and the traditional audio CAPTCHA $(p= 0.000261)$.

Solving the traditional audio CAPTCHA method requires more time to solve than the IRemember CAPTCHA. This may be due to playing the audio challenges several times to memorize the random digits in the correct order. Unlike IRemember CAPTCHA, the participant recognizes the answer once they hear the common phrase, which eliminates the repetition needs.
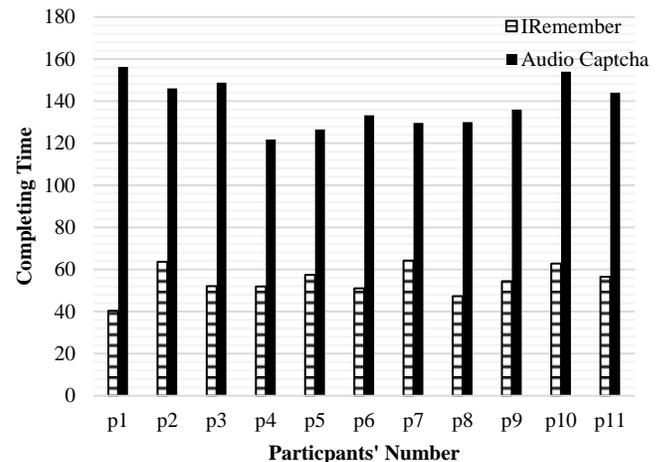
Fig. 4. The Average Time each Participant Spent on Solving IRemember and Audio CAPTCHAs.

C. Memorability Response

The difficulty of remembering the CAPTCHA digits of the traditional audio CAPTCHA was the main cause of error and confusion. Solving CAPTCHA requires memorization of the challenged content. Thus, the primary goal of designing the

IRemember CAPTCHA is to enhance remembering the CAPTCHA responses. So, we asked participants to listen to the CAPTCHA challenges and speak out the response before typing it to figure out the cause of the error, whether it is related to recalling the CAPTCHA response or it is related to the use of the input method.

The result shows that most participants were able to remember the response for the IRememeber CAPTCHA correctly (95%). The reason that caused an error at the stage of speaking the answer out was ignoring the preposition word that is in the provided phrase, for example, the phrase "I waited for a while." Most participants ignored the word "for" and only type "Iwaw". However, the main cause of error for the proposed method was the use of the input method that might lead them to double tap a letter which caused entering the letter twice. As a result, the success rate drops to 84% because the input method is not accessible and error-prone [17].

Regarding the traditional audio CAPTCHA, most participants could not remember the given random digits in the correct order, and sometimes they forgot a digit. Additionally, the other causes of error are that when they type their answer, they also make typing errors using the keyboard as they get confused when they listen to the screen reader and recalling what they should type, they got interfered and confused unlike on the proposed method they can just recall the phrase and determine the letter they should type.

D. Workload (NASA TLX) Questionnaire

The workload questionnaire contains six indicators that are applied to measure mental demand, physical demand, temporal demand, performance, effort, and frustration (see Table II). There is a significant difference based on all six aspects for both CAPTCHA methods. The average score of the mental demand indicator is 33.57 for IRemember CAPTCHA and 87.43 for the audio CAPTCHA method, and it shows the largest significant difference between the IRemember CAPTCHA and the audio CAPTCHA ($F_{(1, 21)} = 262.877$, $p < 0.000159$). Because participants need to memorize the random six digits in the correct order, they must listen to the screen reader when it reads the letter under the user's fingers. The Temporal demand is the second highest significant difference between the IRemember CAPTCHA and the audio CAPTCHA ($F_{(1, 21)} = 302.46$, $p < 0.000708$). The average score of the temporal demand indicator is 33.57 for IRemember CAPTCHA and 84.28 for the audio CAPTCHA method. It means there is high time pressure when solving the audio CAPTCHA. The third highest significant difference between the tested CAPTCHA is the frustration indicator; ANOVA result shows that ($F_{(1, 21)} = 84.309$, $p < 0.000893$). Participants reported a frustration rate of 33.57 for IRemember CAPTCHA and 80.71 for audio CAPTCHA. It means participants feel frustrated when solving the traditional audio CAPTCHA. The fourth highest significant difference between tested CAPTCHA is in the effort indicator ($F_{(1, 21)} = 136.7088$, $p < 0.000645$), where the average score of the effort indicator is 42.85 for IRemember CAPTCHA and 85.71 for the audio CAPTCHA method. It means that participants need to accomplish all tasks at one time, such as listening, memorizing the giving challenge, locating the answer box,

locating letters on the keyboard, and locating submitting button. There was also a statistically significant difference between IRemember CAPTCHA and audio CAPTCHA in terms of performance indicators ($F_{(1, 21)} = 132.923$, $p < 0.000754$). The performance score (80) shows that participants were not satisfied with the performance for audio CAPTCHA. Physical demand was also an essential indicator of workload, and the test showed a significant difference between the Physical demand of IRemember CAPTCHA and audio CAPTCHA ($F_{(1, 21)} = 45.375$, $p < 0.000208$). In general, the t.test values reveal a statistically significant difference between IRemember CAPTCHA and audio CAPTCHA in the six indicators of NASA TLX workload (p=0.000760).

Overall, the interpretation of all six indicators is very high for the traditional audio CAPTCHA, where each indicator receives a very high score, which means users are not satisfied with the audio CAPTCHA. On the other hand, the IRemember CAPTCHA receive a score between 30 and 50 which is interpreted as somewhat high, and users might face some usability issue as well when solving the CAPTCHA.

TABLE II.   AVERAGE NASA TLX SCORES FOR IREMEMBER CAPTCHA AND AUDIO CAPTCHA

TLX aspect	IRemember CAPTCHA	Audio CAPTCHA	T-test
Mental demand	33.5714286	87.4285714	0.00029
Physical demand	52.8571429	84.2857143	0.00170
Temporal demand	33.5714286	84.2857143	0.000377
Performance	45.7142857	80	0.000877
Effort	42.8571429	85.7142857	0.000211
Frustration	33.5714286	80.7142857	0.000596
Average of workload	40.3571429	83.7380952	
Interpretation of score	Somewhat high	Very high	

E. Strengths

The fundamental strength of IRemember CAPTCHA is that it is easy to remember and recall the CAPTCHA challenge response as it provides common phrases that users can easily recall what they just heard, whereas the traditional audio CAPTCHA requires users to remember the six digits as well as the sequence of the digits to pass the CAPTCHA challenge. Another strength of the proposed CAPTCHA is that it reduces the need to locate the repeat button to repeat listening to the CAPTCHA several times. It eliminates the need to repeat the CAPTCHA challenges several times to memorize the digits in the correct order. Users hear the CAPTCHA challenge once and begin typing the response based on memorization. It also removes the need for extra movements like locating the playing button or the response box; users need to interact with the screen to respond to the given challenge. In the proposed CAPTCHA, the cursor is directed at the response box to allow users to type their answers straightforwardly without a requirement to locate the response box.

It also provides a universal design as its interface offers both audio and visual interfaces for sighted people with no

vision. The most important point that makes IRemember CAPTCHA accessible for people with no or low vision is that it overcomes the limitation of memorizing the CAPTCHA content and eliminates the need to locate the position of the play button or the update button. These points improve the efficiency of the proposed CAPTCHA where it enhances the success rate because users can easily remember the answer to CAPTCHA questions in the correct order, as well as reducing the time spent responding to a challenge and minimizing the needed efforts to type the response where users can type three or four digits that are extracted from a given phrase. It also minimizes accidentally pressing an unwanted button, which leads to an error and much extra effort. The IRemember CAPTCHA is also easy to learn, and all participants were able to learn how to accomplish its task from the first trial.

F. Limitations

The main limitation of the proposed CAPTCHA and audio CAPTCHA was that the CAPTCHA challenges and questions were provided in English. Non-native speakers might require more time to memorize the given phrases. Thus, they will spend more time responding to the given challenge. To overcome this limitation, the challenges should be introduced based on the user's native language, preferences, and culture to enhance the user experience. Another drawback associated with the traditional audio CAPTCHA is requiring users to memorize the given challenges in order, which is a difficult task for users with no vision as they cannot see the challenge while typing; they need to depend completely on their memory.

G. Suggestions

At the end of the study, participants suggested multiple recommendations to improve the memorability and usability of IRemember CAPTCHA. They recommended allowing users to customize CAPTCHA challenges based on users' suggestions, personalities, culture, language, and region. Doing so, it can make the CAPTCHA challenge easier to remember. Another critical suggestion was to eliminate the need to type the first character from a preposition word and it is short, and it is the main cause of error in the IRemember CAPTCHA. Finally, they suggested integrating the proposed CAPTCHA into different platforms.

VI. DISCUSSION

The IRemember CAPTCHA method outperformed the traditional audio CAPTCHA method. The success rate for the IRemember CAPTCHA method was higher than the traditional audio CAPTCHA. 82.72% was the success rate for the IRemember CAPTCHA, and it was 48.18% for the traditional method. The solving time was lower for the IRemember CAPTCHA, as well as it required less workload when solving its challenges. 55 seconds users spent to solve the IRemember challenge, and they spent 138.75 seconds solving the traditional audio CAPTCHA. According to prior research, the success rate of the traditional audio CAPTCHA is 46% and the completing time to solve the challenge is 65.64s [27, 28]. These findings show that the IRemember CAPTCHA achieved high performance in terms of success rate and solving time. Overall, the IRemember CAPTCHA achieved high performance in terms of success rate and

solving time. Overall, the IRemember CAPTCHA approach is useable and accessible for visually impaired people, making it suitable for implementation as a real-world CAPTCHA in an application.

According to the NASA TLX scale, the IRemember CAPTCHA method requires fewer mental and physical temporal demands than the traditional audio CAPTCHA. The IRemember CAPTCHA requires users to recall the phrase and type the first letter of each word in the phrase. This task is simple and does not cause cognitive overload and memorization efforts. Thus, all users were able to recall the answer once they heard the given phrase. Unlike traditional audio, CAPTCHA requires users to remember six random digits in the correct order; these steps increase cognitive load as well as physical, temporal, and mental demand. Usually, visually impaired users tend to repeat listing to the given challenge trying to memorize the random digits. Thus, it causes a high level of effort, is demanding and is time-consuming [29, 30]. Therefore, the traditional audio CAPTCHA leads to high frustration once the users are required to repeat until memorize the digits as well as locate the repeat button and then the answer box, which needs more effort and mental demand. Whereas the IRemember CAPTCHA is less frustrating and requires less physical and mental effort than the traditional audio CAPTCHA. It eliminates the causes of confusion and the need to memorize the random CAPTCHA digits by allowing users to hear the common phrase and then recognize the intended character.

Findings show that most existing CAPTCHA requires intensive mental and recalling efforts to remember all digits. Unlike the IRemember CAPTCHA method requires less workload efforts [4-16].

Regarding the security of the proposed method, as the study result shows, IRemember CAPTCHA has the same level of security as the traditional audio CAPTCHA. In addition, IRemember CAPTCHA questions are easy to recall by humans and take only a few seconds to listen to the provided phrase and begin typing the answer [31, 32]. However, software bots find it complicated and time-consuming to solve these questions as the software must be able to determine the first character of each word in the phrase and extract the background noise from the CAPTCHA challenge [33].

VII. CONCLUSION

This paper discusses the implementation and evaluation of a novel audio CAPTCHA method called IRemember CAPTCHA. The IRemember CAPTCHA uses various approaches to build an accessible CAPTCHA for an extensive range of users. The proposed method applies a recognition-based CAPTCHA where the user is asked to listen carefully to the given phrase and determine the first letter of each word from the spoken phrase. Then, users are required to type or record the determined first letters. In addition, users can use swiping gestures to repeat the challenge or submit the response. The result of the evaluation study shows that visually impaired users find the IRemember CAPTCHA easy to learn and solve. However, the most important outcome was that the CAPTCHA was easy to remember for users.

In the experiments, the designed CAPTCHA has demonstrated a higher memorability and recall over the traditional audio method. Through experimental analysis, recording the response for the CAPTCHA method is considered a more effective technique than other input mechanisms like typing or gestures in CAPTCHA design. Moreover, the success rate was high, and the error rate was low. The IRemember CAPTCHA challenges take a different amount of time-based on the length of the phrase and the user's ability to type the response and repeat the challenge. The researchers indicated that the IRemember CAPTCHA is a valuable alternative method as it is a memorable and usable mechanism for people with limited abilities.

In the future, there is a need to improve the CAPTCHA usability by allowing personalization and customization of the presented challenges that are most common and known to the end users and is provided in their own language. Further improvements would make CAPTCHA more accessible by taking into account the impact of culture and language to simplify the way of responding to the given challenges. We also need to integrate the proposed CAPTCHA into a website to ensure its effectiveness and efficiency with the existence of different tasks around the CAPTCHA mechanism.

The study's main limitation is that the sample number is few due to the difficulty of finding a large number of users from this kind of sample demographic of the participants used for the study. We recruited only blind users who have good English language. Thus, there is a need to verify the study results by conducting a longitudinal study with more individuals with no or low vision and sighted users as well to see the effectiveness of the proposed method for all users. Finally, conducting a similar study with users with other types of impairments like mental or mobility impairment.

ACKNOWLEDGMENT

REFERENCES

[1] M. Alnfiai. "Evaluating the accessibility and usability of a universal CAPTCHA based on gestures for smartphones," Universal Access in the Information Society, vol. 20, pp. 817-831, 2020.

[2] S. Roy, S. Nag, I. K. Maitra, and S. K. Bandyopadhyay. International journal of advanced research in computer science and software engineering. International Journal, 3(6), 2013.

[3] A. A. Su¨zen. "Uni-captcha: A novel robust and dynamic user-non-interaction captcha model based on hybrid bilstm+ softmax," Journal of Information Security and Applications, vol. 63, pp. 103-136, 2021.

[4] G. Schryen, G. Wagner, and A. Schlegel. "Development of two novel face- recognition captchas: a security and usability study," Computers & Security, vol. 60, pp. 95–116, 2016.

[5] G. Mori and J. Malik. "Recognizing objects in adversarial clutter: Breaking a visual captcha," In 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 1-1, 2003.

[6] M. Jain, R. Tripathi, I. Bhansali, and P. Kumar. "Automatic generation and evaluation of usable and secure audio recaptcha," In The 21st International ACM SIGACCESS Conference on Computers and Accessibility, pp. 355– 366, 2019.

[7] A. Dattolo and F. L. Luccio. "Accessible and usablewebsites and mobile applications for people with autism spectrum disorders: a comparative study," EAI Endorsed Trans. Ambient Syst., vol. 4, pp. 1-11, 2017.

[8] R. Gafni and I. Nagar. "Captcha: Impact on user experience of users with learning disabilities," Interdisciplinary Journal of e-Skills and Life-long Learning, vol. 12, pp. 207–223, 2016.

[9] A. Jarry, C. Chapdelaine, S. Kurniawan, and W. Wittich. "Blind adults' perspectives on technical problems and solutions when using technology," Journal of blindness innovation and research, vol. 7, pp. 1-10, 2017.

[10] J. Holman, J. Lazar, J. H. Feng, and J. D'Arcy. "Developing usable captchas for blind users," In Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility, pp. 245–246, 2007.

[11] G. Sauer, J. Lazar, H. Hochheiser, and J. Feng. "Towards a universally usable human interaction proof: evaluation of task completion strategies," ACM Transactions on Accessible Computing (TACCESS), vol. 4, pp. 1–32, 2010.

[12] J. Lazar, J. Feng, T. Brooks, G. Melamed, B. Wentz, J. Holman, A. Olalere, and N. Ekedebe. "The soundsright captcha: an improved approach to audio human interaction proofs for blind users," In Proceedings of the SIGCHI conference on human factors in computing systems, vol. A247, pp. 2267–2276, 2012.

[13] M. Shirali-Shahreza and S. Shirali-Shahreza. "Captcha for blind people. In 2007 IEEE International Symposium on Signal Processing and Information Technology," vol. A247, pp. 995–998, 2007.

[14] S. Shirali-Shahreza, H. Abolhassani, H. Sameti, and M. Hassan. "Spoken captcha: A captcha system for blind users," In 2009 ISECS International Colloquium on Computing, Communication, Control, and Management, vol. 1, pp. 221–224, 2009.

[15] V. Fanelle, S. Karimi, A. Shah, B. Subramanian, and S. Das. "Blind and human: Exploring more usable audio CAPTCHA designs," In Sixteenth Symposium on Usable Privacy and Security, pp. 111– 125, 2020.

[16] K. Kuppusamy and G. Aghila. "Human: an accessible, polymorphic and personalized captcha interface with preemption feature tailored for persons with visual impairments," Universal Access in the Information Society, vol. 17, pp. 841–864, 2018.

[17] Yadav, J., Aggarwal, Y., & Goel, V."QR Captcha," International Journal of Engineering Research & Technology. Pp. 398-401, 2020.

[18] M. Alnfiai. "A novel design of audio CAPTCHA for visually impaired users," International Journal of Communication Networks and Information Security, vol. 2, pp. 168–179, 2020.

[19] S. Baheti, H. Shah, R. Sherathia, and P. Waghamode. "CAPTCHA for visually impaired people: A review" vol. 8, pp. 1-6, 2021.

[20] A. Mathai, A. Nirmal, P. Chaudhari, V. Deshmukh, S. Dhamdhere, and P. Joglekar. "Audio captcha for visually impaired," In 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), pp. 1-5, 2021.

[21] S. Kulkarni and H. Fadewar. "Pedometric captcha for mobile internet users," In 2017 2nd IEEE International Conference on Recent Trends in Electron- ics, Information & Communication Technology (RTEICT), pp. 600–604, 2017.

[22] EnglishSpeak. [Online]. "1000 Most Common English Phrases," Retrieved from https://www.englishspeak.com/en/english-phrases, 2022.

[23] Schryen, G., Wagner, G., Schlegel, A. "Development of two novel face-recognition captchas: a security and usability study," Computer Security. vol. 60, pp. 95–116, 2016.

[24] Meriem Guerar, Luca Verderame, Mauro Migliardi, Francesco Palmieri, and Alessio Merlo. "Gotta CAPTCHA 'Em All: A Survey of 20 Years of the Human-or-computer Dilemma," ACM Computer Survey. vol. 54, pp. 1-33, 2021.

[25] A. Mathai, A. Nirmal, P. Chaudhari, V. Deshmukh, S. Dhamdhere and P. Joglekar, "Audio CAPTCHA for Visually Impaired," 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), vol. A247, pp.1-5, 2021.

[26] S. G. Hart and L. E. Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," Advances in psychology, vol. 52, no.5, pp. 139 -183, 1988.

[27] Y. Jyoti, A. Yash and G. Vaishali, "QR Captcha," International Journal of Engineering Research and Technology (IJERT), vol. 09, no. 12. 2020.

[28] S. Graig, L. Jonathan, H. Harry, F. Jinjuan, "Towards a universally usable human interaction proof: Evaluation of task completion strategies," ACM Transactions on Accessible Computing, vol. 32 no. 15, pp.1-15, 2010.

[29] S. G. Hart and L. E. Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," Advances in psychology, vol. 52, no.5, pp. 139 -183, 1988.

[30] J. Lazar, J. Feng and H. Hochheiser, "Research Methods in Human Computer Interaction," John Wiley and Sons, Chichester, UK, 2013.

[31] G. Haichang, W. Wei, Q. Jiao, X. Wang, X. Liu et al. "The robustness of hollow CAPTCHAs'," In Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security, Association for Computing Machinery, New York, NY, USA, pp. 1075-1085, 2013.

[32] Alnfiai, M., Alassery, F. TapCAPTCHA: non-visual CAPTCHA on touchscreens for visually impaired people. J Multimodal User Interfaces 16, 385–398 (2022). https://doi.org/10.1007/s12193-022-00394-2.

[33] G. Sauer, J. Lazar, H. Hochheiser and J. Feng, "Towards A Universally Usable Human Interaction Proof: Evaluation of Task Completion Strategies," ACM Transaction Access Computing, vol. 2, no. 4, pp.1-30, 2010.

Protein Secondary Structure Prediction based on CNN and Machine Learning Algorithms

Romana Rahman Ema[1], Dr. Md. Nasim Adnan[3]
Assistant Professor, Dept. of Computer Science and Engineering
Jashore University of Science and Technology, Jashore, Bangladesh[1, 3]

Mt. Akhi Khatun[2]
(B.Sc Engg.), Dept. of Computer Science and Engineering
Jashore University of Science and Technology
Jashore, Bangladesh

Dr. Syed Md. Galib[5]
Professor, Dept. of Computer Science and Engineering
Jashore University of Science and Technology
Jashore, Bangladesh

Sk. Shalauddin Kabir[4]
Lecturer, Dept. of Computer Science and Engineering
Jashore University of Science and Technology
Jashore, Bangladesh

Dr. Md. Alam Hossain[6]*
Associate Professor, Dept. of Computer Science and
Engineering, Jashore University of Science and Technology
Jashore, Bangladesh

Abstract—One of the most important topics in computational biology is protein secondary structure prediction. Primary, secondary, tertiary, and quaternary structure are the four levels of complexity that can be used to characterize the entire structure of a protein that are totally ordered by the amino acid sequences. The polypeptide backbone of a protein's local configuration is referred to as a secondary structure. In this paper, three prediction algorithms have been proposed which will predict the protein secondary structure based on machine learning. These prediction methods have been improved by the model structure of convolutional neural networks (CNN). The Rectified Linear Units (ReLU) has been used as the activation function. The 2D CNN has been trained with machine learning algorithms, including Support Vector Machine, Naive Bays and Random Forest. The SVM is used to correctly classify the unseen data. Naïve Bays (NB) and Random Forest (RF) are also applied to solve the prediction problems for not only classification problems but also regression problems. The 2D CNN, hybrid of 2D CNN -SVM, CNN-RF and CNN-NB have been proposed in this experiment. These different methods are implemented with the RS126, 25PDB and CB513 dataset. Further, all prediction Q3 accuracy is compared and improved with their datasets.

Keywords—Protein Secondary Structure Prediction (PSSP); Support Vector Machine (SVM); Naive Bays (NB); Random Forest (RF); Convolutional Neural network (CNN)

I. INTRODUCTION

Proteins are the building blocks of amino acid sequences [1] [2]. Generally, there are four types of protein structure's which are primary, secondary, tertiary, and quadratic structure. Secondary structure are the building blocks of the macromolecule structure [1]. Secondary structures can be divided into two categories- the regular and the irregular secondary structures. The regular secondary structure has two types, including α -helices (H) and β-sheet (E) and the irregular secondary structure has more types, including tight

turns, Random coils, Bulges, etc. By using only their basic structure, the PSSP method is a series of bioinformatics technique aimed at predicting the secondary structure of protein sequences or residues[2] [3]. In molecular biology, the most essential and crucial problems are the prediction of the protein secondary structures using machine learning approaches [3]. This proposed work aims to predict the protein secondary structures and come up with a highly accurate solution that would be easily solved by computational biology. The purpose of PSSP is also to categorize the pattern of residues in amino acid sequences as a-helix, B-strand, or coil. To discover the secondary structure of proteins, the researchers must examine hydrogen bonding patterns and geometric limitations, as well as employ the DSSP tool [4]. The prediction of PSSP is done using a variety of machine learning and deep learning techniques[5] [6].

Researchers used a variety of strategies to predict the protein secondary structures in the early years [5][7][8]. Furthermore, compared to the prior years, the prediction accuracy has been improved in this paper. The understanding of protein folding mechanisms are frequently regarded as a crucial objective that will help structural biologists unravel the puzzling connection among the protein sequence, structure, and function. The scientific work will be aided by the ability to estimate protein folding speeds without the requirement for actual experimental study. In this study, the secondary structure prediction is merely enhanced, as it would be challenging to predict the tertiary structure of proteins in the absence of homology.

Convolutional neural network is a form of artificial neural network that is generally used in deep learning for image processing [1][2]. The hidden units of the CNN are frequently the same dimensions or size as the processed data [1]. The data is convolved by the hidden units, which then stored the information from the data. Each hidden unit's information will

*Corresponding Author.

be recorded as a feature map [1] [7]. The number of feature maps that produced equal to the number of hidden units are employed. After that, the pooling stage is performed on the existing feature maps, collecting dense information from them [1][9].

Support vector machine is a machine learning algorithm based on a supervised technique that has been used for different types of classification problems [1]. If a model of SVM is given that the sets of labeled schooling data for each section, it will be enabled to classify the new data. It has been used to analyze the data for not only classification problems but also regression problems. It has also been used to categorize the unlabeled data. The main aim of this is to search hyperplane that can be utilized as a decision surface to close the gap between two classes [1] [10].

Naïve bays is also a supervised machine learning model based on Bays Theorem [3]. Typically, this is a classification technique that predicted with an estimation of independence. Naïve bays model is easy to create and especially applicable to solve the classification problems and huge complex data sets. For that, it can make as a fast prediction [11].

Popular machine learning algorithm, random forest belongs to the supervised approach. This type of algorithm is also applied for both classification as well as regression problems. In RF, each individual tree has been spread as a class prediction and the most voted class can be estimated for prediction and it predicts the final output. It improves the prediction accuracy [12].

In this paper, a prediction method has been represented for the secondary structure of proteins based on CNN and machine learning techniques. In addition, this paper is exploded into six sections. The Section II discusses about the related works. The Section III represents the preliminaries for the proposed models. The Section IV introduces and describes the proposed techniques smoothly. The Section V shows the simulation results. Finally, conclusion is presented in Section VI.

II. RELATED WORK

Vincent Michael Sutanto et al. [1] introduced a hybrid 1-D CNN and SVM for the prediction of protein secondary structures. They fine-tuned CNN and then changed it, therefore, instead of giving an orthogonal label as output, model feature map production. By doing this step, they projected the data into higher levels. They aimed to absorb high dimensional space of SVM as classifier. The modified CNN model created the 3-D array properties as an output and therefore, a conversion must be applied. The modified CNN models used feature maps for the training of SVM models.

Shangxin Xie et al. [10] introduced a new algorithm formed on the increased fuzzy support vector machine (FSVM) for the prediction of secondary structure of proteins. They applied the different classification rules for the prediction. They used this model to improve the fuzzy membership value and prediction accuracy. They implemented this model to derive the approximate optimal division hyperplane in the feature space.

Masood Zamani et al. [13] proposed the evolutionary-based computation method of protein secondary structure classifiers. They evaluated the performance prediction by using the amino acid sequence with the help of the clustering technique. K-means clustering technique is applied to reduce the dimension of the classifier's inputs on sequence component. They also presented PSS classifier stand on genetic programming technique. They evaluated this approach to improve the performance prediction.

Pooja Jain et al. [14] explored the structural classification of protein with the help of supervised learning technique. Firstly, they chose domains for learning the structure classification. This technique considered two types of domains, known and unknown structural classification. They assigned known domains into unknown domains for classification of protein.

Ying Xu et al. [15] proposed a prediction method of secondary structure based on convolutional neural network (CNN) and random forest. After each convolutional layer, Rectified Linear Units (ReLU) activation layer was used to solve the gradient disappearance problems. They used deep CNN to extract the protein features from amino acid sequence. Here, the fully connected layer and SoftMax layer were used to predict the three types of secondary structure (C, H, E).

III. PRELIMINARIES

A. Protein

The three-dimensional arrangement of atoms in an amino acid-chain molecule is known as protein structure [16] [17]. Proteins are polymers – especially polypeptides – made up of amino acid sequences, which are the polymer of monomers [16]. Proteins are involved in a wide range of biological processes from accelerating chemical reactions to construct the architecture of all living things [18]. Despite their diverse activities, all proteins are built up of the same twenty amino acid building blocks. The folding of the protein into its unique final form and function is determined by how these twenty amino acids are organized [8][19][20].
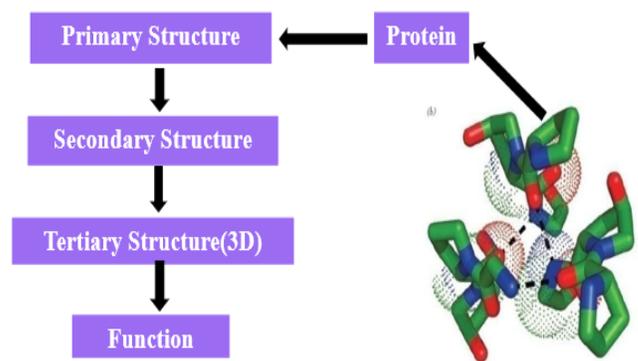
Fig. 1. Protein Structure.

Fig. 1 shows the structure of protein and its function.

B. Primary Structure

The sequence of amino acids in a polypeptide chain is known as primary structure [18]. The primary structure of a

protein is described the starting from the amino-terminal (N) end to the carboxyl-terminal (C) end. Ribosomes are the most frequent organelles in cells that perform protein production [21]. Peptides can also be manufactured in the lab. Primary structures of proteins can be sequenced directly or deduced from DNA sequences [18] [20].

C. Secondary Structure

The pattern of hydrogen bonds between the amino hydrogen and carboxyl oxygen atoms in the peptide backbone is technically characterized as a secondary structure [17] [18]. The hydrophobic side chains are those facing inward. A third of every amino acid has hydrophobic properties [17]. There are hydrogen bonds between the two distinct zones (carbonyl & amino groups). Beta sheets are made up of 5–10 amino acids in each region.

Secondary structure, the next level of protein structure, refers to the local folded structures formed inside a polypeptide as a result of interactions between backbone atoms [22]. (It does not involve R group atoms; the backbone only refers to the polypeptide chain away from the R groups.) Helixes and pleated sheets are the two most common types of secondary structures [18][21][23].

D. Tertiary Structure

The term "tertiary structure" refers to the overall three-dimensional structure of a polypeptide. There are various types of polypeptide chains in it. It interacts with R-groups. Polar R-groups are capable of forming hydrogen bonds. Dipole-dipole interactions and hydrophobic interactions play key roles in the three-dimensional structure [8].

E. Convolutional Neural Network (CNN)

The most popular kind of neural network is the convolutional neural network which is used to solve the image processing problems [1]. CNN is divided into two sections: the hidden layers, also known as the feature extraction section, and the classification section. A series of convolutions and pooling operations are carried out by the hidden layers [7]. To create a map, convolution is applied to the input data using a filter or kernel [7][21].The classification parts assign a probability for the object on the image being what the algorithm is predicted. Multilayer perceptron is regularized variants of CNNs. Normally, these networks are fully connected, which means that every neuron in one layer is connected to every neuron in the layer below. It is susceptible to data overfitting because of the network's "full connectivity." By utilizing the hierarchical structure in the data and assembling patterns of increasing complexity using smaller and simpler patterns imprinted in their filters, CNNs develop a new method of regularization [20]. The bottom end of the connectivity and complexity spectrum is where CNNs fall due to this.

Fig. 2 shows the architecture of CNN, including with three input layers, hidden layer section 1 and section 2 (both containing 4 layers) and finally output layer.

F. Convolutional Layer

It is the principal components of CNN. It has a collection of filters or kernels, which are parameters that must be learned

during training. The convolution 2D layer was employed in this investigation. The size of the filters is typically smaller than the size of the image. An activation map is created when each filter is applied to an image. This has the advantage of reducing parameter usage and allowing the convolution kernel to extract features more effectively. The preprocessed protein data with a size of (13, 20) is used as the input data. To obtain the output, the convolution kernel of 3×3 is twisted in steps of 1 from left to right, from top to bottom, from the upper left corner of the data with the input size of (13, 20). If padding is not used, the output size is $(13 - 3 + 1)/ (20 - 3 + 1)$ [7]. The number of parameters is minimized by the CNN feature and considerably the training speed has been enhanced through it. The data is convolved in the same way by 128 convolution kernels, and each convolution kernel extracts features automatically. Different convolution kernels extract picture edge information, shading information, contours, and other image features automatically in the image domain. The model can extract 128 features automatically in theory. Supposing that the convolution kernel's width is ck and the height is dk. The 2D convolutional equation is:

$$m_{i,j} = \begin{bmatrix} n_{i,j} & n_{i+1,j} & \cdots & n_{i+ck,j} \\ n_{i,j+1} & n_{i+1,j+1} & \cdots & n_{i+ck,j+1} \\ n_{i,j+ck} & n_{i+1,j+dk} & \cdots & n_{i+ck,j+dk} \end{bmatrix} \text{[7][15]} \qquad (1)$$

The ReLU is used as the activation function within the model. The ReLU function has the following expression:

$$F(y) = \max (0, y) \text{ [7]} \qquad (2)$$

G. SoftMax Layer

It is mostly used in artificial neural network as the activation function in the output layer. This is used as classification problems where more than two class labels require class membership [9]. The SoftMax layer has been also used in proposed model as the activation function [7].

H. MaxPooling Layer

It is a procedure for retrieving features from the convolutional layer that reduces their dimensionality [18]. It offers the advantage of reducing the size of the featured image while maintaining the number of feature maps, allowing for data processing after the convolutional layer is output. Simultaneously, significant feature information is stored, reducing the model's computation complexity and increasing calculation speed. Median and average pooling are two other pooling algorithms. The kernel size of MaxPooling layer is 2×2 [9].
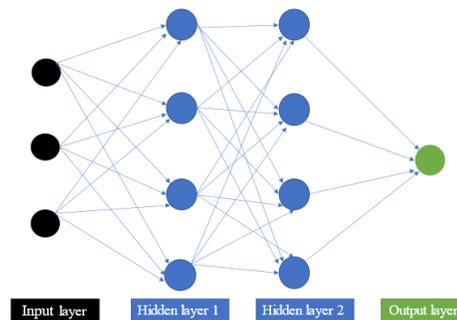
Fig. 2. Convolutional Neural Network.

I. Flatten Layer

It is frequently used to convert multidimensional information into one-dimensional input during the transition from the convolution layer to the fully linked layer. Flattening is the process of converting data into a one-dimensional array for use in the layer below. We flatten the convolutional layer output to create a single, extensive feature vector. The final classification model, sometimes referred to as a fully-connected layer, is connected to it [3].

TABLE I.   POOLED FEATURE MAP AND FLATTENING STEPS

1	1	0	Flattening	1
4	2	1		1
0	2	1		0
Pooled Feature Map				4
				2
				1
				0
				2
				1
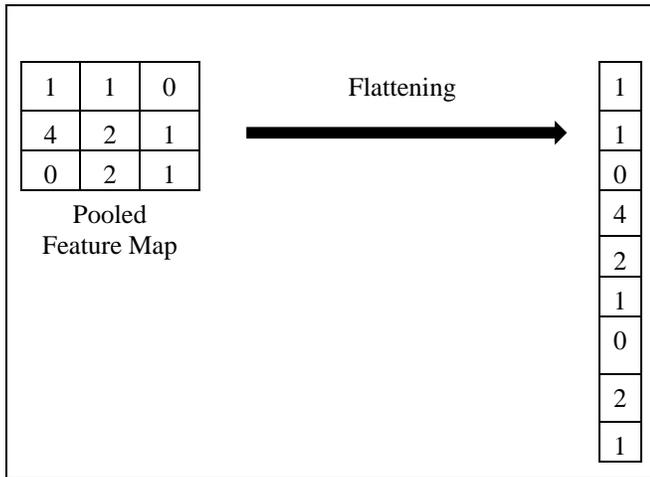
Table I shows that the pooled feature map and flattening steps. The flattening process is used to make a single feature vector and get the next layer.

J. Support Vector Machine (SVM)

It is a type of supervised learning model that analyzes and linearizes data for classification as well as regression in machine learning [18]. SVM algorithm classifies data that is linearly divisible. If it is not linearly divisible, we have to follow the kernel strategy to build decision [24]. This decision program is called support vector. Support Vector uses a subset of training points that makes it efficient in memory [25]. Support vector machine works in following steps:

- Firstly, we have to import the dataset.
- Need to analyze the data.
- Then we have to preprocess the data.
- Split up the dataset.
- Sort out the dataset into training and testing sections.
- Train and test the support vector machine algorithm.
- Build some predictions.
- Finally, compute the accuracy.

Fig. 3 shows the architecture of SVM. In this figure, a hyperplane with maximized margin is created which refers to the distance between the data points.
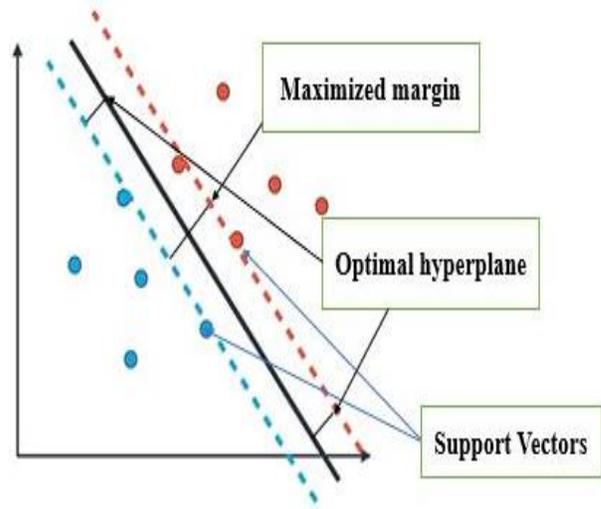
Fig. 3.   The Architecture of Support Vector Machine.

K. Naïve Bays (NB)

Naïve bays is a general machine learning strategy that is supervised using Bayes Theorem [3]. It is a statistical classification technique. The basic innocent way is to assume that each feature makes a distinct and equal one for which it is called a "naïve". This is a strong idea for genuine and unrealistic data. This type of conjecture is indicated as class conditional independence. Naïve Bays algorithm works in following steps such as:

Fig. 4 shows the naïve bays classifier. It is one kind of linear classifier. It's named "Naïve" because it assumes that the feature of datasets is mutually independent.

Naïve Bays algorithm works in following steps such as:

- Divide the class.
- Sum up the dataset.
- Again, sum up data according to class.
- Then calculate density function according to Gaussian Probability.
- Finally compute class probabilities.
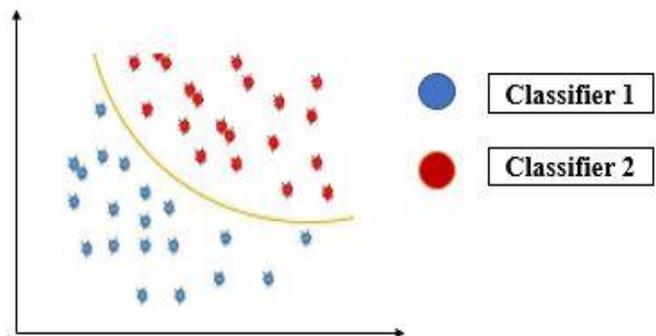
Fig. 4.   The Architecture of Naïve Bays Classifier.

L. Random Forest (RF)

Generally, random forest is a simple, supervised learning technique. It is used for not only classification troubles but also regression troubles like support vector algorithm. It can be learned how to classify data randomly. It extracts several samples from the original sample using the bootstrap resampling approach that works by training a large number of decision trees [23]. The RF output is the class chosen by the majority of trees for the classification task [18].

The random forest algorithm can be explained in following steps:

- Import the dataset.

- Choose the random samples.

- Compute the vote for the predictive result. It can be used "Mode" for "Classification" troubles, and also can be used "Mean" for "Regression" troubles.

- Ultimately, choose the peak voted value for predictive result. And it is the most precious final prediction.
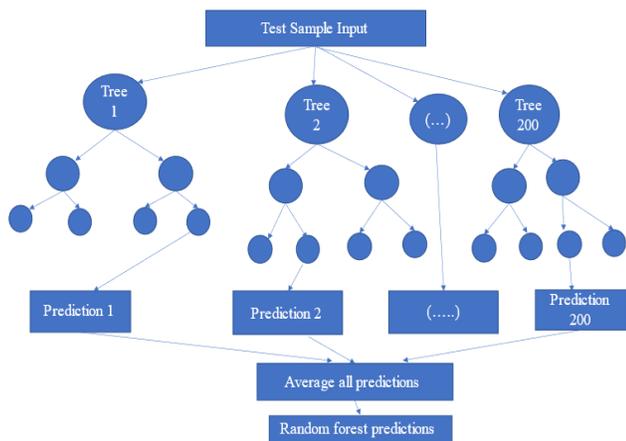
Fig. 5. The Architecture of Random Forest Classifiers.

Fig. 5 shows the random forest classifier. In this figure, the test sample input is tree root. Random forests generate decision trees from randomly chosen data samples, obtain predictions from each tree, then vote on the best option.

IV. PROPOSED METHOD

In this research, the PSSP method has been proposed based on CNN and machine learning algorithms. 2-D CNN and machine learning algorithms i.e., RF, SVM and NB have been used in this paper. Here, the Rectified Linear Unit (ReLU) has been used as an activation function in 2D-CNN. It has been used after each convolutional layer to solve the gradient disappearance problems. This paper used deep CNN to extract the protein features from the amino acid sequence. The fully connected layer, MaxPooling, flatten and SoftMax layer have been used to predict the three types of secondary structure (C, H, E). Further, the pooling function have been used in these layers. The 2D CNN has been combined and trained with SVM, RF, NB. Secondly, the datasets have been trained and tested with SVM, RF and NB. The 2D CNN, the hybrid of 2D

CNN-SVM, CNN-RF and CNN-NB have been implemented. Finally, the prediction accuracy has been calculated.

Fig. 6 shows the flow diagram of protein secondary structure prediction. Firstly, the input (amino acid sequence) has been processed. Secondly, the Con2D layer has been used. Here, the filter has been changed with the process. The softmax layer has been used as an activation function. Also, the maxpooling and other layers have been used in this process. Further, the 2D-CNN has been combined and trained algorithm such as SVM, RF and NB algorithm. Finally, the secondary structure sequence such as Helix (H), Strand (E) and Coil (C) have been predicted from the amino acid sequence.

Secondary structure prediction datasets.

1) RS126 dataset.
2) CB513 dataset.
3) 25PDB dataset.

The first and best-known dataset for the PSSP is the RS126 dataset. Electric Sander as well as Rost came up with the concept [26]. It is one of the best non-homologous datasets for predicting the structure of protein [26]. This dataset is applied for the prediction of protein secondary structure. Its maximum carrying capacity is 23,347 residues, and its typical protein sequence length is 185. RS126 is composed of 47% random coil, 21% -sheet, and 32% -helices (H).[25] Furthermore, the CB513 and 25PDB datasets have also been used. The 2-D structure is predicted using the CB513 dataset. This is a crucial dataset that is excellent for enhancing algorithms and forecasting secondary structures. The CB513 dataset was created by Cuff and Barton and contained 513 sequences and 84,107 residues [24]. This kind of dataset is used in this study to enhance the performance of protein secondary structure prediction. One of the biggest datasets is this. Furthermore, it is utilized to classify data and fold the protein structure. Additionally, the 25PDB dataset was included. The PDB dataset's accuracy is better than some other types of datasets in each of the datasets. All protein datasets have been collected from Research Collaboratory for Structural Bioinformation (RCSB).
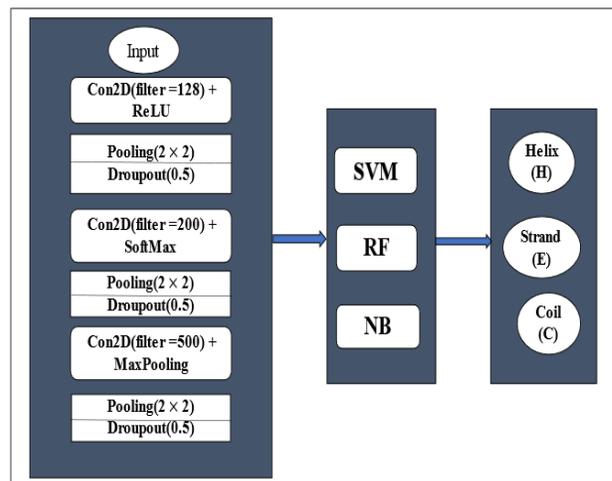
Fig. 6. The Flowchart of Proposed Model of Protein Secondary Structure Prediction.

TABLE II.   PROTEIN SECONDARY STRUCTURES OF 8 AND 3 CLASSES.[22]

8class symbols	8 class names	3class symbols	3class names
H	α-helix	H	Helix
L	Loop/Irregular	C	Coil/Loop
T	B-Turn	C	Coil/Loop
S	Bend	C	Coil/Loop
G	3_{10}-helix	H	Helix
B	B-bridge	E	Sheet
I	π-helix	H	Helix

Table II shows the protein secondary structure of eight and three classes. According to DSSP (Define Secondary Structure of Proteins), eight class secondary structures are converted to three class of secondary structures of proteins.[19].

Pseudocode 1: CNN [1]

Input: Amino acid sequence

Step 1: Sliding window process

Step 2: ESF ⟶ Extract Shadow Feature

Step 3: Normalize by the ESF

Step 4: Feature, data size 13*20

Step 5: Repeat

Step 6: Forward Propagation

CD ⟶ Convolution2D (ESF)

MP ⟶ MaxPooling 2D (CD)

FC ⟶ Fully Connected (MP)

Class label ⟶ SoftMax (FC)

Class label ⟶ Dense(D)

Class label ⟶ Flatten (F)

Step 7: Backward Propagation

Conduct backward propagation with Adam;

Step 8: Use the trained algorithm with SVM, RF and NB.

Pseudocode 2: SVM

Input: M and n filled with schooling labeled data, $\beta \Leftarrow 0$ or $\beta \Leftarrow$ partially trained support vector machine.

Step 1: P ⟸ some value (20 for example)

Step 2: repeat

Step 3: for all {m_i, n_i}, {m_j, n_j} do

Step 4: optimize β_i and β_j.

Step 5: End for

Step 6: until no changes in β or other element constraint criteria met.

Ensure: Just contain the support vectors ($\beta_i > 0$).

Output: Accuracy prediction.

Pseudocode 3: RF [12] [15]

Step 1: Randomly select the "K" attribute from the total "P" attributes. K < P

Step 2: Calculate the node "r" in the "K" properties by applying the largest partition points.

Step 3: Need to use the greatest split node for splitting the node into daughter mode.

Step 4: Steps are repeated 1 through 3 until the "F" number of node is reached.

Step 5: The forest is created by repeating steps 1 to 4 for "N" number times for to build "N" number of trees.

Pseudocode 4: NB [3]

Input: Training dataset D,

P= ($p_1, p_2, p_3 \ldots p_n$) // value of the predictor variable in testing dataset.

Output: A category of testing dataset.

Step 1: Firstly, read the training dataset D.

Step 2: Secondly, compute the mean deviation and standard deviation of the predictor variables in every group;

Step 3: Repeat. Compute the probability of p_i applying the gauss density equation in every group;

Until the probability of predictor variables ($p_1, p_2, p_3 \ldots \ldots, p_n$) has been computed.

Step 4: Compute the likelihood for every class or group;

Step 6: Finally, get the best likelihood.

V. RESULT ANALYSIS AND DISCUSSION

The Q3 approach was employed in this paper to assess the algorithm's effectiveness. Q3 is used to represent the number of residues and is computed by dividing the number of accurately predicted protein residues by the total number of residues in a known protein's secondary structure sequence.

$$Q_3 = \frac{Q_C + Q_H + Q_E}{Q} \qquad (3)$$

Where Q_C, Q_H and Q_E are the number of accurately predicted protein structural class of C, H and E. The total number of amino acids are denoted by Q [7].

$$Q_j = \frac{Q_k}{Q} \qquad (4)$$

Where,

Q_k represents the total number of amino acid residues. K ∈ {C, H, E}. [7].

Q_C, Q_H and Q_E are used to evaluate the experimental result in this research. The Eq. 4 can be used to calculate the values of Q_C, Q_H and Q_E.

Test sequence:

Original sequence:
GGGARSGDDVVAKYCNACHGTGLLNAPKVGDSAAWKTRADAKGGLDGLLAQSLSGLNAMPPKGTCADCSDDELKAAIGKMSGL

Predicted structure:
CCCCCCCHHHHHCCHHHHCCCCCCCCCHHHHHHHHHHHHHHCCCCCCHHHHHCECCECCCCCCCCCCCHHHHHHHHHCCCC

Actual structure:
CCCCCCCHHHHHCCHHHHCCCCCCCCCCCCCHHHHHHHHHHHHCCCCCCHHHHHCECCECCCCCCCCCCCHHHHHHHHHHHHCC

===

Fig. 7.   Prediction from Original Sequence.

Fig. 7 shows the actual structure predicted from the predicted structure and predicted structure predicted from original primary sequence. [Test sequence collected from RS126 dataset].

Table III, IV, V shows the Q_C , Q_H , Q_E and Q_3 accuracy based on RS126, CB513 and PDB25 datasets.

Where Q_C , Q_H and Q_E are used to accurately predict the protein structural class of C, H and E. In these tables, CNN, CNN-SVM, CNN-RF and CNN-Bays methods have been used for the prediction of protein secondary structure. In Q_3 all prediction accuracy, the CNN-SVM method has been achieved the highest prediction accuracy. These achievements are 82.34%, 84.32%, 83.76% based on RS126, CB513 and 25PDB datasets.

Table VI and VII shows the comparison between the previous model and the proposed model. Here, 25PDB and CB513 datasets have been used for the protein secondary structure prediction. It is seen that the previous work of datasets based on the model of CNN-SVM, CNN-RF and CNN-Bays are not satisfactory. The proposed model has improved the performance of protein secondary structure prediction in Q3 accuracy. The CNN-RF, CNN-Bays, CNN and CNN-SVM have been achieved the highest Q3 accuracy of 81.73%, 79.35%, 80.57% and 84.32 based on 25PDB and CB513 datasets which are 2.34%, 2.45%, 1.48%, 2.83% higher than those datasets. It is seen that the proposed model has been achieved the highest prediction Q_3 accuracy than the previous model.

TABLE III.   THE ACCURACY OF RS126 DATASET

Method	Accuracy (%)			
	Q_C	Q_H	Q_E	Q_3
CNN	79.24	76.35	77.43	80.09
CNN-SVM	81.27	78.32	80.35	82.34
CNN-RF	79.76	75.03	78.45	80.24
CNN-Bays	78.20	73.23	75.76	79.27

TABLE IV.   THE ACCURACY OF CB513 DATASET

Method	Accuracy (%)			
	Q_C	Q_H	Q_E	Q_3
CNN	79.54	73.65	77.83	80.35
CNN-SVM	83.08	78.67	81.07	84.32
CNN-RF	78.80	74.09	76.53	79.24
CNN-NB	78.36	74.26	77.34	79.80

TABLE V.   TABLE V. THE ACCURACY OF 25PDB DATASET

Method	Accuracy (%)			
	Q_C	Q_H	Q_E	Q_3
CNN	80.32	74.30	76.04	80.57
CNN-SVM	82.37	75.31	81.09	83.76
CNN-RF	79.05	72.53	74.23	81.73
CNN-NB	77.45	70.63	73.83	79.35

TABLE VI.   THE PERFORMANCE COMPARISON RESULTS BETWEEN THE PREVIOUS AND PROPOSED MODEL

Dataset	Method	Q_3 Accuracy (%)	
		Previous prediction	Proposed prediction
CB513	CNN-SVM	81.49 [1]	84.32
	CNN-RF	-	79.24
	CNN-NB	-	79.80
	CNN	-	80.35

TABLE VII.   THE PERFORMANCE COMPARISON RESULTS BETWEEN THE PREVIOUS AND PROPOSED MODEL

Dataset	Method	Q_3 Accuracy (%)	
		Previous prediction	Proposed prediction
25PDB	CNN	79.09 [15]	80.57
	CNN-RF	79.39 [15]	81.73
	CNN-Bays	76.90 [3]	79.35
	CNN-SVM	-	83.76

VI.   CONCLUSION AND FUTURE WORK

The issue of protein structure prediction must be resolved in the realm of bioinformatics. The protein secondary structure has been utilized in this paper. To understand the protein sequence using CNN and machine learning algorithms was the initial step for solving the PSSP problem. In this work, three identical and separate datasets—25PDB, RS126, and CB513 are used in this study. These datasets are sufficient for solving the problems of prediction. The ReLU layer is used as the activation function (CNN). Further, MaxPooling, SoftMax, dense and flatten layer have been used. The CNN has been integrated with RF, SVM, and NB in this study. The proteins have been categorized using these techniques. These hybrid techniques have successfully overcome the gradient disappearance problems by retaining the significance of

original features data to the maximum extent possible. It is demonstrated that the proposed model has been enabled to capture the long-range interdependencies between the sequence residues. The proposed models reached in high accuracy. The CNN-SVM model has achieved the highest Q3 accuracy of 82.35%, 84.32% and 83.76% on the RS126, CB513 and 25PDB datasets than the previous work. Besides, CNN, CNN-RF and CNN-NB has achieved the highest prediction accuracy. The secondary structure of protein takes part in an important role in protein function and folding. It identifies the similar function where protein sequence varies (only ~50% remote homologies may be identified based on sequence). The proposed method can experimentally perform better than other previous work and this work could be easily understandable by researchers for solving the problem of protein secondary structure prediction. Further, it can help to explain the disease (the effect of mutations and design drugs). As a future scope of research, it would be possible to propose the three-dimensional protein structure prediction using deep learning algorithms.

REFERENCES

[1] V. M. Sutanto, Z. I. Sukma, and A. Afiahayati, "Predicting Secondary Structure of Protein Using Hybrid of Convolutional Neural Network and Support Vector Machine," Int. J. Intell. Eng. Syst., vol. 14, no. 1,pp.232–243,2020, doi: 10.22266/IJIES2021.0228.23.

[2] R. R. Ema, A. Khatun, M. A. Hossain, M. R. Akhond, N. Hossain, and M. Y. Arafat, "Protein Secondary Structure Prediction using Hybrid Recurrent Neural Networks," J. Comput. Sci., vol. 18, no. 7, pp. 599–611, 2022, doi: 10.3844/jcssp.2022.599.611.

[3] Y. Liu, Y. Chen, and J. Cheng, "Feature extraction of protein secondary structure using 2D convolutional neural network," Proc. - 2016 9th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISP-BMEI 2016, pp. 1771–1775, 2017, doi: 10.1109/CISP-BMEI.2016.7853004.

[4] G. B. De Oliveira and H. Pedrini, "Ensemble of Template-Free and Template-Based Classifiers for Protein Secondary Structure Prediction," 2021.

[5] W. Wardah, M. G. M. Khan, A. Sharma, and M. A. Rashid, "Protein secondary structure prediction using neural networks and deep learning: A review," Comput. Biol. Chem., vol. 81, no. December 2018, pp. 1–8, 2019, doi: 10.1016/j.compbiolchem.2019.107093.

[6] W. Yang, K. Wang, and W. Zuo, "A fast and efficient nearest neighbor method for protein secondary structure prediction," 2011 3rd Int. Conf. Adv. Comput. Control. ICACC 2011, no. Icacc, pp. 224–227, 2011, doi: 10.1109/ICACC.2011.6016402.

[7] Y. Zhao, H. Zhang, and Y. Liu, "Protein secondary structure prediction based on generative confrontation and convolutional neural network," IEEE Access, vol. 8,pp.199171–199178,2020,doi: 10.1109/ACCESS.2020.3035208.

[8] P. Kumar, S. Bankapur, and N. Patil, "An enhanced protein secondary structure prediction using deep learning framework on hybrid profile based features," Appl. Soft Comput. J., vol. 86, p. 105926, 2020, doi: 10.1016/j.asoc.2019.105926.

[9] J. Cheng, Y. Liu, and Y. Ma, "Protein secondary structure prediction based on integration of CNN and LSTM model," J. Vis. Commun. Image Represent., vol. 71, p. 102844, 2020, doi: 10.1016/j.jvcir.2020.102844.

[10] S. Xie, Z. Li, and H. Hu, "Protein secondary structure prediction based on the fuzzy support vector machine with the hyperplane optimization," Gene, vol. 642, no. September 2017, pp. 74–83, 2018, doi: 10.1016/j.gene.2017.11.005.

[11] M. O. F. Science, "Prediction of Protein Secondary Structure using Binary Classification Trees , Naive Bayes Classifiers and the Logistic Regression Classifier," no. January, 2015.

[12] C. Kathuria, D. Mehrotra, and N. K. Misra, "Predicting the protein structure using random forest approach," Procedia Comput. Sci., vol. 132, no. Iccids, pp. 1654–1662, 2018, doi: 10.1016/j.procs.2018.05.134.

[13] M. Zamani and S. C. Kremer, "Protein secondary structure prediction using an evolutionary computation method and clustering," 2015 IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol. CIBCB 2015, 2015, doi: 10.1109/CIBCB.2015.7300327.

[14] P. Jain, J. M. Garibaldi, and J. D. Hirst, "Supervised machine learning algorithms for protein structure classification," Comput. Biol. Chem., vol. 33, no. 3, pp. 216–223,2009,doi: 10.1016/j.compbiolchem.2009.04.004.

[15] Y. Xu and J. Cheng, Protein secondary structure prediction using cnn and random forest, vol. 1254 CCIS. 2020. doi: 10.1007/978-981-15-8101-4_25.

[16] M. H. Zangooei and S. Jalili, "Protein secondary structure prediction using DWKF based on SVR-NSGAII," Neurocomputing, vol. 94, pp. 87–101, 2012, doi: 10.1016/j.neucom.2012.04.015.

[17] W. Pirovano and J. Heringa, "Protein secondary structure prediction.," Methods Mol. Biol., vol. 609, pp. 327–348, 2010, doi: 10.1007/978-1-60327-241-4_19.

[18] M. Patel and H. Shah, "Protein secondary structure prediction using support vector machines (SVMs)," Proc. - 2013 Int. Conf. Mach. Intell. Res. Adv. ICMIRA 2013,pp.594–598,2014,doi: 10.1109/ICMIRA.2013.124.

[19] P. Kountouris et al., "A comparative study on filtering protein secondary structure prediction," IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 9, no. 3, pp. 731–739, 2012, doi: 10.1109/TCBB.2012.22.

[20] S. Long and P. Tian, "Protein secondary structure prediction with context convolutional neural network," RSC Adv., vol. 9, no. 66, pp. 38391–38396, 2019, doi: 10.1039/c9ra05218f.

[21] A. K. Mandle, P. Jain, and S. K. Shrivastava, "P Rotein S Tructure P Rediction U Sing," vol. 3, no. 1, pp. 67–78, 2012.

[22] Y. Guo, B. Wang, W. Li, and B. Yang, "Protein secondary structure prediction improved by recurrent neural networks integrated with two-dimensional convolutional neural networks," J. Bioinform. Comput. Biol.,vol.16,no.5,2018,doi:10.1142/S021972001850021X.

[23] Z.Aydin,"Combining Classifiers for Protein Secondary Structure Prediction," pp. 6–10, 2017, doi: 10.1109/CICN.2017.9.

[24] Q. Jiang, X. Jin, S. J. Lee, and S. Yao, "Protein secondary structure prediction: A survey of the state of the art," J. Mol. Graph. Model., vol. 76, pp. 379–402, 2017, doi: 10.1016/j.jmgm.2017.07.015.

[25] Y. F. Chin, R. Hassan, and M. S. Mohamad, "Optimized local protein structure with support vector machine to predict protein secondary structure," Commun. Comput. Inf. Sci., vol. 295 CCIS, no. January 2012, pp. 333–342, 2012, doi: 10.1007/978-3-642-32826-8_34.

[26] B. Rost and C. Sander, "Prediction of protein secondary structure at better than 70% accuracy," Journal of Molecular Biology, vol. 232, no. 2. pp. 584–599, 1993. doi: 10.1006/jmbi.1993.1413.

A Mobility Management Algorithm in the Internet of Things (IoT) for Smart Objects based on Software-Defined Networking (SDN)

Lili Pei

Department of Architectural Engineering
Tangshan Polytechnic College
Tangshan, 063299, China

Abstract—In recent decades, technological advancements have significantly improved people's living standards and given rise to the rapid development of intelligent technologies. The Internet of Things (IoT) is one of the most important research topics worldwide. However, IoT is often comprised of unreliable wireless networks, with hundreds of mobile sensors interconnected. A traditional sensor network typically consists of fixed sensor nodes periodically transmitting data to a pre-determined router. Current applications, however, require sensing devices to be mobile between networks. We need mobility management protocols to manage these mobile nodes to provide uninterrupted service to users. The interactions between the mobile nodes are affected by the loss of signaling messages, increased latency, signaling costs, and energy consumption because of the characteristics of these networks, including constrained memory, processing power, and limited energy source. Hence, developing an algorithm for managing smart devices' mobility on the Internet is necessary. This study proposes an efficient and effective distributed mechanism to manage mobility in IoT devices. Using Software-Defined Networking (SDN) based on the CoAP protocol, the proposed method is intended not only to reduce the signaling cost of messages but also to make mobility management more reliable and simpler.

Keywords—Internet of things (IoT); mobility management; software-defined networking (SDN); CoAP protocol

I. INTRODUCTION

In recent years, Internet of Things (IoT) technology has been used in multiple areas, including health supervision, crisis management, and transportation management. IoT is an information network of physical objects (sensors, machines, devices, etc.) that facilitates communication and cooperation between these objects to reach a specific aim [1-3]. Each device within the IoT can identify, evaluate, and interact with the existing internet infrastructure using its embedded system. In other words, the IoT is a computational concept that can describe a future in which physical objects are connected using the internet and create a network of connections with other objects. In this technology, each entity on the internet is assigned a Unique Identifier (UI) and an Internet Protocol (IP) address through which it can send the data to the designated databases [4-6].

The Internet Engineering Task Force (IETF) has introduced various standards for the interactions between web services and a network of smart objects. For instance, Constrained Application Protocol (CoAP) is an application layer protocol modeled based on the Representational State Transfer (REST) software architectural style that facilitates the communication between the resource-constrained devices and web services within the IoT infrastructure. CoAP is a simple and modified version of the Hypertext Transfer Protocol (HTTP) designed in 2013. In CoAP, two messages are mainly used: Confirmable (CON) to signal secure connections and Non-Confirmable (NON) to signal regular connections. When receiving the CON message, the receiver sends an acknowledgment message to indicate that the message was received correctly. If the acknowledgment message is not received, the CON message is sent out again after a specific time. Therefore, this format provides the ability to resend messages which, in turn, creates an atmosphere of trust within the entire network. Fig. 1 demonstrates the CoAP retransmission mechanism. Chun et al. [7] introduced an algorithm for retransmission management based on the CoAP retransmission mechanism that incorporates the features and message formats of CoAP.

Software-Defined Networking (SDN) has three layers: the infrastructure layer (data plane), the control layer (control plane), and the application layer [8]. The infrastructure layer, also known as the data plane, consists of the packet delivery elements (routers and switches). The control layer is the logical architecture of the software that sets out the delivery codes for the elements within the infrastructure layer while managing the networking and routing tasks. Separating the data and control planes enables the network operator to control the network behavior from the top down. The application layer manages the applications within the network. Software-Defined Networking (SDN) introduces the technology of a centralized network controller that increases network scalability and flexibility by separating the control and data planes. Zhou and Zhang [9] introduce a host-based method that uses SDN technology to improve mobile IP and manage filtering and routing. The problem with this method is that it requires a hostname and IP address change. Furthermore, Raza et al. [10] introduced a network-based retransmission method that sets out an OpenFlow-based Proxy Mobile IPv6 (PMIPv6) protocol. This method separates the signaling control route from the communication data path. However, using Mobile Nodes (MN)

to send Router Solicitation (RS) messages and using the network to send Router Advertisement (RA) messages leads to significant delays within the system. Furthermore, this method uses an IP tunnel instead of an OpenFlow, which requires more network overhead. Chen et al. [11] introduced an SDN-based retransmission protocol that decreases the retransmission timeout. In this method, network switches occur in parallel and simultaneously with layer two switches. At the same time, active currents are delivered to all possible target channels. Network switch configuration takes place through an optimal track method.

Unfortunately, most IP-based standard Transmission Control Protocols (TCPs) are incompatible with IoT infrastructure [12, 13]. IoT structure consists of hundreds of interconnected mobile sensor nodes. Since these sensors have constrained memory, processing ability, computing ability, and energy resources, they introduce multiple challenges to the system that can impact delay-sensitive applications. On the other hand, most standard TCPs, such as Mobile IPv6 (MIPv6), have significant signaling overhead for tunneling and binding. That is why they impose a significant processing overhead on the network [14, 15]. Within the IoT ecosystem, mobile sensor

nodes should be able to deliver the analyzed data to the remote user periodically. As a result, IoT requires a new protocol to control transmission that can meet the various needs of the system based on mobile sensors' features while operating on a constrained energy and sleep mode. In this study, to come up with a solution for this problem, instead of using IP-based standard TCPs, CoAP and SDN-based TCPs are used since they introduce an effective transmission control mechanism for mobile sensor nodes while decreasing signaling costs.

To get a clear overview of the research, short descriptions of the structure of the research paper are summarized as follows:

- Section II: In this section, the details of the proposed solution for managing the mobile nodes of the resource-constrained in the IoT will be examined.

- Section III: This section will evaluate and compare the proposed solution with previous methods in this field.

- Section IV: In this section, the summary and conclusion of the results will be presented.
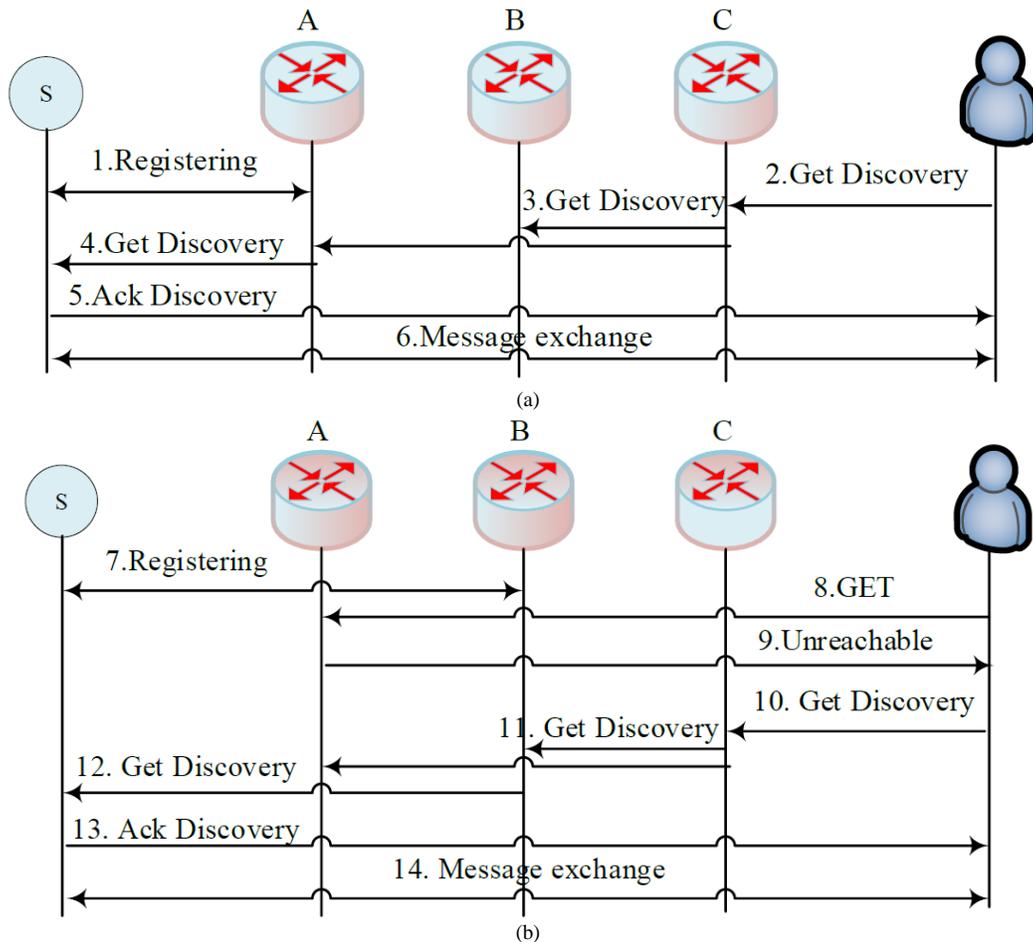
Fig. 1.   CoAP Retransmission Mechanism: (a) Before Retransmission, (b) After Retransmission.

II. PROPOSED SDN-BASED TRANSMISSION CONTROL ALGORITHM

A. Intra-Domain and Inter-Domain Transmission

Fig. 3 demonstrates what transpired before and after the Intra-Domain transmission. The order of the exchanged messages and their related reactions are included in the following:

- Phase 1: Mobile Node (MN) service network delivers the "weak signal" message to the controller.

- Phase 2: After receiving the "weak signal" message, the controller examines the transmission protocol of all neighboring networks to identify the host channels. Then, all the active currents receive the mobile nodes from the current manager and compute the number of routes needed for multicasting.

- Phase 3: After computing the routes, the manager delivers the current change messages to adjust routing tables in the associated networks.

- Phase 4: CoAP-based Mobile Node, while entering a domain, by exchanging the Power-On Self-Test (POST) message, records its IP address within the network and receives an acknowledgment message.

- Phase 5: The network informs the controller of a new mobile sensor connection and sends the mobile sensor's IP address within the report to the controller.

- Phase 6: After receiving the network's message, the controller sends the current change messages to the host networks to eliminate the unnecessary currents.

Fig. 4 demonstrates the order of the exchanged messages in inter-domain transmissions. As demonstrated in Fig. 4, Phases 1, 2, and 3 in this method are similar to intra-domain transmission. The source controller sends out the pre-transmission request to the location server. The location server identifies host networks and sends requests to their domain controllers. When the host network's controller receives the request, it estimates the pre-transmission route and sets out the current tables. In the final phase, by delivering the address report, the target network informs the controller of a new node connection within its domain. The controller transmits the update message to the location server. The location server updates the new location of the mobile sensor node. It sends out the transmission report to Primary Site Controller (PSC). After receiving the report, the controller eliminates the information related to the transmitted mobile node and the extra currents. Fig. 2 shows the SDN controller architecture in the IoT.
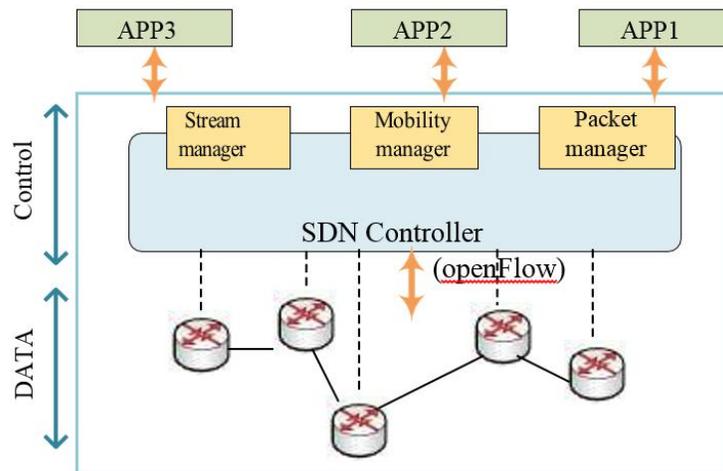
Fig. 2. SDN Controller Architecture in the IoT.
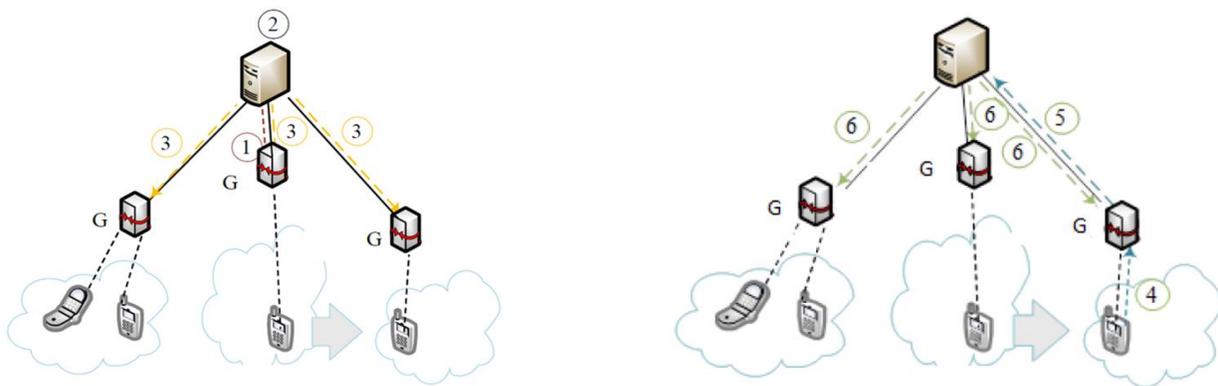
Fig. 3. Intra-Domain Transmission: a) Intra-Domain Pre-Transmission, b) Completed Intra-Domain Transmission.
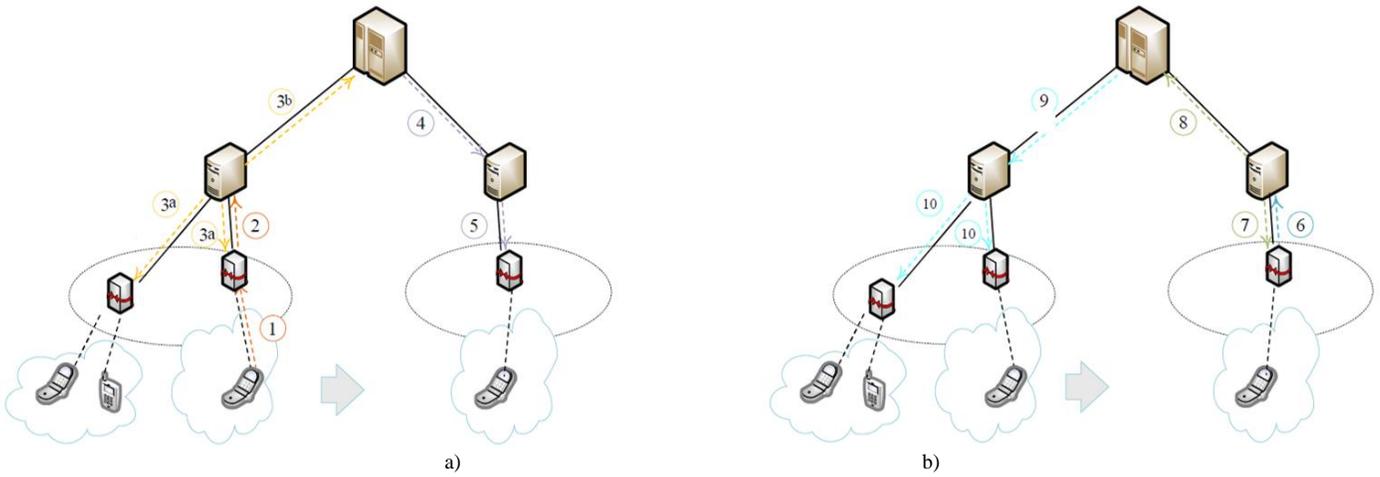
Fig. 4.   Inter-Domain Transmission: a) Inter-Domain Pre-Transmission, b) Completed Inter-Domain Transmission.

B. Transmission Delay and Signaling Costs Analysis

CoAP provides End-to-End Security using CON messages and a Stop-and-Wait Automatic Repeat Request (ARQ) mechanism. If the mobile node does not receive an ACK for its CoAP Confirmable Message (CON) within a specific time, it retransmits the same CON.

Based on the calculations done in [7], the packet loss ratio in the wireless transmission is calculated through the following equation:

$$P = 1 - (1 - S)^{2m} \qquad (1)$$

In which S is the sampling error and m is the packet length.

The Packet Error Rate (PER) in the application layer is calculated through the following equation:

$$E = P^{c+1} \qquad (2)$$

In which c equates to the maximum number of retransmissions in the MAC layer.

When it comes to packet segmentation in CoAP, Packet Error Rate is calculated by:

$$E_s = \sum_{i=1}^{s} E \times (1 - P)^{i-1} \qquad (3)$$

The maximum number of packages is 12.

The probability of message delivery for i times of P_R retransmission of a CoAP message using the s packet is estimated through the following formula:

$$P_R^i = (1 - E_s) \times E_s^i \qquad (4)$$

Therefore, the average retransmission time for each CoAP message is:

$$E(R) = \sum_{i=0}^{m-1} i \times P_R^i + (m-1) \times E_s^m \qquad (5)$$

in which m is the maximum count of message retransmission that the internet-connected system can approve.

If m is the maximum number of message retransmission, then the probability of a CoAP message not being delivered is E_s^{m+1} and the probability of its successful transmission is $\tau = 1 - E_s^{m+1}$. Since n sensors are competing with each other

to transmit the messages within the channel, the probability of having at least one successful transmission in each time slot is:

$$P_s = 1 - (1 - \tau)^n \qquad (6)$$

Therefore, the possibility of successful transmission within a channel provided that only one sensor transmits the message is calculated by the following formula:

$$P_{succ} = \frac{n\tau (1-\tau)^{n-1}}{1-(1-\tau)^n} \qquad (7)$$

At this point, while taking into account Round Trip Time (RTT) and the time required for receiving the Acknowledgment report for each message (T_{ACK}), we calculated the delayed signaling messages in CoAP, which is estimated at:

$$D = \sum_{i=0}^{m-1} P_R^i \times (RTT + T_{ACK}) + E_s^m \times T_{ACK} \qquad (8)$$

RTT is the time from the point of message transmission to the point of receiving a response estimated through the following equation:

$$RTT = \alpha \times old\ RTT + (1 - \alpha) \times new\ RTT \qquad (9)$$

In this equation, α is always between 0 and 1.

Finally, the signaling costs can be estimated from the following formula while taking into account the Packet Length (Lp) and Packet Arrival Rate:

$$C_p = \lambda_s \times L_p \times D \qquad (10)$$

In which λ_s is the packet arrival rate, Lp is the Packet Length, and D is the packet transmission delay.

III.   RESULTS AND DATA ANALYSIS

In this study, we used the two methods of mathematical simulation and analysis to examine the effectiveness of the proposed algorithm. In the simulation method, the study is based on the grounds that it matches the circumstances in real-life situations and operational networks as much as possible. There are various simulators, including OMNet++, NS, and Opnet. Opnet is one of the most powerful and popular simulators that many researchers use to simulate their proposed models. To examine the effectiveness of the proposed

algorithm, we implemented it in the Opnet simulator to compare it with the algorithm introduced by Chun et al. [7]. We used Matlab software to analyze the derived mathematical models.

A. Evaluation Criteria

The proposed algorithm should be assessed based on the basic algorithm to examine its performance of the proposed algorithm. To examine algorithm efficiency, some parameters should be studied that are included in the following:

- Signaling Latency: the time it takes for a signaling message to be transmitted and received indicates the signaling Latency.

- End-to-End Delay: the time it takes to send a request and receive its response is the end-to-end Delay. End-to-End Delay consists of signaling time, request transmission time within the link, and the time required to receive the Acknowledgment message and its response.

- Traffic Load: this parameter indicates the number of packets transmitted within the network at any time.

- Throughput: the average data packet successfully transmitted within the network is called network throughput. Network throughput is typically measured in bits per second or packets per second. In this study, the packet's network throughput is measured per second.

- Signaling Cost: The average costs of the exchanged signaling messages are measured based on the packet length and link latency.

Successful Transmission Probability Rate: successful transmission probability rate is measured by dividing the number of packets that were transmitted successfully within the channels by the total number of packets.

B. Results of the Proposed Algorithm Analysis

Within the IoT ecosystem, retransmission delay and the packet loss ratio are two of the most significant factors in evaluating the transmission algorithm. Channel delay and error rates are calculated by the formula introduced in the previous section. In this part of the study, we implemented Matlab's resulting correlations and calculated the transmission delay and signaling costs. To examine the system's efficiency, we provided networks of sensors that have n mobile sensors connected to the internet using a network. Networks support the OpenFlow protocol through which they connect to the controller.

On the other hand, the networks are connected to the sensor nodes using CON messages in CoAP. The parameters used in this analysis are similar to the ones used by Makaya and Pierre [16] in Table I. The bandwidth of the wireless network and the wired network are 250 kilobits per second and 10 megabits per second, respectively.

TABLE I. THE PARAMETERS USED IN THIS STUDY

Parameter	Value
Wireless link latency	15 ms
Wired link latency	2 ms
Wired bandwidth	10 Mbps
Wireless bandwidth	250 kbps
Average node speed	10 m/s

Fig. 5 demonstrates how RTT affects signaling message latency. As demonstrated in Fig. 5, transmission delay increases signaling latency. However, the Quality of Service (QoS) requirements' impact on signaling latency depends on the internet packet types (IP packet or CoAP packet) and the number of parts within each packet. Fig. 5 demonstrates that small network packets impose fewer network delays than the packets that contain more parts. This parameter begins to increase in network packets that have more parts. That is to say, for more extensive data packets, latency occurs mainly because of the traffic competition in the channel and not because of the multitude of packet parts.
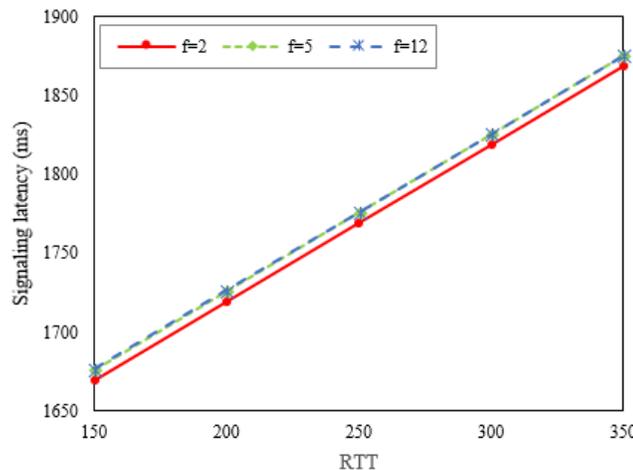
Fig. 5. Signaling Latency Caused by Round Trip Time (RTT) Variations.

Fig. 6 demonstrates signaling costs based on the variation in entry rates. As demonstrated in Fig. 6, when entry rates increase, signaling costs increase, as well. The competition between the nodes for packet transmission within the channel increases the number of packet retransmissions within the channel, which, in turn, raises signaling costs.

The probability of successful transmission based on the number of sensor nodes within the network is demonstrated in Fig. 7. As demonstrated in Fig. 7, the successful transmission rate decreases with increased sensors within the network. This parameter can positively impact the retransmission rate and signaling costs.

C. Proposed Algorithm Simulation

In this study section, we introduce the structure of the simulation modeling of the M-CoSDN transmission mechanisms. An Opnet simulator was employed in this study for simulation modeling. Furthermore, this study provides a comparative analysis of the M-CoSDN proposed algorithm and the algorithm proposed by Chun et al. [7].

Fig. 8 demonstrates the network topology of the inter-domain simulation mainly used in transmission management studies. A 200×200 Wireless Sensor Network (WSN) with a range of 50 meters was chosen for this study. Each trajectory covers around 20 meters. A specific packet is introduced for each exchanged message during the signaling process. A specific packet based on the protocol format was introduced for the controller's response, acknowledgment, and current messages. In this study, the average incoming request time is not changed and is calculated every two milliseconds. Therefore, the incoming requests are transmitted to the controller every two milliseconds.

Fig. 9 demonstrates the End-to-End Delay within M-CoSDN proposed algorithm and the basic algorithm. As predicted, End-to-End Delay within the proposed algorithm is much lower than the basic algorithm. Using a central controller for transmission control that transmits current messages and controls the network's topology and sensors accelerates transmission. That is why the proposed algorithm detects a few delays.
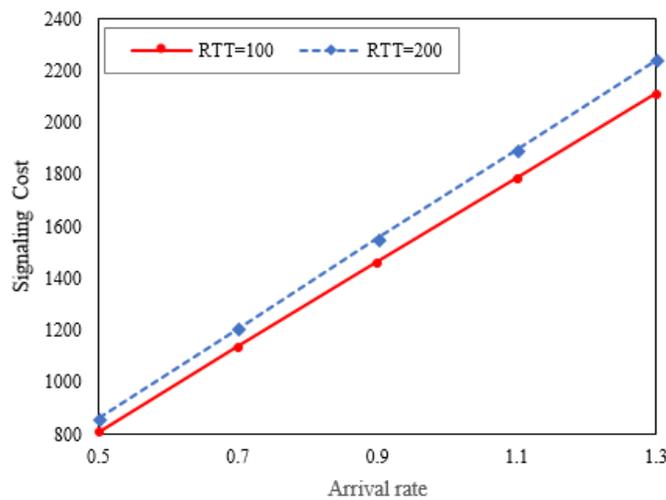
Fig. 6. Signaling Costs Caused by Increasing Entry Rates.
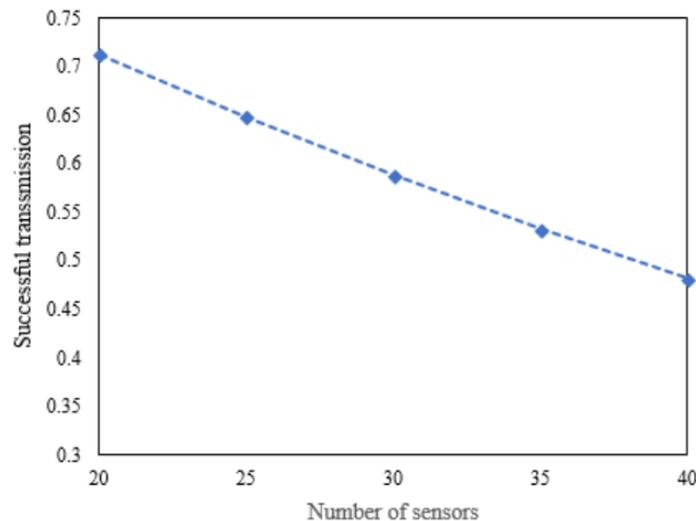
Fig. 7. Successful Transmission Probability within the Channel is Caused by Variance in the Number of Sensor Nodes.
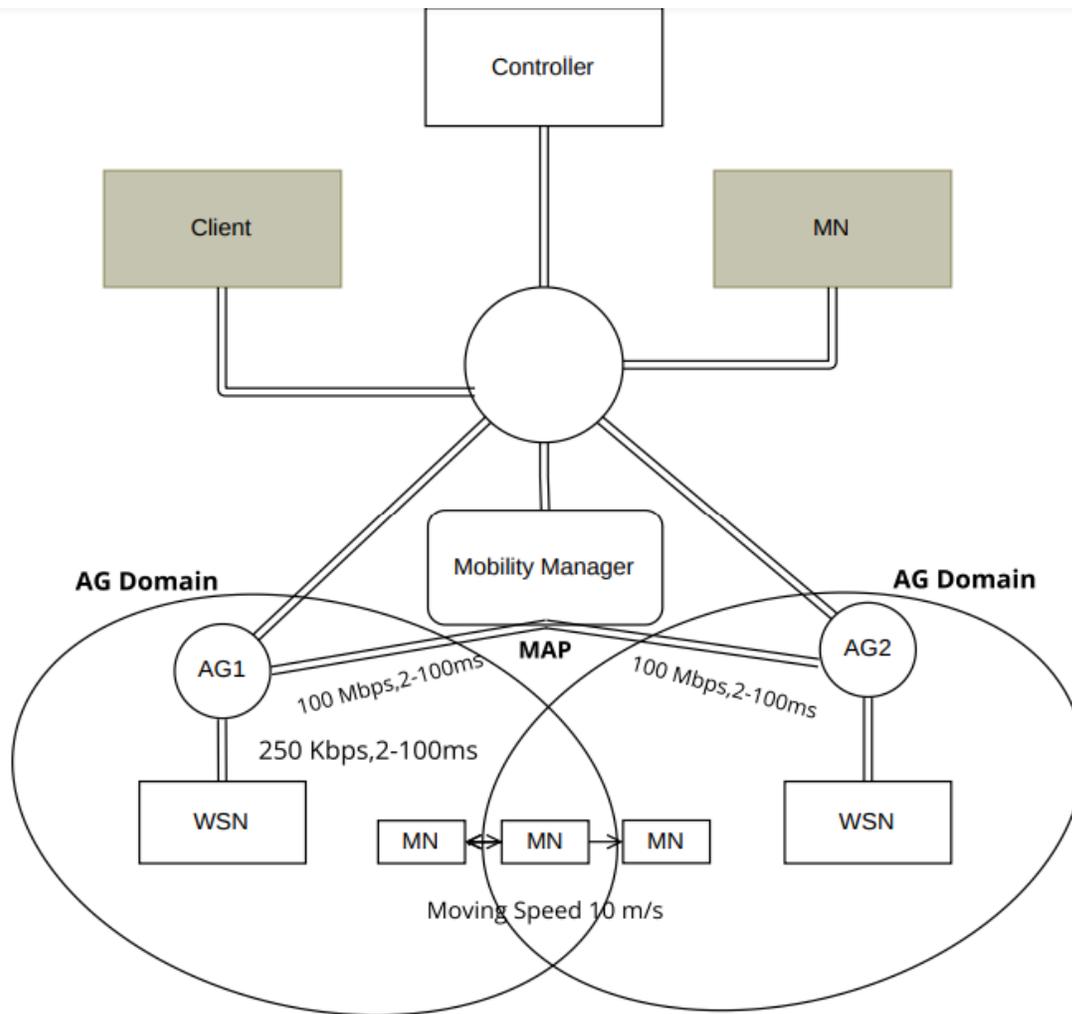
Fig. 8. Network Topology and Simulation Parameters.
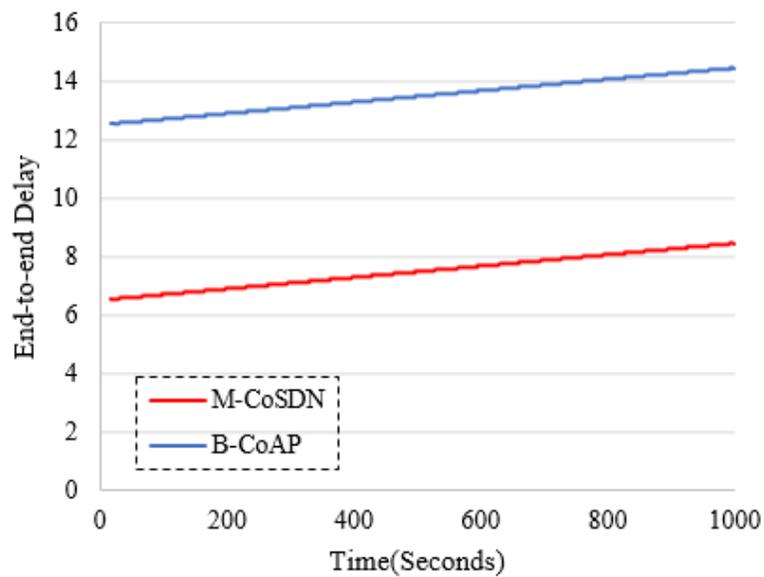
Fig. 9. End-to-End Delay within the Proposed and the basis Algorithm.

Within the basic algorithm, during mobile sensor nodes transmission, the user receives a 'Destination Inaccessibility Message' that informs him of the ongoing transmission. Then, the user finds the node's location through the retransmission of the source discovery message. This process leads to transmission delays. While within the proposed model, the mobile sensor node informs the controller of the domain change using the network. Therefore, the controller can reduce the delays to a bare minimum by updating the current tables and sending them to the network routes. Increasing the simulation time contributes to more delays within both algorithms. The increasing traffic load and transmitted packets within the network cause queuing delays in nodes and other network entities. That is why the End-to-End Delay is bound to increase with time.

Fig. 10 compares network throughput within the two models. In this study, we defined network throughput as the number of successful requests divided by the total number of requests. As observed in Fig. 10, the network throughput in the proposed model is higher than the basic algorithm. The number of signaling messages transmitted between the two entities that might cause delays or collisions within the channel contribute to the higher network throughput in the proposed model. At the beginning of the simulation phase, network throughput was increasing at a high pace. However, with the increasing network traffic loads that cause collisions and errors within the channel, network throughput is decreased at a moderated pace. Therefore, network throughput is dependent on the channel type, errors, and network traffic load.
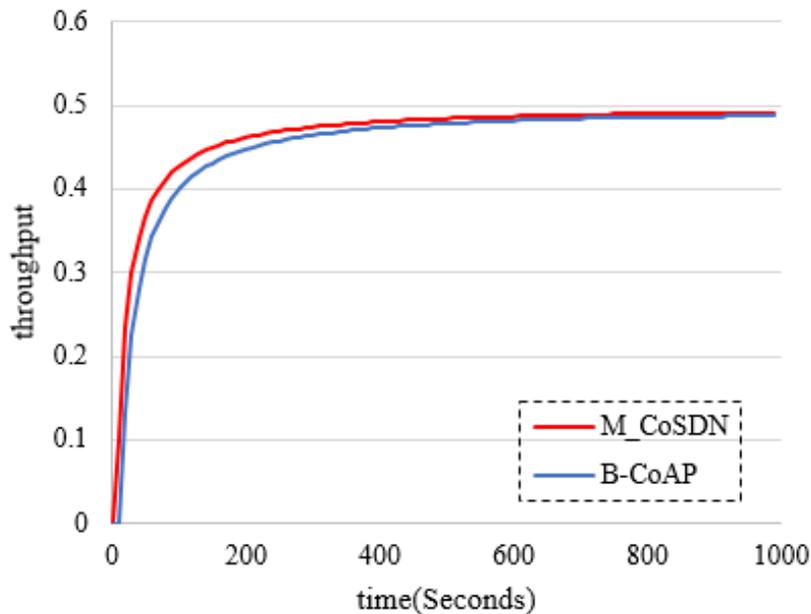
Fig. 11 demonstrates the network traffic load within the proposed and basic algorithms. As demonstrated in Fig. 11, in the basic algorithm, because of the repeated source discovery within each transmission, the exchanged traffic load increases. As predicted, increasing the simulation time leads to an increased network traffic load in the basic algorithm compared with the proposed algorithm that follows a higher positive slope. Traffic load increase, in turn, leads to increased delays and packet loss within the network. Therefore, network traffic load can impact network throughput and overall productivity.

In this part of the research, we have conducted a study on inter-domain routing. To do so, we have used two different domains controlled by two controllers within the network topology. Furthermore, a new entity called 'location server' is added to the network that records the data related to all network sensor nodes. First, the network's End-to-End Delay is studied (Fig. 12).

It is observed that inter-domain routing imposes more delays than intra-domain routing. However, by comparing Fig. 12 with Fig. 9, it is observed that inter-domain routing within the proposed model imposes much fewer delays than the basic algorithm. Both networks produce the same traffic load (Fig. 13) since both timeouts and delays are the same.

Fig. 14 demonstrates that inter-domain network throughput is slightly less than intra-domain network throughput. Increased packet delay and packet collisions in the network route decrease the number of successful requests, which, in turn, causes a slight decrease in inter-domain network throughput compared with intra-domain network throughput.

Fig. 10. Network throughput for the Proposed and basis Algorithms.
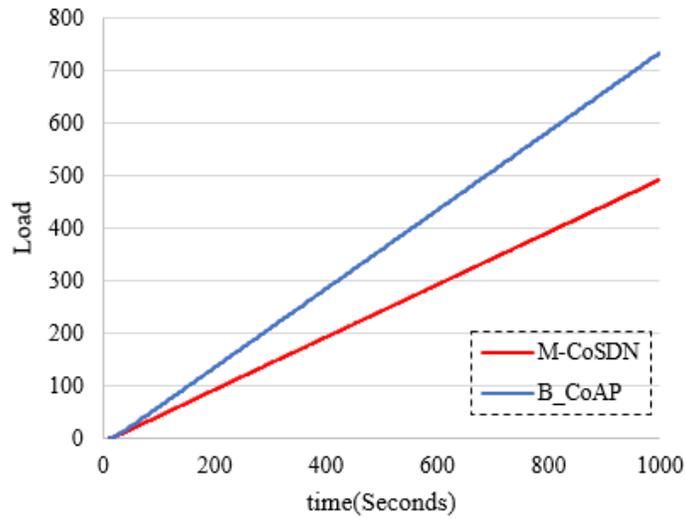
Fig. 11. Network Traffic Load.
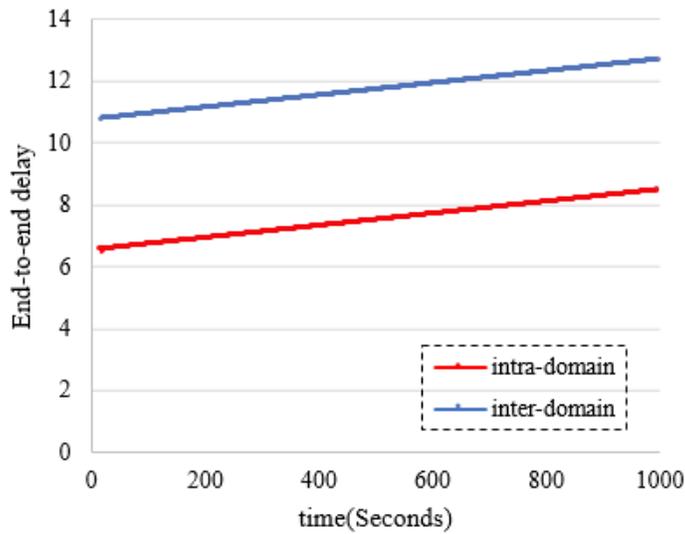
Fig. 12. End-to-End Delay within the Two Inter-Domain and Intra-Domain Scenarios.
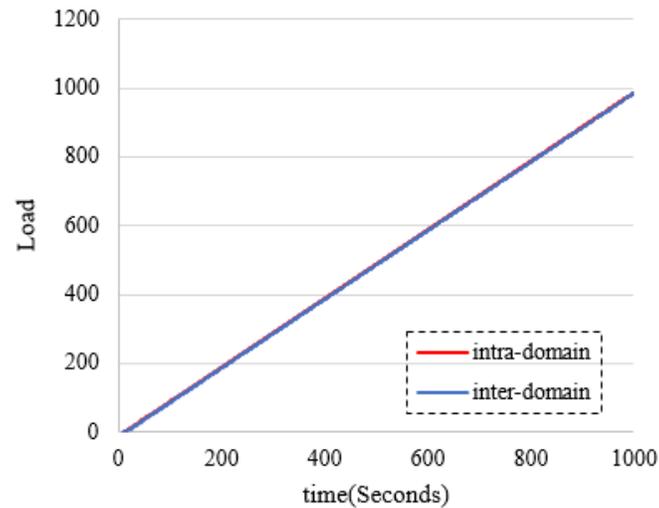
Fig. 13. Network Traffic Load.
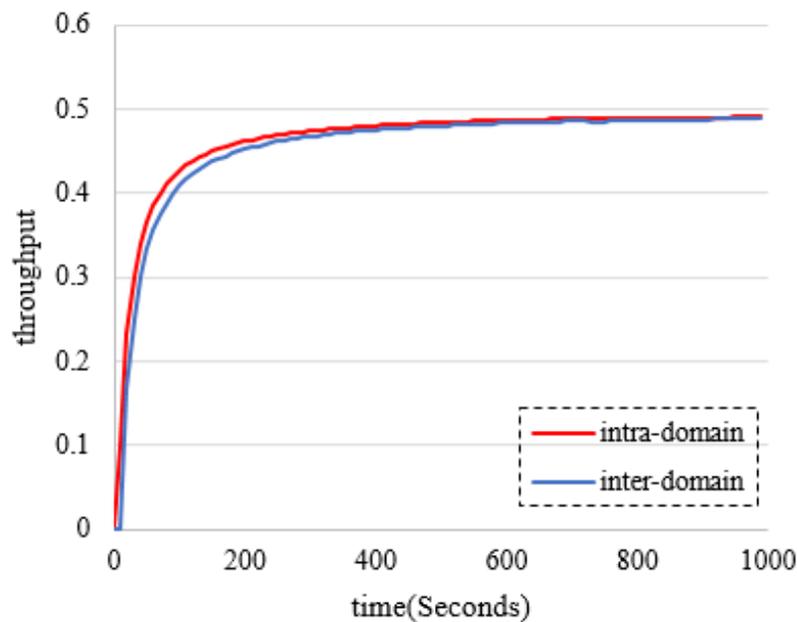
Fig. 14. Two Scenarios for Network throughput.

IV. CONCLUSION

IoT is a concept that connects multiple devices. The purpose of IoT is to facilitate human life and information exchange. IoT has multiple applications that can impact multiple aspects of human life, including personal, business, and industrial sectors. To increase network flexibility and reduce the costs of sensor nodes' transmission, we need to take a practical approach to the sensor nodes' transmission control. Because of the specific structure, infrastructure, and features of IoT, standard TCPs will not be of use. Therefore, it is essential to introduce a protocol based on the specific features of the IoT.

In this study, a secure M-CoSDN transmission model is introduced that would decrease signaling latency and costs in the IoT ecosystem. The proposed model uses SDN technology to impose centralized control on mobile sensor nodes within the network. CoAP CON messages are used to introduce a secure transmission model. Multicast transmission prevents packet loss and transmission delay for mobile sensors. The efficiency analysis results of the proposed algorithm indicate that it is far more effective than other models.

REFERENCES

[1] Jeschke, S., Brecher, C., Song, H., & Rawat, D. B. Industrial Internet of Things: Cybermanufacturing Systems. 2017. ISBN-13.

[2] Lin, J., Yu, W., Zhang, N., Yang, X., Zhang, H., & Zhao, W. (2017). A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. IEEE internet of things journal, 4(5), 1125-1142.

[3] Huang, C., & Huang, Y. (2022). Information Fusion Early Warning of Rail Transit Signal Operation and Maintenance Based on Big Data of Internet of Things. Sustainable Computing: Informatics and Systems, 100763.

[4] Ngu, A. H., Gutierrez, M., Metsis, V., Nepal, S., & Sheng, Q. Z. (2016). IoT middleware: A survey on issues and enabling technologies. IEEE Internet of Things Journal, 4(1), 1-20.

[5] Gazis, V. (2016). A Survey of Standards for Machine-to-Machine and the Internet of Things. IEEE Communications Surveys & Tutorials, 19(1), 482-511.

[6] Zhang, J., Ma, M., Wang, P., & Sun, X. D. (2021). Middleware for the Internet of Things: A survey on requirements, enabling technologies, and solutions. Journal of Systems Architecture, 117, 102098.

[7] Chun, S. M., Kim, H. S., & Park, J. T. (2015). CoAP-based mobility management for the Internet of Things. Sensors, 15(7), 16060-16082.

[8] Braun, W., & Menth, M. (2014). Software-defined networking using OpenFlow: Protocols, applications and architectural design choices. Future Internet, 6(2), 302-336.

[9] Zhou, Q., & Zhang, R. (2013). A survey on All-IP wireless sensor network. In LISS 2012 (pp. 751-756). Springer, Berlin, Heidelberg.

[10] Raza, S. M., Kim, D. S., & Choo, H. (2014, January). Leveraging pmipv6 with sdn. In Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication (pp. 1-8).

[11] Chen, C., Lin, Y. T., Yen, L. H., Chan, M. C., & Tseng, C. C. (2016, April). Mobility management for low-latency handover in SDN-based enterprise networks. In 2016 IEEE wireless communications and networking conference (pp. 1-6). IEEE.

[12] Ghaleb, S. M., Subramaniam, S., Zukarnain, Z. A., & Muhammed, A. (2016). Mobility management for IoT: a survey. EURASIP Journal on Wireless Communications and Networking, 2016(1), 1-25.

[13] Zhang, G., & Navimipour, N. J. (2022). A comprehensive and systematic review of the IoT-based medical management systems: Applications, techniques, trends and open issues. Sustainable Cities and Society, 103914.

[14] Sheng, Z., Yang, S., Yu, Y., Vasilakos, A. V., McCann, J. A., & Leung, K. K. (2013). A survey on the ietf protocol suite for the internet of things: Standards, challenges, and opportunities. IEEE wireless communications, 20(6), 91-98.

[15] Yugha, R., & Chithra, S. (2020). A survey on technologies and security protocols: Reference for future generation IoT. Journal of Network and Computer Applications, 169, 102763.

[16] Makaya, C., & Pierre, S. (2008). An analytical framework for performance evaluation of IPv6-based mobility management protocols. IEEE Transactions on wireless communications, 7(3), 972-983.

SPAMID-PAIR: A Novel Indonesian Post–Comment Pairs Dataset Containing Emoji

Antonius Rachmat Chrismanto[1], Anny Kartika Sari[2]*, Yohanes Suyanto[3]

Department of Computer Science and Electronics, Faculty of Matematics and Natural Science, Universitas Gadjah Mada[1,2,3]
Faculty of Information Technology, Universitas Kristen Duta Wacana, Yogyakarta, Indonesia[1]

Abstract—The detection of spam content is an important task especially in social media. It has become a topic to be continuously studied in Natural Language Processing (NLP) area in the last few years. However, limited data sets are available for this research topic because most researchers collect the data by themselves and make it private. Moreover, most available data sets only provide the post content without considering the comment content. This data becomes a limitation because the post-comment pair is needed when determining the context of a comment from a particular post. The context may contribute to the decision of whether a comment is a spam or not. The scarcity of non-English data sets, including Indonesian, is also another issue. To solve these problems, the authors introduce SPAMID-PAIR, a novel post-comment pair data set collected from Instagram (IG) in Indonesian. It is collected from selected 13 Indonesian actress/actor accounts, each of which has more than 15 million followers. It contains 72874 pairs of data. This data set has been annotated with spam/non-spam labels in Unicode (UTF-8) text format. The data also includes a lot of emojis/emoticons from IG. To test the baseline performance, the data is tested with some machine learning methods using several scenarios and achieves good performance. This dataset aims to be used for the replicable experiment in spam content detection on social media and other tasks in the NLP area.

Keywords—Dataset; natural language processing; spam detection; spamid-pair; post-comment pairs

I. INTRODUCTION

Research on text analysis, especially in context-based detection/classification problems, is increasingly important because of the higher need for system automation. A labeled dataset is needed for supervised text classification to be used as machine learning data. Unfortunately, the datasets for text classification are mainly in English. Other languages (Turki [1], Bangla [2], Chinese [3], Arab [4], and Morocco [5]), including Indonesian, are rare enough [6], [7].

Datasets for text classification can be divided into two types: single-text classification and paired-text classification. Some examples of single-text classification datasets are news classification, sentiment classification, hoax classification, spam, topic classification, and emotion text classification. Examples for paired text classification are text entailment classification, duplicate question classification, text pair similarity classification, including spam comment classification based on a particular post on the social media. One of the challenges in the NLP area is how to understand the context to gets the meaning. Context understanding can also be applied to spam comment detection based on its post by detecting the comment's relevance. If the comment is not related/relavance to its post, it is likely to be categorized as spam. To detect spam comments, machine learning methods require training datasets that can be used according to the context, such as in the context of the language, that are still rare.

The motivation of this research is to overcome the datasets scarcity in Indonesian for the text pairs classification to get the context between two texts in pairs correctly. The authors have collected the dataset for spam comment detection based on social media posts. This dataset is taken from Instagram (IG) based on selected 13 public Indonesian artists/actors inspired by [8], [9]. Each of the public Indonesian artists/actors has more than 15 million followers. Each row of this dataset consists of a post and comments text pair called SPAMID-PAIR[1]. SPAMID-PAIR contains 72874 pairs of posts and comments and breaks down into 53837 non-spam data and 19037 spam data.

This article introduces the SPAMID-PAIR dataset, a novel dataset collected, labeled, validated, and used as training data for spam comment detection based on their posts with several machine learning methods. This dataset is intended to contribute as one of the Indonesian datasets in NLP for text pair classification problems based on the context. The SPAMID-PAIR dataset has an advantage because it contains symbols, special characters, and emojis that are widely available in social media posts and comment texts. This dataset is useful for NLP research because most researchers discard emojis in their classification techniques. Some examples are news article classification [10], Twitter without emoji [11], spam comments from the blog [12], Twitter (removed emoji) [13], SMS and Twitter without emoji [14], Twitter without emoji [15], Youtube comment without emoji [16], video spam comment without emoji [17], Youtube comment without emoji [18]. The emoji is essential because most social media users use emojis to express their feelings, such as to support/deny, show sympathy, joy, sadness, and anger. The emojis in the dataset is needed for research in some fields that learn through emoji expression. This dataset uses the UTF-8 format for post and comment data, so both emojis can be used in emoji pairs expression research.

The contribution of this paper is two-fold; first, the novel SPAMID-PAIR dataset, and second, several machine learning algorithms will be used to implement the supervised text-pair classification using this dataset using the F1 score. This paper is written as follows, firstly, the introduction of SPAMID-PAIR and its purpose. Secondly, the related works of the Indonesian NLP dataset, the experiments and results using this dataset, and finally, the conclusion.

II. Related Works

Datasets are the primary data source in machine/computer learning. Various machine learning and deep learning techniques are in dire need of data sources for system learning. But in reality, not all public datasets are available, especially in Natural Language Processing (NLP). Even though learning datasets in NLP are quite widely available in English, such as IMDB Dataset [19]–[21], SMS Spam UCI [22], FLAIR [23], [24], Twitter Spam [25], [26], YouTube Comments [15], PeerRead [27], and Huggingface Community Datasets [28]. Still, there are few public datasets in other languages, especially Indonesian.

IndoNLU [6] is one of two dataset sources in the field of Natural Language Understanding (NLU) for 12 main tasks that have been attempted to be collected in collaboration with universities and industry. IndoLEM [29], as the second source, is a dataset source in NLP for seven main tasks (post tagging, named entity recognition, parsing, sentiment analysis, summarization, and word prediction). IndoLEM, the second, provides datasets, Indonesian Fasttext, and BERT pre-trained that can be used for other tasks. To the best of our knowledge, unfortunately, for the case of the semantic task in detecting spam comments in social media based on the context of the post in pairs, it has not been found. This article introduces SPAMID-PAIR to enrich the Indonesian NLP dataset collection in spam comment detection based on its post context, which has not been done before.

Spam text detection on social media is mostly done on Twitter [11], [13], [14], [15]. Twitter has a structure that is not in the posts and comments pair structure. Otherwise, Youtube, Facebook, and Instagram are examples of social media with posts and comments pair structures. However, the detection of spam comments in previous studies was not based on paying attention to the post. The previous research used some popular machine learning methods. Septiandri and Wibisono use Naïve Bayes, SVM, and XGBoost to detect spam comments from Instagram, and SVM outperformed the others [30]. Zhang uses the Random Forest to detect Instagram spam posts and achieve good [31]. Research [32] investigated 11 state-of-the-art machine learning methods in text classification using 71 datasets and obtained that Stochastic Gradient Boosting, SVM, and Random Forest were the best methods compared to the others. That research can be used as a reference for the best machine learning in classification.

III. Research Method

This research uses the following steps: data acquisition, dataset construction, data profiling, annotation/labeling, pre-processing, feature extraction/generation, ML algorithm implementation, and evaluation. These stages can be seen in

Fig. 1, while a more detailed explanation is in the following sub-chapters.
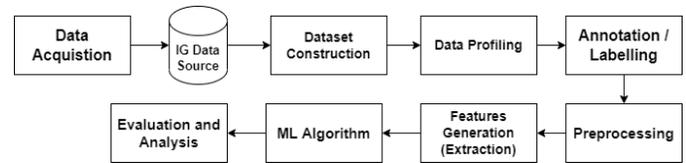
Fig. 1.   The Research Method.

A. Data Acquisition

In the data acquisition stage, IG was chosen because 1) IG has a lot of spam comments, especially on Indonesian public figure accounts [33], [34]. 2) Posting and commenting on IG is in pairs suitable for the pair dataset; 3) IG has a lot of non-formal posts and comments, and it also contains a lot of emojis; 4) IG does not have a spam filtering feature in Indonesian yet. For comparison, on Twitter (TW), a tweet is a post, but replies from other users must always use a mention tag, so the form of the reply is not a comment. The reply data is equivalent to the tweet, not as a child node. On Facebook (FB), a user can create a status/post, and others can comment on it. But on FB, the situation tends to be more formal/serious, so it does not contain much spam and emojis. Nowadays, IG is a famous social media with many young IG users; not as serious and formal as FB.

Comparing three leading social media existing today, e.g., IG, TW, dan FB, IG is the best choice for collecting datasets for spam detection. IG is widely used by public figures such as politicians, artists/actors, and well-known people. Very limited datasets are available in languages other than English and Chinese, especially Indonesian [6], making collecting this dataset more critical. The SPAMID-PAIR dataset from IG contains post-comment pairs from 13 Indonesian artists/actors with more than 15 million followers without stating their account names. It is expected that researchers in the NLP field can use this dataset to replicate research and use it as the dataset reference in the topic of spam detection using various algorithms.

The SPAMID-PAIR dataset was retrieved using several tools such as Instaloader and Chrome Selenium Python driver. For the first planning, the data is taken from the 50 most recent posts, and 120 most recent comments are taken from each post. Hence, it was estimated that 78000 data could be collected. However, the data is not as planned in reality because some posts do not have as many comments as expected. The dataset was collected in September 2020, and after data retrieval was completed, 72874 pairs of post and comment data were obtained, which are ready for further processing.

Table I displays all the artists/actor's usernames used in the SPAMID-PAIR dataset. SPAMID-PAIR contains 72874 pairs of posts and comments and breaks down into 53837 non-spam data (73.87%) and 19037 spam data (26.13%). Details of the number of spam and non-spam labels per artist/actor are highlighted in Table II, analyzed using Python Pandas. Table II also shows that the IG ID 24239929 only has 103 data because the user recently had disabled comments, so the data could not be retrieved anymore. Spam comments are detected in all 13

IG users chosen with varying percentages. The SPAMID-PAIR dataset consists of 11 fields and is available in Excel format (.xlsx) and comma-separated value (CSV) with UTF-8 encoding, as described per field in Table III.

TABLE I. THE 13 PUBLIC FIGURES USED IN THE SPAMID-PAIR DATASET WITH MORE THAN 15 MILLION FOLLOWERS (PER DECEMBER 2021)

Account ID	Followers (millions)
1918078581	54.3
522969993	47.4
225064794	42.4
24239929	36.4
2993265	34.1
361869464	33.6
26444210	33.4
1948416	30.7
305384601	27.3
8115577	27.1
5735890	25.8
4934196	25.2
30585021	15.7

TABLE II. DETAILED STATISTICS OF SPAM AND NON-SPAM DATA PER ACCOUNT ID IN THE SPAMID-PAIR DATASET

Account ID	Count of Non-Spam	Count of Spam	%Non-Spam	% Spam	Sub Total
4934196	4565	2251	66,97	33,03	6816
522969993	5712	1108	83,75	16,25	6820
5735890	3397	691	83,10	16,90	4088
30585021	818	1065	43,44	56,56	1883
2993265	4528	2022	69,13	30,87	6550
1948416	4658	1945	70,54	29,46	6603
361869464	6854	2466	73,54	26,46	9320
225064794	4944	1804	73,27	26,73	6748
24239929	65	38	63,11	36,89	103
1918078581	5045	1557	76,42	23,58	6602
8115577	4818	1971	70,97	29,03	6789
26444210	5537	911	85,87	14,13	6448
305384601	2896	1208	70,57	29,43	4104
Total	53837	19037			72874

Table IV shows that the number of emojis in this dataset reaches 68%, and the number of emojis in the spam category is higher than in the non-spam category. Table V shows detailed data related to emoji statistics in the dataset. Fig. 2 illustrates the distribution of emoji in the SPAMID-PAIR dataset per IG artist ID and tells us how the emoji is related to the spam or non-spam label. Fig. 3 shows some correlation between some attributes of the SPAMID-PAIR dataset. First, it shows a correlation between the length of comments and spam labels. There is also a correlation between the length of comments and the number of emojis. Lastly, there is a correlation between the length of comments and the post length.

TABLE III. DESCRIPTION OF ATTRIBUTES IN THE SPAMID-PAIR DATASET

Attribute	Description
igid	Account ID
comment	Comment on a post
post	Post from an account ID
emoji	Whether the data contains emojis or not (1 or 0)
spam	Whether the data is spam or not (1 or 0)
lengthcomment	The character length of the comment
lengthpost	The character length of the post
countemojicomment	Number of emoji symbol characters in comments
countemojicommentuniq	Number of emoji symbol characters in comments (unique)
countemojipost	Number of emoji symbol characters in posts
countemojipostuniq	Number of emoji symbol characters in the post (unique)

TABLE IV. NUMBER OF EMOJIS IN THE SPAMID-PAIR DATASET

Category	Count	Percentage (%)
Non-Emoji	22710	31,16
Emoji	50164	68,83

TABLE V. NUMBER OF EMOJI IN THE SPAMID-PAIR DATASET PER ACCOUNT ID

Account ID	Count of Non-Emoji	Count of Emoji	% Non-Emoji	% Emoji	Sub Total
4934196	2085	4731	30,59	69,41	6816
522969993	1679	5141	24,62	75,38	6820
5735890	1013	3075	24,78	75,22	4088
30585021	1142	741	60,65	39,35	1883
2993265	2482	4068	37,89	62,11	6550
1948416	1857	4746	28,12	71,88	6603
361869464	3264	6056	35,02	64,98	9320
225064794	1935	4813	28,68	71,32	6748
24239929	2	101	1,94	98,06	103
1918078581	2052	4550	31,08	68,92	6602
8115577	2126	4663	31,32	68,68	6789
26444210	1592	4856	24,69	75,31	6448
305384601	1481	2623	36,09	63,91	4104
Total	22710	50164			72874

SPAMID-PAIR dataset profile generally has an average comment length of 34.23 characters and an average post length of 252.03 characters. The highest number of emojis (non-unique) in a comment is 359 emojis, and the highest number of unique emojis is 112. In the post data, the highest number of emojis (non-unique) is 32 emojis, and the highest number of unique emojis is 14. Complete statistical details of the comment and post data can be seen in Table VI. The maximum length of the comment is 386, and the post is 3938.
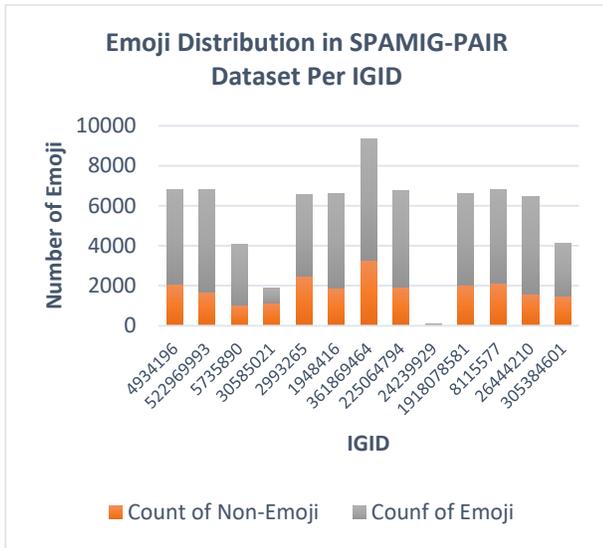
Fig. 2.   Distribution of Emoji in the SPAMID-PAIR Dataset.
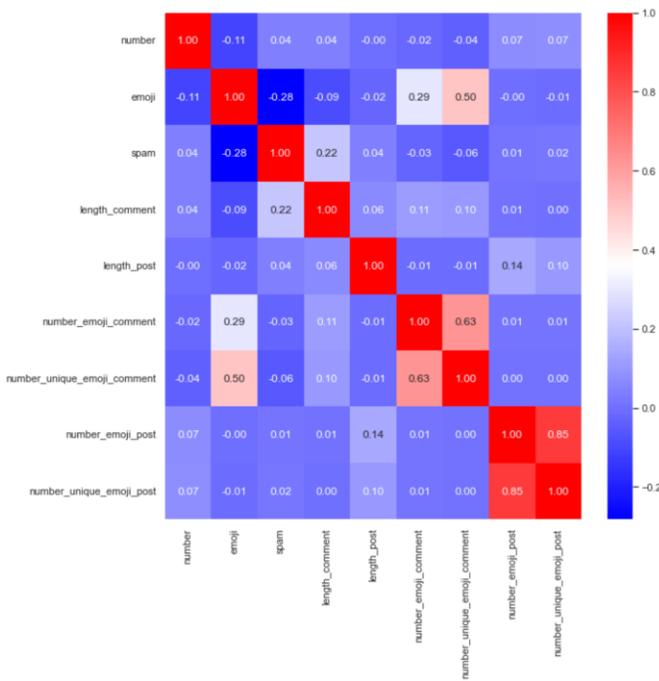
Fig. 3.   Attributes Correlation in the SPAMID-PAIR Dataset.

B. Data Profiling and Labelling

After the dataset has been collected, the next step is data profiling, labeling, and validation. Labeling gives each data a "spam" or "not spam" label. The "spam" criterion is given if the post and comment data, text data, or emojis are irrelevant. On the other hand, the "not spam" criteria will be given if the post and comment data are relevant. Two Indonesian labelers carried out the labeling process. Before starting the labeling process, a joint briefing was held between the two native Indonesian labelers to create a common perception of the meaning of "spam" and "non-spam" labels. After that, labeling was done using an excel formatted dataset that was given to each labeler, and there was one additional column, "label,"

which would be filled with "spam" or "not spam" by each labeler manually. The final label was determined by the final agreement of the two labelers. Based on the Kappa score, the result is the "almost perfect" category with a Kappa score of 0.95, proving that the labeling agreement between the two annotators was relatively easy. The difficulty arises when the comment contains only an emoji, and it is difficult to determine its meaning. However, it can be overcome by looking at the consistency of the type of emoji and the type of "positive" emoji used. Suppose the emojis use "positive" emojis such as expressions of joy, enthusiasm, support, and love. In that case, the label is a high possibility of "not spam." Otherwise, if the post content tends to be "positive" and the comment content tends to be "negative," it is labeled as "spam." Examples of labeling results for data labeled as "spam" and "not spam" can be seen in Table VII.

TABLE VI.   STATISTICAL INFORMATION ON COMMENTS AND POSTS DATA IN THE SPAMID-PAIR DATASET

Statistics of Comments	Average	Max	Statistics of Post	Average	Max
Number of sentences	1,1	29	Number of sentences	2,88	45
Number of characters	34,25	212	Number of characters	2,52	3938
Number of whitespaces	4,4	386	Number of whitespaces	33,13	570
Number of words	4,8	384	Number of words	35,33	602
Number of numbers (as a whole)	0,2	72	Number of numbers (as a whole)	1,1	35
Number of punctuations	1,1	213	Number of punctuations	10,52	192
Number of date format	0,00013	1	Number of date format	0,000618	1

TABLE VII.   EXAMPLE OF LABELING RESULTS

Comment	Post	Label	Reason
😣😣😣😣 😣💧💧💧 💧💧💧💧	Can't argue with the clan 😎! Entertainment Inc presents u @USER. Watch the full version on my video	Spam	The comment contains only emojis that are not consistent, "sad and hot" at the same time about the new post video
cantik bangettt😩🖤 (in English: veryyy pretty 😩🖤)	🌸 Outfits custom @USER Styling @USER Makeup @USER Hair @USER Photographer @USER	Not spam	The comment reply a post about how pretty an artist is because the post shows how beautiful the artist in an outfit

After the labeling had been completed and re-validated, a data profiling step was carried out to determine additional data from the dataset using Python NLP Profiler. It analyses whether there are emojis or not in the posts or comments. It does statistical analysis on the number of sentences, the number of characters, the number of whitespaces, the number of words, the number of words in the form of numbers, and the number of signs. It also reads and counts the number date

format. Data profiling is used to determine the characteristics of the data and assist in determining the appropriate pre-processing steps later.

C. Pre-processing

The pre-processing process consists of the following steps:

1) Generating manual features such as the length, the number of emojis, the number of unique emojis, the number of digits, the number of hashtags, the number of mentions, the number of uppercase letters, the number of special chars, and the number of links.

2) Changing letters to lowercase.

3) Removing spaces and characters that appear excessively.

4) Removing certain punctuation marks unrelated to hashtags, emails, mentions, and URLs.

5) Doing simple normalization as follow:

    a) Repeated words normalization (such as "pergi2" to "pergi-pergi").

    b) Slang words normalization using a dictionary.

    c) Email, hashtag, number, mention to specific TAG (USER, ANGKA, EMAIL, MENTION)

    d) Abbreviation normalization using a dictionary.

    e) Some minor spelling corrections using a dictionary.

6) Performing stopwords removal (using combined stopwords from standard and stopwords generated from the dataset based on their frequency).

7) Performing stemming using the Sastrawi Python library.

8) Saving the final output and passing it to the model.

IV. EXPERIMENTS AND DISCUSSION FOR BASELINE PERFORMANCE

The testing was carried out using the ML method (Nave Bayes, Complement Naïve Bayes, Decision Tree, and Multi-Layer Perceptron) [32], which was partially or fully implemented using the Python Scikit Learn library (Sklearn). The test scenario was carried out in two forms: a dataset with emoji in symbols and emoji in the text. Pre-processing uses tokenization, Indonesian stopwords, and simple normalization and uses the n-gram TF-IDF features, i.e., 1-gram and 2-gram. Table VIII shows the experiment scenario using the machine learning methods. The dataset splits into 80% training data and 20% testing data. The evaluation score used F-measure (F1) with a score between 0-1. The measurement matrix uses the F1 score with 80% training data and 20% testing.

The authors use Naïve Bayes (NB) with an alpha value of 0.01 and other parameters defaulted from sklearn. Complement Naïve Bayes (CNB) was used in the second experiment, which was expected to overcome datasets whose classes are not balanced. Both methods were used as representatives of the probability classifier. The Decision Tree (DT) method was also used, representing the classifier tree with the random_state parameter set and the information gain using the Gini index. Finally, an artificial neural network-based classification method was used: a multi-layer perceptron with a limited

iteration of 300. The F-measure (F1 score) was chosen for the performance evaluation because the F1 score value represents a combination of recall and precision values and can also be used in the unbalanced dataset.

Moreover, the accuracy value alone is inappropriate for the SPAMID-PAIR dataset with an unbalanced number of classes. Table X shows the results of the experimental scenarios using the methods. Fig. 4 to 10 display the confusion matrixes of the models, while Fig. 11 and 12 show ROC curves of the models in testing data. From the confusion matrix in Fig 4(a) and 4(b) (EmojiSymbol NB), all the true positives are higher than the others (true negative, false positive, and false negative). The ability to detect spam comments is good enough, but it also can be seen that the accuracy is better on not-spam comments than on spam comment labels. Fig. 5(a) and 5(b) (EmojiText NB) show that the F1 score is higher than the EmojiSymbol, although the true positives are lower than the EmojiSymbol. From this result, the EmojiText performs better because it can detect spam comment properly in a balanced dataset. Fig. 6(a) and 6(b) (EmojiSymbol CNB) show that the F1 score (based on true positive, true negative, false positive, and false negative) is better than the NB method. CNB works better because it can complement the weight of an unbalanced dataset [34]. Fig. 7(a) and 7(b) (EmojiText CNB) show that the CNB in text format outperforms the NB in EmojiSymbol and EmojiText. Fig. 8(a), 8(b), 9(a), and 9(b) shows the performance of the DT method that also has better F1 in EmojiText but not for the EmojiSymbol. Decision Tree can handle the emoji symbol well. The last, in Fig. 10(a) and 10(b), it can be seen that the confusion matrix shows that MLP (a traditional neural network) has close F1 score to NB and CNB but trains slower than them. But, based on Fig. 11(a) and 11(b), it can be seen that EmojiText in MLP works the best from the other methods.

The authors also extract a list of emojis categorized as 'spam' and 'not spam' based on the SPAMID-PAIR dataset. It can be seen in Table IX. It can be seen that list of spam emojis is more than not spam emojis. The intersection between them is also quite a lot, and the emoji only used in the "not spam" category contain very reasonable emojis (clear emoji meaning). Still, on the other hand, the emoji used only in the "spam" category is quite a lot and very random emojis (not clear emoji meaning).

TABLE VIII. THE TESTING SCENARIO ON SPAMID-PAIR USING THE MACHINE LEARNING METHODS

Test Scenario Using Machine Learning Methods		
Emoji Symbol	Pre-processing: tokenization, stopwords, normalization, stemming, feature: TF-IDF 1 gram and 2 gram	Methods: Naïve Bayes (NB) (alpha: 0.01), Complement Naïve Bayes (CNB) (alpha: 0.01, norm: true), Decision Tree (DT) (random_state: 42, gain: Gini), Multi-layer Perceptron (MLP) (random_state: 42, max_iter: 300)
Emoji in Text	Pre-processing: tokenization, stopwords, normalization, stemming, feature: TF-IDF 1 gram and 2 gram	Methods: Naïve Bayes (NB) (alpha: 0.01), Complement Naïve Bayes (CNB) (alpha: 0.01, norm: true), Decision Tree (DT) (random_state: 42, gain: Gini), Multi-Layer Perceptron (MLP) (random_state: 42, max_iter: 300)

TABLE IX.   EMOJI LIST (SPAM AND NOT SPAM), INTERSECTION, AND DIFFERENCE IN THE SPAMID-PAIR DATASET

List of Emojis	Category
(emoji list)	Spam (S)
(emoji list)	Not Spam (NS)
(emoji list)	S ∩ NS
(emoji list)	NS minus S
(emoji list)	S minus NS
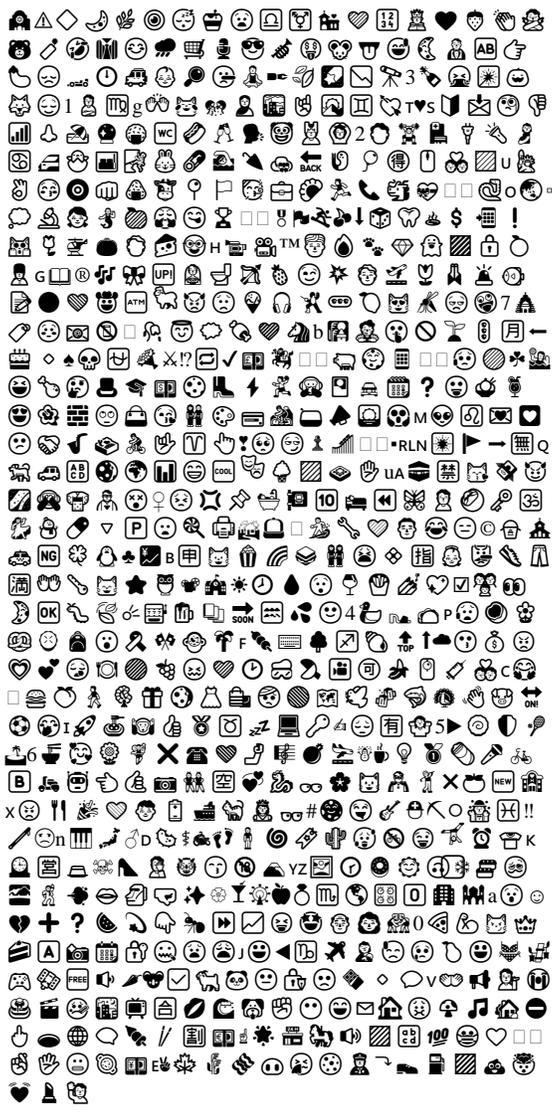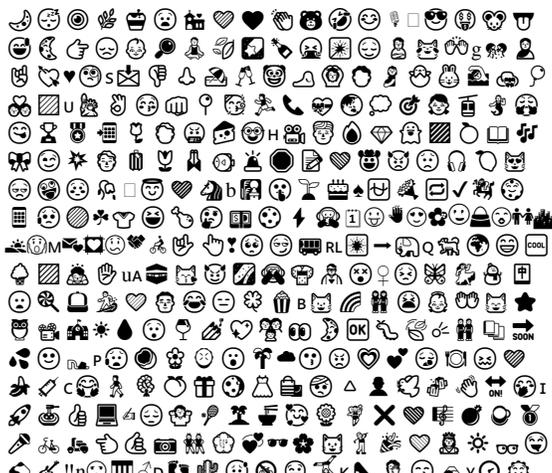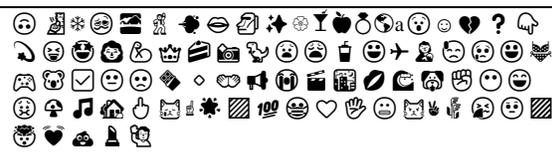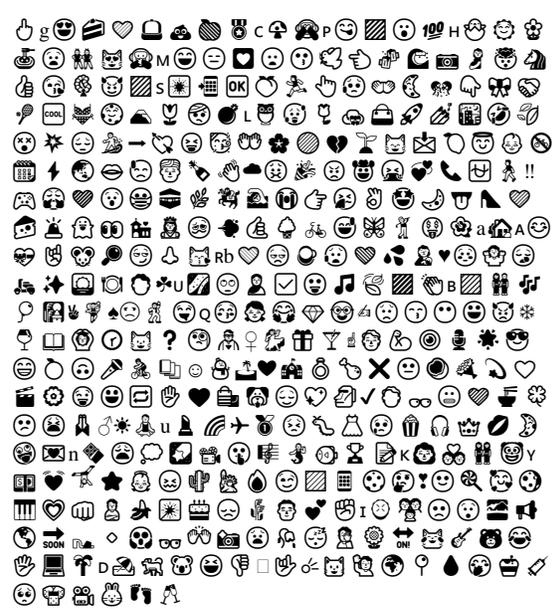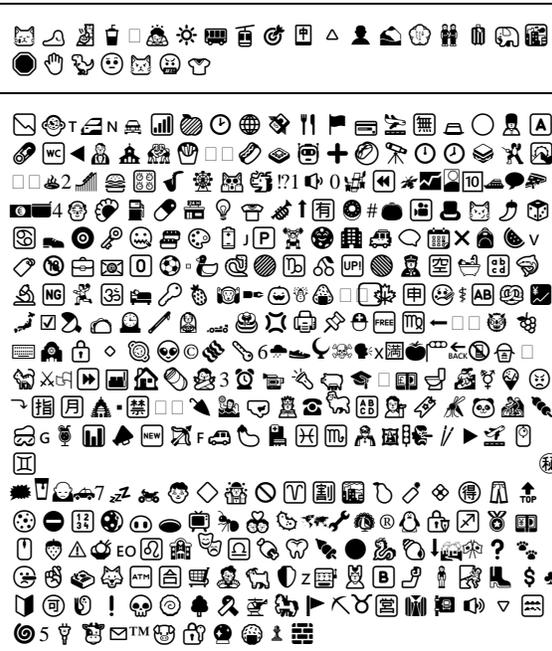
TABLE X.   F-MEASURE (F1) SCORE RESULTS OF THE EXPERIMENTAL SCENARIOS

Scenario	NB	CNB	DT	MLP
EmojiSymbol1GramTFIDF	.74	.75	.72	.74
EmojiSymbol2GramTFIDF	.74	.75	.72	.74
EmojiText1GramTFIDF	.77	.78	.78	.80
EmojiText2GramTFIDF	.78	.80	.78	.80
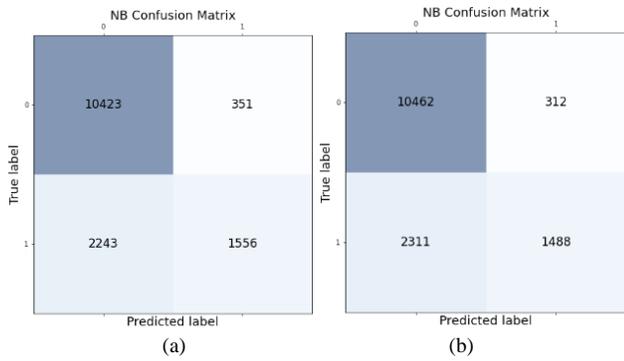
Fig. 4.   Confusion Matrix Naïve Bayes of (a) EmojiSymbol1GramTFIDF (b) EmojiSymbol2GramTFIDF.

Fig. 5.   Confusion Matrix Naïve Bayes of (a) EmojiText1GramTFIDF (b) EmojiText2GramTFIDF.

Fig. 6.   Confusion Matrix Complement Naïve Bayes of (a) EmojiSymbol1GramTFIDF (b) EmojiSymbol2GramTFIDF.

Fig. 7.   Confusion Matrix Complement Naïve Bayes of (a) EmojiText1GramTFIDF (b) EmojiText2GramTFIDF.
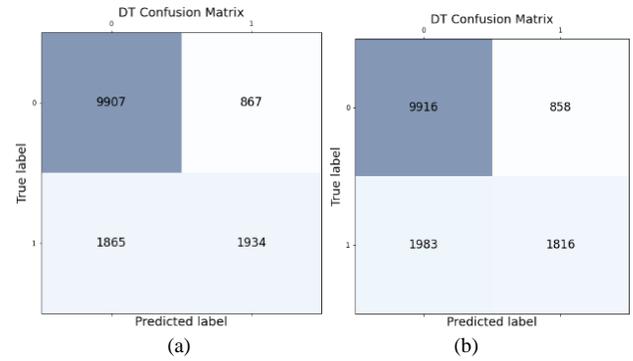
Fig. 8.   Confusion Matrix Decision Tree of (a) EmojiSymbol1GramTFIDF (b) EmojiSymbol2GramTFIDF.
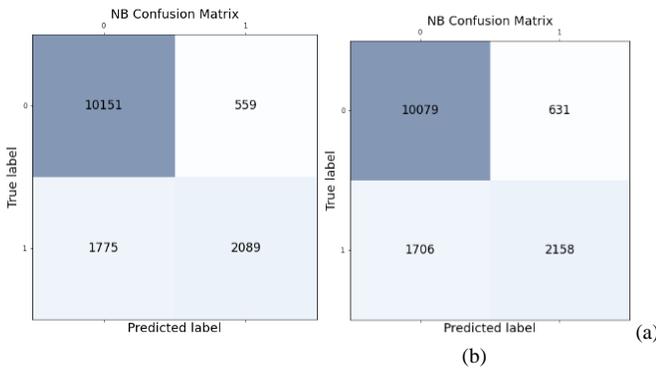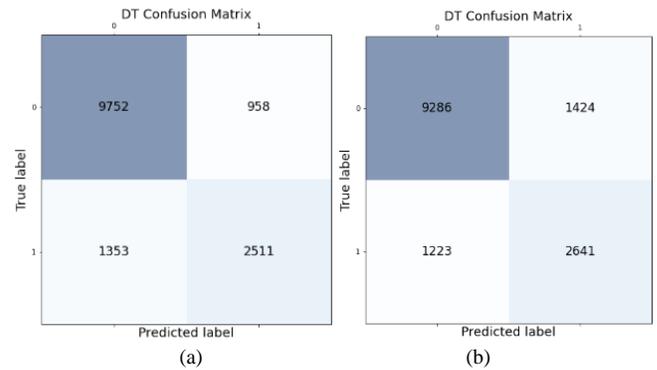
Fig. 9.   Confusion Matrix Decision Tree of (a) EmojiText1GramTFIDF (b) EmojiText2GramTFIDF
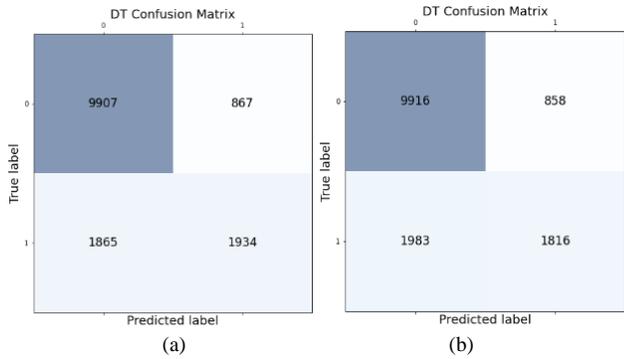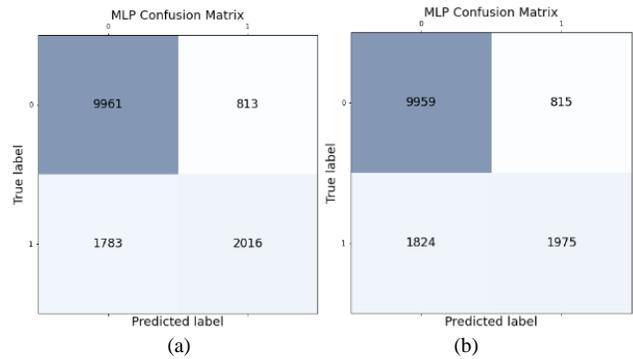
Fig. 10. Confusion Matrix Multi-layer Perceptron of (a) EmojiSymbol1GramTFIDF (b) EmojiSymbol2GramTFIDF
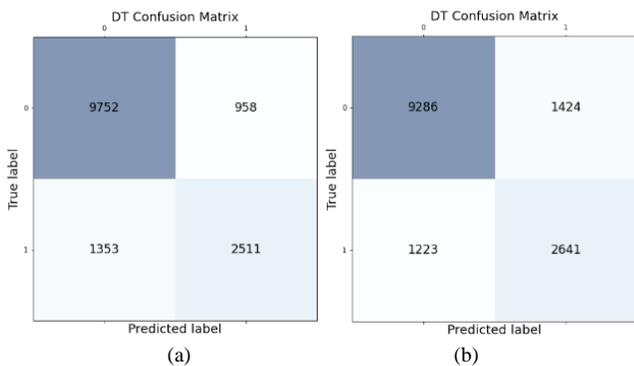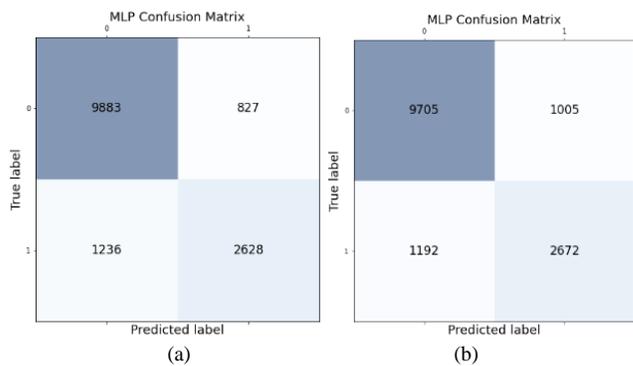
Fig. 11. Confusion Matrix Multi-layer Perceptron of (a) EmojiText1GramTFIDF (b) EmojiText2GramTFIDF
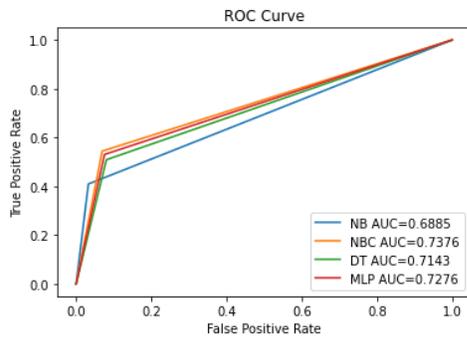
(a) ROC of EmojiSymbol1GramTFIDF.
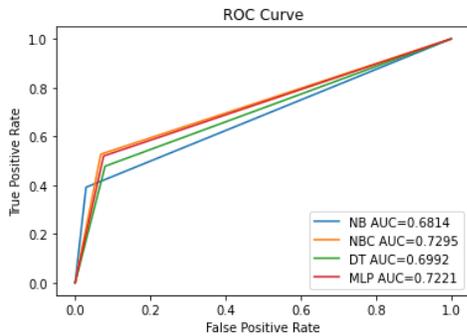
(b) ROC of EmojiSymbol2GramTFIDF.

Fig. 12. ROC Curve of (a) EmojiSymbol1GramTFIDF (NB, CNB, DT, MLP) and (b) EmojiSymbol2GramTFIDF (NB, CNB, DT, MLP).
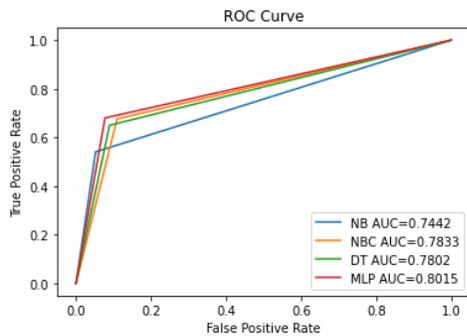
(a) ROC of EmojiText1GramTFIDF.
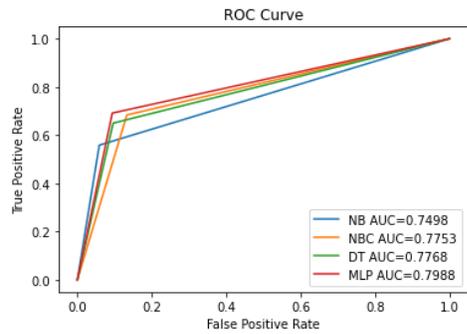
(b) ROC of EmojiText2GramTFIDF.

Fig. 13. ROC Curve of (a) EmojiText1GramTFIDF (NB, CNB, DT, MLP) and (b) EmojiText2GramTFIDF (NB, CNB, DT, MLP)

The results in Table X prove that the SPAMID-PAIR dataset is a dataset that can be used in Indonesian text classification experiments originating from social media. In Fig. 4 to Fig. 10, all the confusion matrixes of the models use 14.573 (20%) data testing. From Table X and Fig. 4 to Fig 12, It can be seen that CNB and MLP are superior to NB and DT. Fig. 13 shows the ROC curve, which explains that the area of the ROC curve in 13(a) is higher than in 13(b). The EmojiText1Gram is better than the EmojiText2Gram because the TFIDF vectors from 1gram have a better weight representing the text's characteristics. The traditional ML can only achieve an F1 score in the range of 0,72-0,78, but a multi-layer perceptron can achieve an F1 score of 0,8. It promises that these results can be improved, such as with the pair context classification approach [35]. Hopefully, this dataset can also be used in other related research and enrich the Indonesian dataset collection, which is still rare. This dataset is also important because it contains pairs of posts and comments that can be related and used in problem sentence pair classification in Indonesian.

V. CONCLUSION

This research collected post and comment pairs data from 13 selected Indonesian public figures (artists) / public accounts with more than 15 million followers. Two persons labeled all pair data as an expert in 72874 data. The dataset is called SPAMID-PAIR, containing post-comment pairs and label in Unicode text (UTF-8) text containing emojis. The dataset does not include any account information except the ID number. Unlike the other existing sentence pair datasets, the SPAMID-PAIR dataset is specifically used to determine the context between comments and posts that have never been collected in a large enough dataset. The objective of this dataset is as the primary data source in machine learning, especially in the NLP area, for spam comments detection based on the post context. This dataset is intended as one of the Indonesian language datasets that also contains many emoji symbols from social media so that it can be used to understand human expressions using emojis.

SPAMID-PAIR proved that it could be used as a training dataset to detect spam comments based on its post. From the experimental research using some ML methods, it can be seen that ML can only achieve an F1 score in the range of 0,72-0,78, but a multi-layer perceptron (MLP) can achieve an F1 score of 0,8. It significantly promises that these results can be improved in future works. The limitation of this dataset is it includes imbalanced data between not spam and spam categories. This dataset can also be enhanced in the future.

REFERENCES

[1] A. E. Yüksel, Y. A. Türkmen, A. Özgür, and A. B. Altınel, "Turkish tweet classification with transformer encoder," in International Conference Recent Advances in Natural Language Processing, RANLP, 2019, vol. 2019-Septe, pp. 1380–1387. doi: 10.26615/978-954-452-056-4_158.

[2]   T. Alam, A. Khan, and F. Alam, "Bangla Text Classification using Transformers," arXiv, Nov. 2020, [Online]. Available: http://arxiv.org/abs/2011.04446.

[3]   C. He and Y. Shi, "Research on Chinese spam comments detection based on Chinese characteristics," in 2018 IEEE 4th International Conference on Computer and Communications, ICCC 2018, 2018, pp. 2608–2612. doi: 10.1109/CompComm.2018.8781051.

[4]   A. M. Gaber, A. M. Gaber, and H. Moussa, "SMAD: Text Classification of Arabic Social Media Dataset for News Sources," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 10, pp. 508–516, 2021, doi: 10.14569/IJACSA.2021.0121058.

[5]   S. Mihi, B. A. BEN Ali, I. EL Bazi, S. Arezki, and N. Laachfoubi, "MSTD: Moroccan sentiment twitter dataset," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 10, pp. 363–372, 2020, doi: 10.14569/IJACSA.2020.0111045.

[6]   B. Wilie et al., "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," arXiv, Sep. 2020, [Online]. Available: https://www.aclweb.org/anthology/2020.aacl-main.85.

[7]   A. R. Chrismanto, A. K. Sari, and Y. Suyanto, "CRITICAL EVALUATION ON SPAM CONTENT DETECTION IN SOCIAL MEDIA," J. Theor. Appl. Inf. Technol., vol. 100, no. 8, pp. 2642–2667, 2022, [Online]. Available: http://www.jatit.org/volumes/Vol100No8/29Vol100No8.pdf.

[8]   C. Mus, "10+ Akun Instagram Dengan Followers Terbanyak Di Indonesia," musdeoranje.net, 2015. http://www.musdeoranje.net/2016/08/akun-instagram-dengan-followers-terbanyak-di-indonesia.html (accessed Oct. 13, 2021).

[9]   Allstars, "10 Artis Followers Terbanyak di Indonesia pada Instagram di 2021," Allstars.id, 2021. https://www.allstars.id/blog/2021/09/23/artis-indonesia-dengan-followers-instagram-terbanyak/ (accessed Oct. 27, 2021).

[10]   R. Wongso, F. A. Luwinda, B. C. Trisnajaya, O. Rusli, and Rudy, "News Article Text Classification in Indonesian Language," Procedia Comput. Sci., vol. 116, pp. 137–143, 2017, doi: 10.1016/j.procs.2017.10.039.

[11]   R. Ghanem and H. Erbay, "Context-dependent model for spam detection on social networks," SN Appl. Sci., vol. 2, no. 9, pp. 1–8, 2020, doi: 10.1007/s42452-020-03374-x.

[12]   M. Li, B. Wu, and Y. Wang, "Comment Spam Detection via Effective Features Combination," 2019. doi: 10.1109/ICC.2019.8761340.

[13]   X. Ban, C. Chen, S. Liu, Y. Wang, and J. Zhang, "Deep-learnt features for Twitter spam detection," 2018 Int. Symp. Secur. Priv. Soc. Networks Big Data, Soc. 2018, pp. 22–26, 2018, doi: 10.1109/SocialSec.2018.8760377.

[14]   G. Jain, M. Sharma, and B. Agarwal, "Spam detection in social media using convolutional and long short term memory neural network," Ann. Math. Artif. Intell., vol. 85, no. 1, pp. 21–44, 2019, doi: 10.1007/s10472-018-9612-z.

[15]   N. M. Samsudin, C. F. B. Mohd Foozy, N. Alias, P. Shamala, N. F. Othman, and W. I. S. Wan Din, "Youtube spam detection framework using naïve bayes and logistic regression," Indones. J. Electr. Eng. Comput. Sci., vol. 14, no. 3, pp. 1508–1517, 2019, doi: 10.11591/ijeecs.v14.i3.pp1508-1517.

[16]   S. Aiyar and N. P. Shetty, "N-Gram Assisted Youtube Spam Comment Detection," Procedia Comput. Sci., vol. 132, pp. 174–182, 2018, doi: 10.1016/j.procs.2018.05.181.

[17]   N. Alias, C. F. M. Foozy, and S. N. Ramli, "Video spam comment features selection using machine learning techniques," Indones. J. Electr. Eng. Comput. Sci., vol. 15, no. 2, pp. 1046–1053, 2019, doi: 10.11591/ijeecs.v15.i2.pp1046-1053.

[18]   R. Abinaya, E. Bertilla Niveda, and P. Naveen, "Spam detection on social media platforms," 2020 7th Int. Conf. Smart Struct. Syst. ICSSS 2020, pp. 31–33, 2020, doi: 10.1109/ICSSS49621.2020.9201948.

[19]   IMDB, "IMDb Datasets," IMDb Datasets, 2022. https://www.imdb.com/interfaces/ (accessed Sep. 30, 2022).

[20]   J. Jang, Y. Kim, K. Choi, and S. Suh, "Sequential Targeting: an incremental learning approach for data imbalance in text classification," 2020, [Online]. Available: http://arxiv.org/abs/2011.10216.

[21]   V. Narayanan, I. Arora, and a Bhatia, "Fast and accurate sentiment classification using an enhanced Naive Bayes model," Int. Data Eng. Autom. Learn. Lect. Notes Comput. Sci., vol. 8206, pp. 194–201, 2013, doi: 10.1007/978-3-642-41278-3_24.

[22]   T. A. Almeida and H. JosÃ, "SMS Spam Collection Data Set," UCI Machine Learning Repository, 2012. https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection (accessed Sep. 30, 2022).

[23]   A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf, "FLAIR: An easy-to-use framework for state-of-the-art NLP," NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Demonstr. Sess., pp. 54–59, 2019.

[24]   L. Zhang and D. Moldovan, "Classification of semantic relations between pairs of nominals using transfer learning," Proc. 32nd Int. Florida Artif. Intell. Res. Soc. Conf. FLAIRS 2019, pp. 92–97, 2019.

[25]   NSCLab, "Twitter Spam," 2014. http://nsclab.org/nsclab/resources/ (accessed Sep. 30, 2022).

[26]   C. Chen, J. Zhang, X. Chen, Y. Xiang, and W. Zhou, "6 million spam tweets: A large ground truth for timely Twitter spam detection," IEEE Int. Conf. Commun., vol. 2015-September, pp. 7065–7070, Sep. 2015, doi: 10.1109/ICC.2015.7249453.

[27]   D. Kang et al., "A Dataset of Peer Reviews (PeerRead): Collection, Insights and NLP Applications," in Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2018, vol. 1, pp. 1647–1661. doi: 10.18653/v1/N18-1149.

[28]   Q. Lhoest et al., "Datasets: A Community Library for Natural Language Processing," in Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2021, pp. 175–184. doi: 10.18653/v1/2021.emnlp-demo.21.

[29]   F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in Proceedings of the 28th International Conference on Computational Linguistics, 2020, pp. 757–770. doi: 10.18653/v1/2020.coling-main.66.

[30]   A. A. Septiandri and O. Wibisono, "Detecting spam comments on Indonesia's Instagram posts," J. Phys. Conf. Ser., vol. 801, no. 012069, pp. 1–7, 2017, doi: 10.1088/1742-6596/755/1/011001.

[31]   W. Zhang and H.-M. Sun, "Instagram Spam Detection," in 2017 IEEE 22nd Pacific Rim International Symposium on Dependable Computing (PRDC), Jan. 2017, pp. 227–228. doi: 10.1109/PRDC.2017.43.

[32]   C. Zhang, C. Liu, X. Zhang, and G. Almpanidis, "An up-to-date comparison of state-of-the-art classification algorithms," Expert Syst. Appl., vol. 82, pp. 128–150, 2017, doi: 10.1016/j.eswa.2017.04.003.

[33]   B. Priyoko and A. Yaqin, "Implementation of naive bayes algorithm for spam comments classification on Instagram," in 2019 International Conference on Information and Communications Technology, ICOIACT 2019, 2019, pp. 508–513. doi: 10.1109/ICOIACT46704.2019.8938575.

[34]   N. A. Haqimi, N. Rokhman, and S. Priyanta, "Detection Of Spam Comments On Instagram Using Complementary Naïve Bayes," IJCCS (Indonesian J. Comput. Cybern. Syst., vol. 13, no. 3, p. 263, Jul. 2019, doi: 10.22146/ijccs.47046.

[35]   R. Yang, J. Zhang, X. Gao, F. Ji, and H. Chen, "Simple and Effective Text Matching with Richer Alignment Features," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 4699–4709. doi: 10.18653/v1/P19-1465.

Cedarwood Quality Classification using SVM Classifier and Convolutional Neural Network (CNN)

Muhammad Ary Murti[1], Casi Setianingsih[2], Eka Kusumawardhani[3], Renal Farhan[4]

School of Electrical Engineering, Telkom University, Bandung, Indonesia[1, 2, 4]
Department of Electrical Engineering, Universitas Tanjungpura, Pontianak, Indonesia[3]

Abstract—**Cedarwood is one of the most sought-after materials since it can be used to create a wide variety of household appliances. Other than its unique aroma, the product's quality is the most important selling attribute. Fiber patterns allow for a qualitative categorization of this wood. Traditionally, workers in the wood-processing business have relied solely on their eyesight to sort materials into several categories. As a result, there will be discrepancies in precision and efficiency, which will hurt the reputation of the regional wood sector. The answer to this issue is machine learning. In this study, we compare the performance of two different cedarwood quality classification systems where both systems use different machine learning methods namely Support Vector Machine (SVM) and Convolutional Neural Network (CNN). Each system will be sent images captured with a Logitech Brio 4K equipped with a joystick and ultrasonic sensors, labeled as belonging to one of five cedar classes (A, B, C, D, or E). In the initial method to learn the wood's pattern and texture, the Histogram of Oriented Gradient (HOG) will be used to identify the material. Meanwhile, the classification method uses a Support Vector Machine (SVM) which will be compared to find the best accuracy and time computation. The first system's experiment achieves 90 percent accuracy with a computation time of 1.40 seconds. For the second, we use a Convolutional Neural Network, a deep learning technique, to classify cedarwood (CNN). Extraction of features occurs in the convolution, activation, and pooling layers. Experimental results demonstrated a considerable enhancement, with an accuracy of 97% and a prediction speed of 0.56 seconds.**

Keywords—Cedarwood classification; convolutional neural network (CNN); HoG feature; SVM classification

I. INTRODUCTION

When it comes to exterior, interior, and home appliance needs, cedarwood is one of the most popularly purchased light wood processing materials. This is not without good reason, the resin content contained in Cedarwood will emit a very fragrant odor, the goal of which is to expel the terms that will eat the wood composition [1]. Cedarwood also has a very high level of resistance to mold because it belongs to the group of hardwoods which incidentally has dense pores so that moist water does not easily enter and cause mold. That is why the quality of cedar is the main point of concern [2]. Like wood in general, cedarwood also has a variety of qualities depending on the factors of sawing and the age of the tree, so these quality factors can be classified directly by paying attention to the color, texture, and pattern of the fiber [3], [4].

Many businesses that deal with processed wood still perform classification by hand, using just their eyes and gut

reactions to determine how many items are alike. Compliance with observations is only indicated as a percentage (50-60%) in that matter [5]. Therefore, it is challenging to avoid the issue of ambiguity and misclassification within each category. The legitimacy of Indonesia's processed wood sector is threatened by subjective standards of accuracy and production time [6]. In light of this, the field of machine learning offers a means of addressing and ultimately resolving these issues. Local wood industries can use machine learning to create uniform wood classes, allowing for reliable automatic classification [7].

The implementation of machine learning to create uniform wood classes, it can use two deep learning approaches into the cedarwood quality categorization system. Support Vector Machine (SVM) and Convolutional Neural Network (CNN) will be utilized to automatically capture and predict when wood enters the categorization area. These approaches will be directly integrated with the small conveyor, Controller, and Logitech Brio 4K webcam. Cedarwood quality data from Classes A through E is employed, and an automated feature extraction process is carried out with a total of 16 layers in CNN and in SVM. The Histogram of Oriented Gradient is then used to determine the grain pattern and texture of the wood during recognition.

II. RELATED WORK

Widespread use of machine learning theory emerged during the era of the 4.0 Industrial Revolution, which also paid close attention to technical advancements. As a result, its implementation in the wood industry is quite probable [8] [9]. Over the years, many types of wood classification systems have been developed, such as wood classification based on PCA, 2DPCA, (2D) PCA and LDA [10], the second method is the Classification of Coconut Wood Quality based on Gray Level Co-Occurrence matric [11], and the third method is the Species Identification of Wooden Material using Convolutional Neural Network [12]. However, in the first method, the highest implementation accuracy was only 86.67% and even then, it had to be tested on a different sample, besides the feature extraction and identification process requires a good dataset structure to obtain a good system performance. In the second method Based on Gray Level Co-Occurrence Matrices, the highest accuracy obtained is only 78.82% so that it still requires a feature selection strategy to improve its performance. Then in the third method which classifies fir, pine, and fir wood with various CNN architectures used, the highest accuracy is 99.4%. However, this classification method can only identify the type of wood in certain parts and cannot make the identification of the whole wood being tested if it is

done in real-time, so pre-processing is still needed to perfect the research.

III. RESEARCH METHODOLOGY

In this research, the CNN and SVM-based cedarwood classification system is broken down into three distinct phases: Acquisition, training, and categorization of images, as well as indication systems. In the first phase, an ultrasonic sensor is used to identify the presence of wood on the conveyor path, and the resulting input signal is used for auto-capture. The Controller relays the signal to the server using serial communication. When the server receives the command, the Logitech brio 4k webcam is activated to take a picture for use in the training and classification process. The second stage is responsible for machine learning and categorizing the collected images through the use of the convolution, activation, and max-pooling layers.

An indicator system made up of five LED lights of varying colors, each of which represents a different category of wood, shows the results of the second stage of learning. The following are the three working states of the cedarwood quality classification system illustrated in Fig. 1.

All process of sending commands between the device controller and the server is done using the serial communication method with a data rate of 8 bits per second. The detection and prediction code that the system has processed must first be converted into binary forms so that the data transmission process can be easily carried out in series.

A. Image Acquisition Process

Image acquisition is one of the subsystems directly related to the process of automatically capturing images when scanning wood. The sensor is placed horizontally next to the direction of the conveyor belt rate to detect wood when it passes by. When wood is detected, the capturing image will be performed by a webcam mounted in the sky of the acquisition box. This section functions as a data gateway between hardware and software on the server.
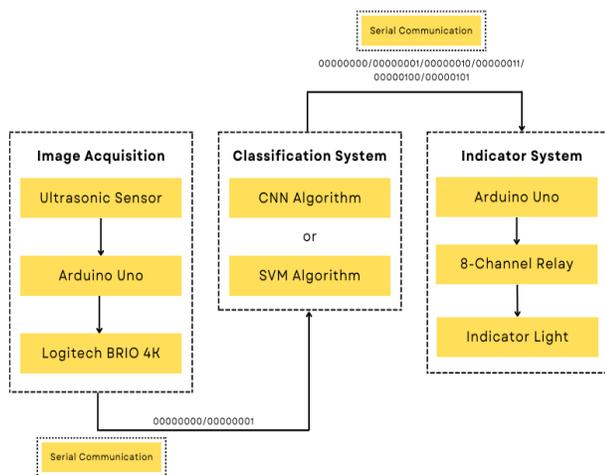
Fig. 1. Block Diagram of Classification System.
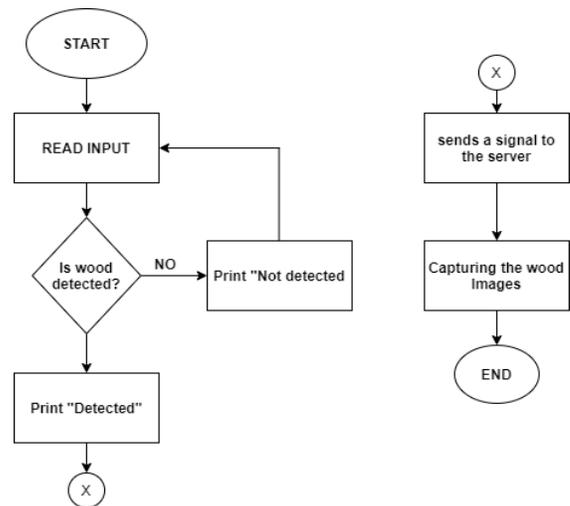
Fig. 2. Flowchart of Images Acquisition Process.

Fig. 2 is a flowchart diagram of the image acquisition process used to give the camera commands to carry out image capture tasks that are integrated between the Controller and the server with serial communication mode. Upon receiving a signal from the HC-SR04 proximity sensor that is incorporated with the Controller, the server will instruct the Logitech brio 4K to automatically capture the cedarwood runs on the conveyor track. The image is captured in a black, matte, 27.6 x 23.6 x 30 cm acquisition box lit from above by LED strips emitting 723 lux. The resulting image has a 640x480 pixel resolution and the following features of the wood used.

Images of cedarwood captured at a zoomed-in percentage of 130%, with a black backdrop, brightness of 26, Contrast of 24, and camera focus of 14 are displayed in Table I.

TABLE I. CHARACTERISTICS OF CEDARWOOD IN EACH CLASS

Class	Fig	Description
A		Magnificently straight fiber, with a distance so short that it is barely perceptible, is the hallmark of this technology.
B		There are a lot of fibers, they're all straight and rather thick, and you can see the separation between them.
C		The space between the fibers is considerable, and there are also some transverse fibers visible.
D		Fiber-to-fiber spacing and transverse pattern are easily visible.
E		Extremely transparent fiber; random fiber patterns are typically not straight.

B. Training and Classification Process

In this section, the image has been baptized and stored on the server, then pre-processing is done to match the

classification architecture used. After a match between the image's dimensions and architecture, the image classification process is carried out using the CNN and SVM methods.

1) Convolutional Neural Network (CNN) Method

a) Feature Extraction: Fig. 3 explains the data processing up to sending commands to the microcontroller. The left is a classification process of wood images captured and stored in a server. On the right is a training process for the Convolutional Neural Network method to get the best classification model used in the system's classification. The first layer of the convolutional neural network design will be fed with the image acquired during the acquisition phase. But first, it needs to go through some pre-processing to ensure that the image and layer have compatible dimensions [13]. The CNN architecture VGG-16 technique is used in the wood classification system. VGG16 is used because it's considered a type of CNN, one of the most advanced computer vision models. VGG16 is an object detection and classification algorithm that can classify 1000 images of 1000 different categories with 92.7% accuracy. It is one of the most widely used algorithms for image classification and is easy to use with transfer learning. It uses a square picture with dimensions of 224 by 224 by 3 pixels as the suitability parameter. Up until the point of delivering commands to the microcontroller, the data processing is described in detail. The left is a classification process of wood images captured and stored in a server. On the right is a training process for the Convolutional Neural Network method to get the best classification model that will be used in the classification of the system. The acquired image will be used as input for the CNN architecture's first layer. However, prior to doing that, the image must be pre-processed so that its dimensions match those of the layer.
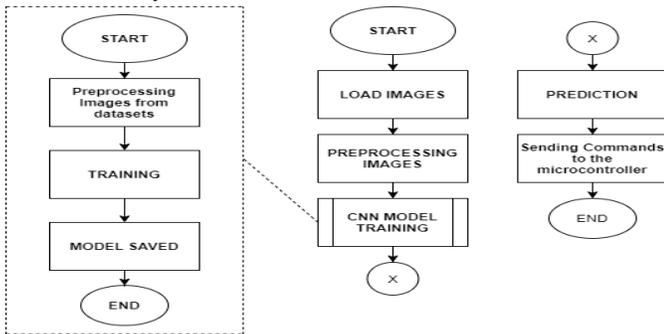
Fig. 3. Flowchart of Training and Classification Process using CNN Method.

In the CNN architecture VGG-16 algorithm [14] employed in the wood classification system, the parameter of appropriateness is a square-shaped image with dimensions of 224 x 224 x 3 pixels.

$$Rz = \frac{P(q,r) + P(q+1,r) + p(q,r+1) + p(q+1,r+1)}{4} \quad (1)$$

Based on Eq. 1, the image will be resized to be smaller, as in Fig. 4. However, so that the original rectangular image can be resized into a new rectangular image, it is necessary to do an additional process by improving the image's aspect ratio or adding padding to the new image to be resized. The step for resizing the image will affect the accuracy of the classification [15]. Also Adding padding can be done in three ways [16]:

- Inserting a zero value at the edge of the dimension.

- A max value at the edge of the dimension.

- Inserting the same value as the image at the edge of the new dimension.
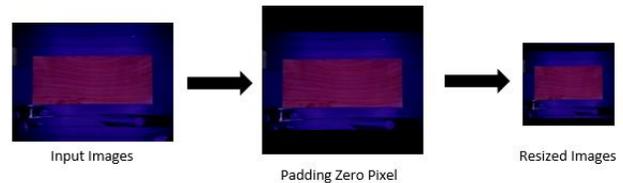
Fig. 4. Resizing Rectangle into Square Images.

A feature extraction technique is carried out to capture features in each wood class after the input image is found to be a match for the first layer of a convolutional neural network. The 2D convolution operation comes at the beginning of the feature extraction procedure. The convolution layer is the initial feature-extracting layer in the CNN architecture, and it does so by convolving the input of a square-resolution picture with a filter, also known as a kernel [17]. In a broad sense, we might say that this 2D convolution procedure uses the sliding window idea to calculate its weight [18].

$$(y, x) = \sum_{p=-m2}^{m2} \sum_{q=-n2}^{n2} h(p + m2 + 1, q + n2 + 1) f(y - p, x - q) \quad (2)$$

If we apply the above equation, we get the following representation of the convolution's final image:
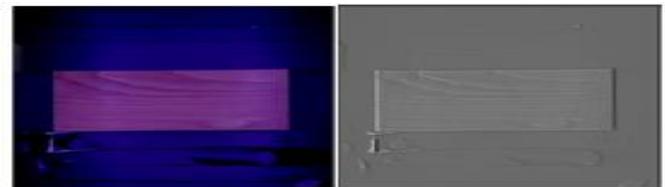
Fig. 5. Visualization of Convolution Layer.

One of the 3 x 3 pixel filters/kernels used in the convolution layer's 2D convolution operations is shown in color in Fig. 5. The study found that 64 features might be used on the first convolution layer when using input dimensions of $224 \times 224 \times 3$. Then, the convolutional layer's output is turned on, which is the second part of the feature extraction procedure. One component of CNN architecture, the activation layer, can prevent vanishing gradients and speed up the convergence process during training [19]. As an activation function, ReLu will swap out negative pixel outputs for zeros without touching the original pixel values [20]. The corresponding mathematical expressions are as follows:

$$ReLu = max\,(y, 0) \quad (3)$$

A new image will be created that is shown in Fig. 6 by plugging the above algorithm into the convolutional layer's output.

Fig. 6. Visibility of the Activation Layer.

Most of the pixels in the final image produced by the above activation are zero, and hence the image is predominantly black. It will shed lighter on the characteristics and make it easier to calculate data for use in either training or testing. Using a pooling layer as the last stage in feature extraction. The pooling layer is the final stage of feature extraction, and its job is to select the most informative aspects of the activation function at the output layer [21]. It aims to simplify features and accelerate computing time. The challenge is to implement the equation to do down sampling on each pixel of the image:

$$Pooling = max(y, 0) \qquad (4)$$

Feature clarity can be achieved using the down sampling method, as seen in Fig. 7. Reduce the size of individual pixels to reduce the time needed to classify data in the fully linked layer. The Fully Connected Layer is a Multilayer Perceptron (MLP) that makes use of the softmax activation function.
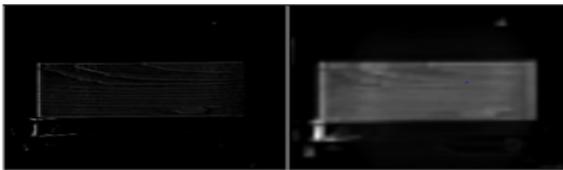
Fig. 7. Visualization of Pooling Layer.

b) Training Process: In this stage, the final probability is computed using the extracted feature. Fully connected layer is an MLP that uses the softmax activation function. It is the responsibility of this layer to take the extracted feature and process it so that the final probability is obtained. This layer also does categorization by selecting the weighted product of the highest value of the flattened results. Using the following simple equation, weight values will be refined until the fewest losses are incurred:

$$W'^l_{n,m} = W^l_{n,m} - \eta \frac{\partial L}{\partial W^l_{n,m}} \qquad (5)$$

As such, a loss value of zero represents the best possible categorization likelihood. As used here, "weight" refers to a property of the objects being categorized. Once losses are zero, the following computation will be used to discontinue updating weight:

$$Loss\ Function = \sum_{k=1}^{n}(Y - \overline{Y})^2 \qquad (6)$$

c) Classification Process: For the record, the multiclass system relies on the Softmax Classifier Function to properly label each class before making a final prediction. One type of function activation is the Softmax classifier. In a multiclass classification problem, the softmax function must be used in the output layer [22]. It's utilized in MLP, and its values range from zero to one, with the overall probability adding up to one [23]. It is undoubtedly essential to evaluate a class prediction

on a machine-learning model because, with a softmax classifier, users will quickly know the level of confidence in the system to make predictions [24]. Sometimes the system predicts classes correctly with a confidence level below 50%, while the system predicts wood classes correctly with a confidence level above 50%. Every chance of a multiclass event occurring with a value of one is expressed in the following equation.

$$fj(y) = \frac{e^{y_i}}{\sum_k e^{y_k}} \qquad (7)$$

For each jth member in the class output vector, the function's result is displayed using the fj notation. For classification purposes, the training model provides the hypothesis y as the argument to the Softmax function.

2) Support Vector Machine (SVM) Method

a) Pre-Processing: During the SVM classification procedure, the acquired image will be converted to grayscale and scaled to 427 x 240 pixels. Several tests utilizing three image sizes, including the original size of 1280 x 720 pixels, 640 x 360 pixels, 427 x 240 pixels, and 320 x 180 pixels, are required. The results demonstrated that the 427 x 240 pixels have an accuracy of 100 percent and a computation time of 3.08 seconds. After determining the optimal size for image capture, the photographs will be converted to grayscale. The straightforward equation for converting RGB to grayscale is as follows [25], [26]:

$$Gray = (Red * 0.3 + Green * 0.59 + Blue * 0.11) \qquad (8)$$

With this equation, the image could be transformed into grayscale (see Fig. 8):

Fig. 8. Grayscale Conversion Image.

b) Feature Extraction: HOG features are the next step. Here, we apply 1-D centering to determine the point of the discrete derivative mask in the x and y directions to compute the gradient value [24] (see Eq. 9 and 10).

$$L_x(r, c) = l(r, c + 1) - l(r, c - 1) \qquad (9)$$

$$L_y(r, c) = l(r, c + 1) - l(r, c - 1) \qquad (10)$$

The equation for the gradient's magnitude is as follows:

$$\mu = \sqrt{lx^2 + ly^2} \qquad (11)$$

Furthermore, gradient orientation is given by:

$$\theta = \tan^{-1} \frac{ly^2}{lx^2} \qquad (12)$$

Bin grouping based on spatial orientation is the next step. The goal of this procedure is totally votes and provide a histogram of cellular activity. To determine the image's overall orientation, each pixel will be assigned to the nearest bin between 0 and 180 degrees. As a next step, the HOG descriptor is used to convert the cellular and histogram data to a vector

space normalization. Using the L2-Hys norm, the block normalization is carried out as follows as the final step:

$$b = \frac{b}{\sqrt{|b|^2 + \varepsilon^2}} \qquad (13)$$

The limit of renormalization is b = 0.2. The Feature used to implement this HOG is depicted in Fig. 9.
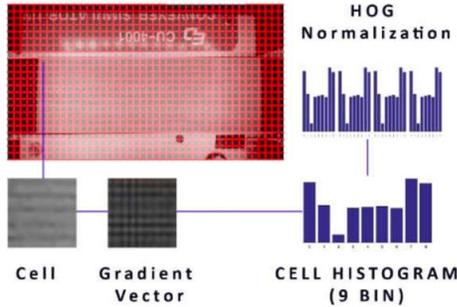
Fig. 9.   Feature Process of HOG to Identify the Wood.

For this experiment, we will partition the detection window of 427 × 240 pixels into 966 chunks, 42 in width and 23 in height. There are four cells in each block, and each cell's histogram has 36 values spread out across nine bins. This increases the total size of the vector to 42 blocks across by 15 blocks vertically by 4 cells/block by 9 bins/histogram, or 22,680 values. The *.csv file format can take this final vector as input.

c) Classification Process

As a final stage, we employ a Support Vector Machine. The goal of this process is to assign a category to each photograph. In addition, SVMs enhance the picture of each class by employing hyperplanes or discriminatory boundaries. The margin at the maximal hyperplane point is used to choose the optimal hyperplane. We can express the linear classification hyperplane as [27]:

$$f_{svm}(x) = \sum i \in N a_i y_i K(x_i, x) + b \qquad (14)$$

$$class = \begin{cases} 1, & f_{svm}(x) \geq 0 \\ -1, & others \end{cases} \qquad (15)$$

C. Indicator System

This section is the last working state of the system that will be created, where the Command that comes from the classification results becomes an input to the microcontroller to function the installed plant, namely the indicator lights with five different colors.

The flowchart above in Fig. 10 shows the last working state in the wood classification system starting from receiving commands sent from the server by the serial communication method to the microcontroller, which will give the state high to each lamp with a relay intermediary. When the prediction results state that the test wood is class A wood, the server will give an Arduino command with the serial communication method to contact the relay so that it becomes NC and makes the red light turn on. Likewise, on the results of other conditions, a similar process will be carried out according to

the predicted wood class— LED lights are employed as the indicator and will light up in response to data from the server. Signals will be transmitted via serial communication to an Arduino that will then activate relays and LED lights as expected. Signals that are transmitted as coded signals based on the classification of the wood, with the following stipulations: Light color codes: "A" for red, "B" for yellow, "C" for white, "D" for green, and "E" for blue.
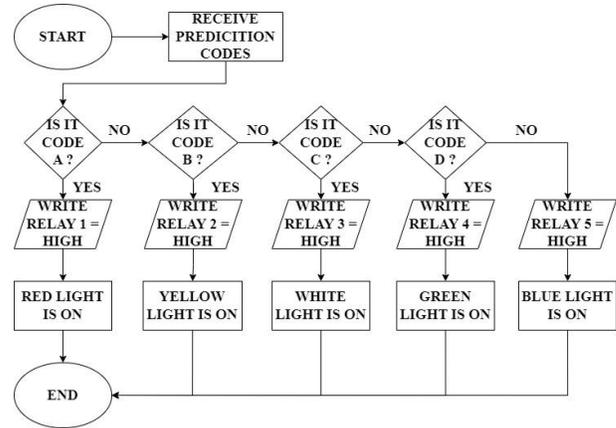
Fig. 10.   Flowchart of Indicator System.

IV.   RESULT AND DISCUSSION

The algorithm utilized by the cedarwood classification system is a CNN and SVM written in Python with the Open CV library. Up to 50 boards of the cedarwood utilized as testing material were examined. The dimensions of the wood for classes A, B, and D are 18.3x6.2x0.45 cm, while the measurements for classes C and E are 18.3x7.6x0.45 cm. The cedarwood training dataset reached 2,830 instances due to data enhancement using a condition that involves rotating through 180 degrees and flipping under the conveyor condition. This section provides an overview of the system. Fig. 11 is the overall design of the system that implemented. The system consists of nine pieces of hardware use like conveyors, ultrasonic sensors, laptop that use for server, Arduino uno, image acquisition boxes, webcams, led strips, indicator lights, and 8 channel of relays.
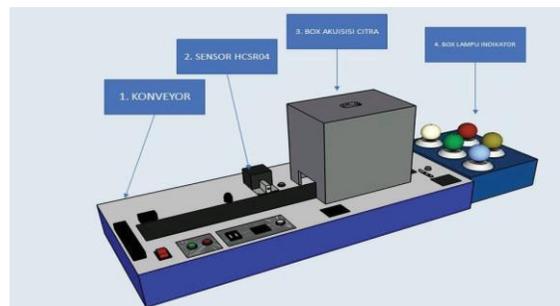
Fig. 11.   Integrated System with Plant.

A. Observation of Data Distribution Effect Result

Observation of data distribution is done to find the most optimal model when the training process on the computer occurs. The augmented dataset is divided into four distribution variations which are illustrated in Table II.

TABLE II.   DATA DISTRIBUTION VARIATIONS

Data Distribution	Training Data	Test Data	Total Datasets
60 : 40	1710	1140	2850
70 : 30	1995	855	2850
80 : 20	2280	570	2850
90 : 10	2565	285	2850

The goal of the data distribution experiment is to evaluate the system's performance in terms of the distribution of the input data using wood from datasets that were chosen at random depending on each class, the experiment was run 20 times, ten times from the front view and ten times from the back perspective, with the following findings from both CNN and SVM.

The experiment result shown in Fig. 12 shows that the 90:10 dataset experiment is the best accuracy in both SVM and CNN methods. It indicates that as the proportion of training data increases, the system will generalize more object attributes of each class, also data variability is more important than the testing data, and identification will be better when the training dataset is more significant than the testing data. So, the 90:10 dataset will be used as the parameter for the next following experiment of both methods.
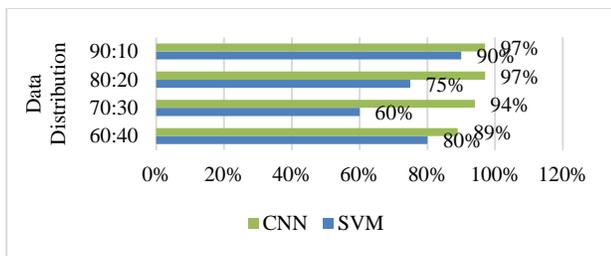
Fig. 12. Data Distribution Experiment Result.

B. Experimental Result of System using SVM Method

There are three experiments carried out to know the system's performance using the SVM method:

- The test angle for the camera position.
- The parameters for the HoG test.
- The parameters for the SVM test.

The first experiment involves changing the angle of the camera position test. The goal of this experiment was to find the best camera angle for this system. On the camera's surface, five angle sites will be established: Five degrees to the right, ten degrees to the right, zero degrees, five degrees to the left, and ten degrees to the left. As a result, the findings are shown in Table III.

Table III demonstrates that a 90 percent success rate can be attained with a camera positioned at a zero-degree angle. Similarly, time computation is important but not always a constrained-range issue. The conveyor belt serves as the background for the training dataset consisting of wood fibers. It will be impacted by the system for it to detect the things. Predicated on the work of Sahki et al. [28], His article concerning HOG-based Fast person detection utilized a two-folder training database from the INRIA pedestrian database.

Each folder contains both "post" and "neg" photos. Negative training or test photos, such as a park, ocean, cityscape, etc., are included in the neg post, while the positive training or test photographs centered on a person with their left-right reflections are included in the post. These conditions may facilitate recognition of the object in a different position inside the image. It might be concluded that 0 degrees Celsius will be employed in this experiment.

TABLE III.   OVERALL RESULT OF ANGLE CAMERA EXPERIMENT

	Right Side		0°	Left Side	
	+10°	+5°		+5°	+10°
Accuracy	0	30	90	40	10
Duration (s)	1.61	1.45	1.58	1.61	1.61

The second experiment is a test of HOG parameters to determine the optimal orientation, block norm, pixel/pixel, and block/pixel parameters. Fig. 13 and 14 depict the outcome of the experiment involving the modification of each HOG indicator. There are several methods for computing precision and time. The ideal parameter for this test is 100% with a computation length of 2 seconds, pixel/cell 10 x 10, and cell block 2 x 2. The cells/block indicator shows that each 2x2 is more essential than a 4 x 4. That data points to a 2 x 2 layout being preferable to a 4 x 4 one in this system. This result is shown in Fig. 13.
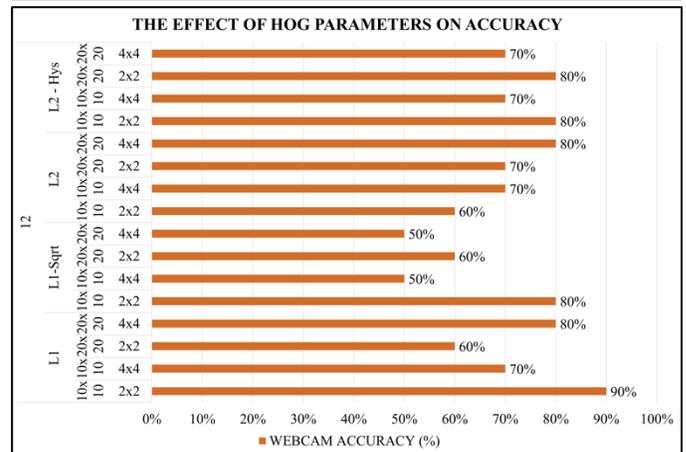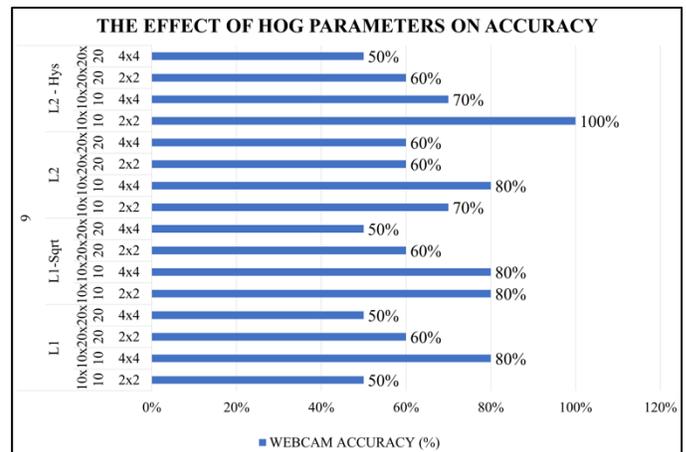
Fig. 13. The Accuracy Rate of HOG Feature (%).

In addition, for the pixel-per-pixel indicator, 10x10 is superior to 20 x 20, indicating that the gradient produces greater detail and precision. In addition, the Block value prevents images from having excessive brightness and contrast. This technique is ideal for controlling cells per 2-by-2 block and pixels every 10-by-10 pixel. Next, the indicator for the block norm has a variable value; it has not yet been affected. Block normalization minimizes brightness and contrast variations in an image depending on neighboring gradient cells. By dividing each vector element by its vector length, it aims to preserve the image's luminance and contrast [29]. The result is shown in Fig. 14.

Last is orientation; the average accuracy of the nine orientation values is 66% for a total of nine values. It indicates that the accuracy is less than 12 orientations. The 12 orientation value has an average performance of 70%. Therefore, if the system employs a greater orientation value, the optimum analysis for picture detail and gradient value will be compromised. It is possible to determine that orientation 9, block norm L2-Hys, pixel/cell 10 x 10, and cell block 2 x 2 are optimal for this system and provide the highest performance rate.

kernel type parameters. The outcome of this experiment is depicted in Fig. 15 and 16.

Fig. 15 and 16 shows that the linear kernel with a one-versus-rest multiclass produced the best performance rate at 1.40 seconds for 90% accuracy. One-to-one accuracy in the radial basis function kernel resulted in the least accurate solution. The highest accuracy can be achieved in this system by using a linear kernel. Due to the database creation process, the image array will become flatter. This implies that the data from each column-major column of the picture array has been moved to a single row or flattened into a 1-D iterator through the array. This database has been converted to text format, with column one indicating the categorization type and columns two through the end carrying picture codes. The soft margin cost function is controlled by a parameter called C or gamma, which controls the effect of each support vector. This process involves trading error penalty for stability. Text-classification-appropriate kernels include linear kernels. Radial Basis Function (RBF) is a nonlinear sample on a higher-dimensional space. In contrast to the linear kernel, it can handle situations when the relationship between class labels and attributes is nonlinear. Nonlinear kernels, such as a polynomial kernel, offer advantages over linear and RBF kernels in terms of speed and accuracy. Since the best performance rate and time for the C = 1 condition using multiclass OVR and a linear kernel are 90% and 1.40 seconds, respectively, this suggests that a linear kernel could be used in this situation.
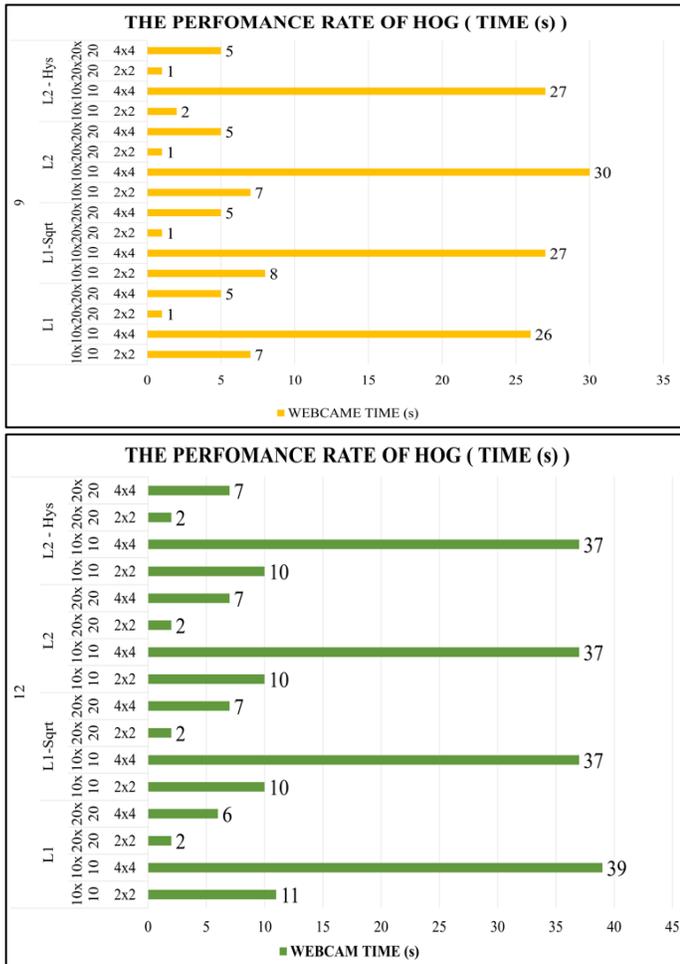
Fig. 14. Time Computation Rate of HOG Feature (s)

In this section, the objective of the third experiment is to determine the multiclass type, gamma, and c value, optimal
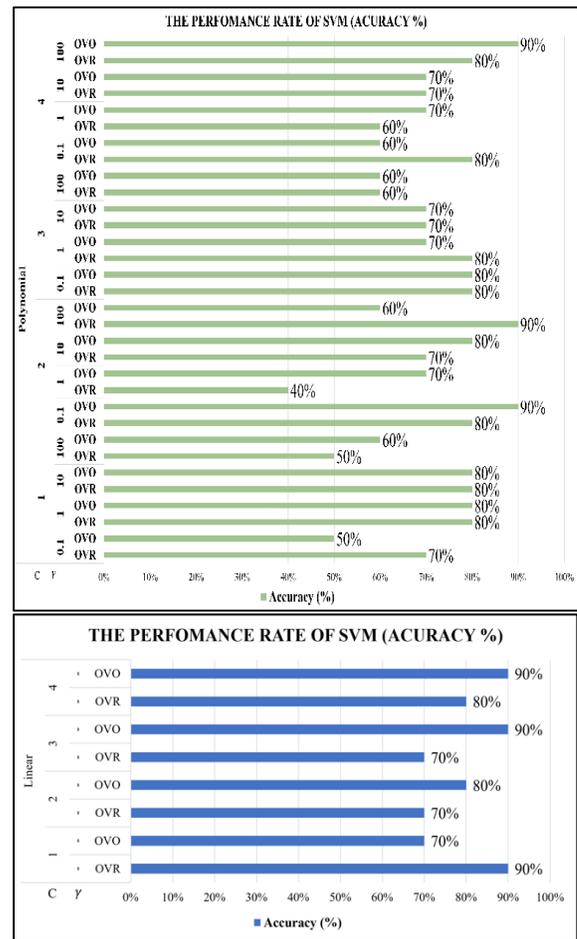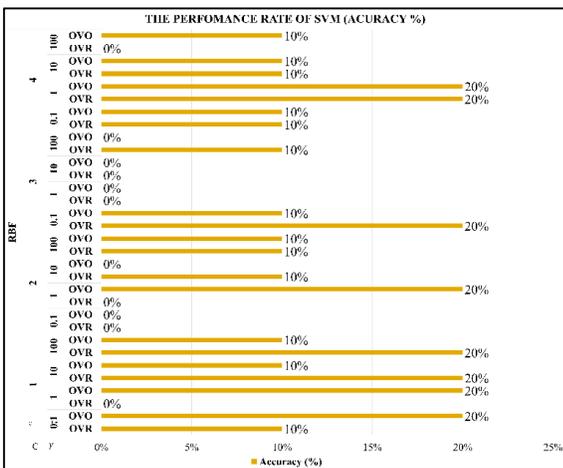
Fig. 15. The Accuracy Rate of SVM (%).
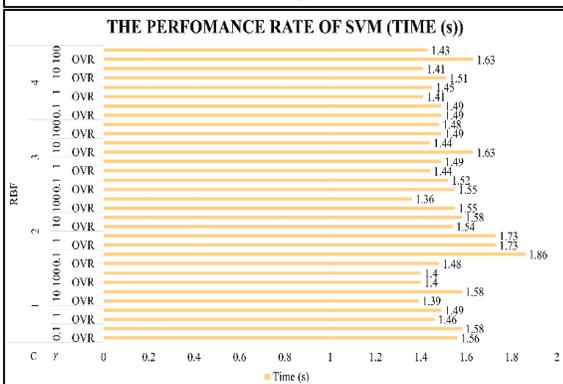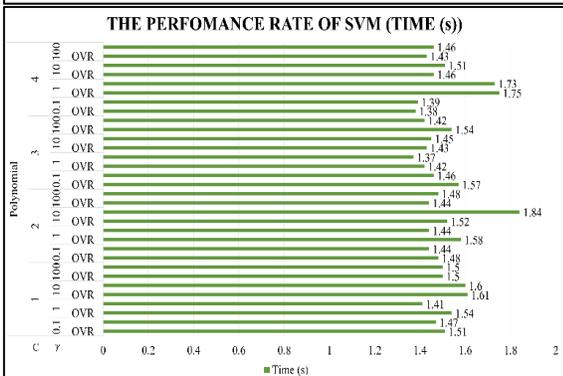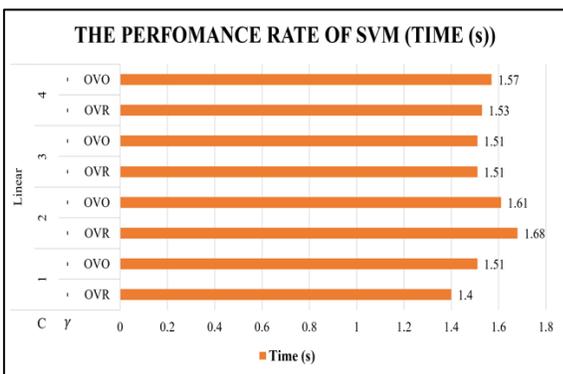
Fig. 16. The Time Computation Rate of SVM (s).

C. *Experimental Result of CNN Method*

To evaluate the performance of the wood classification system using the CNN method training parameter test, which includes observations on number of epochs, batch size, learning rate, and optimizer the CNN training parameter test trials must be conducted. The following Table IV is a scan of CNN algorithm parameter observations.

TABLE IV. CNN TRAINING PARAMETERS

No	CNN Training Parameters			
	Learning Rate	Batch Size	Epoch	Optimizer
1	0.001	8	1	Adam
2	0.0001	16	5	Rmsprop
3	0.00001	32	10	Adagrad
4	0.000001	64	15	SGD

The experiments will be conducted by modifying the Table III parameters. The subsequent experiment evaluates the learning rate as a CNN training parameter. The goal is to measure how quickly the system converges on the global minimum point, where performance is maximized. When the system reaches the greatest levels of accuracy, precision, and recall with variations in the learning rate of 0.000001, 0.00001, 0.0001, and 0.001, then the conditions have been met. Fig. 17 displays the outcomes of the tests.
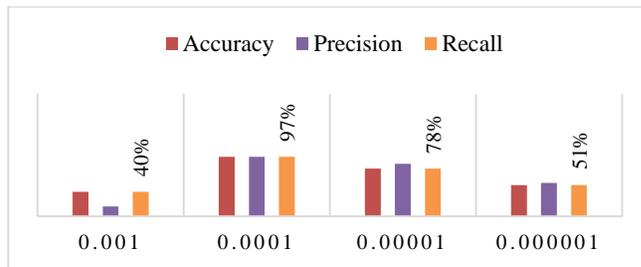
Fig. 17. Observation of Learning Rate Effect.

Fig. 18. Optimization of the Predicted Results from each Learning Rate.

Based on Fig. 17 and 18, which depict the system's hopeful attitude toward the expected object, the system's ideal learning rate is 0.0001. A learning rate of 0.001 is deemed excessively high, as it causes the system to diverge. The alternative is that

training accuracy does not improve as quickly as with a learning rate of 0.000001 since the system cannot reach the global region of minima due to the rapid rise in weight and numerous errors throughout the rebuilding process. The system rejects the input as being too low. Therefore, the length of the step reaches the minimal global area, yielding inferior accuracy, because the iterative step for updating the weight is quite small.

The second experiment of CNN training batch size analyzes the optimal test-time learning rate. Finding the optimal model for generalizing features and speeding the training process for wooden items is the objective. The observed batch sizes are 8, 24, and 32, as shown in Fig. 19, which depicts the experimental outcomes.
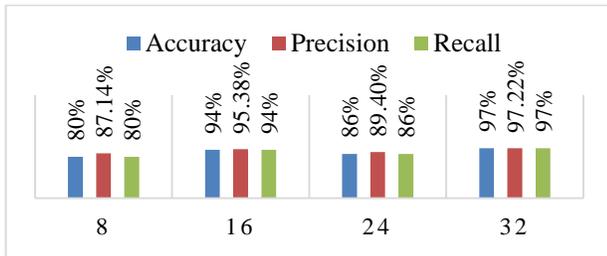
Fig. 19. Observation of Batch Size Effect.
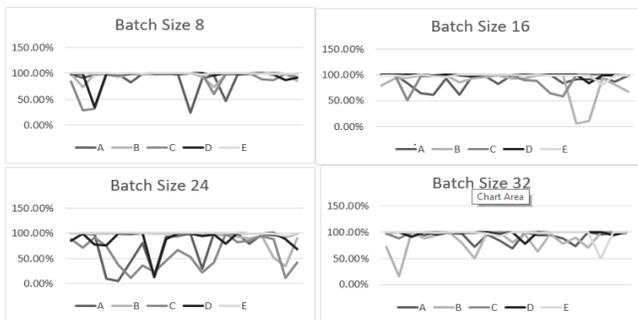
Fig. 20. Optimization of the Predicted Results from each Batch Size.

According to Fig. 19 and 20, the best batch size for a learning rate of 0.0001 is 32. A batch size that is too big can reduce accuracy since changing the weight per batch can result in errors. However, if the batch size is too small, the lengthy training period and system workload will increase. For this reason, it is absolutely necessary to adjust the learning rate parameters in accordance with the batch size in order to obtain precision and the optimum amount of training time [30].

Considering the best parameters from previous experiments, the number of epochs is the subject of the third experiment of CNN training parameters. This test is designed to identify the system with the best performance, one that is neither underfitted nor overfitted. 1, 5, 10, and 15 epochs are utilized, with the performance characteristics depicted in Fig. 21.
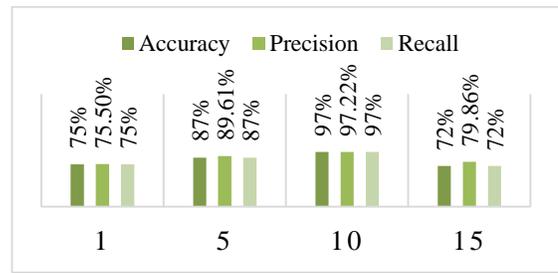
Fig. 21. Observation of Epoch Effect.

Experimentally, training with ten epochs yields the best performance as it can see in Fig. 22. A training cycle for all datasets is one epoch. Too few epochs will induce underfitting in the model we are training, hence the resultant accuracy is typically subpar. In contrast, if there are too many epochs, the training model may become overfit. Consequently, the system is less capable of generalizing the characteristics of the first new timber system site.
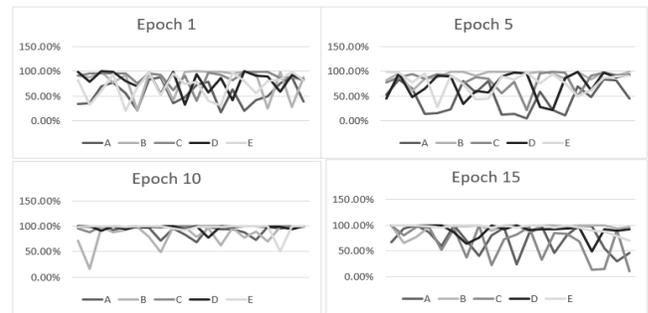
Fig. 22. Optimization of the Predicted Results from each Epoch.

The fourth CNN training parameter is the optimizer, which requires consideration of the preceding three experiment parameters. The optimizer is intimately connected to the system's ability to lower the value of losses and find the most effective actions to reach the global lowest point. The observed optimizer algorithms are SGD, adagrad, rmsprop, and adam, with test results depicted in the Fig. 23.
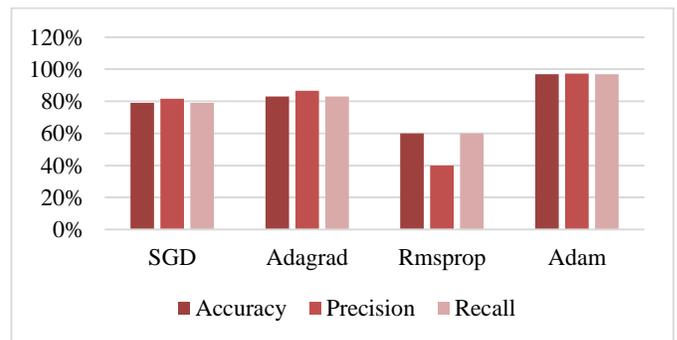
Fig. 23. Observation of Optimizer Effect.
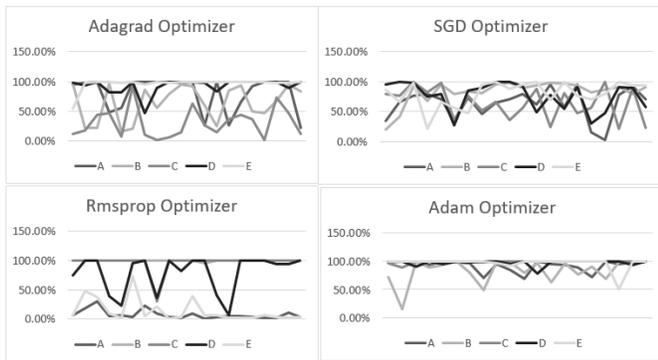
Fig. 24. Optimization of the Predicted Results from each Optimizers.

Based on Fig. 23 and 24, Adam Optimizer provides the most ideal performance. Adam is the most recent optimizer based on the notion of integrating SGD with momentum and rmsprop to tune the learning rate for each system network weight.

V. CONCLUSION

In conclusion, this research shows that a cedarwood classification system based on the CNN and SVM approach correctly recognized five wood classes using the features listed in Table I. SVM's classification output revealed that this system's performance rate is 90%, with a computation time of 1.40 seconds. The value derived from the experiment using the difference in the data set consists of 90% training data and 10% test data. In addition, the camera's angular position is 0°. Utilizing a linear kernel and one-versus-rest multiclass, the SVM classifier used in this study was useful. This condition easily categorizes and identifies the texture and fiber pattern of cedar. In contrast, the CNN improved the system's performance to 97% accuracy with an average computation time of 0.5472s. The value is calculated by comparing training and testing data distributions of 90:10, CNN training parameters of 0.0001, batch size of 32, epoch of 10, adaptive moment estimate as an optimizer and 723 lux of light intensity. It reveals that CNN is a significantly more effective method for image classification than SVM classification utilizing the HOG Feature.

ACKNOWLEDGMENT

REFERENCES

[1] M. I. Taqyudin, B. Irawan, and C. Setianingsih, "Wood Classification Based on Fiber Texture Using Backpropagation Method," in 2019 International Conference on Sustainable Engineering and Creative Computing (ICSECC), Aug. 2019, pp. 245–250. doi: 10.1109/ICSECC.2019.8907197.

[2] J. Feng, P. Dong, R. Li, C. Li, X. Xie, and Q. Shi, "Effects of wood fiber properties on mold resistance of wood polypropylene composites," Int Biodeterior Biodegradation, vol. 140, pp. 152–159, May 2019, doi: 10.1016/j.ibiod.2019.04.005.

[3] V. Bucur, "Properties of Wood Species for Percussion Instruments," in Handbook of Materials for Percussion Musical Instruments, Springer, 2022, pp. 695–785.

[4] S.-W. Hwang and J. Sugiyama, "Computer vision-based wood identification and its expansion and contribution potentials in wood science: A review," Plant Methods, vol. 17, no. 1, p. 47, Dec. 2021, doi: 10.1186/s13007-021-00746-1.

[5] L. Novakova, "The impact of technology development on the future of the labour market in the Slovak Republic," Technol Soc, vol. 62, p. 101256, Aug. 2020, doi: 10.1016/j.techsoc.2020.101256.

[6] S. Hartini, U. Ciptomulyono, M. Anityasari, and Sriyanto, "Manufacturing sustainability assessment using a lean manufacturing tool," International Journal of Lean Six Sigma, vol. 11, no. 5, pp. 943–971, Nov. 2020, doi: 10.1108/IJLSS-12-2017-0150.

[7] T. Han and G. A. Sánchez-Azofeifa, "A Deep Learning Time Series Approach for Leaf and Wood Classification from Terrestrial LiDAR Point Clouds," Remote Sens (Basel), vol. 14, no. 13, p. 3157, Jul. 2022, doi: 10.3390/rs14133157.

[8] G. Bonaccorso, Machine Learning Algorithms: Popular algorithms for data science and machine learning. Packt Publishing Ltd, 2018.

[9] G. Bonaccorso, Machine learning algorithms. Packt Publishing Ltd, 2017.

[10] M. You and C. Cai, "Wood Classification Based on PCA, 2DPCA, (2D)2PCA and LDA," in 2009 Second International Symposium on Knowledge Acquisition and Modeling, 2009, pp. 371–374. doi: 10.1109/KAM.2009.321.

[11] R. A. Pramunendar, C. Supriyanto, Dwi Hermawan Novianto, Ignatius Ngesti Yuwono, G. F. Shidik, and P. N. Andono, "A classification method of coconut wood quality based on Gray Level Co-occurrence matrices," in 2013 International Conference on Robotics, Biomimetics, Intelligent Computational Systems, Nov. 2013, pp. 254–257. doi: 10.1109/ROBIONETICS.2013.6743614.

[12] D. Shustrov, "Species Identification of Wooden Material using Convolutional Neural Networks ," Lappeenranta University of Technology, 2018.

[13] X. Zhuang, D. Zhou, X. L. Yu, and Z. M. Zhao, "An Image Processing and Laser Ranging Approach for Radio Frequency Identification (RFID) Tag Group Reading Performance Prediction.," Lasers in Engineering (Old City Publishing), vol. 51, 2021.

[14] C. B. Gonçalves, J. R. Souza, and H. Fernandes, "CNN architecture optimization using bio-inspired algorithms for breast cancer detection in infrared images," Comput Biol Med, vol. 142, p. 105205, Mar. 2022, doi: 10.1016/j.compbiomed.2021.105205.

[15] W. R. PERDANI, R. MAGDALENA, and N. K. CAECAR PRATIWI, "Deep Learning untuk Klasifikasi Glaukoma dengan menggunakan Arsitektur EfficientNet," ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika, vol. 10, no. 2, p. 322, Apr. 2022, doi: 10.26760/elkomika.v10i2.322.

[16] M. Giménez, J. Palanca, and V. Botti, "Semantic-based padding in convolutional neural networks for improving the performance in natural language processing. A case of study in sentiment analysis," Neurocomputing, vol. 378, pp. 315–323, Feb. 2020, doi: 10.1016/j.neucom.2019.08.096.

[17] W. Huang, J. Cheng, Y. Yang, and G. Guo, "An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis," Neurocomputing, vol. 359, pp. 77–92, Sep. 2019, doi: 10.1016/j.neucom.2019.05.052.

[18] F. Millstein, Convolutional neural networks in Python: beginner's guide to convolutional neural networks in Python. Frank Millstein, 2020.

[19] L. Alzubaidi et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," J Big Data, vol. 8, no. 1, pp. 1–74, 2021.

[20] [20] W. Ouyang, B. Xu, J. Hou, and X. Yuan, "Fabric Defect Detection Using Activation Layer Embedded Convolutional Neural Network," IEEE Access, vol. 7, pp. 70130–70140, 2019, doi: 10.1109/ACCESS.2019.2913620.

[21] Y. H. Liu, "Feature Extraction and Image Recognition with Convolutional Neural Networks," J Phys Conf Ser, vol. 1087, p. 062032, Sep. 2018, doi: 10.1088/1742-6596/1087/6/062032.

[22] S. Maharjan, A. Alsadoon, P. W. C. Prasad, T. Al-Dalain, and O. H. Alsadoon, "A novel enhanced softmax loss function for brain tumour

detection using deep learning," J Neurosci Methods, vol. 330, p. 108520, Jan. 2020, doi: 10.1016/j.jneumeth.2019.108520.

[23] S. Widiyanto, R. Fitrianto, and D. T. Wardani, "Implementation of Convolutional Neural Network Method for Classification of Diseases in Tomato Leaves," in 2019 Fourth International Conference on Informatics and Computing (ICIC), Oct. 2019, pp. 1–5. doi: 10.1109/ICIC47613.2019.8985909.

[24] J. Wu, L. Chang, and G. Yu, "Effective Data Decision-Making and Transmission System Based on Mobile Health for Chronic Disease Management in the Elderly," IEEE Syst J, vol. 15, no. 4, pp. 5537–5548, Dec. 2021, doi: 10.1109/JSYST.2020.3024816.

[25] B. Sugiarto et al., "Wood identification based on histogram of oriented gradient (HOG) feature and support vector machine (SVM) classifier," in 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Nov. 2017, pp. 337–341. doi: 10.1109/ICITISEE.2017.8285523.

[26] V. A. Gunawan, L. S. A. Putra, F. Imansyah, and E. Kusumawardhani, "Identification of Coronary Heart Disease through Iris using Gray Level Co-occurrence Matrix and Support Vector Machine Classification," International Journal of Advanced Computer Science and Applications, vol. 13, no. 1, 2022, doi: 10.14569/IJACSA.2022.0130177.

[27] N. Boyko and R. Hlynka, "Application of Machine Algorithms for Classification and Formation of the Optimal Plan.," in COLINS, 2021, pp. 1853–1865.

[28] M. Kachouane, S. Sahki, M. Lakrouf, and N. Ouadah, "HOG based Fast Human Detection," Jan. 2015, doi: 10.1109/ICM.2012.6471380.

[29] A. Rana, P. Singh, G. Valenzise, F. Dufaux, N. Komodakis, and A. Smolic, "Deep Tone Mapping Operator for High Dynamic Range Images," IEEE Transactions on Image Processing, vol. 29, pp. 1285–1298, 2020, doi: 10.1109/TIP.2019.2936649.

[30] S. L. Smith, P.-J. Kindermans, C. Ying, and Q. v. Le, "Don't Decay the Learning Rate, Increase the Batch Size," Nov. 2017.

Ransomware Detection using Machine and Deep Learning Approaches

Ramadhan A. M. Alsaidi[1], Wael M.S. Yafooz[2], Hashem Alolofi[3], Ghilan Al-Madhagy Taufiq-Hail[4], Abdel-Hamid M. Emara[5], Ahmed Abdel-Wahab[6]

Department of Mathematics, College of Science and Arts in Gurayat, Jouf University Gurayat, Saudi Arabia[1]
Department of Computer Science, College of Computer Science and Engineering, Taibah University, Medina 42353, Saudi Arabia[2]
Bremen University, Bremen, Germany[3]
College of Business (COB), University of Buraimi (UOB), Al Buraimi Governorate, Oman[4]
Department of Computer Science, College of Computer, Science and Engineering, Taibah University, Medina 42353, Saudi Arabia, Department of Computers and Systems Engineering, Faculty of Engineering, Al-Azhar University, Cairo 11884, Egypt[5]
Faculty of Computer Studies, Arab Open University, Riyadh, Saudi Arabia, Department of Computers and Systems Engineering, Faculty of Engineering, Al-Azhar University, Cairo 11884, Egypt[6]

Abstract—**Due to the advancement and easy accessibility to computer and internet technology, network security has become vulnerable to hacker threats. Ransomware is a frequently used malware in cyber-attacks to trick the victim users to expose sensitive and private information to the attackers. Consequently, victims may not be able to access their data any longer until they pay a ransom for stolen files or data. Different methods have been introduced to overcome these issues. It is evident through an extensive literature review that some lexical features are not always sufficient to detect categories of malicious URLs. This paper proposes a model to detect Ransomware using machine and deep learning approaches. This model was introduced as a novel feature for classification using the idea that starts with "https://www." This feature was not considered in the earlier papers on malicious URLs identification. In addition, this paper introduced a novel dataset that consists of 405,836 records. Two main experiments were carried out utilizing malicious URL features to defend Ransomware using the proposed dataset. Moreover, to enhance and optimize the experimental accuracy, various hyper-parameters were tested on the same dataset to define the optimal factors of every method. According to the comparative and experimental results of the applied classification techniques, the proposed model achieved the best performance at 99.8% accuracy rate for detecting malicious URLs using machine and deep learning.**

Keywords—Machine learning; ransomware; URL classification; malicious URLs; deep learning

I. INTRODUCTION

Over the past few years, the growth of ransomware has become an uncontrolled cyber problem and highly profitable criminal business. Ransomware attacks are primarily performed using malicious Uniform Resource Locators (URLs) [1]. On May 7, 2021, Colonial Pipeline [2], an American oil pipeline system that originates in Houston, Texas, suffered a ransomware cyberattack that hit computerized facilities managing the pipeline.

In response, Colonial Pipeline Company stopped all of the pipeline's operations to contain the attack. Colonial Pipeline

paid the demanded ransom (75 bitcoin or $4.4 million) within several hours after the attack. The attackers then sent Colonial Pipeline a software application to restore their network, although it was slow. This example illustrates how problematic Ransomware can be. Therefore, it becomes a necessity to know more about Ransomware [3].

There are essentially two types of ransomware: locker and crypto-ransomware. Locker ransomware works by blocking the victim from arriving at their files by denying access to computing resources, such as locking the desktop or preventing the victim from logging in and, demanding a ransom to regain access to the system. Crypto-Ransomware encrypts all data on the target system until the victim pays the ransom via the Bitcoin currency and obtains the decryption key from the hacker. Some kinds of crypto-ransomware can progressively erase files or release them to the public if the victim declines to pay the ransom on time. Recent ransomware families are essentially based on this type. It can have disastrous effects, primarily on corporate and government agencies, if they do not have a backup to restore the system to the state before the attack. Following the installs of the crypto-ransomware installs on the target system, it can noiselessly search for essential files based on their extensions and starts encrypting them. Then the ransomware will search for files on the local hard disk, the external hard disk, and in the network shares. Many crypto-ransomware users make sure that the files cannot be restored from the backup by totally deleting the backup. Usually, there are three main ways that ransomware can attack a system. They can get it infected, then takes over the files.

The first common way is through browsing the web. The malicious websites usually install the exploit kit to take over the machine and install additional malicious software such as ransomware. The other even more common way is through emails. Emails become a very demanding business tool. Email, more specifically, has two different vectors within itself. It may contain a malicious web link. That is sent within the email message that people click assuming that they are directed to a safe web page location. The second common way

in which an email can affect us is with malicious attachments sent directly in the email body. Once it opens, it can take control over the entire system. The third common way is through free software or games. Most of the free software found on the Internet is malicious software developed by hackers, disguised as legitimate software to access systems. Based on the said, it is necessary to notify the user that some URLs direct users to a malicious website before accessing it. Though numerous cybersecurity techniques defences are developed against Ransomware URLs, the nature of the security still requires further research to improve the entire system. As a remedial response to Ransomware threats, the present paper proposes a model to detect and track malicious URLs using machine learning classifiers and deep learning approaches. To enhance and secure the efficiency of our model, a novel dataset called Ransomware Detection Dataset (RDD) has been introduced. RDD is available for the public on the git hub website [4].

The proposed model achieved a high performance in terms of accuracy using both MCLs and DL, securing 99% and 99% respectively. In addition, through the discussion and analysis of numerous identity algorithms, malicious URLs detection systems were used to identify different forms of known phishing and to examine the appearance of dangerous URLs in websites.

The contribution of this paper can be briefed as follows:

- Proposed a model to detect and track malicious URLs using machine and deep learning approaches.

- Proposed a new feature used in identifying malicious and benign classification using the idea starting with "https://www.

- Introduced a novel dataset containing 405,836 rows of a URL to classify webpages as Malicious or Benign called Ransomware Detection Dataset (RDD) [4]

- Compared performance of the proposed model with that of MLCs and DL approaches.

- Proposed detection algorithms that enhance accuracy in classifying malicious and benign website applications via optimizing URL features and selecting optimal factors for every method.

- Examine three loss functions (Binary Crossentropy,Hinge, and Mean Squared Error respectively) on the RDD using three optimizers (Adam,SGD, and ADAgard).

The rest of this paper is structured as follows: Section II has to do with a brief review of related works; Section III describes the research approach, feature extraction and classifications algorithms; methods and materials will be explained in Section IV; experiments, results and discussion will be presented in Section V; Section VI concludes the paper and provides suggestions for future work.

II. RELATED WORK

To improve accuracy and faster detection, numerous methods have been introduced for malicious URLs detection.

Recently, research has shown that many detection methods can resist the increasingly diverse hackering threats. Popular techniques used for malicious URL detection are machine learning/deep learning-based detection, blacklist-based detection and, heuristic rules-based detection. Blacklist based approach utilizes a massive list of malicious URLs that can be created from numerous sources manually or programmatically. Many commercial products build blacklists utilizing user proprietary and their feedbacks mechanisms to detect malicious URLs [5] and [6], like WOT Web of Trust [7], McAfee's SiteAdvisor [8], Cisco IronPort Web Reputation [9], and Trend Micro Web Reputation Query Online System [10]. On the other hand, diverse websites supply blacklist applications, such as: PhishTank [11], jwSpamSpy [12] and DNS-BH [13]. Although blacklist technique gives faster and low false positive detection, it has some drawbacks. The newly generated malicious URLs are not detected, so the database of blacklisted URL requires timely and frequent updates. As the result, researchers using the heuristic rule-based detection method to overcome this issue, in which generalized rules are used for malicious URL detection. These rules are based on features extracted from the datasets [14]. The limitations of this method are the mapping of weightage to the rules and, the threshold value may be modified based on datasets, it becomes unclear how to fix threshold value for each rule [15].

Later, the machine learning approach has been used in several methods to classify malicious URLs. It operates on two stages: first, extraction of the dataset features and training the model; then, training is used for testing. The best advantage of this technique is that the generated model can automatically assign the weightage of selected features and detect the malicious URL, including newly generated malicious URLs. A survey on malicious URL recognition via machine learning was presented by Doyen et al. in [16]. It scans the cons and pros of various methods used in malicious URLs Authors Suppressed Due to Excessive Length identification literature. Practically, the features utilized in machine learning approach are classified into four broad categories: host-based features, popularity features, content-based features and lexical features. Author in [17] presented a multiclass classification method to identify the malicious URL. It considers 18 hos-based features, 34 content related features and 63 lexical features. Recently, Djaballah et al. [18] presented a new work to recognize the malicious URL in social network such as twitter. In that work, three machine-learning techniques were used namely Random forest, SVM and Logistic Regression for the experiments, giving 5.51%, 93.43% and 90.28% accuracy rate respectively. In [19], the researchers suggested a technique for malicious website detection which utilizes various feature selection methods and diverse machine learning techniques with PCA such as RF, NN, bagging, KNN, NB, and SVM were used for experiments. In [20], the authors presented a method of choosing the optimal features using Chi-Square and ANOVA F-value to detect phish URL. The light weighted technique that uses only a fewer number of lexical features for malicious URL detection was introduced in [14].

III. PROPOSED APPROACH

This paper aims to explore how a URL is identified as malicious or benign by using only its associated URLs lexical features quickly and accurately. This work works on build a URL classifier to predict whether a URL is malicious or benign with a very high degree of recall and precision. When the user sends a DNS (Domain Name System) request, it will be sent first to our URL classifier. If the URL is Benign, then the DNS request will be sent to the DNS server. And in case the URL is malicious, then DNS requests will be dropped and not forwarded to the DNS server.

IV. METHODS AND MATERIALS

This section presents the methods and phases of the proposed model. There are four phases in the presented detection architecture. They are data collection, feature extraction, classification, and model evaluation as shown in Fig. 1.

A. Data Collection

This is the input phase on the mode. Processing and preparing datasets are critical stages. However, getting sizeable balanced data is one of the potential challenges. The efficiency and performance of the system essentially depend on them. The importance of this issue appears when the model deals with imbalanced data that will cause a bias towards predicting the larger class, and on the other hand, may ignore the smaller class altogether. Therefore, one of the challenges that this work is handling a sizeable balanced dataset. Consequently, the dataset was collected from six different Kaggle resources [21, 22, 23, 24, 25] to create an extensive balanced dataset. This effort acquired a total of 405,836 URLs, 202918 benign and 202918 malicious. At the end of this phase, the dataset has been built.
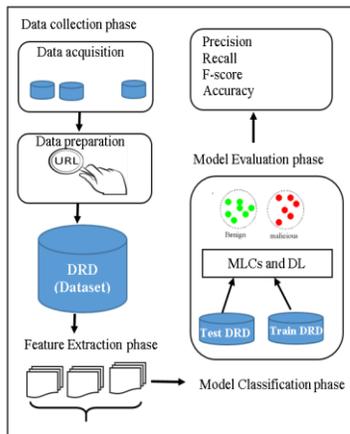
Fig. 1.   Model Architecture.

B. Features Extraction

This is the second phase of the model. It receives input from the first phase. The reliable and discriminative feature is significant for achieving high-performance machine learning classification. The criteria of attributes are essential to present the crucial characteristics of the malicious URL. It should be noted that in the Feature extraction phase, there are three main steps: the length of the feature, count feature, and binary

feature. These steps are performed using the work developed in the open-source code, which Siddharth Kumar provides in [23]. The output of this phase is the features extracted from the URLs. These features are used in the next phase which is classification.

What is a URL? To understand attackers' strategy, firstly, the reader must know the basic components of a URL. The basic structure of a URL is depicted in a URL is typically made up of six or seven components. First is the Protocol, which is used to access resources on the Internet. It could be Hypertext Transfer Protocol Secure (HTTPS) or the less secured HTTP. The other components are subdomain, domain name, top-level domain, path, query, and parameters, as shown in Fig. 2.
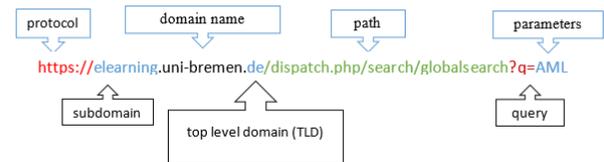
Fig. 2.   URLs Components.

Features Score Initially, this work specified 17 most common lexical features. Since all extracted features are not always suitable for classification, the quality and effectiveness of each lexical feature was examined by testing and training the classifier with each feature group separately. The results are illustrated in Table I.

TABLE I.   FEATURES SCORE

Feature	Score	Feature	Score
Start with https://www	0.9978527	count=	0.546040
count-https	0.980527	count?	0.545931
count-www	0.924533	path length	0.516274
ld length	0.723060	use of ip	0.513758
hostname length	0.681407	count-digits	0.510565
count-http	0.666719	count@	0.510015
count.	0.605214	count%	0.498175
count dir	0.552626	fd length	0.484810
count-	0.549475	count-letters	0.473917

A compelling feature selection will enhance the classifier ability of the model with low computational complexity that reduces the storage consumption and execution time.

Ranking analysis help to find the optimal feature. The effect of prominent features on recognizing accuracy will be addressed in the following:

- Counting HTTPS The first and most crucial feature found was the count-https. The count of HTTPS was calculated in all URLs in the dataset. It has been observed that all benign URLs (202918) use the HTTPS protocols. On the other side, most of the malicious URLs (195014) do not use HTTPS protocols, making this feature one of the best features for training our classifier.

- Counting www The second most essential feature is count-www in which the frequency of www was

counted in all URLs in the dataset. It has been observed that almost all benign URLs (202739) use www whereas most of the Malicious URLs (172290) do not have www.

- Start with https://www Counting https is an essential feature since most malicious URLs do not use HTTPS. However, the problem with this feature in previous work is that it just counts how much time it acquires. It does not take into account the location of the HTTPS. While testing, it was noticed that the result got many false positives because many malicious URLs used HTTPS in the middle of the URL. To solve this problem, this research introduced a new feature called Start with https://www. It is a binary feature that returns one if the URL starts with https://www. And returns zero otherwise. This feature will enhance the classifier ability of the model (i.e., reach an accuracy of 99.78%) with low computational complexity that reduces the storage consumption and execution time. This feature is newly selected and extracted in this paper.

To represent URLs detection issues mathematically. Let us define a function space [26, 27] as:

$$F : u \rightarrow \{0, 1\}$$

Where u represent the URLs. Note that the function is assumed to take values in {0, 1}. Meanwhile, g(u) represents the values of every feature.

$$(F \circ g)(u) = F(g(u)) = F([f1, f2, \dots, f19])$$

$$= \{1\ malicious\ \ 0\ Benign\ \}$$

where g(u) has three operations $\{counting\ compute length\ Boolean\ \}$ and [f1, f2, ..., f19] represents the value of every features.

C. Classification Process

In the third phase, the classification process is carried out by using the most common MLCs and DL approaches used scholarly [28, 29]. In MCLs, the most common classifiers used are the Random Forest, Gradient Boosting Classifier, Neural Net,poly SVM, Decision Tree, AdaBoost, Nearest Neighbors, Stochastic Gradient Descent (SGD), Naive Bayes, and QDA. The general overview of the machine learning classification process is shown in Fig. 3. In deep learning models architecture as shown in Fig. 4, two deep neural networks ANN and LSTM mode are applied. In both experiments, the dataset is divided 30% for training and 70% for testing.

The model performance has been evaluated based on the most common methods using a confusion matrix. For all experiments, the unified performance indicators such as accuracy, precision, recall, and F1 were used to evaluate the performance of the model as represented in the following mathematical formula.
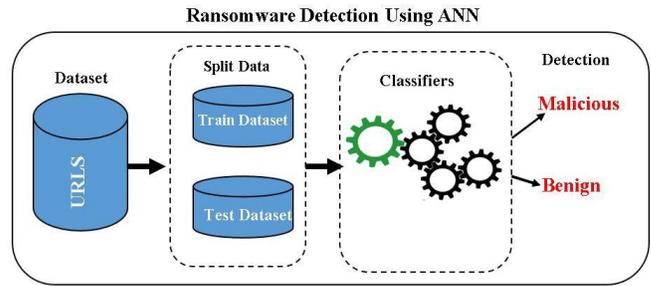
Fig. 3. Overview of Machine Learning Classification Process.
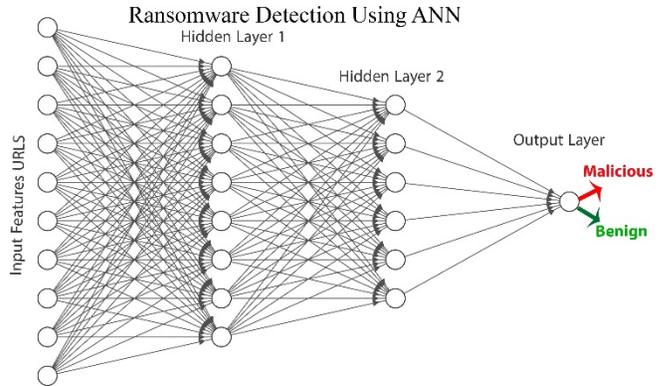
Fig. 4. Deep Neural Networks Architecture.

D. Model Evaluation

The model evaluation has been used in the most popular accuracy measure called confusion matrix to calculate accuracy, precision, recall and, the F1 Score that use in [30].

- Accuracy: measures how much of the data are labeled correctly

$$Accuracy = \frac{T\ P\ + T\ N}{T\ P\ + T\ N + F\ P\ + F\ N}$$

- Precision (specificity): it estimates how many identified targets are indeed relevant (real targets)

$$Specificity = \frac{T\ P}{T\ P\ + F\ P} = \frac{true\ positive}{no.\ of\ predicted\ positive}$$

- Recall (Sensitivity): it measures how good it is at identifying the positive cases.

$$Sensitivity = \frac{T\ P}{T\ P\ + F\ N} = \frac{true\ positive}{no.\ of\ actual\ positive}$$

- F1 Score: F1 is the function of Precision and Recall.

$$F1 = 2\ x\ \frac{Precision\ x\ Recall}{Precision\ +\ Recall}$$

V. RESULTS AND DISCUSSION

This section discusses the experimental results of the proposed model which were used to detect malicious URLs. These experiments are categorized according to their processing architecture; the machine learning experiments and deep learning approach experiments.

A. Machine Learning Experiments

This section presents and demonstrates the results of the experiments of most popular classifiers used in our model [28, 29]. The experiments evaluated in this section used a confusion matrix to show the detection performance of the models. After creating the large balance dataset and extracting features, the implementation of the model is completed in two stages. In the first stage, all the features were used to train the models. In the second stage, we tested the adoption of the presented model with new lexical feature, namely, Start with https://www. This feature enhanced the classifier ability of the model (i.e., reaching an accuracy of 99.78%) with low computational complexity that reduced storage consumption and execution time. For all experiments, 75% of the dataset was utilized for training and 25% for testing. It should be noted that for Preprocessing, Classification, Dimensional reduction, Model selection, and evaluation Scikit-learn (sklearn) library has been used. To find the best parameter values, the GridSearchCV, provided by sklearn, has been used. The results of the experiments are shown in Tables I and II.

TABLE II. EVALUATION METRICS

Sr. No.	Classification Algorithm	Accuracy	precision	recall	f1-score
1	Random Forest	0.998115	0.9999	0.996	0.9981
2	GradientBoosting Classifier	0.99809	0.9997	0.997	0.9981
3	Neural Net	0.997967	0.9996	0.996	0.998
4	poly SVM	0.997893	0.9997	0.996	0.9979
5	Decision Tree	0.997893	0.9996	0.996	0.9979
6	AdaBoost	0.997893	0.9998	0.996	0.9979
7	Nearest Neighbors	0.997856	0.9994	0.996	0.9979
8	Stochastic Gradient Descent (SGD)	0.997647	0.9998	0.996	0.9976
9	Naive Bayes	0.995232	0.9953	0.995	0.9952
10	QDA	0.935577	0.9944	0.876	0.9316

B. Deep Learning Experiment

This section presents the results of the deep neural network experiment built on the basis of a Multi-layer Perceptron. Two Deep learning models were applied. They are an artificial neural network (ANN) and long short-term memory (LSTM). All the experiments were executed using Keras by Jupyter platform. In the ANN experiment, the model consists of five layers of one input layer, three hidden layers, and one output layer. The input layer consists of seventeen input dimensions. The features and activation function is "Relu" and the output are 256 neurons. The three hidden layers consist of 256, 128, 64 neurons, and the activation function is "Relu" with a drop out of 30%, while the output layer consists of a single layer

with "sigmoid" as the activation function. The loss function used in the testing stage was binary crossentropy while three optimizers were used. They are: Adam, ADAgard, and SGD in which the learning rate was recorded at 0.01, 0.001 and 0.0001. Both training and testing for the models were applied, and the dataset was divided into 70% for training, and 30% for testing utilizing Adam optimizer with the three categories of the learning rate. In training and validation, the model achieved a high-performance rate of 99.16% in terms of accuracy while for validation it achieved 98.10%. Twenty epochs were included, and the batch size was 64. The loss function noticeably decreased. The model accuracy of ANN for learning rates of 0.01, 0.001, and 0.0001 are shown in Fig. 5 to 7 for training and validation.
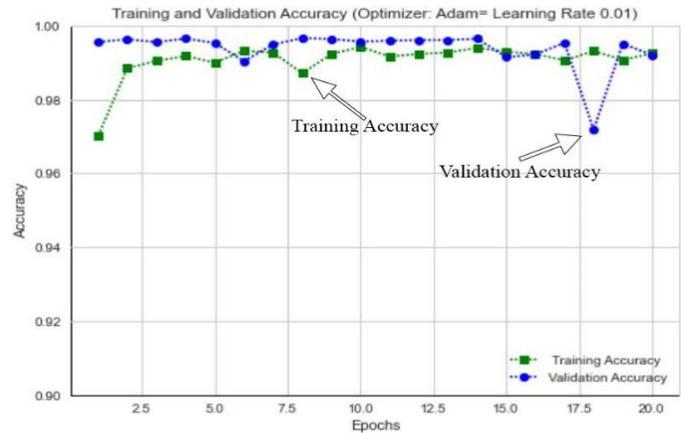
Fig. 5. Model Accuracy for Validation and Testing using Adam Optimizer (LR 0.01).

While loss function for testing and validation is shown in Fig. 8 to 10. The best accuracy rate has been recorded using a learning transfer of 0.0001.
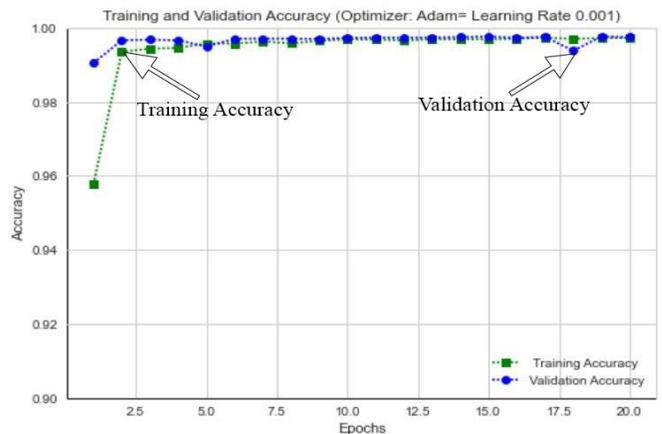
Fig. 6. Model Accuracy for Validation and Testing using Adam Optimizer (LR: 0.001).
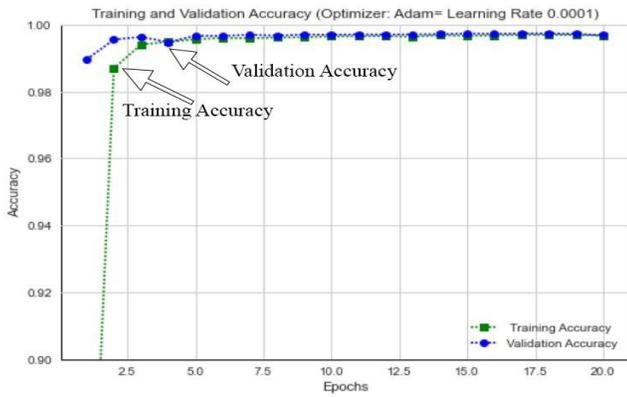
Fig. 7. Model Accuracy for Validation and Testing using Adam Optimizer
(LR: 0.0001).

Using the same experiment setting and same model hyperparameter "SGD" and "ADAgrad" optimizers were applied. The model performance was calculated as shown in Fig. 11 and Fig. 12 while the loss for the validation and accuracy is presented in Fig. 13 and Fig. 14. Additionally, three loss functions were used. The results as shown in Table III, IV, and V, Binary Crossentropy, Hinge, and Mean Squared Error respectively.

Fig. 8. Loss of Validation and Testing using Adam (LR0.01).
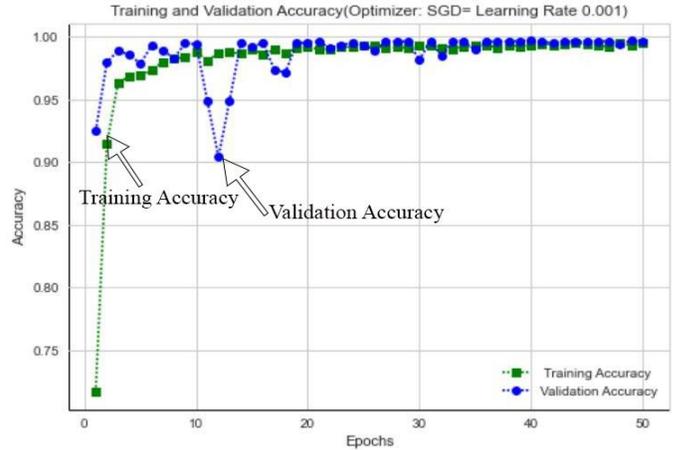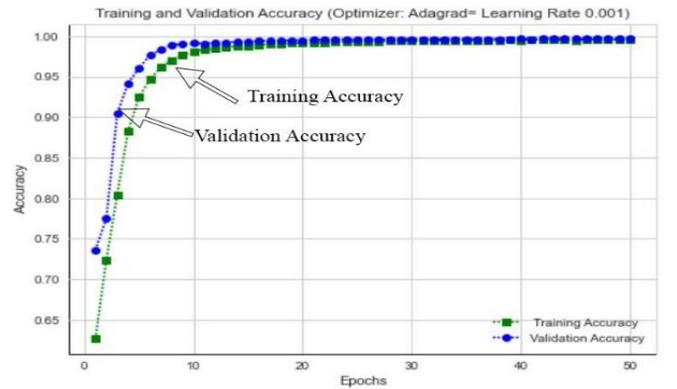
Fig. 11. Model Accuracy for Validation and Testing using SGD Optimizer.

Fig. 9. Loss of Validation and Testing using Adam (LR: 0.001).

Fig. 12. Model Accuracy for Validation and Testing using ADAgard
Optimizer.

Fig. 10. Loss of Validation and Testing using Adam (LR: 0.0001).
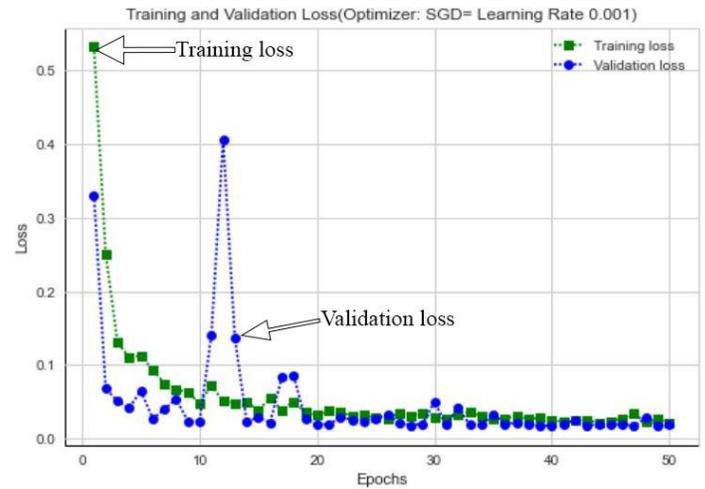
Fig. 13. Loss of Validation and Testing using SGD Optimizer.

TABLE III.   ADAM OPTIMIZER WITH THREE LOSS FUNCTIONS

Loss function	TR	Loss	Acc.TR	Acc.VAL	Acc.TE
Binary Crossentropy	0.01	0.0203	0.9961	0.9965	0.997
	0.001	0.0161	0.9969	0.9973	0.997
	0.0001	0.0175	0.9967	0.9974	0.9967
Hinge	0.01	0.9985	0.5008	0.4997	0.499
	0.001	0.5117	0.9875	0.9933	0.993
	0.0001	0.5035	0.9959	0.9968	0.9968
Mean Squared Error	0.01	0.1887	0.8108	0.8858	0.885
	0.001	0.0028	0.9970	0.9969	0.996
	0.0001	0.0034	0.9964	0.9969	0.9969

TABLE V.   ADAGARD OPTIMIZER WITH THREE LOSS FUNCTIONS

Loss function	TR	Loss	Acc.TR	Acc.AVL	Acc.TE
Binary Crossentropy	0.01	0.02	0.9968	0.9971	0.997
	0.001	0.091	0.9824	0.9929	0.992
	0.0001	0.707	0.5557	0.6493	0.649
Hinge	0.01	0.504	0.996	0.9957	0.995
	0.001	0.642	0.8932	0.922	0.921
	0.0001	0.922	0.6003	0.689	0.689
Mean Squared Error	0.01	0.004	0.996	0.9967	0.996
	0.001	0.068	0.9389	0.9771	0.977
	0.0001	0.227	0.6314	0.7093	0.709

Fig. 14. Loss of Validation and Testing using ADAgard Optimizer.

Fig. 15. Model Accuracy for Validation and Testing using LSTM.

In the LSTM experiment, the model consists of three layers. They are the input layer, LSTM layers, and output layer. In the input layer, the model receives a training dataset of features as an input dimension with 'Relu' activation function including 128 neurons.

In the second and third layers of LSTM, the number of neurons is 64, 32 respectively. The activation function is "Relu" for both layers. The dataset was divided into 70% for training and 30% for testing. In the training, the model reached a performance rate of 99.76% in terms of accuracy with twenty epochs using 'Adam' optimizer with a learning rate of 0.0001 and the validation reached to 99.66% is shown in Fig. 15. The loss function of validation and testing is "binary-cross entropy" which decreased with each training epoch is shown in Fig. 16.

Fig. 16. Loss of Validation and Testing using LSTM.

TABLE IV.   SGD OPTIMIZER WITH THREE LOSS FUNCTIONS

Loss function	TR	Loss	Acc.TR	Acc.VAL	Acc.TE
Binary Crossentropy	0.01	0.1463	0.9584	0.9592	0.959
	0.001	0.1543	0.9585	0.9757	0.975
	0.0001	0.5725	0.6909	0.7331	0.733
Hinge	0.01	0.5487	0.9514	0.9780	0.977
	0.001	0.7697	0.7383	0.7495	0.749
	0.0001	0.9257	0.5611	0.6394	0.639
Mean Squared Error	0.01	0.0251	0.9718	0.7260	0.725
	0.001	0.0984	0.9096	0.9518	0.951
	0.0001	0.2272	0.6247	0.7110	0.710

VI.   CONCLUSION

The study explored the possibility of distinguishing benign and malicious URLs by using machine learning classification algorithms. A new feature was proposed for classifying and identifying malicious and benign using the idea of start with https://www. Experimental results using this feature were found significant. To evaluate the proposed feature, a sizeable balanced dataset has been created with a total of 405,836 URLs, 202918 benign, and 202918 malicious. Machine learning classifiers and deep learning approaches were applied, particularly ANN and LSTM. The proposed approach

gave a high classification accuracy rate of 99.81%, and a very high degree of classification precision at 99.99%. In the future, more focused research on modeling URL detection can be made to cope with the most sophisticated cyber-attack techniques.

REFERENCES

[1] L. W. Al-Yaseen, "Improving intrusion detection system by developing feature selection model based on firefly algorithm and support vector machine." IAENG International Journal of Computer Science, vol. 46, no. 4, pp. 534-540, 2019.

[2] M. Sigalos, "Colonial Pipeline cyberattack is no cause for panic – here's why," CNBC, May 14, 2021. https://www.cnbc.com/2021/05/14/colonial-pipeline-hack-doesnt-mean-more-ransomware-attacks-critical-infrastructure.htm

[3] S. Sachdeva, R. Jolivot and W. Choensawat, "Android malware classification based on mobile security framework," IAENG International Journal of Computer Science, vol. 45, no. 4, pp. 514–522, 2018.

[4] R. Alsaidi, "Ransomware detection dataset (RDD) dataset," Ransomware detection dataset (RDD) dataset. [Online]. Available: kaggle, https://www.kaggle.com/ramdhanamalsaidi/a-novel-dataset-containing-405836-url, [Accessed November 28, 2021].

[5] H. Choi, B. B. Zhu, and H. Lee, "Detecting malicious web links and identifying their attack types," in 2nd USENIX Conference on Web Application Development, 2011, vol. 11.

[6] Y. Sun Y. Chen Y. P and L. Wu, "Android malware family classification based on deep learning of code images," IAENG International Journal of Computer Science, vol. 46, no. 4, pp. 524–532, 2019.

[7] "Web of Trust community-based safe surfing tool," Website Safety Check and Phishing Protection Web of Trust, 2021. [Online]. Available: https://www.mywot.com.

[8] "Service for reporting the safety of web sites," McAfee SiteAdvisor, 2006. [Online]. Available: http://www.siteadvisor.com.

[9] "IronPort Web Reputation: Protect and Defend Against URL-Based Threats," Ironportstore.com. [Online]. Available: https://www.ironportstore.com/search.php?q=Protect+and+defend+against+URL-based+threat. [Accessed: 25-Nov-2022].

[10] "Web reputation query - online system," TREND MICRO, 2021. [Online]. Available: http://reclassify.wrs.trendmicro.com.

[11] "Free community site for anti-phishing service," PHISHTANK, 2021. [Online]. Available: http://www.phishtank.com.

[12] J. Wein, "jwSpamSpy - E-mail spam filter for Microsoft Windows TM," Jwspamspy.net. [Online]. Available: http://www.jwspamspy.net. [Accessed: 25-Nov-2022].

[13] "Malware Prevention through domain blocking," yumpu.com. [Online]. Available: https://www.yumpu.com/en/document/read/44684944/malware- prevention-through-domain-blocking. [Accessed: 25-Nov-2022].

[14] A. Saleem Raja, R. Vinodini, and A. Kavitha, "Lexical features based malicious URL detection using machine learning techniques," Materials Today: Proceedings, Apr. 2021.

[15] H. Tupsamudre A. K. Singh and S. Lodha, "Everything is in the name–a url based approach for phishing detection," in International symposium on cyber security cryptography and machine learning, 2019, pp. 231–248.

[16] D. Sahoo, C. Liu, and S. C. H. Hoi, "Malicious URL Detection using Machine Learning: A Survey," arXiv [cs.LG], 2017.

[17] D. Patil and J. Patil, "Feature-based malicious url and attack type detection using multi-class classification," in The ISC International Journal of Information Security, 2018, pp. 141–162.

[18] K. A. Djaballah, K. Boukhalfa, Z. Ghalem, and O. Boukerma, "A new approach for the detection and analysis of phishing in social networks: the case of Twitter," in 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS), 2020.

[19] A. Zamir, H.U. Khan, T. Iqbal, N. Yousaf, F. Aslam, A. Anjum, M. Hamdani, "Phishing web site detection using diverse machine learning algorithms," The Electronic Library, vol. 38, no. 1, pp. 65–80, Jan. 2020.

[20] H. M. Junaid Khan, Q. Niyaz, V. K. Devabhaktuni, S. Guo, and U. Shaikh, "Identifying generic features for malicious URL detection system," in 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2019.

[21] Remosin, "Bot_Detection," Bot Detection. [Online]. Available: kaggle, https://www.kaggle.com/datasets/remosin/bot-detection. [Accessed November 24, 2022].

[22] A. Singh, "Dataset of Malicious and Benign Webpages," Dataset of Malicious and Benign Webpages. [Online]. Available: kaggle, https://www.kaggle.com/datasets/aksingh2411/dataset-of-malicious-and-benign-webpages. [Accessed November 24, 2022].

[23] S. Kumar, "Detect Malicious URL using ML," Detect Malicious URL using ML. [Online]. Available: kaggle, https://www.kaggle.com/code/siddharthkumar25/detect-malicious-url-using-ml. [Accessed November 24, 2022].

[24] T. Tiwari, "Phishing Site URLs," Phishing Site URLs. [Online]. Available: kaggle, https://www.kaggle.com/datasets/taruntiwarihp/phishing-site-urls. [Accessed November 24, 2022].

[25] J. Antony, "Malicious_n_Non-Malicious URL," Malicious_n_Non-Malicious URL. [Online]. Available: kaggle, https://www.kaggle.com/antonyj453/urldataset. [Accessed November 24, 2022].

[26] A. R. Alsaidi, "Impact of the various measures of similarity on the statistic hierarchical neural response method." Journal of Theoretical and Applied Information Technology, vol. 99, no. 21, 2021.

[27] R. A. M. Alsaidi, H. Li, Y. Wei, R. Khaji, and Y. Y. Tang, "Hierarchical sparse method with applications in vision and speech recognition," Int. J. Wavelets Multiresolut. Inf. Process., vol. 11, no. 02, p. 1350016, 2013.

[28] Alhujaili, R. F., & Yafooz, W. M. (2022, May). Sentiment Analysis for YouTube Educational Videos Using Machine and Deep Learning Approaches. In 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI) (pp. 238-244). IEEE.

[29] Alhejaili, R., Alsaeedi, A., & Yafooz, W. (2023). Detecting Hate Speech in Arabic Tweets During COVID-19 Using Machine Learning Approaches. In Proceedings of Third Doctoral Symposium on Computational Intelligence (pp. 467-475). Springer, Singapore.

[30] Alhejaili, R., Alhazmi, E. S., Alsaeedi, A., & Yafooz, W. M. (2021, September). Sentiment Analysis of The Covid-19 Vaccine For Arabic Tweets Using Machine Learning. In 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 1-5). IEEE.

Evaluation of Online Teaching in the Covid Period using Learning Analytics

Jolana Gubalova

Department of Quantitative Methods and Information Systems
Faculty of Economics, Matej Bel University
Banská Bystrica, Slovakia

Abstract—**The article compares education at the Faculty of Economics Matej Bel University before the pandemic and during the coronavirus pandemic. At the same time, it tries to outline what the education will look like after this situation is over. It finds out how the situation during the corona affected the education of economists and to what extent the changes it brought will be preserved in the future. The comparison of face-to-face and distance learning in 2019 and 2020 was made. This is because teaching in 2019 was carried out in a "classic", face-to-face manner, and on the contrary, in 2020, after the closure of schools in March 2020, teaching at Matej Bel University was carried out only distance online method. To get the best possible view of the researched topic, several research methods were used: the examination of the LMS Moodle with using of various Learning Analytics tools and Questionnaire Research. The results showed that face-to-face education before the Covid pandemic and after this pandemic will no longer be the same because distance online education will also cause changes in face-to-face education in the post-pandemic period. Questionnaire research showed that up to 78% of part-time students and 61% of full-time students would like their study program to use elements of distance education in full-time study as well. Since this is a large group of students, their opinion will be considered in the future when fully returning to face-to-face teaching.**

Keywords—Distance online learning; learning management system; moodle; collaboration platform microsoft teams

I. INTRODUCTION

The COVID-19 pandemic in Slovakia is part of the global pandemic of this infectious disease. The first case was confirmed in our country on March 6, 2020. From March 16, 2020, schools were closed and at the same time a state of emergency began to apply, which lasted with smaller or larger breaks until February 22, 2022. However, the state of emergency remains in force. The pandemic caused a revolution in education at all levels of schools, including university education. It also significantly influenced the education of economists at Matej Bel University (MBU) in Banská Bystrica. Before the pandemic, it was a classic face-to-face education, which was carried out either in large lecture rooms or specialized ones, e.g., computer classrooms.

The Matej Bel Virtual University portal is available at MBU, based on the LMS Moodle, from 2012.

Moodle is a free software, a learning management system providing a platform for e-learning, and it helps the various educators considerably in conceptualizing the various courses, course structures and curriculum thus facilitating interaction with online students [19].

Teachers use LMS (Learning Management System) to publish teaching materials in courses for individual subjects and to communicate with students, but not everyone had used this occasion in the past. Various communication platforms were also sporadically used - e.g., Skype, Zoom, Google Hangouts, etc. for communication, mainly with external students. Some teachers went further and prepared multimedia teaching materials for students - e.g., video sequences with the solution of tasks in the Camtasia Studio program, which they published as part of courses in LMS Moodle or in YouTube. Therefore, after the outbreak of the pandemic and the complete transition to distance online education, they had no problem quickly adapting to the new situation, when all teaching was moved to the online space. After the state of emergency was established, every teacher had the obligation to teach online with the support of the communication platform MS Teams, which was introduced as a unified platform for the entire university according to the previously prepared schedule for face-to-face teaching [20]. They also had to publish their teaching materials for subjects in courses in LMS Moodle. Invaluable help in these difficult times was provided by the MBU Institute of Automation and Communication, which organized courses for teachers who needed to learn/improve their work with LMS Moodle and MS Teams.

The exam period also took place online with the support of LMS Moodle and MS Teams. Most of the exam dates looked like this: students logged into MS Teams, where they had a channel ready for the test and had to turn on their webcam. At the same time, they signed up for the subject course in LMS Moodle, where they had a test with tasks prepared. They downloaded file with tasks to the desktop and, after processing, uploaded them to the storage in the LMS Moodle. In addition, many tests were also prepared directly in LMS Moodle through the Test activity.

The article compares the features of face-to-face and distance online education and their suitability for university students with support of Learning Analytics tools and questionnaire research.

Learning Analytics is a research area that focuses on the use of quantitative methods in learning research in and outside of the virtual environment [2]. The first definition of Learning Analytics was published by G. Siemens (on his blog) in 2010,

which was used in a modified form at the 1st international conference focused on Learning Analytics in 2011. Learning analytics is the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs [3].

But it's not just access to data that helps us make smarter decisions, it's the way we analyze it. That's why it's important to understand the four levels of analytics: descriptive, diagnostic, predictive and prescriptive [22].

A. Descriptive Analytics

Descriptive (also known as observation and reporting) is the most basic level of analysis. This includes compiling reports and presenting what has happened in the past.

B. Diagnostic Analytics

Diagnostic analytics is where we get to the why. We move beyond an observation and get to the "what" that is making it happen. This is where the ability to ask questions about the data and tie those questions back to objectives is most important.

C. Predictive Analytics

Predictive analytics allows to predict different decisions, test them for success, find areas of weakness, make more predictions—and so forth. This flow allows institutions to see how the first three levels can work together. Predictive analytics involves technologies like machine learning, algorithms, and artificial intelligence, which gives it power because this is where the data science comes in.

D. Prescriptive Analytics

Prescriptive analytics exist at a very advanced level and is the most powerful and final phase, and truly encompasses the "why" of analytics. It's when the data itself prescribes what should be done. Data-driven decision making is tied most closely to predictive and prescriptive analytics, even though these are the most advanced.

All four levels create the puzzle of analytics: describe, diagnose, predict, prescribe. When all four work together, we can truly succeed with a data and analytical strategy.

II. LITERATURE REVIEW

A significant number of authors are engaged in research focused on the use of LMS Moodle with various type of plug-ins to improve the quality of teaching, especially in the Covid period. A greater number of authors use Learning Analysis tools for this purpose. Their work was a springboard for our research.

M. C. S. Manzanares at al. used a module UBU Monitor in Learning Management System for monitoring and detecting students at risk of dropping out during a lockdown caused by COVID-19 [6].

R. Kuo et al. designs a Moodle plug-in that not only can visualize students' learning behavior patterns from the log but also can cluster students into different groups based on their behavior patterns. The annotations made by the teacher can be a support for researchers to further analyze and design

mechanism and algorithm to automatically recognize and identify a student's characteristics and conditions like learning styles, preferences, at-risk, and potential required assistances via the features extracted from a learning pattern and notify the teacher or administrative staff automatically [7].

Research's aim of [8] is to compare Moodle LMS usage before, during and after the first wave of COVID-19 pandemic, what resources and activities were used and what future tendencies have come from this scenario. Results show a huge growth of the LMS platform usage during the first wave and an increase afterwards, showing that teachers and students behaviors have changed, and technologies can complement traditional on premises classes, improving teaching and learning methods.

The study [9] illustrates the development of a learning analytics dashboard that can improve learning outcomes for educators and students.

The work [10] reinforces that local educational data analysis is feasible, opens new ways of analyzing data without data transfer to third parties while generating debate around the "local technologies first" approach adoption.

Learning management systems (LMSs) that incorporate hypermedia Smart Tutoring Systems and personalized student feedback can increase self-regulated learning, motivation, and effective learning. LMS with hypermedia Smart Tutoring Systems in Moodle increased the effectiveness of student learning outcomes, above all in the individual quiz-type tests. It also facilitated personalized learning and respect for the individual pace of student-learning [11].

Paper [12] presents an empirical study and related activity system analysis regarding the implementation and use of Moodle specifically, and learning management systems in general, in problem-based learning. The research involved an exploration of the characteristics that defined use of Moodle at a Danish university, the reasons why Moodle was or was not used in specific contexts and the way in which Moodle use was perceived by students. The investigation uncovered several reasons for the lack of focus on problem-based learning in Moodle structures and content and explored them through the contradictions identified within the activity systems and between the double contextual frame surrounding the interacting activity system.

Study [13] sought to determine the impact of using Moodle in teaching university courses on students' future anxiety and psychological happiness. It shows that implementing Moodle technology into teaching had a positive impact in reducing future anxiety and increasing psychological happiness among university students.

Paper [14] presents research work conducted at the University in Sri Lanka, to solve facilitate students learning in fully online and blended learning environments using Learning Analytics. The system was designed as a Moodle Plugin. As a result of the system, students could track their current progress and performance compared to the peers, which helps to improve their motivation to engage more in the course. Also, the increased engagement in the course enhances the student's self-confidence since the student can see

continuous improvement of his/her progress and performance which in turn improves the student's grades.

Qualitative case study [16] explores the perceptions and experiences of embedded experts in a global learning community that occurred over a 12-year period. The study was designed using the Online Collaborative Learning Framework developed by the authors in 2006. The goal of the study was to provide a nuanced understanding of embedded experts in online discussion that engage in real world issues related to today's diverse and digital classrooms. From the thematic analysis of the data, the following three implications emerged: Purposeful selection of technology; orientation and supports for the experts; and design of an organic environment that fosters the development of community including embedded experts.

The study [17] examines the effects of interactive and learning structures enabled by different Learning Management Systems (LMS) on satisfaction and learner engagement in online courses. An LMS can support or hinder active engagement, meaningful connections between segments of the course, easy communication, and formative feedback by making it easier or more difficult for faculty to communicate course requirements, provide open-ended feedback, and place course elements that are used together contiguous to one another.

Drawing on design experience of developing learning analytics and inducting others into its use, the article [18] presents a model that researchers have used to address five key challenges they have encountered. In developing this model, they recommend: a focus on impact on learning through augmentation of existing practice; the centrality of tasks in implementing learning analytics for impact on learning; the commensurate centrality of learning in evaluating learning analytics; inclusion of co-design approaches in implementing learning analytics across sites; and an attention to both social and technical infrastructure.

The paper [28] examines how training support and LMS readiness factors influence the capability of faculty to adopt e-learning and student perceived benefits. The results reveal that training support and LMS readiness have a positive influence on the faculty's capability to adopt e-learning, which leads to enhancing students' perceived benefits. By identifying the factors that influence e-learning adoption, universities can provide enhanced e-learning services to students and support faculty through providing adequate training and powerful e-learning platform.

III. METHODOLOGY

The comparison of face-to-face and distance online education was carried out for years 2019 and 2020. The reason was the teaching in 2019 took place "classically", face-to-face, and on the contrary, in 2020, after the schools closed on 16/03/2020, teaching continued solely online. The year 2021 was different from this point of view and there was an irregular alternation of face-to-face and distance education depending on the current pandemic situation, and therefore it is not suitable for our research.

To get the best possible view of the researched topic, we combine a few research methods:

1) The examination of the LMS Moodle (version 3.11.8) with using of Learning Analytics tools:

a) Google Analytics,

b) Data analysis from the MySQL Standard Edition Database System (Data storage of LMS Moodle),

2) Questionnaire research.

During the data processing we followed the recommendations given in the sources [1], [4], [5], [15], [23].

Google Analytics is a website traffic analysis application that provides real-time statistics and analysis of user interaction with the website. Google analytics enables website owners to analyze their visitors, with the objective of interpreting and optimizing website's performance [21]. In this research, we use Google Analytics to explore data from LMS Moodle at MBU.

By default, LMS Moodle at MBU uses the relational database system MySQL (version 8.0.29) for data storage. Given that, we planned to prepare SQL queries in MS SQL Server 2019, we first converted the data from MySQL to MS SQL Server, using the tool MS SQL Server Migration Assistant for MySQL (SSMA) [26], [27].

SQL Server Migration Assistant (SSMA) is a free supported tool from Microsoft that simplifies database migration process from MySQL to SQL Server. SSMA automates all aspects of migration including migration assessment analysis, schema and SQL statement conversion, data migration as well as migration testing. This download includes a GUI client-based application to manage migration process. A separate extension pack will install functionalities in SQL Server to emulate MySQL features not natively supported in SQL Server [24].

We have more user experience with MS SQL Server than with MySQL, which is why we are more comfortable working with it. We used the Microsoft SQL Server Management Studio graphical interface to communicate with SQL Server.

SQL Server Management Studio (SSMS) is an integrated environment for managing any SQL infrastructure. SSMS provides tools to configure, monitor, and administer instances of SQL Server and databases [25].

The investigated database consisted of more than 460 tables and its size was 9.36 GB and its preview is in the Fig. 1.
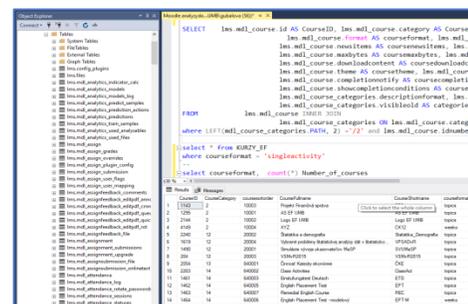
Fig. 1. The Moodle Database in the SSMS Tool.

IV. RESULTS AND DISCUSSION

The research of the LMS Moodle at MBU with support of Google Analytics tools shows how the number of users in LMS Moodle increased during the pandemic in 2020, compared to the previous year 2019, when "classic" face-to-face teaching was taking place.

As shown in Fig. 2, the curves had an approximately same course until March 2020. The turning point occurred on March 16, 2020, after the declaration of a state of emergency and the closing of schools, when teaching was fully moved into the online space and continued in this mode until the end of 2020.

The number of users in 2020 increased by 118.21% compared to the previous year 2019, and the number of new users increased by up to 132.08%. The number of sessions more than doubled - 210.67%. The number of sessions per user increased by 42.13%, and on the contrary, the number of exits - Bounce Rate decreased (-16.19%). The total number of displayed pages (Pageviews) increased almost 3.5 times (342.37%), while the average number of pages viewed by users during one session (Pages/Session) increased by 42.39% and the average duration of one session (Avg. Session Duration) increased by 42.13%.

The enormous increase in users results in higher hardware and software requirements for LMS and next learning-related systems to be reckoned with in similar crisis situations.

During both monitored years, users worked with LMS Moodle most often with the support of these browsers. Based on Fig. 3 we state that most sessions were made through the multi-platform Chrome browser - 69.63%. Another 10.89% of sessions were made through the Safari browser for the operating system Mac OS X and for Windows. The third in the order is the Firefox browser with a 7.88% share. It is followed by Edge with a share of 3.64%, which also has the lowest Bounce Rate - 18.13%. A slightly higher Bounce Rate have Firefox (19.27%) and Chrome (23.04%).

Browser	Acquisition			Behavior		
	Users	New Users	Sessions	Bounce Rate	Pages / Session	Avg. Session Duration
	405,354 % of Total: 100.00% (405,354)	402,380 % of Total: 100.06% (402,138)	3,274,375 % of Total: 100.00% (3,274,375)	24.02% Avg for View: 24.02% (0.00%)	8.45 Avg for View: 8.45 (0.00%)	00:07:38 Avg for View: 00:07:38 (0.00%)
1. Chrome	230,108 (56.98%)	229,388 (57.01%)	2,280,059 (69.63%)	23.04%	8.43	00:07:44
2. Safari	92,467 (22.90%)	91,374 (22.71%)	356,437 (10.89%)	30.37%	6.48	00:04:51
3. Firefox	30,317 (7.51%)	30,587 (7.60%)	257,970 (7.88%)	19.27%	10.92	00:09:59
4. Edge	14,694 (3.64%)	14,717 (3.66%)	179,081 (5.47%)	18.13%	10.51	00:09:47
5. Internet Explorer	12,229 (3.03%)	12,138 (3.02%)	24,891 (0.76%)	47.68%	8.86	00:06:19
6. Opera	8,412 (2.08%)	8,387 (2.08%)	119,215 (3.64%)	28.50%	7.79	00:07:46
7. Samsung Internet	6,375 (1.58%)	6,516 (1.62%)	37,887 (1.16%)	42.25%	4.88	00:03:15
8. Safari (in-app)	3,331 (0.82%)	3,329 (0.83%)	4,139 (0.13%)	52.55%	2.67	00:02:01
9. Android Webview	3,320 (0.82%)	3,405 (0.85%)	5,037 (0.15%)	52.11%	2.98	00:02:22
10. Mozilla Compatible Agent	839 (0.21%)	835 (0.21%)	842 (0.03%)	99.52%	1.01	00:00:02

Fig. 3. Browsers Overview in LMS Moodle via Google Analytics Tools (Jan. 1, 2019 – Dec. 31, 2020).

The Pages/Session indicator is the highest for Firefox (10.92), followed by Edge (10.51) and Internet Explorer (8.86) with Chrome (8.43). These browsers also have the highest Avg. Session Duration in this order: Firefox (00:09:59), Edge (00:09:47), Opera (00:07:46) and Chrome (00:07:44). Therefore, we note that these browsers are most used for online study and should be given more attention. Electronic learning materials available over the web (e.g., in LMS Moodle) should be adapted mainly to these browsers.

Other browsers for mobile devices (Samsung Internet, Safari (in-app), Android WebView or Mozilla Compatible Agent) are used to search for information in LMS Moodle, but not for e-learning because of their small screen. There is no need to prepare e-learning materials tailored to these browsers.

Fig. 4 shows the Devices Overview in LMS Moodle in the examined period.

Fig. 2. Audience Overview in LMS Moodle via Google Analytics Tools (Jan 1, 2020 – Dec 31, 2020, Compared to: Jan. 1, - Dec. 31, 2019).

Device Category	Acquisition			Behavior		
	Users	New Users	Sessions	Bounce Rate	Pages / Session	Avg. Session Duration
	405,354 % of Total: 100.00% (405,354)	402,380 % of Total: 100.06% (402,138)	3,274,375 % of Total: 100.00% (3,274,375)	24.02% Avg for View: 24.02% (0.00%)	8.45 Avg for View: 8.45 (0.00%)	00:07:38 Avg for View: 00:07:38 (0.00%)
1. desktop	222,699 (55.17%)	222,786 (55.37%)	2,497,557 (76.28%)	19.47%	9.44	00:08:55
2. mobile	175,643 (43.51%)	174,268 (43.31%)	760,057 (23.21%)	38.60%	5.23	00:03:28
3. tablet	5,346 (1.32%)	5,326 (1.32%)	16,761 (0.51%)	40.77%	5.78	00:04:40

Fig. 4. Devices Overview in LMS Moodle via Google Analytics Tools (Jan. 1, 2019 – Dec. 31, 2020).

We note that 76.28% of all sessions were made from desktop devices. The relatively high share of mobile devices (23.21%) was a surprise, so we will focus on it in the next analysis (Fig. 5). Tablets with a share of 0.51% represent rarely used devices.

The most widespread mobile platform is represented by Android-fcm (2360 devices), followed by iOS-fcm (2222), Android (265) and iOS (125).

```
SELECT [PLATFORM], COUNT(*) NUMBER_OF_DEVICES
FROM [LMS].[MDL_USER_DEVICES]
GROUP BY [PLATFORM]
ORDER BY 2 DESC
```

	platform	Number_of_Devices
1	Android-fcm	2360
2	iOS-fcm	2222
3	Android	265
4	iOS	125

Fig. 5. Overview of Mobile Device Platforms in LMS Moodle through Data Analysis of the Moodle Database (Jan. 1, 2019 – Dec. 31, 2020).

Fig. 6 shows the number of newly created courses during the entire operation of LMS Moodle at UMB in the years 2012 – 2022.

```
SELECT YEAR(DATEADD(SECOND, STARTDATE + 8*60*60,
'19700101')) [YEAR], COUNT(*) NUMBER_OF_NEWCOURSES
FROM LMS.MDL_COURSE
GROUP BY YEAR(DATEADD(SECOND, STARTDATE + 8*60*60,
'19700101'))
ORDER BY 1
```

	Year	Number_of_NewCourses
1	2012	72
2	2013	399
3	2014	156
4	2015	208
5	2016	229
6	2017	103
7	2018	109
8	2019	411
9	2020	2082
10	2021	1131
11	2022	11

Fig. 6. Comparison of the Number of Newly Created Courses in LMS Moodle in Individual Years.

We note that the number of newly created courses increased more than five times in the pandemic year 2020 (2082) compared to the previous year 2019 (411). Next, we analyze their formats in more detail in the Fig. 7.

We constate that the biggest increase (26 times) occurred in the format Collapse Topics (topcoll), but its share in the total number of course formats is minor (1.2%). By researching courses with this format, we found that the reason for choosing this type of format was the large number of sections with learning materials that needed to be expanded or collapsed for clarity. The second highest increase (15 times) was recorded by the Single Activity format, but its share is also minor (0.7%). Teachers used it to present one study obligation - e.g., term paper, case study or seminar paper. The Topics format follows with a 7.37-fold increase and a 46% share of the total number of course formats. The lowest increase occurred in the case of the Weeks format (3.88 times), which still dominated with its 52% share in the number of course formats. During the pandemic, course creators chose the Topics format more often than in the past.

```
--2019
SELECT [FORMAT], COUNT(*) NUMBER_OF_COURSES_2019
FROM LMS.MDL_COURSE
WHERE YEAR(DATEADD(SECOND, STARTDATE + 8*60*60,
'19700101'))=2019
GROUP BY FORMAT
ORDER BY 2 DESC
--2020
SELECT [FORMAT], COUNT(*) NUMBER_OF_COURSES_2020
FROM LMS.MDL_COURSE
WHERE YEAR(DATEADD(SECOND, STARTDATE + 8*60*60,
'19700101'))=2020
GROUP BY FORMAT
ORDER BY 2 DESC
```

	format	Number_of_Courses_2019
1	weeks	279
2	topics	130
3	singleactivity	1
4	topcoll	1

	format	Number_of_Courses_2020
1	weeks	1082
2	topics	958
3	topcoll	26
4	singleactivity	15
5	tiles	1

Fig. 7. Comparison Formats of Newly Created Courses in LMS Moodle in 2019 and 2020.

At the pandemic time, the testing and verification of knowledge has fully moved to the online space. As Fig. 8 shows, in 2020 the number of newly created quizzes increased 75.73 times compared to 2019.

```
SELECT YEAR(DATEADD(SECOND, TIMEOPEN + 8*60*60,
'19700101')) AS [YEAR], COUNT(*) NUMBER_OF_NEWQUIZZES
FROM LMS.MDL_QUIZ
GROUP BY YEAR(DATEADD(SECOND, TIMEOPEN + 8*60*60,
'19700101'))
ORDER BY 1
```

	Year	Number_of_NewQuizzes
1	2012	45
2	2013	16
3	2014	30
4	2015	46
5	2016	8
6	2017	27
7	2018	21
8	2019	15
9	2020	1136

Fig. 8. Comparison of the Number of Newly Created Quizzes in LMS Moodle in Individual Years.

Therefore, at the beginning of the pandemic, the LMS Moodle portal collapsed under the onslaught of testing activities. The schedule had to be prepared in such a way that several testing activities did not take place at the same time. The limit was max. 250 test users simultaneously.

Fig. 9 presents that in 2020 there was also a 26-fold increase in the number of Questionnaires, compared to 2019. By researching in the LMS Moodle, we found out that it was mainly about surveys of satisfaction with teaching, finding suitable dates for an exams or defense of seminar papers by students.

```
SELECT YEAR(DATEADD(SECOND, TIMEMODIFIED + 8*60*60,
'19700101')) AS [YEAR], COUNT(*)
NUMBER_OF_NEWQUESTIONNAIRES
FROM [LMS].[MDL_QUESTIONNAIRE]
GROUP BY YEAR(DATEADD(SECOND, TIMEMODIFIED +
8*60*60, '19700101'))
ORDER BY 1
```


	Year	Number_of_NewQuestionnaires
1	2012	19
2	2013	20
3	2014	86
4	2015	117
5	2016	8
6	2017	15
7	2018	7
8	2019	5
9	2020	130
10	2021	49
11	2022	1
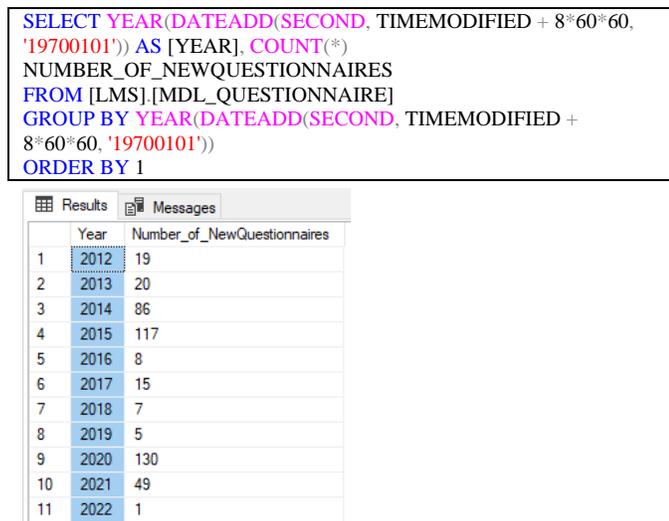
Fig. 9. Comparison of the Number of Newly Created Questionnaires in LMS Moodle in Individual Years.

After the end of the winter term, during the month of February 2021, we prepared an anonymous questionnaire in LMS Moodle for students, regarding the evaluation of the progress and organization of online teaching at Faculty of Economics MBU. There were 231 students (respondents) – 202 full-time and 29 externals, answered the questionnaire.

The questions Q1 - Q5 were aimed at surveying satisfaction with distance online education and utilized a five-point Likert scale (1-Very dissatisfied, 2- Rather dissatisfied, 3-Neutral, 4- Rather satisfied, 5-Very satisfied). For the analysis, a Likert score of 1-2 was regarded as a negative response, 4-5 as a positive response and 3 as a neutral response. The Likert scores for the questions Q1 - Q5 are presented in Table I.

We constate that the highest satisfaction was with the collaboration platform MS Teams (4.31), educational support of LMS Moodle (4.28) and with studying at Faculty of Economics MBU generally (4.05).

TABLE I. LIKERT SCORES FOR QUESTIONS Q1 – Q5

	How satisfied are/were you:	Likert Scores
Q1	with studying at Faculty of Economics MBU?	4.05
Q2	with the process and organization of distance online education in year 2020?”	3.90
Q3	with educational support of LMS Moodle	4.28
Q4	with the provision of teaching through collaboration platform MS Teams	4.31
Q5	with e-learning materials from teachers	3.14

Q7 investigated how students prefer individual forms of lectures with using of five-point Likert scale (1- I don't prefer it at all, 2- I don't prefer it, 3- neutral attitude, 4- I prefer it, 5- I strongly prefer it).

TABLE II. LIKERT SCORES – EVALUATION OF LECTURES

	Form of lectures	Likert Scores
L1	Publication of study materials in the LMS	4.82
L2	Online video lectures with recording	4.82
L3	Off-line video lectures (video recordings)	4.75
L4	Audio recording to the published presentation	3.25
L5	Online video lectures without recording	3.05

Based on the results in Table II, we conclude that students prefer persistent forms of lectures for which there is a record and to which they can return in the future if necessary.

Q8: “In what form were the lectures carried out?”

TABLE III. FORM OF LECTURES IN THE MONITORED PERIOD

Form of Lectures	Very Often & Often	Less Often & Not at all	I don't know
L1 - Publication of study materials in the LMS	79.65%	12.12%	8.23%
L5 - Online video lectures without recording	79.22%	12.55%	8.23%
L2 - Online video lectures with recording	28.14%	63.64%	8.23%
L4 - Audio recording to the published presentation	8.23%	83.55%	8.23%
L3 - Off-line video lectures (video recordings)	7.79%	83.98%	8.23%

It follows from Table III that in period of online teaching, lectures were very often provided to students in the form of publishing in LMS Moodle, the second most common form was “online video lectures without recording” and, to a much lesser extent – “online video lectures with recording”.

Q9 investigated which individual forms of seminars/exercises students prefer with using of five-point Likert scale (1- I don't prefer it at all, 2- I don't prefer it, 3- neutral attitude, 4- I prefer it, 5- I strongly prefer it).

According to Table IV, students prefer persistent forms of Seminars/Exercises (with recording).

TABLE IV. FORM OF SEMINARS/EXERCISES IN THE MONITORED PERIOD

x	Form of Seminars/Exercises	Likert Scores
S1	Online video seminars with recording	4.82
S3	Off-line video seminars (video recordings)	4.55
S4	Online video seminars without recording	3.35
S2	Assignments you could consult with teacher	3.25
S5	Published assignments with solution procedure	3.20
S7	Writing term papers/case studies/projects, etc.	3.16
S6	Posted assignments with results only	2,56

Q7: "In what form were the seminars/exercises carried out?"

TABLE V. FORM OF SEMINARS/EXERCISES IN THE MONITORED PERIOD

Form of Seminars/Exercises	Very Often & Often	Less Often & Not at all	I don't know
S4 – Onlin e video seminars without recording	75.76%	16.02%	8.23%
S7- Writing term papers/case studies/projects, etc.	57.58%	34.20%	8.23%
S2 - Assignments you could consult with teacher	50.22%	41.56%	8.23%
S5 - Published assignments with solution procedure	35.50%	56.28%	8.23%
S1- Online video seminars with recording	27.27%	64.50%	8.23%
S6 - Posted assignments with results only	21.65%	70.13%	8.23%
S3 - Off-line video seminars (video recordings)	9.96%	81.82%	8.23%

Table V presents that during online teaching, the seminars/exercises were carried out particularly in form of "online video exercises without recording", at second position

was "Writing term papers/case studies/projects" and at third position "Assignments you could consult with teacher".

The answers to questions Q8, Q9 and Q10 differed significantly for full-time and external (part-time) students, so we evaluate them separately.

Q8: "I acquired comparable knowledge as during face-to-face teaching."

Q9: "The method of evaluating subjects took into account the specifics of distance education."

Q10: "There is a lack of live contact with the teacher (the teacher usually sends texts and assignments; students send completed assignments)."

As for the answers, students could choose from the options: Majority of subjects, Minority of subjects, I don't know.

Based on the Fig. 10 and Fig. 11, we conclude that distance learning is perceived more positively by external students. It is also because the main way of their education in the past was self-study. The Covid era brought them a benefit in the form of a large amount of e-learning teaching materials and better support in LMS Moodle and MS Teams.
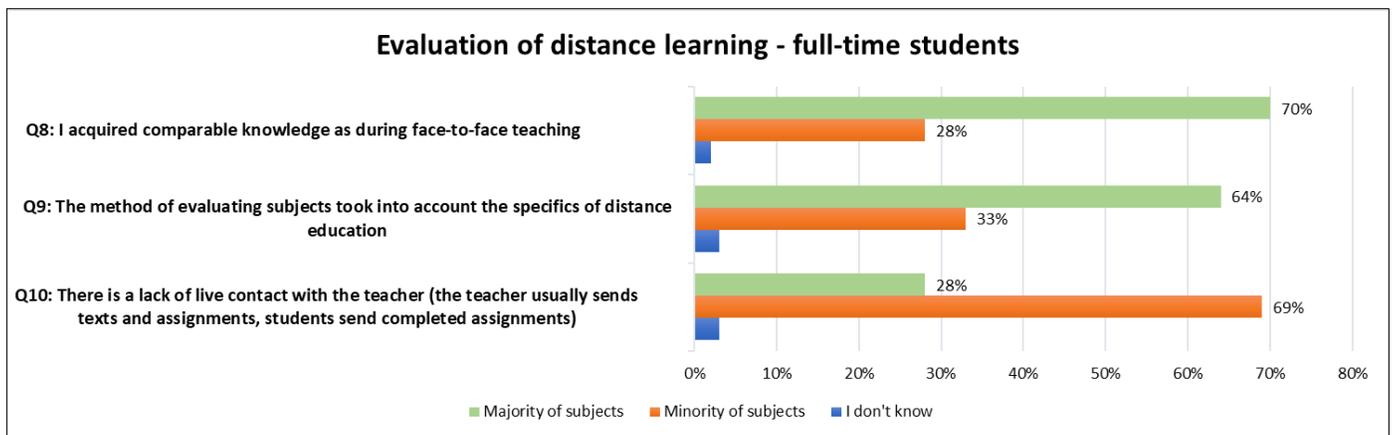
Fig. 10. Questions Q8, Q9, Q10 - Answers of Full-time Students.
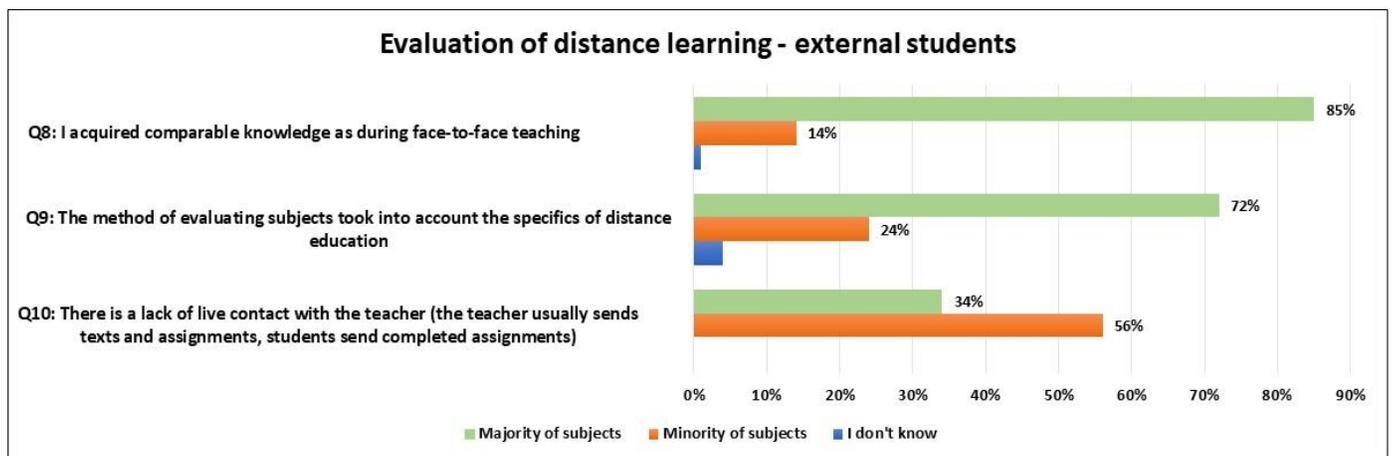
Fig. 11. Questions Q8, Q9, Q10 - Answers of External Students.

TABLE VI. ADVANTAGES AND DISADVANTAGES OF DISTANCE LEARNING

Advantages:	Disadvantages:
1 reduction of travel and accommodation costs (students could use the saved time for longer sleep, self-study or sports activities), 2 the opportunity to learn in the home comfort, 3 the possibility of education even with a weaker illness, 4 the possibility to replay the learning activity from the record, 5 the possibility to use new information and communication technologies in education, 6 there is no need to look for someone to supervise the children, as was the case with full-time studies (in case of external students).	1 absence of social contact with teachers and classmates, 2 inability to maintain long-term attention when working with a PC, 3 lack of motivation to study, procrastination, 4 health problems (e.g. vision impairment, back pain), 5 increased demands on technology, technical problems, especially poor internet connection on the part of students, 6 inability to understand the subject matter as in face-to-face studies, 7 less movement, 8 during online classes they can eat, look at their mobile phone, etc. When a person is in a comfortable environment, it is easy to lose focus, 9 external distractions (family, pets, couriers, housework).

Q11: "Even after the pandemic, my study program should use elements of distance education". We note that 78% of external (part-time) students and 61% of full-time students agreed with this statement.

Q12: Prepare list the advantages of the distance learning method from your point of view. Table V lists the most frequently mentioned advantages and disadvantages.

Questionnaire research showed that up to 78% of part-time students and 61% of full-time students would like their study program to use elements of distance education in full-time study as well. Since this is a large group of students, we will take their opinion into account in the future when fully returning to face-to-face teaching.

Based on the results of the research, we declare the following recommendations for face-to-face education in the post-Covid period:

- In time of building university information systems for e-learning (e.g., Learning Management Systems, Communication Platforms), emphasize their expandability and scalability (the system's ability to advantageously use additional resources such as processors, memory, or disk space) in case of need.

- To adapt the prepared e-learning teaching materials to the most used internet browsers and desktops.

- To enable students who cannot participate in face-to-face teaching (e.g., due to a minor illness) to participate in it online, via a communication platform - the so-called hybrid method of teaching.

- To prepare records from educational events (lectures, seminars/exercises) so that students can return to them in the future, e.g., when preparing for the exam.

- To provide external students with the option of participating in classes either face-to-face or online (and not only in the case of e.g., illness).

- Regularly implement training for teachers, focused on working with modern information and communication technologies (ICT), which are applied in teaching, so that the next pandemic or similar event does not catch them unprepared.

- To organize trainings of effective work with ICT for teachers and students.

V. CONCLUSION

The article deals with the comparison of face-to-face education and distance online education in Covid period. In frame of research, we combine a few research methods: the examination of the LMS Moodle with using of Learning Analytics tools: Google Analytics, Data analysis from the MySQL Standard Edition Database System and Questionnaire research.

The results showed that face-to-face education before the Covid pandemic and after this pandemic will no longer be the same because distance online education will also cause changes in face-to-face education in the post-pandemic period.

Future research is planned to determine the extent to which these recommendations will be implemented in the UMB environment and what their impact on education will be.

REFERENCES

[1] M. S. Brown, "Transforming Unstructured Data into Useful Information" chapter in Big data, Mining, and Analytics, pp. 213-215, Auerbach Publications, New York, 2014.

[2] L. Juhanak, J. Zounek, "Learning analytics: a new approach to exploring learning (not only) in virtual environments" in Pedagogical Orientation, ISSN 1805-9511, vol.26, No.3 (2016), Muni Ped Press, Brno, pp. 560-584.

[3] "Message from the LAK 2011 General & Program Chairs" in Proceedings of the 1st International Conference on Learning Analytics and Knowledge, February 27–March 1, 2011, Banff, Alberta, Canada.

[4] J.A. Larusson, B. White et al., Learning Analytics: From Research to Practice. ISBN: 978-1-4614-3305-7, New York Springer, 2014.

[5] F. J. García-Peñalvo, A. M. S. Pardo et al., Online Tutor 2.0: Methodologies and Case Studies for Successful Learning. ISBN: 97814666583252014. IGI Global, 2014.

[6] M. C. Sáiz-Manzanares et al., "Monitoring of Student Learning in Learning Management Systems: An Application of Educational Data Mining Techniques". In Applied Sciences, ISSN: 2076-3417, 2021, 11, 2677.

[7] R. Kuo et al., "Behaviour Analytics - A Moodle Plug-in to Visualize Students' Learning Patterns". In Proceedings Intelligent Tutoring Systems, 17th International Conference on Intelligent Tutoring Systems, ITS 2021, pp. 232–238.

[8] L. Pereira, J. Guerreiro, "Evaluation on Moodle LMS Data Usage During the First Wave of Covid-19's Pandemic". In Proceedings Universal Access in Human-Computer Interaction, HCII 2021, Part II, International Conference on Human-Computer Interaction, pp.154-166.

[9]   O. K. Xin, D. Singh, "Development of Learning Analytics Dashboard based on Moodle Learning Management System". In International Journal of Advanced Computer Science and Applications, Vol. 12, No. 7, 2021.

[10]   D. Amo et al., "A Privacy-Oriented Local Web Learning Analytics JavaScript Library with a Configurable Schema to Analyze Any Edtech Log: Moodle's Case Study". In Sustainability, Vol.13, Issue 9 - Special Issue "Information Systems, E-learning and Knowledge Management", 2021.

[11]   M. C. Saiz Manzanares et al., "Does the Use of Learning Management Systems With Hypermedia Mean Improved Student Learning Outcomes?". In Frontiers in Psychology, the Research Topic: Psychology, Technological Innovation, and Entrepreneurship, 2019.

[12]   R. Ørngreen et al., "Moodle and Problem-Based Learning: Pedagogical Designs and Contradictions in the Activity System". The Electronic Journal of e-Learning, Vol. 19 Issue 3, 2021, pp. 133-146.

[13]   R. Al-kreimeen, O. Murad, "Using Moodle in University Courses and Its Impact on Future Anxiety and Psychological Happiness". In The Electronic Journal of e-Learning, Vol. 20, No. 2, 2022.

[14]   R. Jayashanka et al., "Technology Enhanced Learning Analytics Dashboard in Higher Education", In The Electronic Journal of e-Learning, Vol. 20, No. 2, 2022.

[15]   J. Spirkova, P. Kral, "How to calibrate a questionnaire for risk measurement?", In Advances in fuzzy logic and technology 2017: proceedings of conference EUSFLAT 2017 and on international workshop IWIFSGN 2017, 11.-15.09.2017, Warsaw. Vol. 643. - 1. ed. - Heidelberg : Springer, 2018. - ISBN 978-3-319-66826-0. - ISSN 2194-5357. - Pp. 348-360.

[16]   J. Lock, P. Redmond, "Embedded experts in online collaborative learning: A case study". In The Internet and Higher Education, Vol. 48, January 2021.

[17]   B. Rubin et al., "The effect of learning management systems on student and faculty outcomes". In The Internet and Higher Education, Vol. 13, January 2010.

[18]   S. Knight et al., "Implementing learning analytics for learning impact: Taking tools to task". In The Internet and Higher Education, Vol. 45, April 2020.

[19]   Moodle, <https://moodle.org/?lang=sk> , [accessed 08.06.2022].

[20]   MS Teams, <https://www.microsoft.com/sk-sk/microsoft-teams/>, [accessed 09.06.2022].

[21]   Google Analytics, <https://analytics.google.com>, [accessed 08.06.2022].

[22]   PLURALSIGHT, "4 levels of analytics you need for better decision making". <https://www.pluralsight.com/blog/data-professional/data-informed-decisions>, [accessed 02.07.2022].

[23]   M. Kanderova, M. Boda, "Correcting the information ratio for first-order autocorrelation", In: Lecture notes in management science, management innovation and business innovation: 2nd international conference on management innovation and business innovation (ICMIBI 2014), December 8-9, 2014, Bangkok, Thailand / ed. Ying Zhang. - Singapore : Singapore management and sports science institute, 2014. - ISBN 978-981-09-1685-5. - ISSN 2251-3051. - Vol. 44 (2014), pp. 49-58.

[24]   SQL Server Migration Assistant for MySQL (MySQLToSQL), <https://docs.microsoft.com/en-us/sql/ssma/mysql/sql-server-migration-assistant-for-mysql-mysqltosql?view=sql-server-ver16>, [accessed 02.03.2022].

[25]   What is SQL Server Management Studio (SSMS)?, <https://docs.microsoft.com/sk-sk/sql/ssms/sql-server-management-studio-ssms?view=sql-server-ver16>, [accessed 01.03.2022].

[26]   MySQL, <https://www.mysql.com/>, [accessed 02.03.2022].

[27]   MS SQL Server, <https://www.microsoft.com/en-us/sql-server/sql-server-2019>, [accessed 02.03.2022].

[28]   A. S. Alshehri, S. A. Alahmari, "Faculty e-Learning Adoption During the COVID-19 Pandemic: A Case Study of Shaqra University". In International Journal of Advanced Computer Science and Applications, Vol. 12, No. 10, 2021.

Multi-Feature Extraction Method of Power Customer's Portrait based on Knowledge Map and Label Extraction

Wentao Liu*, Liang Ji

Big Data Center of State Grid Corporation Limited

Beijing 100053, China

Abstract—In order to realize the visualization of power customer characteristics and better provide power services for power customers, a multi-feature extraction method of power customer's portrait based on knowledge map and label extraction is studied. The power customer's portrait construction model is designed, which uses the knowledge map construction link to collect the power customer related data from the power system official website and database, and clean and convert the data; In the multi-feature analysis section, natural language processing technology is used to analyze the characteristics of power customers through Chinese word segmentation, vocabulary weight determination and emotion calculation; Based on the feature analysis results, the portrait label is extracted to generate the power customer's portrait. The power customer's portrait is used to realize the application of power customer's feature visualization, power customer recommendation, power customer evaluation and so on. The experimental results show that this method can effectively construct the knowledge map of power customers, accurately extract the characteristics of power customers, generate labels, and realize the visualization of power customer's portraits.

Keywords—Knowledge map; label extraction; power customer's portrait; multi-feature extraction; natural language processing; feature visualization

I. INTRODUCTION

At present, China's power grid has retained nearly 500 million customer data, which come from data collected by power grid energy meters, customer telephone service system data, power grid management data, etc. [1]. Based on these data, using big data analysis technology to analyze power customers, we can build a multi-level and multi-dimensional power customer's portrait [2]. The customer portrait can help relevant personnel quickly and accurately identify and recognize customers, and formulate targeted, refined and personalized service plans by quantifying sensitivity, so as to improve service quality and service efficiency [3].

The current research on portraits can be roughly divided into two categories according to the different objects of the portraits: one is the user portraits of scientific and technological experts, library users, website users and other characters, which are mainly used to recommend content to the research objects and improve the experience of the research objects; The other is the portrait of non-human things such as cities and books [4], which is mainly used to provide information services for groups other than the research object based on the portrait of things. Among them, the research on user portrait is relatively more extensive, and has been extended to various fields for different application scenarios. A lot of research has been done on the generation methods of user portraits at home and abroad. Wang and Zhu et al. used user portraits sharing representative opinions to predict emotion, and considered the impact of different characteristics of user groups on emotion analysis from three aspects: attribute characteristics, interest characteristics and emotional expertise [5]. Guo and Wei et al. used user portraits to automatically obtain user behavior data in web server logs in a big data environment by using a collection system; used PrefixSpan algorithm to build user portrait model, establish user behavior feature labels through frequent sequence mining, and realize abnormal feature extraction [6]. Chicaiza et al. used knowledge maps to build user portraits, and designed a recommendation system based on user portraits to achieve a good recommendation function [7].When the above three methods process massive power customer data, it is difficult to achieve rapid data extraction, and to provide better power service, and the application value is not high. Based on the previous research results, the multi-feature extraction method of power customer's portrait based on knowledge map and label extraction is studied. Knowledge atlas aims to describe the entities in the objective world we know and the direct relationships between entities in a structured form. It provides a better way to summarize and manage massive data in the Internet. The application of knowledge map is a good thing for the semantic search of the Internet. It plays an absolute role in promoting the development of semantic search. At the same time, it also shows its special and amazing ability in intelligent question answering. In the era of big data, knowledge map will have its place. Big data is the foundation and knowledge map is the tool. Combined with deep learning, it will develop better and better. Based on the analysis of user characteristics, tag extraction abstracts user portraits into user-friendly or computer-readable labels. The label has generality and condenses the key information in user characteristics. The content of user portrait labels is diverse, which can be words, phrases or concepts. The visualization methods of labels can be vectors, description charts and label clouds. Through the full analysis of user features and the accurate extraction of labels, we can achieve a deeper characterization of user portraits. By accurately constructing the power customer's

---

*Corresponding Author.

portrait, we can better provide power services for power customers.

II. MATERIALS AND METHODS

A. *Construction of Power Customer's Portrait*

In the context of big data, single data cannot fully reveal the characteristics of things, and the decision support services are insufficient and incomplete. Multi-source big data can describe things from different perspectives. Data complement and cross verify each other, breaking the "data island" [8]. Based on the understanding and elaboration of domestic and foreign scholars on the concept of power customer's portrait, it is considered that power customer's portrait refers to a model that can describe and characterize the features of power customers from different dimensions after abstracting and summarizing the basic statistical characteristics of power customers, power category characteristics, contact preference characteristics, power grid sensitive type characteristics and other characteristics from multi-source big data. On this basis, a model based on power customer's portrait is proposed, as shown in Fig. 1.
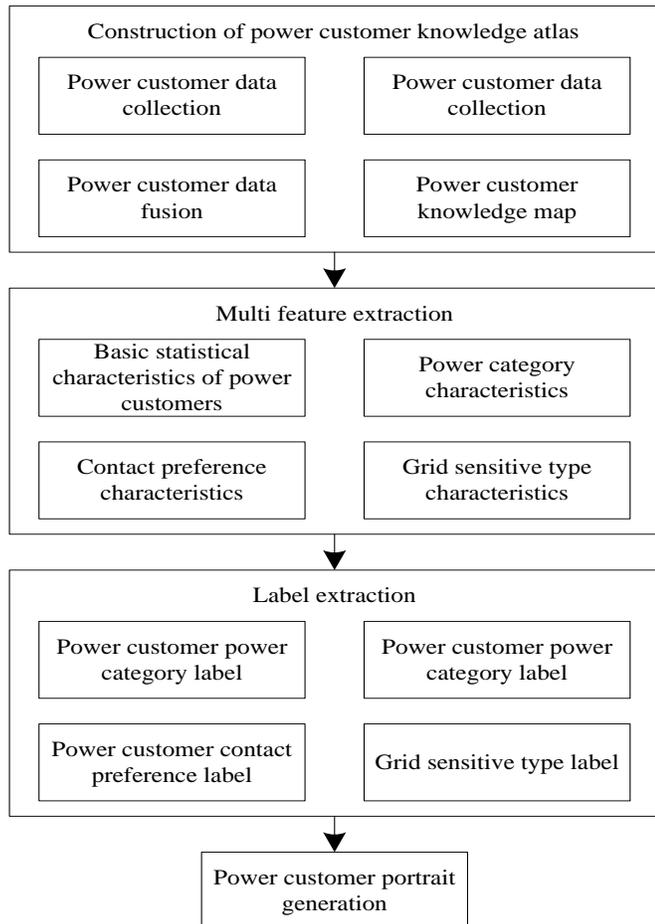
Fig. 1. Power Customer's Portrait Construction Model.

The power customer's portrait construction model can be roughly divided into three links: knowledge map construction, multi-feature analysis and portrait label extraction.

The knowledge map construction link collects relevant data of power customers from the power system official website, encyclopedia website and literature database; Data cleaning, data specification, data integration, data conversion and other data processing work are carried out for the relevant data of power customers. In the step of data fusion, the relevant data of power customers obtained after processing are stored according to the basic statistical characteristics of power customers, the characteristics of power categories, the characteristics of good contacts, the characteristics of power grid sensitive types and other categories to realize data processing; Multi-feature analysis links analyze the characteristics of power customers; Based on the feature analysis results, the portrait label is extracted to generate the power customer's portrait. The power customer's portrait is used to realize the application of power customer's feature visualization, power customer's recommendation, power customer's evaluation and so on.

B. *Construction of Power Customer's Knowledge Map*

The data extraction and processing of power customers in the process of building the knowledge map of power customers is to collect the relevant data of power customers from the power system website, encyclopedia website and relevant structured knowledge base, and build the original tourism route database.

The perfection of the construction of power customer's knowledge map directly affects the quality of the generation results of power customer's portraits [9]. Therefore, it is necessary to collect relevant data in the official website of the power system, encyclopedia websites and literature databases, and analyze the collected data through big data technology to determine that it meets the construction standard of knowledge map. On this basis, the collected power customer's related data is analyzed with the current existing data [10], and the data that meets the threshold standard is defined as valid data, which is stored in the original database.

Fig. 2 shows the construction process of power customer's knowledge map. The construction process of power customer's knowledge map can be roughly divided into four links:

1) Power customer data collection: The main function of this link is to collect relevant data of power customer's entities in various websites, and realize noise elimination through the preprocessing process in the data crawling process. On this basis, based on the differences of data types (structured, semi-structured and unstructured), the collected data are stored and processed, and the power customer entity database is constructed.

2) Power customer data extraction: The main function of this link is to use the data extraction model to extract different data needed to build the power customer's knowledge map from all kinds of data collected.

3) Power customer data fusion: The main function of this link is to use thesaurus to fuse the extracted data, and in this process, to complete the elimination of entity / attribute ambiguity and synonymous relationship merging [11].

4) Construction of power customer knowledge map: The main function of this link is to use the fused data to generate entity / attribute / relationship triples, so as to finally complete the construction of power customer knowledge map.
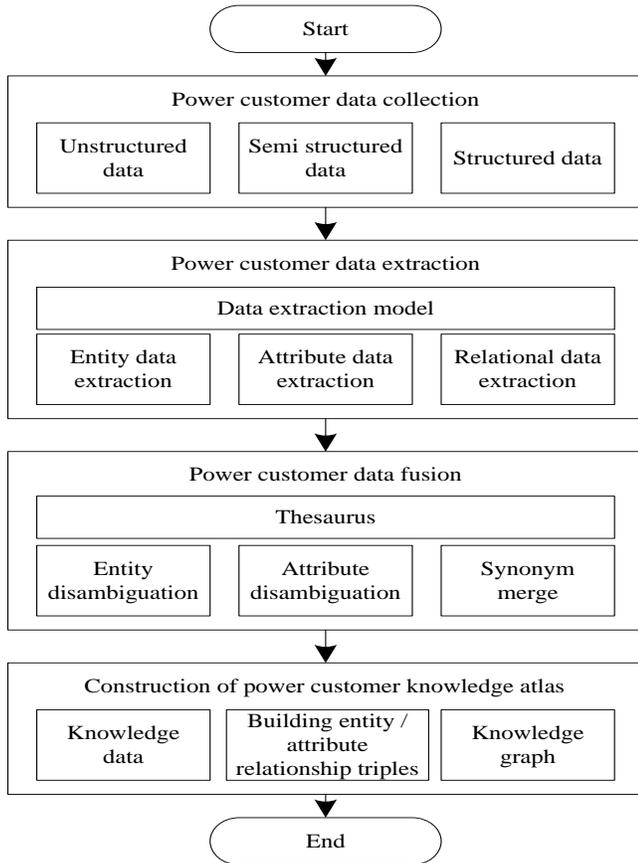
Fig. 2.   Construction Process of Power Customer Knowledge Map.

In the data extraction process, a timestamp incremental data extraction model based on variable window is used, which depends on the given time window for incremental data extraction. It is normal to assume that exceptions occur in the process of data extraction, and the occurrence of exceptions often leads to the inconsistency between the target table data and the source system data. To solve this problem, before incremental data extraction, it firstly gives two time windows (source time window and target time window, which are respectively used to extract data from various websites and power customer entity databases), then uses the given time window to extract corresponding data from various websites and power customer entity databases, and finally de-duplicates these data to produce the final result set. Considering the impact of abnormal occurrence on the consistency of power customer data in the process of power customer data extraction [12], it is necessary to use the data maintenance lead time to make minor adjustments to the given time

window before power customer data extraction. Here, the formal definition of timestamp incremental data extraction model with variable window is given:

$$M = \left(S_R, d_s, gmt, \Delta t, T_W, O_P, clean, d_u \right) \tag{1}$$

In equation (1), S_R is an n-tuple, $S = \left(s_1, s_2, \cdots, s_n \right)$ represents n data sources; d_s refers to the information database table of power customer, which requires an attribute column that marks the storage time of records; gmt represents the maximum value of the warehousing time of the data obtained from table d_s; Δt refers to the lead time of power user's data maintenance, which is used to adjust the size of the time window. T_W represents the time window information required for data extraction, T_W can be represented by a binary $T_W = \left(T_{W,S}, T_{W,T} \right)$, where $T_{W,S}$ represents the source time window and $T_{W,T}$ represents the target time window. In model M, the determination of $T_{W,S}$ and $T_{W,T}$ in T_W is very important to ensure the data consistency between the source system and the power customer's information database, especially when the data extraction process is started again after an exception occurs in the data extraction process, the determination of $T_{W,S}$ in T_W is directly related to whether the data that cannot be extracted due to the exception can be extracted. O_P refers to the operation defined on the data source, O_P is also a multivariate group, and the dimension of tuples is related to the operation defined on the data source, such as $O_P = \left(INNERJOIN_{s_1-s_2}, \cdots \right)$, which means that the power user data extracted from data source s_1 and data source s_2 will be internally connected. $clean$ represents a simple cleaning operation for the power user data extracted from S_R, that is, to remove the redundant time field value (the non-minimum data warehousing time of each record from S_R); d_u means to de duplicate the power user data extracted from S_R and d_s.

The working process of incremental data extraction model based on variable time window is described as follows:

1) Perform gmt operation on the target database table d_s to obtain the maximum value $\max Time$ of the storage time of power user data in d_s;

2) Set the value of the start time node of data extraction in $T_{W,S}$ and $T_{W,T}$ time windows to $\max Time$ respectively, and then adjust the time windows $T_{W,S}$ and $T_{W,T}$ according to the given data maintenance lead time to obtain $T_{W,S}{}'$ and $T_{W,T}{}'$;

3) Take $T_{W,S}{}'$ as the new time window, extract the data from the power user data source $S_{R,i}$, and operate the extracted power user data O_P, and then make $clean$ operation to remove the redundant time field values in the data to obtain the result set $tempRS\_S_R$;

4) Then take $T_{W,T}{}'$ as the target time window, extract the power user data from d_s, and get the result set $tempRS\_d_s$;

5) Perform d_u operation on $tempRS\_S_R$ and $tempRS\_d_s$ to obtain the final result set;

6) Load the final result set into d_s.

C. Multi-feature Extraction based on Natural Language Processing Technology

Multi-feature extraction of power grid sensitive customer's portrait under natural language technology, that is, using the methods of word segmentation, word vector conversion and word weight calculation in natural language processing technology to extract the features in the power grid sensitive customer's portrait label.

1) Chinese word segmentation: Letting the computer understand the label text information is the basis of multi-feature extraction of sensitive customer's portrait in power grid. There are a large number of Chinese words in the label text translated from Chinese. When using computer tools to comprehensively analyze the text information of portrait labels, the more common statistical word segmentation method is usually used [13] to preprocess the text information of portrait labels of power grid sensitive customers.

The N-ary grammar model in the statistical analysis method is selected. The conditions for the establishment of this model are: the production of the N-th word is affected by the first N-1-th word; There is no correlation with other words except this word and the N-1-th word; The probability of a whole sentence in the label text is the product of the production rates of different words in the sentence. Using the N-ary grammar model to judge the scientificity of the word segmentation plan is based on the probability value of the sequence of N words [14]. For example, suppose a sentence Y in the label text of power grid sensitive customer's portrait has two word segmentation plans, which are defined as Y_1 and Y_2 respectively, and the detailed process of word segmentation processing using the N-ary grammar model is as follows:

a) If the word order in Y is $C^{(1)}, C^{(2)}, C^{(3)}, \cdots, C^{(n)}$, it is expressed as $Y_1 = \left\{ C_1^{(1)}, C_1^{(2)}, C_1^{(3)}, \cdots, C_1^{(n)} \right\}$ and $Y_2 = \left\{ C_2^{(1)}, C_2^{(2)}, C_2^{(3)}, \cdots, C_2^{(n)} \right\}$ respectively.

b) The joint probability distribution expression of Y_1 and Y_2 is:

$$P(Y_m) = P\left(R_m^{(1)}\right) \times P\left(C_m^{(2)} \middle| C_m^{(1)}\right) \times P\left(C_m^{(3)} \middle| C_m^{(1)} C_m^{(2)}\right)$$
$$\times \cdots \times P\left(C_m^{(n)} \middle| C_m^{(1)} C_m^{(2)} \cdots C_m^{(n-1)}\right) \tag{2}$$

In equation (2), $P\left(C_m^{(n)} \middle| C_m^{(1)} C_m^{(2)} \cdots C_m^{(n-1)}\right)$, $P\left(C_m^{(n)}\right)$ and $P\left(C_m^{(n)} \middle| C_m^{(n-1)}\right) C_m^{(n)}$ respectively represent the joint probability distribution of sentence Y_m, the probability of word and the basis of word $C_m^{(n-1)}$. Through the public corpus, we can determine the probability of corresponding words, and then we can determine $P\left(C_m^{(n)}\right)$ and $P\left(C_m^{(n)} \middle| C_m^{(n-1)}\right)$.

c) The $P(Y_1)$ probability value is compared with the $P(Y_2)$ probability value, and the higher probability value is the correct word segmentation plan of the label text sentence T of power customer's portrait. After the word segmentation plan is determined, word vector conversion is performed on the label text.

2) Determination of vocabulary weight: Different words form sentences, and different sentences form the label text of the sensitive customer's portrait in power grid. The importance of each word in different label texts is different, and the importance of the same word in different label texts is also different. Therefore, the determination of vocabulary weight is very important for the extraction of multiple features of the sensitive customer's portrait in power grid. Statistical knowledge is used to determine the vocabulary weights that have been transformed into vectors in the text of sensitive customer's portrait labels in power grid, that is, the vocabulary weights are determined according to the text statistical information such as word frequency [15]. The above statistical process is based on Shannon's informatics theory: assuming that a certain word in all texts has a high word frequency, its information entropy is small; On the contrary, the lower the word frequency of a word in all texts is, the greater the information entropy is, that is, the inverse relationship between the word frequency and its information entropy.

As a typical information retrieval and data mining weighting technology, the term frequency-inverse document frequency (TF-IDF) algorithm can act on the vocabulary weight determination of the label text of the sensitive customer portrait in power grid [16], which is conducive to extracting the characteristic vocabulary in the label text information and realizing the multi-feature extraction of the power grid sensitive customer's portrait. The calculation process of TF-IDF algorithm is as follows, in which equation (3) and equation (4) calculate word frequency and inverse text frequency index respectively.

$$F_{i,j} = \frac{x_{i,j}}{\sum_k x_{k,j}} \times \ell$$

(3)

$$DF_i = \log \frac{T}{j : r_i \in t_j} \times \ell$$

(4)

$$F - DF = F_{i,j} \times DF_i \times \ell$$

(5)

In the above equation, $x_{i,j}$ and $\sum_k x_{k,j}$ respectively represent the occurrence frequency of word i in the power customer's portrait label text t_j and the total number of words in t_j; $|T|$ and $\left|\left\{ j : r_i \in t_j \right\}\right|$ respectively represent the number of all tagged texts in the corpus and the number of all tagged texts containing the word r_i in the corpus; ℓ represents the correction factor. Equation (3) is multiplied by equation (4), that is, equation (5) can exclude the vocabulary commonly existing in the text of power customer's portrait labels, and retain the vocabulary that can reflect a certain feature.

3) Emotion calculation: Using the feature words retained after the vocabulary weight is determined, the sentence emotion in the image label text of power grid sensitive customers can be determined. When determining sentence emotion based on vocabulary level, in order to ensure the efficiency of emotion determination, feature dimensionality reduction needs to be implemented, which will cause a large loss of feature words and lead to the deviation of emotion determination results. In order to avoid such problems, the Latent Dirichlet Allocation (LDA) model is used to determine the sentence emotion in the label text of the sensitive customer's portrait in power grid [17].

$T = \left\{ t_1, t_2, \cdots, t_T \right\}$ and $t_i = \left\{ y_1^i, y_2^i, \cdots, y_k^i \right\}$ are used to represent the text set of the power customer's portrait label and the sentence set in the text t_i respectively, so $T = \left\{ y_1, y_2, \cdots, y_N \right\}$ can be used to describe T as the set of all sentences in t_i, where N represents the number of all sentences in D.

The model of T is constructed by using LDA model. After parameter estimation, the distribution of sentences on characteristic topics and the distribution of characteristic topics on vocabulary are obtained respectively [18], and expressed by $\rho(\phi|\alpha)$ and $\rho(w_n|\phi,\partial)$, where ϕ, α and ∂ represent vectors, k-dimensional Dirichlet parameters and a $k \times v$ matrix respectively. Based on $\rho(\phi|\alpha)$ and $\rho(w_n|\phi,\partial)$, the multi-feature extraction of power customer's portrait is carried out. The specific process is as follows.

4) Density features: describe the maximum number of sentences with emotional words in the sentence, expressed by density (S).

5) Range features: describe the number of words in the sentence, expressed by range (S).

6) Quantitative features: describe the number of emotional words, expressed by the number (S).

7) Polarity features: q and $p_o(q)$ are used to represent emotional words and their polarity respectively. After calculation, it can determine the polarity of the sentence as:

$$p_o - s(S) = \sum_{i=1}^{count(S)} p_o(q_i) \times \beta$$

(6)

Where β is any constant.

8) Degree features: assuming the degree of q is $e_x(q)$, the degree of the sentence after calculation is:

$$e_x \_ s(q) = \max_{i=1,count(S)} e_x(q_i) \times \beta \tag{7}$$

Through Chinese word segmentation, vocabulary weight determination and emotion calculation, the basic statistical features of power customers, power category features, contact preference features and power grid sensitive type features are extracted. According to the feature extraction results, the basic statistical labels of power customers' portraits, power category labels, contact preference labels and power grid sensitive type labels are extracted.

D. Label Extraction and Portrait Generation of Power Customer's Portrait

1) Basic statistical label extraction of power customer's portrait.

$$D_e = \langle B_a, E_d, W_o, C_o \rangle \tag{8}$$

Where, B_a refers to the basic information of power customers, including name, gender and date of birth; E_d refers to the education experience of power customers, including education background, graduation college and graduation time; W_o refers to the work experience of power customers; C_o refers to the contact information of power customers, including office phone, mobile phone, email and personal web address.

2) Electricity category label extraction of power customer

$$S_p = \langle s_1, s_2, s_3, \ldots, s_n \rangle \tag{9}$$

Where, s refers to the power category feature word extracted from the power consumption data of power customers.

3) Contact preference label extraction of power customer

$$I_n = \{(t_1, tf_1), (t_2, tf_2), (t_3, tf_3), \ldots, (t_n, tf_n)\} \tag{10}$$

Where, t refers to the keyword in the contact preference of power customers, and tf_n refers to the number of times the preference word appears.

4) Extraction of grid sensitive type label

$$l_a = \langle l_{a,T}, R_d, R_{di} \rangle \tag{11}$$

Where, $l_{a,T}$ is the node label in the grid use partnership of power customer, R_d is the contribution of power customer d in the power grid use partnership, and R_{di} is the weight of the correlation between power customer d and power customer i.

Power customer's portrait is a label combination model of multi-dimensional and multi-level power customer [19]. According to the type of relevant data of power customers, it can define a vector space to represent the power customer's portrait [20]:

$$d = D_e, S_p, I_n, l_a \tag{12}$$

III. EXPERIMENTAL RESULTS

A. Experimental Environment

In order to verify the application effect of the multi-feature extraction method of power customer's portrait based on knowledge map and label extraction studied in this paper in the actual application process, the open data set of 2020 "Customer portrait" competition is taken as the experimental object, which includes the power use data of millions of grid customers in a province in 2019. This paper determines three types of information, a total of 12 fields (Table I) for the experiment.

The experimental environment is as follows: the server and client exist in the same computer, and the computer processor, memory and operating system are (Intel)i9-9900KF core eight core CPU, 8 GB memory and Windows 8 operating system respectively.

TABLE I. DESCRIPTION OF EXPERIMENTAL DATA

Experimental information category	Number of fields	Content description
Power system work order information	6	Name, gender, contact information, contact address, customer nature, etc
Registration form information	3	Business hall, store, communication reasons, etc
Data acquisition information of power meter	7	High energy consumption type, electricity fee guarantee type, total electricity consumption, paid in amount, payment, etc

B. Construction Results of Power Customer's Knowledge Map

The construction of power customer's knowledge map is the basis of power customer's portrait generation. According to the experimental data, the proposed method is used to build the customer knowledge map. Fig. 3 shows the customer knowledge ontology model built by this method.
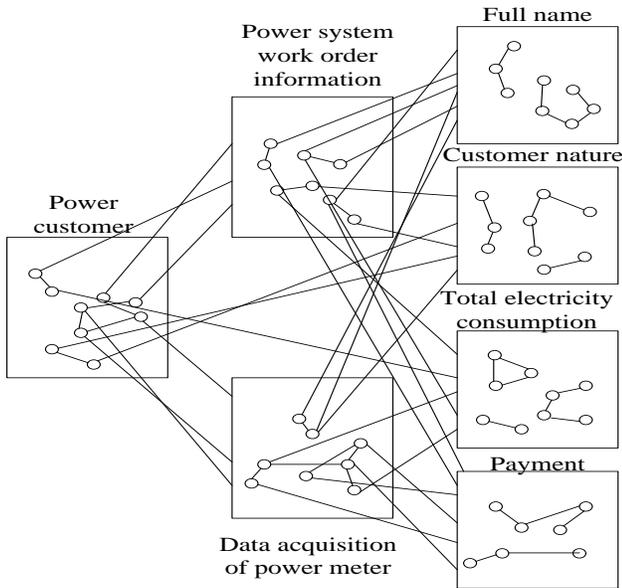
Fig. 3.   Experiment Object Knowledge Ontology Model.

The knowledge ontology model of power customer shown in Fig. 3 contains the work order information of the power system of the power customer and the data collection information of the power meter, in which the name, customer nature, total electricity consumption and payment are the low redundancy entity summaries generated by the proposed method. The different kinds of data contained in the body of knowledge come from experimental data.

C. Multi-feature Extraction Test of Power Customer

The multi-feature extraction of power customers is the most critical and time-consuming link in the generation of power customer's portraits. Therefore, the multi-feature extraction test of power customers is mainly analyzed from two aspects: extraction accuracy and extraction efficiency.

In the process of extracting accuracy test, we mainly use accuracy, precision rate, recall rate and F1 value as evaluation indicators to evaluate the effect of experimental feature extraction.

According to the combination of the actual feature extraction results of the experimental object and the extraction results of the proposed method, the real examples, false positive examples, true negative examples and false negative examples are defined. The confusion matrix of the feature extraction results during the experiment is shown in Table II.

TABLE II.   CONFUSION MATRIX OF FEATURE EXTRACTION RESULTS

Actual feature extraction results		Positive example	Counterexample
This method extracts the knot	Positive example	True example	False positive cases
	Counterexample	False counterexample	True counterexample

As the most critical and widely used evaluation index in the process of multi-feature extraction of power customers, accuracy can be understood as the proportion of the correct number of samples for feature extraction in the number of text samples of all power customers' portraits, and its expression is:

$$\text{Accuracy} = \frac{\text{ture positive example+false negative example}}{\substack{\text{ture positive example+false negative example+}\\ \text{ture negative example+false positive example}}} \quad (13)$$

The index describing the correct extraction probability of feature extraction into the positive example is the precision rate, and its expression is:

$$\text{Precision rate} = \frac{\text{ture positive example}}{\text{ture positive example+false positive example}} \quad (14)$$

The index describing the probability that the features in the positive example are correctly extracted is recall rate, which is expressed as:

$$\text{Recall rate} = \frac{\text{ture positive example}}{\text{ture positive example+false negative example}} \quad (15)$$

The harmonic mean of precision and recall is defined as F1 value, and its expression is:

$$\text{F1} = \frac{2 \times \text{ture positive example}}{\substack{2 \times \text{ture positive example−false positive example+}\\ \text{ture negative example+false negative example}}} \quad (16)$$

Among the above four indicators, there is an inverse relationship between recall rate and precision rat, that is, the higher the recall rate in the proposed method is, the lower the precision rate is. According to the prediction parameters of feature extraction, the evaluation results of multi-feature extraction of this method are obtained, as shown in Fig. 4.
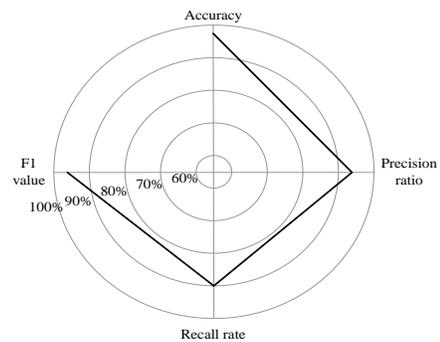
Fig. 4.   Performance Analysis of Multi-feature Extraction Results.

By analyzing Fig. 4, we can see that the comprehensive performance of the four index evaluation results of the proposed method has certain advantages, indicating that the multi-feature extraction effect of this method is good.

In order to test the feature extraction efficiency of the proposed method under the condition of different sample numbers, the number of feature texts to be extracted in the experiment is set to 1000, 3000, 5000 and 10000 respectively, and the time consumed by the proposed method under different text numbers is compared. The results are shown in Table III.

TABLE III.   NALYSIS OF MULTI-FEATURE EXTRACTION EFFICIENCY OF THIS METHOD

Elapsed time /s	Number of power customer information texts / piece			
	1000	3000	5000	10000
Chinese participle	6.33	18.97	54.16	97.85
Vocabulary weight determination	2.53	8.04	18.35	26.45
Affective computing	0.86	0.96	1.22	1.31

It can be seen from Table III that the time used for text emotion calculation is relatively short when the method is used to process the power customer information text. When the sample size is 10000, the running time is less than 1.31s; It takes a long time for Chinese word segmentation. When the sample size is 10000, the running time is about 97.85s, and it grows in gradient with the gradual increase of the number of power customer portrait label texts. Among them, the time consumed by this method in Chinese word segmentation and the determination of vocabulary weight fluctuates significantly, specifically in a linear rising state; The time consumed by emotional calculation fluctuates slightly. Since the determination of Chinese word segmentation and vocabulary weight is only conducted once in the initial stage, the impact of the above two steps on the subsequent feature extraction process is not significant. The experimental results show that the increase in the number of power customer portraits has no significant impact on the efficiency of multi feature extraction, indicating that the method is suitable for power grid big data environment.

D. Label Extraction Results

Different data sources such as power system registration form information, power system customer telephone system work order and power meter information collection are the data sources of label design of power customer's portrait. These data sources generally have the characteristics of massive, multi-directional and multi-dimensional content, high dispersion, non-uniform format and so on. To extract labels from these data sources, data association analysis and preprocessing should be carried out first. For example, by analyzing the customer information, guarantee information and charge control information in the power grid customer data, we can fully understand the attribute data of relevant types of power grid customers; By analyzing the work order information of the customer's telephone system, the actual electricity charge collection information, the electricity charge collection information, etc., we can understand the customer behavior data of the power grid. Only after fully understanding the data, it is convenient to extract the characteristics. Because the relevant data of the same power customer exists in different data tables, that is, there are several data of the same customer in different data tables, in order to better describe the power customer's portrait, it is necessary to integrate the multi-dimensional information of the power customer, and form the portrait label of the power customer based on the multi-feature extraction results of the power customer. The label structure is shown in Table IV.

TABLE IV.   POWER CUSTOMER LABEL EXTRACTION RESULTS

Label name	Label details
Basic statistics of power customer's portrait	Age
	Gender
	Industry
	Income
	Contact information
	Contact address
	Town / village
	Customer nature
Power customer power category	Important type
	High energy consumption type
	Multi power type
	Special line variant
	Electricity charge guarantee type
	Prepaid controlled
Power customer contact preferences	Store business hall
	Network business hall
	Terminal
	APP
	Telephone system
Grid sensitive type	Non sensitive
	Class A sensitivity
	Class B sensitivity
	Class C sensitivity
	Class D sensitivity

By analyzing Table IV, we can get that the labels in the relevant information of power customers can be extracted comprehensively and accurately by using the proposed method based on the multi-feature extraction results of power customers.

E. Visualization of Power Customer's Portrait

This paper uses Tagul to realize the visualization of power customer's portrait, imports the label of power customer's portrait into Tagul, and sets the size of the label according to the weight of the tag. The weight of each label in each label of power customers is set to 0.2, and the product of the frequency of each label in this feature and the feature weight are taken as the weight of the label. Because there may be the same label in several features, this paper adds the weight of the same label as the total weight of the label, and makes the power customer's portrait of "Li Boming", as shown in Fig. 5.

According to Fig. 5, it can get the portrait results of the power customer "Li Boming". According to this portrait, it can get the power customers who are more similar to the portrait of the power customer "Li Boming". On this basis, we can provide recommendations for these power customers who are more similar to "Li Boming", improve the allocation efficiency of power application resources, and bring more convenient power consumption experience and more effective power supply services to power customers.
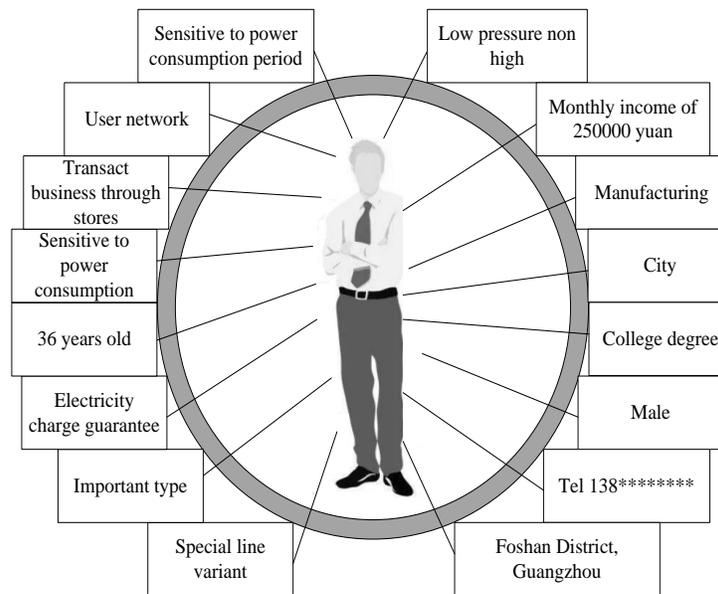
Fig. 5.   Visualization Results of Power Portrait.

IV. CONCLUSION

In this paper, a multi-feature extraction method based on knowledge atlas and label extraction is proposed, and experimental research is carried out on this basis. In this paper, a multi feature extraction method based on knowledge atlas and label extraction is proposed, and experimental research is carried out on this basis. This method can realize the multi feature extraction of power users, and the calculation efficiency is high. When the sample size is less than 10000, the time for emotion calculation is less than 1.31s, the time for Chinese word segmentation is less than 97.85s, and the time for vocabulary measurement is less than 26.45s, which has certain research value. The innovation of this method lies in:

1) A construction model of power customer's portraits based on knowledge map and label extraction is proposed, and the extracted labels are used to calculate the similarity between power customer's portraits, which provides a research idea for personalized service recommendation and accurate power supply services to power customers through the similarity of power customer's portraits.

2) It combs the data sources, main contents, data types and collection methods of building power customer's portraits. By integrating multi-source data, it depicts the characteristics of power customers as people who enjoy power services in the actual power application process from a more comprehensive perspective, enriches the connotation of power customer's portraits, presents power customer's portraits in a visual way, and promotes the innovative development of smart power.

However, there are some limitations in the research of this paper. In the future, two main works will be carried out: expanding the sample to the data of all power customers, further enriching and constructing the group of power customer's portraits; using PLSI, LDA, BTM and other data mining methods in topic mining to mine information in unstructured text data in various forms.

REFERENCE

[1]   H. Liang, and J. Ma, "Data-driven resource planning for virtual power plant integrating demand response customer selection and storage," IEEE Transactions on Industrial Informatics, 2021, PP(99), pp. 1-1.

[2]   F. Rahdari, N. Movahhedinia, M.R. Khayyambashi, and S. Valaee, "Qoe-aware power control and user grouping in cognitive radio ofdm-noma systems," Computer Networks, 2021, 189(2), pp. 107906.

[3]   Z. Zhao, D. Wang, H. Zhang, and H. Sang, "Joint user pairing and power allocation scheme based on transmission mode switching between noma-based maximum ratio transmission and mmse beamforming in downlink miso systems," Mobile Information Systems, 2021(3), pp. 1-21.

[4]   J. Fei, Q. Yao, M. Chen, X. Wang, and J. Fan, "The abnormal detection for network traffic of power iot based on device portrait," Scientific Programming, 2020(9), pp. 1-9.

[5]   B. Wang, E. Wang, Z. Zhu, Y. Sun, Y. Tao, and W. Wang, "An explainable sentiment prediction model based on the portraits of users sharing representative opinions in social sensors," International Journal of Distributed Sensor Networks, 2021, 17(10), pp. 3323-3330.

[6]   N. Guo, R. K. Wei, and Y. P. Shen, "Abnormal Feature Extraction Method in Large Data Environment Based on User Portrait," Computer Simulation, 2020, 37(8), pp. 332-336.

[7]   J. Chicaiza, and P. Valdiviezo-Diaz, "A comprehensive survey of knowledge graph-based recommender systems: technologies, development, and contributions," Information (Switzerland), 2021, 12(6).

[8]   J. Zhang, W. Huang, D. Ji, Y. Ren, "Globally normalized neural model for joint entity and event extraction," Information Processing & Management, 2021, 58(5), pp. 102636.

[9]   H. J. Kim, J. W. Baek, and K. Chung, "Associative knowledge graph using fuzzy clustering and min-max normalization in video contents," IEEE Access, 2021, PP(99), pp. 1-1.

[10]   J. Li, S. Liu, A. Liu, and R. Huang, "Knowledge graph construction for sofl formal specifications," International Journal of Software Engineering and Knowledge Engineering, 2022, 32(04), pp. 605-644.

[11]   J. T. Ma, J. W. Yan, t. C. Xue, and Q. Q. Ya, "Ngdcrm:a numeric graph dependency-based conflict resolution method for knowledge graph," 2021, 27(2), pp. 10.

[12] Z. H. Han, X. S. Chen, X. M. Zeng, Y. Zhu, and M. Y. Yin, "Detecting proxy user based on communication behavior portrait," The Computer journal, 2019, 62(12), pp. 1777-1792.

[13] X. Li, K. Zhang, Q. Zhu, Y. Wang, and J. Ma, "Hybrid feature fusion learning towards chinese chemical literature word segmentation," IEEE Access, 2021, PP(99), pp. 1-1.

[14] S. Ying, W. Li, B. He, W. Wang, and Y. Wan, "Chinese segmentation of city address set based on the statistical decision tree," Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomatics and Information Science of Wuhan University, 2019, 44(2), pp. 302-309.

[15] Z. Kong, C. Yue, Y. Shi, J. Yu, and L. Xie, "Entity extraction of electrical equipment malfunction text by a hybrid natural language processing algorithm," IEEE Access, 2021, PP(99), pp. 1-1.

[16] F. Dornaika, A. Baradaaji, and Y. E. Traboulsi, "Semi-supervised classification via simultaneous label and discriminant embedding estimation," Information Sciences, 2020, 546(1).

[17] Z. Taskin, and U. Al, "Natural language processing applications in library and information science," Online Information Review, ahead-of-print 2019, (4), pp. 676-690.

[18] L. Niu, J. Cai, A. Veeraraghavan, and L. Zhang, "Zero-shot learning via category-specific visual-semantic mapping and label refinement," IEEE Transactions on Image Processing, 2019, 28(2), pp. 965-979.

[19] Z. C. Sha, Z. M. Liu, C. Ma, and J. Chen, "Feature selection for multi-label classification by maximizing full-dimensional conditional mutual information," Applied Intelligence, 2021, 51(22).

[20] T. Sun, C. Zhang, Y. Ji, and Z. Hu, "A latent-label denoising method for relation extraction with self-directed confidence learning," Intelligent Data Analysis, 2020, 24(1), pp. 101-117.

Student Acceptance Towards Online Learning Management System based on UTAUT2 Model

Masitah Musa[1], Mohd. Norasri Ismail[2], Suhaidah Tahir[3], Mohd. Farhan Md. Fudzee[4], Muhamad Hanif Jofri[5]

Department of Multimedia, Faculty of Computer Science and Information Technology[1, 2, 4]
Universiti Tun Hussein Onn Malaysia, Johor, Malaysia[1, 2, 4]
Faculty of Education & Liberal Studies, City University Malaysia, Selangor, Malaysia[3]
Department of Information Technology, Centre for Diploma Studies, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia[5]

Abstract—**Recently, education has changed from physical learning to online and hybrid learning. Furthermore, the outbreak of COVID-19 makes them more significant. An online learning management system (LMS) is one of the most prevalent approaches to online and distance learning. The acceptance of the students towards the LMS is significant and it can give either bad or good responses to ensure the success of LMS. However, the Universiti Tun Hussein Onn Malaysia (UTHM) has not yet implemented any study to examine their LMS. The Unified Theory of Acceptance and Usage of Technology (UTAUT2) model is used in this study to investigate students' Behavioral Intention and Use Behavior when using the LMS in UTHM. This study also introduces a new construct in UTAUT2 named Online Learning Value. 376 respondents took part in this survey. Descriptive Statistics, Reliability Analysis, Pearson Correlation Coefficient, and Multiple Linear Regression analysis were all used to analyze survey data. The outcome of this research is Performance Expectancy ($\beta=0.129$, $p=0.014$), Hedonic Motivation ($\beta=0.221$, $p=0.000$), Online Learning Value ($\beta=0.109$, $p=0.036$) and Habit ($\beta=0.513$, $p=0.000$) has influence on students' intention to use LMS. Besides that, Facilitating Conditions ($\beta=0.481$, $p=0.000$) are the most important factors in students' use behavior toward the LMS followed by Habit ($\beta=0.343$, $p=0.000$) and Behavioral Intention ($\beta=0.239$, $p=0.000$). By utilizing the UTAUT2 model, the constructs of technology acceptance related to students' adoption of LMS have been identified and may become a reference to the stakeholders for future enhancement.**

Keywords—*Online learning management system; technology acceptance; unified theory of acceptance and usage of technology 2; online learning value*

I. INTRODUCTION

Currently, almost all universities in Malaysia rely on their online learning management system (LMS) to support academic activities including teaching and learning. Besides, online learning has become the best solution for students and lecturers in academic activities during the COVID-19 worldwide outbreak. The pandemic has forced changes in new normal academic activities where the institutions have the potential to build their own LMS [1]. The LMS is considered one of the best approaches toward online learning [2], [3] which provides benefits to the students and lecturers where it provides unlimited access to the learning materials, can track student progress and performance from assessment activities, organize the sources into the same shared data center, collaboration and discussion, and provides a different experience for both students and lecturers [4].

The use of LMS also is convenient and can boost students' efficiency throughout the process of teaching and learning because this system is compatible with any device, such as a smartphone or laptop [5]–[7]. Hence, students can gain knowledge and information by joining the learning management system that has been provided by the university at anytime and anywhere. There are many campuses that had been developed and used their own LMS including Universiti Tun Hussein Onn Malaysia (UTHM) [8].

In UTHM, an online LMS named Author has been implemented and was used before the COVID-19 pandemic as a mixed or blended learning tool. This LMS has been the primary tool for instruction and learning throughout recent years. However, the acceptance and usage of LMS in UTHM have not been empirically investigated. Due to the recent increases in the usage of LMS, it is valuable to understand and distinguish the factors that influence students' acceptance of LMS usage. Therefore, this research is important to examine the variables influencing students' Use Behavior as well as their Behavioral Intention of using the LMS. The study's research question is structured as shown in the following Table I.

TABLE I. RESEARCH QUESTIONS

No.	Research Questions
RQ1	What are the factors influencing UTHM students' behavioural intention and use behaviour towards using the LMS?
RQ2	Which main variables influence students' use behaviour towards using the LMS in UTHM?
RQ3	Does online learning value construct influence students' behavioural intention toward using the LMS in UTHM?

This paper is arranged as follows: The subsequent part will go through relevant research, namely the UTAUT2 model used in this study. Afterwards, the following section presents the research method where one new variable of UTAUT2 is proposed. Following that, the research's findings and analyses will be discussed. Finally, we address the conclusion and future works in this study.

II. RELATED WORKS

In order to obtain fewer risks during the adoption of new technology, many theories and models have been developed by previous researchers according to technology acceptance and usage.

A. Models Related to Technology Acceptance and Usage

There are eight user acceptance models: the Theory of Reason Action (TRA), the Technology Acceptance Model (TAM), the Motivational Model (MM), the Theory of Planned Behavior (TPB), the Combined TAM and TPB (C-TAM-TPB), the Model of PC Utilization (MPCU), the Innovation Diffusion Theory (IDT), and the Social Cognitive Theory (SCT). These models were then combined into a new framework called as the Unified Theory of Acceptance and Use of Technology (UTAUT)[9]. The UTAUT model is a well-improved model that can explain the technology acceptance behavior as it combines the eight models (TRA, TAM, MM, TPB, C-TAM-TPB, MPCU, IDT, and SCT) to form the determinants of behavioral intention and use behavior towards technology [10]. In this study, we utilized the UTAUT2 model, an expanded version of UTAUT, to investigate student acceptance of online LMS at UTHM. The UTAUT2 model focuses on the individual perspective of technology adoption compared to the UTAUT model. Therefore, this model will give a better understanding of technology acceptance by users.

The constructs in the UTAUT model are Performance Expectancy, Effort Expectancy, Social Influence, and Facilitating Conditions. An expanded edition of the UTAUT model with additional constructs and relationships was proposed to give an effective acceptance model in the consumer use context [11]. The added constructs in the UTAUT2 model are Hedonic Motivation, Price Value, and Habit. We proposed the Online Learning Value construct as an extension of the UTAUT2 in this study. This construct was classified into two types: dependent variables and independent variables. Use Behavior and Behavioral Intention are the dependent variables, whereas the independent variables are Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions, Hedonic Motivation, Online Learning Value, and Habit.

Use Behavior explains the constancy of the technology usage and the degree to which technology may improve a user's knowledge and skills[12]. Use Behavior also is referred to the level of technology use among students such as their use of mobile devices as learning aids [13].

The adherence of the user to keep using the system and how long the user intends to use the system is referred to as Behavioral Intention [12]. Furthermore, Behavioral Intention is the extent to which students expect to utilize technology such as smartphones in their studies and how long they will continue to do so [13]. Also, it can be referred to person's willingness to engage in a specific behavior [14]. Behavioral Intention also is described as individuals' intents to replace traditional methods with the new systems in the future [15].

Performance Expectation is the degree to which an individual believes that their work performance would be improved when utilizing a system [9]. It is also specified as the expectation of performance improvement due to the use of technology [16]. Similarly in education, it is also defined as the extent of a belief that their performance in academics will be improved by using LMS [17].

The level of easiness associated with the application of a system is characterized as Effort Expectancy [9]. Therefore, the student's level of education and information technologies knowledge will not involve any kind of physical or mental exertion in the usage of technologies in their studies [13]. It is also mentioned that the perceived easiness of using LMS is considered Effort Expectancy [18].

The degree to which an individual considers that essential persons think they have to utilize technology is referred to as Social Influence [9]. Furthermore, the degree to which students believe that important persons, such as colleagues, friends, and university lecturers, think they must apply technologies such as smartphones in their studies is also referred to as Social Influence [13]. Next, the social impact is the belief that others' perspectives on utilizing online learning management systems are essential for instructional activities [18].

The amount to which someone believes that the adoption of technology comes with technological and organizational infrastructure is referred to as the Facilitating Conditions [9]. It also refers to the tools and assistance available for using technology [16]. Facilitating Conditions also is defined as students' belief that technology like smartphones can be used by students as supplemental learning tools in their coursework since there is adequate organizational and technical infrastructure [13].

Hedonic Motivation is described as happiness or joy brought on by technology [11]. It is also described as the perceived enjoyment acquired when using LMS in the education sector [18]. Previous studies also stated that within the framework of mobile learning adoption and usage, Hedonic Motivation is referred to envisioned as a sense of enjoyment [16].

Habit is defined as a human action that is repeatable due to knowledge [11]. Habit also is referred to the extent to which the student uses the online learning management system platform automatically [18]. Previous studies also mentioned that the extent to which a user feels their technology usage is automatic or instilled is referred to as habit [16].

Learning Value is referring to an intellectual exchange among the alleged advantages of the applications and the period and effort spent on utilizing them [19]. Learning Value also is related to the student's learning achievement by using a smart campus [20].

B. Students' Acceptance towards Technology Adoption in Campus using UTAUT2

There are many adoptions of technology on campuses to provide better life experiences to the campus community, especially students. As such, research has been made to study student acceptance of technology adoption on campus by using the UTAUT2 model. Table II provides an overview of related studies regarding to the students' acceptance towards technology in campus.

TABLE II.   SUMMARY OF RELATED WORKS

Author & Year	Domain of Measure	Variables	Results
Ali et al. (2016) [21]	Factors influencing how well students in hospitality and tourism institutions embrace and use computer-based collaborative classrooms	Performance expectancy, effort expectancy, social influence, facilitating conditions, price value, hedonic motivation, habit	The admission of pupils is significantly influenced by all the elements. The way that students utilize technology is determined by their intention, their habits, and the facilitating conditions that are in place.
Farooq et al. (2017) [22]	Relationship between the UTAUT2's current notions of personal innovativeness (PI), intention, and use behavior regarding lecture capture systems (LCS)	Performance expectancy, effort expectancy, social influence, facilitating conditions, price value, hedonic motivation, habit, personal innovativeness	The acceptability and usage of LCS by students is significantly influenced by all variables.
Arain *et al.* (2019) [16]	Factors influence mobile learning acceptance in context of higher education	Performance expectancy, hedonic motivation, habit, ubiquity, satisfaction, information quality, system quality, appearance quality, effort expectancy, facilitating conditions, social influence	Performance expectancy, hedonic motivation, habit, ubiquity, satisfaction have the significant impact on students' behavioral intention. Information quality, system quality and appearance quality have significant impact on satisfaction of the students. Effort expectancy, facilitating conditions and social influence does not have significant impact on students' behavioral intention towards mobile learning acceptance.
Samsudeen, & Mohamed (2019) [15]	Factors that affect how university students in Sri Lanka intention and use behavior when using e-learning systems	Performance expectancy, effort expectancy, social influence, hedonic motivation, facilitating conditions, work life quality, internet experience	Students' behavioral intentions are extensively influenced by performance expectancy, effort expectancy, social influence, work life quality, internet experience and hedonic motivation. The LMS usage is influenced by behavioral intention and facilitating conditions.
Sharif et al. (2019) [20]	Combining the UTAUT2 and Task Technology Fit (TTF) model to analyze behavioral intention while adopting learning management systems.	Social influence, facilitating conditions, hedonic motivation, learning value, habit, performance expectancy, effort expectancy, task characteristics, technology characteristics.	The acceptance of LMS is significantly influenced by social influence, facilitating conditions, hedonic motivation, learning value and habit. However, the intention to employ a LMS is not significantly impacted by performance expectancy and effort expectancy. TTF will motivate students as well as influences user-friendliness and performance.

Based on the UTAUT2 model, variables influencing hospitality and tourism students' acceptance and use of classroom technology are explained by Ali et al. [21]. According to the study, performance expectancy, effort expectancy, social influence, facilitating conditions, price value, hedonic motivation, and habit all have a substantial impact on student acceptance. Furthermore, students' use behavior differs depending on their intention to use and their habit of using classrooms technology, as well as the facilitating conditions accessible to the students.

Next, pivotal correlation between existing constructs of UTAUT2, personal innovativeness (PI) as the new variables extends in UTAUT2, intention and use behavior towards lecture capture systems (LCS) presented by Farooq et al. [22]. The studies also revealed that student acceptance and usage of LCS are significantly influenced by performance expectancy, effort expectancy, social influence, facilitating conditions, price value, hedonic motivation, and habit. Personal innovativeness, a new variable included to the UTAUT2 model, was also found to have a substantial positive influence on student acceptance and usage of LCS. It has a large influence on students' intention and use behavior toward the LCS.

Sharif et al. [20] integrate the UTAUT2 and the Task Technology Fit (TTF) model, to describe individuals' behavioral intention to embrace learning management. The findings revealed that social influence, facilitating conditions, hedonic motivation, learning value, and habit had a substantial impact on the acceptability of the LMS in the UTAUT2 model. However, there was no significant influence of performance expectancy or effort expectancy on the desire to use a LMS. Furthermore, the TTF has an impact on user-friendliness and performance in addition to motivating student adoption of the system. Hence, the relationship between UTAUT2 and TTF will provide advanced performance and effort expectancy when students use the technology based on a match between its characteristics and the task's requirements.

III. PROPOSED METHODOLOGY

The UTAUT2 model was used to design a questionnaire in accordance with the recommendations from authors of earlier research [13], [20] to study student acceptance of using the LMS in UTHM. The research model employed in this work is depicted in the following Fig. 1.
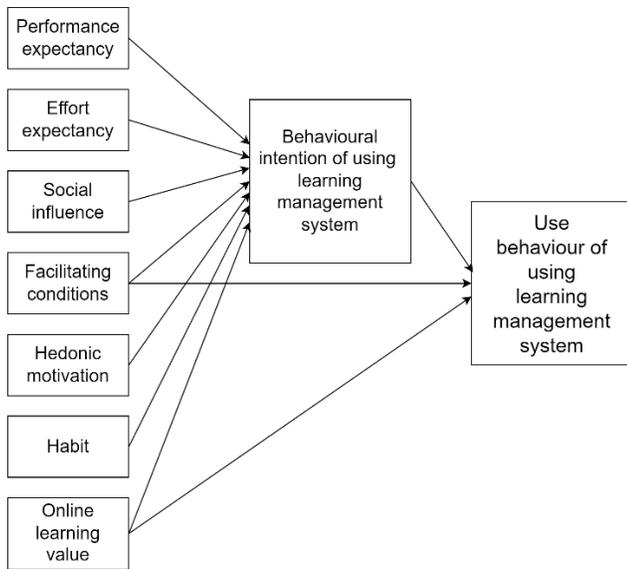
Fig. 1. Extended UTAUT2 Research Model.

A. Questionnaire Design

Generally, this study was conducted by distributing the questionnaire through an online survey to the participants. The respondents for this study are UTHM students as we need to study the acceptance of LMS among UTHM students. The questionnaire consists of two parts comprising the demographic and UTAUT2 variables. An online tool known as Google Forms was used as a mode of survey delivery. These items were measured using a five-point Likert scale, ranging from "strongly agree" to "strongly disagree" [11].

Table III shows the list of measurement items included in the questionnaire which was derived from prior research. The total number of measurement items produced for this study were 30. The measurement item codes of PE1-PE3, EE1-EE4, SI1-SI3, FC1-FC4, HM1-HM3, HT1-HT3, and BI1-BI3 have been modified from Venkatesh et al. [11]. The measurement item codes of OLV1-OLV3 were modified from Ain et al. [23] while the measurement item codes of USE1-USE4 were adapted from Nikolopoulou et al. [13]. Each of these measurement items has been tailored to the perspective of this study.

B. Pilot Study

Table IV shows Cronbach's alpha results for every construct in Pilot Study from 43 respondents. According to the findings, all of the constructs have Cronbach's alpha values more than 0.70, which is deemed good [24]. Thus, the questionnaire was considered reliable to proceed with the actual study.

C. Data Collection

In order to reach the respondents, a link to the questionnaires was shared with them via social media, including Twitter, Instagram, and WhatsApp. The questionnaire link remained active until the targeted sample size has been reached. The data then were collected and analyzed using SPSS software.

TABLE III.   LIST OF MEASUREMENT ITEMS

Item Code	Question
PE1	I find that using a learning management system is beneficial to my learning.
PE2	Using a learning management system allows me to complete tasks faster.
PE3	Using a learning management system improves my understanding of my courses.
EE1	It is simple for me to learn how to use a learning management system.
EE2	My interaction with the learning management system is simple and straightforward.
EE3	I found the learning management system to be simple to use.
EE4	It is easy for me to become skillful at using learning management system.
SI1	People who are significant to me believe that I should use the learning management system.
SI2	People that have an impact on my behavior believe that I should use a learning management system.
SI3	People whose opinions I appreciate advise me to use the learning management system.
FC1	I have the necessary resources to use the learning management system.
FC2	I have the essential knowledge to use the learning management system.
FC3	Other technologies that I use are compatible with the learning management system.
FC4	I can seek support from others once I am having difficulty utilizing the learning management system.
HM1	It is fun to use a learning management system.
HM2	It is enjoyable to use a learning management system.
HM3	It is very entertaining while using learning management system.
OLV1	Learning using a learning management system is more valuable than the time and effort invested.
OLV2	The learning management system enables me to share my knowledge rapidly and easily with others in less time.
OLV3	The learning management system allows me to expand my knowledge and regulate my achievement.
HT1	Using a learning management system has become a habit for me.
HT2	I'm hooked on using the learning management system.
HT3	I must utilize the learning management system.
BI1	In the future, I intend to continue utilizing the learning management system.
BI2	In my daily life, I will continuously attempt to use a learning management system.
BI3	I want to make extensive use of the learning management system.
USE1	I regularly use learning management system in my studies.
USE2	Learning management system usage is a pleasant experience.
USE3	I am now using a learning management system to help me with my academics.
USE4	I spend a lot of time using a learning management system during my studies.

TABLE IV.   CRONBACH'S ALPHA COEFFICIENT

Construct	Cronbach's Alpha	Number of Items
Performance expectancy	0.942	3
Effort expectancy	0.974	4
Social influence	0.961	3
Facilitating conditions	0.969	4
Hedonic motivation	0.978	3
Habit	0.927	3
Online Learning value	0.971	3
Behavioral intention	0.972	3
Use behavior	0.977	4

The sample size for this study was obtained by applying Adam's sample size calculation [25]. The calculation is applicable to be used in determining the optimum sample size at all levels of confidence for both continuous and categorical variables. Equation (1) shows the sample size formula by Adam [25].

$$n = \frac{N}{1 + N\varepsilon^2} \tag{1}$$

Where.

n is the minimum size of returned samples.

N is the size of population.

ε is the adjust margin of error [$\varepsilon = (\frac{\rho e}{t})$].

e is the degree of precision given as a proportion.

ρ is the number of standard deviations that would encompass all feasible range values.

t is the t-value for the chosen alpha level of confidence.

According to Adam [25], the ρ value recommended for continuous variables is four and for categorical variables is two, whereas according to Krejcie & Morgan [26], the e value recommended for continuous variables is 0.03 and 0.05 for categorical variables. Besides, there are three confidence levels which are 90% confidence level with a 10% chance to be wrong, 95% confidence level with a 5% chance to be wrong, and 99% confidence level. The population size of UTHM students includes 13,895 students. The ρ value calculated in this formula is two, while the e value used in this formula is 0.05. The 95% confidence level was used in this study because the 90% confidence level would be narrower, and the 99% confidence level would be wider. Thus, the t-value calculated in this formula is 1.96. Adam [25] stated that a minimal required sample size for categorical variables for a population size of 10,000 is 370 and for continuous variables is 260. Based on the calculation, the minimum returned sample size required for this study is 374 (2.7%) respondents.

IV. DATA ANALYSIS AND RESULTS

Descriptive Statistics, Pearson's Correlation Coefficient, and Multiple Linear Regression were used to analyze the data in this study. Each of it will be presented in the following sections.

A. Descriptive Statistic

To summarize the data from the demographic section, a descriptive analysis was performed. Table V shows the demographic analysis.

TABLE V.   DEMOGRAPHIC ANALYSIS

Items	Category	Frequency	Percentage (%)
Gender	Male	161	48.82
	Female	215	57.18
Age	Below 20	130	34.57
	21-30	246	65.43
	31-40	-	-
	41-50	-	-
	>50	-	-
Faculty	FKAAB	18	4.79
	FKEE	14	3.72
	FKMP	28	7.45
	FPTP	31	8.24
	FPTV	26	6.91
	FSKTM	187	49.73
	FAST	16	4.26
	FTK	10	2.66
	PPD	46	12.23
Last Time Using	A days ago	332	88.30
	A weeks ago	18	4.79
	A months ago	14	3.72
	A years ago	12	3.19
How Often Using	Once every week	8	2.13
	2-3 times each week	60	15.96
	Always	282	75.00
	Rarely	26	6.91

The total number of questionnaires filled by the respondents after the data preparation was 376 consisting of 161 males and 215 females, representing a proportion of 42.82% and 57.18% respectively. Additionally, most responders fall within the age range of 21-30 years old with a percentage of 65.43%, followed by the age below 20 years old with a percentage of 34.57%. The results of this survey, which covered nine faculties, show that the Faculty of Computer Science and Information Technology (FSKTM) provided half of the replies, with a total of 187 respondents. Also, most of the respondents used LMS days ago with a percentage of 88.30%, while the least respondents used LMS years ago with a percentage of 3.19%. Lastly, most of the respondents are always using a LMS with a percentage of 75%, while the minority of the respondents use a LMS once a week with a percentage of 2.13%. In this regard, most of the students often use LMS because it has been fully used by the academic staff as a medium in the teaching and learning process instead of

using other systems or applications related to online learning platforms.

Construct	Measurement Item	Mean	Std. Deviation
Performance Expectancy	PE1	2.54	1.377
	PE2	2.63	1.283
	PE3	2.81	1.152
Effort Expectancy	EE1	2.46	1.397
	EE2	2.53	1.358
	EE3	2.44	1.441
	EE4	2.53	1.433
Social Influence	SI1	2.76	1.195
	SI2	2.86	1.158
	SI3	2.85	1.166
Facilitating Conditions	FC1	2.62	1.289
	FC2	2.52	1.338
	FC3	2.61	1.338
	FC4	2.67	1.343
Hedonic Motivation	HM1	2.71	1.197
	HM2	2.73	1.167
	HM3	2.81	1.156
Online Learning Value	OLV1	2.78	1.139
	OLV2	2.79	1.187
	OLV3	2.72	1.154
Habit	HT1	2.65	1.383
	HT2	3.16	1.152
	HT3	2.82	1.248
Behavioral Intention	BI1	2.73	1.179
	BI2	2.90	1.119
	BI3	2.83	1.160
Use Behavior	USE1	2.48	1.403
	USE2	2.64	1.214
	USE3	2.58	1.290
	USE4	2.78	1.300

Table VI shows the average and Standard Deviation (SD) for each measurement item. Based on the table, the mean value ranges from 2.44 to 3.16, which indicates that the center of the data is within the range of Likert scale between "agree" and "not sure". The standard deviation for each measurement item also shows that the data were spread out closely to the mean.

B. Analysis of Pearson Correlation Coefficient

Table VII shows the Pearson correlation coefficient analysis result. The Pearson correlation coefficient was utilized to evaluate the linear relationship between the independent and dependent variables. A value of 0 implies there is no link between the two variables, a value of +1 suggests there is a strong relationship, and a value of -1 indicates there is a weak association [27].

The multicollinearity in the data can be detected by determining the correlation matrix, tolerance value, and variance inflation factor (VIF). If the correlation coefficient in Pearson's correlation does not exceed 0.90, then the data do not have any multicollinearity problem [28]. Also, Table VII illustrates that all variable's correlation values do not exceed 0.90; as a result, there is no multicollinearity in the data. Additionally, Multicollinearity in the data may be recognized if the tolerance value is less than 0.10 and the variance inflation factor (VIF) value is more than 10 [29]. According to the study, all tolerance levels above 0.10, and all VIF values were under 10. This shows that the data do not have any multicollinearity problem.

Var	PE	EE	SI	FC	HM	OLV	HT	BI	UB
PE	1								
EE	0.846	1							
SI	0.766	0.724	1						
FC	0.818	0.876	0.725	1					
HM	0.707	0.611	0.647	0.669	1				
OLV	0.771	0.700	0.714	0.777	0.792	1			
HT	0.638	0.632	0.639	0.636	0.670	0.667	1		
BI	0.707	0.645	0.666	0.671	0.753	0.736	0.810	1	
UB	0.804	0.817	0.724	0.823	0.685	0.764	0.778	0.781	1

C. Analysis of Multiple Linear Regression

Multiple linear regression analysis was conducted to measure the significance of two or more independent variables with the dependent variables. The following equation shows the multiple linear regression formula used in this study.

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_1 x_1 + \cdots + +\beta_k x_k \qquad (2)$$

where,

$\hat{y}$ is the dependent variable.

x_k is the independent variable.

β_0 $\beta_0 = \bar{y} - \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2$

$\beta_1 = \dfrac{[(\sum x_2{}^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)]}{[(\sum x_1{}^2)(\sum x_2{}^2) - (\sum x_1 x_2)^2]}$

$\beta_2 = \dfrac{[(\sum x_1{}^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)]}{[(\sum x_1{}^2)(\sum x_2{}^2) - (\sum x_1 x_2)^2]}$

Using SPSS software, multiple linear regression analysis was conducted to assess the significance of two or more independent variables in relation to the dependent variables of behavioral intention and use behavior.

We changed R Square in the model summary for the first dependent variable, Behavioral Intention, to estimate the total variability proportion in the dependent variable explained by the independent variables [30]. The modified R Square value was 0.752, indicating that the independent factors explain 75.2 percent of the variability in behavioral intention to utilize the LMS. This implies that the research model might be utilized to

explain the variables influencing students' behavioral intention to use the LMS. Table VIII shows an analysis of the multiple linear regression for the Behavioral Intention.

TABLE VIII. MULTIPLE LINEAR REGRESSION : BEHAVIORAL INTENTION

Independent Variable	Coefficient β	p-value
Performance Expectancy	0.129	0.014
Effort Expectancy	-0.032	0.515
Social Influence	0.037	0.369
Facilitating Conditions	0.013	0.812
Hedonic Motivation	0.221	0.000
Online Learning Value	0.109	0.036
Habit	0.513	0.000

The ANOVA table's Sig. column displays the p-value. The dependent variable may be predicted by the model if the p-value is less than 0.05. According to the results of the ANOVA, the p-value was 0.000, which is less than 0.05. As a result, there is significant evidence that the model is beneficial in understanding the factors influencing students' intentions to use LMS and that the independent variables can predict students' Behavioral Intention.

The coefficient β is used to identify the strength of each independent variable towards Behavioral Intention. This means that changes by one unit in the independent variable that keeps other independent variables constant can determine the number of changes in independent variables [31]. Habit has the highest coefficient β value with 0.513. As a result, this factor has the biggest influence on students' Behavioral Intention to use the LMS which means that when students keep using the learning management system, it will influence their habit to use the LMS.

Furthermore, the p-value should be less than 0.05 for the variables to be significant. Evidently, Performance Expectancy, Hedonic Motivation, Online Learning Value, and Habit each has a p-value under 0.05. As a result, each of these variables affects the dependent variable. When the student also enjoy using the learning management system, it will influence their intention to use the LMS as we can see hedonic motivation is the second important factor that influence student's behavioral intention towards LMS. Furthermore, when the students keep using the LMS, they recognize that LMS help them in their learning performance as in this study the performance expectancy is the third important factor. As the LMS assist in their learning performance, they tend to explore more on the benefits of the learning management system. This means that the learning value that is obtain by using the LMS will influence students' intention to use the LMS.

In the meanwhile, the p-values for effort expectancy, social influence, and facilitating conditions were all more than 0.05. This shows that each of these factors does not influence the dependent variable. The facilitating conditions that are provided and the ease of use by using the LMS does not significantly influence their behavioral intention to use the LMS. Students also do not intent to use the LMS by

depending on others opinion. As a result, none of these indicators can predict students' behavioral intention to use the LMS at UTHM.

For the Use Behavior, the output shows that the value of adjusted R Square was 0.798; hence, 79.8% of the variability in the students' Use Behavior towards the learning management system is explained by the independent variables. This also means that the research model may be used to describe the factors influencing students' Use Behavior in the learning management system.

TABLE IX. MULTIPLE LINEAR REGRESSION: USE BEHAVIOR

Independent Variable	Coefficient β	p-value
Facilitating Conditions	0.481	0.000
Habit	0.343	0.000
Behavioural Intention	0.239	0.000

Table IX shows the coefficient β and the p-value for the multiple linear regression analysis of the dependent variable: Use Behavior. Facilitating Conditions yielded a higher coefficient β value of 0.481. This means that this factor has the strongest effect on students' Use Behavior toward the LMS. Furthermore, the p-value should be less than 0.05 for the variables to be significant. Evidently, Facilitating Conditions, Habit, and Behavioral Intention all have p-values of less than 0.05. Hence, each factor influences the dependent variable. The result also shows that behavioral intention has the least coefficient β which means that their intention to use the LMS did not fully influence their use of LMS. However, because the students must continue to use the LMS, it has become a habit for them. As a result, habit has emerged as the second most significant element influencing student LMS usage behavior. Fig. 2 shows the research model after analysis.
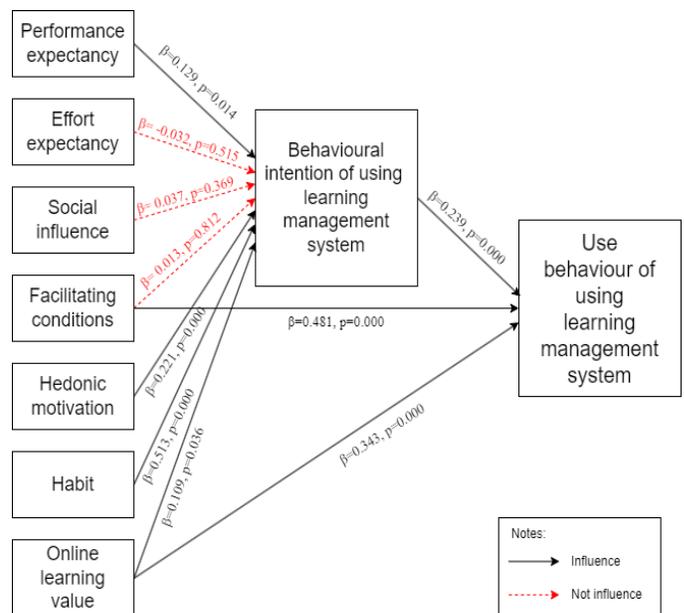
Fig. 2. Research Model after Analysis Result.

Overall, the results have shown that performance expectancy, hedonic motivation, habit, and online learning value are the significant factors that influence students' intention to use LMS, whereas facilitating conditions, habit, and behavioral intention are the significant factors that influence students' use behavior towards LMS.

V. CONCLUSION AND FUTURE WORKS

This study is focused on the acceptance of UTHM students of their online learning management system based on the UTAUT2 model. A detailed explanation on the use of the UTAUT2 model and the background information related to this study has been provided. The proposed methodology also has been discussed. Then the data analysis and results has been presented using Descriptive Analysis, Pearson's Correlation and Multiple Linear Regression. The objectives of this study have been achieved.

Firstly, from the results, the factors that have the most influence on UTHM students' behavioral intentions toward using the LMS are Performance Expectancy, Hedonic Motivation, Online Learning Value, and Habit. Also, we may infer from the results that students think implementing a learning management system would improve their academic achievement. Next, students have fun and joy derived from using the LMS. Furthermore, students dedicated their time and effort to using the LMS and the perceived value that they get from using the LMS. Lastly, students used the LMS automatically and become habitual in their learning process.

Secondly, the outcome also demonstrates that Habit, Behavioral Intention, and Facilitating Conditions are the elements that affect UTHM students' Use Behavior when using LMS. From the results, we conclude that having access to a few sources and assistance boosts students' intention to utilize the LMS, making it the most important factor influencing students' LMS Use Behavior. The next reason why students use the LMS is Habit, and it is the final element that has the least impact on how students use the LMS is whether they want to keep using it to further their education.

Thirdly, Online Learning Value construct was proposed as the new variable in the UTAUT2 model to study students' behavioral intention toward the LMS in UTHM. The result from Table VIII shows that the Online Learning Value construct (β =0.109, p=0.036) also is a significant factor in determining students' behavioral intention toward using the LMS in UTHM.

Several recommendations for future works based on the findings of this study can be explored. Currently, this study only focuses on the variables stated in the UTAUT2 model along with the new construct which is Online Learning Value. Other variables may be explored to identify other factors that can influence student acceptance of using LMS in UTHM. This may further give insight to LMS administrator and developer to enhance the LMS in UTHM. Also, choosing other stakeholders such as academic staff and administrators for the acceptance of the LMS also can be done in the future. This work is also important considering that other stakeholders' acceptance may also play an important part in the LMS's success.

REFERENCES

[1] Y. Wiratomo and F. Mulyatna, "Use of Learning Management Systems in Learning Efforts during a Pandemic," Journal of Mathematical Pedagogy (JoMP), vol. 1, no. 2, 2020.

[2] M. A. Almaiah, A. Al-Khasawneh, and A. Althunibat, "Exploring the critical challenges and factors influencing the E-learning system usage during COVID-19 pandemic," Educ Inf Technol (Dordr), vol. 25, no. 6, 2020, doi: 10.1007/s10639-020-10219-y.

[3] S. Balkaya and U. Akkucuk, "Adoption and use of learning management systems in education: The role of playfulness and self-management," Sustainability (Switzerland), vol. 13, no. 3, 2021, doi: 10.3390/su13031127.

[4] S. Hamid, S. N. Ismail, M. Hamzah, and A. W. Malik, "Developing Engagement in the Learning Management System Supported by Learning Analytics," Computer Systems Science and Engineering, vol. 42, no. 1, 2022, doi: 10.32604/CSSE.2022.021927.

[5] I. Maslov, S. Nikou, and P. Hansen, "Exploring user experience of learning management system," International Journal of Information and Learning Technology, vol. 38, no. 4, 2021, doi: 10.1108/IJILT-03-2021-0046.

[6] B. J. Santiago et al., "Learning management system-based evaluation to determine academic efficiency performance," Sustainability (Switzerland), vol. 12, no. 10, 2020, doi: 10.3390/su12104256.

[7] A. Z. A. Kadir and N. S. Aziz, "Learning Management System of higher education institution," Indian J Sci Technol, vol. 9, no. 9, 2016, doi: 10.17485/ijst/2016/v9i9/88717.

[8] M. Musa, M. N. Ismail, and M. F. M. Fudzee, "A survey on smart campus implementation in Malaysia," International Journal on Informatics Visualization, vol. 5, no. 1, 2021, doi: 10.30630/joiv.5.1.434.

[9] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," MIS Q, vol. 27, no. 3, 2003, doi: 10.2307/30036540.

[10] G. D. M. N. Samaradiwakara and C. G. Gunawardena, "Comparison of Existing Technology Acceptance Theories and Models to Suggest a Well Improved Theory / Model," International Technical Sciences Journal, vol. 1, no. 1, 2014.

[11] V. Venkatesh, J. Y. L. Thong, and X. Xu, "Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology," MIS Q, vol. 36, no. 1, 2012, doi: 10.2307/41410412.

[12] Meyliana, H. A. E. Widjaja, S. W. Santoso, S. Petrus, Jovian, and Jessica, "The Enhancement of Learning Management System in Teaching Learning Process with the UTAUT2 and Trust Model," in Proceedings of 2019 International Conference on Information Management and Technology, ICIMTech 2019, 2019. doi: 10.1109/ICIMTech.2019.8843828.

[13] K. Nikolopoulou, V. Gialamas, and K. Lavidas, "Acceptance of mobile phone by university students for their studies: an investigation applying UTAUT2 model," Educ Inf Technol (Dordr), 2020, doi: 10.1007/s10639-020-10157-9.

[14] X. Huang, "Social media use by college students and teachers: An application of UTAUT2," ProQuest Dissertations and Theses, 2018.

[15] S. N. Samsudeen and R. Mohamed, "University students' intention to use e-learning systems: A study of higher educational institutions in Sri Lanka," Interactive Technology and Smart Education, vol. 16, no. 3, 2019, doi: 10.1108/ITSE-11-2018-0092.

[16] A. A. Arain, Z. Hussain, W. H. Rizvi, and M. S. Vighio, "Extending UTAUT2 toward acceptance of mobile learning in the context of higher education," Univers Access Inf Soc, vol. 18, no. 3, 2019, doi: 10.1007/s10209-019-00685-8.

[17] P. Jakkaew and S. Hemrungrote, "The use of UTAUT2 model for understanding student perceptions using Google Classroom: A case study of Introduction to Information Technology course," in 2nd Joint International Conference on Digital Arts, Media and Technology 2017: Digital Economy for Sustainable Growth, ICDAMT 2017, 2017. doi: 10.1109/ICDAMT.2017.7904962.

[18] J. A. Kumar and B. Bervell, "Google Classroom for mobile learning in higher education: Modelling the initial perceptions of students," Educ Inf Technol (Dordr), vol. 24, no. 2, 2019, doi: 10.1007/s10639-018-09858-z.

[19] E. Mohamed, M. Nasef, N. Megat, M. Zainuddin, R. Ibrahim, and S. A. Shariff, "Proposed Model of Students Acceptance of Massive Open Online Courses Article history," 2019.

[20] A. Sharif, S. Afshan, and M. A. Qureshi, "Acceptance of learning management system in university students: An integrating framework of modified UTAUT2 and TTF theories," International Journal of Technology Enhanced Learning, vol. 11, no. 2, 2019, doi: 10.1504/IJTEL.2019.098810.

[21] F. Ali, P. K. Nair, and K. Hussain, "An assessment of students' acceptance and usage of computer supported collaborative classrooms in hospitality and tourism schools," J Hosp Leis Sport Tour Educ, vol. 18, 2016, doi: 10.1016/j.jhlste.2016.03.002.

[22] M. S. Farooq et al., "Acceptance and use of lecture capture system (LCS) in executive business studies: Extending UTAUT2," Interactive Technology and Smart Education, vol. 14, no. 4, 2017, doi: 10.1108/ITSE-06-2016-0015.

[23] N. U. Ain, K. Kaur, and M. Waheed, "The influence of learning value on learning management system use: An extension of UTAUT2," Information Development, vol. 32, no. 5, 2016, doi: 10.1177/0266666915597546.

[24] S. Kilic, "Cronbach's alpha reliability coefficient," Journal of Mood Disorders, vol. 6, no. 1, 2016, doi: 10.5455/jmood.20160307122823.

[25] A. M. Adam, "Sample Size Determination in Survey Research," J Sci Res Rep, 2020, doi: 10.9734/jsrr/2020/v26i530263.

[26] R. v. Krejcie and D. W. Morgan, "Determining Sample Size for Research Activities," Educ Psychol Meas, vol. 30, no. 3, 1970, doi: 10.1177/001316447003000308.

[27] H. Akoglu, "User's guide to correlation coefficients," Turkish Journal of Emergency Medicine, vol. 18, no. 3. 2018. doi: 10.1016/j.tjem.2018.08.001.

[28] K. Moorthy et al., "Behavioral Intention to Adopt Digital Library by the Undergraduates," International Information and Library Review, vol. 51, no. 2, 2019, doi: 10.1080/10572317.2018.1463049.

[29] M. A. Schroeder, "Diagnosing and Dealing with Multicollinearity," West J Nurs Res, vol. 12, no. 2, 1990, doi: 10.1177/019394599001200204.

[30] J. Miladinovic and H. Xiang, "A Study on Factors Affecting the Behavioral Intention to use Mobile Shopping Fashion Apps in Sweden," 2016.

[31] D. M. Levine and D. F. Stephen, Even You Can Learn Statistics and Analytics: An Easy to Understand Guide to Statistics and Analytics, 3rd ed. Upper Saddle River: New Jersey: Pearson FT Press, 2014.

Object Pre-processing using Motion Stabilization and Key Frame Extraction with Machine Learning Techniques

Kande Archana[1]

Research Scholar
Department of CSE
Jawaharlal Nehru Technological University (JNTU)
Hyderabad, India, Telangana State

V Kamakshi Prasad[2]

Professor
Department of CSE
Jawaharlal Nehru Technological University (JNTU)
Hyderabad, India, Telangana State

Abstract—Video information processing is one of the most important application areas in research and to solve in various pre-processing issues. The pre-processing issues such as unstable video frame rates or capture angle, noisy data and large size of the video data prevent the researchers to apply information retrieval or categorization algorithms. The video data itself plays a vital role in various areas. This work aims to solve the motion stabilization, noise reduction and key frame extraction, without losing the information and in reduced time. The work results into 66% reduction in key frame extraction and nearly 6 ns time for complete video data processing.

Keywords—Information loss preventive; mean angle measure; key frame extraction; moving average; dynamic thresholding

I. INTRODUCTION

Key-frame extraction from first-person vision (FPV) is a fundamental tool for highlighting and memorizing meaningful moments from a person's life. Selecting pivotal frames is challenging when using head-mounted FPV cameras due to the scene's inherent uncertainty. Because of the inherent instability of head-mounted cameras, FPV footage is noisier than TPV footage. In order to extract key frames, most algorithms consider TPV scenes to be static. Extraction of key frames from noisy first-person-view videos is still in its infancy. While human motion is ubiquitous in daily life, most key-frame fractionation algorithms rely on information from first-person videos. FPV videos rely heavily on motion capture integrated with visual scenes due to the rapid nature of scene transitions. The authors of this study propose a new key-frame extraction technique for FPV videos to attenuate background noise and identify potentially important events. Both the sparse-based and graph-based multi-sensor integration models proposed for key-frame extraction perform well on the shared space. Experiments with multiple datasets support the claim that the proposed key-frame extraction techniques enhance "extraction precision and coverage of entire video sequences" [1].

When applied to dynamic regions of interest, video SAR can enhance automatic retrieval of information. Extracting key frames from a video efficiently processes large amounts of data. Background subtraction using computer vision is suggested for automatic key-frame selection in Video SAR scattering. Key frames for scattering in Video SAR are revealed by a universal parameterization model. To differentiate the scattering key-frame state of transient persistence and disappearance, we propose the use of the "sub aperture energy gradient (SEG) and a modified statistical and knowledge-based object tracker (MSAKBOT)". Multitemporal video sequences are no match for the proposed SEG-MSAKBOT method, which evaluates pelting key frames in a more comprehensive and adaptable manner. Experimental findings and performance evaluation on two actual airborne Video SAR datasets with coherent integration angles [2].

VSUMM is widely used for processing large amounts of video information. Frames that best capture the essence of a video's subject are chosen by VSUMM. A novel framework is proposed for elucidating content and motion to solve these problems. A cross - functional and cross motion curve is created by Capsules Net, which has been taught to extract spatiotemporal information. Next, we propose an approach to automatically extract individual shots from continuous video feeds by analyzing their sensitivity to a variety of transition effects. Key-frame patterns in under shots can be selected using a self-attention model, key static images can be chosen for video content categorization, and optical movements can be calculated for video motion summarization. Results from experiments show that the method achieves competitive results [3].

An essential part of video retrieval is the extraction of key frames. For videos with multiple scenes and actions, the current state of the art in key frame extraction falls short. This article proposes a model for image saliency extraction that is aided by deep prior information. The paper proposes a saliency extraction algorithm to find important details in a given image. Here we describe a novel approach to finding the most important moments in a video. In this article, we build an image-integrated model of visual attention to track a combination of multiple bottom-level features and the skin color confidence map of the target. Extraction of the moving targets. The algorithm has been shown to effectively grasp pedestrian information in moving videos and provide samples of motion targets for post-processing [4] based on experimental results.

The rest of the work is organized such as in the Section – II, III and IV the foundational methods for motion stabilization, noise reduction and key frame extractions are discussed, further based on the baseline methods, the recent research improvements are discussed in the Section – V, further the persistent research problems are discussed in the Section – VI and based on the analysis of the problem, the proposed solutions are discussed in Section – VII and Section VIII, the obtained results from the proposed solutions are discussed in the Section – IX and again compared with the other benchmarked algorithms in Section – X. The research presents the conclusion in the Section – XI.

II. Foundational Method for Motion Stabilization

Firstly, in this section of the work, the baseline method for motion stabilization is analyzed. The process for motion stabilization is highly crucial as the unstable video data influences the decisions during video stabilization process.

Assuming that, the complete video data is V [] and each frame is denoted as f_i. Then, for n number of total frames, the relation can be formulated as,

$$V[] = <f_1, f_2, f_3 .... f_n> \tag{1}$$

Further, assuming that the function ϕ is an arbitrary function to extract the dimension, $d_X[]$ and angle of video capture, a_X from the frames. Thus, this can be furnished as,

$$\phi\{f_X\} \Rightarrow [d[]_X, a_X] \tag{2}$$

Thus, for two independent frames, as f_i and f_j, the formulation can be realized as,

$$\phi\{f_i\} \Rightarrow [d[]_i, a_i] \tag{3}$$

$$\phi\{f_j\} \Rightarrow [d[]_j, a_j] \tag{4}$$

The detection of the unstable video can be detected if the following condition appears in any two consecutive frames as,

$$a_i \neq a_j \tag{5}$$

During such situations, the baseline method recommends calculating the mean angle, a_{Mean} of the total video and further change the angle of the unstable frames with the mean angle. This can be realized as,

$$a_{Mean} \Leftarrow mean\{\sum_{i=1}^{n} a_i\} \tag{6}$$

And further replace with the existing angles, as,

$$a_i \leftarrow a_{Mean} \tag{7}$$

$$a_j \leftarrow a_{Mean} \tag{8}$$

Regardless to mention, the base line method has seen many improvements in recent times. The improvements are discussed in the further section of this work.

III. Foundational Method for Noise Reduction

Secondly, in this section of the work, the baseline method for noise reduction is analyzed. The noise reduction is yet another highly important task before processing the video data for information extraction. Any noisy video data can wrongly influence the decision-making tasks during the processing of the video data.

The baseline method for noise reduction or removal process is furnished here. After realizing the Eq. 1, assuming that, λ is an arbitrary function to extract the size of the pixel information, s_X and list of objects with details, $obj_X[]$ from each frame. This can be formulated as,

$$\lambda\{f_X\} \Rightarrow [s_X, obj_X[]] \tag{9}$$

The baseline method demonstrates that, the extraction of similar frames, F[], with similar objects is the initial step towards the noise reduction as,

$$F[] \Leftarrow \prod_{obj_X[]=V[i].obj_i[]} V[i] \tag{10}$$

Now assuming that two frames, fi and fj are part of F[] set and contain similar number of objects. Hence, the information size, s_i and s_j, must be same. If the information sizes are different, then it is natural to realize that the frames contain noise as per the baseline method. As,

$$iff \ s_i \neq s_j$$
$$then, s_i \ \& \ s_j \Rightarrow Noise \tag{11}$$

Further, the baseline method indicates to replace the pixel information of the noisy frames with the mean information, S_{Mean} from the similar frames. As,

$$s_i \Leftarrow Mean\{\sum_{k=1}^{Length\{F[]\}} F[k]\} \tag{12}$$

$$s_j \Leftarrow Mean\{\sum_{k=1}^{Length\{F[]\}} F[k]\} \tag{13}$$

Regardless to mention, that the base line method has seen many improvements in recent times. The improvements are discussed in the further section of this work.

IV. Foundational Method for Key Frame Extraction

Thirdly, in this section of the work, the baseline method for the key frame extraction is furnished. The video information processing is a highly time complex process in general and to reduce the time complexity of the processing algorithms, the researchers have adapted to method to reduce the size of the video data without losing the critical information from the data. This process is identified as key frame extraction process.

The baseline method indicates, that the total video data must be broken into set of frames, F[], based on a given time threshold as,

$$F[] = \frac{d}{dt}V[] \qquad (14)$$

Further, the key frame extraction process, as per the baseline method, is very simple. The consecutive frames contain different information must be considered as key frames, KF[]. This can be formulated as,

$$KF[] \Leftarrow \prod_{F[i] \neq F[i+1]} F[] \qquad (15)$$

The terminating condition for such methods is that the length of the extracted key frame set must be less than the total video data as,

$$Length\{KF[]\} \; \Box \;\; Length\{V[]\} \qquad (16)$$

Regardless to mention, that the base line method has seen many improvements in recent times. The improvements are discussed in the further section of this work.

V. RECENT RESEARCH REVIEWS

After realizing the baseline methods, in this section of the work, the recent research improvements over the baseline methods are discussed.

To address the problems of miss-election and misselection due to poor video key frame detection, Z. Wang et al. [5] propose an algorithm based on motion vectors to locate the crucial frames in a video. Using the sum of the entropy of the difference between adjacent frames and the entropy of the image in two dimensions, we can quantify frames. Second, the lens boundary can be obtained with the help of statistical tools. ViBe is an algorithm that can recognize the object in the foreground of a video and extract transformation features that are not affected by the size of the image. The motion vector is found by first segmenting two consecutive frames, and then matching those segments block by block. The video's level of motion is reflected in the magnitude of the motion vector, allowing us to identify "active and inactive motion" regions. Video frames are compared using a predetermined formula to determine which regions of frames are most similar, and then video frames are extracted from those regions. Video findings with rich motion information are improved by the proposed detection algorithm, as demonstrated by experimental results. The outcomes of VR key frame extraction are also analyzed in this article.

Determining whether or not a video frame was intentionally deleted or altered, is a crucial part of video forensics. Existing methods are unable to process videos with varying levels of motion. Interfering frames are ignored by these methods. C. Feng et al. [6] aim to develop an interference-free, motion-adaptive forensic technique. Researchers analyze statistically the frames that interfere with one another, like moved I-frames to pinpoint the frames that should be removed (FDPs). The adaptability of the fluctuation feature to various levels of motion is enhanced by the elimination of intra-predictions. The improvement is quantified with the aid of moving window detectors. We propose a post-processing technique to eliminate jarring transitions in brightness, focus, and frame rate. When applied to videos with varying motion intensities and interfering frames, the algorithm has a true positive rate of 90% and a false alarm rate of 0.3%. The proposed technique has potential applications in video forensics.

CRC mortality rates can be lowered through early diagnosis. Polyps are the first stage of cancer. Disease diagnosis and comprehension require analysis of the most crucial endoscopy frames. Sasmal et al. [7] use deep learning to select laparoscopic video key-frames. This technique employs transfer learning due to the scarcity of high-quality polyp depth maps. During an endoscopy, many images are taken. Discarding frames from an endoscopic video that are of poor quality or have no diagnostic value is necessary for making a clinical diagnosis. Key-frame selection is suggested using polyp depth information. This method intelligently chooses key-frames by analyzing the significance of edges, punctuation, and other image features. So that the surgeon can separate the polyp from the mucosa layer, a real-time 3D image of the polyp's surface is provided. Polyps can be more easily pinpointed with the aid of depth maps.

Recognizing anomalies in video footage, which is necessary for uses such as surveillance, is not easy. The state-of-the-art methods for detecting video anomalies typically rely on deep rebuilding models, but their results fall short because there isn't enough of a difference between the reconstruction errors for normal and anomalous video frames to fully optimize the models. Anomaly detection using frame predictions shows promise. Unsupervised content anomaly detection was proposed by X. Wang et al. [8] using frame prediction. To deal with semantically informative objects and regions of varying scales and to capture spatial-temporal dependencies in everyday videos, the proposed method employs a "multipath ConvGRU-based frame prediction" network. In order to lessen the impact of ambient noise, training lowers tolerance for it. The proposed method achieves better results than the current state-of-the-art approaches.

To achieve a higher frame rate in videos, interpolation is used to synthesize new frames between the existing ones. Existing techniques rely on pairing adjacent frames to create intermediate frames, but they struggle with issues like high-velocity motion, occlusion, and blurring. In order to make the most of the spatial and temporal data at their disposal, H. Zhang et al. [9] proposed "a multi-frame pyramid refinement network". There are three technical advancements that would benefit the proposed network. The first step is the proposal of a coarse-to-fine framework for improving optical flows across multiple frames. Estimates can be made for both large-scale motion and occlusion. Second, spatial and temporal context is unearthed, and texture is restored. Third, we use perceptual loss in multiple steps to keep the finer details of the intermediate frames intact. For interpolation between multiple frames, this technique can be used. Overall, 80K frame groups are used in the training process. The method has been shown to outperform state-of-the-art methods and handle challenging cases on multiple independent datasets.

H. Bhuyan et al. [10] extract the key frames from the dance video and the motion frames from the video itself. New,

easy, and efficient, the localization method is suggested. The basic structure of KFs varies depending on the dance style and the dancer. It's not easy to establish a universal threshold for movement that applies to all dancers and performances. The threshold was previously determined using iterative methods.

W. Zhang et al. [11] investigated spatio-temporal video super-resolution. The researchers first propose a cross-frame transformer-based network for end-to-end spatio-temporal video super-resolution. To reconstruct high resolution and frame rate results from coarse to fine, the researchers propose "a multi-level residual reconstruction module" that makes use of the maximum similarity and similarity coefficient matrices produced by the cross-frame transformer. Compared to the standard two-stage network, this method has fewer training parameters while providing improved performance.

The process of "video object extraction (VOE)" identifies and isolates region of interest from a video. To solve the issues of imprecise foreground object extraction and superfluous small-scale motion interference, Y. Guo et al. [12] proposed a novel VOE approach based on "spatiotemporal consistency saliency" detection. The proposed method's main innovation is comprised of three parts: first, the spatiotemporal gradient field (SGF) is built a tempor Experiments on the public video saliency data sets ViSal and SegtrackV2 demonstrate efficiently and accurately identify the salient object in a video sequence.

Semantic sections of a video are extracted using video segmentation, with each segment corresponding to a user-defined concept. The goal of the user has not been taken into account in previous research on video segmentation. With dimension reduction and temporal clustering, X. Peng et al. [13] present a "two-stage user-guided video segmentation" framework. During dimension reduction, coarse-grained features are extracted using ImageNet. During the temporal clustering phase dimensionally, reduced frames is used to segment videos on the time domain based on the user's intended viewing path. To better understand videos, users can employ hierarchical clustering to divide them into smaller, more manageable chunks.

To facilitate the retrieval, it takes advantage of the fact that certain video-related image features tend to exhibit temporal correlations. Key frame identification and key area localization are made possible through the use of video imprint representation. The video imprint tensor is created in the framework developed by Z. Gao et al. [14] by removing redundancy across frames with a specialized feature alignment module. The proposed reasoning network employs a memory-inspired attention mechanism. The latent structure of the reasoning network identifies key frames from the video imprint that can be used to reconstruct the sequence of events. Event retrieval is enhanced over the state-of-the-art techniques thanks to the "compact video representation aggregated from the video" imprint.

Observing moving objects is made possible thanks to video satellite's ability to produce rich actionable information. Time resolution is more important than spatial clarity in video satellite images. The quality of video satellite images is greatly enhanced by super-resolution. "Video satellite image SR reconstruction" was proposed by Z. He et al. [15]. Oftenest calculates the LR optical flow from multiple image frames. Next, an upnet is constructed to enhance the resolution of the input frame and the LR optical flow. Motion compensation is executed in accordance with HR optical flows. Through the use of the HR cube's compensation, the ARLnet produces SR results.

Intelligent service robots with video analysis capabilities are used for complex computing in the cloud. To monitoring human activity, intelligent service robots' film in a continuous loop. "Action classification, recognition, abnormal event detection, and crowd emotion sensing" all require action extraction from unrestricted continuous video. With three components—spatial location estimation, temporal action path searching, and spatial-temporal action compensation. H. Guo et al. [16] proposed a novel approach for action extraction in unconstrained continuous video. It is possible to pinpoint one's location based on a person's outward appearance and their motion. Then, considering missed sightings and false alarms, the results of the spatial action proposal are used to formulate the space - time action trail scanning as an optimal probability model. "The researchers advocate for and demonstrate the convergence of the "Markov Chain Monte Carlo" algorithm.

Because of developments in video and image processing, video tampering forensics is extremely difficult to conduct. Object reduction video forgery requires the use of passive forensics techniques, which are crucial in the courtroom. The "spatial rich model (SRM) and 3D convolution (C3D)" were the basis for the proposal of a spatiotemporal trident network by Q. Yang et al. [17], which is used to extract tampering traces from video. It can enhance the detection and identification of tampered regions, and it has three distinct branches. The "spatiotemporal trident network" served as inspiration for the development of a time-based detector and a space-based locator for locating instances of video manipulation. The temporal detector used 3D convolutional neural networks (CNNs) with three different types of encoders and decoders. C3D-ResNet12 was developed as the spatial locator's encoder. The loss functions of both algorithms were optimized by the researchers.

The future of the space information network will rely heavily on satellite video because of the dynamic information it provides on large spatial and temporal scales. This study employs a two-stream approach to extract EOI from satellite video scenes. Still images, such as scenes, are captured in each frame of a satellite video, while the order of these frames is what establishes motion. In light of these details, we propose a brand new two-stream EOI detection framework. One stream uses AlexNet to pull static spatial data from satellite videos, while the other uses local trajectories analysis to pull data on motion. Before anything else, the video scene is cut up into spatial-temporal patches where EOI and non-EOI regions are labelled. The next step is to take the 3-D satellite video cubes from the event scene patches and extract the trajectories. Finally, this weak-supervision trajectory classification problem is solved by sparse dictionary learning. The experimental outcomes demonstrate the efficacy of the two-stream method in EOI detection and its potential utility in "satellite video analysis" and comprehension. The approach

taken by Y. Gu et al. [18] is superior to current video analysis models.

CNNs can dehaze individual images, as discovered by W. Ren et al. [19]. Scientists investigate the feasibility of using a network to defog video footage. Dehazing a video can make use of the wealth of data available in adjacent frames. Based on the estimated transmission map, a haze-free scene model is created. Since the semantic information of a scene is a powerful prior for image restoration, the team suggests using "global semantic prior to regularize the transmission" maps, making the estimated maps continuous only between objects and smooth within each object. To train this network, scientists generate synthetic hazy and clear videos. Scientists have demonstrated that the features of this dataset can clear up haze in videos of outdoor scenes.

For supervised video summarization, W. Zhu et al. [20] proposed DSNet. Anchors are used and ignored by DSNet. While the anchor-based approach eliminates pre-defined temporal proposals and predicts importance scores and segment locations, the anchor-free approach in contrast to preexisting supervised video summarization techniques, the interest detection framework makes an initial attempt to exploit temporal consistency. Before extracting their long-range temporal features for location regression and importance prediction, researchers in the anchor-based approach. Segments are assigned as positive or negative to ensure the summary is accurate and comprehensive. The anchor-free method directly predicts the significance of video frames and segments, sidestepping the drawbacks of temporal proposals. This framework is compatible with existing supervised video summarization tools. Scientists evaluate both anchor-based and anchor-free methods on SumMe and TVSum. Both the anchor-based and anchor-free strategies have been verified by experiments.

Object recognition is the mainstay of computer vision. There is hope in the use of correlation filters. Because each target is so small and the target and background are so similar, the "kernel correlation filter (KCF) tracker" has trouble keeping up with moving objects in satellite videos. To better locate objects in satellite footage, B. Du et al. [21] suggested combining the KCF tracker with a three-frame-difference algorithm. For a robust tracker, it is suggested to combine the KCF tracker with the three-frame-difference algorithm. Three satellite videos demonstrate the superiority of the proposed method, as demonstrated by the researchers.

Further, based on the recent research improvements, in the next section of this work, the persistent research problems are furnished.

VI. PROBLEM FORMULATION – MATHEMATICAL MODEL

After the understanding of the baseline methods and detailed analysis of the recent research outcomes, in this section, the persistent research problems are discussed.

Firstly, the problem of information loss due to the change of angle for video stabilization is discussed.

Continuing from Eq. 3 and 4, assuming that, due to the change of the angle with a_X, the frames f_i and f_j, are expected

to change the dimensions from $d_i[]$ and $d_j[]$ to $d_X[]$. Naturally, which are different from the original dimensions. This can be realized as,

$$d_i[H_i, W_i] \neq d_X[H_X, W_X] \tag{17}$$

$$d_j[H_j, W_j] \neq d_X[H_X, W_X] \tag{18}$$

Where H and W are the height and width respectively also, it is natural to realize that, if the new dimensions are less than the original dimensions, then the information loss is non-preventable. As,

$$iff \ H_i > H_X \ Or \ W_i > W_X$$
$$Then, \lambda(f_i) > \lambda(f_i') \tag{19}$$

Where, f' is the modified frame after the motion angle change. Hence, it is natural to observe the information loss due to the angle change as per the traditional existing methods.

Secondly, the problem of undetected object mismatch during the noise reduction is realized.

The second issue with the traditional existing methods are to be realized with a condition that, the shape and number of the objects are same, however the position of the objects are different. This can be realized from Eq. 9. Assuming that, two frame f_i and f_j contain the set of objects in the frame as $obj_i[]$ and $obj_j[]$. As,

$$\lambda\{f_i\} \Rightarrow [s_i, obj_i[]] \tag{20}$$

$$\lambda\{f_j\} \Rightarrow [s_j, obj_j[]] \tag{21}$$

Here, the objects in the respective frames are same and the size of the frames are also same as,

$$s_i = s_j \tag{22}$$

$$obj_i[] = obj_j[] \tag{23}$$

Nonetheless, the positions of the objects are different as,

$$obj_i[k].(H, W) \neq obj_j[k].(H, W) \tag{24}$$

Thus, naturally, the information from the frames is also not equal. As,

$$\varpi\{f_i\} \neq \varpi\{f_j\} \tag{25}$$

Where, ϖ is an arbitrary function to extract knowledge from the video frame. Henceforth, the loss of object unique locations over in the noisy frames will be lost as per the Eq. 13.

Finally, the problem of higher time complexity during the key frame extraction is analyzed. As per the initial assumptions, the total number of frames in the video data is "n" and as per the Eq. 15, the total time complexity, T, can be formulated as,

$$T = n*(n-1) \tag{26}$$

Or,

$$T = n^2 \tag{27}$$

Which implies,

$$T(n) = O(n^2) \tag{28}$$

And, during a situation of $n \rightarrow High$, naturally, $T(n) \rightarrow Very\ High$. Thus, this problem also must be addressed.

Further, the proposed solutions to these identified research problems are furnished in the next section of this work.

VII. PROPOSED SOLUTIONS

After the detailed analysis of the problems with baseline methods and the recent improvements observed to the baseline methods, in this section of the work, proposed methods are furnished here.

Firstly, the proposed video stabilization method without the information loss is realized. The initial process is to identify the variation of the angles in each frame. Using Eq. 2, extracting all the angles from each frame and build the angle collection, A[], as,

$$A[] = \prod V[].a \tag{29}$$

Further calculate the mean angle, A_{Mean} from the angle collection as,

$$A_{Mean} \Leftarrow Mean\{\sum_{i=1}^{n} A[i]\} \tag{30}$$

Further, build another set of frames, FA[], where two consecutive frames angle is different as,

$$FA[] \Leftarrow \prod_{a_i \neq a_j} F[] \tag{31}$$

Further, apply the mean angle to the FA[] collection frames and check the number of frames, where information loss is observed. As,

$$FA[].a \Leftarrow A_{Mean}$$
$$iff\ \lambda(FA[i]) > \lambda(FA'[i]) \tag{32}$$
$$then, C[] \Leftarrow FA[i]$$

Where, C[] is collection of frames, where after motion stabilization information is lost. Now, if for the maximum number of frames information loss is observed, then the mean angle must be recalculated. As,

$$iff\ length\{C[]\} \approx length(FA[])$$
$$then, A_{Mean} \Leftarrow Max\{FA[].a\} \tag{33}$$

The proposed method indicates to repeat the process until $length\{C[]\} \rightarrow 0$. Henceforth, the loss of information is almost reduced to zero.

Secondly, the proposed noise reduction method with object information persistent and background separation is furnished.

The proposed method indicates to separate the background and foreground information from each frame. Assuming that, X is an arbitrary function to perform the separation task, then the following model can be formulated,

$$\Delta\{f_x\} \Rightarrow \{f_x.FG, f_x.BG\} \tag{34}$$

Further, build the set of frames, K[][], with similar background and similar foreground as,

$$K[][] \Leftarrow \prod_{f_x.FG=F[].FG, f_x.BG=F[].BG} F[] \tag{35}$$

The mean value of the information extracted from the K[][] collection for the background, BG_{Mean}, foreground, FG_{Mean}, must be calculate to substitute for the noisy frames. As,

$$BG_{Mean} = Mean\{\sum_{i=1, j=1}^{length\{K[][]\}, n} \lambda\{K[i][j].BG\}\} \tag{36}$$

$$FG_{Mean} = Mean\{\sum_{i=1, j=1}^{length\{K[][]\}, n} \lambda\{K[i][j].FG\}\} \tag{37}$$

Once the mean values are calculated, then the pixel value information must be replaced with the mean values. Hence, continuing from Eq. 15.

$$f_x.BG \Leftarrow BG_{Mean} \tag{38}$$

$$f_x.FG \Leftarrow FG_{Mean} \tag{39}$$

Hence, now the frames with similar object positions and similar background information can be denoised without missing the specific objects information.

Finally, the proposed variable threshold based key frame extraction method is realized.

The concept of the variable threshold used here is primarily to indicate that as the video information changes frame by frame, thus the selection of the key frames also must be decided dynamically. Hence this proposed method indicates to calculate the keyframe selection method using moving average of the threshold.

Continuing from the Eq. 9 and Eq. 14, the threshold, TH, for the information frame must be calculated as,

$$TH \Leftarrow \lambda\{f_i\} \tag{40}$$

and, further,

$$KF[] \Leftarrow \prod_{F[i].TH > Mean\{\sum_{i=1}^{n} TH(i)\}} F[] \tag{41}$$

Hence, this proposed key frame extraction process reduces the time complexity to O(n) as the mean value calculation is an iterative process along with the same steps of key frame extraction.

Further, based on the proposed methods, in the next section of this work, the proposed algorithms are furnished along with the proposed framework.

VIII. PROPOSED ALGORITHMS AND FRAMEWORKS

After the finalization of the proposed mathematical models in the previous section of this work, in this section, the proposed algorithms based on the mathematical models are presented.

Firstly, the Information Loss Preventive Video Stabilization using Mean Angle Measure (ILP-VS-MAM) Algorithm is discussed.

Algorithm - I: Information Loss Preventive Video Stabilization using Mean Angle Measure (**ILP-VS-MAM**) Algorithm

---

Input:
Video data as V[]

Output:
Motion Stabilized Video data as V1[]

Process:
Step - 1. Read the total video data as V[]
Step - 2. For every frame in V[] as V[i]
    a. Extract the angle as A[i] using Eq. 2.
    b. If A[i] != A[i-1]
    c. Then, C[j] = V[i]
Step - 3. For every element in V[] as V[k]
    a. Calculate the mean angle as MA from A[]
    b. Apply V[k].A = MA and generate V_Temp[i]
    c. If V_Temp[k] contains less information then V[k] using Eq. 32
    d. Then, remove A[k] and C[k] = 0
    e. Else, V1[k] = V[k]
    f. Stop if Length((C[]) = =0)
Step - 4. Return V1[]

In order to prevent visual quality loss, video stabilization technology minimizes unintentional jitters and shakes of an object capturing equipment without affecting moving subjects or purposeful camera panning. This is especially important for handheld imaging devices because of how much more susceptible they are to vibrations. Unwanted changes in camera position led to unstable image sequences, but controlled movement commonly generate unstable images.

Secondly, the Noise Reduction using Object Separation and Background Normalization (NR-OS-BN) Algorithm is discussed.

Algorithm - II: Noise Reduction using Object Separation and Background Normalization (**NR-OS-BN**) Algorithm

---

Input:
Motion Stabilized Video data as V1[]

Output:
Noise Reduced Video data as V2[]

Process:
Step - 1. Read the total video data as V1[]
Step - 2. For every frame in V1[] as V1[i]
    a. Extract the Background as BG[i]
    b. Extract the Foreground as FG[i]
Step - 3. Calculate the BG_Mean and FG_Mean from BG[] and FG[]
Step - 4. For each information set in BG[] as BG[k]
    a. If BG[k] == BG[k+1] and FG[k] == FG[i+1]
    b. Then, K[][]= V1[k]
Step - 5. For each element in K[][] as K[j][]
    a. If K[j][].BG > BG_Mean and K[j][].FG > FG_Mean
    b. Then,
        i. K[j][].BG = BG_Mean and K[j][].FG = FG_Mean
        ii. V2[]=K[j][]
Step - 6. Return V2[]

The process of reducing noise from a signal is known as noise reduction. Both audio and picture noise reduction methods exist. Algorithms for noise reduction may slightly skew the signal. As with common-mode rejection ratio, noise rejection refers to a circuit's capacity to separate an undesirable signal component from the desired signal component.

Finally, the Key Frame extraction using Moving Average Dynamic Thresholding (KFE-MA-DT) Algorithm is discussed.

Algorithm - III: Key Frame extraction using Moving Average Dynamic Thresholding (**KFE-MA-DT**) Algorithm

---

Input:
Noise Reduced Video data as V2[]

Output:
Reduced Key Frame set as KF[]

Process:
Step - 1. Read the total video data as V2[]
Step - 2. For each frame in V2[] as V2[p]
    a. Extract the frame threshold as TH[p]
    b. Calculate the Mean Threshold as MT using Eq. 41
    c. If TH[p] > MT
    d. Then, KF[r] = V2[p]
    e. Else, Discard the frame
Step - 3. Return KF[]

Further, based on the proposed algorithms, the framework to automate the process is furnished [Fig. 1].
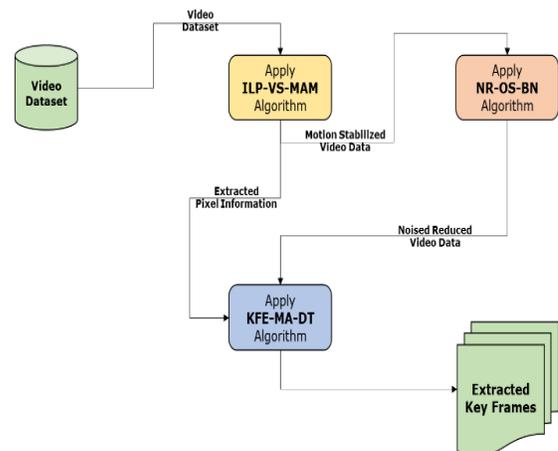
Fig. 1.   Proposed Framework.

Furthermore, in the next section of this work, the obtained results are discussed.

IX. RESULTS AND DISCUSSIONS

After the detailed discussions on problems and proposed solutions, in this section of the work, the obtained results are discussed. Firstly, the dataset information is furnished [Table I].

TABLE I. DATASET INFORMATION [22]

Attributes	Number (#)
Number of Videos	1900
Average number of Frames	7247
Number of Categories	8

The number of samples in the dataset are 1900, however due to representation purposes, only 10 samples are listed. Secondly, the video data stabilization outcomes are furnished [Table II].

TABLE II. VIDEO MOTION STABILIZATION OUTCOMES

Dataset ID	Before Processing		Dataset ID	After Processing		Dataset ID
	Mean Heights	Mean Width		Mean Heights	Mean Width	
1	240	320	1	240	320	1
2	240	320	2	240	320	2
3	240	320	3	240	320	3
4	240	320	4	240	320	4
5	240	320	5	240	320	5
6	240	320	6	240	320	6
7	240	320	7	240	320	7
8	240	320	8	240	320	8
9	240	320	9	240	320	9
10	240	320	1	292	292	2

It is natural to realize that, due to the change of camera angle for each frame, the images are converted to a square image frame and the camera angles are justified. The same results are visualized graphically here [Fig. 2].
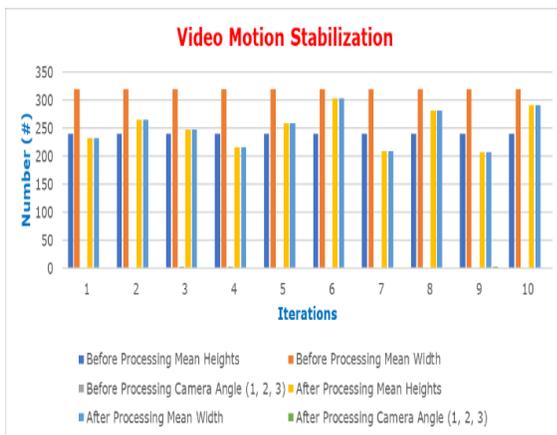
Fig. 2. Video Motion Stabilization Outcomes.

For a single video data, few frame-by-frame analysis are furnished here [Table III].

TABLE III. SINGLE VIDEO FRAME-BY-FRAME ANALYSIS

Dataset ID	Before Processing				After Processing			
	Mean Heights	Mean Width	Camera Angle (1, 2, 3)	FPS	Mean Heights	Mean Width	Camera Angle (1, 2, 3)	FPS
1	240	320	3	10	265	229	2	10
	240	320	3	10	259	269	3	10
	240	320	3	10	212	251	2	10
	240	320	3	10	282	186	1	10
	240	320	3	10	293	173	2	10
	240	320	3	10	276	114	2	10
	240	320	3	10	201	247	2	10
	240	320	3	10	293	235	3	10
	240	320	3	10	131	175	2	10
	240	320	3	10	248	168	2	10

Thirdly, the analysis of the noise reduction process is analyzed [Table IV].

TABLE IV. VIDEO MOTION STABILIZATION OUTCOMES

Dataset ID	Mean Initial Noise (dB)	Mean Reduced Noise (dB)	Time (ns)
1	5.7684727	2.7311466	1.722
2	4.3735304	1.9569633	7.556
3	5.286506	3.096948	7.395
4	4.7074976	2.7789779	7.110
5	4.768707	2.8522859	3.160
6	4.387303	2.6379135	9.517
7	4.337146	2.4204903	6.397
8	3.9589996	2.2259648	8.053
9	3.7900214	2.1423862	1.512
10	3.7343879	2.0677133	6.911

Thus, it is natural to realize that the noise reduction [Fig. 3] with a minimal time complexity [Fig. 4] is achieved.
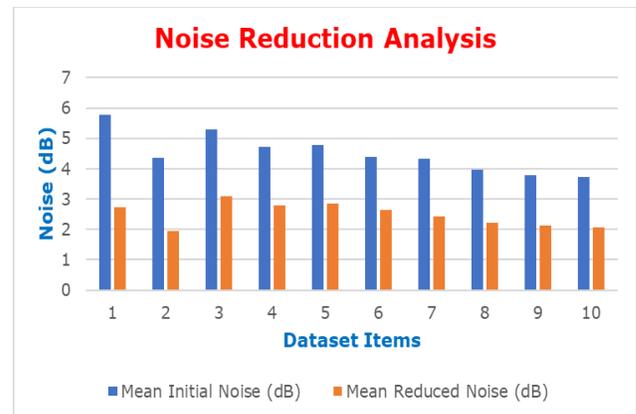
Fig. 3. Noise Reduction Analysis.
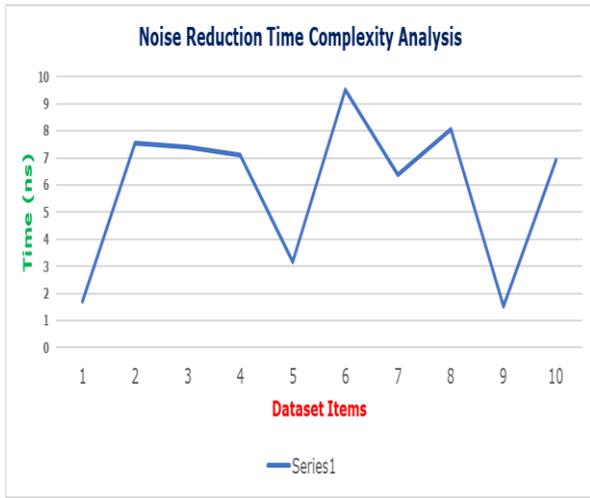
Fig. 4.   Noise Reduction Time Complexity Analysis.

Finally, the key frame extraction outcomes are analyzed [Table V].

TABLE V.   KEY FRAME EXTRACTION ANALYSIS

Dataset ID	Initial Number of Frames	Number of Key Frames	Adaptive Threshold	Time (ns)
1	150	119	90456	8.788
2	61	56	96709	4.229
3	181	83	98426	6.131
4	141	32	99201	7.822
5	131	61	94871	3.244
6	81	16	95059	6.798
7	151	148	95206	5.393
8	71	64	91944	6.577
9	111	61	90952	7.671
10	151	72	90036	2.254

Henceforth, the reduction in terms of key frames [Fig. 5] is clearly visible with finite number of iterations, which again reflects in the minimal time complexity [Fig. 6].

Fig. 5.   Key Frame Extraction Analysis.
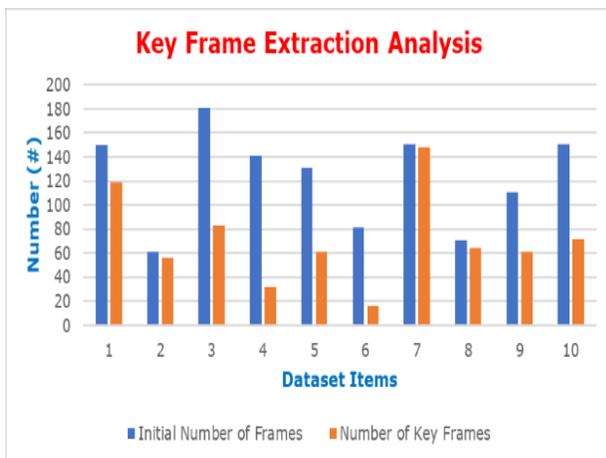
Fig. 6.   Key Frame Extraction Time Complexity Analysis.

Henceforth, with the complete confirmation of the achieved results discussed in this section of the work, in the next section the obtained results are compared with other parallel research works.

X.   COMPARATIVE ANALYSIS

After the detailed discussion on the proposed solutions, algorithms and the results, this section of the work, is dedicated to summarizing the outcomes and further compare with significant parallel research works [Table VI].

TABLE VI.   COMPARATIVE ANALYSIS

Author, Year	Proposed Method	Key Frame Reduction (mean) (%)	Model Complexity	Time Complexity (mean) (ns)
Y. Zhang et. al. [2], 2020	Key Frame Extraction	49%	$O(n^2)$	9.58
P. Sasmal et. al. [7], 2021	Key Frame Extraction	51%	$O(n^2)$	7.55
X. Wang et. al.[8], 2022	Key Frame Extraction	63%	$O(n^2)$	7.85
Proposed Method	Motion Stabilization, Noise Reduction, Key Frame Extraction	65%	$O(n)$	5.89

It is natural to realize that the proposed method has outperformed the parallel research works. Henceforth, in the next section of the work, the research conclusion is presented.

XI. CONCLUSION

Over the past decade, research into video information processing has been one of the most concentrated efforts, with many studies focusing on pre-processing concerns. Researchers are unable to use information retrieval or categorization algorithms due to pre-processing problems such as inconsistent video frame rates or collection angles, noisy data, and enormous file sizes. Due to the popularity of using video as a means of presenting information, addressing these

issues is equally crucial. The purpose of this study is to automate the processes of motion stabilization, noise reduction, and key frame extraction so that these tasks can be completed in less time and with fewer losses of information. This work initially applies the Information Loss Preventive Video Stabilization using Mean Angle Measure (ILP-VS-MAM) Algorithm to stabilize the motion across the frames, secondly applies the Noise Reduction using Object Separation and Background Normalization (NR-OS-BN) Algorithm to reduce the noise from each frame and from the overall video data and finally, applies the Key Frame extraction using Moving Average Dynamic Thresholding (KFE-MA-DT) Algorithm to extract the minimal amount of key frames with reduced time to complete all the processes. The study leads in a 66% reduction in key frame extraction time and a duration of roughly 6 ns for processing all of the video data.

REFERENCES

[1] Y. Li, A. Kanemura, H. Asoh, T. Miyanishi and M. Kawanabe, "Multi-Sensor Integration for Key-Frame Extraction From First-Person Videos," in IEEE Access, vol. 8, pp. 122281-122291, 2020.

[2] Y. Zhang, D. Zhu, H. Bi, G. Zhang and H. Leung, "Scattering Key-Frame Extraction for Comprehensive VideoSAR Summarization: A Spatiotemporal Background Subtraction Perspective," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 7, pp. 4768-4784, July 2020.

[3] C. Huang and H. Wang, "A Novel Key-Frames Selection Framework for Comprehensive Video Summarization," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 2, pp. 577-589, Feb. 2020.

[4] Q. Zhong, Y. Zhang, J. Zhang, K. Shi, Y. Yu and C. Liu, "Key Frame Extraction Algorithm of Motion Video Based on Priori," in IEEE Access, vol. 8, pp. 174424-174436, 2020.

[5] Z. Wang and Y. Zhu, "Video Key Frame Monitoring Algorithm and Virtual Reality Display Based on Motion Vector," in IEEE Access, vol. 8, pp. 159027-159038, 2020.

[6] C. Feng, Z. Xu, S. Jia, W. Zhang and Y. Xu, "Motion-Adaptive Frame Deletion Detection for Digital Video Forensics," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 12, pp. 2543-2554, Dec. 2017.

[7] P. Sasmal, A. Paul, M. K. Bhuyan, Y. Iwahori and K. Kasugai, "Extraction of Key-Frames From Endoscopic Videos by Using Depth Information," in IEEE Access, vol. 9, pp. 153004-153011, 2021.

[8] X. Wang et al., "Robust Unsupervised Video Anomaly Detection by Multipath Frame Prediction," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 6, pp. 2301-2312, June 2022.

[9] H. Zhang, R. Wang and Y. Zhao, "Multi-Frame Pyramid Refinement Network for Video Frame Interpolation," in IEEE Access, vol. 7, pp. 130610-130621, 2019.

[10] H. Bhuyan, P. P. Das, J. K. Dash and J. Killi, "An Automated Method for Identification of Key frames in Bharatanatyam Dance Videos," in IEEE Access, vol. 9, pp. 72670-72680, 2021.

[11] W. Zhang, M. Zhou C. Ji, X. Sui and J. Bai, "Cross-Frame Transformer-Based Spatio-Temporal Video Super-Resolution," in IEEE Transactions on Broadcasting, vol. 68, no. 2, pp. 359-369, June 2022.

[12] Y. Guo, Z. Li, Y. Liu, G. Yan and M. Yu, "Video Object Extraction Based on Spatiotemporal Consistency Saliency Detection," in IEEE Access, vol. 6, pp. 35171-35181, 2018.

[13] X. Peng, R. Li, J. Wang and H. Shang, "User-Guided Clustering for Video Segmentation on Coarse-Grained Feature Extraction," in IEEE Access, vol. 7, pp. 149820-149832, 2019.

[14] Z. Gao, L. Wang, N. Jojic, Z. Niu, N. Zheng and G. Hua, "Video Imprint," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 12, pp. 3086-3099, 1 Dec. 2019.

[15] Z. He, J. Li, L. Liu, D. He and M. Xiao, "Multiframe Video Satellite Image Super-Resolution via Attention-Based Residual Learning," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-15, 2022, Art no. 5605015.

[16] H. Guo, X. Wu and N. Li, "Action Extraction in Continuous Unconstrained Video for Cloud-Based Intelligent Service Robot," in IEEE Access, vol. 6, pp. 33460-33471, 2018.

[17] Q. Yang, D. Yu, Z. Zhang, Y. Yao and L. Chen, "Spatiotemporal Trident Networks: Detection and Localization of Object Removal Tampering in Video Passive Forensics," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 10, pp. 4131-4144, Oct. 2021.

[18] Y. Gu, T. Wang, X. Jin and G. Gao, "Detection of Event of Interest for Satellite Video Understanding," in IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 11, pp. 7860-7871, Nov. 2020.

[19] W. Ren et al., "Deep Video Dehazing With Semantic Segmentation," in IEEE Transactions on Image Processing, vol. 28, no. 4, pp. 1895-1908, April 2019.

[20] W. Zhu, J. Lu, J. Li and J. Zhou, "DSNet: A Flexible Detect-to-Summarize Network for Video Summarization," in IEEE Transactions on Image Processing, vol. 30, pp. 948-962, 2021.

[21] B. Du, Y. Sun, S. Cai, C. Wu and Q. Du, "Object Tracking in Satellite Videos by Fusing the Kernel Correlation Filter and the Three-Frame-Difference Algorithm," in IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 2, pp. 168-172, Feb. 2018.

[22] Dataset: M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis. Learning temporal regularity in video sequences. In CVPR, June 2016.

A Review on Approaches in Arabic Chatbot for Open and Closed Domain Dialog

Abraheem Mohammed Sulayman Alsubayhay[1], Md Sah Hj Salam[2], Farhan Bin Mohamed[3]

School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Johour, Malaysia

Abstract—**A Chatbot is a computer program which facilitates human-to-human communication between an artificial agent and humans. The Arabic language unlike the other languages has been used in Natural Language Processing in relatively fewer works owing to the lack of corpus along with the complexity of the language which has a number of dialects extending across various countries across the world. In the current scenario, little research has been conducted in the case of Arabic chatbots. This study presents a review about the existing literature on Arabic chatbot studies to determine knowledge gaps and suggests areas that require additional study and research. Additionally, this research observes that all relevant research relies on pattern matching or AIML techniques. The searching process was conducted utilizing keywords like 'utterance' 'chatbot', 'ArabChat', 'chat agent', 'dialogue', 'interactive agent', 'chatterbot', 'conversational robot', 'artificial conversational', and 'conversational agent'. Further the study deals with the existing studies and the various approaches in Open and Closed domain dialog system and their working in the case of Arabic Chatbots. The study identified a severe lack of studies on Arabic chatbots, and it was observed that the majority of those studies were retrieval-based or rule-based.**

Keywords—Arabic chatbot; artificial intelligence; arabchat; human-machine interaction; conversational agent

I. INTRODUCTION

Artificial Intelligence (AI) research works like natural language processing (NLP) aids in developing new software solutions in computer science. One of them is the Chatbot or Chatterbot which is a high-intelligence technology. It was created to increase human-computer interaction by simulating a true human conversation by the use of diverse technologies and computer applications. Numerous areas of life, including customer support and companions for mental health, have seen extensive usage of chatbots. Arabic language chatbots powered by NLP and AI have been relatively rare because of the Arabic language complexity, despite advances in attaining human-like interactions. NLP advances are more readily accessible to computing resources and the communication techniques have facilitated the chatbot's rapid growth and deployment in several industries. One of the earliest systems of dialogue is Eliza which was created in 1966. Later, Weizenbaum created the ELIZA chatbot to mimic a psychotherapist[1]. PARRY attempted to imitate a paranoid agent in 1971[2], ALICE (Artificial Linguistic Internet Computer Entity) established by Wallace in 1995[3], inspired by ELIZA. They were created to mimic human behaviour in a text-based discussion and to pass the Turing Test within a certain domain.

A. Research Problem

The literature review reveals the chatbot development goals range from specific to wide range. In particular domains, such as healthcare, education, and industry, Chatbots primarily focused on specific topics. Thus, this paper aims to summarize classic, modern and dialect Chatbots for Arabic language. The work comprises of chatbots from close domain as well as open domain. Though English is the most frequently used language for chatbots, the technology has extended to other languages as well. Lack of Arabic corpuses has led to the reduced utilization of the language in several natural language processing technologies like chatbots. Moreover, the complexity of the Arabic language like rich morphology, higher intent of ambiguity, frequent orthographic variants and the number of dialects are also the reason behind the reduced usage of Arabic in Natural Language Processing.

B. Research Question

What are the techniques utilized to overcome the existing researches?

What are the keywords utilized in this study for searching process?

What are the purposes of open and closed domain dialog system?

C. Research Objectives

The study utilized Pattern matching and an AIML technique which documents the chatbots that interacted with users while speaking Arabic. The searching process was conducted utilizing keywords like 'utterance' 'chatbot', 'ArabChat', 'chat agent', 'dialogue', 'interactive agent', 'chatterbot', 'conversational robot', 'artificial conversational', and 'conversational agent'. The closed domain is used in situations where the amount of information needed to produce an appropriate response to an input is restricted. Similar to human conversation, the open domain will grow over time to support multiple conversation domains

D. Research Significance

The purpose of this study is to analyse the body of literature on Arabic chatbot studies, identify knowledge gaps, and suggest areas for further investigation. Additionally, this research notes that pattern matching or AIML approaches are used in every pertinent research. The study also discusses previous research and different Open and Closed domain dialogue system methodologies, as well as how they apply to Arabic chatbots. The study found a severe paucity of research

on Arabic chatbots and found that most of the research was retrieval - or rule-based.

The standard orthographic representation of Arabic not only determines the intended pronunciation of a written word in the language but also sets a special diacritic which is required to indicate the corresponding pronunciation. Various diacritics produce different kinds of words with different possible meanings for the same spelling form. In the case of most genres of written Arabic, however, these diacritics are usually omitted, resulting in an extensive ambiguity in the case of pronunciation and even meaning (in some cases). Even though native learners can typically infer the desired meaning as well as intonation from the sense, Arabic is frequently difficult to analyse automatically. While native speakers can quickly recognise the intended purpose and intonation from the background context, automated Arabic processing is often hindered by the lack of diacritics [4]. The subject of this paper is 'Arabic Chatbots'. The well-established approaches in closed and open domain dialogue system for Arabic language and the Chatbot applications are available for research in Arabic are numerous in number which are studied to a certain extent.

II. BACKGROUND

The Chatter Bot's system design comprises three main components:

- Chat Interface - It is the component that communicates with users directly. Its primary role is to take in chat text as from users, pre-process them, and then send it to the Knowledge Engine and Conversation Engine for processing. It then relays chat text again to the users after receiving input from engines. It features a number of sub modules that make this work easier.

- Knowledge Engine - The variety of topics, which fall inside the purview of the chatbot's conversation are organized in the Knowledge Engine's Topic Hash Table. Every topic covers a wide range of information relevant to the specific domain being used. A series of contextual maps are used to encrypt specific data about every topic item in the hash table. Various goal fulfillment maps have been now used to realize the context maps.

- Conversation Engine - Based on the pre-coded criteria with in goal-fulfillment mappings chosen by the Knowledge Engine; moreover, the Conversation Engine component regulates the conversation path. This controls how the dialogue moves forward [3].

In Artificial Intelligence-Based Systems, machine learning algorithms are frequently utilized to generate replies to the input of the user. Since this system is primarily based on machine learning techniques, it is regarded to be more efficient than rule-based systems. Artificial intelligence-based systems can be categorized into two types[5]. They are:

- Retrieval-based model.

- Generative-based model [2].

Retrieval-based chatbots benefit from providing knowledgeable and fluent answers; they use response selection algorithms to find the appropriate reply for the present discussion from a repository [3]. According to Wang et al., most of the researches on retrieval-based chatbots emphasize on answer assortment of single-turn dialogues which take into account the most recent input message [4]. This kind of chatbot is well suited for use in domain-specific conversational systems [2]. According to Arfan Ahmed et al., scoping review, most of the chatbots are developed based on retrieval-based model [49]. Despite the fact that there are many chatbots in use today, Arabic-language chatbots are rather rare. Furthermore, compared to chatbots for other languages, Arabic chatbots have become less sophisticated. Moreover, the research suggested that extensive generative-based models, input and output modes, and NLP-based Arabic chatbots must be developed by researchers.

The generative learning model is a knowledge-based unsupervised learning paradigm. Large volumes of conversational information can be used by generative-based chatbots to start a new discussion. The generative-based model can combine different learning methods, such as supervised learning, adversarial learning, unsupervised learning, and reinforcement learning. Some trials and investigations were started when Wittrock first presented the sub-model in 1974 [6]. The system is trained so that it can create a reply in the process of a new query. While the Support Vector Machine (SVM) method generates output based on stated rules or database matching, a vast amount of data for training is used by generative model. The system then tries to produce newer replies in response to a new inquiry.

A. Dialogue System

A dialogue seems to be a discussion between two or more entities, whether they are artificial or human. Human-human conversation and human-computer conversation are the two main areas of dialogue studies. The latter participates in a dialogue system, a computerised system its aim is to converse in natural language with people. The development of dialogue systems nowadays includes graphic, spoken, written, and multimodal platforms. The systems which are adapted for the purpose of on-the-job learning are the dialogue frameworks. They have the ability to communicate with the users and ensure that proper feedbacks are received for further advancements[8]. Dialogue Systems are studied in a number of papers[4, 8-13].

1) Dialogue system design: As compared to the other speech and language processing domains, the user holds a great part in the dialogue systems. This makes it closely inclined to Human-Computer Interaction (HCI). The term Voice Strategy is used to describe the error messages, dialogue strategies and prompts. It follows the principles which is user-centred[15].The first of the principles is the studying the user and task concerned. Similarly, after the process, building of simulations and prototypes must be done followed by an iterative testing of the design on the users. The step by interpretation is shown in the form of a Fig. 1 as shown.
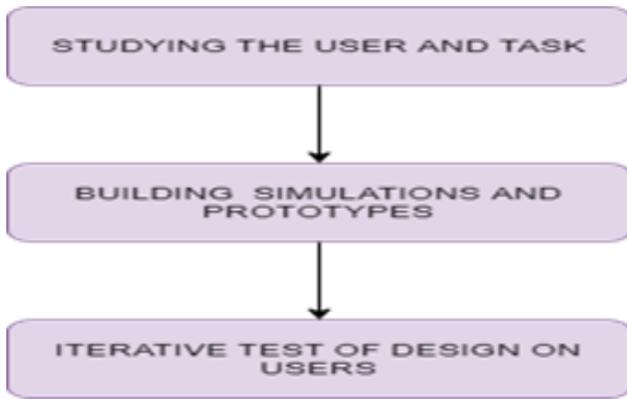
Fig. 1.   Steps in Dialogue System Design.

*2) Classification of dialogue systems.*The dialogue systems' classification is an important aspect of all the research papers involved[4, 9, 16-18]. The Dialogue systems are basically divided into two types[14]. It is shown in the Fig. 2.

- Task-Oriented and
- Non-Task Oriented/ Chatbots.
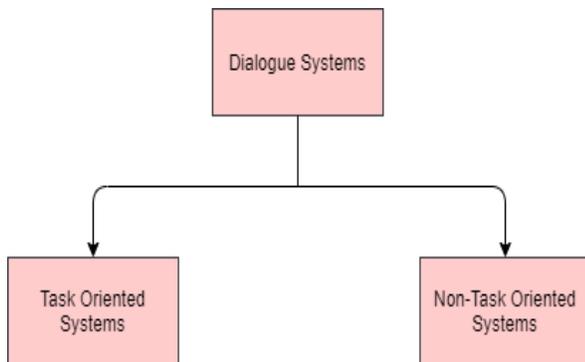
Fig. 2.   Types of Dialogue Systems.

According to (Gao et al., 2019; Zhang et al., 2020) the conversational systems are classified into three main types as shown in Fig. 3. They are

- Answering Agents for questions.
- Task-Oriented Dialogue Agents.
- Chatbots.
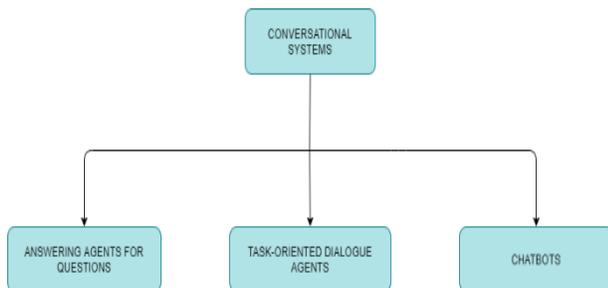
Fig. 3.   Classification of Conversational Systems.

The classification of Dialog system has been studied by many researchers. Lee and Dernoncourt studied the recurrent and convolutional neural network with regard to preceding short texts. In the case of dialog prediction it achieves results based upon three different datasets. The model though produces predictions and performance of better quality, it is restricted to the sequential model it uses(Lee and Dernoncourt, 2016). The study makes it very evident that the Dialogue system seems to be an excellent instrument for user communication with any programme. This is a useful tool that can be employed with a variety of gadgets, including smartphones, PCs, and telephones. It might be a useful tool for helping users navigate websites that deal with topics like stock transactions, online shopping, information search, remote banking, route planning, etc.

B. Chat bots: An Overview

The human lives are integrated with Artificial Intelligence and this is evident from the creation of intelligent software as well as hardware along with the analysis of the same. One of them is the work of Himanshu and Rizwan who studied the mental model in Human Computer Interaction[19]. It has the capability to understand multiple languages of humans through the technology of NLP[20]. The functioning of AI systems were studied in[20], the turning tests and the concerned problems thereby redefining the machine intelligence. The importance of the work relies on its discussion over businesses like e-commerce and the other related domains. A much more detailed working of the chatbot system with the language processing system is needed in the work[21].

Chatbots can be categorized into Rule-Based or Corpus-Based systems, which are discussed as follows:

3) Rule-based chatbots: The Rule-based Chatbots or the Decision Tree Bots work based upon a series of well-defined rules or principle. A sample conversation of a Rule-based Chatbot is shown in Fig. 4. Conversations are mapped out by these chatbots in order to lay out a plan meeting the expectation of the customer and the chatbots must respond. Some of the famous rule based chatbots are ELIZA and PARRY[22].

4) Corpus-based chatbots: The Corpus-based Chatbots unlike the rule based chatbos mine the human-human conversation. A sample conversation of a Corpus-based chatbot is discussed by various researchers in one of his papers on spoken dialogue systems[18]. It is shown in Fig. 5. One of the advantages of using them is they are not based on framed out rules. Hence they require huge amounts of data for training[23].

General intent Chatbot's knowledge base must have been small, simple, and user-friendly to understand. Despite the fact that certain of the commercial solutions have only recently appeared, advancements must be done to discover a standard method for creating Chatbots.

The majority of corpus-based chatbots respond to user turns in contexts either utilising retrieval techniques (employing information retrieval to find a reply from a corpus,

which is suitable provided the dialogue contexts) or generation techniques (employing a language prototype or encoder-decoder to produce the reply provided the dialogue contexts). For interactions that are brief enough to fit into a particular model's frame, systems typically produce a solitary respond turn that is acceptable given the entirety of the dialogue up to that point. They are frequently referred to as response generation algorithms for this purpose. Thus, corpus-based chatbot methods borrow from question answering framework methods, which also emphasise single replies while disregarding longer-term conversational objectives.
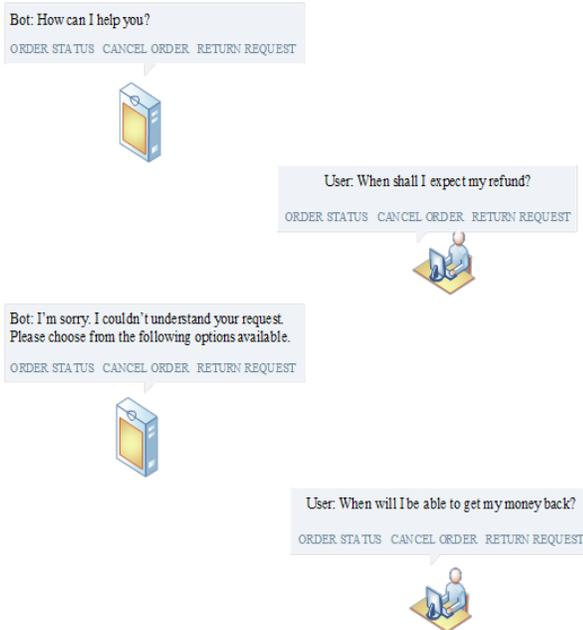
Fig. 4. Sample Conversation of Genius.

Fig. 5. Conversation of a Corpus based Chatbot.

A part of the research survey, databases of different literatures were used similar to the literature review of any paper[24]. The bibliography was compiled using widely referenced computer science resources such as Google Scholar, IEEE, Science Direct, Springer, and ACM. The searching process was conducted utilizing keywords like 'utterance' 'chatbot', 'ArabChat', 'chat agent', 'dialogue', 'interactive agent', 'chatterbot', 'conversational robot', 'artificial conversational', and 'conversational agent'. Papers were collected in the years ranging between 2004 and 2021 and reviewed based on their title and abstract, omitting those that did not demonstrate an Arabic chatbot implementation, which has been from the Google Scholar, Springer libraries, and IEEE. Studies documenting chatbots that interacted with users while speaking Arabic were considered in this investigation, irrespective of the users' categories, input and output methods, platforms, or topic of interaction. Nevertheless, chatbots that have been managed by a human operator were not included in this study.

III. RELATED WORK OF ARABIC CHATBOT

Despite the fact that Arabic chatbot implementations have been built, little research has been done on them. Al-Haj Bot, El-Kahwagy, Msa3ed, Theyabi, Rammas, and other services that offer Arabic chatbot applications created for commercial use seem to be some instances of the former. Additionally, there have been platforms like Watson by Telegram Bot, Messenger Bot via Facebook, IBM, and PandoraBot that offer creators coding assistance. Due to the paucity of published studies on these systems and apps, it was difficult to conduct a fair analysis and comparison. This study's objective is to showcase cutting-edge Arabic chatbot investigation. The Arabic language is classified into three dialects:

- Classical Arabic - Old Written type.

- Modern Standard Arabic (MSA) – The commonly used type.

- Dialectal Arabic – The dialect of natives across various regions.

Since the Middle Ages, the language's written form has deviated from dialectic Arabic. Although Modern Standard Arabic is basically written, the regional dialects are frequently used in spoken communication and daily interactions. In contrast to MSA, the dialects lack written grammatical rules and have their own morphology, syntax, and phonetics [4]. Chatbots are often created using either generative or retrieval-based methods. While retrieval-based methods profit from relevant and eloquent replies, generative approaches can produce additional relevant replies that wouldn't have surfaced in the corpus [25]. Pattern matching is a common technique for constructing the conversational agent in a retrieval-based model. In question-and-answer chatbots, where the machine compares the inputs to a specified framework and creates a reply, pattern matching has been most frequently used. Additionally, chatbots can have either brief interactions (one reply is generated for each Q&A session) or lengthy interactions (large number of data exchanged throughout interactions). Chatbots might be open (i.e., strive to serve various sorts of communication on various topics) or closed (i.e., fulfill a specific goal or topic; the information necessary for output answer is restricted) [49]. Chatbots have been

further splitted into two subcategories predicated on their domain of usage. They are:

- Closed domain.

- Open domain.

A. Closed Domain

The closed domain is used in situations where the amount of information needed to produce an appropriate response to an input is restricted. The creation of the ArabChat has taken a significant amount of time and effort. Table I shows a list of the closed conversation chatbots that has been discussed. In [26] Brini, Ellouze, Mesfar, and Belguith suggest the Question-Answering System for Arabic Language (QASAL) as an Arabic Question-Answering (Q-A) framework. A natural language question typed in Modern Standard Arabic (MSA) is provided to QASAL, which then outputs the best possible response. The system consists of three subsystems: a question assessment module, a text retrieval subsystem, and a reply extracting subsystem. These three components were processed using the language development tools known as the NooJ Platform. The content of the Quran restricts the conversation's scope. The foundations of the chatbot database are therefore retrieval-based. The group of the greatest crucial terms from the "Ayhas" that indeed java implementation pulls restricts the chatbot's interactivity and responses. The dialogue is brief since the chatbot responds to single user input with a single response.

Shawar demonstrates the usage of a chatbot in accessing Arabic Web Question Answering (QA) corpus without resorting to advanced natural language processing or inference. Java application was developed for converting textual corpus to AIML[27] chatbot linguistic model format, and the chatbot dataset was retrieved from multiple sources. Since it is impossible to predict regardless of whether a client will query the ALICE information base with the same queries, the standard file has been built using the initial and most important word techniques. Users could use Spanish, English, and Arabic to engage FAQ chatbot platforms. A modest collection of 412 Arabic QA spanning five themes was also created from a variety of Web sites, including fasting and its linked health concerns, women and their reproductive wellness, dental care, blood illnesses like diabetic and cholesterol, and donating blood. Additionally, the user submits a textual query in MSA to the chatbot regarding one of the available domains, and the chatbot responds with the appropriate response[28] [29]. The user enters a text inquiry in MSA regarding one of the allowed domains, as well as the chatbot responds with the appropriate response even without usage of complicated NLP. The text corpus has also been converted using Java software to produce the explosive and standard AIML documents. The queries and responses that occur in the corpus are contained in the binary file. The utilization of the usual file ensures that the customer's query is translated to the relevant query kept in the information base. Additionally, the file is constructed utilising the first word as well as most important word method. The most important word seems to be the least common in the query, while the first word serves as a classifier for the query. After tokenizing the query, the latter is accomplished by creating a list of frequently asked questions. The terms in the query, together with respective frequencies, are included in the created list. The method then takes the two least common terms from the list—which are also the two most prominent words—and uses them as keywords to connect the inquiry to a response. The most crucial word strategy was used to boost the anticipated output's rate. The chatbot was evaluated with fifteen questions, with a success rate of 93 percent. The primary disadvantage of this paradigm is that if the structure of the question is in an altered form in the knowledge base, then what will be stored, then the replies of the chatbot will be an incorrect one. The chatbot doesn't employ a heuristic to choose the appropriate response but instead relies on a direct retrieval mechanism. Additionally, when the Arabic query is rewritten in its original form, one of two things occurs: either no response is generated, or a similar but erroneous response is obtained. Moreover, a success rate of 93 percent is unjustifiable when using a sample of only fifteen questions.

Abdullah [4] implements an approach known as Pattern Matching. Abdullah CITS has been basically an Arabic Conversational Intelligent Tutoring System (CITS) which educates about 10 to 12 fundamental Islamic concepts. This online system is capable of communicating with MSA students. These pupils are given a series of questions and then addressed their responses in Classical Arabic, citing evidence from the Hadith and Quran or the traditions and sayings of the Prophet Muhammed of Islam. The system communicates with pupils via pictures and sound effects, and it is capable of determining a student's level of expertise and guiding the dialogue. Abdullah CITS is capable of distinguishing between a user's questions and responses. A comprehension base with subject-specific data, a communicative agent coding language for providing tutorial conversations to trainees, and a tutorial comprehension base for assessing every student's abilities of subject-specific insight make up the system, which uses a Pattern Matching method [30].

A. Shawar and E. Atwell published one of the early studies on Arabic chatbot applications [31]. They communicated with the ALICE chatbot (Artificial Linguistic Internet Computer Entity). Here, the techniques of Machine learning (ML) were employed to create an Arabic chatbot that receives input of the user in Arabic and answers with Quranic quotes. The Quran is comprised of 'Ayahs' or 6236 verses and 114 'Surahs', or collections of verses. The user provides the Arabic words with 'Tashkil' or diacritics in the form of phonetic guidance. The chatbot then responds by searching the Quran for "Ayahs," containing the user's comments. Since Quran text is not conversational, Java software was designed to establish a learning mechanism. The learning process is guided by the Ayah, the most critical word in the AIML file and prototype's "Ayha" category. The Java software creates an AIML file in Arabic. The chatbot dataset origins are retrieval-based. However, feedback from Arabic evaluators indicates that entering vowels in Arabic words is difficult. Another argument that does not satisfy users is the random selection of an object from the list; if the word(s) is repeated in more than one Ayya, one is chosen randomly. They also found that not all words have a response related to the least frequently used word technique. The content of the Quran also limits the

dataset; hence this is a retrieval-based model. This project aims to demonstrate how the ALICE chatbot has evolved to learn from non-conversational content. The chatbot's ability to engage and respond is constrained by the set of "Ayhas" words, which the java programme has identified as being the most important.

Hijjawi, Bandar [32] develops ArabChat, the first version of Chatbot of Arabic. ArabChat is a web-based conversational agent system. The chatbot conversation is Arabic task-oriented, and it was created to support students at Jordan's Applied Science University. Pattern Matching is used to manage user conversations in the system for creating the Arabic Conversational Agent (ArabChat). The hybrid rule mechanism and the user classification approach were used to solve issues in the first version of ArabChat.

ArabChat Mobile [33] is a mobile-based conversational agent which serves as a student counselor at Amman's Applied Science University based on ArabChat's original code[32]. It is a streamlined Android version of ArabChat. The ArabChat CA and Mobile ArabChat system are rule-based CAs that include a scripting language, a temporal memory, an information container brain, a scripting engine, and a user interface. Additionally, the Mobile ArabChat employs the Pattern Matching PM method to handle Arabic textual chats. Although users in Arab nations suffer several obstacles, including poor and inconsistent internet connections and limited capacity, this program operates despite these limitations. Mobile ArabChat employed a pattern matching method predicated on text. In terms of its programming engines, programming language, as well as knowledge repository, this architecture is similar to ArabChat. 96 percent of customers concur that utilizing Mobile ArabChat on a smart phone is preferable to doing so on a computer. On the other hand, mobile ArabChat requires internet connection for working. This result may appear satisfactory at first glance, but for the most reliable results, manual analysis of the Mobile ArabChat logs is required. The outcome revealed that 83.2 percent of the inputs corresponded to the anticipated outcome. However, Mobile ArabChat requires an internet connection to function.

In [34] an Arabic interaction platform or conversational interface is presented by Moubaiddin, Shalbak, Hammo, and Obeid. It is intended to communicate with hotel guests and generate responses about booking hotel rooms and other operations. The technology makes use of text-based human language conversation. The two key components in this architecture are the analyzer and the dialogue controller. They have chosen to use the Government and Binding (GB) theory to construct a GB-based analyzer [35]. Users can bargain a desired appointment, as well as question about the rooms and amenities offered by the hotel. They report the findings of an actual study that involved 500 participants who've been total newcomers to the method. The system was to be used by the customers, who have been then asked to score the conversations as "extremely bad," "bad," "moderate," "good," or "extremely good." 92.3 % of the conversations received "good" or "extremely good" ratings, making up 66.92 percentage of the total. The usage of an Arabic conversation system to solve the issue of participatory Arabic dialogues is

supported by these outcomes. The system, which allows users to make hotel reservations using written Arabic text, could be improved. The system does not employ a morphological analyzer to provide important word characteristics such as clitics, number, and gender. Additional criteria for including more phrase patterns and other grammatical elements, including subject-verb alignment on numbers and gender, aren't also included.

Al-Ajmi, A. H., & Al-Twairesh, N. suggest a hybrid rule-based/data-driven method in [36] for a text-based flight booking DS ability of handling consumer utterances. The suggested DS was developed using the natural language interface Wit.ai. The Wizard of Oz technique was used to establish the conversation flow, and crowdsourced training samples were used to develop the DS intents and entities. The evaluation outcomes indicate that the designed system has been capable of comprehending user words and self-feeding effectively. They proposed a hybrid method to data science development that addresses the difficulty of establishing a data science system with limited training data. Furthermore, they designed a DS that utilises a hybrid method to fulfil flight booking tasks by leveraging current DS programming frameworks. Additionally, they evaluated the method in two steps to ascertain the DS's capacity for self-feeding. They use a pipeline architecture methodology and a combination of rule-based and data-driven methodologies. Users can text-book airline tickets in Arabic using this method. According to the assessment results, the suggested methodology was a quick and easy way to book tickets as users gained experience. As more information about member booking times was provided, the system learned. Effectiveness of the DS was severely hampered by the lack of training examples, which researchers tried to lessen by giving the DS self-feeding capacities. They did not, however, imitate real-world flight booking systems. In addition, they did not integrate their DS with an existing reservation system. Although the total experience of ordering a flight ticket via the created flight booking DS was favourable, 52 per cent of participants reported having difficulties comprehending Telegram commands, believing they needed to restart the discussion in order to begin booking. As a result of the misunderstanding, the orders' explanation message is displayed. Additionally, there are no buttons or a greeting message to send following the description of the commands. Another issue was that their system did not notify the user when to begin the booking.

Hijjawi, Bandar [37] and Hijjawi, Bandar [38] developed a previous research project called ArabChat. ArabChat closed domain in Jordan's Applied Science University was chosen as a knowledge point guide for their innate Arabic students in the chat. They use a hybrid rule mechanism as well as a user classification approach. After that, the revisions were included in the final output. The utterance classification function makes an attempt to distinguish between utterances that are and are not inquiries. It accomplishes keyword matching by supplementing the pattern of the question-based rule with additional keywords. The second function is Hybrid Rule, which is concerned with responding to and dealing with an utterance that encompasses multiple themes. Manual examination generated more precise results and shown

performance enhancement, even if ArabChat provided a Ratio of Matched Utterances to the Total (RMUT) outcome that was more precise than the enhanced one brought about by unintellectual users. While carefully reviewing logs, Improved ArabChat correctly processes 82% of phrases with two subjects, and this proportion decreases as the amount of items in the phrase increases. Al Humoud and Al Wazrah [7] claim that is manually classifying utterances reveal a large proportion of query-based phrases because of three factors: the area chosen, the users' demands, which lead them to ask rather than discuss, and the complexity of scripting multiple rules.

TABLE I.   SHOWS A LIST OF THE CLOSED CONVERSATION CHATBOTS

Author	Language	Domain	Approach
Shawar, B. A., & Atwell, E 2004	English/ classical Arabic	Closed (Quran book)	AIML
Shawar, B. A. 2011 and Brini, W., Ellouze, M., Mesfar, S., & Belguith, L. H. 2009	Arabic MSA	Closed Medical care	AIML
Shawar, A., & Atwell, E. S. 2013	Classical Arabic	close domain, quran book	AIML
Alobaidi, O. G., Crockett, K. A., O'Shea, J. D., & Jarad, T. M. 2013	Classical Arabic /MSA	Close(teach Islam for children)	Pattern matching
M. Hijjawi, Z. Bandar, K. Crockett, & D. Mclean 2014	Arabic MSA	Closed (for students of Applied Science University)	Pattern matching Rule-base
Hijjawi, M., Qattous, H., & Alsheiksalem, O 2015	Arabic MSA	Closed (for students of Applied Science University)	Pattern matching Rule-base
Moubaiddin, A., Shalbak, O., Hammo, B., & Obeid, N. 2015	Arabic MSA	Closed Hotel reservation	GBbased parser
M. Hijjawi, Z. Bandar, & K. Crockett 2016	Arabic MSA	Closed (for students of Applied Science University)	AIML
Aljameel, S. S., O'Shea, J. D., Crockett, K. A., Latham, A., & Kaleem, M. 2018	English/ Arabic MSA	Close (for children with Autism Spectrum Disorder)	Pattern matching Rule-base
Al-Ghadhban, D. and N. Al-Twairesh 2020	Arabic MSA	Close for student of King Saud University Information Technology (IT) acting as an academic advisor	AIML
Al-Ajmi, A. H., & Al-Twairesh, N. (2021)	MSA	One domain for flight booking	hybrid rule-based and data-driven

LANA was developed to help autistic children with Autism Spectrum Disorder (ASD) aged 10-16 years old learn more effectively by adapting Visual Auditory Kinaesthetic (VAK) learning styles. They have a basic understanding of the mechanics of Arabic writing and can use MSA to teach them science topics. For kids with ASD, conventional education is challenging since the teacher can't cater to the requirements of every single student. Using pattern matching as well as a short text similarity method, LANA seems to be equivalent to Abdullah CITS (Arabic Conversational Intelligent Tutoring Sys-tem), but still it endorses a wide range of learning genres, such as visual, hearing, and proprioceptive. This allows kids to practise training skills autonomously predicated on their unique needs [39].

In [40] created "Nabiha," a chatbot that can converse with King Saud University Information Technology (IT) students employing the Saudi Arabia language. Nabiha is accessible on Twitter, web, and Android, among other platforms. As an academic adviser, the Nabiha can converse with students and answer their questions about the IT department's courses or their academic achievement at KSU. Thirteen students practice the Nabiha chatbot and offer replies via a survey to ensure the chatbot's usefulness. Twitter have tried Android and web technologies because its textual area only supports a limited number of characters. Several learners who were asked to envision who Nabiha was thought of her as a professional in the IT division, whereas others thought of her as a graduate candidate with substantial training. Nabiha is a robot, according to only one human. The AIML files were created by converting the readable text from the corpus into AIML format using a java program. Nabiha chatbot was launched on the Pandorabots platform after ALIM files were generated, and it was later integrated with Android, Twitter, and the Web. However, even though the dataset contained 1104 categories, Nabiha still needs to be enhanced. HTML tags have an issue, and some sentences are incompatible with Twitter's text area scale, which is very large.

Bendjamaa and Nora proposed a discussion method based on Quranic ontology. The ontology under consideration comprises Quranic chapters and verses and each Quranic term, its origin, and lemma. The system is an Arabic-language natural language communication system. The user's input must be processed first, followed by segmentation, and finally by establishing a semantic route. Additionally, a module is required to access the ontology and retrieve the data. The system admits as input an Arabic spelling string (arrived by the client). This project, which was developed in Java, was implemented using the CoreNLP plugin. Its purpose is to make Islamic knowledge more participatory. The performance, however, is either in Arabic or English, depending on the ontology and the response to the inquiry. If an Arabic version of the answer is available, the output will be in Arabic; otherwise, it will be in English [41]. Although this project permits easy admit to Qur'anic data, it needs to add more querying functionalities. As a result, it does not help a lot of non-Arabic speakers. Syntactic mistakes in the inquiry are some other problems that are not addressed. Additionally, it does not extensively cover Islamic scholarship and legal sources.

B. Open Domain

Similar to human conversation, the open domain will grow over time to support multiple conversation domains.[42] Makatchev and co-authors Hala discusses a bilingual (English

and Arabic) robot administrator at Qatar's Carnegie Mellon University. The investigators compare Hala's English conversation vocabulary to that of a comparable monolingual robotic on the campuses of Carnegie Mellon University in Pittsburgh to verify the installation of Hala (named "Tank"). Hala's purpose is to provide information on campus directions, weather, and area events and respond to inquiries about her personal life. A rule-based conversation manager, which consists of an information base with prepared statements and criteria that trigger responses in English as well as Arabic, has been used by the robot to interact with people. The English information base currently has a much wider range of topics covered. In addition to Hala's spoken answer, a text appears on the monitor next to her image.

IbnSina [43, 44] is a dual-language conversational robot capable of communicating Arabic and English. The user controls it via text or voice inputs. The IbnSina robot responds with audio output in the language specified by the user. The IbnSina robot develops human-to-human contact discourse by utilizing Wikipedia and a locally stored Quran database. As a result, the Chabot of the IbnSina robot can address a wide variety of subjects. As a result, it responds to general inquiries by translating terms, providing online resources, or referring to books stored in its database. Additionally, it warns the user when certain pieces of information are missing or incorrectly spelled. As a result, IbnSina's approach to dialogue is open and extensive. However, because it relies on the information that has already been described in the database to react with the proper output, it does not produce new responses. The IbnSina dialogue system is built on object-oriented methods like the Quran class and Wikipedia class, which permit the robot to respond with the required response, as well as the chatterbot class, which allows for simple discussion and user query responses. Additionally, a chatterbot module has been developed that responds to human input.

In [32] and [45] the ArabChat's effectiveness was increased by differentiating between statements that contained questions and those that did not by classifying Arabic phrases as conversation activities predicated on structural traits found in Arabic function phrases. To study the approaches for categorizing statements into inquiries and non-inquiries using function words, a database of statements has been compiled from diverse sources. These sites were chosen to reflect various issues in the Arabic language, including politics, religion, sports education, and business. 1000 statements make up the resultant database, known as the "CA Database" (500 queries and 500 non-queries). The artificial non-inquiry phrases and the indirect inquiry, in contrast side, were absent from the "CA Database." As the first step in creating the ArabChat categorization method, the "CA Database" was loaded into a ML toolkit for categorization and rule creation. Researchers chose WEKA (Waikato Environment for Information Analysis) as that of the ML toolbox because of its open-source nature. The outcome reveals that 73.56 percent of the inputs corresponded to the anticipated output.

However, ArabChat requires a technique that distinguishes between query-based and non-query-based statements that resolves the same rule. This is because each utterance type (query-based and non-query-based) requires a unique response, even though they target the same rule. Additionally, the number of predefined patterns per rule increases, reducing ArabChat's ability to respond with appropriate responses. Additionally, there is the requirement for a technique that deals with utterances discussing various themes, which necessitates the firing of several rules (each rule has one topic). Finally, it is beneficial to keep the number of required patterns to write to a minimum.

The authors introduce "BOTTA," an Arabic dialect chatbot, in[46]. BOTTA seems to be a communication partner that mimics friendly human interaction by speaking Egyptian Arabic (Cairene). BOTTA is the first Arabic-language chatbot in the world. BOTTA strives to be the Rosie of Arabic dialects; Rosie is a chatbot that communicates in English. BOTTA was created using artificial intelligence and is now available on the Pandorabots platform. The information base of BOTTA has been made up of set data contains themed terms and phrases,, mapping documents storing pairings of linked words and sentences, and AIML documents containing categories including its replies to user inputs. Additionally, the BOTTA chatbot collects essential information about the user during each discussion via questions, resulting in an open dialogue due to the chatbot's ability to react to various topics. It is constructed based on a retrieval-based paradigm. It uses a retrieval-based approach rather than adding new answers or updating the information base. It makes use of algorithms to select a suitable response from a range of pre-written ones. Furthermore, before answering properly, the chatbot doesn't perform text normalisation on the user input. When performing orthographic changes, it also fixes common spelling errors in user input. By employing this technique, BOTTA has been able to fix 85.1% of typical spelling mistakes in Arabic typing. Three people also put BOTTA towards the test. One Levantine Arabic person and two native Egyptian Arabic learners make up the total, which does not meet the requirements for a passing grade.

The authors of [47] give the Arabic Reading Comprehension Dataset (ARCD), which contains 1,395 questions posed by crowd workers on Wikipedia articles, as well as a machine translation of the Stanford Question Answering Dataset (Arabic-SQuAD). The system for open-domain question answering in Arabic (SOQAL) is composed of two components: (1) a document retrieval component that utilizes the hierarchical TF-IDF approach, and (2) a neural reading comprehension component that utilizes the pre-trained bidirectional transformer BERT [24]. Their ARCD trials demonstrate the efficiency of their methodology, with their BERT-based reader attaining a 61.3 F1 score and their open-domain system SOQAL attaining a 27.6 F1 score. However, this approach should increase the size of ARCD. Additionally, this work is not focused on paragraph selection but rather on improving the end-to-end system.

To construct a CA in the Arabic Gulf dialect, T. Alshareef and M. A. Siddiqui used a deep-learning architecture dubbed the Seq2Seq neural network. They framed the CA problem as a machine translation task and trained and evaluated the model using post–reply to tweets. They employed the Bilingual Examination Understudy (BLEU) scores and professional reviewers to assess how pre-trained encoders affected a

convolutional neural network's efficiency. The model fared better than earlier deep learning systems since it was trained on more databases and in more languages [48]. The method has attained the BLEU score as 25.1. However, they do not concentrate on the mechanism for developing a baseline against which this work can be compared. Additionally, when doing the automatic measurement, BLEU, it is necessary to compare the result to many references, as each sentence does not have a single accurate response. Table II shows the table type styles. To improve the model's ability to engage in fluent discourse (i.e., a dialogue with numerous turns), rather than simply reacting to a single remark.

TABLE II. TABLE TYPE STYLES

Author	Language	Domain	Approach	Interaction	
				Input	*output*
Maxim Makatchev , Imran Fanaswala, Ameer Abdulsalam, Brett Browning 2010	English / Arabic (Arabizi / MSA)	Open	AIML Rule-base	Text	Text /voice
Riek, L. D., Mavridis, N., Antali, S., Darmaki, N., Ahmed, Z., Al-Neyadi, M., & Alketheri, A 2010 ,and Mavridis, N., AlDhaheri, A., AlDhaheri, L., Khanii, M., & AlDarmaki, N 2011	**English / Arabic (MSA)**	Open	**Pattern Matching**	Text /voice	**Voice**
Hijjawi, M., Bandar, Z., & Crockett, K 2013 Hijjawi, M., Bandar, Z., Crockett, K., & Mclean, D. 2014	Arabic MSA	Open	**Pattern Matching**	Text	Text
Ali, D. A., & Habash, N. 2016	**Egyptian Arabic dialect**	Open	AIML	Text	Text
Hijjawi, M., Bandar, Z., & Crockett, K 2016&2013	Arabic MSA	Open	Hybrid Rule	Text	Text
Mozannar, H., et al. 2019	Arabic (MSA)	Open	hierarchical TF-IDF and BERT	Text	Text
Alshareef, T., & Siddiqui, M. A. 2020	Arabic Gulf dialect	Open	Rule-based	Text	Text

According to the evaluated papers, retrieval-based modeling is used in all of the work that has been done on Arabic chatbot implementations. In other words, the information pool from AIML documents, databases, or web pages is what the chatbot answers are built on. The key finding of this study is the scarcity of Arabic chatbot literature. Most chatbots were built using a rule-based model. This once more demonstrates how little Arabic is used in more sophisticated systems. Because of the language intricacy, Arabic NLP has been developing more slowly than other languages. Relatively fewer Arabic chatbots use NLP methods. Future research might likely concentrate on figuring out the best way to combine many approaches to chatbot development in order to maximize the usefulness of each approach, even the most straightforward.

IV. DISCUSSION

According to the evaluated papers, retrieval-based modelling is used in all of the work that has been done on Arabic chatbot applications. In other words, the data pool from AIML files, databases, or web pages is what the chatbot responses are built on. It may restrict the chatbot's functionality and usability. Additionally, it should be noted that all relevant work relies on AIML or pattern matching techniques. This could result in 1) a tiny dataset for the chatbot as well as a limitation to closed domains. 2) Limits the chatbot's response to the user, in which the right answer can only be obtained if the user's input fits the chatbot database. Additionally, one of the causes Arabic chatbots in the literature lack certain features is due to the intricacy of Arabic grammar and user errors in spelling and grammar, which explains why there aren't many Arabic chatbot programmes available in both text and speech.

V. CONCLUSION

A review on the available Arabic chatbots is done in the study. The papers collected were basically divided into two categories based on the form of chatbot conversation or interaction as closed and open domain. Because of the complexity of the language and the lack of Arabic Corpuses, the available chatbots in the language is low compared to that of English and hence limited papers are available for the research. This study discovered the dearth of researches on Arabic chatbots and among those works, it is found that the major part of the available literature is retrieval-based or rule-based. Furthermore, the majority of the Arabic Dialog Systems discussed earlier is text-based and employ a rule or pattern-matching technique. Only a fewer of them have developed enhanced system functionality through the use of Short Text Similarity in conjunction with a rule-based approach. Additionally, it has been suggested[30] that the pattern matching approach is preferred despite its limitations due to the difficulties associated with developing Dialog Systems in Arabic. Future studies should perhaps focus on determining the optimal way to integrate many ways to chatbot development in order to maximise the utility of each method, even the simplest.

REFERENCES

[1] Weizenbaum, J., ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM, 1966. 9(1): p. 36-45.
[2] Colby, K.M., S. Weber, and F.D. Hilf, Artificial paranoia. Artificial Intelligence, 1971. 2(1): p. 1-25.

[3] Chakrabarti, C. and G.F. Luger. A semantic architecture for artificial conversations. in The 6th International Conference on Soft Computing and Intelligent Systems, and The 13th International Symposium on Advanced Intelligence Systems. 2012. IEEE.

[4] Elmadany, A.A., S.M. Abdou, and M. Gheith, A Survey of Arabic Dialogues Understanding for Spontaneous Dialogues and Instant Message. arXiv preprint arXiv:1505.03084, 2015.

[5] Surendran, A., R. Murali, and R.K. Babu, Conversational AI-A Retrieval Based Chatbot. 2020, EasyChair.

[6] Jaakkola, T.S. and D. Haussler, Exploiting generative models in discriminative classifiers. Advances in neural information processing systems, 1999: p. 487-493

[7] Hochreiter, S. and J. Schmidhuber, Long short-term memory. Neural computation, 1997. 9(8): p. 1735-1780.

[8] Veron, M., et al., Evaluate on-the-job learning dialogue systems and a case study for natural language understanding. arXiv preprint arXiv:2102.13589, 2021.

[9] Blache, P., et al. Two-level classification for dialogue act recognition in task-oriented dialogues. in Proceedings of the 28th International Conference on Computational Linguistics. 2020.

[10] Bothe, C., et al., A context-based approach for dialogue act recognition using simple recurrent neural networks. arXiv preprint arXiv:1805.06280, 2018.

[11] Joukhadar, A., et al. Arabic dialogue act recognition for textual chatbot systems. in Proceedings of The First International Workshop on NLP Solutions for Under Resourced Languages (NSURL 2019) co-located with ICNLSP 2019-Short Papers. 2019.

[12] Ribeiro, E., R. Ribeiro, and D.M. de Matos, A multilingual and multidomain study on dialog act recognition using character-level tokenization. Information, 2019. 10(3): p. 94.

[13] Serban, I.V., Lowe, R., Henderson, P., Charlin, L., Pineau, J, A Survey of Available Corpora For Building Data-Driven Dialogue Systems. Public Knowledge Project, 2018(dad 9): p. 1–49.

[14] Chen, H., et al., A survey on dialogue systems: Recent advances and new frontiers. Acm Sigkdd Explorations Newsletter, 2017. 19(2): p. 25-35.

[15] Gould, J.D. and C. Lewis, Designing for usability: key principles and what designers think. Communications of the ACM, 1985. 28(3): p. 300-311.

[16] Elmadany, A., S. Abdou, and M. Gheith, Improving dialogue act classification for spontaneous arabic speech and instant messages at utterance level. arXiv preprint arXiv:1806.00522, 2018.

[17] Raheja, V. and J. Tetreault, Dialogue act classification with context-aware self-attention. arXiv preprint arXiv:1904.02594, 2019.

[18] Su, P.-H., et al., On-line active reward learning for policy optimisation in spoken dialogue systems. arXiv preprint arXiv:1605.07669, 2016.

[19] Bansal, H. and R. Khan, A review paper on human computer interaction. International Journals of Advanced Research in Computer Science and Software Engineering, 2018. 8: p. 53-56.

[20] Khanna, A., et al., A study of today's AI through chatbots and rediscovery of machine intelligence. International Journal of u-and e-Service, Science and Technology, 2015. 8(7): p. 277-284.

[21] Shawar, B.A. and E. Atwell. Chatbots: are they really useful? in Ldv forum. 2007.

[22] AbuShawar, B. and E. Atwell, ALICE chatbot: Trials and outputs. Computación y Sistemas, 2015. 19(4): p. 625-632.

[23] Serban, I.V., et al., A survey of available corpora for building data-driven dialogue systems. arXiv preprint arXiv:1512.05742, 2015.

[24] Al Humoud, S., A. Al Wazrah, and W. Aldamegh, Arabic chatbots: a survey. Int. J. Adv. Comp. Sci. Appl., 2018: p. 535-541.

[25] Ji, Z., Z. Lu, and H. Li, An information retrieval approach to short text conversation. arXiv preprint arXiv:1408.6988, 2014.

[26] Shawar, A. and E. Atwell. An Arabic chatbot giving answers from the Qur'an. in Proceedings of TALN04: XI Conference sur le Traitement Automatique des Langues Naturelles. 2004. ATALA.

[27] Wallace, R., The elements of AIML style. Alice AI Foundation, 2003. 139.

[28] Shawar, B.A., A Chatbot as a natural web Interface to Arabic web QA. International Journal of Emerging Technologies in Learning (iJET), 2011. 6(1): p. 37-43.

[29] Brini, W., et al. An Arabic Question-Answering system for factoid questions. in 2009 International Conference on Natural Language Processing and Knowledge Engineering. 2009. IEEE.

[30] Alobaidi, O.G., et al. Abdullah: An intelligent arabic conversational tutoring system for modern islamic education. in Proceedings of the World Congress on Engineering. 2013.

[31] Shawar, B.A. and E. Atwell. Accessing an information system by chatting. in International Conference on Application of Natural Language to Information Systems. 2004. Springer.

[32] Hijjawi, M., et al. ArabChat: An arabic conversational agent. in 2014 6th International Conference on Computer Science and Information Technology (CSIT). 2014. IEEE.

[33] Hijjawi, M., H. Qattous, and O. Alsheiksalem, Mobile Arabchat: An Arabic Mobile-Based Conversational Agent. Int. J. Adv. Comput. Sci. Appl. IJACSA, 2015. 6(10).

[34] Moubaiddin, A., et al., Arabic dialogue system for hotel reservation based on natural language processing techniques. Computación y Sistemas, 2015. 19(1): p. 119-134.

[35] Moubaiddin, A., et al. Investigating the syntactic structure of Arabic sentences. in 2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA). 2013. IEEE.

[36] Al-Ajmi, A.-H. and N. Al-Twairesh, Building an Arabic Flight Booking Dialogue System Using a Hybrid Rule-Based and Data Driven Approach. IEEE Access, 2021. 9: p. 7043-7053.

[37] Hijjawi, M., Z. Bandar, and K. Crockett, The Enhanced Arabchat: An Arabic Conversational Agent. International Journal of Advanced Computer Science and Applications, 2016. 7.

[38] Hijjawi, M., et al., A novel hybrid rule mechanism for the Arabic conversational agent ArabChat. Global Journal on Technology, 2015.

[39] Aljameel, S.S., et al. Development of an Arabic conversational intelligent tutoring system for education of children with ASD. in 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA). 2017. IEEE.

[40] Al-Ghadhban, D. and N. Al-Twairesh, Nabiha: An Arabic dialect chatbot. Int. J. Adv. Comput. Sci. Appl, 2020. 11(3): p. 1-8.

[41] Bendjamaa, F. and T. Nora. A Dialogue-System Using a Qur'anic Ontology. in 2020 Second International Conference on Embedded & Distributed Systems (EDiS). 2020. IEEE.

[42] Makatchev, M., et al. Dialogue patterns of an arabic robot receptionist. in 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI). 2010. IEEE.

[43] Riek, L.D., et al. Ibn sina steps out: Exploring arabic attitudes toward humanoid robots. in Proceedings of the 2nd international symposium on new frontiers in human–robot interaction, AISB, Leicester. 2010.

[44] Mavridis, N., et al. Transforming IbnSina into an advanced multilingual interactive android robot. in 2011 IEEE GCC Conference and Exhibition (GCC). 2011. IEEE.

[45] Hijjawi, M., Z. Bandar, and K. Crockett. User's utterance classification using machine learning for Arabic Conversational Agents. in 2013 5th International Conference on Computer Science and Information Technology. 2013. IEEE.

[46] Ali, D.A. and N. Habash. Botta: An arabic dialect chatbot. in Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations. 2016.

[47] Mozannar, H., et al., Neural arabic question answering. arXiv preprint arXiv:1906.05394, 2019.

[48] Alshareef, T. and M.A. Siddiqui. A seq2seq Neural Network based Conversational Agent for Gulf Arabic Dialect. in 2020 21st International Arab Conference on Information Technology (ACIT). 2020. IEEE.

[49] Ahmed, Arfan, et al. "Arabic Chatbot Technologies: A Scoping Review." Computer Methods and Programs in Biomedicine Update (2022): 100057.

Facial Emotion Detection using Convolutional Neural Network

Pooja Bagane, Shaasvata Vishal, Rohit Raj, Tanushree Ganorkar, Riya

Department of Computer Science and Engineering, Symbiosis Institute of Technology (SIT), Pune, India

Abstract—Non-verbal specialized strategies, e.g. look, eye development, and motions are utilized in numerous uses of human-PC connection, among them facial feeling is generally utilized as it conveys the enthusiastic states and sensations of people. In the machine learning calculation, a few significant separated highlights are utilized for displaying the face. As a result, it won't get a high accuracy rate for acknowledging that the highlights rely on prior knowledge. Convolutional Neural Network (CNN) has created this work for acknowledgment of facial feeling appearance. Looks assume an essential part in nonverbal correspondence which shows up because of the inner sensations of an individual that thinks about the countenances. This paper has utilized the calculation to distinguish features of a face such as eyes, nose, etc. This paper identified feelings from the mouth, and eyes. This paper will be proposed as a viable method for distinguishing outrage, hatred, disdain, dread, bliss, misery, and shock. These are the seven feelings from the front-facing facial picture of people. The final result gives us an accuracy of 63% on the CNN model and 85% on the ResNet Model.

Keywords—Feature extraction; convolutional neural network; resnet; emotion recognition; emotion detection; facial recognition

I. INTRODUCTION

For ideal human-PC interfaces (HCI) would likely want that equipment to produce usable and safe systems [1]. This examination is definitely about how precisely PCs could distinguish feelings correctly from the distinct sensors. This try-out has been applied as a cosmetic picture as a new medium to follow human inclination [2]. Seven important feelings are most inclusive to men and women. Specifically, unbiased, mad, disdain, dread, happy, miserable, and impact and these vital feelings can always be perceived from a new human's look. This specific examination proposes a new compelling way for figuring out unbiased, cheerful, gloomy, and shocking these kinds of four feelings by front-facing cosmetic inclination. During many previous years, different techniques have been proposed for feeling acknowledgment. Numerous calculations were recommended to foster framework applications that can identify feelings well indeed [3]. PC applications could better convey by changing reactions as per the enthusiastic condition of human clients in different connections. The feeling of an individual is not set in stone by discourse or his face or even one's signal [4]. The work introduced in this paper investigates the acknowledgment of looks from the face. Concerning feeling suggestion, the customary procedures ordinarily consider another face picture that is perceived coming from your information picture, and facial broken expressions or accomplishments as a rule are

seen from the specific face regions. Just after that novel space and normal qualities are disengaged out of their facial helpings. At last dependent on the secluded attributes a classifier, with respect to example, the Keras index, sporadic forest, is typically ready to supply affirmations results.

Each of our facial thoughts is expressed through account activation of specific systems of facial muscles. These are sporadic, subtly complex alarms in an expression that frequently contains a wealth of data [5]. Through facial sentiment recognition, we are able to assess the effects that content and services have on the audience/users. We used a machine learning model like CNN to achieve these aims [6].

The essential motive of this research is to discover facial feelings. Emotion recognition will be a significant step in identifying many thoughts on the face. Through face emotion recognition, all of us can gauge the effects content materials and services bestow on users. The opportunity that devices can interpret human beings' emotions is exactly what motivated us to generate this particular research.

Convolutional Neural Network is one type of Deep Learning algorithm which is used for working with images and videos. Images are used as inputs, and the system collects their features before categorizing those using previously learned features. The various filters available in the model help extract features at each layer such as edges, and shapes (vertical, horizontal, round), and combine them all to be used as identification for the image.

The CNN model works in mostly two stages:

Feature Extraction – The phase where filters and layers work for extracting information from the images to be passed on to the next layer for classification.

Classification – Classifying the images based on the target variable is carried out.

In Section II, we describe the literature review followed by the methodology and model specifications in Section III that have been implemented in this paper. Moving forward we have discussed the outcome of our model in Section IV and have concluded that in Section V.

II. LITERATURE REVIEW

1) Bartlett et al. carried out the initial work that outlines a method to complete the task [7]. The goal of the article was to develop a system that could automatically identify frontal faces in video streams and categorize the emotions those faces were expressing based on their facial expressions [8]. The

article defined facial expressions as expressing happiness, sorrow, surprise, contempt, fear, rage, or neutral emotions.

2) With the help of the Radial Basis Function network (RBFN) for classification and Fisher's Linear Discriminant (FLD), Singular Value Decomposition (SVD), for feature selection, an algorithm for identifying a person's emotional state through facial expressions like anger, disgust, and happiness is demonstrated. It attempts to represent the face analytically in order for the feature vector can be input into a classifier. The system's overall success mostly hinges on the accurate recognition of the face or specific facial features like the eyes, brows, and lips [9].

3) Authors proposed to use of a revolutionary EEG-based emotion recognition method. This method uses multi-channel EEG data to examine the usage of three-Dimensional Convolutional Neural Networks (3D-CNN) for emotion identification. The proposed 3D-CNN approach's performance is improved by the data augmentation step. And using the multi-channel EEG signals as the data input for the proposed 3D-CNN model, a 3D data representation is created [10].

4) The VGG Net architecture was adopted by Yousif Khaireddin and Zhuofa Chen, who tuned its hyperparameters and experimented with other optimization techniques. To their experience, the model does not require additional training data to attain state-of-the-art single-network accuracy of 73.28% on FER2013 [11].

5) The initial step of visual processing, the encoding of facial features, and the decoding of facial expression connotation can all be considered as stages in the process of facial emotion identification, which was seen in schizophrenia patients [12].

6) In their study, M. Kalpana Chowdary, D. Jude Hemanth, and Tu N. Nguyen discuss the use of transfer learning techniques to recognise emotions. Pre-trained Resnet50, vgg19, Inception V3, and Mobile Net networks were employed in this study. After removing the completely connected layers from the pre-trained ConvNets, their own fully connected layers that are appropriate for the number of instructions in our task were added. The CK + database was used to conduct the experiment [13].

7) The technology of convolutional neural network model is used in an effective deep learning technique for face emotion detection to classify emotions from facial photographs and accurately identify age and gender from facial expressions [14].

8) A pre-trained DCNN model is adopted by replacing its dense higher layer(s) compatible with FER in one study's proposal for very Deep Convolution Neural Network (DCNN) modelling through Transfer Learning (TL) technique. The model is then tweaked using facial emotion data. A new approach is shown, which leads to a progressive increase in the model's accuracy by first tweaking each of the pre-trained DCNN blocks individually, after training the dense layer(s). Eight alternative pre-trained DCNN models (VGG-16, VGG-19, ResNet-18, ResNet-34, ResNet-50, ResNet-152, Inception-v3, and DenseNet-161) as well as the well-known KDEF and JAFFE facial image datasets are used to verify the proposed FER system [15].

9) A pre-trained DCNN model is adopted by replacing its dense higher layer(s) compatible with FER in one study's proposal for very Deep Convolution Neural Network (DCNN) modelling through Transfer Learning (TL) technique. The model is then tweaked using facial emotion data. A new approach is shown, which leads to a progressive increase in the model's accuracy by first tweaking each of the pre-trained DCNN blocks individually, after training the dense layer(s). Eight alternative pre-trained DCNN models (VGG-16, VGG-19, ResNet-18, ResNet-34, ResNet-50, ResNet-152, Inception-v3, and DenseNet-161) as well as the well-known KDEF and JAFFE facial image datasets are used to verify the proposed FER system [16].

10) The authors of the cited paper have suggested a two-tower CNN architecture to categorize images into seven fundamental types of emotion in light of the significance of edges in an image. In addition to the other tower, the planned CNN contains a tower called the edge-tower, which has a more straightforward architectural design and employs edge pictures as inputs [17].

III. METHODOLOGY AND MODEL SPECIFICATIONS

This activity looks at the ongoing challenge facing machine learning and the whole program is part of the training. The location where the system should educate using real-time individual face replies to data. For example, if the system has to discover a furious face the very first system must learn about an angry face and if the program should see a happy face, then the first program should get acquainted with a happy face. To precede a program with these types of emotions, a retraining method was used. Retraining data is then collected from surveys and people. The most difficult part of the program was the iterating component out of various other components of the system. Machine learning works as an advanced tool to analyze data from a wide range of expertise and slow motion [18]. This gives you the ability to get more accurate emotions [19] and provides real-time feedback. The machine did not hold out for a future result, not to be preserved. With the help of modern personal computers, neoteric data exploration techniques can examine thousands of data quickly and save several hours [20].

Other than that, using and installing such programs is very costly. Properly covered, it will withstand a high deal of adverse situations. This work offended the common and possible framework of emotional data mining to verify emotional patterns with the use of machine learning. This kind of paper proposes software based on typically the in-depth reading style and emotional conception of computer eyesight. This proposed approach uses the CNN algorithm in this document. This has triggered a much even more advanced approach as compared to the one which observed only seven feelings about CNN [21].

CNN Implementation: Convolutional Neural Network is one type of Deep Learning algorithm which is used for

working with images and videos. Images are used as inputs, and the system collects their features before categorizing those using previously learned features [22]. The various filters available in the model help extract features at each layer such as edges, and shapes (vertical, horizontal, round), and combine them all to be used as identification for the image [23].

The CNN model works in mostly two stages as seen in Fig. 1:

Feature Extraction – The phase where filters and layers work for extracting information from the images to be passed on to the next layer for classification [24].
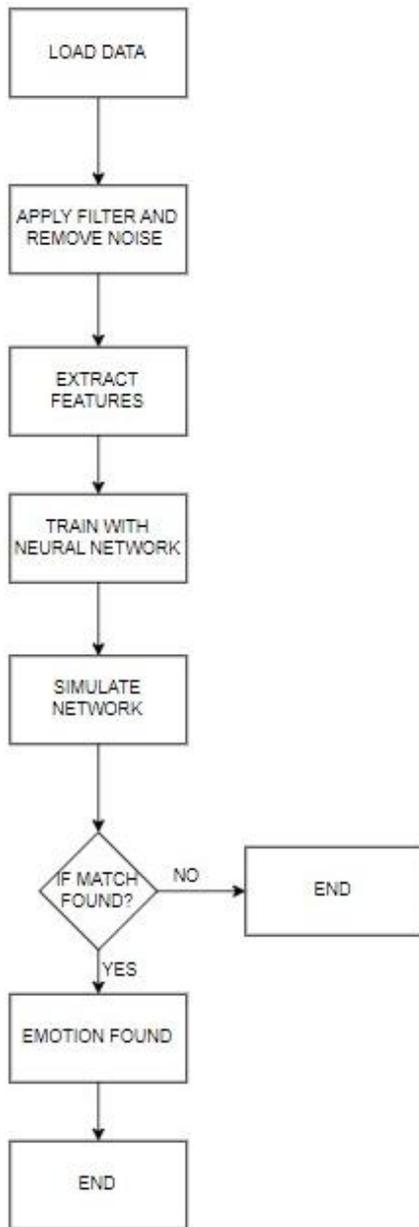
Classification – Classifying the images based on the target variable is carried out.

CNN Algorithm: The convolution neural network works by taking input images and assigning them weights and biases according to the importance of the objects identified. Here, we take the Facial Emotion Recognition 2013 dataset available publicly as our input data. Various pre-processing of input data is carried out as converting the images to grayscale, resizing them, reshaping the data, etc. The data is then divided into training and testing sets for model implementation. Various accuracy parameters are then taken into consideration for checking model performance and prediction.

The architecture of the CNN Model Fig. 2:

Fig. 1. Flowchart.

(a)

(b)

(c)

(d)
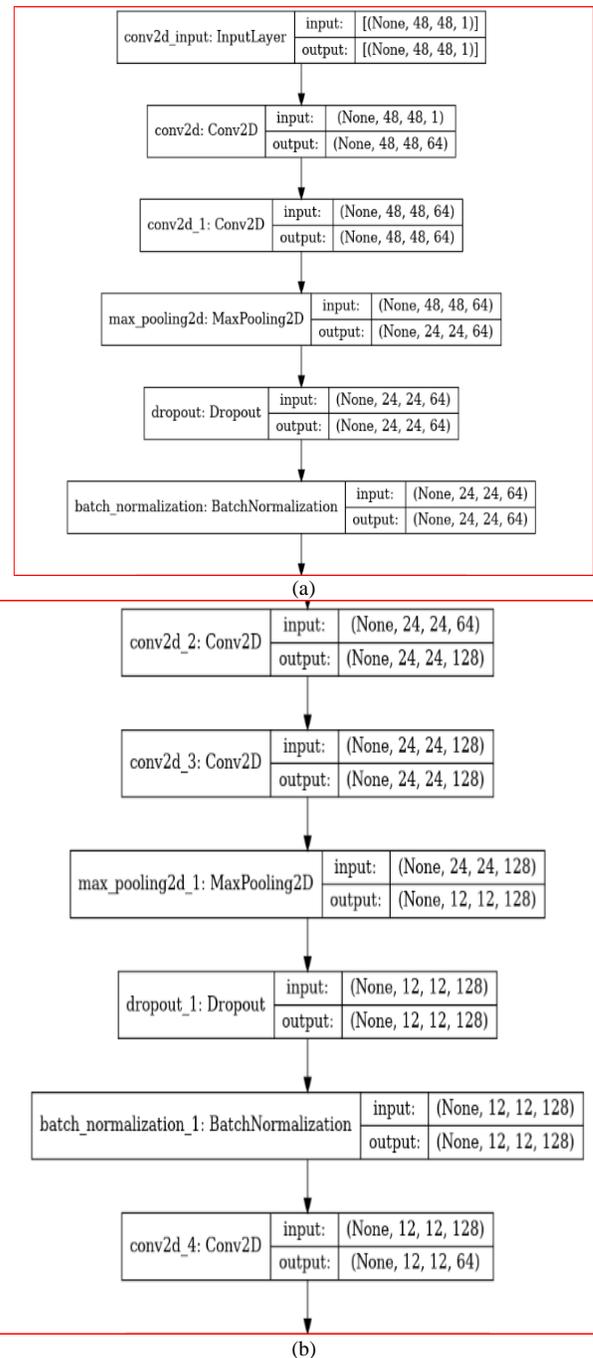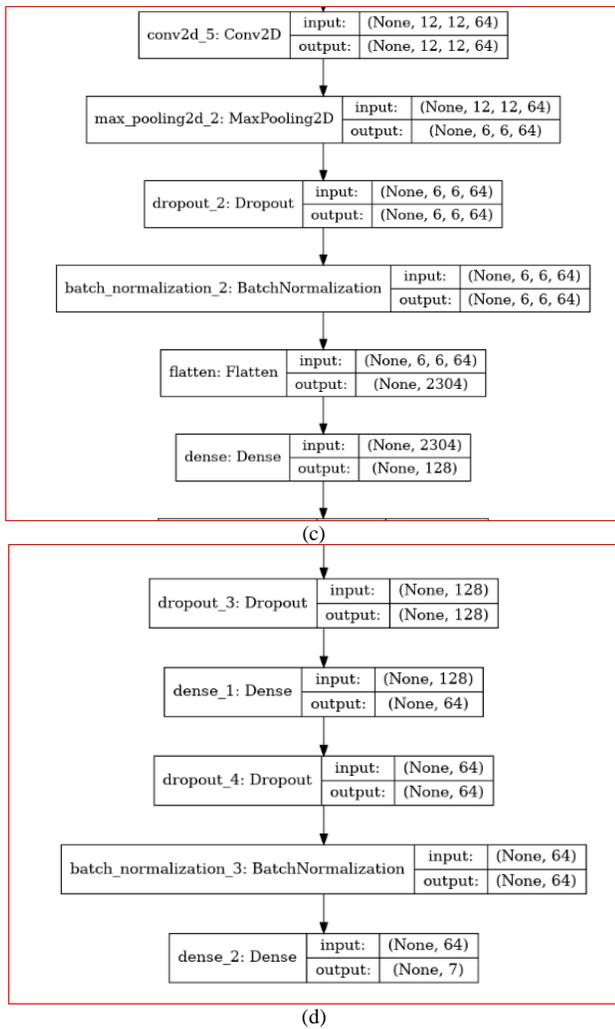
Fig. 2.   Model Architecture (a) - (d).

1) ResNet implementation: Residual Networks, or ResNet, learn residual functions with reference to the layer inputs, instead of learning unreferenced functions. The residual is the amount or number by which you have to change your prediction to meet the actual value. Rather than skipping every few stacked layers directly, to fit a desired underlying mapping, residual nets let these layers fit a residual mapping. Residual blocks are stacked on top of each other to form a network. The model ResNet-50 has fifty layers using these blocks.

To produce classification results, ResNet-50 first performs a convolution operation on the input, then it applies residual blocks, and finally, it applies a full connection operation. CNNs typically end with fully connected (FC) layers that summarize the characteristics of the preceding connected layers [25]. The Fully Connected layer can be viewed as feature weighting and uses earlier convolution and pooling as feature engineering, local amplification, and local feature extraction.

The structure of the Fully Connected layer is a way to quickly learn the non-linear combinations of advanced attributes generated by the convolutional layer. The layer learns a possible non-linear function. First, the image, which has been converted into a suitable format, is flattened into column vectors and fed back to the feed-forward neural network. Every training iteration uses flattened data. In this approach, the model is able to discriminate between some low-level aspects in the image and the image's main features, classifying those using methods like SoftMax [26]. The categorization outcomes for the seven expressions are output here.

2) ResNet Algorithm: The ResNets work on the principle of building deeper networks compared to other existing network architectures and simultaneously finding an optimized number of layers to negate the vanishing gradient problem. Here, we take the Facial Emotion Recognition 2013 dataset available publicly as our input data. Various pre-processing of input data is carried out as converting the images to grayscale, resizing them, reshaping the data, etc. The data is then divided into training and testing sets for model implementation. The model works on a 50-layer deep residual neural network that stacks residual blocks on top of each other. Various accuracy parameters are then taken into consideration for checking model performance and prediction performance.

IV. RESULTS

The very first challenge that we faced here was the smaller number of images in the dataset for emotions like disgust and fear. As a result of this, the model would be biased towards the emotions that have a larger number of images while training. Machine learning algorithms work well with datasets having a large number of segments and information. Image data generator helps deal with this by creating augmented images by making some modifications to the images in the form of flipping or rotating at various angles.

A. CNN Result

The model used by us here uses a convolution neural network (CNN) where we work by classifying images according to the emotion that it represents [27]. This is done with the help of feature extraction and various layers of the model such as max pool and dense [28]. The dropout layers help decrease the complexity of the models. This model is helpful in classifying the images with an accuracy level of approximately 62%. This however is not useful and very accurate so we try handling this problem with the help of transfer learning [29]. We can see the prediction outcome in Fig. 3.

In Fig. 4, we see how accurately our model performs by plotting the loss and accuracy metrics for the training and validation datasets. The loss that occurs, can be seen as decreasing as we increase the number of epochs the model iterates over. The accuracy for the model can be seen increasing as the model learns and the validation data accuracy can be seen as that of training i.e., 62%.
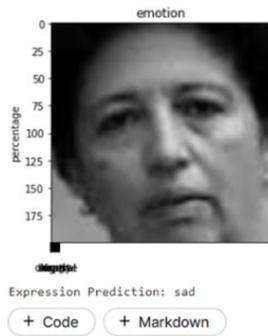
Fig. 3.   Expression Predicted: Sad.

Fig. 4.   Loss and Accuracy.

B. ResNet Result

Transfer learning involves the use of pre-trained models or information that is already stored and with the help of it we train new models. It helps in the reduction of time complexity as well as space complexity. We have chosen Resnet50 as the transfer learning model to solve our problem. Resnet50 makes use of its pre-existing information for feature extraction in determining emotions [30]. The model is then fine-tuned on different parameters and is able to reach an accuracy of 85% for both training and validation data. With an AUC score of approximately 70%, the model using Resnet50 seems to be more promising as compared to CNN. We can see the prediction outcome in Fig. 5.

In Fig. 6, we see how accurately our ResNet model performs by plotting the loss and accuracy metrics for the training and validation datasets. The loss that occurs, can be seen as decreasing as we increase the number of epochs the model iterates over. The losses for validation data can also be seen as decreasing. The accuracy for the model can be seen increasing as the model learns and the validation data accuracy can be seen as that to training after finetuning as well to reach around 85%.

Fig. 5.   Expression Predicted: Neutral.

Fig. 6.   History of Accuracy and History of Loss.
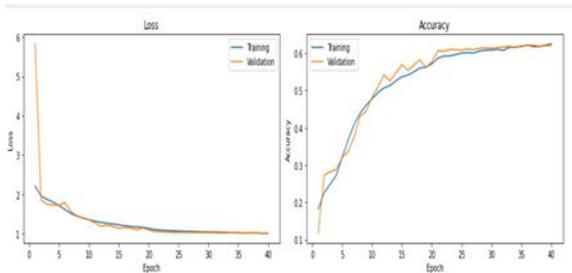
V.   CONCLUSION

A seasoned human can often identify another human's thoughts by inspecting and searching his/her face. Yet, machines are becoming more intelligent day by day as we are moving towards more advancements. Machines with capacities for acting similar to humans are introduced. Machines are being trained to behave on behalf of human opinion. Alternatively, if machines can identify the sentiment, it can prevent lots of incidents too. Data exploration can facilitate appropriate expression patterns by allowing machines to find an action more like humans with additional proficiency and error-free computation sentiment.

We have discussed a method for expression recognition in static pictures in this paper. In order to develop this emotion analysis system, CNN and Haarcascade were used for feature selection, and CNN and ResNet models for classification. This study uses average values generated from training samples to identify emotional expressions on human faces. We tested the system's ability to recognize the photographs and correctly interpret the expressions from the images.

REFERENCES

[1]   Hashim Abdulsalam, Wisal & al-hamdani, dr & Al Salam, Mohammed. (2018). Facial Emotion Recognition: A Survey. 7. 771-779. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 7, Issue 11, November 2018, ISSN: 2278 – 1323.

[2]   Liu Y, Li Y, Ma X, Song R (2017) Facial expression recognition with fusion features extracted from salient facial areas. Sensors 17(4):712.

[3]   Anggraeni Pitaloka, Diah & Wulandari, Ajeng & Basaruddin, T. & Liliana, Dewi Yanti. (2017). Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition. Procedia Computer Science. 116. 523-529. 10.1016/j.procs.2017.10.038.

[4]   R. Cowie et al., "Emotion recognition in human-computer interaction," in IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 32-80, Jan 2001, doi: 10.1109/79.911197.

[5]   Jonathan, Jonathan & Lim, Andreas & Paoline, & Kusuma Negara, I Gede Putra & Zahra, Amalia. (2018). Facial Emotion Recognition Using Computer Vision. Indonesian Association for Pattern Recognition. 46-50. 10.1109/INAPR.2018.8626999.

(IJACSA) International Journal of Advanced Computer Science and Applications,
Vol. 13, No. 11, 2022

[6] Mehendale, N. Facial emotion recognition using convolutional neural networks (FERC). SN Appl. Sci. 2, 446 (2020). Springer.https://doi.org/10.1007/s42452-020-2234-1.

[7] M. S. Bartlett, G. Littlewort, I. Fasel and J. R. Movellan, "real time face detection and facial expression recognition: development and applications to human computer interaction," Computer Vision and Pattern Recognition Workshop, vol. 5, July 2003.

[8] Viola, Paul & Jones, Michael. (2001). Robust Real-time Object Detection. International Journal of Computer Vision. 57. 137-154.

[9] Arumugam, Devi & Srinivasan, Purushothaman. (2011). Emotion Classification Using Facial Expression. International Journal of Advanced Computer Science and Applications - IJACSA. 2. 10.14569/IJACSA.2011.020714.

[10] Shawky, Elham & El-Khoribi, Reda & Shoman, Mahmoud & Wahby Shalaby, Mohamed. (2018). EEG-Based Emotion Recognition using 3D Convolutional Neural Networks. International Journal of Advanced Computer Science and Applications. 9. 329.

[11] Yousif Khaireddin, Zhuofa Chen,Cornell University, arXiv:2105.03588

[12] Gao Z, Zhao W, Liu S, Liu Z, Yang C and Xu Y (2021) Facial Emotion Recognition in Schizophrenia. Front. Psychiatry 12:633717. doi: 10.3389/fpsyt.2021.633717.

[13] Chowdary, M.K., Nguyen, T.N. & Hemanth, D.J. Deep learning-based facial emotion recognition for human–computer interaction applications. Neural Comput & Applic (2021). https://doi.org/10.1007/s00521-021-06012-8.

[14] Khattak, A., Asghar, M.Z., Ali, M. et al. An efficient deep learning technique for facial emotion recognition. Multimed Tools Appl 81, 1649–1683 (2022). https://doi.org/10.1007/s11042-021-11298-w.

[15] Akhand, M.A.H.; Roy, S.; Siddique, N.; Kamal, M.A.S.; Shimamura, T. Facial Emotion Recognition Using Transfer Learning in the Deep CNN. Electronics 2021, 10, 1036. https://doi.org/10.3390/electronics10091036.

[16] Bodavarapu PNR, Srinivas PVVS (2021) Facial expression recognition for low resolution images using convolutional neural network and denoising techniques. Indian Journal of Science and Technology 14(12): 971-983. https://doi.org/10.17485/IJST/v14i12.14.

[17] Arkaprabha Bhandari, Nikhil R. Pal, Can edges help convolution neural networks in emotion recognition?, Neurocomputing, Volume 433, 2021, Pages 162-168, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2020.12.092.

[18] Rafaqat Hussain Arain, Abdullah Maitlo, Sadaquat Ali Ruk, Hidayatullah Shaikh. Madan Lal, Kamlesh Kumar Study of Face Recognition Techniques: A Survey. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 6, 2018.

[19] Kak, Shakir & Mustafa, Firas & Valente, Pedro. (2018). A Review of Person Recognition Based on Face Model. 4. 157-168.

[20] 10.23918/eajse.v4i1sip157. Eurasian Journal of Science & Engineering ISSN 2414-5629 (Print), ISSN 2414-5602 (Online).

[20] X. Han and Q. Du, "Research on face recognition based on deep learning," 2018 Sixth International Conference on Digital Information, Networking, and Wireless Communications (DINWC), 2018, pp. 53-58, doi: 10.1109/DINWC.2018.8356995.

[21] Amal Azazi, Syaheerah Lebai Lutfi, Ibrahim Venkat, Fernando Fernández-Martínez, Towards a robust affect recognition: Automatic facial expression recognition in 3D faces, Expert Systems with Applications, Volume 42, Issue 6,2015, Pages 3056-3066, ISSN 0957-4174.

[22] Huang, H.-F & Tai, S.-C. (2012). Facial expression recognition using new feature extraction algorithm. Electronic Letters on Computer Vision and Image Analysis. 11. 41-54.

[23] Morshed, Golam & Ujir, Hamimah & Hipiny, Irwandi. (2021). Customer's spontaneous facial expression recognition. Indonesian Journal of Electrical Engineering and Computer Science. 22. 1436. 10.11591/ijeecs.v22.i3.pp1436-1445.

[24] Prof. S. Graceline Jasmine, Shivam Singh. Face Recognition System. International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV8IS050150 Vol. 8 Issue 05, May-2019.

[25] KH Teoh et al 2021 J. Phys.: Conf. Ser. 1755 012006. Face Recognition and Identification using Deep Learning Approach. 5th International Conference on Electronic Design (ICED) 2020 Journal of Physics: Conference Series 1755 (2021) 012006 IOP Publishing doi:10.1088/1742-6596/1755/1/012006.

[26] Nazeer, Shahrin & Omar, Normah & Khalid, Marzuki. (2007). Face Recognition System using Artificial Neural Networks Approach. Proceedings of ICSCN 2007: International Conference on Signal Processing Communications and Networking. 420 - 425. 10.1109/ICSCN.2007.350774.

[27] Sajid M, Ali N, Dar SH, Iqbal Ratyal N, Butt AR, Zafar B, Shafique T, Baig MJA, Riaz I, Baig S (2018) Data augmentation-assisted makeup-invariant face recognition. Math Probl Eng 2018:1–10.

[28] TANOY DEBNATH et al. Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity. Scientific Reports, [s. l.], v. 12, n. 1, p. 1–18, 2022. DOI 10.1038/s41598-022-111730.

[29] Zhang H, Jolfaei A, Alazab M. A face emotion recognition method using convolutional neural network and image edge computing. IEEE Access. 2019; 7: 159081-159089. 10.1109/ACCESS.2019.2949741.

[30] Deshmukh, S.; Patwardhan, M.; Mahajan, A. Survey on real-time facial expression recognition techniques. IET Biom. 2016,5, 155–163. Sensors (Basel). 2019 Apr; 19(8): 1863.Published online 2019 Apr 18. doi: 10.3390/s19081863. Multidisciplinary Digital Publishing Institute (MDPI).

173 | P a g e
www.ijacsa.thesai.org

Research on Sentiment Analysis Algorithm for Comments on Online Ideological and Political Courses

Xiang Zhang[1]

Guizhou University of Traditional Chinese Medicine
Guiyang, Guizhou 550025, China

Xiaobo Qin[2]

Guiyang University
Guiyang, Guizhou 550005, China

Abstract—The online course teaching platform provides a more accessible and open teaching environment for teachers and students. The sentiment tendency reflected in the online course comments becomes an essential basis for teachers to adjust the course and students to choose the course. This paper combined two deep learning algorithms, i.e., a convolutional neural network (CNN) algorithm and a long short-term memory (LSTM) algorithm, to identify and analyze the emotional tendency of comments on online ideological and political courses. Moreover, the CNN+LSTM-based sentiment analysis algorithm was simulated in MATLAB software. The influence of the text vectorization method on the recognition performance of the CNN+LSTM algorithm was tested; then, it was compared with support vector machine (SVM) and LSTM algorithms, and the comments on online ideological and political courses were analyzed. The results showed that the recognition performance of the CNN+LSTM-based sentiment analysis algorithm adopting the Word2vec text vectorization method was better than that adopting the one-hot text vectorization method; the recognition performance of the CNN+LSTM algorithm was the best, the LSTM algorithm was the second, and the SVM algorithm was the worst in terms of the performance of recognizing the sentiment of comment texts; 86.36% of the selected comments on ideological and political courses contained positive sentiment, and 13.64% contained negative sentiment. Relevant suggestions were given based on the negative comments.

Keywords—Online courses; comment; sentiment tendency; long short-term memory

I. INTRODUCTION

Education is an important cause of social development, and good education can provide more excellent talents for society and further promote the benign development of society [1]. Students are still in the stage of shaping their value system. The information they can receive at that stage is mixed, and the existence of the Internet has exacerbated this process [2]. Once the information has a negative impact on student's value system, it will seriously affect their future growth. Therefore, in addition to learning the necessary knowledge, students need to establish the correct value system through ideological and political education [3]. Traditional classroom ideological and political education aims to popularize education and is student-oriented. Although there are excellent teachers who can take into account the personalized training of different students in classroom teaching, the number of excellent teachers is limited; moreover, as the fixed time and place of traditional classroom

education is fixed, students only do homework after school hours [4]. With the development of Internet technology, online teaching platforms have emerged, which are not bound by time and place compared with traditional classroom teaching. In general, the teaching resources provided by the platform are pre-uploaded teaching texts and recorded videos [5], and the platform users can browse and learn anytime and anywhere. Online teaching also has shortcomings. Firstly, the lack of direct interaction between teachers and students makes it difficult for teachers to get feedback. Secondly, students are the main subjects when choosing teaching resources, and the diversity of teaching resources in the platform makes it difficult for students to make effective choices [6]. The comments on teaching resources in the platform can alleviate the shortcomings mentioned above to a certain extent. Teachers can correct the problems of their courses according to the emotional preferences of the comments, and students can make preliminary judgments on teaching resources [7]. However, since there are hundreds of comments for a course and even more comments when comparing different courses, it is difficult to collect and organize the comments by manual means alone. In order to quickly collect and process the sentiment tendencies of online course comments to assist instructors and students, intelligent algorithms are needed to process the comment texts. Intelligent algorithms can analyze the sentiment in the comment area under online ideological and political courses efficiently. Students can select courses based on the tendency of sentiment distribution in the comment area, while teachers can improve courses based on it. This paper studied the sentiment analysis algorithm for analyzing the sentiment tendency of comments on online ideological and political courses, which combines two deep learning algorithms, the convolutional neural network (CNN) algorithm and long short-term memory (LSTM). The CNN algorithm obtained the local and global features of the comment text, and then the sentiment tendency was calculated by the LSTM based on the comment text features extracted by the CNN. Finally, the CNN+LSTM algorithm was compared with support vector machine (SVM) and LSTM algorithms.

II. LITERATURE REVIEW

Related studies are as follows. Yasmina et al. [8] used an unsupervised machine learning algorithm to classify sentiment based on a corpus of data constructed using YouTube comments. They used a point-by-point mutual information

metric to calculate the similarity of the sentiment to every target sentiment and found that the algorithm had a global accuracy of 92.75%. Atmaja et al. [9] recognized sentiment in natural language by fusing acoustic and text networks using a support vector machine (SVM). The experimental results showed that this two-stage late-fusion approach achieved higher performance than any one-stage processing. Bernhard et al. [10] proposed sent2affect, a tailored form of transfer learning for affective computing. The results showed that recurrent neural network (RNN) and transfer learning consistently outperformed traditional machine learning. Zhi et al. [11] proposed an application framework and design of an intelligent system for emotion recognition and topic mining, aiming at intelligent and personalized learning analysis for Massive Open Online Courses (MOOC). The system could predict the popularity of every course and obtain emotional topic feedback about the course content so that teachers could analyze and improve their teaching strategies. It could also obtain the emotional topic feedback platform support of the relevant course topic so that administrators could improve user experience in the platform. Jin et al. [12] proposed a hybrid deep neural network model for Chinese text sentiment recognition to identify the sentiment tendency of Chinese medical reviews. The results showed that the model was able to obtain better text features than the reference model and had a text classification accuracy of 99%. Pan et al. [13] proposed an emotion recognition algorithm based on synchronized time comments as the basis for video clip recommendation and found that the model could effectively analyze the complex emotions of different types of text information.

III. METHODS

A. Basic Process of the CNN+LSTM Algorithm

This paper analyzed the sentiment in the comments of online courses with the CNN and LSTM algorithms [14]. The CNN algorithm was used to extract features from comment texts. The LSTM algorithm analyzed the sentiment in the extracted features based on its feature of capable of contacting the context. The flow of the combined algorithm is shown in Fig. 1.
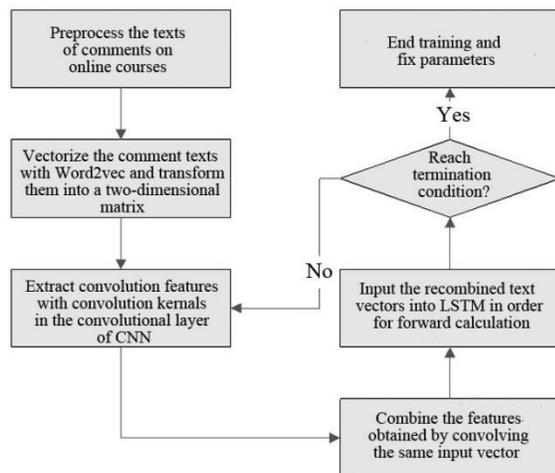
Fig. 1. Training Process of the CNN+LSTM-Based Sentiment Analysis Algorithm for Comments on Online Ideological and Political Courses.

1) The comment text about online ideological and political courses is pre-processed by denoising, filtering, and word segmentation [15]. Denoising refers to removing meaningless characters such as numbers and punctuation from the comment text; filtering means removing comments that are irrelevant to the course topic with a clustering algorithm; word segmentation means dividing sentences into phrases or words [16] for later text vectorization.

2) The pre-processed comment text is vectorized to transform natural language into a text vector that computers can understand. The skip-gram model [17] in Word2vec is used for text vectorization. The dimensionality of the text vector is determined according to the computational requirements, and usually, the larger the dimensionality, the more accurate the distribution of semantic features of the words that the vector can reflect, but the more difficult the computation. Then, the vectorized text is transformed into a two-dimensional matrix. Taking a five-dimensional text vector as an example, it is assumed that there are only four words in the short comment. The two-dimensional matrix of the text vector of the comment is shown in Fig. 2, where every row represents a five-dimensional text vector of one word.

3) After transforming the text vector of the comment into a two-dimensional matrix, the text vector of the comment is regarded as an image, and every dimensional vector of words is regarded as a pixel point in the image. The two-dimensional matrix is input into the CNN, and plural convolution kernels slide on the two-dimensional matrix in a certain step length in the convolutional layer, and convolution operation is performed using the convolution kernel in every slide [18]. The convolution formula is:

$$Y_i = f(X_i \otimes W_i + b_i),$$
(1)

where Y_i is the convolutional output eigenvalue of the i-th convolutional kernel, X_i is the input vector of the i-th convolutional kernel, W_i is the weight of the i-th convolutional kernel, and b_i is the bias of the i-th convolutional kernel.

4) Several convolutional characteristic faces of the two-dimensional matrix are obtained by the convolutional operation of the CNN convolutional layer. The characteristic part obtained by convolving the same input vector is selected from the convolutional characteristic faces extracted by every convolutional kernel and joined in rows to obtain a new vector [19].

$$
\begin{bmatrix}
x_{11} & x_{12} & x_{13} & x_{14} & x_{15} \\
x_{21} & x_{22} & x_{23} & x_{24} & x_{25} \\
x_{31} & x_{32} & x_{33} & x_{34} & x_{35} \\
x_{41} & x_{42} & x_{43} & x_{44} & x_{45}
\end{bmatrix}
$$
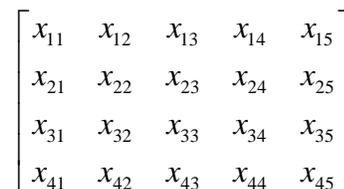
Fig. 2. Schematic Diagram of the Two-Dimensional Matrix of Text Vectors.

5) The new combined vector obtained is input to the LSTM for forward computation by the input gate, forgetting gate, and output gate [20]. The equations are as follows:

$$\begin{cases} i_t = g(\omega_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(\omega_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \\ f_t = g(\omega_f \cdot [h_{t-1}, x_t] + b_f) \\ o_t = g(\omega_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \cdot \tanh(C_t) \end{cases}, \qquad (2)$$

where the current cell input is x_t, the previous hidden layer state is h_{t-1}, the last cell state is C_{t-1}, i_t is the proportion that determines the newly added information that can be remembered, $\tilde{C}_t$ is the cell state of the newly added information, C_t is the current cell state after the new information is added, ω_i and ω_t are the corresponding weights [21], b_i and b_C are the corresponding biases, f_t is the weight of the information not to be forgotten in C_{t-1}, ω_f is the corresponding weight, b_f is the corresponding bias, o_t is the proportion that determines the final output information volume, and h_t is the final output or the next hidden state[22]. The output result of the output gate of the LSTM is calculated using the softmax function [23] in the fully connected layer to obtain the sentiment classification result.

The actual results obtained from the forward calculation are compared with the expected results of the training samples. Cross-entropy is used to calculate the error between them. Then, whether the training of the algorithm reaches the termination conditions, including the number of iterations reaching the set number and the error converging to the set range, is determined. If the algorithm reaches the termination condition, the training will be stopped, and the algorithm's parameters will be fixed; if the algorithm does not reach the termination condition, the parameters in the algorithm will be adjusted reversely, and then it returns to step *3)*.

B. Simulation Experiment on the CNN+LSTM Algorithm

1) Experimental Data and Setup

The comment text data used in the experiment came from the China University MOOC website. Twelve thousand comments were crawled from the comment area of ideological and political courses using crawler software. After identification, the emotional tendency of these comments was divided into positive and negative, and their numbers were close. Then, 60% of the comments were used as training samples, and the remaining 40% was used as testing samples.

The parameters of the CNN+LSTM-based sentiment analysis algorithm [24] are as follows. The dimension number of the Word2vec vector for vectorizing the comment text was set as 100. The parameters of the CNN part are as follows. There was a convolutional layer with 64 convolutional kernels.

The specification of every convolutional kernel was 2×100, and the moving step length of the convolutional kernel was 1. The specification of the input data in the input layer, i.e., the size of the two-dimensional matrix, was $M \times 100$, where M is the maximum number of the segmented words. The parameters of the LSTM part are as follows. There were 64 hidden neurons. The weights were initialized using glotot_normal. The bias was set as 0. Training stopped after a maximum of 1,000 iterations.

In order to further verify the effectiveness of the CNN+LSTM algorithm, it was compared with a machine learning algorithm, SVM, and a deep learning algorithm, LSTM, by experiments. As to the relevant parameters of the SVM algorithm, the sigmoid function was used as the kernel function, and the penalty parameter was set as 1. The relevant parameters of the LSTM algorithm were the same as the LSTM part of the CNN+LSTM algorithm.

2) Experimental Project

a) The Impact of Comment Text Vectorization Methods on Sentiment Analysis Algorithms

The Word2vec method was used to vectorize the comment text, but the one-hot method was also applicable in addition to this method. The CNN+LSTM algorithm used the one-hot method and the Word2vec method, respectively, and the recognition performance of sentiment analysis algorithms under these two text vectorization methods was compared.

b) Performance Differences between Different Sentiment Analysis Algorithms

All three sentiment analysis algorithms used the Word2vec method for text vectorization. Then, they were trained using the same training set and tested by the same training set.

The evaluation indexes used to compare the performance were precision [25], recall rate, and F-value, which were calculated by the following formulas:

$$\begin{cases} P = \dfrac{TP}{TP + FP} \\ R = \dfrac{TP}{TP + FN} \\ F = \dfrac{2PR}{P + R} \end{cases}, \qquad (3)$$

where P is the precision, R is the recall rate, F is a combination of the precision and recall rate, TP is the number of positive samples that are predicted as positive, FP is the number of negative samples that are predicted as positive, and FN is the number of positive samples that are predicted as negative.

(3) Analysis of comments on online ideological and political courses with the validated sentiment analysis algorithm

The CNN+LSTM-based sentiment analysis algorithm performed sentiment analysis on online ideological and political courses on the Chinese University MOOC website.

The top ten ideological and political courses with the highest number of plays were selected from the corresponding classification in the MOOC platform. Comments on these ten courses were crawled using crawler software. The sentiment analysis algorithm identified the sentiment classification of the collected comments, and relevant course improvement suggestions were given based on the negative comments.

IV. RESULTS

Fig. 3 shows the recognition performance of the CNN+LSTM sentiment analysis algorithm under different text vectorization methods. It was seen from Fig. 3 that the sentiment analysis algorithm using the one-hot method for text vectorization had a precision of 84.48%, a recall rate of 84.57%, and an F-value of 84.52%; the sentiment analysis algorithm using Word2vec for text vectorization had a precision of 94.26%, a recall rate of 95.24%, and an F-value of 94.75%. The comparison of the performance between the algorithms under the two text vectorization methods revealed that the sentiment analysis algorithm using Word2vec had a more excellent sentiment recognition performance.
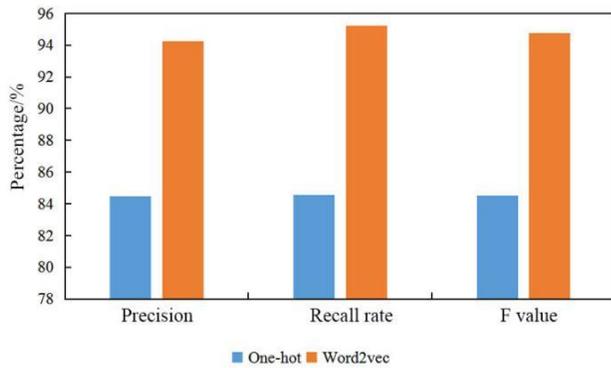
Fig. 3. Recognition Performance of the CNN+LSTM Sentiment Analysis Algorithm under Different Text Vectorization Methods.

Fig. 4 shows the comparison of the recognition performance of three sentiment analysis algorithms, SVM, LSTM, and CNN+LSTM algorithms. It was noticed in Fig. 4 that the precision, recall rate, and F-value of the SVM algorithm were 77.36%, 76.58%, and 76.97%, respectively; the precision, recall rate, and F-value of LSTM were 90.12%, 89.63%, and 89.87%; the precision, recall rate, and F-value of the CNN+LSTM algorithm was 94.26%, 95.24%, and 94.75%, respectively. It was intuitively seen that the SVM algorithm had the worst recognition performance, the LSTM algorithm had the medium performance, and the CNN+LSTM algorithm performed the best.

Fig. 5 shows the results of sentiment identification using the CNN+LSTM-based algorithm for the ten comments selected from the comment section of ideological and political courses. It was seen from Fig. 5 that 86.36% of the comments in the comment section contained positive sentiment and 13.64 % contained negative sentiment. For students, positive comments can help them understand the advantages of the course, while negative comments can make students view the online ideological and political course as objectively as possible; for teachers, positive comments can encourage teachers, but more

importantly, negative comments can help them search for defects in online ideological and political courses and make better improvements. Due to space limitation, this paper only presented some comments containing negative sentiment and the corresponding key negative words, as shown in Table I. It was seen that users' negative comments on online ideological and political courses were "boring content," "lack discipline in online courses," "lack of interactivity in online ideological and political courses," "one-size-fits-all," etc.
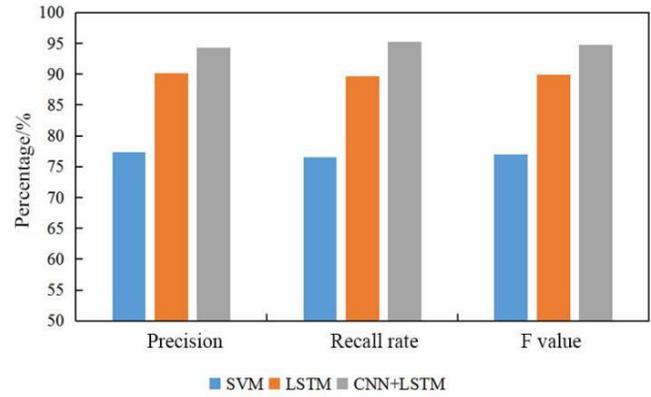
Fig. 4. Recognition Performance of Three Sentiment Analysis Algorithms.
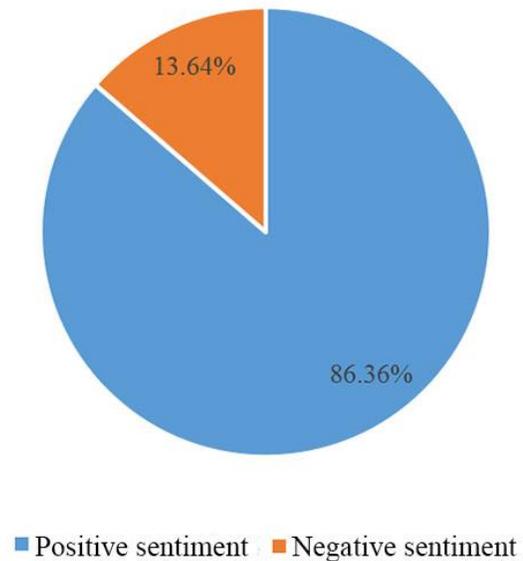
Fig. 5. Sentiment Recognition Results of Comments on Online Ideological and Political Courses by the CNN+LSTM-Based Sentiment Analysis Algorithm.

TABLE I. SOME NEGATIVE COMMENTS AND THEIR KEY NEGATIVE WORDS

Part of the original texts of some negative comments	Key negative words
The course content is straightforward and boring	Boring content
Online courses lack discipline, and students have low autonomy	Lack discipline
Lack of interactivity between teacher and students in online courses	Lack of interactivity

V. Discussion

Online teaching courses are gradually increasing with the popularity of the Internet. Compared with traditional offline teaching courses, online teaching courses have advantages of not constrained by time and space and a higher degree of autonomy for students. However, online teaching courses also have disadvantages such as difficulties in teacher-student interaction and students' difficulties in choosing due to the variety of teaching resources in the platform. The platforms where online teaching course resources are located often have comment areas. These comments on online courses can, to a certain extent, alleviate the problems of communication difficulties and diversification of choices for teachers and students, but the number of comments in the comment areas is so large that it is difficult for them to quickly obtain the emotional tendency from these comments. In order to solve the above problems, this paper combined two deep learning algorithms, CNN and LSTM, and then used them for sentiment analysis of online course reviews. In the simulation experiment, the effects of different text vectorization methods on the CNN+LSTM algorithm were tested, and it was compared with SVM and LSTM algorithms. Finally, the CNN+LSTM algorithm was applied to the actual MOOC platform to recognize the sentiment of comments in ten online ideological and political course comment areas in the platform. The final experimental results have shown in the last section.

The CNN+LSTM algorithm that adopted Word2vec performed better in sentiment recognition than the CNN+LSTM algorithm that adopted the one-hot method. The reason was analyzed. The one-hot method sets the length of the text vector according to the total number of words in the coding dictionary. The position of the word to be transformed in the dictionary determines that its corresponding position in the vector is one and the other positions are zero. This method is simple and intuitive, but in practice, the number of words in the dictionary is huge, resulting in a high dimensionality of the text vector, and only one position in the text vector is 1, making the text vector very sparse, so the algorithm wastes much computational power when training. In addition, the one-hot method encodes the words independently and does not contact the context. The Word2vec method vectorizes words by mapping them into a word space. Semantically similar words are clustered, i.e., the distribution expression of words is realized based on the context. In this paper, the skip-gram model of the Word2vec method predicted the probability distribution of the context based on the intermediate word and used it as its vector representation. Compared with the one-hot method, Word2vec made full use of the context of the words to be transformed and used a smaller word space dimension, so the sentiment analysis algorithm using Word2vec was better in terms of recognition performance.

The results of the comparative experiment on SVM, LSTM, and CNN+LSTM algorithms showed that the CNN+LSTM algorithm had better sentiment recognition performance. The reason is as follows. The SVM algorithm used the kernel function to map the text vector to a high-dimensional space, then searched for the hyperplane that could divide the space in the high-dimensional space, and used the hyperplane to classify the comment text. The SVM algorithm was still difficult to fit completely though it used the high-dimensional mapping of kernel functions to linearize the nonlinear laws in the text as much as possible. The LSTM algorithm is a derivative of the RNN algorithm. The activation function in the hidden layer effectively fit the nonlinear law, and the LSTM algorithm contacted the context based on the historical information, so its recognition performance was much stronger than the SVM algorithm. The CNN+LSTM algorithm is a combination of CNN and LSTM algorithms. The convolutional kernel of the CNN algorithm extracted local features from text vectors, and then the features were reorganized in text order. The LSTM algorithm identified the sentiment from the reorganized text features. The local features provided by the CNN algorithm made up for the defect that the LSTM algorithm could not utilize local features; thus, the CNN+LSTM algorithm was better than the LSTM algorithm in recognition.

The results of the sentiment recognition of the ten comment areas in the actual MOOC platform using the CNN+LSTM algorithm showed that most of the comments were positive. The following suggestions are given according to the negative comments in the comment areas.

1) Teachers should not use too many text descriptions when making online ideological and political courses but can combine teaching contents with pictures and videos.

2) For the lack of discipline when students take online courses, teachers can assist by setting hard targets and focusing on cultivating autonomy in the daily teaching process.

3) Teachers can enhance the interactivity of online ideological and political courses through the live streaming function in the MOOC platform.

VI. Conclusion

This paper combined two deep learning algorithms, CNN and LSTM algorithms, to identify and analyze the sentiment tendency of comments on online ideological and political courses and then simulated the CNN+LSTM-based algorithm in MATLAB software to test the recognition performance difference under two different text vectorization methods, one-hot and Word2vec methods. The following results are obtained. Moreover, the CNN+LSTM-based algorithm was compared with SVM and LSTM algorithms, and the comments on online ideological and political courses were also analyzed. The following results are obtained. The CNN+LSTM sentiment analysis algorithm using Word2vec for text vectorization had better sentiment recognition performance than the algorithm using the one-hot method. The SVM algorithm had the worst recognition performance, the LSTM algorithm had the medium performance, and the CNN+LSTM algorithm had the best performance. 86.36% of the selected comments on the ideological and political courses contained positive sentiment, and 13.64% contained negative sentiment. Relevant suggestions were given based on the negative comments.

References

[1] J. H. Kim, and S. P. Lee, "Multi-modal Emotion Recognition using Speech Features and Text Embedding," T. Korean Instit. Electr. Eng., vol. 70, pp. 108-113, August 2021.

[2] J. Li, and L. Xiao, "Emotions in online debates: Tales from 4Forums and ConvinceMe," Proc. of the Association for Information Science and Technology, vol. 57, October 2020.

[3] M. H. Su, C. H. Wu, K. Y. Huang, and Q. B. Hong, "LSTM-based Text Emotion Recognition Using Semantic and Emotional Word Vectors," 2018 First Asian Conf. on Affective Computing and Intelligent Interaction (ACII Asia), Beijing, China, pp. 1-6, May 2018.

[4] B. Wang, "Automatic Network Emotion Recognition Method and Its Application in Online Text Emotion Classification," Rev. Fac. Ing., vol. 32, pp. 18-22, January 2017.

[5] Q. Cai, G. C. Cui, and H. X. Wang, "EEG-based Emotion Recognition Using Multiple Kernel Learning," Res. Mach. Intell. English, vol. 19, pp. 472-484, September 2022.

[6] A. Wang, X. Liu, X. Sun, and J. Wang, "Research of internet public opinion based on hybrid algorithm of LDA and VSM," C e Ca, vol. 42, pp. 1508-1513, April 2017.

[7] Z. Hao, "Emotion recognition simulation of Japanese text based on FPGA and neural network," Microprocess. Microsy., pp. 103384, November 2020.

[8] D. Yasmina, M. Hajar, and A. M. Hassan, "Using YouTube Comments for Text-based Emotion Recognition," Proc. Comput. Sci., vol. 83, pp. 292-299, December 2016.

[9] B. T. Atmaja, and M. Akagi, "Two-stage dimensional emotion recognition by fusing predictions of acoustic and text networks using SVM," Speech Commun., vol. 126, pp. 9-21, October 2021.

[10] B. Kratzwald, S. Ilic, M. Kraus, S. Feuerriegel, and H. Prendinger, "Deep learning for affective computing: text-based emotion recognition in decision support," Decision Support Systems, vol. 115, pp. 24-35, November 2018.

[11] Z. Liu, W. Zhang, J. Sun, H. N. H. Cheng, X. Peng, and S. Liu, "Emotion and Associated Topic Detection for Course Comments in a MOOC Platform," Int. Conf. on Educational Innovation Through Technology, pp. 15-19, October 2017.

[12] Q. Jin, X. Xue, W. Peng, W. Cai, Y. Zhang, and L. Zhang, "TBLC-rAttention: A Deep Neural Network Model for Recognizing the Emotional Tendency of Chinese Medical Comment," IEEE Access, vol. 8, pp. 96811-96828, May 2020.

[13] Z. Pan, X. Li, L. Cui, and Z. Zhang, "Video clip recommendation model by sentiment analysis of time-sync comments," Multimed. Tools Appl., vol. 79, pp. 33449-33466, December 2020.

[14] B. Sun, F. Tian, and M. Jia, "Emotion recognition method of Tibetan micro-blog text based on sentiment dictionary," J. Phys. Conf. Ser., vol. 1314, pp. 012182, October 2019.

[15] S. R. Adarsh, "Enhancement of text based emotion recognition performances using word clusters," Int. J. Res. - GRANTHAALAYAH, vol. 7, pp. 238-250, January 2019.

[16] P. Thakur, and R. Shrivastava, "A Review on Text Based Emotion Recognition System," Int. J. Adv. Trend. Comput. Sci. Eng., vol. 7, pp. 67-71, October 2018.

[17] C. Jeyalakshmi, B. Murugeswari, and M. Karthick, "Recognition of emotions in Berlin speech: A HTK based approach for speaker and text independent emotion recognition," Pak. J. Biotechnol., vol. 14, pp. 63-69, January 2017.

[18] P. Singh, R. Srivastava, K. Rana, and V. Kumar, "A multimodal hierarchical approach to speech emotion recognition from audio and text," Knowl.-Based Syst., vol. 229, pp. 1-17, October 2021.

[19] Z. Peng, "The Approaches of Internet Public Opinion Research," Libraly J., vol. 35, pp. 63-68, December 2016.

[20] E. Batbaatar, M. Li, and K. H. Ryu, "Semantic-Emotion Neural Network for Emotion Recognition from Text," IEEE Access, vol. 7, pp. 111866-111878, August 2019.

[21] F. Zhang, S. C. Li, and Y. Guan, "Hot issues about drug price in china: from the view of internet public opinion monitoring," Value Health, vol. 19, pp. A273-A274, May 2016.

[22] H. Zhu, P. Liu, and X. Shan, "Analysis of internet-based public opinion in China, 2012," J. Mol. Neurosci., vol. 49, pp. 614-617, January 2015.

[23] N. Alswaidan, and M. Menai, "Hybrid Feature Model for Emotion Recognition in Arabic Text," IEEE Access, vol. 8, pp. 37843-37854, February 2020.

[24] S. Kiruthika, and V. V. Priyan, "Detection of Online Fake Reviews Based on Text Emotion," J. Phys. Conf. Ser., vol. 1916, pp. 1-8, May 2021.

[25] L. Li, "Research on the Transfer Rules of Internet Users' Negative Emotional State in Financial Public Opinion," Open J. Bus. Manag., vol. 08, pp. 282-301, January 2020.

Why do Women Volunteer More than Men? Gender and its Role in Voluntary Citizen Reporting Applications Usage and Adoption

Muna M. Alhammad

Management Information System Dept

King Saud University

Riyadh, Saudi Arabia

Abstract—By researching why citizens are eager to participate in citizen reporting applications, this study contributes to the understanding of citizen-government interaction in open government. Self-determination theory, gender role theory, and social role theory were employed to evaluate the impact of various motivational factors on individual behavioural intentions to participate in citizen reporting applications, as well as the role of gender in moderating their effects. The model was quantitatively tested by collecting 499 responses through a questionnaire from citizens who had previously utilized citizen reporting applications. The model was validated using partial least squares. The findings reveal that social responsibility, output quality, self-concern, and revenge are the motivational antecedents that have the most influence on individuals' motivation to participate in citizen reporting applications managing to explain 65.9% of behavioural intention variances. Social responsibility is the most significant driver when compared to the others. The study also revealed that gender differences moderate the impact of social responsibility and revenge on user involvement in citizen reporting apps. The current study adds to the existing literature on citizen reporting adoption and usage by examining the motivational factors that affect citizens' engagement across multiple contexts and evaluating the effect of gender in moderating the influence of social responsibility and revenge. Government institutions need to consider gender differences when designing their citizen reporting applications and their associated marketing campaigns.

Keywords—*Self-determination theory; gender role theory; social role theory; motivation; amotivation; gender diversity; social responsibility; citizens reporting application*

I. INTRODUCTION

Transparent public information, participatory decision-making procedures, and innovative platform-based forms of collaborative activities are examples of government efforts to strengthen citizen-government connections [1]. Crowdsourcing applications are one form of open government initiative that aims to accomplish a task by soliciting of external community contributions via an online platform [2]. Crowdsourcing involves two types of actors: individuals or organizations seeking assistance in carrying out tasks and solving problems, and crowd members who offer their services as a contribution [3]. Citizen reporting applications are one type of crowdsourcing application that aims to incorporate citizen participation in reporting incidents to develop public services. Originally, it was mainly used to report infrastructure issues in specific regions through the utilization of geolocation technology [4]. Citizen reporting applications are now being employed for several purposes, such as reporting environmental disasters, controlling pandemic outbreaks [5], and reporting commercial or security incidents.

Many publications on citizen reporting focus on either reviewing the structural concept of citizen reporting applications, such as the study of Linders [6], or concentrate on the ultimate design of citizen reporting applications, such as the study of Lönn, et al. [7]. Few studies were found investigating motivational factors behind citizens' willingness to voluntarily engage in citizen reporting activities in a specific context. Wijnhoven, et al. [8] investigated various motivational factors behind users' willingness to engage in open government, which includes citizen reporting applications as one variant of the open government initiative. Although the study is useful in giving general insight about factors motivating citizens to participate, it only studies citizens' hypothetical interest in participating as the majority of the survey respondents did not actually use this kind of application before. Therefore, further studies are needed to confirm their results using a representative sample. Abu-Tayeh, et al. [4] studied the impacts of two drivers, i.e., self-concern and other-orientation, on citizens' actual participation in citizen reporting applications in the context of smart cities. Although the study found these factors to be significant, the validity of their findings is limited to the context of reporting issues and damages related to the infrastructure in Zurich, which cannot be generalized to other contexts. Alhammad, et al. [9] studies the motivational factors behind citizens' engagement in citizen reporting applications to report commercial incidents. They found that self-concern, revenge, and output quality play significant roles in users' engagement. However, their study was limited to the context of only reporting commercial incidents. Motivational factors behind citizens' willingness to engage in citizens' reporting applications related to different contexts, such as reporting commercial, health, or security issues and violations, remain unstudied. In addition, there is a scarcity of empirical research on studying gender differences when studying what motivates application users to dedicate time and resources to assisting

governments through their participation in citizen reporting applications.

Gender has a significant impact on how people organize their identities and how they interact with different stimuli surrounding them [10]. When it comes to personality, men and women are very different [11]. Common standards for men encourage rivalry, independence, status striving, and toughness, but women are expected to be caring, communal, and modest, rather than controlling, aggressive, dominant, or stubborn [12]. In the context of individuals' engagement in citizen reporting applications, Abu-Tayeh, et al. [4] found that gender is related to the number of reports. In particular, men were found to contribute substantially more frequently to the citizen reporting applications compared to women. This implies that the incentive to participate in citizen reporting is not distributed equally among genders. However, no previous studies investigated the role of gender in moderating the relationships between antecedents of motivations and individuals' engagement in citizen reporting applications. Gender is potentially important to our understanding of user acceptance of citizen reporting since it could play an important role in determining how users make their decisions about adopting and using such applications.

Hence, in order to bridge the gaps in the contemporary literature on citizen reporting applications stated above, the present study aims to investigate gender differences in the overlooked context of studying the impact of motivational factors behind the use of citizen reporting applications. In light of this, this research will employ the self-determination theory to elicit the main reasons that enhance citizens' participation in citizen reporting applications. The theory will be expanded to consider factors impacting users' intrinsic and extrinsic motivations to participate, as well as gender as a moderator factor. As such, this paper contributes to enhancing our understanding of the motivational factors behind citizens' participation in such applications, which are increasingly being used as part of many open government initiatives.

The reminders of this paper are organized as follows: the following section continues a review of relevant literature about citizen sourcing and reporting applications with reviewing the motivations behind citizen reporting behaviour. Next, the model and hypothesis development are presented, followed by the research methodology used in this research. We then continue with presenting the analysis and results. The final section discusses the results and concludes the study.

II. Literature Review

A. Citizens Sourcing and Reporting Applications

Three components of crowdsourcing, as identified by [13], are interacting to bring about the rise of a "collective intelligence system": the organization, the crowd, and the platform, which are represented in the context of this study by the government, citizens, and technology, respectively [14]. Crowdsourcing is sometimes called as 'citizen-sourcing' or 'citizen-reporting' in the context of open governments, as considered in this study [15]. Citizen reporting applications allow citizens to share information directly with the concerned governmental agencies via web-based or mobile platforms [4].

By offering this kind of applications, citizens can play an important role in providing situational awareness by sharing information with the government [6]. Governments can then employ the information gathered through these applications to enhance the quality of public services. Therefore, the goal of governments usage of citizen sourcing is to co-produce knowledge with citizens while the government remains fully responsible for carrying out its activities.

Evidence shows that government agencies are capable of utilizing citizen reporting applications to improve public services at a lower cost, carry out policy innovations, and boost public engagement [16]. Therefore, several governments have developed citizen reporting applications for different purposes. For example, in Switzerland, the "FixMyStreet" application allows citizens residing in Zurich to report issues regarding the infrastructure of the city to local authorities [4]. The Portuguese government has also developed a mobile application, "Citzens@City", that allows citizens to report city-related problems, such as potholes, poor road lighting, or the lack of accessibility of wheelchairs for people with disabilities, to the local authority. Citizens can easily download the application and use it to report the problem, which is categorized by a subject, description, location, and optional picture of the spot [17]. Additionally, the Saudi government has also created a set of national citizen reporting applications such as "Kulluna Amn" (which translates to "We are All Safe") to enable citizens to report security incidents, and "Balagh" (which translates to "report app"), to enables customers to report any commercial violations.

In order for these applications to be effective, citizens must be encouraged to participate in and engage with them. These apps will not provide the intended benefits unless citizens participate. As a result, this article will discuss the motivators that drive user participation and engagement in citizen reporting systems.

B. Motivations and Citizens Engagement in Citizen Reporting Applications

The motivation for human behaviour is one of the most important current subjects in psychology. Motivation stems from the dynamic relationship that occurs between an individual and their environment, which causes certain behaviours to occur [18]. Motivation is more than just the basic concept of motive, which remains relatively steady throughout an individual's lifespan. Motivation is primarily about the interaction between an individual's personal motives and a setting that may spark positive behaviour [19, 20].

Even though motivation is present in all activities, the effect and orientation of these drives vary. Motivation is grouped into three types according to the Self-Determination Theory (SDT) [21]: extrinsic motivations, intrinsic motivations, and amotivation. From an extrinsic motivation perspective, individuals' behaviour is driven through the use of external sources such as attaining better/more valuable outcomes or gaining reward [21]. On the other side, from an intrinsic motivation perspective, individual behaviour is self-driven, and the motivation comes from internal drives within the individuals themselves. Individuals' core values and interests are examples of intrinsic motivations, rather than the

value of the activity itself. Enjoyment and happiness, though important, are not the primary motivations for choosing to perform certain behaviour [22]. According to Waterman [23], only positive subjective states should be considered part of intrinsic motivation, whereas activities that are more related to hedonic enjoyment should be labelled as "hedonic motivation". Waterman [23], therefore, emphasised the importance that the term "intrinsic motivation" be redefined to relate specifically to activities that involve both eudaimonia and hedonic motivation. Lastly, from the amotivation perspective, amotivated people do not appear to have particular aims and goals, and they do not appear to approach their goals in a systematic manner [24]. They are not motivated either internally or externally [25].

In the context of this study, it is essential for governments to know what motivates citizens to engage and participate in citizen reporting applications especially as this kind of engagement is completely voluntary behaviour. Citizens are not obligated to participate and report observed incidents. In fact, citizens using this kind of application work as voluntary sensors to report issues and share knowledge with governments to help them perform their jobs and provide better services. Lin [26] examined the motivations behind individuals' participation in voluntary knowledge sharing within an organization. The study examined extrinsic motivations (i.e., expected organizational rewards and reciprocal benefits) and intrinsic motivations (i.e., knowledge self-efficacy, and enjoyment in helping others) as key factors influencing individuals' intentions to share knowledge. The results show that three motivational factors (i.e., reciprocal benefits, knowledge self-efficacy, and enjoyment in helping others) are significantly associated with an individual's intention to share knowledge. However, expected organizational rewards do not significantly influence individuals' behavioural intentions toward knowledge-sharing. Thapa, et al. [27] found that citizens with relevant expertise are self-motivated and have the courage and sense of responsibility to collaborate and engage in their field of expertise. They also found that rewards, though working as incentives, are not essential to ensure citizens' engagement in reporting applications. Another study conducted by Schmidthuber, et al. [1] found that the level of citizen motivation to engage in citizen reporting applications varies based on the types of users, i.e., proactive, interactive, and passive types of users. The study also found that citizens who always report offline tend to be more likely to report incidents online using citizen reporting applications. This indicates that regardless of the channel of reporting, citizens' motivations and personalities play a significant role in their willingness to report to the government. Additionally, Abu-Tayeh, et al. [4] studied the motivational factors behind citizens' engagement to use the "Zueri wie neu" application (which translates to "FixMyStreet") in the city of Zurich, Switzerland. The study indicates that both self-concern and other-orientation motivate citizens to voluntarily support the government by reporting incidents related to infrastructure using this platform, though self-concern is a slightly stronger driver. According to Alam, et al. [2], intrinsic motivations that are related to personal interest, fun, and community service are the main drivers for individuals' participation in voluntary crowdsourcing

applications and they only impact users short-term engagement. On the other hand, individuals' extrinsic motivations, such as recognition and rewards, are what actually drive long-term engagement. However, by reviewing other relevant studies in the field [8, 9, 28], one can indicate that both intrinsic and extrinsic motivational categories are to varying degrees, responsible for human behaviour.

In this study, amotivation, intrinsic and extrinsic motivations, and their antecedents will be considered as the main predictors of citizen participation in the citizen reporting application. Individual differences, particularly gender differences, will also be taken into account as they are theorized to moderate different relationships within models studying information system adoption in general [29].

III. HYPOTHESES AND MODEL DEVELOPMENT

The self-determination theory (SDT) is feasible in the context of this investigation. It will be used to investigate the motivating elements that drive user adoption and engagement in citizen reporting applications (see Fig. 1). The SDT theory can be expanded to include aspects that may elicit both intrinsic and extrinsic motivations, influencing an individual's behavioural intention to use such apps. Extrinsic motivation is driven by the user's perceptions of usefulness and benefits that will be gained from using an application [29, 30]. On the other hand, intrinsic motivation is driven by the user's perceptions of satisfaction and pleasure when performing a behaviour [31, 32]. This sense of pleasure motivates citizens to help and support the government by voluntarily reporting incidents in online citizen reporting applications [33]. Interestingly, Van der Heijden [34] found that although both types of motivation are significant predictors, antecedents of intrinsic motivation are stronger determinants of an individual's behaviour and intention to perform certain tasks compared to antecedents of extrinsic motivation. Psychological research found that a higher level of intrinsic motivation leads to users' inclination to dedicate more time to perform the task due to the self-satisfaction feeling they experience while performing the behaviour [21]. However, both types of motivations have been verified in the literature to be a strong incentive for users to participate in crowdsourcing projects and open government initiatives [8, 9, 27, 35]. Hence, this leads to the following hypotheses:

H1: Intrinsic motivation will positively impact users' behavioural intention to use citizen reporting applications.

H2: Extrinsic motivation will positively impact users' behavioural intention to use citizen reporting applications.

Amotivation must also be included in order to completely comprehend human behaviour [21]. Amotivation, which refers to the lack of motivation to perform certain activities, is a good predictor of human behaviour. According to Deci and Ryan [36], individuals are experiencing the amotivation state if they believe there is a lack of consistency between their behaviour and its outcomes, or when they feel incompetent and out of control [36]. Legault, et al. [37] further illustrate that amotivation is a result of an individual's feeling of lack of ability, the difficulty of the work, task characteristics, and the value derived from executing the task. Amotivated people

may feel fragmented or detached from their actions and, as a result, expend little effort or energy in carrying out assigned tasks. Such people will consider their behaviour as being beyond their control. Therefore, as individuals that are unmotivated simply do not demonstrate a willingness to participate in an activity, we hypothesis that:

H3: Amotivation will negatively impact users' behavioural intentions to use citizen reporting applications.

A. Antecedents of Intrinsic Motivation in Citizens Reporting Applications

Engagement in citizen reporting applications is in part driven by antecedents of intrinsic motivation, as the majority of citizen reporting applications are voluntary and users are typically not rewarded for their participation [38]. Based on the reviewed literature, the main antecedents of intrinsic motivation that are considered in this study are settling an issue related to self-concern or satisfying the feeling of revenge.

According to [22], revenge initiate a feeling of pleasure resulting from the relief of a painful tension. Although revenge is considered inappropriate and discouraged in modern society [39], it remains an emotionally and politically powerful force in society [40]. Revenge forms an intrinsic motive that drive individuals to take actions to relief the dissatisfaction feeling [41]. [42] found that customer retaliation entails a customer causing harm to a company in exchange for perceived losses committed by the company. Additionally, [43] has found that a vengeance motive (a strong desire to cause damage) increases the chance of "tangible" revenge behaviour. The emphasis on a desire for vengeance is vital since users are not always able to translate their desire into acts [44]. Citizen's sourcing applications, which allow citizens to report incidents, empower citizens and open an opportunity for citizens to report incidents that harm them directly. Hence, it offers a way for satisfying the revenge feeling. Therefore, the desire for revenge will motivate citizens to participate in citizens sourcing applications. In light of this, we hypothesis that:

H4: Revenge will positively impact users' intrinsic motivation towards using citizen reporting applications.

According to [4], self-concern is a strong intrinsic motivator for citizen reporting engagement. Self-concern is defined as the inclination to form one's behaviour with respect to the craving to secure and improve one's self-interest. Citizens may engage in citizen reporting applications in the hope of solving their own issues and get the most personal benefits from the platform [8]. An example of this is when the users are reporting an incident that impacts them personally. On the other hand, when self-concern is the main motive for using the platform, self-concern can be raised by setting higher aspirations [45]. In addition, several other studies [1, 9, 46, 47] found that self-concern is a significant predictor of citizens' participation in crowdsourcing projects, which include citizen reporting applications. Drawing on these studies, we hypothesis the following:

H5: Self-concern will positively impact users' intrinsic motivation towards using citizen reporting applications.

B. Antecedents of Extrinsic Motivation in Citizens Reporting Applications

Extrinsic motivation is behaviour that is instigated by external benefits such as monetary rewards, promotions, and other tangible rewards. Citizens' participation in the citizen's sourcing application is influenced in part by extrinsic motivational factors such as having a sense of and appreciation for social responsibility, expecting rewards, or receiving better service quality as a result of reporting an incident. Literature shows that this kind of motivation is thought of as a significant stimulus for adopting and using information systems in general [29, 48-50].

In the context of citizen reporting applications, [4] found that other-orientation (aka social responsibility) is a significant extrinsic motivational driver of citizen reporting engagement. Other-orientation is defined as the attempt to help others by reporting their issues to the government with the aim of solving their problems. In this case, altruistic motivation is the driving force behind citizens' voluntary engagement to use this kind of collaboration platform [51]. [46] study revealed that the main reason for individuals' participation in open-source platforms is the desire to help others. According to [1], individuals who are altruistic and interested in helping their community engage in more citizen sourcing activities more frequently compared to others. Caring and having love for the community initiate an altruistic motivation to engage in crowdsourcing projects [52]. However, feeling obligated to participate in order to be a good community member can occur [53]. In many cultural settings, especially those characterised as being collectivistic, social standing is determined by what one gives away rather than what one owns [54]. Hence, intensifying the obligation feeling can raise citizens' extrinsic motivation to participate as well as their direct willingness to engage in citizens reporting applications. Therefore, we hypothesis the following:

H6: Social responsibility will positively impact users' extrinsic motivation towards using citizen reporting applications.

H7: Social responsibility will positively impact users' behavioural intentions towards using the citizen reporting application.

The second antecedent of extrinsic motivation to be considered in this research is output quality. The second version of technology acceptance model (TAM2) posits that output quality is a determinant of users' perceptions of a system's usefulness [30]. This construct refers to the performance related consequences of doing the task [55]. Venkatesh and Davis [30] suggest that output quality judgments take the form of a profitability test, "in which, given a choice set containing multiple relevant systems, one would be inclined to choose a system that delivers the highest output quality" (pp.192). Citizens are more likely to participate in government citizen reporting applications if the system's expected output quality is high. In other studies by Winkler, et al. [56] and , Alhammad, et al. [9] they find that output quality takes on greater importance than any other variables related to extrinsic motivation. In light of the aforementioned, we derive the following hypothesis.

H8: Output quality will positively impact users' extrinsic motivation towards using citizen reporting applications.

Crowdsourcing research also discovered that reward and prizes were especially important in determining citizens' behaviour to participate in crowdsourcing, including citizen reporting applications [35, 57]. Monetary rewards can increase participants' willingness to report to the government as many participants treat crowdsourcing applications as a kind of employment [35]. According to Garcia Martinez and Walton [58], increasing the monetary compensation can attract more participants and therefore increase the success of the crowdsourcing project. In non-government crowdsourcing projects, such as Amazon Mechanical Turk, studies show that more than half of the crowd dedicates about eight hours to work on the platform as an additional source of income [57]. Another study conducted by Assegaff et al. (2016) found that if employees believe they can obtain organizational rewards for sharing their knowledge, they will be more willing to use virtual communities of practice (VCoPs) and to share their knowledge. In the context of the public sector, the U.S. Office of Management and Budget in 2009 encouraged the use of prizes and rewards to encourage citizen participation in open government initiatives [59]. This is because offering rewards and prizes initiates incentives and motives for citizens to participate and increases the chance of project success. On the other hand, while offering excessive tangible rewards increases participants' extrinsic motivations, it may weaken their intrinsic motivation [18]. This is known as the "over-justification" effect phenomenon. To simplify, it explains that if the behaviour is already intrinsically rewarding, offering extrinsic motivation will eliminate the enjoyment gained from performing it. Rewards as an extrinsic motivation should be offered when an individual needs to perform an unpleasant task. Therefore, government sectors should evaluate the right rewards to be offered for individuals participating in crowdfunding projects. As most of the offered citizen reporting applications do not incorporate enjoyment factors into their design, it would be expected that rewards would play a positive role in citizens' motivation and willingness to participate in this kind of project. Hence, in this study, we hypothesis the following:

H9: Rewards will positively impact users' extrinsic motivation towards using citizen reporting applications.

C. The Moderating Role of Gender

In the context of the proposed model, gender role theory [60] and social role theory [61] are employed to establish gender's differences as a moderator in the proposed model. According to these theories, gender differences are likely to moderate the impact of social responsibility and revenge. Literature demonstrates that gender differences exist when it comes to caring for other people and the environment, with women being more caring, concerned about environmental issues, and having more environmentally friendly ideas and beliefs [62]. Females not only care about environmental degradation and climate change, but they also maintain the awareness and abilities necessary to discover local solutions [63]. The authors in [64] stated that females, in general, are interested in acquiring and cultivating social behaviors such as

helping and caring for others. Revenge, on the other hand, is more associated with males than females, as females mostly assume a submissive attitude and avoid aggression and retaliation, whereas men are encouraged to demonstrate violent behaviours and seek revenge [65, 66]. [67] found that men had more vengeance dreams than women, whereas women thought vengeance was pointless. Hence, this study will hypothesis that:

H10: Gender moderates the relationship between social responsibility and users' behavioural intentions towards using citizen reporting applications.

H11: Gender moderates the relationship between revenge and users' intrinsic motivation towards using citizen reporting

These hypotheses are presented comprehensively in a proposed model shown in Fig. 1.
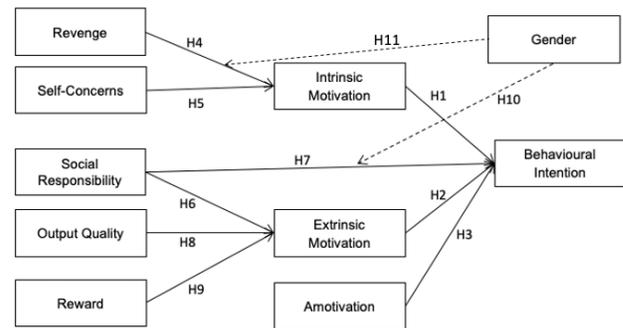
Fig. 1. Proposed Model.

IV. RESEARCH METHODOLOGY

A. Sample and Data Collection

The purpose of this empirical study is to investigate the direct effect of motivational factors on citizen participation in citizen reporting applications. It also intends to investigate the moderating effect of gender on some of the proposed relationships. To test the proposed hypotheses, this study applies the questionnaire survey method, which is a widely accepted method for model testing in the field of information systems [68]. A closed-ended structured questionnaire was designed using the SmartSurvey website. Trap questions were used to identify unengaged responses. The questionnaire was pilot-tested in order to assess its correctness, meaningfulness, and clarity. Specifically, 50 respondents participated in the pilot study before administering the survey to the targeted audience to certify the face validity of the applied measurement items. A slight refinement was made based on their feedback. After ensuring the quality of the questionnaire, the survey link was distributed to 1000 individuals living in Saudi Arabia through online channels such as emails, WhatsApp, and social media accounts. Only the respondents who had previously used one of the available citizen reporting applications were allowed to participate. After removing duplicate responses and unengaged responses, a total of 499 valid responses were received for analysis. Table I shows the demographic information of the respondents.

TABLE I.   RESPONDENTS DEMOGRAPHIC INFORMATION (N = 499)

Variable	Value	Frequency	Percentage
Age	Under 18	18	3.6 %
	18-24	72	14.43 %
	25-34	134	26.85 %
	35-54	237	47.49 %
	Over 55	38	7.62 %
Gender	Male	124	24.85 %
	Female	375	75.15 %
Educational level	High school	96	19.24 %
	Bachelor degree	328	65.73 %
	Master degree	48	9.62 %
	PhD.	11	2.2 %
	Other	16	3.2 %

B. Measurements

All of the items used in this study were derived from previously validated items and changed to match the goal of this study (see Table II). Behavioural intention and intrinsic and extrinsic motivation measurement items were adapted from [50] and [29]. Amotivation was measured using items adapted from [69]. Self-concern was measured using items adapted from [4] while social responsibility was measured using items adapted from both [4] and [1]. Items measuring rewards were adapted from [8]. Output quality measurement items were adapted from [55] and [29]. For revenge, measurement items, which were originally developed by [70] and adapted by [44], were used after alteration to suit the context of this study. All of measurement items were measured and operationalised using a five-point Likert scale ranging from strongly disagree (1) to strongly agree (5). The questionnaire was in English and translated into Arabic to ensure respondents understood the questions. Three experts used back translation to validate the translation's accuracy.

TABLE II.   MEASUREMENT ITEMS

VBs	Items	Source
Behavioural Intention	BI1: I am considering using this app to report incidents. BI2: I would seriously contemplate using this app. BI3: It is likely that I am going to use this app. BI4: I am likely to make future reports using this app.	Adapted from [50] and [29]
Extrinsic motivation	EM1: I find using this application useful. EM2: Using this application enables me to report incidents more efficiently. EM3: I can forward my concerns to local government directly.	
Intrinsic motivation	IM1: I participate in this application because I think that this participation is interesting. IM2: I participate in this application because this participation is fun IM3: I participate in this application because I feel good when doing this reporting	
Amotivation	AM1: There may be good reasons to do this activity, but personally I don't see any AM2: I do this activity but I am not sure if it is worth it	Adapted from [69].

output quality	OQ1: The use of this application will improve the quality of provided services. OQ2: The use of this application will contribute to the development of offered services. OQ3: Using this application will enhance the overall quality and efficiency of the provided services.	Adapted from [55]
Reward	RW1: My willingness to participate in this application would increase if there were monetary rewards RW2: I will really like to participate in this application if I would receive monetary rewards in return for my knowledge sharing.	Adapted from [8]
Revenge	RV1: My feeling of anger towards violators pushes me to use the application and report them RV2: I use this app to publicize the practices of violators and punish them. RV3: I submit a complaint via the application to avenge violators	Adapted from [70]
Self-concern	SC1: I took part in "Balag" because I could report problems that concerned me personally. SC2: I took part in "Balag" because I could report problems that prevented me from fulfilling my needs SC5: I participate in this application because I believe that this kind of reporting is important for me	Adapted from [4]
Social responsibilities	SR1: I took part in Balag application because it gives me the opportunity to protect others from fraud. SR2: I took part in this application because I could help the community by doing so. SR3: I want to contribute to the development of the services provided in my city by using this application. SR4: I would feel bad about myself if I don't share information about commercial violations with the relevant authorities. SR5: I participate in this application because I feel that this is something that I have to do it for the society.	Adapted from both [4] and [1]

V.   DATA ANALYSIS AND RESULTS

The study uses the partial least squares structure equation Modelling (PLS-SEM) method to analyse the collected data due to its ability to validate the measurement and test structural models at the same time. PLS is a comprehensive variance-based structural equation modelling for analysing skewed non-multinormal distribution, and its variance-based approach is more averse to overestimating relationships between constructs compared to the common SEM [71]. The PLS is also useful for analysing complex models with multiple constructs [72, 73]. Therefore, PLS was the ideal method to be used in this study due to the relative complexity of the proposed model. The analysis was done using the SmartPLS 4 for Mac OSX to examine the measurement and structural models.

A. Measurement Model

To validate the measurement model, the reliability and validity of the measurement model are evaluated. Table II shows that all of the measurement items show a good level of factor loading ranging from 0.560 to 0.963 which exceeds the minimum acceptable value of 0.5 [74]. Two measurement items, one from amotivation and one from reward, had to be removed because their factor loading values were less than 0.5. All of the items were loaded to their relevant construct than on any other constructs with the t-values showing significant results (p < 0.001). Therefore, these items were maintained for the analysis. In addition, to assess the

reliability of the maintained measurement items, Cronbach's alpha (α) and composite reliability (CR) were applied. The Cronbach's alpha (α) values for most of the constructs were above 0.7 which indicates a good internal consistency level of each construct [75]. Amotivation and revenge have slightly lower reliability level with Cronbach's alpha (α) values of 0.663 and 0.617 respectively. However, these Cronbach's alpha (α) values still exceed the minimum threshold value of 0.6 [76] and demonstrates an acceptable level. Composite reliability (CR) values of all of the constructs are above the recommended value of 0.7 [77]. Hence, all the constructs demonstrate acceptable internal consistency.

TABLE III.   THE MEASUREMENT MODEL STATISTICS

Variable	Items	Loading	T-value	α	CR	AVE
Behavioural Intention	BI1	0.841	47.041	0.867	0.909	0.715
	BI2	0.854	43.247			
	BI3	0.839	39.678			
	BI4	0.850	55.164			
Amotivation	AM1	0.917	18.233	0.663	0.773	0.636
	AM2	0.657	5.702			
Extrinsic motivation	EM1	0.832	52.969	0.750	0.856	0.665
	EM2	0.807	31.543			
	EM3	0.809	30.046			
Intrinsic motivation	IM1	0.769	22.409	0.721	0.834	0.626
	IM2	0.786	25.439			
	IM3	0.818	45.401			
output quality	OQ1	0.871	64.121	0.824	0.895	0.739
	OQ2	0.840	32.563			
	OQ3	0.867	53.018			
Reward	RW1	0.963	3.054	0.811	0.702	0.570
	RW2	0.560	1.274			
Revenge	RV1	0.786	27.894	0.617	0.797	0.567
	RV2	0.765	27.240			
	RV3	0.707	20.702			
Self-concern	SC1	0.749	24.626	0.721	0.824	0.541
	SC2	0.640	13.936			
	SC3	0.759	29.309			
	SC4	0.786	29.174			
Social responsibilities	SR1	0.825	34.577	0.861	0.900	0.643
	SR2	0.804	38.671			
	SR3	0.786	28.045			
	SR4.	0.762	32.847			
	SR5	0.830	33.293			

To assess the convergent validity of the constructs, the average variance extracted (AVE) was calculated. Table III shows that all of the constructs have an AVE above the recommended value of 0.5 [78] showing satisfactory convergent validity. On the other hand, discriminant validity was also assessed using the criterion of Fornell and Larcker [79] where the square root of AVE is calculated. Table IV shows that the square root of AVE for each construct is greater than its correlation with the other constructs, indicating that the measurement items used meet the discriminate validity conditions. Additionally, the Heterotrait-Monotrait ratio of correlations (HTMT), which is a new method for assessing the discriminant validity of the measurement model, was also assessed. Table V indicates that the HTMT values for each of the two distinct constructs in the proposed model are less than one. Thus, based on the HTMT criterion [80], discriminant validity is achieved, and the structural model can be evaluated.

TABLE IV.   DISCRIMINANT VALIDITY: SQUARE ROOT OF AVE (FORNELL-LARCKER CRITERION)

Variables	AM	BI	EM	IM	OQ	RV	RW	SC	SR
Amotivation	**0.798**								
Behavioural Intention	-0.219	**0.846**							
Extrinsic Motivation	-0.248	0.706	**0.816**						
Intrinsic Motivation	0.008	0.482	0.425	**0.791**					
Output Quality	-0.220	0.714	0.750	0.435	**0.860**				
Revenge	-0.018	0.512	0.505	0.559	0.489	**0.753**			
Reward	0.110	0.047	0.098	0.202	0.104	0.198	**0.755**		
Self-concern	-0.102	0.601	0.620	0.616	0.564	0.641	0.199	**0.736**	
Social Responsibilities	-0.208	0.785	0.767	0.490	0.784	0.532	0.056	0.629	**0.802**

TABLE V.   DISCRIMINANT VALIDITY: HETEROTRAIT-MONOTRAIT RATIO (HTMT)

	AM	BI	EM	IM	OQ	RV	RW	SC	SR
Amotivation									
Behavioural Intention	0.324								
Extrinsic Motivation	0.392	0.869							
Intrinsic Motivation	0.235	0.559	0.524						
Output Quality	0.339	0.843	0.843	0.494					
Revenge	0.253	0.692	0.736	0.795	0.678				
Reward	0.403	0.050	0.068	0.315	0.085	0.303			
Self-concern	0.219	0.746	0.835	0.792	0.728	0.867	0.268		
Social Responsibilities	0.320	0.808	0.856	0.542	0.936	0.725	0.081	0.784	

B. Structural Model Assessment

The second step is to analyse the structural model to test the proposed hypotheses and the theoretically established path. The bootstrapping procedure with 5000 samples set on PLS was used to obtain the path coefficients. The significance level was set to 5% to determine the significance levels for each of the path coefficients. The squared multiple correlation (R^2)

and significance of paths were used to assess the predictive power of the model and test the proposed hypotheses. The R^2 of extrinsic motivation is 0.652 which implies that social responsibilities, output quality, and rewards are able to explain the variance in extrinsic motivation by 65.2%. Likewise, revenge and self-concern are able to explain 42% of the variance of intrinsic motivation (R^2 = 0.42). The multi-mediation model explains 65.9% of behavioural intention variances. The results indicate that the proposed model has good predictive power, as all of the R^2 values are greater than the specified threshold value of 0.1 (Falk & Miller, 1992). In addition, the model fit was evaluated using SRMR, which, according to Henseler, et al. [81], can be used to avoid model misspecification. The SRMR value was 0.093, which is less than the required threshold value of 0.1, suggesting satisfactory model fit [71].

After assessing the proposed model's goodness of fit, hypotheses were evaluated to determine the significance of the identified relationships. As hypothesised in H1, H2, H3, and H7, intrinsic motivation (β = 0.122, t = 3.806, p < 0.001), extrinsic motivation (β = 0.213, t = 3.213, p < 0.005), amotivation (β = -0.045, t = 1.714, p < 0.05), and social responsibility (β = 0.812, t = 6.360, p < 0.001) exhibit significant impact on behavioural intention. As expected, intrinsic motivation, extrinsic motivation, and social responsibility all have a positive impact on behavioural intention, with social reasonability having the highest impact on variance, while amotivation has a negative impact. Therefore, H1, H2, H3, and H7 are supported. In addition, H4 and H5 evaluate whether revenge and self-concern have a significant positive impact on intrinsic motivation. The results indict that revenge (β = 0.278, t = 5.971, p < 0.001) and self-concern (β = 0.435, t = 9.369, p < 0.001) positively influence intrinsic motivation, thus validating H4 and H5. Additionally, extrinsic motivation is found to be positively influenced by social responsibilities (β = 0.494, t = 9.984, p < 0.001) and output quality (β = 0.354, t = 7.058, p < 0.001) with the social responsibility having the highest impact, supporting H6 and H8. Rewording, on the other hand, was found to have no effect on extrinsic motivation (β = 0.031, t = 0.790, p > 0.05), rejecting H9.

For the moderate relationships, the results demonstrate that gender moderates the relationship between social responsibility and behavioural intention. At (β = -0.159, t = 2.132, p < 0.01), this relationship is statistically significant. Thus, H10 is supported. Furthermore, it moderates the relationship between revenge and intrinsic motivation, which is statistically significant at (β = 0.303, t = 2.918, p < 0.005); thereby supporting H11. Fig. 2 depicts the results of the structural model, and Table VI shows the hypotheses test results.
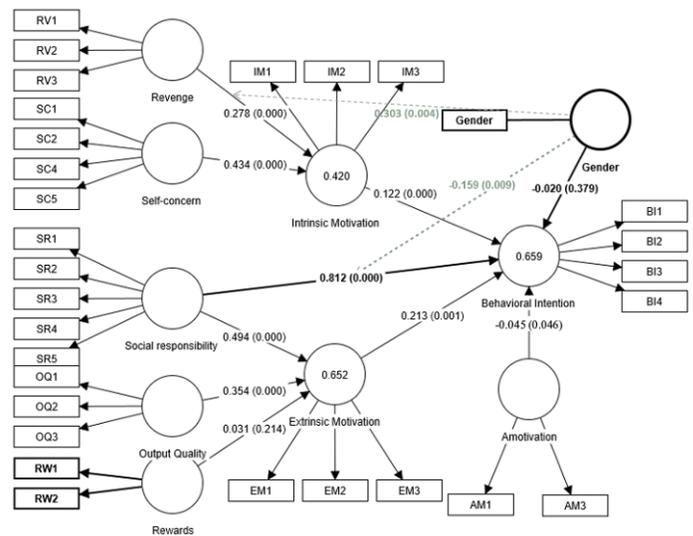
Fig. 2. The Results of the Empirical Study.

TABLE VI. OVERVIEW OF THE HYPOTHESES TEST RESULTS

H	Hypotheses	β	SD	T-value	P-Values	5.00%	95.00%	Supported / not supported
H1	Intrinsic Motivation → Behavioural Intention	0.122	0.034	3.806	0.000	0.074	0.189	supported
H2	Extrinsic Motivation → Behavioural Intention	0.213	0.066	3.213	0.001	0.106	0.324	supported
H3	Amotivation → Behavioural Intention	-0.045	0.028	1.714	0.046	-0.096	-0.006	supported
H4	Revenge → Intrinsic Motivation	0.278	0.047	5.971	0.000	0.204	0.357	supported
H5	Self-concern → Intrinsic Motivation	0.435	0.046	9.369	0.000	0.359	0.513	supported
H6	Social Responsibilities → Extrinsic Motivation	0.494	0.049	9.984	0.000	0.41	0.574	supported
H7	Social Responsibilities → Behavioural Intention	0.812	0.124	6.360	0.000	0.582	0.992	supported
H8	Output Quality → Extrinsic Motivation	0.354	0.050	7.058	0.000	0.273	0.439	supported
H9	Reward → Extrinsic Motivation	0.031	0.039	0.790	0.214	-0.051	0.079	Rejected
H10	Social Responsibilities *Gender → Behavioural Intention	-0.159	0.102	2.132	0.009	-0.387	-0.05	supported
H11	Revenge*Gender → Intrinsic Motivation	0.303	0.104	2.918	0.004	0.088	0.493	supported

VI. DISCUSSION AND IMPLICATIONS

As mentioned earlier, the aim of this study is to examine the motivational factors behind the usage of citizen reporting applications and investigate the role of gender differences in moderating the proposed relationships. The statistical analysis indicates that both intrinsic and extrinsic motivations positively influence individuals' behavioural intention, with extrinsic motivation being the highest predictor. On the other hand, amotivation, as expected, was found to negatively influence individuals' behavioural intentions. Surprisingly, these findings support the current and efficacy of the three types of motivations identified in self-determination theory [21]. Individuals experiencing a low amount of self-determination (amotivation) are less likely to engage in the citizen reporting applications, whereas individuals having a high amount of self-determination (intrinsic or extrinsic motivation) are more likely to participate. Several awareness efforts should be launched by government organizations to raise citizens' understanding of the benefits of their collaboration and participation in the use of citizen reporting applications. Hence, enhancing their self-determination level.

Self-concern and revenge are found to significantly influence individuals' intrinsic motivation, with self-concern being more influential. The significant impact of self-concern is consistent with previous research findings such as Abu-Tayeh, et al. [4], Schmidthuber, et al. [1], Alhammad, et al. [9], Wu, et al. [46], Oreg and Nov [47]. It stands to reason that citizens will report concerns that directly affect them in order to solve their own problems and meet their own demands. Individuals that are preoccupied with self-concern are proven to be more persistent contributors. Similarly, when an individual feels vengeance, it increases his/her self-determination level to impose harm on the offender. By offering citizen reporting applications to report incidences in a variety of fields, the government provides a means for individuals to express their rage without causing physical harm. Interestingly, the results show that revenge boosts men's willingness to harm the offender as compared to women, who are less prone to revenge. This corresponds to the personal attributes of men and women highlighted in the gender role theory [60] and the social role theory [61]. It also supports Mullins, et al. [65] illustration that females are more submissive and tend to avoid aggression and reprisal, whereas men, in several cultures, are raised to engage in violent behavior and seek vengeance.

Extrinsic motivation was found to be significantly impacted by social responsibility. Social responsibility is not only the strongest predictor of extrinsic motivation, but it is also the strongest direct predictor of an individual's behavioural intentions to use citizen reporting applications. This result is in line with the findings of Abu-Tayeh, et al. [4] and Schmidthuber, et al. [1]. Hence, the impact of social responsibility should not be underestimated. Altruistic people often go above and beyond in order to serve others and protect their community. The government should promote collectivism's social principles in order to create a loving community that looks out for one another. Creating such a community will boost the number of people who participate in citizen reporting applications. Furthermore, while women were found to have a higher sense of social responsibility than men, the impact of social responsibility principles in driving men to participate in citizen reporting applications was considerably sharper compared to women, where the impact is steady. This could be because men have a strong sense of duty and dedication when they believe in something [60]. Additionally, the majority of women, according to the analysis, have a greater level of social responsibility and actively participate in citizen reporting applications, making the analysis of this variable less sensitive to covariance analysis.

Output quality was also found to significantly impact extrinsic motivation. Many citizens participate in citizen reporting applications with the hope that they will receive better public services and an improved environment. Several technology acceptance theories, such as TAM2 [30] and UTAUT [48], emphasise the role of output quality in determining user acceptance of technology. The government should report on its progress as a result of citizen participation in citizen reporting applications. If citizens who participate cannot see the outcomes of their reporting, they may be reluctant to take part again.

Although many studies [35, 57] emphasised the importance of rewards in increasing individuals' willingness to perform certain tasks, rewards were not found to influence citizens' extrinsic motivation. The finding, though not expected, is not surprising, as some studies (e.g. Deci, et al. [18]) found that reward might produce negative consequences by weakening participants' intrinsic motivation. Hence, offering rewards will not intensify citizens' motivation to participate. Rewards should only be offered in citizen reporting applications if the reporting process is unpleasant and will not result in a benefit for the reporter or his/her community.

The current study makes numerous contributions to the recent literature. For example, the current study adds to the existing citizen reporting adoption and usage literature by investigating the motivational factors that influence citizens' engagement across multiple contexts, whereas the majority of previous studies have focused on citizen reporting adoption and usage in a single context. Furthermore, it investigates the impact of amotivation on users' behavioural intentions to use citizen reporting applications, which has not previously been investigated to the best of the author's knowledge. Similarly, the current study is the first to add considerably to the existing citizen reporting literature by studying the role of gender in moderating the influence of social responsibility and revenge.

VII. CONCLUSION

We conclude that extrinsic motivation, internal motivation, and amotivation all influence citizen involvement in citizen reporting applications. According to our findings, the motivating antecedents that strongly influence individuals' motivation to participate in citizen reporting applications are social responsibility, output quality, self-concern, and revenge. When compared to other drivers, social responsibility is the most powerful. The study also discovered that there are gender differences that influence the relationship between social responsibility and citizens' behavioural intentions, as well as

the relationship between revenge and intrinsic motivation. Future research should look into the function of other socioeconomic characteristics in moderating the observed motivational factors and their impact on individuals' behavioural intentions to participate in citizen reporting applications.

REFERENCES

[1] L. Schmidthuber, D. Hilgers, T. Gegenhuber, and S. Etzelstorfer, "The emergence of local open government: Determinants of citizen participation in online service reporting," Government Information Quarterly, vol. 34, no. 3, pp. 457-469, 2017/09/01/ 2017, doi: https://doi.org/10.1016/j.giq.2017.07.001.

[2] S. L. Alam, R. Sun, and J. Campbell, "Helping yourself or others? motivation dynamics for high-performing volunteers in GLAM crowdsourcing," Australasian Journal of Information Systems, vol. 24, pp. 1-25, 2020, doi: 10.3127/AJIS.V24I0.2599.

[3] M. Weiss, "Crowdsourcing Literature Reviews in New Domains," Technology Innovation Management Review, vol. 6, no. 2, pp. 5-14, 2016, doi: 10.22215/timreview963.

[4] G. Abu-Tayeh, O. Neumann, and M. Stuermer, "Exploring the Motives of Citizen Reporting Engagement: Self-Concern and Other-Orientation," Business & Information Systems Engineering, 2018, doi: 10.1007/s12599-018-0530-8.

[5] A. Desai et al., "Crowdsourcing a crisis response for COVID-19 in oncology," Nature Cancer, vol. 1, no. 5, pp. 473-476, 2020, doi: 10.1038/s43018-020-0065-z.

[6] D. Linders, "From e-government to we-government: Defining a typology for citizen coproduction in the age of social media," 2012, doi: 10.1016/j.giq.2012.06.003.

[7] C.-M. Lönn, E. Uppström, A. Nilsson, and R. Lönn, "DESIGNING AN M-GOVERNMENT SOLUTION: ENABLING COLLABORATION THROUGH CITIZEN SOURCING DESIGNING AN M-GOVERNMENT SOLUTION: ENABLING COLLABORATION THROUGH CITIZEN SOURCING," Association for Information Systems AIS Electronic Library (AISeL), 2016. [Online]. Available: http://aisel.aisnet.org/ecis2016_rp http://aisel.aisnet.org/ecis2016_rp/86.

[8] F. Wijnhoven, M. Ehrenhard, and J. Kuhn, "Open government objectives and participation motivations," Government Information Quarterly, vol. 32, no. 1, pp. 30-42, 2015, doi: 10.1016/J.GIQ.2014.10.002.

[9] M. M. Alhammad, L. Hajar, S. Alshathry, and M. Alqasabi, "Motivational Factors Impacting the Use of Citizen Reporting Applications in Saudi Arabia: The Case of Balagh Application," International Journal of Advanced Computer Science and Applications, vol. 12, no. 6, 2021.

[10] V. Venkatesh and M. G. Morris, "Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior," Management Information Systems Quarterly, vol. 24, no. 1, pp. 115-139, 2000.

[11] M. E. Heilman, "Description and Prescription: How Gender Stereotypes Prevent Women's Ascent Up the Organizational Ladder," Journal of Social Issues, vol. 57, no. 4, pp. 657-674, 2001, doi: https://doi.org/10.1111/0022-4537.00234.

[12] C. A. Moss-Racusin, J. E. Phelan, and L. A. Rudman, "When men break the gender rules: status incongruity and backlash against modest men," Psychology of Men & Masculinity, vol. 11, no. 2, p. 140, 2010.

[13] Y. Zhao and Q. Zhu, "Evaluation on crowdsourcing research: Current status and future direction," Information Systems Frontiers, vol. 16, no. 3, pp. 417-434, 2014.

[14] K. Cupido and J. Ophoff, "A Model of Fundamental Components for an e-Government Crowdsourcing Platform," Electronic Journal of e-Government, vol. 12, no. 2, pp. 141-156, 2014.

[15] E. Estellés-Arolas and F. González-Ladrón-de-Guevara, "Towards an integrated crowdsourcing definition," Journal of Information Science, vol. 38, no. 2, pp. 189-200, 2012/04/01 2012, doi: 10.1177/0165551512437638.

[16] P. Dutil, "Crowdsourcing as a new instrument in the government's arsenal: Explorations and considerations," Canadian Public Administration, vol. 58, no. 3, pp. 363-383, 2015, doi: https://doi.org/10.1111/capa.12134.

[17] A. M. Ribeiro, R. P. Costa, L. Marcelino, and C. Silva, "Citizens@City Mobile Application for Urban Problem Reporting," in ENTERprise Information Systems, Berlin, Heidelberg, M. M. Cruz-Cunha, J. Varajão, P. Powell, and R. Martinho, Eds., 2011// 2011: Springer Berlin Heidelberg, pp. 141-150.

[18] E. L. Deci, R. Koestner, and R. M. Ryan, "Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again," Review of Educational Researchl Spring, vol. 71, no. 1, pp. 1-27, 2001, doi: 10.3102/00346543071001001.

[19] M. Fishbein and I. Ajzen, Belief, attitude, intention and behaviour: An introduction to theory and research. Reading, Mass: Addison-Wesley, 1975.

[20] I. Ajzen, "The theory of planned behavior," Organizational behavior and human decision processes, vol. 50, no. 2, pp. 179-211, 1991.

[21] E. L. Deci and R. M. Ryan, Intrinsic Motivation and Self-Determination in Human Behavior. Springer, 1985.

[22] E. Fromm, Man for himself; an inquiry into the psychology of ethics (Man for himself; an inquiry into the psychology of ethics.). Oxford, England: Rinehart, 1947, pp. xiv, 254-xiv, 254.

[23] A. S. Waterman, "When effort is enjoyed: Two studies of intrinsic motivation for personally salient activities," Motivation and Emotion, vol. 29, no. 3, pp. 165-188, 2005.

[24] V. Barkoukis, H. Tsorbatzoudis, G. Grouios, and G. Sideridis, "The assessment of intrinsic and extrinsic motivation and amotivation: Validity and reliability of the Greek version of the Academic Motivation Scale," Assessment in Education: Principles, Policy & Practice, vol. 15, no. 1, pp. 39-55, 2008, doi: 10.1080/09695940701876128.

[25] L. G. Pelletier, K. M. Tuson, M. S. Fortier, R. J. Vallerand, N. M. Briére, and M. R. Blais, "Toward a New Measure of Intrinsic Motivation, Extrinsic Motivation, and Amotivation in Sports: The Sport Motivation Scale (SMS)," Journal of Sport and Exercise Psychology, vol. 17, no. 1, pp. 35-53, 1995, doi: 10.1123/jsep.17.1.35.

[26] H. F. Lin, "Effects of extrinsic and intrinsic motivation on employee knowledge sharing intentions," Journal of Information Science, vol. 33, no. 2, pp. 135-149, 2007, doi: 10.1177/0165551506068174.

[27] B. E. P. Thapa, B. Niehaves, C. E. Seidel, and R. Plattfaut, "Citizen involvement in public sector innovation: Government and citizen perspectives," Information Polity, vol. 20, pp. 3-17, 2015, doi: 10.3233/IP-150351.

[28] G. V. Krogh, S. Haefliger, S. Spaeth, and M. W. Wallin, "Carrots and rainbows: motivation and social practice in open source software development," MIS Q., vol. 36, no. 2, pp. 649–676, 2012.

[29] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," Management Information Systems Quarterly, pp. 425-478, 2003.

[30] V. Venkatesh and F. D. Davis, "A theoretical extension of the technology acceptance model: Four longitudinal field studies," Management science, vol. 46, no. 2, pp. 186-204, 2000.

[31] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," Contemporary educational psychology, vol. 25, no. 1, pp. 54-67, 2000.

[32] V. Venkatesh, "Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model," Information systems research, vol. 11, no. 4, pp. 342-365, 2000.

[33] R. E. Schumacker and R. G. Lomax, A Beginner's Guide to Structural Equation Modeling: Third Edition. Taylor & Francis, 2012.

[34] H. Van der Heijden, "User acceptance of hedonic information systems," MIS quarterly, pp. 695-704, 2004.

[35] H. K. Liu, "Crowdsourcing Government: Lessons from Multiple Disciplines," Public Administration Review, vol. 77, no. 5, pp. 656-667, 2017, doi: 10.1111/puar.12808.

[36] E. L. Deci and R. M. Ryan, "Self-determination theory." Thousand Oaks, CA: Sage Publications Ltd, 2012, pp. 416-436.

[37] L. Legault, I. Green-Demers, and L. Pelletier, "Why do high school students lack motivation in the classroom? Toward an understanding of academic amotivation and the role of social support," Journal of educational psychology, vol. 98, no. 3, p. 567, 2006.

[38] H. Wang, J. E. Chung, N. Park, M. L. McLaughlin, and J. Fulk, "Understanding Online Community Participation," Communication Research, vol. 39, no. 6, pp. 781-801, 2012, doi: 10.1177/0093650211408593.

[39] L. H. Grobbink, J. J. Derksen, and H. J. van Marle, "Revenge: An analysis of its psychological underpinnings," International journal of offender therapy and comparative criminology, vol. 59, no. 8, pp. 892-907, 2015.

[40] P. Fischer, S. A. Haslam, and L. Smith, ""If you wrong us, shall we not revenge?" Social identity salience moderates support for retaliation in response to collective threat," Group Dynamics: Theory, Research, and Practice, vol. 14, no. 2, p. 143, 2010.

[41] N. N. Bechwati and M. Morrin, "Outraged consumers: Getting even at the expense of getting a good deal," Journal of Consumer Psychology, vol. 13, no. 4, pp. 440-453, 2003.

[42] Y. Grégoire, F. Ghadami, S. Laporte, S. Sénécal, and D. Larocque, "How can firms stop customer revenge? The effects of direct and indirect revenge on post-complaint responses," Journal of the Academy of Marketing Science, vol. 46, no. 6, pp. 1052-1071, 2018/11/01 2018, doi: 10.1007/s11747-018-0597-2.

[43] H. Zourrig, J.-C. Chebat, and R. Toffoli, "Consumer revenge behavior: A cross-cultural perspective," Journal of Business Research, vol. 62, no. 10, pp. 995-1001, 2009, doi: https://doi.org/10.1016/j.jbusres.2008.08.006.

[44] Y. Grégoire, D. Laufer, and T. M. Tripp, "A comprehensive model of customer direct and indirect revenge: Understanding the effects of perceived greed and customer power," Journal of the Academy of Marketing Science, vol. 38, no. 6, pp. 738-758, 2010, doi: 10.1007/s11747-009-0186-5.

[45] C. K. W. De Dreu and A. Nauta, "Self-interest and other-orientation in organizational behavior: Implications for job performance, prosocial behavior, and personal initiative," vol. 94, ed. De Dreu, Carsten K. W.: Department of Psychology, University of Amsterdam, Roetersstraat 15, Amsterdam, Netherlands, 1018 WB, c.k.w.dedreu@uva.nl: American Psychological Association, 2009, pp. 913-926.

[46] C.-G. Wu, J. H. Gerlach, and C. E. Young, "An empirical analysis of open source software developers' motivations and continuance intentions," Information & Management, vol. 44, no. 3, pp. 253-262, 2007, doi: https://doi.org/10.1016/j.im.2006.12.006.

[47] S. Oreg and O. Nov, "Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values," Computers in Human Behavior, vol. 24, no. 5, pp. 2055-2073, 2008, doi: https://doi.org/10.1016/j.chb.2007.09.007.

[48] V. Venkatesh, J. Thong, and X. Xu, "Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology," Management Information Systems Quarterly, vol. 36, no. 1, pp. 157-178, 2012.

[49] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User acceptance of computer technology: a comparison of two theoretical models," Management science, vol. 35, no. 8, pp. 982-1003, 1989.

[50] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "Extrinsic and intrinsic motivation to use computers in the workplace1," Journal of applied social psychology, vol. 22, no. 14, pp. 1111-1132, 1992.

[51] M. Kube, D. Hilgers, G. Koch, and J. Füller, "Explaining voluntary citizen online participation using the concept of citizenship: an explanatory study on an open government platform," Journal of Business Economics, vol. 85, no. 8, pp. 873-895, 2015/11/01 2015, doi: 10.1007/s11573-014-0756-y.

[52] D. C. Brabham, "MOVING THE CROWD AT THREADLESS," Information, Communication & Society, vol. 13, no. 8, pp. 1122-1145, 2010/12/01 2010, doi: 10.1080/13691181003624090.

[53] K. R. Lakhani and E. von Hippel, "How open source software works: "free" user-to-user assistance," Research Policy, vol. 32, no. 6, pp. 923-943, 2003/06/01/ 2003, doi: https://doi.org/10.1016/S0048-7333(02)00095-1.

[54] G. Hofstede, G. J. Hofstede, and M. Minkov, Cultures and Organizations: Software of the Mind, 3rd Edition ed. USA: McGraw-Hill, 2010.

[55] D. R. Compeau and C. A. Higgins, "Computer self-efficacy: Development of a measure and initial test," Mis Quarterly, vol. 19, no. 2, pp. 189-211, 1995.

[56] T. J. Winkler, O. Günther, and G. Trouvilliez, "Participatory urban sensing: Citizens' acceptance of a mobile reporting service," Ecis, no. 2012, pp. 1-10, 2012. [Online]. Available: http://aisel.aisnet.org/ecis2012/106.

[57] N. Kaufmann, T. Schulze, and D. Veit, "More than fun and money. Worker Motivation in Crowdsourcing – A Study on Mechanical Turk," Proceedings of the Seventeenth Americas Conference on Information Systems, vol. 4, no. 2009, pp. 1-11, 2011, doi: 10.1145/1979742.1979593.

[58] M. Garcia Martinez and B. Walton, "The wisdom of crowds: The potential of online communities as a tool for data analysis," Technovation, vol. 34, no. 4, pp. 203-214, 2014/04/01/ 2014, doi: https://doi.org/10.1016/j.technovation.2014.01.011.

[59] I. Mergel, "Social media adoption and resulting tactics in the U.S. federal government," Government Information Quarterly, vol. 30, no. 2, pp. 123-130, 2013, doi: 10.1016/J.GIQ.2012.12.004.

[60] S. Shimanoff, "Gender role theory," in Encyclopedia of communication theory, vol. 1, S. W. Littlejohn and K. A. Foss Eds.: Sage, 2009.

[61] A. H. Eagly, W. Wood, and A. B. Diekman, "Social role theory of sex differences and similarities: A current appraisal," The developmental social psychology of gender, vol. 12, no. 174, pp. 9781410605245-12, 2000.

[62] M. X. Yang, X. Tang, M. L. Cheung, and Y. Zhang, "An institutional perspective on consumers' environmental awareness and pro-environmental behavioral intention: Evidence from 39 countries," Business Strategy and the Environment, vol. 30, no. 1, pp. 566-575, 2021, doi: https://doi.org/10.1002/bse.2638.

[63] N. Ahmad et al., "Corporate social responsibility at the micro-level as a "new organizational value" for sustainability: Are females more aligned towards it?," International Journal of Environmental Research and Public Health, vol. 18, no. 4, p. 2165, 2021.

[64] J. F. Dovidio, J. A. Piliavin, D. A. Schroeder, and L. A. Penner, The social psychology of prosocial behavior. Psychology Press, 2017.

[65] C. W. Mullins, R. Wright, and B. A. Jacobs, "Gender, streetlife and criminal retaliation," Criminology, vol. 42, no. 4, pp. 911-940, 2004.

[66] S. L. Feld and R. B. Felson, "Gender norms and retaliatory violence against spouses and acquaintances," Journal of Family Issues, vol. 29, no. 5, pp. 692-703, 2008.

[67] L. Goldner, R. Lev-Wiesel, and G. Simon, "Revenge fantasies after experiencing traumatic events: Sex differences," Frontiers in psychology, vol. 10, p. 886, 2019.

[68] A. Lee, "Researching MIS," in Rethinking Management Information Systems : An Interdisciplinary Perspective: An Interdisciplinary Perspective, W. Currie and B. Galliers Eds.: OUP Oxford, 1999.

[69] F. Guay, R. J. Vallerand, and C. Blanchard, "On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS)," Motivation and Emotion, vol. 24, no. 3, pp. 175-213, 2000, doi: 10.1023/A:1005614228250.

[70] K. Aquino, T. M. Tripp, and R. J. Bies, "How employees respond to personal offense: the effects of blame attribution, victim status, and offender status on revenge and reconciliation in the workplace," (in eng), The Journal of applied psychology, vol. 86, no. 1, pp. 52-9, Feb 2001, doi: 10.1037/0021-9010.86.1.52.

[71] J. F. Hair, G. T. M. Hult, C. Ringle, and M. Sarstedt, A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). SAGE Publications, 2013.

[72] W. W. Chin, "Issues and opinions on structural equation modeling," Mis Quarterly, vol. 22, no. 1, pp. 7-16, 2003.

[73] K. K.-K. Wong, "Partial Least Squares Structural Equation Modeling (PLS-SEM) Techniques Using SmartPLS," Marketing Bulletin, vol. 24, no. 1, pp. 1-32, 2013. [Online]. Available: http://marketing-bulletin.massey.ac.nz/v24/mb_v24_t1_wong.pdf%5Cnhttp://www.resear

chgate.net/profile/Ken_Wong10/publication/268449353_Partial_Least_Squares_Structural_Equation_Modeling_(PLS-SEM)_Techniques_Using_SmartPLS/links/54773b1b0cf293e2da25e3f3.pdf.

[74] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, 7th, Ed. Multivariate data analysis. Englewood Cliffs, NJ: Prentice Hall, 2009.

[75] R. P. Bagozzi and Y. Yi, "On the evaluation of structural equation models," Journal of the Academy of Marketing Science, vol. 16, no. 1, pp. 74-94, 1988, doi: 10.1007/BF02723327.

[76] J. C. Nunnally, Psychometric Theory. New York: McGraw-Hill, 1978.

[77] R. P. Bagozzi, Y. Yi, and L. W. Phillips, "Assessing Construct Validity in Organizational Research," Administrative Science Quarterly, vol. 36, no. 3, pp. 421-458, 1991, doi: 10.2307/2393203.

[78] R. P. Bagozzi, "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error: A Comment," Journal of Marketing Research, vol. 18, no. 3, pp. 375-381, 1981, doi: 10.2307/3150979.

[79] C. Fornell and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," Journal of marketing research, vol. 18, no. 1, pp. 39-50, 1981.

[80] J. F. Hair, C. William, B. Babin, and R. Anderson, "Multivariate data analysis," ed: Upper Saddle River, NJ: Prentice Hall, 2010.

[81] J. Henseler, G. Hubona, and P. A. Ray, "Using PLS path modeling in new technology research: updated guidelines," Industrial management & data systems, 2016.

Effect of Visuospatial Ability on E-learning for Pupils of Physics Option in Scientific Common Trunk

Khalid Marnoufi[1], Imane Ghazlane[2], Fatima Zahra Soubhi[3], Bouzekri Touri[4], Elhassan Aamro[5]

Laboratory of Sciences and Technologies of Information and Education, STCED Research Team, FSBM, Hassan II University of Casablanca, Morocco[1,2,3,4,5]

Higher Institute of Nursing and Health Technology Professions Casablanca-Settat, Morocco Street Faidouzi Mohamed, 20250 Casablanca, Morocco[2]

Abstract—**This study aims to reveal the existence of a relationship between the visuospatial capacity of pupils with a specialization in physics, with high educational performance, and the capacity for E-learning. To achieve the study, we used the Wechsler intelligence test of cognitive ability. Our sample is composed of 204 adolescents, whose average age is 15 years, 12 months, and 11 days, with a standard deviation of 00 years, 1 month and 19 days. The selection criterion was based on the general results and specifically the physics science mark. The results of the study showed the existence of a significant relationship between visuospatial ability and scientific thinking, and statistically significant homogeneity attributed to specialization in visuospatial ability and creative thinking.**

Keywords—Visuospatial; e-learning; physics; intelligence

I. INTRODUCTION

For more than two thousand years, psychological studies have been interested in mental abilities and have occupied a prominent place among cognitive psychologists because of the importance of this topic in the professional, academic and social life of an individual. Mental abilities are of great importance in guiding the individual to studies appropriate to his abilities and preparing him to integrate into it, something that will allow him to know his faculties. The holder of high mechanical ability is qualified to study mechanical engineering with distinction, the holder of high linguistic ability is qualified to study linguistics with distinction, and the holder of distinguished ability to understand mathematical problems is qualified to study successfully engineering and mathematics [1]. Indeed, spatial visual ability has a significant impact on pupils, as it develops their mental abilities, and helps them to understand study material and solve problems they face from different method, so as to raise their level of success [2]. This visuo-spatial ability is defined as the ability to represent and transform non-linguistic symbolic information [3]. Spatial visual ability requires the ability to mentally rotate models, as well as short-term visual memory, in addition to a series of sequential operations [4].

A lot of studies indicate the relationship between visuospatial ability, performance, and academic achievement, and have shown the superior performance of high-performing pupils in measures of visuospatial ability. Various studies have also shown that academic performance and success in many subjects is associated with high visuospatial ability. Studies indicate that the visuospatial ability of pupils when solving mathematical problems is related to their abilities for academic achievement, [5].

Some experimental research shows the importance of the role of visuospatial experience in the development of mental abilities, and the importance of visuospatial experience in explaining the differences between individuals in abilities [6]. A study indicates that movement and displacement in the environment is one of the most important sources of learning visuospatial skills [7], which reinforces the role of the environment in abilities development.

Visuospatial experience can be acquired through academic specialization, and this is confirmed by a study that found that pupils with math and engineering backgrounds go on to study science and engineering in college and succeed in advantages than others. They did a long way in shortening many visuospatial concepts and thinking about their performance, compared to others that still need exposure to experiences to improve that performance [8]. This confirms that differences in pupils' visuospatial abilities date to their earliest school and extracurricular experiences [9]. The study specify that shape visualization abilities increase whenever pupils are exposed to appropriate experiences, in which the teacher plays a major role [10].

Intelligence tests are the best tests for measuring visuospatial ability, however it is rare in our environment that this skill is studied, or indeed measured, before pupils choose the best academic orientation. As a result, many pupils who fail visuospatial skills may complete their studies in fields that require such skills, such as science and technology, or vice versa, where pupils who have visuospatial skills complete their studies in fields such as than languages. It doesn't depend on the high school, but also on the university course.

Intelligence is one of the important factors that help the pupil achieve success. The most intelligent pupils are more capable than others of learning and acquiring the experiences provided by the environment [11] [12] [13] [14].

Pupils with learning difficulties are considered a heterogeneous group, whether in terms of intelligence or achievement in academic subjects. In terms of intelligence for Moroccan pupils, some pupils' intelligence is average, and others' intelligence is above average; This means that they possess mental abilities that help them to learn what their ordinary peers are learning while making some necessary

adaptations, especially in the teaching methods used in their education, which is reflected in their development in university research [15] [16] [17]. Thus, the use of techniques and electronic media in learning requires skills and tasks on the part of the learner such as the quality of knowledge of E-learning, where visuospatial intelligence exceeds the capabilities required for successful learning. E-Learning in period of COVID, is among the causes of this study, since learning was online where many new educational technologies were applied, which makes E-Learning possible, such as visual digital media that makes teaching effective. The visuospatial interaction of the learner with electronic media is the first to control performance and academic success.

This study is the first in Morocco, this research work aims to research the types of abilities via a cognitive and E-learning test with pupils with a scientific orientation, physics option, who excel in performance and academic success. The results of this study will provide on the one hand to researchers and specialists a theoretical basis on the role of visuospatial intelligence in the success of E-learning, on the other hand to evaluate educational programs which aim the development of abilities and the skills of his pupils.

II. METHOD

A. Participants

Our sample is composed of many adolescents' pupils from the Physics Option in the scientific common trunk in the city of Safi in Morocco. The number of pupils who participated in our survey was 204 pupils with a mean age is 15 years 12 months 11 days with a standard deviation of 00 years 1 month 19 days which corresponds to a very important period in a person's life, when the child undergoes a range of changes, such as physical, emotional, social, cognitive, perceptual, and other developments. The study included the same number of participants of both sexes.

B. Material

In order to calculate Visuospatial ability, two subtests of a cognitive test were choice from the Wechsler Intelligence Scale for Children and Adolescents. The Visuospatial intelligence scale is calculated by both the Cubes and Visual Puzzles subtests. The Visuospatial Intelligence Scale subtests were administered individually during vacant hours. The administration of the scale was carried out by the use of the cubes and the images in the booklet of stimuli which calculates the Visuospatial intelligence. The attribution and rating were carried out according to the procedure of the administration and rating manual [18] [19]. The questionnaire given to pupils includes questions on the social and cultural environment of the family as well as the visual and auditory problems which the pupils could suffer.

C. Data Collection and Analysis

The overall raw scores obtained after running each subtest were transmitted for unlimited correction online to Q-global instead of manual conversion, which requires more time and may lead to miscalculations. The final results were obtained in the form of Standard Scores and Composite Scores. In order to analyse the study data, the program Statistical Package for

Social Sciences (SPSS version 25) was used for research of reliability, statistical descriptive and t-test results.

D. Socio-Economic Status and Level of Education

The cultural characteristics of the social environment of the participants include the level of education of the child and his parents which reflect on the social level of the families of the participants. The study conducted represents the Moroccan public school which suffers from the absence of children from the high earning class and the absence of children from an integrated family educational environment, in terms of the presence of a high level of education. The level of parental education is very low, and all participants haven't a telephone or a computer, at least one member of their family have a telephone that can be connected to the internet.

III. RESULTS

Pupils participating in this study get an average Composite Score equal to the theoretical average in visuospatial intelligence of 100.43, with a standard deviation half of the theoretical average of 8.98. Thus, girls' pupils and boys' pupils get equal marks.

Concerning the ranking of the results studies, we note that important number of the participating pupils amounted to 65.70% on average, 19.60% of the participating pupils on average strong, and 14.70% of the participating pupils on average weak. The reliability measured by Cronbach's alpha in this test was 0.860.

The average and the standard deviations of the standard marks of the two subtests Cubes and Visual Puzzles which builds the visuospatial intelligence are successively with the theoretical average. The Cubes subtest average is 10.01 with a standard deviation of 1.86, also the Visual Puzzles subtest average is 10.22 with a standard deviation of 2.33.
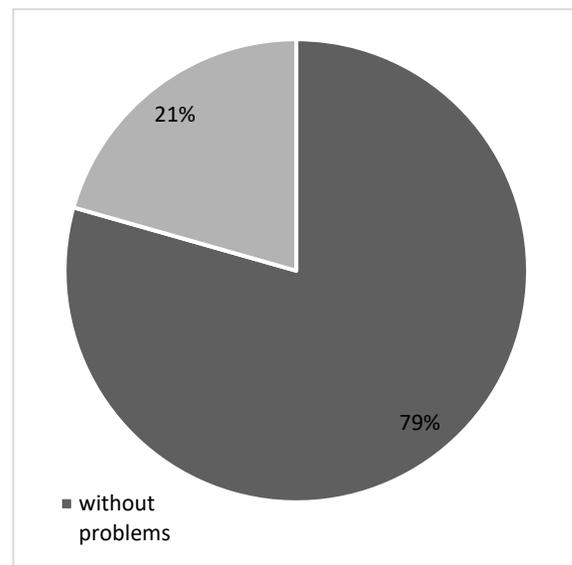
Fig. 1. Percentage of Pupils with Problems among All Participants.

A fifth of the participants indicated in the questionnaire that they suffer from hearing or visual disorders or both Fig. 1. The percentage of girl's pupils and boys' pupils participating in this

study is equal, but the percentage of girls who report suffering from a hearing or visual disorders or both is 22%, and 78% do not suffer from any disorders.
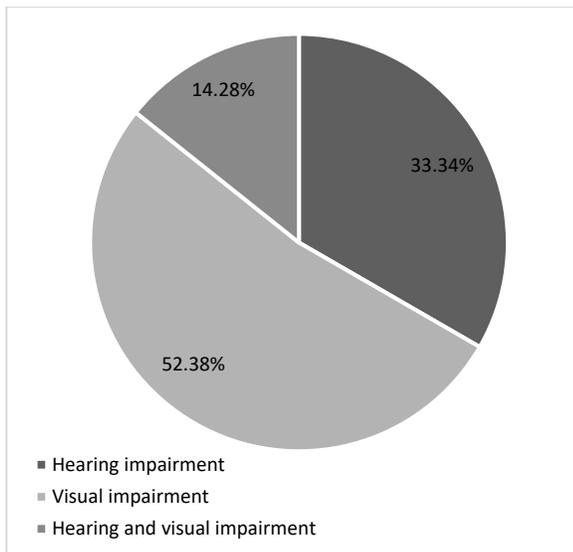
Fig. 2. Percentage of Pupils with Hearing or Visual Disorders or Both.

Fig. 2 shows the percentage of pupils with hearing or visual disorders or both. A half of pupils who have disorders confirmed having visual disorders. While a third of pupil's report having a hearing disorder. The lowest percentage are pupils who have hearing and visual disorders.

According to gender, the large percentage of female who suffer from problems 45.45%, and this percentage concerns female who have visual problems, followed by a percentage of 36.36% of female who report that they have visual problems, hearing problems; the low percentage of 18.18% is among female reported that they have hearing and visual problems. A large percentage of male who have problems is 60%, is linked to male who have vision problems, with a percentage of 30% of male who suffer from hearing problems, the 10% of male reported that they have hearing and visual problems.

TABLE I. CORRELATIONS BETWEEN VISUOSPATIAL INTELLIGENCE SCORES AND DEVELOPMENTAL AND COGNITIVE IMPAIRMENT FACTORS

Factor	r	p
The quality of family earnings	-0,240	0,015
The quality of the family's level of education	0,244	0,013
Submission to the preschool period	0,167	0,093
Got at least one phone in the family	0,326	0,001
Got at least one computer in the family	0,474	0,000
Participate in e-learning (the Covid-19 pandemic)	0,540	0,000
Desire for E-learning	0,339	0,000

The correlation between visuospatial intelligence scores and participation in e-learning (Covid-19) is the highest according to the Table I which presents the correlation between visuospatial intelligence scores and the seven factors. The high correlation is followed by the correlation between the availability of electronic communication tools such as computer and telephone and visuospatial intelligence. The desire for e-learning is also great. Social and educational factors are also important.

IV. DISCUSSION

Hybrid learning remains a promising technological approach that reduced the existing difference in a disrupted world, the covid 19 pandemic brought different constraints and changed the voices maintained for normal life. However, the success of the E-Learning process depends on a set of skills and cognitive abilities of the learners. Visuospatial intelligence is no longer considered one of the components of the intelligence quotient of each individual but also an essential component to complete one's learning. The study indicates that the results are equal between the female and male gender, this result explains that the programs don't distinguish between two genders in terms of preparation, and that the teachers treat their pupils equally regardless of their gender, and this treatment is reflected in their performance in examinations and general tests [20][21].

In general, the average theoretical visuospatial intelligence is one hundred. The results of this study indicate that the participants received a composite average score equal to the theoretical average of 100.43, it's superior to all intelligence studies in Morocco [22]. This result is between M=99.1 and M=105.6 in the study which conducted in America twice over a same sample belonging to the same age group as the study sample [23]. The average result derived from the theoretical standard and the results of various studies conducted in the United States is concord with the results obtained. The situation of 1,607 Caucasian child with 830 girls and 777 boys aged 6 to 16, with a Visuospatial of 103.49 and a standard deviation of 14.46 (girls with M=102.97 and SD= 13.85; boys with M=104.05 and SD=15.12). Two other studies reveal average inferior. The first concerns 409 African - American child: 221 girls and 188 boys, aged 6 to 16, with a Visuospatial of 90.69 and a standard deviation of 13.71 (girls with M=90.95 and SD=13.67; boys with M= 90.39 SD = 13.76). The second study concerns 621 Hispanic child: 313 girls and 308 boys, from 6 to 16 years old, with a Visuospatial of 97.17 and a standard deviation of 13.41 (girls with M = 95.83 SD = 12.75; boys with M = 98.54 and SD = 14.09) [24]. The difference in study results can be explained by the nature of the development of minds and the use of imagination while they ponder, analyse and reflect on school subjects, in addition to the nature of schools and cultural level and social of parents.

Despite cultural differences and demographic characteristics, the results obtained in our study show similarities with most studies. It's important to emphasize that the samples of American studies serving as a point of comparison with our study, differ from the samples of our study from the point of view of the educational composition of the parents (holders of a bachelor's degree or more), which very logically leads, a set of social and economic privileges for their children [25].

The Moroccan pupils are science pupils, which suggests why they excelled in the test which is due to family orientation and continuous educational development since childhood, unlike pupils who tend to acquire languages, who perform less well on tests.

Suggestion of interventions in the curriculum is necessary to become more adapt to pupil, through the development of

curricula that contain activities that develop visuospatial ability with the diversification of teaching strategies to include both visuospatial methods, and not limited to traditional methods of presenting subject. Using visuospatial skills help pupils find multiple solutions to the same problem during their studies which leads to an improvement in their level of creative thinking and achievement.

The results of the present study indicate that there is a correlation between the development of visuospatial intelligence of participants, and the rate of their participation in E-learning. Also, the acquired spatial visual skills reflected in academic achievement and contribute to the acquisition of learning skills at an early age, have relatively eliminated the comparison between the social and economic privileges of the Moroccan and American learner.

V. CONCLUSION

Based on the results of the research, where spatial ability isn't affected by gender, it can be concluded that public school pupils who excel in the physics course have a high level of spatial ability and that they have a higher level in demonstrating their spatial capability. This is due to the development of their abilities due to the early experiences they are exposed to inside and outside of school, which affects academic achievement. It also shows the strength of the correlation between E-learning and spatial ability, which is reflected in the quality of nature of instructional orientation for this sample, as well as fat status in performance and academic achievement based on e-learning.

VI. RECOMMENDATION

The pupils participating in this study have very good academic performance. Didactic aids to increase and develop the visuospatial skills of all pupils may include providing activities to support visuospatial strengths, activities that build creative structures, provide many educational digital games, and support visuospatial strengths, provide visuospatial activities; encourage the child to engage in visual-spatial tasks, such as putting together puzzles; creating maps; drawing or playing with construction toys. Provide visual learning opportunities to help understand and remember new ideas. When new information is presented in the classroom, visual aids should be presented to supplement the content presented orally. This means decreasing or increasing the number of visual displays with manipulatives, drawings, diagrams, and charts may overwhelm the pupil. It is necessary to explain in words, all new skills and concepts, graphics and visual tasks.

REFERENCES

[1] Schank, D. (1991). Learning Theories: An educational perspective. New Jersey: Merrill.

[2] Galean, B. -C. (1988). Mind Sight: Learning Through Imaging, Center for Integrative Learning. Berkeley, California, USA: 3rd printing.

[3] Chan, Da. (2007). Gender Differences in Spatial Ability: Relationship to Spatial Experience Among Chinese Gifted Students in Hong Kong. Roeper Review, 29, 4pp.

[4] Carroll, J. B. (2005). The three- stratum theory of cognitive abilities. Dans D. P. (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp. 2nd ed., pp. 69–76). New York: Guilford.

[5] Garderen, D. (2006). Spatial visualization, visual imagery, and mathematical problem solving of students with varying abilities. Journal of Learning Disabilities, 39(6), 496-506.

[6] Yang, E. M., Andre, T., Greenbowe, T. J., & Tibell, L. (2003). Spatial ability and the impact of visualization/animation on learning electrochemistry. International Journal of Science Education, 25(3), 329-349.

[7] Munroe, R. H., Munroe, R. L., & Brasher, A. (1985). Precursors of spatial ability: A longitudinal study among the Logoli of Kenya. The Journal of social psychology, 125 (1), 23-33.

[8] Bickley-Green, C. A. (1995). Math and art curriculum integration: A post-modern foundation. Studies in Art Education, 6-18.

[9] Gardner, H. (2006). The development and education of the mind: The collected works of Howard Gardner. London: Routledge.

[10] Stix, A. (1995). The Link between Art and Mathematics, paper presentedat the Annual Conference of the National Middle School Association. Cincinnati: OH.

[11] Flanagan, D. P., & Kaufman, A. S. (2009). Essentials of WISC-IV assessment (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[12] Flanagan, D., & Kaufman, A. (2004). Essentials of WISC-IV assessment. New Jersey: John Wiley & Sons Inc.

[13] Kaufman, A. S., & Lichtenberger, E. O. (2006). Assessing adolescent and adult intelligence (3rd ed.). Hoboken, NJ: John Wiley & Sons.

[14] Raiford, S. E., & Coalson, D. L. (2014). Essentials of psychological assessment. Dans S. E. (Eds.), Essentials of WPPSI™-IV assessment (pp. p. 1–43). John Wiley & Sons Inc.

[15] Ghazlane, I., Marnoufi, K., Daaif, J., & Touri, B. (2022). The Relationship between Critical Thinking Skills, Portfolio Models and Academic Achievement of Moroccan Midwifery Students. Journal of Educational and Social Research, 12. 20.

[16] Marnoufi, K., Touri, B., Bergadi, M., Ghazlane, I., (2020). Intelligence Analysis among Rural Learners in Morocco. Dans Psychology Applications & Developments VI (pp. 367-375). inScience Press.

[17] Marnoufi, K., Touri, B., Bergadi, M., Ghazlane, I., & Soubhi, Fz., (2022). Evaluation of Intelligence Scores among Students From Moroccan Urban Areas. Dans Psychology Applications & Developments VI (pp. 359, 366). inScience Press.

[18] Wechsler, D. (2016a). WISC–V. Echelle d'intelligence de Wechsler pour enfants-5e édition, manual d'interprétation. Paris, France: Pearson France-ECPA.

[19] Wechsler, D. (2016b). WISC–V. Echelle d'intelligence de Wechsler pour enfants-5e édition, manual d'interprétation. Paris, France: Pearson France-ECPA.

[20] Weiss, L. G., Locke, V., Pan, T., Harris, J. G., Saklofske, D, H., & Prifitera, A. (2016). WISC-V Use in Societal Context. Dans L. G. Weiss, WISC-V Assessment and Interpretation Scientist–Practitioner Perspectives (pp. 123-185). Elsevier Inc.

[21] Weiss, L. G., Saklofske, D, H., Holdnack, J. A., Prifitera, A. (2016). WISC-V Assessment and Interpretation: Scientist–Practitioner Perspectives. Academic Press.

[22] Lynn, R., Vanhanen, T. (2012). intelligence: a unifying construct for the social sciences. Ulster Institute for Social Research.

[23] Wechsler, D., Raiford, S. E., Holdnack, J. A. (2014). Technical and Interpretive Manual Supplement: Special Group Validity Studies with Other Measures and Additional Tables. Bloomington, MN: NCS Pearson.

[24] Scheiber, C. (2016). Is the Cattell–Horn–Carroll-Based Factor Structure of the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V) Construct Invariant for a Representative Sample of African – American, Hispanic, and Caucasian Male and Female Students Ages 6 to 16 Ye. J Pediatr Neuropsychol, 2, 79–88. doi:10.1007/s40817-016-0019-7.

[25] Sattler, J. M. (2008a). Assessment of children: Cognitive foundations (5th ed.). San Diego, CA: Author.

Mobile Devices Supporting People with Special Needs

Tihomir Stefanov, Silviya Varbanova, Milena Stefanova
Department of Mathematics and Informatics
University of Veliko Tarnovo
Veliko Tarnovo, Bulgaria

Abstract—**Over the years, various devices designed for people with special needs have been used for a while and then replaced with modern devices to make everyday life easier. The development of mobile devices is especially improved today and through them different everyday activities are facilitated, not only by people with special needs. The purpose of this paper is to present some of the modern mobile devices with an analysis of their operating systems, functionalities, applications and design. Based on the research, their usability for both sighted and visually and hearing-impaired users is described. Attention is paid to the preferences formed among users when using specialized applications developed for mobile devices. Based on a survey of specific target user groups, the paper provides summary results to support the thesis on the importance of the facilities, offered by modern mobile devices.**

Keywords—Mobile devices; mobile operating systems; Android; iOS; special needs; visually impaired; hearing impaired; e-learning

I. INTRODUCTION

The last few years have seen increased use of mobile devices by users of all ages to accomplish a variety of tasks [2]. In the recent past, their main purpose has been solely for communication. Today, however, using smartphones and tablets, people around the world not only communicate with each other, but work and play by sending and receiving information of all kinds, quickly and easily. It is not yet possible to draw an accurate conclusion as to what percentage of users have completely replaced PCs and laptops with smartphones, and in some professional fields, this is already a fact.

Nowadays, there is an increased interest of people of different nationalities and with diverse health status in professions related to programming and IT. This has contributed to the updating of curricula at universities to include new courses in computer science. The use of computer techniques is particularly important for the laboratory lessons in these subjects. In a number of educational establishments, there are also some students with special needs, which determines the implementation of modern ways and devices in order to provide better and more effective learning. However, not all teachers receive specialized training in teaching the blind or hearing impaired, and the implementation of new ways and the integration of a variety of devices would support the educational process.

The study aims to observe the learning process in leading computer courses for students with visual or hearing impairments. Learning, when supported by the use of smartphones and tablets provides alternative solutions for the hearing impaired or visually impaired – such as special educational tools within a real learning environment or a learning resource center [12].

In our country over the past decade, there has been an increased use of mobile devices and apps among people with special needs, including those who are blind or hard of hearing. The size and weight of the devices are of particular importance for these user groups, especially in daily carrying and use anytime and anywhere possible.

Comparing today's devices with those of ten years ago, there has been a significant change in the external design and the replacement push-button keypads with touch screen displays, which is both a convenience and a drawback for some user groups [13]. Older users have used button keyboards in the past, and touch screens today. Younger users have grown up with mainly modern devices with touch functions

II. RELATED WORK

It took less than a century to advance technology in the areas of hardware and software, and to replace John Atanasoff's large-scale computer with Steve Jobs' smartphone. Over the years, the Internet has established itself as the primary medium for searching, storing and sending data. Users visit leading websites and social networks on a daily basis. The World Wide Web has become popular and has encompassed users' lives completely, not only in their work but also in their leisure time.

This has led to the development, testing and optimization of various applications for computer and mobile devices by specialists in computer companies. The demand for laptops and computers designed for different activities – for office or home, professional, gaming, multimedia has increased. However, there is a preference for the use of smartphones or tablets among 'different' users, namely those with special needs.

According to the World Health Organization, nearly 2.2 million people are visually impaired worldwide. It is particularly important for them to navigate and move around safely, both in open and in unfamiliar indoor spaces [8]. Navigating corridors, sidewalks, stairways, and narrow walkways makes it difficult for the blind. After years of research, navigation aids and systems for mobile devices such as GPS systems have been developed to replace 3D printed maps with protruding dots or magnetic boards [10].

With the development and deployment of voice assistants such as Siri for Apple, Alexa for Amazon, Cortana for Microsoft, Assistant for Google and others, visually impaired

users are greatly facilitated. They now have the ability to do a variety of activities not only through computers, but also through smartphones. Especially when searching for literary resources. In this regard, libraries in educational institutions and localities could provide library services and materials through specialized applications for mobile devices, in order to facilitate their use by the blind [6].

Regarding reading and writing texts, blind people can use Braille in several different ways: with a Braille typewriter or tablet, via a computer or a mobile device [9]. Applying Braille to read and understand text has been the only traditional way for the blind to receive information in the recent past [18]. According to Radulov, however, learning it was an easier process than applying it [16]. Today, braille keyboards and braille tablets greatly facilitate the use of devices as priority tools of accessing needed information [14].

The application of innovative technologies is important for the quality of the learning process among general groups, which include both able-bodied and students with special needs. Usually students with problems are given fewer assignments as they fail to cope and slow down the learning process. Specialized applications have been developed to help them and the teachers to facilitate the learning process of the tasks [3]. The Scrum methodology is one of the most suitable ways to implement software development management, as the purpose of this methodology is to work together and adapt to changes that occur within the project. The application developed within the research of Boza-Chua, Gabriel-Gonzales and Andrade-Arenas, targets students with visual impairment in order to improve the quality of teaching for student teachers [3].

In recent years, due to various reasons, the number of people with hearing problems in the World has been increasing. Currently, there are about 120,000 people with hearing impairment in Bulgaria. In the future, it is expected that one in ten will have some degree of hearing impairment. Everyone with a hearing impairment also has the right to be perceived as healthy. They should be given the opportunity to be educated and to work regardless of the difficulties they face. To communicate with each other, these users use a special language called sign language based on the dactylic alphabet. There are around 25,000 sign language speakers in Bulgaria.

Surveys conducted with users on a predefined or random topic in conversation, and with a trained associate have found that users who made the call without a mobile device were more reliable. However, for people with special needs, the results were different – it was found that through the mobile device they perceived the information better due to mimic gestures [7].

One of the main requirements for admission of students to the University is to pass an entrance examination. However, many higher education institutions do not have the facilities to conduct entrance examinations for visually or hearing impaired people. The need for tools to enable them to perform optimally in the entrance examination are important. To this end, the implementation of web applications that could provide online application capabilities is needed. These applications could

also be beneficial for any of the ongoing examinations during the training [4].

The provision of resource support and out-of-class and learning activities through mobile devices is also supportive of the hearing impaired and is widely applicable in Asia. With the help of resource teachers, pupils and students with hearing impairment are motivated to participate in the learning process and activity. Last but not least, the assistance provided by a parent or relative in the home environment is also of particular importance for improving learning efficiency [17].

In the learning process, the hearing impaired people need proper communication regardless of their age, social status and nationality. Specialized applications are being developed to recognize terms from different languages. The added learning features in these applications are designed to intensively train users in the use of sign language. This can be achieved by voice recognition of words or by typing the words in the selected spoken language. Next, the application displays the appropriate written language images. The research of Setiawardhana, Hakkun and Baharuddin is aimed at developing an Android-based application that directly interprets the sign language, provided by the sign language, into written speech. The translation process starts from hand detection using Opencv and translation of hand signals with K-NN classification [17].

In life, these users often use social networks and video conferencing to communicate with each other. They use the language of the deaf to transmit and perceive information. For convenience, through mobile phones and the Internet, they can exchange information anywhere, anytime.

III. ANALYSIS OF POPULAR MOBILE OPERATING SYSTEMS

New phone models developed in recent years feature not only convenient designs, but also improved versions of both operating systems and software applications. This has led to an increase in users replacing desktops and laptops with modern phones. Improvements in new smartphones are directly aimed at helping people with special needs, and many apps have been developed to help these users in due course.

Just ten years ago, the fact that blind people could use computers, tablets and smartphones was not accepted. Today, visual impairments are no barrier to this, given that blind people perform important work, learning and mobility tasks using their mobile phones. The built-in accessibility features available in smartphones make the process much easier [15]. For example, for getting around, reading emails, online shopping, banking or booking, listening to music. Screen readers such as VoiceOver, TalkBack are also used for the services listed. Blind people can also perceive images accompanied by descriptions through them. Can also successfully use speech-to-text functions when sending messages [5].

Years ago, the popular mobile operating systems were Symbian, Palm, Windows Phone and BlackBerry. The leading operating systems current in modern mobile devices are Android, iOS, and Harmony OS. In this report, only the capabilities of the first two will be discussed, due to the minimal spread of the Harmony OS in Bulgaria at the moment.

In modern smartphones, the 'Accessibility' mode includes options related to screen readers and display settings such as screen magnification, pronunciation of on-screen elements, voice reading, volume up. These features are particularly useful for people with special needs.

A. Features of iOS 16

In iOS, facial recognition of people is implemented through the built-in camera application, which is also helpful for visually impaired users. iOS was originally introduced in 2007, with built-in features to help people with special needs available with the very first versions. Although it supports as many popular languages such as English, Spanish, German, for many years, it has not offered a Bulgarian version.

Localization for Bulgaria was first introduced on September 7, 2022, and the new iOS-16 download became available to users on September 12, 2022 [1]. This version is backwards compatible with all Apple phones – since the iPhone 8, which was introduced in September 2017. This allows people with special needs to take advantage of the opportunity provided without needing devices in the high price range.

In Fig. 1, functionalities of OS are shown in separate groups – for the hearing impaired, visually impaired and physically limited, respectively. The 'VoiceOver' feature speaks out items on the screen and provides helpful hints, where it can also be controlled using simple gestures. On initial activation of this feature, a wealth of detailed information is presented which is useful to the disabled user and their social assistant. Customization includes the ability to set the voice, pronunciation, tone and speed of utterance from the device. For convenience, the additional option of recognizing the phone owner's voice is also provided. The mobile OS also offers Bluetooth connectivity to Braille devices.

Fig. 1.   Functionalities of iOS 6 for People with Special Needs.

Partially sight impaired users are provided by: the ability to zoom the screen, increase text size, adjust color schemes, negative and black and white mode. Hearing impaired people offer the possibility to connect with hearing aids via Bluetooth and a sound recognition option. In the 'Audio/Visual' menu, settings for Mono Audio, Notifications, 'LED Flash for Alerts'

and 'warnings' are collated. In addition, the smartphone supports and adds subtitles and captions for user ease.

The Physical and Motor section includes 'Touch', 'Face ID & Attention', 'Switch Control', 'Voice Control', 'Control Nearby Device' and 'Apple TV Remote'. This section adds options for specific keyboard and side button settings on the device. With all the above-mentioned conveniences of iOS 16 for users, the use of Siri voice assistant can be added, which, however, does not yet support the Bulgarian language.

B. Features of Android 12

The first version of Android was introduced at the end of 2007, with the first T-Mobile G1 (HTC Dream) device running it being introduced in October 2008. In just a few years, Android became the most mainstream OS for mobile devices [23]. Unlike devices at the high end of the price spectrum running iOS, Android is affordable for users with minimal financial means. Another advantage of the Android mobile OS, besides being free, is that every smartphone manufacturer is licensed to use it. The first version of the operating system localized in Bulgarian was Android 2.3 Gingerbread, introduced in 2010.

The menu for people with special needs is located in the 'Accessibility' section (Fig. 2.). It includes 'Recommended for You', 'Visibility Enhancement' and 'Hearing Enhancement', 'Interaction and dexterity', 'TalkBack'.

The TalkBack feature provides speech feedback so the device can be used without looking at the screen. Apart from this feature, visibility enhancement settings are available in Android, which include high contrast theme, fonts, button coloring, animation removal, color inversion.
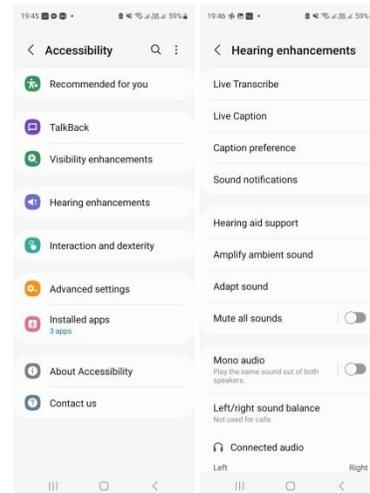
Fig. 2.   The Menu in Android 12 for People with Special Needs.

For the hearing impaired, there are personalization options that cover quickly "finding the relevant written symbols of speech sounds", sound notifications, hearing aid support, amplification of external noise, muting of all sounds, left/right sound balance. Unlike iOS, no option has been added in Android to adapt the device to recognize the user's voice. The Physical Restrictions and Movements section houses options for voice control, answering and ending calls, touch duration,

and more. English Android users are facilitated with Google's voice assistant, but only when setting commands in their native language. Full localization of Google Assistant in Bulgarian is not yet available. In addition to the capabilities of the mobile OS, smartphones extend their functions for people with disabilities through dozens of specialized applications. A more extensive exploration of their capabilities would be a topic for future work.

IV. MOBILE APPLICATIONS TO SUPPORT STUDENTS IN CODE DEVELOPMENT

The use of mobile devices to carry out various daily tasks is almost inevitable in today's information society. Integrating modern technologies and devices into the learning process of students and pupils is important. The inclusion of mobile devices as an aid to learning also greatly facilitates people with special needs, especially those who are hearing or visually impaired. Thanks to this, access to educational resources is becoming a reality for an increasing number of such learners [12]. Recent years have seen increased use of the Android operating system in combination with specialized applications for creating programming code and the ability for beginners to write source code in popular programming and markup languages, scripting languages, and layout.

Leading subjects such as Computer Science, Information Technology, Web Technology, Internet Programming in Higher Education and Schools are categorized as complex. The formation of algorithmic thinking and code design and implementation habits by learners, especially those with special needs, requires time, guidance and persistence. To provide the ability to input the code in markup languages, scripting languages, layout styles such as HTML, JavaScript, CSS, PHP and other technologies, specialized text editors are needed for both desktop and mobile devices. In terms of functionality, they have options for quick debugging, automatically adding code or working with more than one file at a time, and existing customization options in terms of themes. The main task of students is to adapt to using these editors as a medium for writing and reading code. Only then, they could independently implement codes according to predefined criteria.

As a practical experience, it can start with easier tasks involving tagging, basic elements such as headings, paragraphs, images, links, buttons, forms, and more. Moving on to the layout of the elements using styles and defining their dynamics. The idea is to gradually build practical skills in using popular text editors for mobile and desktop devices.

During classes, a common practice in recent years is for students to use their personal mobile devices to write programming code as well. Screens of the two of the mobile applications used for this purpose are presented on Fig. 3.

Fig. 3.   Screens of Web Master's HTML Editor Lite and Droid Edit.

Some of the popular mobile applications for writing programming code are:

- *Web Master's HTML editor Lite* – popular and easy to use editor for Android. Provides the ability to write codes in HTML, CSS, PHP, and JavaScript [22].

- *Droid Edit* – intended for Android analogue of the NotePad++ editor for Windows supporting not only HTML, PHP, CSS, JavaScript, but also C++, C#, Java, Python [20].

- *AnWriter* – widely applicable editor, shown on Fig. 4. with the ability to enter codes in jQuery, Angular, Bootstrap, SQL, LateX [19].
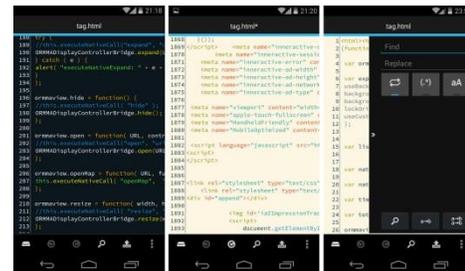
Fig. 4.   AnWriter Mobile App Screens.

- *Quoda CODE Editor* – specialized editor (Fig. 5.) for programming languages such as Haskell, Lisp, HTML, C#, C++, Ruby, Python, SQL, XML [21, 22].
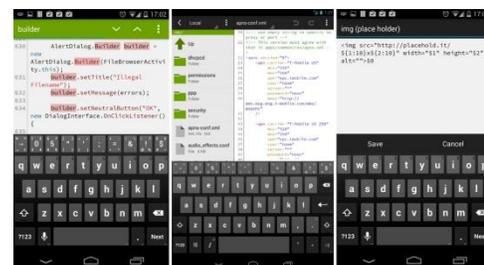
Fig. 5.   Quoda CODE Editor, Supporting Functional Programming Too.

V. MOBILE DEVICES AND THEIR ROLE IN EDUCATION

Every blind person has the right to adequate education. However, not all blind people in Bulgaria and around the world have the opportunity to complete their studies. One important factor is the need for appropriate technological tools. Based on studies carried out involving both blind students from Latin America and their parent-assistants, applications for mobile devices have been developed to improve their social life and promote their education. The study points to the design of applications for developing computer games, even by the blind themselves, as a good practice [3].

Until recently, it was impossible to use the phones in educational institutions. In recent years, however, there have been cardinal changes that have gradually necessitated the use of mobile devices by both teachers and learners [2]. The need for quick access to electronic resources has led to their increased use.

During the pandemic of Covid-19, e-learning established itself as an alternative form of attendance in many countries around the world. Everyone involved in online learning – teachers, students, pupils, even parents – had to work in virtual classrooms, which took considerable time to adapt to. Through mobile devices, all learners are greatly facilitated – including those with special needs – when they need to access electronic resources for reading, editing, sending and archiving theoretical information or practical assignments [11].

Schools and universities have made a rapid transition to electronic platforms and this has resulted in the need for students to have the appropriate computer equipment and devices for greater ease of communication. Gradually, their use has become established in almost every discipline. A variety of disciplines focusing on different areas are being taught in leading universities in the country and often programming, web design, multimedia, giving basic knowledge in the IT field are more preferred by the learners.

In recent years, more and more people with special needs are completing their higher education. The fact that mobile devices are finding application in the learning process, which facilitates the perception of information by these groups of users, contributes to this. In the Faculty of Mathematics and Informatics at the St Cyril and St Methodius University of Veliko Tarnovo, the leading specialties in the Bachelor's degree are Computer Science, Informatics, and Software Engineering. Students of these specialties receive the necessary knowledge from qualified university lecturers and experts from computer companies. The faculty also trains students with special needs who successfully complete their education, regardless of whether they are hearing impaired or visually impaired.

The problems encountered during the E-learning of students with hearing impairments were mainly related to the perception of voice information. On the other hand, the provision of electronic resources by lecturers in the form of text files and multimedia presentations contributed to overcoming such difficulties.

VI. MOBILE DEVICES AND THEIR USE IN ACCOMPLISHING SPECIFIC TASKS

For the implementation of specific tasks, the considered editors for smartphones can be used by the visually and hearing impaired without significant difficulties.

For the purpose of the study, in some of the laboratory classes of the leading courses in the Faculty of Mathematics and Informatics, students were given three sample problems with the aim of deriving results related to the topic of this paper, which are summarized in Table I. Both able-bodied students and students with visual and hearing impairments were included as user groups.

The following tasks, described below, were set to be completed within two to four class periods.

First Task: to implement a web page (portfolio) including personal information about a student using the web technologies, scripting languages, markup languages and layout using a personal mobile device and an online text editor. Web page to include elements such as images, headings, paragraphs, buttons, navigation, forms.

The time to complete the assignment is two class periods. As a result, the number of students who completed the task according to the criteria set for its implementation is summarized. The assignment is provided for a group of 10 to 15 students. In completing this assignment, students use a text editor for HTML and CSS of their choice. Some of the more famous editors for Android are presented next section of the report.

In the 'UX/UI Design' classes, students develop prototypes using modern web technologies such as HTML 5.0, CSS 3, Java Script, TypeScript, Bootstrap 4, JQuery, React, AngularJS, and others. The applications being developed are for educational purposes and are not yet available online. One of the conceptual projects is a website for selling mobile devices for all users, including blind and hearing impaired people. The project is designed for both desktop screens and mobile devices (Fig. 6).
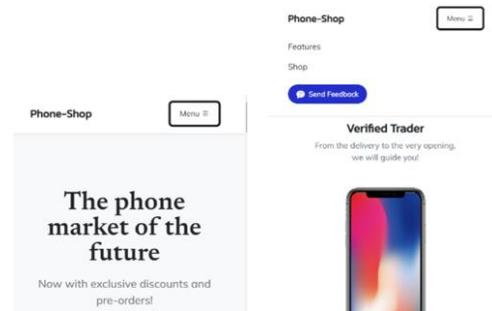
Fig. 6.   Design of the Developed Website for Mobile Devices.

The student integrated an embedded audio file reader into the project homepage, the code for which is presented in Fig. 7. A user-friendly and easily reachable navigation of the website has been achieved, which also offers a contact form.

```
<audio id="features" src="./audio/features.mp3"></audio>
<audio id="intuitive-audio" src="./audio/intuitive-solution.mp3"></audio>
<audio id="jump-shop-audio" src="./audio/jump-to-shop.mp3"></audio>
<audio id="phone-market-audio" src="./audio/phone-market-of-the-future.wav"></audio>
<audio id="phone-shop-audio" src="./audio/phone-shop.wav"></audio>
<audio id="send-feedback-audio" src="./audio/send-feedback.wav"></audio>
<audio id="shop-button-audio" src="./audio/shop.mp3"></audio>
<audio id="stop-waiting-audio" src="./audio/stop-waiting.mp3"></audio>
<audio id="wide-variety-audio" src="./audio/wide-variety.mp3"></audio>
<audio id="exclusive-audio" src="./audio/exclusive.mp3"></audio>
<audio id="mobile-app-audio" src="./audio/mobile-app.mp3"></audio>
<audio id="new-age-audio" src="./audio/new-age.mp3"></audio>
<audio id="devices" src="./audio/devices.mp3"></audio>
<audio id="flexible-use" src="./audio/flexible-use.mp3"></audio>
<audio id="referal" src="./audio/referal.mp3"></audio>
<audio id="verified-trader" src="./audio/verified-trader.mp3"></audio>
```

Fig. 7.   Part of the Source Code of the Created Audio File Reader.

Another of the student projects is a web-based quiz application designed for blind users and implemented for both desktop systems and mobile devices (Fig. 8.).
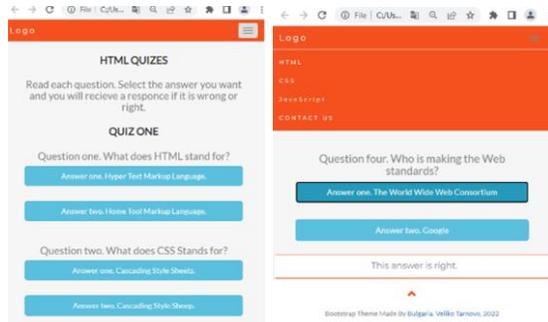
Fig. 8.   Screenshots from the Mobile Version of the Developed Application.

In the tests made, blind users easily and quickly adapted to the information presented in the application. When viewed with the Google Chrome browser and the Screen Reader turned on, the reader tells out the text part and thus made it easier for blind users to get the content of each page. When viewed with browsers such as Opera, Avast Secure Browser, Firefox, a pre-configured NVDA or JAWS voice reader is required. The template project is a Web Technology related quiz and includes two key questions and the option for two answers. When one of the answers is selected, its correctness is displayed at the bottom of the page. Two functions implemented in JavaScript verify the received answers (Fig. 9.).

```
<script>
function rightAnswer() {
  document.getElementById("foo").innerHTML = "This answer is right.";
  document.getElementById("foo").focus();
}
function wrongAnswer() {
  document.getElementById("foo").innerHTML = "This answer is wrong.";
  document.getElementById("foo").focus();
}
</script>
```

Fig. 9.   The Function Code is Implemented using JavaScript.

Task Two: Each student should familiarize themselves with the curriculum of a given discipline through the E-student system. For this purpose, a personal mobile device should be used.

The guiding criterion for the task is to apply the UX/UI design principle for the number of clicks (up to three times).

As a result of the task performance, the number of students who successfully comply with the set principle is counted – on the first and second visit to the system, respectively. The task is assigned in groups of 10 to 15 students. The number of students who completed the assignment according to the guidelines is reported as the result.

Task Three: Implement an online market research and simulation of a product or service purchase, optionally via a mobile device or computer, using existing websites.

The task is set in a group of 10 to 15 students. Students who completed the task using the pre-set criteria of using a mobile device, laptop or desktop system are scored. Students' opinion on which device they would prefer in a repeat market research for online purchase is recorded, with 72.73% choosing their personal mobile device over a laptop or desktop system.

A summary of the results is presented in the following Table I.

TABLE I.   TASKS ASSIGNED COMPLETED BY STUDENTS

Student's results	Task 1	Task 2		Task 3	
	One solution	*1st option*	*2nd option*	*1st option*	*2nd option*
Users with special needs	3	1	2	3	2
Regular users	47	96+36	125	117	98

VII.   CONCLUSION

The idea of this report is not only to promote modern devices, their functionalities, purposes and applications, but also to present the opportunities for some specific groups of users to use them as an aid. In everyday life, education, science and work, people with special needs experience communication difficulties in perceiving and transmitting information that should not be a barrier to their current and future development.

The design of assistive applications for mobile phones is the subject of research, analysis and testing, and implementation in the future, in order to support visually impaired users. People with hearing impairments find it difficult to perceive 'talking' information such as music, news, and multimedia information. However, the perception of textual information in the form of emails, subtitles and messages facilitate communication and data perception.

Screen readers have been developed for the visually impaired to make it easier for them to use mobile devices. Blind people cannot perceive visual information; rather they 'listen fast'. For the hearing impaired, the perception of 'talking' information is partial or impossible. They adapt to perceiving visual information. Regardless of whether the disorder is congenital or acquired, it results in impaired motor, visual and auditory functions.

According to Stephen Hawking, the ability of people to integrate to changes is important. This applies in full force to all people with and without special needs. As the main task of devices is to serve them.

REFERENCES

[1]   Apple Press Release. Apple unveils an all-new Lock Screen experience and new ways to share and communicate in iOS 16, June 2022. https://www.apple.com/bg/newsroom/2022/06/apple-unveils-new-ways-to-share-and-communicate-in-ios-16/. Accessed Oct 2022.

[2]   Asylah Suwadi, N., Aini Abd Majid, N., Chun Lam, M., Hashimah Jalaluddin, N., Aisyah Abdullah, A., Husairi Hussain, A., Ahmad, A., Zairi Maarof, D., Kasdan, J. (2021). Expert Review on Mobile

Augmented Reality Applications for Language Learning, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. 10, 2021. http://dx.doi.org/10.14569/IJACSA.2021.0121037.

[3] Boza-Chua, A., Gabriel-Gonzales, K., Andrade-Arenas, L. (2021). Inclusive Education: Implementation of a Mobile Application for Blind Students. In: International Journal of Advanced Computer Science and Applications, West Yorkshire Vol. 12, Iss. 11, 2021. DOI: 10.14569/IJACSA.2021.0121189.

[4] Carrion-Silva, A., Diaz-Nuñez, C., Andrade-Arenas, L, (2020). Admission Exam Web Application Prototype for Blind People at the University of Sciences and Humanities, International Journal of Advanced Computer Science and Applications (IJACSA), 11(12), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0111246.

[5] Holly. How do blind and visually impaired people use a mobile phone?, February 2019. https://lifeofablindgirl.com/2019/02/03/how-do-blind-and-visually-impaired-people-use-a-mobile-phone/. Accessed Oct 2022.

[6] Hoy, M.B. (2018). Alexa, Siri, Cortana, and more: An introduction to voice assistants. Medical Reference Services Quarterly, 37: 1, 81-88, DOI: 10.1080/02763869.2018.1404391.

[7] Ismaili, J., Ibrahimi, E.H.O. Mobile learning as alternative to assistive technology devices for special needs students. Educ Inf Technol 22, 883–899 (2017). https://doi.org/10.1007/s10639-015-9462-9.

[8] Jabnoun, H., Hashish, M.A., Benzarti, F. (2020). Mobile Assistive Application for Blind People in Indoor Navigation. In: Jmaiel, M., Mokhtari, M., Abdulrazak, B., Aloulou, H., Kallel, S. (eds) The Impact of Digital Technologies on Public Health in Developed and Developing Countries. ICOST 2020. Lecture Notes in Computer Science (), vol 12157. Springer, Cham. https://doi.org/10.1007/978-3-030-51517-1_36.

[9] Karastoyanov, D., Stoimenov, N. Gyoshev, S. (2019). Methods and Means for Education of People with Visual Impairments, IFAC-PapersOnLine, Volume 52, Issue 25, 2019, Pages 539-542, ISSN 2405-8963, https://doi.org/10.1016/j.ifacol.2019.12.601.

[10] Kuriakose, B., Shrestha, R., Sandnes, F.E. (2022). Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review, IETE Technical Review, 39:1, 3-18, DOI: 10.1080/02564602.2020.1819893.

[11] Lazo-Amado, M., Cueva-Ruiz, L., & Andrade-Arenas, L. (2022). Prototyping a Mobile Application for Children with Dyscalculia in Primary Education using Augmented Reality International Journal of Advanced Computer Science and Applications (IJACSA), 13(10), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0131085.

[12] Liu, C.-C. and Hong, Y.-C. (2007). Providing hearing-impaired students with learning care after classes through smart phones and the GPRS network. British Journal of Educational Technology, 38: 727-741. https://doi.org/10.1111/j.1467-8535.2006.00656.x

[13] Martiniello, N., Eisenbarth, W., Lehane, C., Johnson A., Wittich W. (2022). Exploring the use of smartphones and tablets among people with visual impairments: Are mainstream devices replacing the use of traditional visual aids?, Assistive Technology, 34:1, 34-45, DOI: 10.1080/10400435.2019.1682084.

[14] Misra, S., Cheng, L., Genevie, J., & Yuan, M. (2016). The iPhone Effect: The Quality of In-Person Social Interactions in the Presence of Mobile Devices. Environment and Behavior, 48(2), 275–298. https://doi.org/10.1177/0013916514539755.

[15] Nield, D. All the sensors in your smartphone, and how they work, June 2020. https://gizmodo.com/all-the-sensors-in-your-smartphone-and-how-they-work-1797121002. Accessed Oct 2022.

[16] Radulov, V. History of the pedagogy of the visually impaired. Publishing house 'Phenomen', Sofia, 2017.

[17] Setiawardhana, R. Y. Hakkun and A. Baharuddin, (2015). Sign language learning based on Android for deaf and speech impaired people, 2015 International Electronics Symposium (IES), 2015, pp. 114-117, doi: 10.1109/ELECSYM.2015.7380825.

[18] Stefanov, T., Varbanova, S., Stefanova, M. (2022). An Overview of some popular devices and technologies designed for blind and visually impaired people. In: 2022 International Conference Automatics and Informatics, October 06-08, 2022, Varna, Bulgaria (ICAI'22).

[19] http://www.anwriter.com/.

[20] http://www.droidedit.com/.

[21] https://quoda.en.uptodown.com/android.

[22] https://webmasterlite.en.aptoide.com/app.

[23] https://www.statista.com/topics/876/android/.

Vision based 3D Object Detection using Deep Learning: Methods with Challenges and Applications towards Future Directions

A F M Saifuddin Saif[1]
School of Engineering
Aalto University
Finland

Zainal Rasyid Mahayuddin[2]
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
Selangor, Malaysia

Abstract—For autonomous intelligent systems, 3D object detection can act as a basis for decision making by providing information such as object's size, position and direction to perceive information about surrounding environment. Successful application using robust 3D object detection can hugely impact robotic industry, augmented and virtual reality sectors in the context of Fourth Industrial Revolution (IR4.0). Recently, deep learning has become potential approach for 3D object detection to learn powerful semantic object features for various tasks, i.e., depth map construction, segmentation and classification. As a result, exponential development in the growth of potential methods is observed in recent years. Although, good number of potential efforts have been made to address 3D object detection, a depth and critical review from different viewpoints is still lacking. As a result, comparison among various methods remains challenging which is important to select method for particular application. Based on strong heterogeneity in previous methods, this research aims to alleviate, analyze and systematize related existing research based on challenges and methodologies from different viewpoints to guide future development and evaluation by bridging the gaps using various sensors, i.e., cameras, LiDAR and Pseudo-LiDAR. At first, this research illustrates critical analysis on existing sophisticated methods by identifying six significant key areas based on current scenarios, challenges, and significant problems to be addressed for solution. Next, this research presents strict comprehensive analysis for validating 3D object detection methods based on eight authoritative 3D detection benchmark datasets depending on the size of the datasets and eight validation matrices. Finally, valuable insights of existing challenges are presented for future directions. Overall extensive review proposed in this research can contribute significantly to embark further investigation in multimodal 3D object detection.

Keywords—3D object detection; deep learning; vision; depth map; point cloud

I. INTRODUCTION

3D object detection provides precise representation of objects in the format of semantically meaningful 3D bounding boxes. 3D object detection aims to categorize and localize objects from various sensors data, i.e., monocular and stereo cameras [1, 2], LiDAR point clouds [3], to understand the 3D visual world and associated semantic labels for objects in 3D scenes, has attracted increasing attention from vision community. In addition, advances of deep learning facilitate the rapid progress of 3D object detection indicates strong application demands which can serve numerous applications, i.e., robotics, autonomous driving, augmented reality, virtual reality, robot navigation, enabling systems to understand their environment and react accordingly. As a result, there has been a surge of interest for developing improved 3D object detection pipeline. Although current methods show impressive performance despite the facts that various problems form different viewpoints were observed and illustrated by this research.

This research identified four major approaches along with deep neural networks for 3D object detection, i.e., monocular images, stereo images, LiDAR and Pseudo LiDAR based approaches. 3D object detection from monocular frames is a fertile research area due to potentially vast impact, ubiquity of cameras, low expense, easy implicated solution with one camera [4]. However, estimation of depth from single monocular images is an ill-posed inverse problem causes accuracy of 3D detection from only monocular images is lower than that from LiDAR or stereo images [2, 5]. In addition, loss of significant information during calibration is another reason for degraded performance on the same 3D object detection benchmarks comparing with LiDAR and stereo methods. Advances of deep neural network facilitate immensely the progress of 3D object detection using LiDAR and stereo based methods. The inclusion of depth information allows capturing the three-dimensional structure of the object's environment, is a key feature for ensuring robustness while maintaining high accuracy. Modern LiDAR acquisition sensors provide meaningful information not only for avoiding imminent collisions, but also to perceive the environment as good as image-based data and even surpass it under poor lighting conditions. For stereo images, calibration issues between two camera rigs [6] and occupation of more pixels for nearby objects than far way objects during perspective projection are considered as major challenges. Besides, high cost of LiDAR sensor encourages researchers to look for alternatives such as Pseudo LiDAR based approaches. In this context, Pseudo LiDAR based approaches uses pre-trained depth network to compute an intermediate point cloud representation to mimic LiDAR data and then fed to a 3D detection network [7]. The strength of Pseudo LiDAR based approaches is that they monotonically improve with depth estimation quality although for long distance object Pseudo

LiDAR approaches could not provide expected detection outcome. However, accountability of the depth estimation is the prime gap between Pseudo-LiDAR and LiDAR based 3D object detection. In this context, simpler end-to-end monocular 3D detectors shows strong promise as an option where lack of same scalability from unsupervised pre-training for their one stage nature is the main drawbacks in this context. Besides, existing datasets for validating 3D object detection was not generalized well for different weather conditions and geographical locations, i.e., any method trained on Waymo datasets [8] suffer from dramatic performance drop on KITTI dataset [1]. Therefore, approaches to effectively adaptive 3D object detection method are highly demanded for practical applications where environment varies significantly. In this context, recent success for various 3D object methods mostly depends on larger datasets of 3D scenes where annotations are done carefully and remains as main bottleneck in the context of available datasets.

In summary, the contributions of this work are:

1) Six key areas are identified to analysis existing 3D object detection methods from different viewpoints to guide future development and evaluation by bridging the gaps using various sensors, i.e., cameras, LiDAR and Pseudo- LiDAR.

2) This research demonstrates extensive experimental analysis based on existing research depending on three major aspects, i.e., comprehensive insights on hardware and software, analysis and challenges of using six datasets with details specification, illustration of eight performance metrics required for validation of 3D object detection methods.

3) Based on comprehensive previous research investigation on existing 3D object detection methods, six key observations are elaborated for improving future 3D object detection methods.

II. BACKGROUND

3D object detection methods mostly depend on four types of input pattern, i.e., monocular image, stereo image, LiDAR point cloud and pseudo-LiDAR signal. Research in [9] also used monocular RGB images as input and mapped image features into an orthographic 3D space by describing deep learning architecture for estimating 3D bounding boxes. They performed front-end feature extractor to extract features from monocular RGB images to form 3D features space and the transformation from 2D features maps to 3D features map was termed as Orthographic Feature Transform (OFT). However, validation on large variations of scales and distances could provide real time usage of their proposed method. Research in [10] used monocular RGB images as input to predict 3D human locations by learning data ambiguity without supervision which leads to predict confidence intervals with point estimation. They used Laplace loss to model Aleatoric Uncertainty and multivariate Gaussians to model Epistemic Uncertainty. However, loss of multiview visual characteristics and spatial structure characteristics from 2D human poses for 3D localization might be the reason for low accuracy compared with other research results.

Research in [1] used stereo data as input to produce set of 3D object proposals which run through convolutional neural network for high quality 3D object detection. Their proposed method generated 3D proposals using energy minimization function to encode object size priors, context of ground plane and features for depth information. However, they used 3D integral images which are not suitable for real time 3D object detection. Research in [4] used stereo data to predict sparse key point for estimating 3D bounding box. They proposed Stereo R-CNN to improve overall performance. However, recent advancement of R-CNN like Faster R-CNN should be implicated to justify the effectiveness of their proposed method. Research in [2] used stereo imagery as input to estimate depth using convolutional neural net (CNN) [11] to compute point clouds. They used proposal generation problem as inference in Markov Random Field (MRF) to encode high density in the point cloud. However, overall performance of their proposed method depends on accurate depth estimation.

Research in [2] used LiDAR point cloud and RGB images as input to generate 3D proposals and projected them to multiple views for feature extraction. They used region-based fusion network to deeply integrate Multiview information for classification of 3D proposals. However, region based fusion network works-based region using convolutional neural network [12] requires more computation overheads for real time detection. In addition, LiDAR sensor is expensive whereas optical camera could be a potential alternative for their proposed method. Research in [7] generated 3D proposal by using raw point cloud instead of generating proposals from RGB image or projecting point cloud to bird's view or voxels. They used PointNet++ to learn point-wise features for describing the raw point clouds which later used for foreground point cloud segmentation. However, for sparse convolutions, alternative point-cloud network structures, such as VoxelNet could be investigated as their backbone network. Research in [13] used RGBD data as input for raw point clouds and proposed framework for RGB-D data-based 3D object detection called Frustum PointNets. Their frustum region is based on 2D region proposal indicates that no 3D object will be detected without 2D region proposal or 2D detection. However, aggregation of image feature after extracting features using based on their backbone network could improve overall 3D object detection performance. Although, point clouds can provide detailed geometry and capture 3D structure of the scene, on the other hand, point clouds are irregular, which cannot be processed by powerful deep learning models, such as convolutional neural networks directly [14]. Research in [5] used both monocular images and stereo images and estimated depth and disparity for monocular and stereo images respectively to generate point clouds. They proposed two step approach by first extracting dense pixel depth from stereo or monocular imagery followed by back-projecting pixels into a 3D point cloud to view the representation as pseudo-LiDAR signal. However, for long distance objects pseudo-LiDAR signal could not provide expected detection results comparing with short distance objects.

III.　ANALYSIS OF METHODS BASED ON SIX ASPECTS

This research identified six significant key areas to analysis existing methods for 3D object detection mentioned in Fig. 1 and comprehensively illustrated in the next subsequent sections.
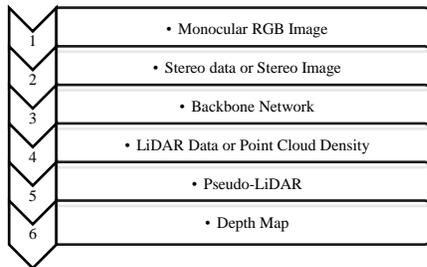
Fig. 1.　Six Key Areas to Analysis Existing Methods for 3D Object Detection.

A. Monocular rgb Images

Although 3D point cloud extracted from LiDAR provides superior performance, previously less effective alternative was monocular RGB images that were collected from single view RGB image as input for monocular 3D object detection [15]. Usage of monocular RGB images can facilitate domain adaptation for multimodal big data management and significantly aids for 3D model retrieval and classification when compared with multiview image sets [16]. Research in [4] used a single monocular image to generate class specific object proposals to run through standard CNN pipelines for 3D object detection. However, they assumed that objects should be on a ground plane which makes the overall proposed methodology uncomfortable to be used for other mediums such as detection from UAV, UGV or Krane. Research in [10] used monocular RGB images to tackle ill-posed problem of 3D human localization. They used Laplace distribution to address ambiguity by predicting confidence intervals of 3D bounding boxes. However, they used 2D human poses for 3D localization which might be the reason for low accuracy compared with other research results.

Research in [5] used both monocular and stereo images to mimic LiDAR signal and depth map constructions using the Pseudo-Signal hence called Pseudo-LiDAR and provided a milestone instead of using LiDAR sensor for point cloud generation. However, Pseudo-LiDAR approach required additional post processing steps due to the range issue of objects from the source of the camera platform.

If these pose processing can be eradicated, then Pseudo-LiDAR can be considered as a potential option to generate point cloud and depth map. Research in [9] used orthographic feature transforms from monocular image as part of an end-to-end deep learning architecture to map image-based features into an orthographic 3D space. However, fruitful experimentation on large variations of scales and distances can provide real time usage of their proposed method. There are several challenges exists for using monocular images for 3D object detection, i.e., significant information loss, object pose inconsistency, complex background of monocular images [16], accurate depth information [6]. Monocular images lose significant information during calibration such as multiview

visual characteristics and spatial structure characteristics. Overall scenarios of using monocular images are mentioned in Fig. 2.
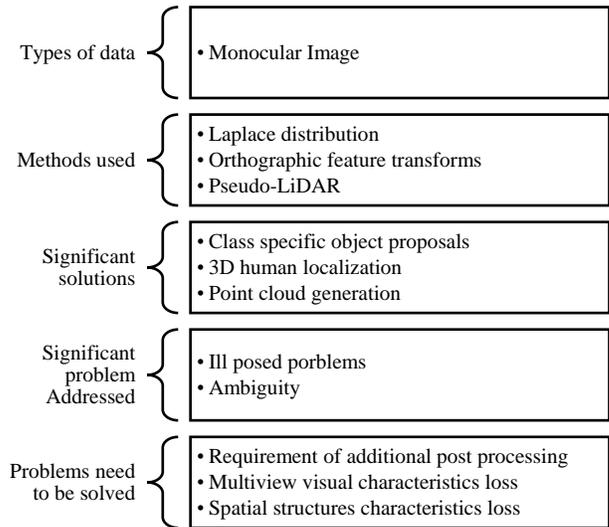
Fig. 2.　Scenarios of using Monocular RGB Images for 3D Object Detection.

B. Stereo rgb Images

Stereo cameras work in a manner like human binocular vision which is cost less and have higher resolutions for which they have gathered significant attention in academia and industry [17]. Disparity cues are provided by stereo images to enable better depth estimation compared to monocular images [18]. Research in [1] exploited stereo imagery for high quality 3D object proposals using energy minimization function to encode object size, localization of objects on the depth informed features to reason free space, point cloud densities and distance to the ground. They used CNN to exploit depth information for regressing 3D bounding box coordinates and object pose. However, they used 3D integral images which might cause additional computational cost during training. Research in [4] used stereo R-CNN for detecting associated objects in the stereo images. To achieve the aim, they proposed a 3D box estimator to exploit stereo box key points and dense region-based photometric alignment method to improve 3D object localization accuracy. However, due to later advancement of R-CNN like Faster R-CNN could be better option to investigate in their proposed method. Research in [5] proposed a twostep approach using stereo imagery data, i.e., estimation of depth map from stereo and usage of existing LiDAR-based 3D object detection pipelines. Although they used both monocular and stereo images, it is not clear based on the elaboration whether they used combined depth maps for both types of images or not. Research in [2] used contextual models by exploiting stereo information by reasoning and placing in 3D proposals in the form of 3D bounding boxes. Although, they implicated their proposed method for autonomous driving, they claimed better performance on other object classes such as Cyclist and Pedestrians. Several other shortcomings for using stereo images are observed by this research although stereo image based detection shown promising results, i.e. calibration issues between two camera rigs [2], memory expensive 3D cost

volume learning in image space due to the image's high resolution and requirement of additional dimension for features [18], depth estimation in image space while downstream detection is performed in 3D space [18], existence of many pixels among nearby objects due to perspective projection than faraway objects which lead to biased depth estimation with degraded long-range detection performance [18], dependency on anchor-based 2D detectors with association approach [17].

For stereo images, in image space, 3D cost volume learning is costly in memory due to high computation and extra dimension of the features. Besides, previous lest computation time for one frame was 0.5 seconds which can be challenging for critical applications such as collision avoidance [17, 20, 21]. Besides, in stereo images, depth estimation is performed in 2D image space whereas detection tasks take place in 3D space. Existence of more pixels in nearby objects than faraway objects due to perspective projection led to imbalance biased depth map. Overall scenarios of using stereo images are mentioned in Fig. 3.

C. LiDAR Data/ Point Cloud Density

Research in [1] extended 3D object proposal with class independent variant and neural network to grab both appearance and depth features. Later, they used point clouds obtained via LiDAR followed by giving comparison of the stereo, LIDAR and hybrid settings. However, class specific module they proposed, needs further elaboration to justify their claim. Research in [3] used sensory-fusion framework to take both LiDAR point cloud and RGB images as input. They encoded sparse 3D point cloud for compact multiview representation.

Types of data	• Stereo Image
Methods used	• Energy minimization function • Stereo R-CNN • Depth map estimation • Contextual models
Significant solutions	• Improved 3D Object localization accuracy • Depth map estimation
Problems need to be solved	• Additional computational cost • Biased depth map • Dependancy on anchor based 2D detectors • High image resolution and extra dimension

Fig. 3. Scenarios of using Stereo Images for 3D Object Detection.

Their proposed deep learning network was composed of two subnetworks, i.e., one for 3D object proposal generation and other one for multiview features fusion. For multiple views and interactions between intermediary layers they used deep fusion schemes to combine these two parts. However, region based fusion network works-based region using

convolutional neural network requires more computation overheads for real time detection. In addition, LiDAR sensor is expensive whereas optical camera could be a potential alternative for their proposed method. Research in [4] used raw point cloud for generating small number of high-quality 3D proposals via segmenting the point clouds of the whole scene into foreground points and background. They performed 3D box refinement later by combing local spatial features with global semantic features [22] for each point cloud. However, later version of RCNN such as Fast RCNN and Faster RCNN could be aligned later to find the suitability of the proposed method with most recent deep learning architectures in case of using CNN for raw point cloud achieved from LiDAR.

D. Pseudo-LiDAR Generation

Previous LiDAR methods formulated LiDAR point cloud as point [7], pillar [23] and voxel representation [29,30]. Despite the remarkable performance of these methods, LiDAR sensors are expensive sensor and for the far distance, LiDAR sensor-based detection can potentially make 3D object detection tasks difficult which initiate the need for some approaches to mimic LiDAR point cloud due to high accuracy called Pseudo-LiDAR based point cloud representation. Research in [26] proposed self-supervised learning schemes by mimicking latent spatial features representation based on point-based module. However, their proposed method depends on 2D-3D detection pairs. Accurate depth map is one of the key challenges for 3D object detection from monocular image as lack of prior information is the main issue and eradicated recently by deep learning approaches [25, 27]. By constructing robust depth map, Pseudo-LiDAR point cloud can be constructed to mimic LiDAR point cloud based on pre-calibrated intrinsic camera parameters [28,29]. However, heavy computation is the bottleneck for this research. Research in [30] generated pseudo-ground without the need of LiDAR point clouds by proposing a statistical shape model to address the challenge of disparity annotations in training. However, for monocular 3D object detection, disparity map generation will not be possible for their proposed method. In addition, their proposed research depends on shape analysis which may provide poor performance in case of obstacles such as shadow. Research in [31] combined 2D object detection with Pseudo-LiDAR point cloud data generated from stereo images to investigate the boost of performance with existing six different 3D object detectors. However, distance calculation for their proposed method can be costly. Besides, they did not investigate high consistency issue for the point clouds generated by Pseudo-LiDAR approach. Research in [32] constructed a pseudo-LiDAR feature volume (PLUME) to estimate depth map and 3D object detection in 3D metric space. The main purpose for their proposed PLUME is to avoid biased depth estimation with degraded long-range detection for stereo images. However, for low light and extreme weather conditions, their proposed method may encounter due to the need of accurate depth map estimation.

E. Depth Map

Recent algorithms for stereo depth estimation can produce surprisingly accurate depth maps [33]. However, inferring depth of pedestrians from monocular images is a fundamentally ill-posed problem [10]. This additional

challenge is due to human variation of height. If every pedestrian has the same height, there would be no ambiguity [10]. LiDAR point clouds has been dominated approach in the existing research for 3D object detection due to the availability of accurate depth map [3, 13, 34, 35, 36, 37, 38, 39]. However, performance of image-only methods lacks in absolute depth information of LiDAR. In this context, optical cameras are highly affordable than LiDAR, operate at a high frame rate, and provide a dense depth map rather than 64 or 128 sparse rotating laser beams that LiDAR signal is inherently limited to [5]. Research in [40] proposed first study for estimating depth map in outdoor environments where they combined single RGB image and LiDAR point cloud. However, due to large learning network, powerful architectures for training were needed to use their proposed method. Research in [5] converted estimated depth map from stereo or monocular imagery into a 3D point cloud referred to as pseudo-LiDAR as it mimics the LiDAR signal. Then they took the advantage of existing LiDAR-based 3D object detection pipelines [13, 34] to train directly on the pseudo-LiDAR representation using deep convolutional neural networks to obtain unprecedented increase in accuracy of image-based 3D object detection algorithms. However, usage of LiDAR based method may not be suitable for other image-based classifier for overall 3D object detection. Besides, their proposed pseudo-LiDAR fails to detect far-away objects precisely due to inaccurate depth estimation. Research in [10] detected 2D joints using PifPaf and used MonoDepth [41] to estimate depth for a set of 9 pixels around each key point followed by consideration of minimum depth as their reference value. Later, they calculated distance from normalized image coordinates of the centre of the bounding box using the estimated minimum depth rather than using the average one. However, their proposed method needs further investigation for monocular image as their method worked only for stereo images. Overall scenarios for depth map estimation using different sensors and images are shown in Fig. 4.
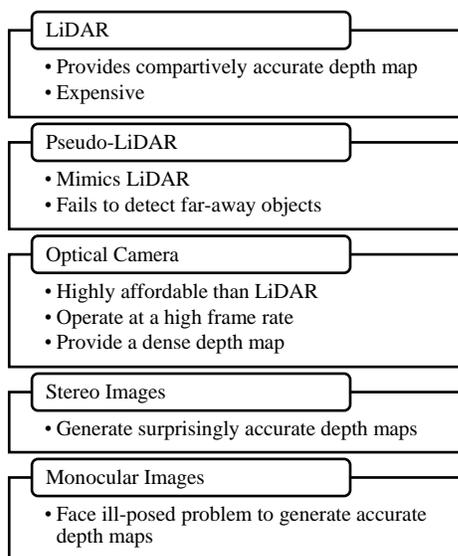
Fig. 4. Overall Scenario for Depth Map Estimation using Different Sensors and Types of Images.

F. Backbone Network

For 3D object detection various backbone networks has been used for feature representation shown in Fig. 5. Research in [42] preferred image features in ResNet-101 [4, 17, 43] for block 1 to maintain a high spatial resolution and avoid redundancy of same features. However, ResNet as backbone network contains too many layers and are not very efficient [4]. Darknet-53 [44] as a backbone network are used with fewer floating-point operations and more speed. Darknet-53 is better than ResNet-101 and 1.5× faster. In addition, Darknet-53 has similar performance to ResNet-152 and is 2× faster. Darknet-53 also achieves the highest measured floating-point operations per second which means the network structure better utilizes the GPU, making it more efficient to evaluate and thus faster.
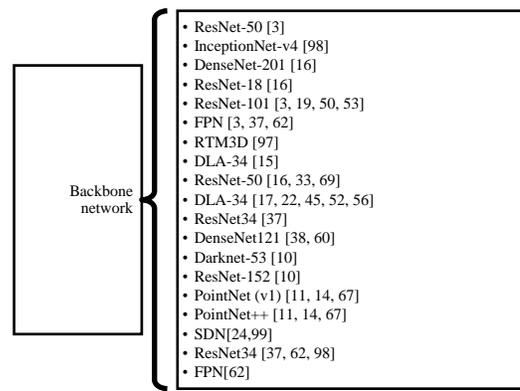
Fig. 5. Backbone Networks for 3D Object Detection.

Research in [13] used PointNet (v1) and PointNet++ (v2) backbone due to much cleaner in design. While out of the scope for their work, sensor fusion in terms with aggregation of image feature for 3D object detection after extracting features using PointNet (v1) and PointNet++ could further improve their results. Research in [7] extracted point-wise features encoded by PointNet++ as backbone point cloud network, they appended one segmentation head for estimating the foreground mask and one box regression head for generating 3D proposals. However, alternative point-cloud network structures, such as VoxelNet with sparse convolutions could be investigated as their backbone network. Research in [16] used ResNet-50 as the backbone for feature extraction followed by a fully connected neural network [45] with one or two hidden layers. The number of hidden nodes is tuned with 64, 100, 128, and 192 to generate the candidate results which cause additional processing due to too many layers. Research in [63] used SDN [46] as backbone network to estimate dense depth map and fine-tuned on the KITTI datasets. However, their predicted depth was not accurate enough since the ground truth is very sparse. Research in [47] used ResNet34 [48, 49] with Feature Pyramid Network (FPN) [48] as backbone network. They replaced BatchNorm+ReLU layers with the synchronized version of InPlaceABN activated with LeakyReLU with negative slope 0.01 as proposed in [50] to

free up a significant amount of GPU memory for scaling up the batch size or input resolution. However, top down and bottom-up features fusion in feature pyramid network might create additional processing due to combine ResNet34 and FPN. Existing methods mostly used ResNet-50 as the backbone network for both monocular and stereo images due to contain too many layers [4]. InceptionNet-v4 [49] was previously used for visual feature extraction, however, in case of 3D models for multiview mages, InceptionNet-v4 needs to be further investigated for 3D object detection overall pipeline. DenseNet-201 [16] was used to extract the visual features of monocular images and ResNet-18 [16] was for multiview feature extraction of 3-D models. However, combination two backbone network at two different phases, may increase overall computational overheads. In case of monocular images and multiview fusion, four key aspects can ensure efficient use of backbone network for 3D object detection mentioned in Fig. 6 [16].

Domain adaptation is the first concerns for backbone network to be used with any method or strategy to solve cross domain problem. In this context, improved datasets developed with the concerns about cross domain ease the task for backbone network. For monocular and multiview aspects different fusion strategies for multiview views are needed for the extracted features using backbone network, such as pooling, concatenation. As monocular images lose multiview visual characteristic and spatial structure characteristic causes significant information loss, various function needs to be designed during features learning process. Efficient implication of the above three aspects helps for efficient pair wise similarity measurement.
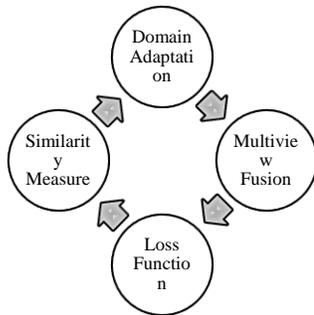
Fig. 6. Four Key Aspects for Efficient Use of Backbone Network for 3D Object Detection.

IV. REVIEW ON EXPERIMENTAL ANALYSIS

A. Hardwares and Software Specification for Experimentation

Previously researcher used sensors such as Grayscale cameras [51], optical cameras [5], color cameras [51] and LiDAR like Velodyne HDL-64E LIDAR [5, 51] for own datasets development for validating 3D object detection evaluation. For training, various GPUs were used mentioned in Fig. 7. PyTorch machine learning framework [42, 52, 53, 54, 55, 56] have been mostly used by existing research.
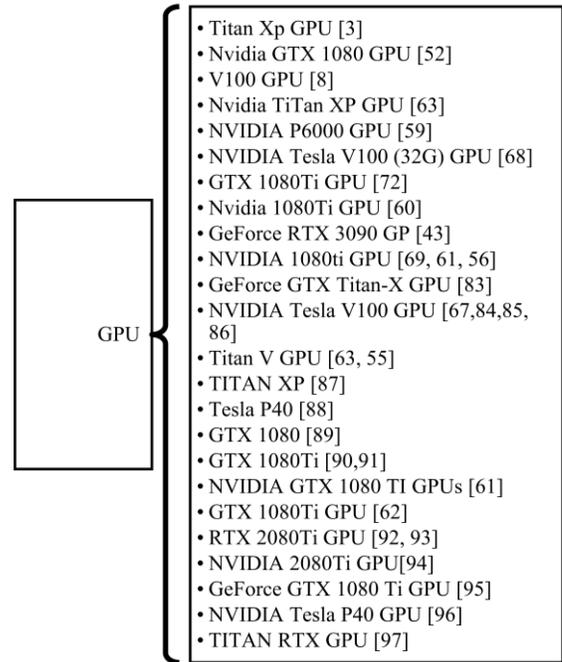
Fig. 7. GPUs Used in Previous Research.

B. Datasets

Domain gaps for different datasets depend on object size, weather condition, specific locations, and orientation [57]. Overall gaps can be categorized in to two categories mentioned below.

1) Content gap such as object, weather condition due to locations during data capture depending on time.

2) Point distribution gap owing to different LiDAR types such as number of beam ways, beam range, vertical inclination, horizontal and vertical angularity estimation of LiDAR. Existing datasets used for 3D object detection purpose are mentioned in Fig. 8.
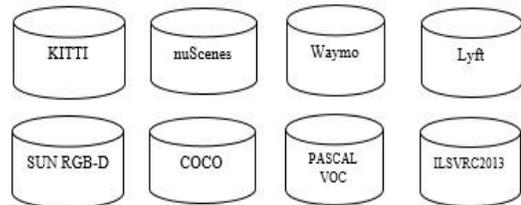
Fig. 8. Datasets for 3D Object Detection.

1) KITTI dataset: KITTI is the most used dataset for 3D object detection. However, existing datasets such as nuScenes, Waymo, Lyft, SUN RGB-D, COCO, PASCAL VOC and ILSVRC2013 shown in Fig. 8 has been used by the research for more robust validation.

The KITTI 3D dataset [58, 59] is the most used benchmark in the 3D object detection task and it provides left camera images, calibration files, annotations for 3D detection [45].

KITTI datasets are the widely used benchmark for validating 3D object detection includes 2D object detection, Average Orientation Similarity (AOS), Bird's Eye View (BEV) [13, 60, 61]. Samples in KITTI datasets include 3D point clouds, images and Camera-LiDAR calibration data [52]. Images in KITTI datasets are captured in the same city using same camera [10]. 3D bounding boxes for various object classes are provided in KITTI datasets which includes cars, vans, trucks, pedestrians and cyclists labelled manually in 3D point clouds depending on calibrated camera's information [62]. Number of training and test images or point clouds in KITTI datasets are 7481 and 7518 images respectively containing three classes, i.e., car, pedestrian and cyclist [1,2, 8, 42, 52, 54, 60, 61, 63, 64, 65, 66, 75]. Each class is annoted by camera Field of Vision (FOV) with the 3D bounding boxes [64]. Evaluation for each class depends into three categories, i.e., easy, moderate, hard according three aspects, i.e., object size, occlusion state and maximum truncation levels of objects [10, 42, 53, 55]. For ranking the completion of the methods, moderate category is used in the benchmark [2]. Easy object is indicated with minimum pixel height as 40px within 28m as vehicles correspondence. Besides, 25px are the limit for moderate and hard level objects within minimum distance of 47m [62].

In another way, there are three commonly used data splits in the KITTI dataset, i.e., split for testing, validation category 1, and validation category 2 mentioned in Fig. 9.
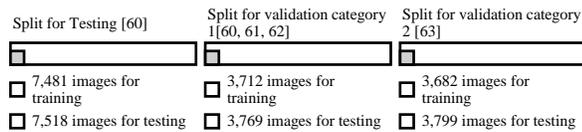
Fig. 9. Split of Images for Training and Testing in KITTI Datasets.

In KITTI datasets, object detection validation is estimated mostly through average precision (AP) and IOU (Intersection over union) with threshold 0.7 for car class [8, 53, 42, 67, 62], 0.5 for pedestrian [60, 67, 62] and 0.5 for cyclist [60, 67, 62]. Six illustrative factors for using KITTI datasets is shown in Fig. 10.

In addition, this research identified ten challenges to use KITTI dataset mentioned in Fig. 11 although KITTI dataset is the commonly used dataset for 3D object detection performance validation.

2) nuScenes dataset: nuScenes dataset contains 1000 segments of 20 seconds each for 3D object detection where 750, 150 and 150 segments for training, validation and testing, respectively [10, 68, 69, 57, 70]. Annotation rate is 2Hz for which 28k, 6k and 6k annotated frames are available in these datasets for training, validation and testing respectively. This dataset contains more classes comparing with KITTI datasets which is 10 and evaluation metrics are mean average precision and nuScenes detection score (NDS) [68]. BEV center distance is the true positive metric for this dataset instead of IoU which is another significant difference with KITTI

datasets. However, camera extrinsic information is not also available in this dataset like in KITTI datasets [69].

Fig. 10. Six Illustrative Factors to use KITTI Datasets for 3D Object Detection.

Fig. 11. Ten Challenges to use KITTI Datasets for 3D Object Detection Validation Performance.

3) Waymo dataset: Waymo Open Dataset is more recently released datasets consists of 798 training sequences and 202 validation sequences [8, 53, 71, 42, 72, 57]. Waymo dataset provides object labels in the full 360◦ field of view with a multi-camera rig which is the advantage over KITTI and NuSceneses dataset. However, Waymo dataset includes only 150 test sequences without ground truth data. In addition, there is no published depth results on Waymo open dataset [15].

In addition, with KITTI, nuScenes and Lyft datasets, for 3D object detection other datasets such Lyft [68, 57], PASCAL VOC [73, 74], ILSVRC2013 [73, 74] dataset was previously used for validation. However, these datasets use different annotation rules for validation, i.e., large number of objects outside the road were not annotated for validation.

C. Validation Metrics

Overall framework for 3D object detection has been validated based on two tasks, i.e., object detection and then object detection with orientation estimation [75,103]. To validate accurate object detection, Average Precision (AP) metric was used by most of the existing research, Average Orientation Similarity (AOS) was used mostly for object detection and orientation estimation. The true positive metric is based on 2D/3D IoU. This research identified eight major performance metrics for validating any method for 3D object detection mentioned in Fig. 12.
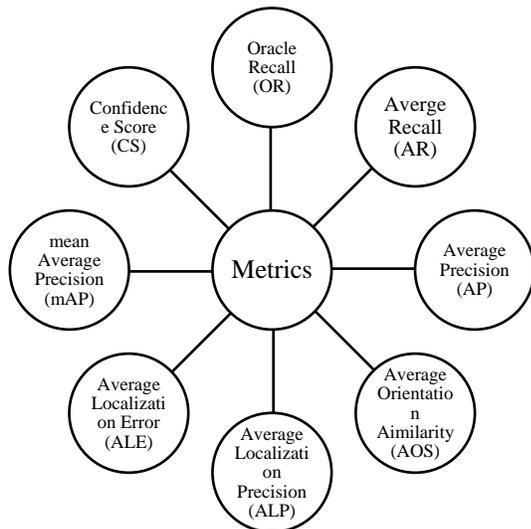
Fig. 12. Major Eight Performance Metrics for Validation for 3D Object Detection.

1) Oracle Recall: Oracle recall computes the percentage of the recalled ground truth objects to receive recall rate [75, 2, 76, 104,105]. For certain threshold, if at least one proposal overlaps with IoU, then ground truth object is said to be recalled [1]. Recall is measured for both 2D bounding box and later for 3D bounding box for overall performance.

2) Average Recall (AR): For 3D object detection performance, Average Recall (AR) [32] is highly correlated metrics that needs to be measured for both 2D bounding box and 3D bounding box for overall performance [1,4,75,32]. For

stereo AR metrics, Average Recall needs to be measured for both left and right images [3].

3) Average Precision (AP): For various sampled points, average precision (AP) extracts average value of precision at various recall threshold values [62].

In other words, Average Precision (AP) indicates the average precision value for recall over 0 to 1. For precision and recall, ideal value is 1 [62]. However, for the real time scenarios, any method is assumed to be good enough if the precision and recall metrics gets closer to 1. In this context, this research observed that there is a trade-off between precision and recall, i.e. if more optimizations can be done on precision, recall gets lower, oppositely if recall can be improved, precision value becomes lower. So, this research recommends to balance at the point of fixing threshold point. Average Precision (AP) is used for 2D and 3D object detection for both monocular and stereo images [1, 4, 8, 53, 62, 63,75,77, 83,102]. Like AR, stereo AP metric needs to be evaluated for both on left and right images.

4) Average Orientation Similarity (AOS): Average Orientation Similarity (AOS) indicates perfect prediction between 0 and 1 [1,2, 75]. For 3D object detection, AOS has been used for orientation estimation task for 2D object detection performance towards 3D object detection [75, 15].

5) Average Localization Precision (ALP) [1,10]: Average precision and recall are the requirement to calculate Average Localization Precision (ALP) which can be computed similarly to AP except that 3D localization precision needs to be replaced in pace of bounding box overlap [1]. In other words, ALP provides a prediction to be correct depending on the error between predicted distance and ground truth is smaller than threshold [10]. In this context, predicted 3D location is to be correct if the distance to the ground truth 3D location is smaller than certain threshold [1].

6) Average Localization Error (ALE) [10]: For misaligned bounding box, Average Localization Error (ALE) is estimated from the target category [10]. In other words, ALE provides variation of the actual and estimated value of each location in the localization process.

7) mean Average Precision (mAP): mean Average Precision (mAP) is calculated from Average Precision (AP) from all classes for the IoU thresholds depending on various problems scenario [42,78, 101].

8) Confidence Score: Confidence score indicates the optimum threshold to categorize false positives to ensure the predicted bounding box contains minimum standard score and often used for model performance evaluation [19,79,106]. Non optimal settings for any proposed model requires more minimized confidence score for precise bounding box detection for 3D object detection. 3D box confidence estimation for 3D object detection realized by the previous research. Research in [80] calculated 3D IoU from the predicted 3D box and ground truth involves 3D object dimensions. Research in [47] used 3D box loss to represent 3D detection. Research in [81] introduced self-balancing

confidence loss for generating confidence score from relatively achievable samples. However, all these methods considered loss function for confidence score. To overcome this problem, research in [82] considered the relationship between 3D objects and associated 2D boxes to decompose confidence mechanisms. However, in case of weak transformation for 2D to 3D, their proposed confidence decomposition may result in weakness for their proposed methodology. Research in [10] addressed ill posed problems for predicting confidence intervals to account aleatoric and epistemic uncertainties. They estimated pose to obtain 2D joints which later were used as input to feedforward network and output the 3D location along with a confidence interval. However, their proposed method is suited only for small training data [100]. Research in [83] propagated information from the labelled to unlabeled training set in the form of pseudo-labels contains significant noise for which they introduced confidence-based filtering mechanism for 3D object detection. Their confidence proposals were based on predicted objectless and class probability to filter low quality pseudo labels. However, their proposed confidence intervals depend on category specific thresholds.

V. Observations and Future Research Directions

1) There is a tradeoff between precision and recall, i.e. if more optimizations can be done on precision, recall gets lower, oppositely if recall can be improved, precision value becomes lower. So, this research recommends to balance at the point of fixing threshold point for IoU.

2) Design of experiments should cover practical 3D domain adaptation scenarios mentioned below:

a) Adaptation from label rich domains to label insufficient domains,

b) Adaptation across domains with different data collection locations and time (e.g., Waymo → KITTI, nuScenes → KITTI), and

c) Adaptation across domains with a different number of the LiDAR beams (i.e., Waymo → nuScenes and nuScenes → KITTI). Therefore, domain adaptive evaluation needs to be done for validating 3D object detection models on the following four adaptation tasks: Waymo → KITTI, Waymo → Lyft, Waymo → nuScenes and nuScenes → KITTI.

3) Some ill-posed settings that is not suitable for evaluation needs to rule out. For example, KITTI datasets lacks in ring view annotations (less practical) and Lyft uses very different annotation rules (i.e., many objects outside the road are not annotated).

4) Comparing with stereo images, monocular images lose multiview visual characteristic and spatial structure characteristic causes significant information loss demands for robust depth construction for 3D object detection which can lead to 3D localization.

5) Due to the significant visual gap between object-centric images (various texture with complex background) and multiview images of 3-D models (gray model appearance with clean background), monocular and stereo image-based 3D object detection toward localization or tacking domain adaptation is a challenging task requires further investigation. In lieu of current existing datasets, development of novel dataset for object centric monocular and stereo image based is required to advocate the use of 3D object detection towards 3D localization for real world applications. In this context, few possible query images should be the primary key which can be the significant implication for domain adaptation for overall 3D object detection pipeline.

6) To choose the appropriate backbone network depending on problem specific 3D object detection, appropriate fusion strategies need to be designed in lieu of robust loss function to ensure efficient similarity measurement for final classification tasks of 3D object detection.

VI. Conclusion

3D object detection is the basis of many autonomous intelligent applications. This research demonstrates comprehensive and critical reviews on existing 3D object detection methods using RGB images and other fusion based detection methodology based on LiDAR and Pseudo-LiDAR. Some existing methods detected objects with 2D bounding box to recognize position of the objects which is not sufficient for perfect autonomous system. Therefore, predicting 3D object's position is similarly important as determining the 2D position of object in the image. In this research, sensor modality for the overall review is categorized in four types, i.e., monocular image, stereo image, point clouds obtained from LiDAR and Pseudo-LiDAR and fusion of both where advantages and disadvantages were addressed for each type. Depth summary with relative challenges for eight datasets are critically highlighted by this research. In this context, KITTI benchmark are not suitable for monocular methods for 3D object detection due to lack of depth information and prevents accurate 3D positioning which encourages to use maximum number of datasets to ensure robustness for any 3D object detection method. Besides, comprehensive details for eight evaluation metrics are illustrated to evaluate 3D object detection methods. This research observed that 3D object detection is not matured as 2D object due to large gap existing between them. Existing methods still did not achieve the benchmark performance for real time autonomous applications initiates the need for fast and reliable 3D object detection system for wide range of real time applications. Besides, recent trend for using point cloud processing was observed by this research provides effective solution for 3D object detection but LiDAR is an expensive sensor and further geometrical relationship needs to be discovered among points. Besides, some fusion based methods, i.e., RGB images either with LiDAR point cloud or depth images from RGB-D data could not confirm their superiority than other methods in multimodal datasets to ensure robust validation which indicates that more focus is needed to develop multimodal methods for 3D object detection. In addition, lack of large scale annotated training data, more datasets and fusion methods are expected in near future for indoor and outdoor scenarios to form a unified 3D object detection framework. From this study, this research remarks that 3D object detection has

gained many successes, but remains as potential and fertile research problem which requires more exploration. Demonstrated critical review by this research is expected to serve as a supportive significant reference and forms an important endorsement to the related research community.

REFERENCES

[1] X. Chen, K. Kundu, Y. Zhu, H. Ma, S. Fidler, R. Urtasun, 3d object proposals using stereo imagery for accurate object class detection, IEEE transactions on pattern analysis and machine intelligence, 40 (2017) 1259-1272.

[2] X. Chen, K. Kundu, Y. Zhu, A.G. Berneshawi, H. Ma, S. Fidler, R. Urtasun, 3d object proposals for accurate object class detection, in: Advances in Neural Information Processing Systems, Citeseer, 2015, pp. 424-432.

[3] X. Chen, H. Ma, J. Wan, B. Li, T. Xia, Multi-view 3d object detection network for autonomous driving, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2017, pp. 1907-1915.

[4] P. Li, X. Chen, S. Shen, Stereo r-cnn based 3d object detection for autonomous driving, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7644-7652.

[5] Y. Wang, W.-L. Chao, D. Garg, B. Hariharan, M. Campbell, K.Q. Weinberger, Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 8445-8453.

[6] Z. Liu, D. Zhou, F. Lu, J. Fang, L. Zhang, Autoshape: Real-time shape-aware monocular 3d object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 15641-15650.

[7] S. Shi, X. Wang, H. Li, Pointrcnn: 3d object proposal generation and detection from point cloud, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 770-779.

[8] H. Sheng, S. Cai, Y. Liu, B. Deng, J. Huang, X.-S. Hua, M.-J. Zhao, Improving 3d object detection with channel-wise transformer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 2743-2752.

[9] T. Roddick, A. Kendall, R. Cipolla, Orthographic feature transform for monocular 3d object detection, arXiv preprint arXiv:1811.08188, (2018).

[10] L. Bertoni, S. Kreiss, A. Alahi, Monoloco: Monocular 3d pedestrian localization and uncertainty estimation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6861-6871.

[11] A.S. Saif, Z.R. Mahayuddin, Moment Features based Violence Action Detection using Optical Flow, International Journal of Advanced Computer Science and Applications, 11 (2020).

[12] A.S. Saif, Z.R. Mahayuddin, Robust Drowsiness Detection for Vehicle Driver using Deep Convolutional Neural Network, International Journal of Advanced Computer Science and Applications, 11 (2020).

[13] C.R. Qi, W. Liu, C. Wu, H. Su, L.J. Guibas, Frustum pointnets for 3d object detection from rgb-d data, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 918-927.

[14] X. Pan, Z. Xia, S. Song, L.E. Li, G. Huang, 3d object detection with pointformer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 7463-7472.

[15] L. Wang, L. Zhang, Y. Zhu, Z. Zhang, T. He, M. Li, X. Xue, Progressive coordinate transforms for monocular 3d object detection, Advances in Neural Information Processing Systems, 34 (2021).

[16] D. Song, W.-Z. Nie, W.-H. Li, M. Kankanhalli, A.-A. Liu, Monocular Image-Based 3-D Model Retrieval: A Benchmark, IEEE Transactions on Cybernetics, (2021).

[17] Y. Chen, S. Liu, X. Shen, J. Jia, Dsgn: Deep stereo geometry network for 3d object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12536-12545.

[18] Y. Wang, B. Yang, R. Hu, M. Liang, R. Urtasun, PLUMENet: Efficient 3D Object Detection from Stereo Images, in: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, pp. 3383-3390.

[19] D. Rukhovich, A. Vorontsova, A. Konushin, Imvoxelnet: Image to voxels projection for monocular and multi-view general-purpose 3d object detection, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 2397-2406.

[20] C. Li, J. Ku, S.L. Waslander, Confidence guided stereo 3D object detection with split depth estimation, in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2020, pp. 5776-5783.

[21] A.S. Saif, Z.R. Mahayuddin, H. Arshad, Vision-Based Efficient Collision Avoidance Model Using Distance Measurement, in: Soft Computing Approach for Mathematical Modeling of Engineering Problems, CRC Press, 2021, pp. 191-202.

[22] A.S. Saif, Z.R. Mahayuddin, Edge Feature based Moving Object Detection Using Aerial Images: A Comparative Study, in: Edge Feature based Moving Object Detection Using Aerial Images: A Comparative Study, IEEE Press, 2021.

[23] A.H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, O. Beijbom, Pointpillars: Fast encoders for object detection from point clouds, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 12697-12705.

[24] Y. Zhou, O. Tuzel, Voxelnet: End-to-end learning for point cloud based 3d object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4490-4499.

[25] S. Shi, Z. Wang, J. Shi, X. Wang, H. Li, From points to parts: 3d object detection from point cloud with part-aware and part-aggregation network, IEEE transactions on pattern analysis and machine intelligence, (2020).

[26] D. Feng, S. Han, H. Xu, X. Liang, X. Tan, Point-Guided Contrastive Learning for Monocular 3-D Object Detection, IEEE Transactions on Cybernetics, (2021).

[27] Z. Liu, D. Zhou, F. Lu, J. Fang, L. Zhang, Autoshape: Real-time shape-aware monocular 3d object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 15641-15650.

[28] R. Qian, D. Garg, Y. Wang, Y. You, S. Belongie, B. Hariharan, M. Campbell, K.Q. Weinberger, W.-L. Chao, End-to-end pseudo-lidar for image-based 3d object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5881-5890.

[29] J.M.U. Vianney, S. Aich, B. Liu, Refinedmpl: Refined monocular pseudolidar for 3d object detection in autonomous driving, arXiv preprint arXiv:1911.09712, (2019).

[30] L. Chen, J. Sun, Y. Xie, S. Zhang, Q. Shuai, Q. Jiang, G. Zhang, H. Bao, X. Zhou, Shape Prior Guided Instance Disparity Estimation for 3D Object Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2021).

[31] F. Negahbani, O.B. Töre, F. Güney, B. Akgun, Frustum Fusion: Pseudo-LiDAR and LiDAR Fusion for 3D Detection, arXiv preprint arXiv:2111.04780, (2021).

[32] Y. Wang, B. Yang, R. Hu, M. Liang, R. Urtasun, PLUMENet: Efficient 3D Object Detection from Stereo Images, in: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, pp. 3383-3390.

[33] J.-R. Chang, Y.-S. Chen, Pyramid stereo matching network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5410-5418.

[34] J. Ku, M. Mozifian, J. Lee, A. Harakeh, S.L. Waslander, Joint 3d proposal generation and object detection from view aggregation, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2018, pp. 1-8.

[35] L. Peng, F. Liu, S. Yan, X. He, D. Cai, OCM3D: Object-Centric Monocular 3D Object Detection, arXiv preprint arXiv:2104.06041, (2021).

[36] S. Wirges, T. Fischer, C. Stiller, J.B. Frias, Object detection and classification in occupancy grid maps using deep convolutional networks, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2018, pp. 3530-3535.

[37] X. Du, M.H. Ang, S. Karaman, D. Rus, A general pipeline for 3d detection of vehicles, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018, pp. 3194-3200.

[38] K. Minemura, H. Liau, A. Monrroy, S. Kato, LMNet: Real-time multiclass object detection on CPU using 3D LiDAR, in: 2018 3rd Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), IEEE, 2018, pp. 28-34.

[39] J. Beltrán, C. Guindel, F.M. Moreno, D. Cruzado, F. Garcia, A. De La Escalera, Birdnet: a 3d object detection framework from lidar information, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2018, pp. 3517-3523.

[40] J. Qiu, Z. Cui, Y. Zhang, X. Zhang, S. Liu, B. Zeng, M. Pollefeys, Deeplidar: Deep surface normal guided depth prediction for outdoor scene from sparse lidar data and single color image, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3313-3322.

[41] C. Godard, O. Mac Aodha, G.J. Brostow, Unsupervised monocular depth estimation with left-right consistency, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 270-279.

[42] C. Reading, A. Harakeh, J. Chae, S.L. Waslander, Categorical depth distribution network for monocular 3d object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 8555-8564.

[43] Y. Liu, Y. Yixuan, M. Liu, Ground-aware monocular 3d object detection for autonomous driving, IEEE Robotics and Automation Letters, 6 (2021) 919-926.

[44] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, arXiv preprint arXiv:1804.02767, (2018).

[45] A.S. Saif, Z.R. Mahayuddin, Vision based 3D Gesture Tracking using Augmented Reality and Virtual Reality for Improved Learning Applications, International Journal of Advanced Computer Science and Applications, 12 (2022) 631-638.

[46] Y. You, Y. Wang, W.-L. Chao, D. Garg, G. Pleiss, B. Hariharan, M. Campbell, K.Q. Weinberger, Pseudo-lidar++: Accurate depth for 3d object detection in autonomous driving, arXiv preprint arXiv:1906.06310, (2019).

[47] A. Simonelli, S.R. Bulo, L. Porzi, M. López-Antequera, P. Kontschieder, Disentangling monocular 3d object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1991-1999.

[48] A. Simonelli, S.R. Bulo, L. Porzi, M.L. Antequera, P. Kontschieder, Disentangling monocular 3d object detection: From single to multi-class recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2020).

[49] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[50] S.R. Bulo, L. Porzi, P. Kontschieder, In-place activated batchnorm for memory-optimized training of dnns, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5639-5647.

[51] C. Chen, L.Z. Fragonara, A. Tsourdos, RoIFusion: 3D Object Detection From LiDAR and Vision, IEEE Access, 9 (2021) 51710-51721.

[52] A. Paigwar, D. Sierra-Gonzalez, Ö. Erkent, C. Laugier, Frustum-pointpillars: A multi-stage approach for 3d object detection using rgb camera and lidar, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 2926-2933.

[53] J. Yang, S. Shi, Z. Wang, H. Li, X. Qi, ST3D: Self-training for Unsupervised Domain Adaptation on 3D Object Detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10368-10378.

[54] A. Kumar, G. Brazil, X. Liu, GrooMeD-NMS: Grouped Mathematically Differentiable NMS for Monocular 3D Object Detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 8973-8983.

[55] J. Noh, S. Lee, B. Ham, HVPR: Hybrid Voxel-Point Representation for Single-stage 3D Object Detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14605-14614.

[56] L. Peng, F. Liu, S. Yan, X. He, D. Cai, OCM3D: Object-Centric Monocular 3D Object Detection, arXiv preprint arXiv:2104.06041, (2021).

[57] J. Yang, S. Shi, Z. Wang, H. Li, X. Qi, ST3D++: Denoised Self-training for Unsupervised Domain Adaptation on 3D Object Detection, arXiv preprint arXiv:2108.06682, (2021).

[58] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? the kitti vision benchmark suite, in: 2012 IEEE conference on computer vision and pattern recognition, IEEE, 2012, pp. 3354-3361.

[59] T. Guan, J. Wang, S. Lan, R. Chandra, Z. Wu, L. Davis, D. Manocha, M3detr: Multi-representation, multi-scale, mutual-relation 3d object detection with transformers, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 772-782.

[60] X. Ma, Y. Zhang, D. Xu, D. Zhou, S. Yi, H. Li, W. Ouyang, Delving into Localization Errors for Monocular 3D Object Detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4721-4730.

[61] Y. Zhang, X. Ma, S. Yi, J. Hou, Z. Wang, W. Ouyang, D. Xu, Learning Geometry-Guided Depth via Projective Modeling for Monocular 3D Object Detection, arXiv preprint arXiv:2107.13931, (2021).

[62] A. Sagar, Aa3dnet: Attention augmented real time 3d object detection, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 628-635.

[63] Y. Lu, X. Ma, L. Yang, T. Zhang, Y. Liu, Q. Chu, J. Yan, W. Ouyang, Geometry uncertainty projection network for monocular 3d object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 3111-3121.

[64] J. Li, Y. Sun, S. Luo, Z. Zhu, H. Dai, A.S. Krylov, Y. Ding, L. Shao, P2V-RCNN: Point to Voxel Feature Learning for 3D Object Detection from Point Clouds, IEEE Access, 9 (2021) 98249-98260.

[65] G. Brazil, G. Pons-Moll, X. Liu, B. Schiele, Kinematic 3d object detection in monocular video, in: European Conference on Computer Vision, Springer, 2020, pp. 135-152.

[66] T. Jiang, N. Song, H. Liu, R. Yin, Y. Gong, J. Yao, VIC-Net: Voxelization Information Compensation Network for Point Cloud 3D Object Detection, in: 2021 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2021, pp. 13408-13414.

[67] Y. Shi, Y. Guo, Z. Mi, X. Li, Stereo CenterNet-based 3D object detection for autonomous driving, Neurocomputing, 471 (2022) 219-229.

[68] C. Reading, A. Harakeh, J. Chae, S.L. Waslander, Categorical depth distribution network for monocular 3d object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 8555-8564.

[69] Y. Zhou, Y. He, H. Zhu, C. Wang, H. Li, Q. Jiang, Monocular 3D Object Detection: An Extrinsic Parameter Free Approach, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 7556-7566.

[70] S. Pang, D. Morris, H. Radha, Fast-CLOCs: Fast Camera-LiDAR Object Candidates Fusion for 3D Object Detection, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 187-196.

[71] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, Scalability in perception for autonomous driving: Waymo open dataset, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2446-2454.

[72] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, Scalability in perception for autonomous driving: Waymo open dataset, in: Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2446-2454.

[73] X. Chen, H. Ma, C. Zhu, X. Wang, Z. Zhao, Boundary-aware box refinement for object proposal generation, Neurocomputing, 219 (2017) 323-332.

[74] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580-587.

[75] X. Chen, K. Kundu, Z. Zhang, H. Ma, S. Fidler, R. Urtasun, Monocular 3d object detection for autonomous driving, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2147-2156.

[76] J. Hosang, R. Benenson, P. Dollár, B. Schiele, What makes for effective detection proposals?, IEEE transactions on pattern analysis and machine intelligence, 38 (2015) 814-830.

[77] F. Yu, D. Wang, E. Shelhamer, T. Darrell, Deep layer aggregation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 2403-2412.

[78] R. Qian, X. Lai, X. Li, Boundary-Aware 3D Object Detection from Point Clouds, arXiv preprint arXiv:2104.10330, (2021).

[79] Z. Miao, J. Chen, H. Pan, R. Zhang, K. Liu, P. Hao, J. Zhu, Y. Wang, X. Zhan, PVGNet: A Bottom-Up One-Stage 3D Object Detector With Integrated Multi-Level Features, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3279-3288.

[80] X. Liu, N. Xue, T. Wu, Learning Auxiliary Monocular Contexts Helps Monocular 3D Object Detection, arXiv preprint arXiv:2112.04628, (2021).

[81] G. Brazil, G. Pons-Moll, X. Liu, B. Schiele, Kinematic 3d object detection in monocular video, in: European Conference on Computer Vision, Springer, 2020, pp. 135-152.

[82] Q. Lian, B. Ye, R. Xu, W. Yao, T. Zhang, Geometry-aware data augmentation for monocular 3D object detection, arXiv preprint arXiv:2104.05858, (2021).

[83] H. Wang, Y. Cong, O. Litany, Y. Gao, L.J. Guibas, 3DIoUMatch: Leveraging iou prediction for semi-supervised 3d object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14615-14624.

[84] Z. Zou, X. Ye, L. Du, X. Cheng, X. Tan, L. Zhang, J. Feng, X. Xue, E. Ding, The devil is in the task: Exploiting reciprocal appearance-localization features for monocular 3d object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 2713-2722.

[85] C.-H. Wang, H.-W. Chen, L.-C. Fu, VPFNet: Voxel-Pixel Fusion Network for Multi-class 3D Object Detection, arXiv preprint arXiv:2111.00966, (2021).

[86] Z. Liu, D. Zhou, F. Lu, J. Fang, L. Zhang, Autoshape: Real-time shape-aware monocular 3d object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 15641-15650.

[87] S. Shi, X. Wang, H. Li, Pointrcnn: 3d object proposal generation and detection from point cloud, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 770-779.

[88] Y. Chen, S. Liu, X. Shen, J. Jia, Fast point r-cnn, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 9775-9784.

[89] Z. Wang, K. Jia, Frustum convnet: Sliding frustums to aggregate local point-wise features for amodal 3d object detection, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2019, pp. 1742-1749.

[90] S. Shi, C. Guo, L. Jiang, Z. Wang, J. Shi, X. Wang, H. Li, Pv-rcnn: Point-voxel feature set abstraction for 3d object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10529-10538.

[91] R. Qian, X. Lai, X. Li, Boundary-Aware 3D Object Detection from Point Clouds, arXiv preprint arXiv:2104.10330, (2021).

[92] Y. Zhang, J. Lu, J. Zhou, Objects are Different: Flexible Monocular 3D Object Detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3289-3298.

[93] Z. Miao, J. Chen, H. Pan, R. Zhang, K. Liu, P. Hao, J. Zhu, Y. Wang, X. Zhan, PVGNet: A Bottom-Up One-Stage 3D Object Detector With Integrated Multi-Level Features, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3279-3288.

[94] X. Liu, N. Xue, T. Wu, Learning Auxiliary Monocular Contexts Helps Monocular 3D Object Detection, arXiv preprint arXiv:2112.04628, (2021).

[95] J. Lei, T. Guo, B. Peng, C. Yu, Depth-Assisted Joint Detection Network For Monocular 3d Object Detection, in: 2021 IEEE International Conference on Image Processing (ICIP), IEEE, 2021, pp. 2204-2208.

[96] Y. Li, S. Yang, Y. Zheng, H. Lu, Improved Point-Voxel Region Convolutional Neural Network: 3D Object Detectors for Autonomous Driving, IEEE Transactions on Intelligent Transportation Systems, (2021).

[97] S. Shi, L. Jiang, J. Deng, Z. Wang, C. Guo, J. Shi, X. Wang, H. Li, PV-RCNN++: Point-Voxel Feature Set Abstraction With Local Vector Representation for 3D Object Detection, arXiv preprint arXiv:2102.00463, (2021).

[98] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[99] Q. Lian, B. Ye, R. Xu, W. Yao, T. Zhang, Geometry-aware data augmentation for monocular 3D object detection, arXiv preprint arXiv:2104.05858, (2021).

[100] S.R. Bulo, L. Porzi, P. Kontschieder, In-place activated batchnorm for memory-optimized training of dnns, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5639-5647.

[101] X. Chen, H. Ma, Learning a compact latent representation of the bag-of-parts model, in: 2014 IEEE International Conference on Image Processing (ICIP), IEEE, 2014, pp. 5926-5930.

[102] N.A.M. Mai, P. Duthon, L. Khoudour, A. Crouzil, S.A. Velastin, Sparse LiDAR and Stereo Fusion (SLS-Fusion) for Depth Estimationand 3D Object Detection, arXiv preprint arXiv:2103.03977, (2021).

[103] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? the kitti vision benchmark suite, in: 2012 IEEE conference on computer vision and pattern recognition, IEEE, 2012, pp. 3354-3361.

[104] Y. Xiang, W. Choi, Y. Lin, S. Savarese, Subcategory-aware convolutional neural networks for object proposals and detection, in: 2017 IEEE winter conference on applications of computer vision (WACV), IEEE, 2017, pp. 924-933.

[105] F. Yu, D. Wang, E. Shelhamer, T. Darrell, Deep layer aggregation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 2403-2412.

[106] P. Li, H. Zhao, P. Liu, F. Cao, Rtm3d: Real-time monocular 3d detection from object keypoints for autonomous driving, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16, Springer, 2020, pp. 644-660.

Emotion Estimation Method with Mel-frequency Spectrum, Voice Power Level and Pitch Frequency of Human Voices through CNN Learning Processes

Taiga Haruta[1], Mariko Oda[2], Kohei Arai[3]

Graduate School of Engineering, Kurume Institute of Technology, Kurume City, Japan[1, 2]

Applied AI-Research Laboratory, Kurume Institute of Technology, Kurume City, Japan and Saga University[3]

Abstract—**Emotion estimation method with Mel-frequency spectrum, voice power level and pitch frequency of human voices through CNN (Convolutional Neural Network) learning processes is proposed. Usually, frequency spectra are used for emotion estimation. The proposed method utilizes not only Mel-frequency spectrum, but also voice pressure level (voice power level) and pitch frequency to improve emotion estimation accuracy. These components are used through CNN learning processes using training samples which are provided by Keio University (emotional speech corpus) together with our own training samples which are collected by our students in emotion estimation processes. In these processes, the target emotion is divided into two categories, confident and non-confident. Through experiments, it is found that the proposed method is superior to the traditional method with only Mel-frequency by 15%.**

Keywords—*e-Learning; emotion estimation; Mel-frequency spectrum; fundamental frequency (pitch frequency); sound pressure level (voice power level)*

I. INTRODUCTION

Since the spread of the new coronavirus, many schools have introduced e-learning, a learning system using information and communication technology (ICT) that allows students to study independently.

Unlike regular classes for many students, e-learning has the advantage that a history is kept for everyone, and learning can proceed according to the student's level of proficiency. However, self-study through e-learning is generally a challenge for students to maintain their motivation to learn. Children and students with intellectual disabilities need psychological support advice beside the disabled child and support to change the questions according to the learner's situation.

Because of "Annual Report on Government Measures for Persons with Disabilities (Summary) 2020", the total number of mentally retarded persons under the age of 18 is 225,000 [1]. In the Covid-19 pandemic, we believe it is necessary to develop an e-Learning system that considers disabilities to realize "fair, individualized, and optimized learning that leaves no one behind" for a diverse group of children.

As a result of speech change by emotion, it has a significant effect on the continuous speech estimation engine. These effects are caused by the fact that the acoustic models of speech estimation engines currently in general use are designed

assuming calm speech. It is necessary to take measures such as switching the acoustic model every time.

Emotion estimation from speech can be divided into methods that use information obtained from the utterance content (context), that is, methods that perform speech estimation and use the results, and methods that use signal processing methods such as prosody, amplitude, and stress. However, it has already been mentioned that the former is difficult to use as the first stage of a large-vocabulary continuous speech estimation system in terms of computational complexity. Therefore, in order to actually estimate emotions, it is realistic to use the latter method of extracting features from emotional speech and estimating emotions. Methods using neural networks and methods using discriminant analysis are conceivable for estimating emotions using the extracted feature values.

It is important that the feature values used in this method reflect the differences between emotions well, and various feature values have been studied so far. It appears in three aspects: prosodic structure such as frequency and formant frequency, amplitude structure such as sentence stress and accent, and temporal structure such as sentence length. After the speech data is transformed into data in the frequency domain by Discrete Fourier Transformation (DFT), it is then subjected to inverse DFT by taking the logarithm. The basic idea of this fundamental frequency extraction method is to select a certain number in descending order and connect these peaks smoothly by Dynamic Programming (DP).

Although the traditional methods for emotion estimation use a combination among the aforementioned features, the most commonly used feature is frequency components (processed from them), and has long been utilized in speech recognition research [2], also the estimation accuracy is not all good. In order to improve the accuracy, the feature combination among frequency components (Mel-Frequency), voice pressure level and pitch frequency is proposed in this paper. One of the targeted applications of the proposed emotion estimation method is a confidence level estimation of students through remote lectures. Actually, interactive lighting using voice emotion recognition is actually being studied [3]. The final goal of this study is the development of learning support system which Artificial Intelligence (AI) plays the role of teacher to understand the learner's emotions and prompt their self-discipline learning. This paper makes clear how can

we determine whether they are confident in their answers from their utterances during learning.

The related research works are described in the following section. After that, the proposed method for emotion estimation with voices is described followed by experiments. Then conclusion is described with some discussions.

II. Related Research Works

Although voice recognition is now world widely available, recognition performance is not good enough for normal conversations. For instance, voice recognition performance of the typical Hidden Markov Model: HMM based method [4] (this is referred to the conventional voice recognition hereafter) with the feature of Formant is less than 50 % when the signal to noise ratio is below 5dB. In other words, voice recognition performance is totally affected by noise. In normal conversation among us, not only voice but also mouth movement is used for recognitions. Mouth movement video analysis makes voice recognition much better performance. The lip-reading method is for improvement of voice recognition performance.

Usually, Hidden Markov Model based method or neural network-based method is used for voice recognitions as well as optical flow [5]-[12] based analysis of the mouth movement videos. Forward direction (from the past to the future) of optical flow is usually used for mouth movement analysis. Voice recognition performance can be improved by adding backward direction (from the future to the past) of optical flow for correction of voice recognition errors through a confirmation of recognized results. In this process, two voice elements are treated as a unit for the proposed backward optical flow. The conventional forward direction of optical flow recognizes by voice element by voice element, though. In order to make sure the recognized results, two voice elements are much easier and efficient manner. This is because transient between voice element and voice element is so important for voice recognitions. This is the basic idea of the proposed lip-reading method.

As for the voice recognition methods, there are the following related research works,

E-learning system which allows students' confidence level evaluation with their voice when they answer to the questions during achievement tests is developed [13]. On the other hand, voice recognition method with mouth movement videos based on forward and backward optical flow is proposed and validated [14]. Mobile device based personalized equalizer for improving hearing capability of human voices for elderly persons is developed and realized [15]. Meanwhile, hearing aid method by equalizing frequency response of phoneme extracted from human voice is proposed and validated [16].

English pronunciation practice system using voice and video recognitions based on optical flow is developed and validated [17]. The eye based domestic helper robot allowing patient to be self-services through voice communications is also developed and validated its usefulness [18].

As mentioned in the previous section, speech features are essential for speech recognition. Recently, in the frequency component, Mel frequency magnitude coefficient was able to recognize emotions with up to 95.25% accuracy on a multi-class support vector machine [19].

Also, Unsupervised feature selection combining Gammatone Cepstral Coefficients (GTCC) and Power Normalized Cepstral Coefficients (PNCC) is used as speech features to recognize emotion from speech in the presence of various noises if the signal-to-noise ratio from -5 dB to 20 dB is 15 dB or higher [20].

After extracting some statistical features from the fundamental frequencies, we applied different AI methods to examine and study whether they have information about the speaker's emotional state, and found that 89.74% for two emotions, 76.14% for a set of three emotions, and 62.99% for a set of four emotions equal accuracy has been achieved [21].

III. Proposed Method

A. Method Configuration

As mentioned above in introduction section, three features combination, frequency components (Mel-Frequency), voice pressure level and pitch frequencywhich are extracted from the studnets voices and from the speech corpus is used in the emotion estimaion. Jupyter notebook provided by anaconda is used. The programming language used is Python. The following Deep Learning used in the emotion estimation, Convolution Neural Network: CNN + Rectified Linear Unit: ReLu + CNN + ReLu + Pooling + CNN + ReLu + CNN + ReLu + Pooling + CNN + ReLu + CNN + ReLu + Pooling + Fully Connected Layers.

B. Voice Data for Deep Learning

There are two datasets of training samples. One is voice data produced by 17 students of Kurume Institute of Technology to verbally answer "yes" or "no" to 20 questions which are shown in Fig. 1(a) in Japanese, including current events, riddles, and simple calculation problems, and used voice data recorded from their voices.

We also used Keio-ESD[1] [22], a speech corpus provided by the Speech Resources Consortium of the National Institute of Informatics, to perform a binary classification of the recorded speech into positive and negative categories.

C. Feature Extraction

This time, a comparison was made between a model trained with the Mel-frequency spectrogram alone and a model trained with the fundamental frequency and sound pressure level in addition to it. Therefore, each feature was extracted using Python programming, and the numerical data was stored in csv format and then in a Numpy array.

The resulting Mel-frequency spectrograms, fundamental frequencies, and sound pressure levels were scaled differently for each feature, so the scales were aligned by standardizing the data.

---

[1] Keio-ESD: Keio University Japanese Emotional Speech Database, http://research.nii.ac.jp/src/index.html, A set of human speech with vocal emotion spoken by a Japanese male speaker and a set of artificial speech that were synthesized by a system that had been developed using the subset of this database for training.

In addition, the number of feature data is proportional to the voice time. This is because the feature data itself is time series data. Therefore, all students recording data was trimmed every 0.5 seconds. In addition, the number of features for each of the 1 voice was matched to the fundamental frequency with the lowest number of features extracted. This is because if the number of features is not aligned, there will be a bias toward the features with the highest number of features during training. Fig. 1(b) is one of the audio waveform data recorded this time. Fig. 1(c) is the Mel-frequency spectrogram extracted from Fig. 1(b). Fig. 1(d) is the fundamental frequency extracted from Fig. 1(b). Fig. 1(e) is the sound pressure level extracted from Fig. 1(b), respectively.

15

$$\int e^x dx = e^x + C$$

である。ただしeはネイピア数、Cは積分定数とする。
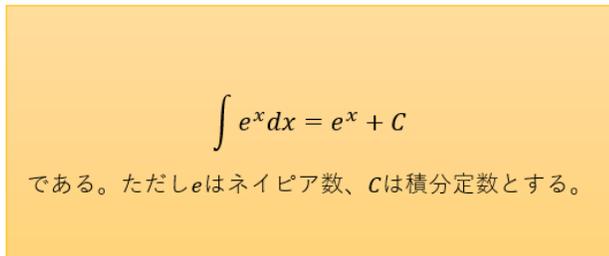
(a) Some of the questions that were asked
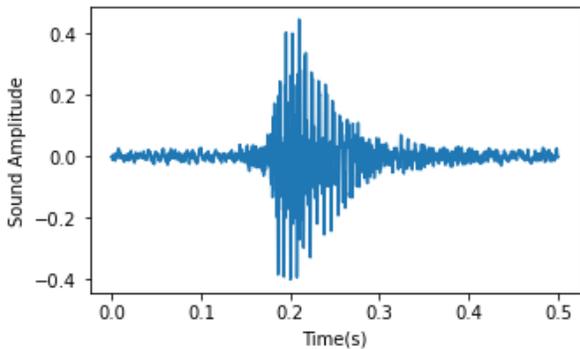
(b) One of the audio waveform data recorded this time.
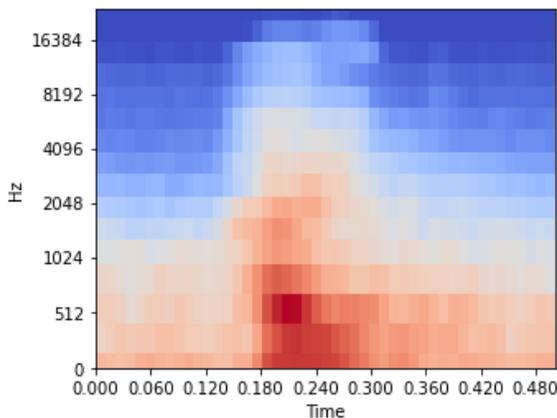
(c) The Mel-frequency spectrogram extracted from Fig. 1(b)
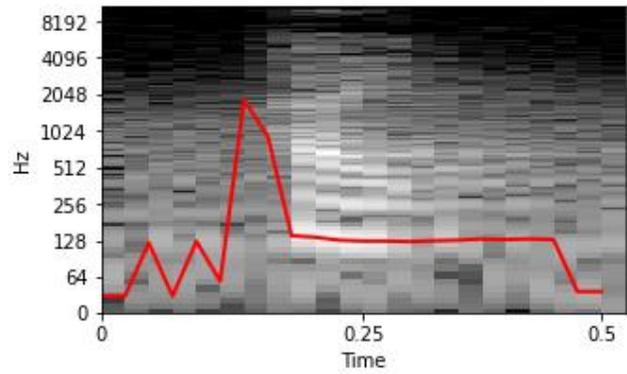
(d) The fundamental frequency extracted from Fig. 1(b)
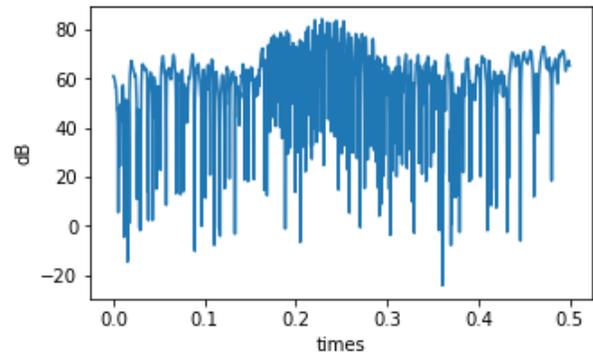
(e) The sound pressure level extracted from Fig. 1(b)

Fig. 1. A Portion of the Training Samples Extracted from the Students' Voices.

D. Preparation for Supervised Learning and Learning Model

The speech corpus "Keio-ESD" contains 20 words with 47 different emotions. Among those, we labelled 12 positives: "pleasure", "kind", "gentle", "tolerance", "glad", "admiration", "pride", "love", "satisfaction", "expectation", "happiness", and "like". Then we labelled 12 negatives: "repugnance", "complaint", "disappointment", "sorrowful", "fear", "indifferent", "lamentation", "boredom", "dislike", "Nope." "dejection", and "anxiety". We used the speech for the emotions "expectation," "happiness," "like," "Nope." "Dejection", and "anxiety" as test data and the speech data for the other emotions as training data. However, we did not use the last 20% of the recorded data for training but used it as validation data to measure the goodness of the model parameters after each training.

As for the students' confidence classification of the recorded data, they were asked to answer in advance their confidence after their answers. We then used the recorded data of 13 of the 17 students in our study data as training data. We then used the recordings of the remaining four individuals as test data. We did not use these data for the last 20% of recordings but used them as validation data too.

We used the Function API for CNN with Keras to build our learning model. We trained the Keio-ESD emotion classification 200 times and the student confidence classification 20 times, then evaluated each test data set and compared the results.

IV. EXPERIEMNTAL RESULTS

A. Binary Classification of Emotional Speech Corpus into Positive and Negative

As for the binary classification of speech into positive and negative for the emotional speech corpus, Fig. 2 shows the transition graph of the loss function for each epoch.
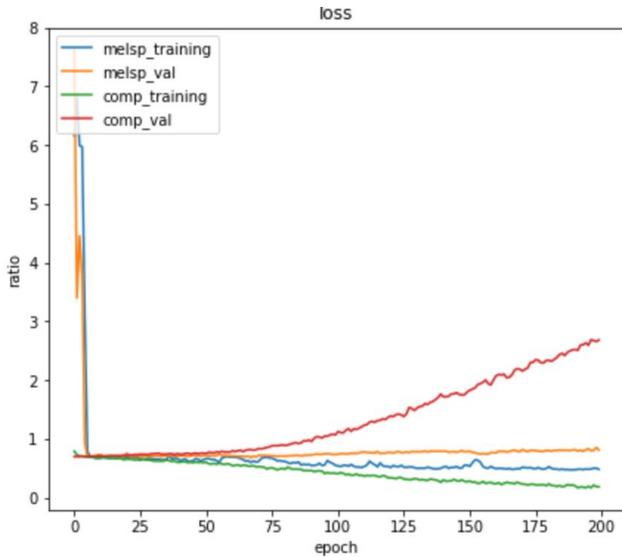
Fig. 2. The Loss Function of Binary Classification of Emotional Speech Corpus into Positive and Negative.

The vertical axis shows the percentage, and the horizontal axis shows the number of epochs. The precedents are discussed below:

- "melsp_training" shows the result of training with only the training data of the Mel-frequency spectrogram in the input layer.

- "melsp_val" shows the results that were checked against the validation data of the Mel-frequency spectrogram after each study.

- "comp_training" shows the result of training with only the training data of the Mel-frequency spectrogram, fundamental frequency, and sound pressure level in the input layer.

- "comp_val" shows the results that were checked against the validation data of the Mel-frequency spectrogram fundamental frequency, and sound pressure level after each study.

Figures are basically rounded to two decimal places.

Regarding the loss function, it was quite large immediately after the Mel-frequency spectrogram was put in the input layer, but after each epoch, melsp_training and melsp_val converged around 0.48 and 0.81, respectively. Concerning comp_training and comp_val, comp_training eventually converged around 0.18, comp_val eventually converged around 2.68.

Fig. 3 is a graph of the percentage of accuracy. With regard to the training data, both melsp_training and comp_training showed an upward trend with each epoch. Finally,

melsp_training and comp_val showed 0.76 and 0.93, respectively. In other words, comp_training is about 16% better at learning than melsp_training. As for the validation data, the melsp_val and comp_val finally converged to 0.46 and 0.43, respectively.
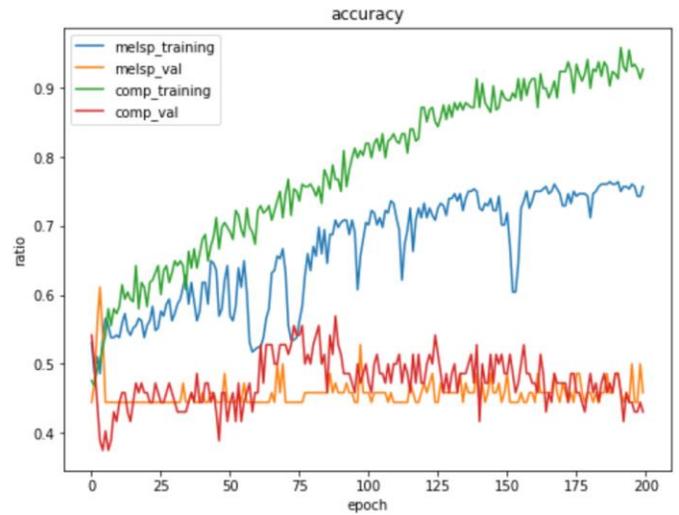
Fig. 3. The Accuracy of Binary Classification of Emotional Speech Corpus into Positive and Negative.

B. Confidence Classification of the Student Recorded Data

Fig. 4 is the loss function, and Fig. 5 is the accuracy. Regarding the loss function, as in the previous page, the values of the loss functions for melsp_training and melsp_data were quite large, but after the fourth training, they showed values below 1 and finally converged to 0.69. Regarding the validation data, comp_training eventually converged to 0.55 and comp_val converged to 0.83. comp_val exceeded melsp_val by binary classification of emotional speech corpus into positive and negative same amount.
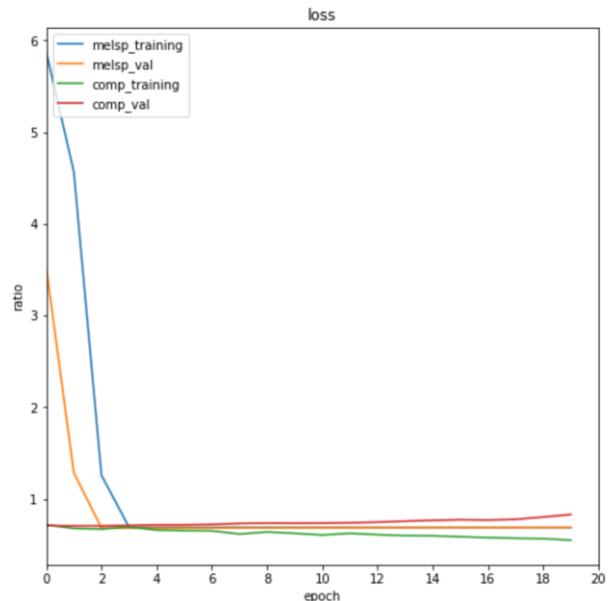
Fig. 4. The Loss Function of Confidence Classification of Student Recorded Data.
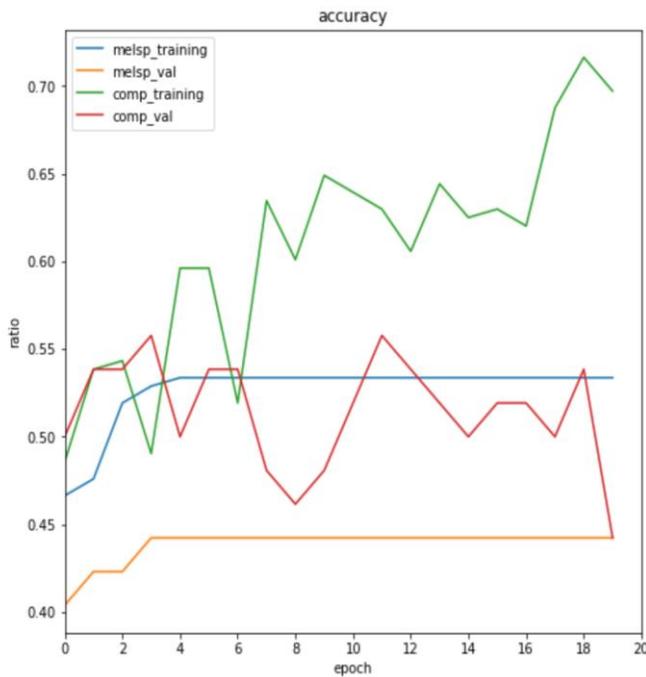
Fig. 5. The Accuracy of Confidence Classification of Student Recorded Data.

With respect to the accuracy, the value of melsp_training remained unchanged from 0.53 after the fifth training, and comp_training increased from 0.49 at the first training to a final value of 0.70. The value of melsp_val also remained unchanged from the fifth study and finally converged to 0.44. Comp_val showed a large difference in the percentage of correct responses from epoch to epoch and finally converged to 0.44.

C. Evaluation with Test Data

We finally evaluated models trained only on the Mel-frequency spectrogram, plus the fundamental frequency and sound pressure level, on test data for each classification. The table of it is Table I and Table II.

TABLE I. EVALUATION WITH TEST DATA (KEIO-ESD)

Binary classification of emotional speech corpus into positive and negative	
Test 1 loss function	1.765992283821106
Test 2 loss function	2.4016263484954834
Test1 accuracy	0.5083333253860474
Test2 accuracy	0.5333333611488342

TABLE II. EVALUATION WITH TEST DATA (CONFIDENCE CLASSIFICATION)

Confidence classification of the student recorded data	
Test 1 loss function	0.6947100162506104
Test 2 loss function	0.7546933889389038
Test1 accuracy	0.48750001192092896
Test2 accuracy	0.637499988079071

Note that Test 1 in the following table shows the results of evaluating the model with only the Mel-frequency spectrogram in the input layer, and Test 2 shows the results of evaluating the model trained with the Mel-frequency spectrogram, fundamental frequency, and sound pressure level.

The loss function for Keio-ESD's emotional voice classification was much higher than 1 for both Test 1 and Test 2, with a difference of about 0.64, while the loss function for the student's confidence classification was also lower than 1, and the difference in the loss function was about 0.06, indicating that the difference between the two functions had narrowed.

On the other hand, Test 2 outperformed Test 1 in both accuracies, with a 15% improvement in accuracy, especially in the classification of students' self-confidence.

V. CONCLUSION

In this study, to develop a learning support system for physically challenged children that applies speech and emotion recognition, which is an applied technology of AI, we attempted to discriminate the confidence of the respondent's speech using deep learning and to classify the emotion of the emotional speech corpus into two emotion values. In this connection, we constructed a model in which only the Mel-frequency spectrogram was learned as the speech feature charge, firstly. Then, not only Mel-frequency spectrum, but also another two types of models were learned: fundamental frequency and sound pressure level. A comparison of the change in the percentage of correct responses between the two models showed that using a model trained with three types of features improved accuracy by up to about 15%. Therefore, it may say that the proposed method is superior to the conventional method with only Mel-frequency spectrum by 15%.

FUTURE RESEARCH WORKS

Further investigation has to be made for improving voice-based emotion recognition accuracy with further experiments by the other testers. The other features related to voice recognition have to be added to the proposed method for improving the recognition accuracy.

REFERENCES

[1] Cabinet Office, Japan "Annual Report on Government Measures for Persons with Disabilities (Summary) 2020 (Japanese) " (2020) P241.

[2] Mengyao Zhang, Kai Cheng, (2021) "Considerations on Speech Feature Extraction for Humming Retrieval", The 83rd National Convention of Information Processing Society of Japan , 251-252,, Information Processing Society of Japan.

[3] Yuichiro Okado, Kumiko Kushiyama, Wataru Kurihara, (2021) "Kotodama : Using Emotion Recognition with Voice Input Interactive lighting production" Proceedings of the Entertainment Computing Symposium (JAPAN), 147-150.

[4] Hongbing Hu, Stephen A. Zahorian, (2010) "Dimensionality Reduction Methods for HMM Phonetic Recognition," ICASSP 2010, Dallas, TX,

http://bingweb.binghamton.edu/~hhu1/paper/Hu2010Dimensionality.pdf .(accessed on September 14 2012).

[5] Huston SJ, Krapp HG (2008). Kurtz, Rafael. ed. "Visuomotor Transformation in the Fly Gaze Stabilization System". PLoS Biology 6 (7): e173. doi:10.1371/journal.pbio.0060173. PMC 2475543. PMID 18651791. http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.00601 73.(accessed on September 14 2012).

[6] Andrew Burton and John Radford (1978). Thinking in Perspective: Critical Essays in the Study of Thought Processes. Routledge. ISBN 0416-85840-6. http://books.google.com/?id=CSgOAAAAQAAJ&pg= PA77&dq=%22o ptical+flow%22+%22optic+flow%22+date:0-1985.(accessed on September 14 2012).

[7] David H. Warren and Edward R. Strelow (1985). Electronic Spatial Sensing for the Blind: Contributions from Perception. Springer. ISBN 90-247-2689-1. http://books.google.com/?id=I_Hazgqx8QC&pg=PA4 14&dq=%22optical+flow%22+%22optic+flow %22+date:0-1985.(accessed on September 14 2012).

[8] S. S. Beauchemin , J. L. Barron (1995). The computation of optical flow. ACM New York, USA http://portal.acm.org/ft_gateway.cfm? id=212141&type=pdf&coll=GUI DE&dl=GUIDE&CFID=72158298& CFTOKEN=85078203.(accessed on September 14 2012)

[9] David J. Fleet and Yair Weiss (2006). "Optical Flow Estimation". In Paragios et al.. Handbook of Mathematical Models in Computer Vision. Springer. ISBN 0-387-26371-3. http://www.cs.toronto.edu/~fleet/ research/Papers/flowChapter05.pdf.(ac cessed on September 14 2012)

[10] John L. Barron, David J. Fleet, and Steven Beauchemin (1994). "Performance of optical flow techniques". International Journal of Computer Vision (Springer). http://www.cs.toronto.edu/~fleet/research/ Papers/ijcv-94.pdf.(accessed on September 14 2012)

[11] B. Glocker, N. Komodakis, G. Tziritas, N. Navab & N. Paragios (2008). Dense Image Registration through MRFs and Efficient Linear Programming. Medical Image Analysis Journal. http://vision.mas.ecp.fr/pub/mian08.pdf.(accessed on September 14 2012)

[12] Christopher M. Brown (1987). Advances in Computer Vision. Lawrence Erlbaum Associates. ISBN 0-89859-648-3. http://books.google.com/?id=c97huisjZYYC&pg=PA133&dq=%22optic +flow%22++%22optical+flow%22. (accessed on September 14 2012)

[13] Kohei Arai, E-learning system which allows students' confidence level evaluation with their voice when they answer to the questions during achievement tests, International Journal of Advanced Computer Science and Applications, 3, 9, 80-84, 2012.

[14] Kohei Arai, Voice recognition method with mouth movement videos based on forward and backward optical flow, International Journal of Advanced Research in Artificial Intelligence, 2, 2, 48-52, 2013.

[15] Kohei Arai, Takuto Konishi, Mobile device based personalized equalizer for improving hearing capability of human voices in particular for elderly persons, International Journal of Advanced Research on Artificial Intelligence, 4, 6, 23-27, 2015.

[16] Kohei Arai, Takuto Konishi, Hearing aid method by equalizing frequency response of phoneme extracted from human voice, International Journal of Advanced Computer Science and Applications IJACSA, 8, 7, 88-93, 2017.

[17] Kohei Arai, Shinji Matsuda and Mariko Oda, English pronounciation practice system using voice and video recognitions based on optical flow, Proceedings of the International Conference on Information Technology Based Higher Education and Training, Kumamoto, 2001.

[18] Kohei Arai, R. Mardiyanto, The eye based domestic helper robot allowing patient to be self-services through voice communications, Proceedings of the 260th conference in Saga of Image and Electronics Engineering Society of Japan, 139-142, 2012.

[19] J. Ancilin , A. Milton, (2021) "Improved speech emotion recognition with Mel frequency magnitude coefficient", Applied Acoustics Volume 179, article 108046

[20] Surekha Reddy Bandela, T. KishoreKumar, (2021) "Unsupervised feature selection and NMF de-noising for robust Speech Emotion Recognition", Applied Acoustics Volume 172, article 107645

[21] Teodora DIMITROVA-GREKOW, Aneta KLIS, Magdalena IGRAS-CYBULSKA (2019), "Speech Emotion Recognition Based on Voice Fundamental Frequency", Archives of Acoustics, 44, 2, pp. 277–286,

[22] Tsuyoshi Moriyama, Shinya Mori, Shinji Ozawa, A Synthesis Method of Emotional Speech Using Subspace Constraints in Prosody, Journal of Information Processing, pp.1181-1191, Vol.50, No.3, 2009.

AUTHORS' PROFILE

Taiga Haruta: He received BE degree in 2022. He also receives the Kurume Institute of Technology President's Award. He is now working on AI-applied e-Learning research and voice recognition in Master's Program at Kurume Institute of Technology.

Mariko Oda: She graduated from the Faculty of Engineering, Saga University in 1992, and completed her master's and doctoral studies at the Graduate School of Engineering, Saga University in 1994 and 2012, respectively. She received Ph.D(Engineering) from Saga University in 2012. She also received the IPSJ Kyushu Section Newcomer Incentive Award.In 1994, she became an assistant professor at the department of engineering in Kurume Institute of Technology; in 2001, a lecturer ; from 2012 to 2014, an associate professor at the same institute; from 2014, an associate professor at Hagoromo university of International studies; from 2017 to 2020, a professor at the Department of Media studies, Hagoromo university of International studies. In 2020, she was appointed Deputy Director and Professor of the Applied of AI Research Institute at Kurume Institute of Technology. She has been in this position up to the present.She is currently working on applied AI research in the fields of education.

Kohei Arai: He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January 1979 to March 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post-Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science in April 1990. He is now an Emeritus Professor of Saga University since 2014. He was a council member for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is a Science Council of Japan Special Member since 2012. He is an Adjunct Professor of University of Arizona, USA since 1998 and is an Adjunct Professor of Nishi-Kyushu University as well as Kurume Institute of Technology/AI Application Laboratory since 2021. He also is Vice Chairman of the Science Commission "A" of ICSU/COSPAR since 2008 then he is now award committee member of ICSU/COSPAR. He wrote 77 books and published 678 journal papers as well as 550 conference papers. He received 66 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA and IJISA. http://teagis.ip.is.saga-u.ac.jp/index.html

Cybersecurity in Deep Learning Techniques: Detecting Network Attacks

Shatha Fawaz Ghazal[1]

Department of Administrative Sciences, Finance and
Computer, Al- Balqa Applied University (BAU)
Zarqa, Jordan

Salameh A. Mjlae[2]

Prince Abdullah bin Ghazi Faculty of Information
Technology, Al- Balqa Applied University (BAU)
AL- Salt, Jordan

Abstract—**Deep learning techniques have been found to be useful in a variety of fields. Cybersecurity is one such area. In cybersecurity, both Machine Learning and Deep Learning classification algorithms can be used to monitor and prevent network attacks. Additionally, it can be utilized to identify system irregularities that may signal an ongoing attack. Cybersecurity experts can utilize machine learning and deep learning to help make systems safer. Eleven classification techniques, including eight machine learning algorithms (Decision Tree, Random Forest, and Gradient Boosting) and one statistical technique, were employed to examine the popular HTTP DATASET CSIC 2010. (K-Means). Along with XGBoost, AdaBoost, Multilayer Perceptrons, and Voting, three deep learning algorithms are Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and LSTM plus CNN. To evaluate the performance of such models, precision, accuracy, f1-score, and recall are often used metrics. The results showed that when comparing the three deep learning algorithms by the aforementioned metrics, the LSTM with CNN produced the best performance outcomes in this paper. These findings will show that our use of this algorithm allows us to detect multiple attacks and defend against any external or internal threat to the network.**

Keywords—HTTP DATASET CSIC 2010; deep learning; cybersecurity attacks; detection attacks; network attacks

I. INTRODUCTION

Cybersecurity (Cyber_Security) [1] protects computers, servers, and networks from intrusion, theft, and other forms of harm to their data, hardware, and software, as well as from interruptions in service or misuse of the resources they provide. Our reliance on computers, the Internet, and wireless network standards like Bluetooth and Wi-Fi, as well as the proliferation of "smart" devices like smartphones, televisions, and the myriad other devices that make up the Internet of things, have all contributed to the Internet of Things growing significance (IOT) [1]. Due to the complication of its political application and technological implementation, cybersecurity is one of the greatest issues of the twenty-first century. The system's key objectives are dependable operation, integrity of data, and safeguarding of sensitive information. Due to the complication of its political application and technological implementation, cybersecurity is one of the greatest issues of the twenty-first century. The system's key objectives are dependable operation, intact data, and the security of sensitive information [2].

A cyber-attack is any hostile attempt to hack an IT system or its users in order to gain unauthorized access to the system's data or information [3]. Cyber attackers are often crooks looking to profit financially from the attack in the great majority of instances. In other cases, the intention is to stop operations by preventing users from accessing IT systems or destroying physical equipment [3]. State agents or cybercriminals working for them are frequently involved in state-sponsored cybercrime. Cyberattacks can be wide in scope, impacting a number of businesses across several regions and nations, or they can be directed at specific companies or persons [4]. Targeted attacks frequently expand beyond their intended targets, posing a threat to all firms. A state-sponsored strike against Ukrainian banks and utilities most likely caused the global NotPetya outbreak in June 2017. The clean-up achieved the desired effect on Ukraine, but it also had a worldwide impact, costing almost $10 billion in IT system recovery and lost productivity, according to publications documenting the cleanup [2, 3, and 4].

Deep learning can be classified as a sort of machine learning. Research into algorithms is crucial to the growth and development of this discipline. Deep learning employs artificial neural networks created to simulate human cognitive processes, whereas machine learning focuses on generalization and empirical data. Until recently, the available computational power limited the complexity of neural networks [5]. Systems in the cybersecurity sector may use deep learning to recognize and learn from patterns in order to foil future attacks and adapt to evolving adversarial tactics. It can improve the speed with which cybersecurity teams respond to actual attacks and mitigate risks [5]. Because it reduces the amount of time spent on routine tasks, it allows businesses to better allocate their resources. In conclusion, deep learning can make cybersecurity easier, more preventative, cheaper, and more effective [4, 5]. However, this will only be possible if the data that drives machine learning provides an accurate depiction of the setting. As the adage goes, "if you put garbage in, you'll get garbage out." [5].

The following are some key takeaways from this paper that outline its contribution:

1) This thesis makes use of a publicly available dataset pertinent to the topic at hand, allowing a wide range of classification algorithms to be put to use.

2) This category includes deep learning and machine learning. Building a robust model to recognize different network assaults utilizing a variety of machine learning and deep learning methods.

Examining the ability of each method to distinguish between the numerous network assaults that have been documented in the dataset. Several classification techniques, including machine learning and deep learning algorithms, are applied to a well-known network intrusion dataset in this study in order to determine whether or not each sample is typical and to identify attacks that are not typical. The remainder of this study is organized as follows: The second component of this paper provides a bibliography of relevant archival materials for this investigation. The methods employed are thoroughly detailed in Section III. Some examples are provided to illustrate the results in Section IV. Section V presents our conclusion and suggestions for future research.

II. RELATED WORK

Several authors applied machine learning algorithms in the cybersecurity field to detect different types of attacks based on real-time datasets or existing datasets from several resources. Kim et al. [6] using a Convolutional Neural Network (CNN) with long short-term memory (LSTM) and a Deep Neural Network (DNN), the sample was identified as normal or abnormal (DNN). They used the 61,065 instances and 16 features from the HTTP DATASET CSIC 2010 that were classified as either normal (36,000 samples) or anomalous (25,065 samples). The results demonstrated that the CNN equipped with LSTM was 91.54 percent accurate. Vartouni et al. [7] used a novel algorithm known as a Stacked Auto-Encoder to determine whether the sample was typical or out of the ordinary. They used the 61,065 instances and 16 features from the HTTP DATASET CSIC 2010 that were classified as either normal (36,000 samples) or anomalous (25,065 samples). According to the findings, the Stacked Auto-Encoder was able to achieve an accuracy of 88.32 percent.

Betarte et al. [8] The sample was classified as normal or abnormal using three distinct machine learning techniques: Random Forest, K-Nearest Neighbors (K-NN) with a k value of three, and Support Vector Machine (SVM). They utilized the HTTP DATASET CSIC 2010 dataset for their analysis, which covers normal (36,000 samples) and abnormal (61,065 cases) categories (25,065 samples). Compared to the other methods, Random Forest achieved 91.54 percent more accuracy. Tuan et al. [9] examined their five machine learning techniques using the common dataset UNSW-NB15, which includes numerous types of network assaults.

This dataset consists of nine separate attack types (DoS, Reconnaissance, Backdoor, Fuzzers, Analysis, Exploits, Worms, Shellcode, and Generic) and common attacks (44 characteristics). The machine learning algorithms are SVM, ANN, Naive Bayes (NB), and Unsupervised Learning (USML). Using this dataset, they demonstrated the USML's high degree of precision (94.78 percent). Anwer et al. [10] used four machine learning techniques to detect malicious network traffic based on a well-known dataset. They used a popular cybersecurity dataset named NSL-KDD that has 148,517 samples divided into training (125,973) and testing (22,544) datasets. There are a total of five classes in this dataset, split evenly between normal attacks and non-normal attacks, with subclasses within each subclass. For example, DoS attacks (Smurf, Back, Land, Processstable, Neptune, Pod, Apache2, Udpstorm, Worm, Teardrop) and R2L attacks (Xsnoop, Guess Password, Named, Ftp write, Phf, Multihop, Imap, Warezmaster, Warezclient, Snmpgetattack, Spy, Xlock, Snmpguess, Httptunnel, Sendmail) are all in existence. Machine learning algorithms include things like the Support Vector Machine (SVM), Gradient Boosted Decision Trees (GBDT), and Random Forest (RF). These algorithms were evaluated on four fronts: accuracy, specificity, training time, and prediction time. The 85.34 percent accuracy achieved by RF was the highest of any of the tested algorithms.

Su et al. [11] presented the BAT model, a deep learning technique for spotting hostile network infiltration. Utilization of the NSL-KDD dataset, which is commonly utilized in the investigation of network intrusion. The same number of data points were used for training and testing, totaling 125,973 points (22,544). Each of the five classes in this dataset is further subdivided into two additional groups: normal attacks and non-normal attacks. To name just a few examples, there are denial-of-service (DoS) attacks (Smurf, Back, Land, Processstable, Neptune, Pod, Apache2, Udpstorm, Worm, Teardrop), reconnaissance-to-leak (R2L) attacks (Xsnoop, Guess Password, Named, Ftp write, Phf, Multihop, Imap, Warezmaster, Warezclient, Snmpgetattack, Spy, Xlock The intrusion detection accuracy of this model was 84.25%.

Xu et al. [12] proposed a new model based on a five-layer autoencoder (AE) that is more adept at detecting network anomalies. Because of its prominence in the realm of network assault, the NSL KDD dataset was deployed. The same number of data points were used for training and testing, totaling 125,973 points (22,544). Each of this dataset's five classifications is then separated into two additional groups: normal attacks and non-normal attacks. There are denial-of-service (DoS) assaults (Smurf, Back, Land, Processstable, Neptune, Worm, Pod, Apache2, UDPstorm, and others). Teardrop), reconnaissance-to-leak (R2L) attacks (Xsnoop, Guess Password, Named, Ftp write, Phf, Multihop, Imap, Warezmaster, Warezclient, Snmpgetattack, Spy, Xlock), and reconnaissance-to-attack (R2A) attacks (Xsnoop, Guess Password, Named, Ftp write, Phf, Multihop, I. The study demonstrated that the model's accuracy in identifying intrusions is 90.61 percent.

Kavitha et al. [13] presented a new technique for network intrusion detection based on One Class Support Vector Machine (OCSVM). They utilized the popular NSL-KDD dataset, which consists of a total of 148,517 samples split evenly between a training set of 125,973 samples and a testing set of 22,544 samples. This dataset has a total of five classes, evenly divided between normal and non-normal attacks, with subclasses inside each subclass. New technique for network intrusion detection based on One Class Support Vector Machine (OCSVM). They utilized the popular NSL-KDD dataset, which consists of a total of 148,517 samples split evenly between a training set of 125,973 samples and a testing

set of 22,544 samples. This dataset has a total of five classes, evenly divided between normal and non-normal attacks, with subclasses inside each subclass. To name just a few examples, there are denial-of-service (DoS) attacks (Smurf, Back, Land, Processtable, Neptune, Pod, Apache2, Udpstorm, Worm, Teardrop), reconnaissance-to-leak (R2L) attacks (Xsnoop, Guess Password, Named, Ftp write, Phf, Multihop, Imap, Warezmaster, Warezclient, Snmpgetattack, Spy, Xlock. The model was found to be 81.29 percent accurate when used for intrusion detection.

Ferriyan et al. [14] developed several machine learning models for detecting cyberattacks, and used a new dataset called ALLFLOWMETER HIKARI2021. Background, Benign, Bruteforce, Bruteforce-XML, Probing, and XMRIGCC CryptoMiner are the six types of attacks represented among the 555,278 instances and 86 features extracted by Zeek [https://zeek.org/] in this dataset. The normal attacks include Bruteforce, Bruteforce-XML, Probing, and XMRIGCC CryptoMiner, while the malicious attacks include Background and Benign. KNN, SVM, RF, and MultiLayer Perceptron are the ML models to use (MLP). They demonstrated that these models can achieve detection accuracy of roughly 0.99 percent as shown in Table I.

TABLE I.   PREVIOUS RELATED WORK FOR CYBERSECURITY ATTACKS DETECTION

Ref	Year	Attack	Cybersecurity Dataset	Algorithm	Result
[6]	2020	Normal Anomalous	CSIC 2010 dataset	CNN and LSTM DNN	CNN and LSTM accuracy = 91.54%
[7]	2018	Normal Anomalous	CSIC 2010 dataset	Stacked Auto-Encoder	Accuracy = 88.32
[8]	2018	Normal Anomalous	CSIC 2010 dataset	Random Forest KNN-3 SVM	Random Forest accuracy = 72%
[9]	2019	• Analysis • Reconnaissance • DoS • Exploits • Fuzzers • Generic • Normal • Worms • Backdoor • Shellcode	UNSW-NB15	SVM, ANN, NB, DT, and USML	USML accuracy = 94.78%.
[10]	2021	• R2L • DoS • U2R • Probe • Normal	NSL-KDD	SVM, GBDT, and RF	RF Accuracy = 85.34%
[11]	2020	• R2L • DoS • U2R • Probe • Normal	NSL-KDD	BAT model	Accuracy = 84.25%
[12]	2021	• R2L • DoS • U2R • Probe • Normal	NSL-KDD	5-layer autoencoder (AE)-based model	Accuracy = 90.61%
[13]	2021	• R2L • DoS • U2R • Probe • Normal	NSL-KDD	One Class SVM	Accuracy = 81.29%
[14]	2021	• Background • Benign • Bruteforce • Bruteforce-XML • Probing • XMRIGCC CryptoMiner	ALLFLOWMETER HIKARI2021	KNN, SVM, RF, and MLP	Accuracy = 0.99

III. PROPOSED METHODOLOGY

Fig. 1 depicts the proposed strategy utilized in this study to identify threats in a wide variety of cybersecurity datasets. The subsequent sections demonstrate how the proposed method operates in practice.

A. Dataset Description

In this study, we utilized the widely-used and exhaustive HTTP DATASET CSIC 2010 dataset, which comprises a variety of cyberattack kinds. The types of cyberattacks, the total number of Instances in the respective datasets, and the total number of characteristics are detailed in Table II.

Our department's online store's traffic is logged in the HTTP dataset CSIC 2010 for analysis. Registering with this web app requires the input of personal data, and once registered, users can make purchases from a cart interface. Due to the fact that it is a Spanish-language website, the collected information may contain some Latin characters [15]. In the dataset, 36,000 requests are legitimate, while another 25,000 were created fraudulently by bots. Threats like cross-site scripting, CRLF injection, server-side inclusion, parameter manipulation, and SQL injection are all part of the package. There have been successful web detection experiments using this dataset [15]. Fig. 2 depicts the frequency of attacks on the HTTP dataset CSIC 2010, while Table III details the characteristics of this dataset.

B. Preparing Dataset

In order to apply different deep learning algorithms to each dataset, we encode their non-numerical properties into numerical features using a standard method [16]. Label Encoder is a technique for converting data that isn't numerical into a form that computers can understand by assigning each value a number starting at zero [16]. All the features in this dataset are categories, so we need to transform them into numbers. We used the Hold out method to divide the data set. As a result, in our experiments, we use only 0.20 percent of the total dataset for evaluation purposes, while the remaining 0.80 percent serves as training data.

C. Classification Algorithms

In this paper, we detail the detection procedure and the classification algorithms that enabled it. Some machine learning and deep learning formula examples are provided.

TABLE II. INFORMATION OF CYBERSECURITY DATASETS

Dataset Name	No. of Instances	No. of Features	Cybersecurity Attacks	
HTTP DATASET CSIC 2010	61,065	16	Normal	36,000
			Malicious	25,065

TABLE III. FEATURES OF THE HTTP DATASET CSIC 2010.

No.	Feature
1	Method
2	User-Agent
3	Pragma
4	Cache-Control
5	Accept
6	Accept-encoding
7	Accept-charset
8	language
9	host
10	cookie
11	content-type
12	connection
13	length
14	content
15	URL
16	label
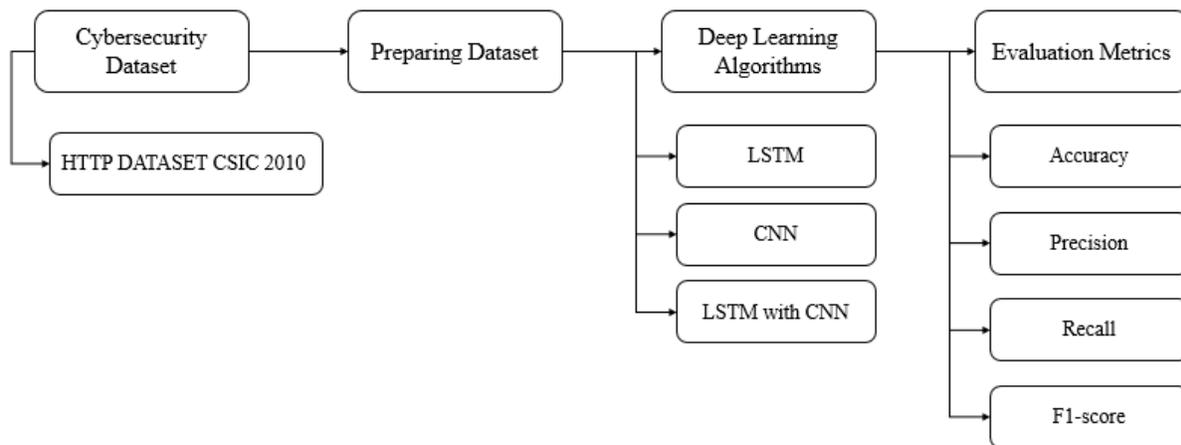
Fig. 1. Illustrate Flow-chart of the Proposed Methodology.
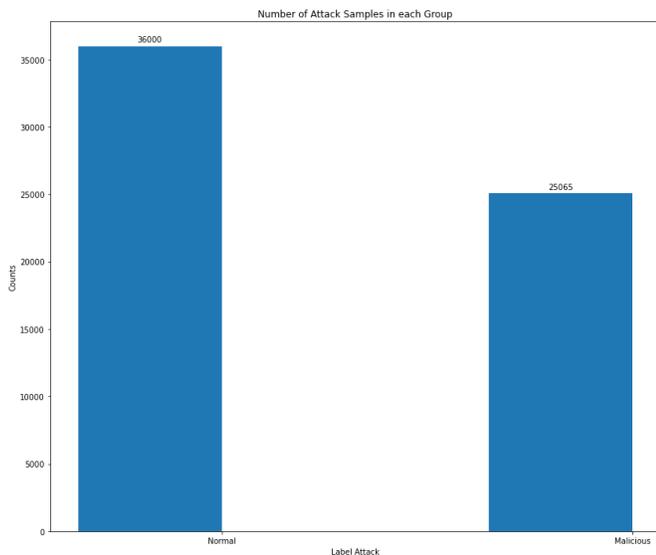
Fig. 2.   Attacks Frequency in HTTP Dataset CSIC 2010.

1) Machine learning algorithms: After preparing the dataset, it is fitted to seven machine learning algorithms, each with a test size of 0.1, in order to detect the aforementioned cybersecurity attacks. This was accomplished by utilizing a hold-out approach to divide the dataset into training and testing datasets, with 0.9% of the total datasets acting as training datasets and 0.1% serving as testing datasets. Multiple machine learning models are constructed utilizing the training dataset, and their efficacy is assessed utilizing the testing dataset.

- Random Forest Algorithm (RF)

Random forests, an ensemble supervised learning method, trains many decision trees to perform regression and classification. The most commonly selected class by trees is the one that is expected to be the solution to classification problems. When it comes to regression jobs, we give you the typical prediction from all of our trees [17]. Overfitting the training set is a common problem for decision trees; random decision forests are capable of mitigating this effect. Therefore, we utilize the Random Forest classifier to categorize each cyber-attack dataset [18, 19]. Due to the discontinuous nature of our experiment's label, we employed a random forest of the classification variety with the following parameters: n_estimators (number of decision trees) = 100, max_features = sqrt, max_depth = None, random_state = 42.

- Decision Tree Algorithm (DT)

Decision Tree is a supervised machine learning system that, similar to people, makes decisions based on predefined criteria. [20]. The fields of data mining, statistics, and machine learning all use decision tree learning (or induction of decision trees) as a method of predictive modeling [21]. The use of a decision tree, it goes from observing a sample (representing the trunk) to drawing conclusions about the sample's target value (representing the leaves, which are attack types) [22].

Classification trees are a type of tree model where the target variable is discrete, and where the "leaves" represent the different types of attacks and the "branches" represent characteristics of the dataset that aid in predicting class labels [23, 24]. One type of decision tree, called a regression tree, has a continuous objective variable (typically real numbers) [23]. One of the most well-known machine learning algorithms, decision trees are prized for their clarity and ease of use. Using a Decision Tree with a classification type was necessary due to the discrete nature of the label in our experiment. These are the parameters we used for DT: criterion = gini and random_state = 42.

- Multilayer Perceptron Algorithm (MLP)

The feedforward artificial neural network is based on the multilayer perceptron (MLP). Many-layer Perceptron Networks (MLPs) are a type of feedforward artificial neural network (ANN) that can also be thought of as "deep neural networks" (with threshold activation). Decision Tree is a supervised machine learning system that, similar to people, makes decisions based on predefined criteria [20].

Data mining, statistics, and machine learning all use decision tree learning as a technique for predictive modeling (or induction of decision trees) [21]. The use of a decision tree Input, hidden, and output layers are the three types of node layers an MLP needs to operate. Except for the input nodes, all other nodes are neurons with nonlinear activation functions. [25, 26]. As part of its training process, MLP uses backpropagation, a supervised learning method. An MLP differs from a linear perceptron because to its several layers and non-linear activation. It can distinguish between data that can be cleanly categorised and data that cannot [27]. Since our experiment's label is discrete, we used an MLP of the classification variety; the corresponding MLP parameters were as follows: activation = relu and random_state = 42.

- eXtreme Gradient Boosting Algorithm (XGBoost)

Shortened to "XGBoost," Extreme Gradient Boosting is a popular method for both classification and regression. Gradient boosting has been parallelized and optimized [28]. Through the use of parallelization, the training time for the boosting process can be cut in half. As an alternative to training a single optimal model on the entire dataset, we train hundreds of models on different subsets of the dataset [28]. Afterwards, have a vote to determine which model did the best (as in traditional approaches). As opposed to more traditional gradient boosting methods, XGBoost often yields better results [29]. The Python implementation provides access to numerous hidden characteristics that can be changed to increase accuracy and precision [30].

This algorithm's overarching goal is to strengthen weak learners (decision trees) so they can generate the final prediction label (the weighted average of each weak classifier's predictions) [30]. Among the many notable features of the XGBoost [28, 29, 30] are: The model is 1) parallelized, meaning it can run in parallel on multiple CPU cores. 2) Regularization, XGBoost offers a variety of regularisation penalties to prevent overfitting. A properly trained model can successfully generalize thanks to

regularizations with penalties that improve training. XGBoost can recognize and learn from non-linear patterns in data. 3) it's possible to use cross-validation right now because it's built in. 4) XGBoost's ability to run on distributed servers and clusters like Hadoop and Spark makes it possible to manage massive amounts of data. C++, Java, Python, and Julia are just some of the supported programming languages. Because the label in our experiment is discrete, we utilised this approach with classification type. The XGBoost parameters we used are as follows:

colsample_bylevel = 1, learning_rate = 0.1, gamma = 0, n_estimators = 100, and random_state = 42.

- AdaBoost Algorithm

Adaptive boosting, or the statistical classification meta-algorithm AdaBoost, is an example of such a program. It complements a variety of learning strategies to boost efficiency and effectiveness. In a boosted classifier, the final output is a weighted sum of the outputs of other learning algorithms, or "weak learners" [31]. AdaBoost is adaptive because it adjusts subsequent weak learners to prioritize instances that earlier classifiers incorrectly labeled. It has the potential to be more resistant to the overfitting problem than other learning algorithms. Even if individual learners perform badly, the model as a whole will eventually converge on a highly effective learner if and only if it outperforms random guessing [32]. Combining powerful base learners, such as deep decision trees, with AdaBoost has been shown to work well to create a more accurate model [33]. It is common practice to use AdaBoost to combine weak base learners (such as decision stumps). While some learners may perform poorly, the model as a whole will eventually converge to a highly effective learner if and only if it can outperform random guessing [32]. Combining powerful base learners, such as deep decision trees, with AdaBoost has been shown to work well to create a more accurate model [33]. It is common practice to use AdaBoost to combine weak base learners (such as decision stumps).

Most learning algorithms are better suited to certain problem classes than others, and the optimal performance of an algorithm on a dataset depends on a wide range of tuning parameters and settings. Since it employs decision trees as weak learners, AdaBoost has been deemed by many to be the best out-of-the-box classifier [31, 32, 33]. AdaBoost, when combined with decision tree learning, employs information obtained at each stage about the relative "hardness" of each training sample to instruct subsequent trees to prioritize the most difficult-to-classify samples [33]. Because the label was discrete, we employed AdaBoost with classification-type training data in our experiment. Our final GB settings are as follows: algorithm = SAMME.R, learning_rate = 1.0, n_estimators = 50 and random_state = 42.

- Gradient Boosting Algorithm (GB)

Classification and regression are just two examples of the many applications of the machine learning technique known as "gradient boosting." It provides a predictive model in the form of a series of decision trees, which are in general unreliable [34]. Using a technique called gradient boosting, we

can combine the strengths of multiple less-effective learners (decision trees) into a single, more robust one. In this setting, individual decision trees are inefficient learners [35]. Each successive tree in the line is linked to the one before it and functions to correct the defect of the one before it in the chain. Boosting algorithms require extensive training time but provide high accuracy thanks to this causal connection. Statistic learning favors models with a slower learning rate [34, 35].

The following are our final GB settings: As the model improves, each new member of the group of slow learners can be accommodated within the residuals of the previous stage. The final model synthesizes insights from all three to create a robust learner. The residuals are calculated by using a loss function. For instance, in classification, mean square error (MSE) can be used, and in regression, logarithmic loss (log loss) is employed. No alterations occur when a new tree is added to the model.

The residuals of the existing model fit well with the additional decision tree [34, 35, and 36]. We used a GB with categorization type because our experiment's label is discrete. Here are our preferences for configuring GB: subsample= 1.0, learning_rate = 0.1, criterion= friedman_mse, n_estimators = 100, and random_state = 42.

- Voting Algorithm

As stated by [37], an ensemble machine learning model known as a "voting classifier" forecasts a class based on the likelihood that the output (class) corresponds to the target class. The only thing it does is add the results of each classifier that was fed into the voting classifier to predict which output class will receive the greatest number of votes [38]. Instead of building and analysing numerous specialised models, we propose a single model that trains on many models and predicts output based on the total number of votes for each output class. There are two voting processes. [39, 40] that are compatible with Voting Classifier. The projected output class with the most votes, and thus the highest likelihood of being predicted by each classifier, is the one chosen by "hard voting," as described in (1). The second method is known as "soft voting," and it involves basing the forecast on the output class rather than a majority vote.

Given the dichotomous nature of the labels in our experiment, we opted for voting with classification, setting the parameters as follows: estimators = DT, RF, and XGB, and voting type = hard.

2) Deep learning algorithms: Once the dataset is ready, it is possible to use three deep learning algorithms to identify cybersecurity attacks using the dataset. There are three methods: LSTM, CNN, and LSTM combined with CNN.

- Long Short-Term Memory (LSTM) Algorithm

Long Short-Term Memory (LSTM) neural networks are used in the fields of deep learning and artificial intelligence. Because the LSTM can develop itself through feedback connections, it has a distinct advantage over traditional feedforward neural networks. Long Short-Term Memory (LSTM) neural networks are used in the fields of deep

learning and artificial intelligence. Because the LSTM can develop itself through feedback connections, it has a distinct advantage over traditional feedforward neural networks. This recurrent neural network can analyse complete data sequences in addition to individual data points (such as photos, speech or video). LSTM has demonstrated success in the fields of healthcare, video game creation, healthcare analytics, and networked, unsegmented handwriting recognition. The longest short-term memory (LSTM) model of neural networks has been the subject of the most research over the past 100 years [41]. The remaining components of a typical LSTM unit are made up of cells, input gates, output gates, and forget gates. The values stored inside the cell can be retained there indefinitely, and the flow of data entering and leaving the cell is controlled by the cell's three gates. [41, 42]. LSTM networks are excellent for jobs involving the categorization, processing, and prediction of time series data because delays of variable lengths may exist between significant occurrences in a time series. Since the training of standard RNNs can result in the vanishing gradient problem, long short-term memory (LSTMs) was created as a solution. Because it is less sensitive to gap length, LSTM outperforms RNNs, hidden Markov models, and other sequence learning strategies. [42].

Fig. 3 depicts the two layers of the LSTM architecture that we employed: 1) a Bi-LSTM layer made up of an LSTM layer, 64 units, and relu as the activation function; and 2) a dense layer made up of a single unit and sigmoid as the activation function.

```
Model: "sequential"
_____
Layer (type)              Output Shape            Param #
=================================================================
bidirectional_2 (Bidirectio  (None, 512)          557056
nal)

dense_2 (Dense)           (None, 1)               513

=================================================================
Total params: 557,569
Trainable params: 557,569
Non-trainable params: 0
_____
```

Fig. 3.   LSTM Architecture.

- Convolutional Neural Network (CNN) Algorithm

Visual imaging evaluation typically makes use of an ANN class called a convolutional neural network (CNN, or ConvNet). Feature maps, which are translation-equivariant responses generated by convolution kernels or filters based on their shared-weight design and slide along input features, CNNs are also known as artificial neural networks that are shift- or space-invariant (SIANN). Contrary to common opinion, the majority of convolutional neural networks down sample the input, they do not translate invariantly. They are used in natural language processing, picture and video recognition, brain-computer interfaces, classification, segmentation, recommender systems, medical image analysis, and financial time series [43].

The structure of the visual cortex of animals served as inspiration for the development of convolutional networks. In particular, a cortical cell will only respond to stimuli that fall within the cell's receptive field [44]. All of the visible world is covered by the partially overlapping receptive fields of many neurons. Comparatively, CNNs require much less pre-processing than other image classification methods. The implication is that, unlike conventional methods, the network figures out how to optimize the filters (or kernels) on its own. First and foremost, feature extraction doesn't need any context or human input [43, 44].

The five-layered LSTM architecture utilized in our experiment is depicted in Fig. 4. Two Conv1Ds with the following parameters are provided: 128-element filters, three-element kernels, the same amount of padding, and the relu activation function. There are a total of three layers, including two max pooling layers and one dense layer containing one unit and a sigmoid activation function.

```
Model: "sequential"
_____
Layer (type)              Output Shape            Param #
=================================================================
conv1d (Conv1D)           (None, 15, 128)         5888

max_pooling1d (MaxPooling1D  (None, 7, 128)       0
)

conv1d_1 (Conv1D)         (None, 7, 128)          49280

max_pooling1d_1 (MaxPooling  (None, 3, 128)       0
1D)

dense (Dense)             (None, 3, 1)            129

=================================================================
Total params: 55,297
Trainable params: 55,297
Non-trainable params: 0
_____
```

Fig. 4.   CNN Architecture.

- LSTM with CNN Algorithm

We integrated the layers of LSTM and CNN to improve each algorithm's performance results. Fig. 5 depicts the LSTM with CNN architecture and includes: Two Conv1D are included: filters are 128, kernel-size = 3, padding = same, and relu as an activation function. One Bi-LSTM layer with an LSTM layer, 64 units, and relu as an activation function, as well as two max pooling layers and one dense layer with one unit and sigmoid as an activation function.

```
Model: "sequential"
_____
Layer (type)              Output Shape            Param #
=================================================================
conv1d_2 (Conv1D)         (None, 15, 128)         5888

max_pooling1d_2 (MaxPooling  (None, 7, 128)       0
1D)

conv1d_3 (Conv1D)         (None, 7, 128)          49280

max_pooling1d_3 (MaxPooling  (None, 3, 128)       0
1D)

bidirectional (Bidirectiona  (None, 128)          98816
l)

dense_1 (Dense)           (None, 1)               129

=================================================================
Total params: 154,113
Trainable params: 154,113
Non-trainable params: 0
_____
```

Fig. 5.   LSTM with CNN Architecture.

IV. RESULTS AND DISCUSSION

Based on four evaluation metrics, this section summarizes the experimental findings for each cybersecurity dataset for each deep learning method.

A. Evaluation Metrics

There are several assessment measures to examine the machine learning algorithms that were utilized, includes f1-score, recall, accuracy, and precision [45]. These metrics' formulas are as follows where:

- TP: The outcome is a true positive when the algorithm accurately predicts the positive class.

- TN: When the algorithm correctly predicts the negative class, the result is said to be "True Negative."

- FP: False positive results occur when the algorithm forecasts the positive class incorrectly.

- FN: A False Negative result occurs when the algorithm guesses the negative class incorrectly.

1) Accuracy: The ratio of correctly predicted samples to total samples, which is just a ratio of correctly predicted samples to total samples, is the most evident performance metric.

$$Accuracy = \frac{TP+TN}{TP + TN + FP + FN} \tag{1}$$

2) Precision: is the proportion of positively anticipated tweets that actually occurred to all positively predicted samples.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

3) Recall: is the ratio of positive samples that were accurately predicted to those that were generally forecast to be positive samples.

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

4) F1-score: consider the weighted average of Precision and Recall.

$$F1\text{-}score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

B. Machine Learning Results

Table IV and Fig. 6 provide effectiveness measurements for the used machine learning techniques (precision, accuracy, f1-score, and recall). Voting and DT classifiers perform superiorly when it comes to the detection attack technique.

C. Deep Learning Results

The aforementioned three deep learning approaches were implemented with the following parameters: accuracy as the evaluation metric, binary cross entropy as the loss function, Adam as the optimizer function, 15 epochs as the number of training sessions, and 128 as the batch size. Table V and Fig. 7 exhibit the experimental outcomes for various methods. Each algorithm's performance is measured by four different criteria: accuracy, f1-score, recall, and precision:

LSTM: 0.93, 0.96, 0.87, and 0.91.

CNN: 0.92, 0.93, 0.88, and 0.91.

LSTM with CNN: 0.995, 1.0, 0.99, and 0.995.

The LSTM with CNN gave the best performance results, which means that this algorithm can detect these attacks efficiently and effectively compared with LSTM and CNN.

TABLE IV.   BINARY CLASSIFICATION PERFORMANCE RESULTS

Models	Accuracy	Precision	Recall	F1-score
DT	0.894056	0.878958	0.864066	0.871448
RF	0.879974	0.857993	0.852246	0.85511
GB	0.829376	0.733313	0.92632	0.818593
XGB	0.888489	0.852908	0.884161	0.868253
AdaBoost	0.774685	0.731106	0.724192	0.727633
MLP	0.735713	0.747059	0.550433	0.633848
Voting	0.893074	0.869463	0.873916	0.871684
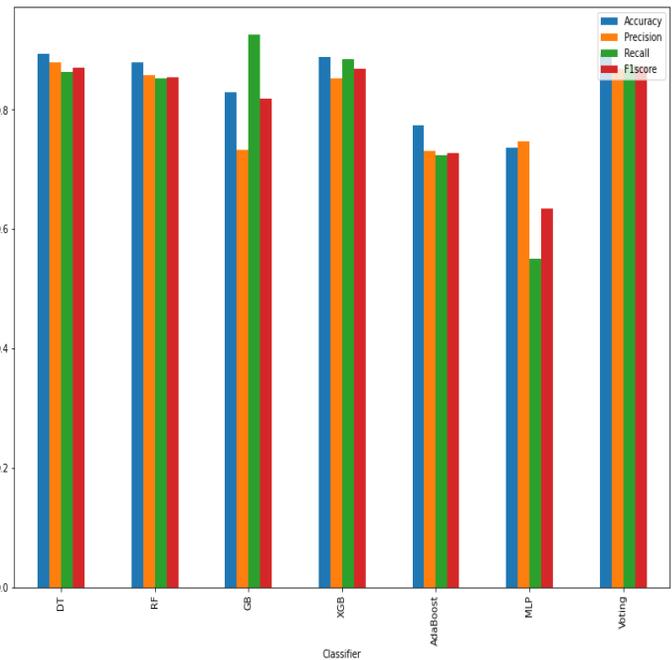
Fig. 6.   Performance Results of Binary Classification.

TABLE V.   DEEP LEARNING RESULTS

Algorithm	Accuracy	Precision	Recall	F1-score
LSTM	0.93	0.96	0.87	0.91
CNN	0.92	0.93	0.88	0.91
LSTM with CNN	**0.995**	**1.0**	**0.99**	**0.995**

Deep Learning Results
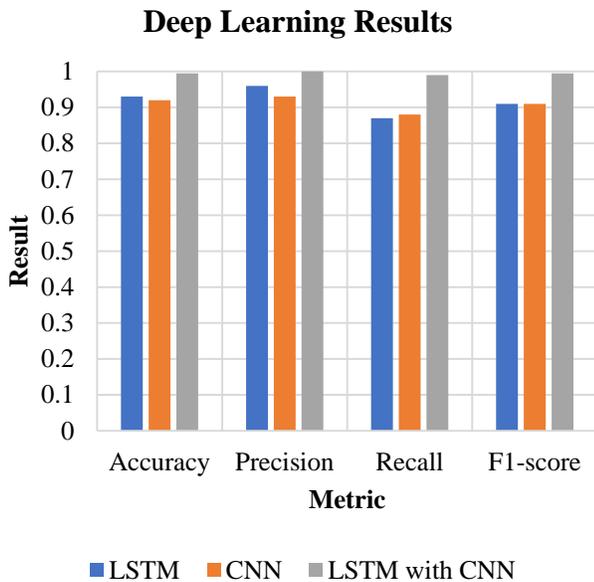
Fig. 7. Deep Learning Results.

V. Conclusion and Future Work

Eleven classification techniques, including three deep learning approaches [long short-term memory, convolutional neural networks, and long short-term memory + convolutional neural networks] [Decision Tree, Random Forest, Gradient Boosting, XGBoost, AdaBoost, Multilayer Perceptron, and Voting] were used in this study to identify suspicious assaults. Correctly predicted sample ratio is the most straightforward performance metric, as it is simply the predicted sample ratio. Using the Label Encoder method, we transformed the dataset from text to numbers. Algorithms are evaluated using the f1-score, along with precision, accuracy, and recall. In terms of detecting assaults, the LSTM with CNN outperformed the other models with scores of 0.995, 1.0, 0.99, and 0.995 for accuracy, precision, and recall, respectively. In this test, DT and voting classifiers outperformed the top machine learning methods. The results imply that a range of assault kinds can be detected using machine learning and deep learning. In a future study, we intend to assess the efficacy of these algorithms on a distinct dataset. In addition, we plan to implement ML, ML models, and DL approaches in our practice.

References

[1] Seemma, P. S., Nandhini, S., and Sowmiya, M. (2018), "Overview of cyber security", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 7 No. 11, pp.125-128.

[2] Ervural, B. C., and Ervural, B. (2018), "Overview of cyber security in the industry 4.0 era", In Industry 4.0: managing the digital transformation, pp.267-284.

[3] Chowdhury, A. (2016), "Recent cyber security attacks and their mitigation approaches–an overview", In International conference on applications and techniques in information security, pp.54-65.

[4] El-Rewini, Z., Sadatsharan, K., Selvaraj, D. F., Plathottam, S. J., and Ranganathan, P. (2020), "Cybersecurity challenges in vehicular communications", Vehicular Communications, Vol. 23.

[5] Berman, D. S., Buczak, A. L., Chavis, J. S., and Corbett, C. L. (2019), "A survey of deep learning methods for cyber security", Information, Vol. 10 No.4.

[6] Kim, A., Park, M., and Lee, D. H. (2020), "AI-IDS: Application of deep learning to real-time Web intrusion detection", IEEE Access, Vol. 8, pp.70245-70261.

[7] Vartouni, A. M., Kashi, S. S., and Teshnehlab, M. (2018), "An anomaly detection method to detect web attacks using stacked auto-encoder". In 2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), pp. 131-134.

[8] Betarte, G., Pardo, Á., and Martínez, R. (2018), "Web application attacks detection using machine learning techniques", In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp.1065-1072.

[9] Tuan, T. A., Long, H. V., Kumar, R., Priyadarshini, I., and Son, N. T. K. (2019), "Performance evaluation of Botnet DDoS attack detection using machine learning", Evolutionary Intelligence, pp.1-12.

[10] Anwer, M., Farooq, M. U., Khan, S. M., and Waseemullah, W. (2021), "Attack Detection in IoT using Machine Learning", Engineering, Technology and Applied Science Research, Vol. 11 No. 3, pp.7273-7278.

[11] Su, T., Sun, H., Zhu, J., Wang, S., and Li, Y. (2020), "BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset", IEEE Access, Vol. 8, pp.29575-29585.

[12] Xu, W., Jang-Jaccard, J., Singh, A., Wei, Y., and Sabrina, F. (2021), "Improving Performance of Autoencoder-Based Network Anomaly Detection on NSL-KDD Dataset", IEEE Access, Vol. 9, pp.140136-140146.

[13] Kavitha, S., and Uma Maheswari, N. (2021), "Network Anomaly Detection for NSL-KDD Dataset Using Deep Learning", Information Technology in Industry, Vol. 9 No. 2, pp.821-827.

[14] Ferriyan, A., Thamrin, A. H., Takeda, K., and Murai, J. (2021), "Generating Network Intrusion Detection Dataset Based on Real and Encrypted Synthetic Attack Traffic", Applied Sciences, Vol. 11 No. 17.

[15] Giménez, C. T., Villegas, A. P., and Marañón, G. Á. (2010), "HTTP data set CSIC 2010", Information Security Institute of CSIC (Spanish Research National Council).

[16] Hancock, J. T., and Khoshgoftaar, T. M. (2020), "Survey on categorical data for neural networks", Journal of Big Data, Vol. 7 No.1, pp.1-41.

[17] Pal, M. (2005). Random forest classifier for remote sensing classification. International journal of remote sensing, 26(1), 217-222.

[18] Farnaaz, N., & Jabbar, M. A. (2016). Random forest modeling for network intrusion detection system. Procedia Computer Science, 89, 213-217.

[19] Idhammad, M., Afdel, K., & Belouch, M. (2018). Detection system of HTTP DDoS attacks in a cloud environment based on information theoretic entropy and random forest. Security and Communication Networks, 2018.

[20] Kingsford, C., & Salzberg, S. L. (2008). What are decision trees?. Nature biotechnology, 26(9), 1011-1013.

[21] Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1(1), 81-106.

[22] De Ville, B. (2013). Decision trees. Wiley Interdisciplinary Reviews: Computational Statistics, 5(6), 448-455.

[23] Kotsiantis, S. B. (2013). Decision trees: a recent overview. Artificial Intelligence Review, 39(4), 261-283.

[24] Amor, N. B., Benferhat, S., & Elouedi, Z. (2004, March). Naive bayes vs decision trees in intrusion detection systems. In Proceedings of the 2004 ACM symposium on Applied computing (pp. 420-424).

[25] Noriega, L. (2005). Multilayer perceptron tutorial. School of Computing. Staffordshire University.

[26] Tang, J., Deng, C., & Huang, G. B. (2015). Extreme learning machine for multilayer perceptron. IEEE transactions on neural networks and learning systems, 27(4), 809-821.

[27] Ramchoun, H., Ghanou, Y., Ettaouil, M., & Janati Idrissi, M. A. (2016). Multilayer perceptron: Architecture optimization and training.

[28] Mitchell, R., & Frank, E. (2017). Accelerating the XGBoost algorithm using GPU computing. PeerJ Computer Science, 3, e127.

[29] Pan, B. (2018, February). Application of XGBoost algorithm in hourly PM2. 5 concentration prediction. In IOP conference series: earth and environmental science (Vol. 113, No. 1, p. 012127). IOP publishing.

[30] Dong, W., Huang, Y., Lehane, B., & Ma, G. (2020). XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. Automation in Construction, 114, 103155.

[31] Hu, W., & Hu, W. (2005, September). Network-based intrusion detection using Adaboost algorithm. In The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05) (pp. 712-717). IEEE.

[32] Jabri, S., Saidallah, M., El Alaoui, A. E. B., & El Fergougui, A. (2018, December). Moving vehicle detection using Haar-like, LBP and a machine learning Adaboost algorithm. In 2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS) (pp. 121-124). IEEE.

[33] Yuan, L., & Zhang, F. (2009, July). Ear detection based on improved adaboost algorithm. In 2009 International Conference on Machine Learning and Cybernetics (Vol. 4, pp. 2414-2417). IEEE.

[34] Son, J., Jung, I., Park, K., & Han, B. (2015). Tracking-by-segmentation with online gradient boosting decision tree. In Proceedings of the IEEE international conference on computer vision (pp. 3056-3064).

[35] Peter, S., Diego, F., Hamprecht, F. A., & Nadler, B. (2017). Cost efficient gradient boosting. Advances in neural information processing systems, 30.

[36] Lusa, L. (2017). Gradient boosting for high-dimensional prediction of rare events. Computational Statistics & Data Analysis, 113, 19-37.

[37] Kumar, U. K., Nikhil, M. S., & Sumangali, K. (2017, August). Prediction of breast cancer using voting classifier technique. In 2017 IEEE international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM) (pp. 108-114). IEEE.

[38] El-Kenawy, E. S. M., Ibrahim, A., Mirjalili, S., Eid, M. M., & Hussein, S. E. (2020). Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images. IEEE Access, 8, 179317-179335.

[39] Khan, M. A., Khan Khattk, M. A., Latif, S., Shah, A. A., Ur Rehman, M., Boulila, W., ... & Ahmad, J. (2022). Voting classifier-based intrusion detection for iot networks. In Advances on Smart and Soft Computing (pp. 313-328). Springer, Singapore.

[40] Mahabub, A. (2020). A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers. SN Applied Sciences, 2(4), 1-9.

[41] Monner, D., and Reggia, J. A. (2012), "A generalized LSTM-like training algorithm for second-order recurrent neural networks", Neural Networks, Vol. 25, pp.70-83.

[42] Gers, F. A., Eck, D., and Schmidhuber, J. (2002), "Applying LSTM to time series predictable through time-window approaches", In Neural Nets WIRN Vietri-01, pp.193-200.

[43] Sun, Y., Xue, B., Zhang, M., Yen, G. G., and Lv, J. (2020), "Automatically designing CNN architectures using the genetic algorithm for image classification", IEEE transactions on cybernetics, Vol. 50 No. 9, pp.3840-3854.

[44] Dalianis, H. (2018), "Evaluation metrics and evaluation", In Clinical text mining, pp.45-53.

Permission and Usage Control for Virtual Tourism using Blockchain-based Smart Contracts

Muhammad Shoaib Siddiqui, Toqeer Ali Syed, Adnan Nadeem, Waqas Nawaz, Ahmad Alkhodre
Faculty of Computer and Information Systems
Islamic University of Madinah, Madinah
Kingdom of Saudi Arabia

Abstract—**Virtual Tourism (VT) is a booming business with potential perspectives in the entertainment and financial industry. Due to travel restrictions, safety concerns, and expensive travelling the younger generation is showing interest in virtual tourism instead of traditional tourism. However, virtual tourism does not financially benefit the service providers as compared to traditional tourism stakeholders. An online system is essential to provide a central point of access to various tourism sites along with usage, permission, and payment control. In this paper, a secure blockchain-based broker service for users and content providers is proposed, which allows tourism sites to announce their virtual tours and provide accessibility and accountability. Meanwhile, it enables users to register, subscribe, access, and be billed according to their usage. The permission control module ensures authentication and authorization, while the usage control provides accountability to the predefined service level agreement. The transactions are stored on the blockchain to ensure the integrity of data and smart contracts are used to ensure automatic usage and permission control. An implementation on Hyperledger Fabric is provided as a proof of concept with performance measurements as a case study.**

Keywords—*Virtual tourism; permission control; usage control; access control; blockchain*

I. INTRODUCTION

The imaginative ability of the human mind can be perceived and experienced by the latest Virtual Reality Technologies [1]. People can visualize and interact with virtual objects in a realistic or enhanced environment. They can build and experience their imaginative environment, which allows them to live, roam, and interact in an ideal, purpose-built world. In general, it is not easy to define the limits of the human imagination, but with the cutting-edge technology of computer-based systems [2], it is possible to create human-inspired imaginary worlds. With the development of technology, the imagination of the virtual world can be diversified from the real world to provide a more realistic experience.

Industry 4.0's arsenal accompanies one of the most promising technologies, Virtual Reality, which is one of the modern technologies in the current generation that yielded a rising market for VR application development. The extensive advancements in VR and other mixed realities have considerably changed experiences in different areas of interest. The implications and uses of these technologies are broad and are being actively used to boost the level of experiences of clients. The massive adoption of these technologies is found in healthcare departments [3] [4], education sectors [5] [6], manufacturing and logistics [7], smart cities [8], Museum [9] [10], social media [11], etc. In retail, virtual reality enables products to be visualized in 3D for consumers [13] [12] that ultimately leads to customer satisfaction and maximum sales.

One of the breakthrough applications of VR in recent times is Virtual Tourism [16], which enables a tourist to explore nature, attractions, destinations, ruins, buildings, and other travel destinations without physical interactions. The immersive experience is provided by the emerging technologies of VR and remote sensing. Virtual tours may include 360-degree view, guided tour of the virtual and augmented world, and video tours of remote destinations, which provides the ability to interact with and experience the art, nature, culture, or ancient ruins [14]. Worldwide, the VT industry was worth USD five Billion in 2021 and it is anticipated to reach USD 24 Billion in 2022 [15], as shown in Fig. 1.

VT provides a cheap replicate, motivation, and a marketing pretense for the tourism industry [16]. Due to the recent coronavirus pandemic (global pandemic), social distancing was recommended that made people's daily lives a bit boring with monotonous experiences. VT enables tourism experiences within the boundaries and guidelines of epidemic restrictions. Users can use specialized equipment from the comfort of their homes to create immersive experiences in virtual or real environments. Another benefit of VT is that it preserves cultural and natural heritage in environments vulnerable to tourist visitation and urbanization [17].

VT is enabled by VR/AR/MR technologies and applications. Without the implementation and support of the right tools, this promising industry cannot develop and thrive [16]. At one end, the virtual and augmented content is created using 3D cameras, scene modelling and composition, and image rendering using computer graphics; while at the extreme end, display technologies with embedded tracking technologies are used to provide an immersive experience to the user. The realistic experience of the user is the selling point of these virtual tours. The availability of tours as online or computer-based services requires accountability for permission, rights, access, and usage to ensure justified financial benefits for the service and content providers. Without this accountability and billing model, the services are not financially feasible.
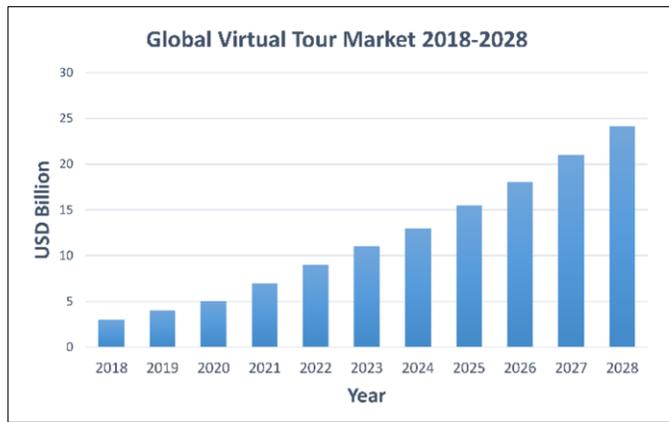
Fig. 1.   Market Growth of Virtual Tourism [15].
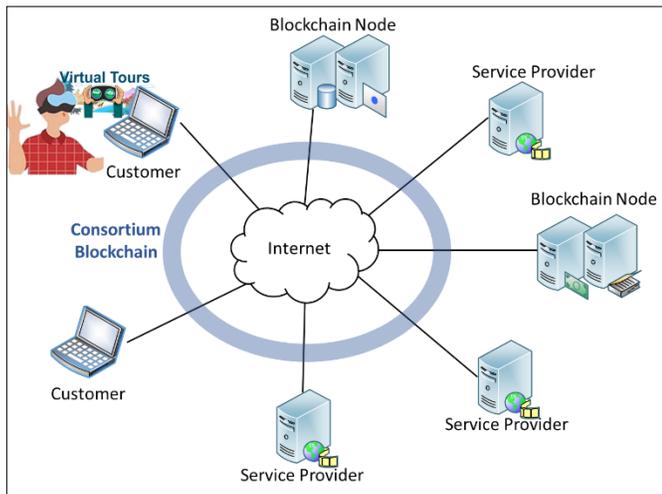
Fig. 2.   The Blockchain based Content Distribution System using a Consortium of Blockchain Nodes, and Registered Service Providers and Customers.

To facilitate the virtual tourism industry and its customers, this paper presents an online system that provides a consortium blockchain-based decentralized access to various tourism sites and content providers, as shown in Fig. 2. Along with the permanent nodes of the blockchain, service providers and customers can become members of the blockchain, which maintains usage, permission, and payment control. It allows the tourism sites to announce their virtual tours and provides accessibility and accountability, at the same time, allows users to register, subscribe, access, and be billed according to their usage. The permission control module ensures authentication and authorization, while the usage control provides accountability and billing. The transactions are stored on the blockchain to ensure the integrity of data and smart contracts are used to ensure automatic usage and permission control.

The rest of the paper is structured as follows. Section II discusses the literature review. Section III presents the proposed blockchain-based permission and usage control system. Section IV presents the implementation details on Hyperledger fabric and evaluation of the system and Section V concludes the paper and provides future directions.

II.   RELATED WORK

VT is used by various industries from museums to ancient ruins and from universities to stadiums. VT is an efficient marketing edge that any business can rely on for expanding clientage and market penetration.   Tourism agencies use virtual tours for marketing their destinations and attractions. According to a survey in 2017 by TIG Global, 50% of tourists base their selection of a destination on virtual tours [18]. A study by Google in 2015, identified a 50% increase in online interest and a 40% increase in booking reservations [19]. According to Panomatics' statistics, 75% of potential buyers consider virtual tours as a major factor in their purchasing decision [20]. Another study found that among the students of culture, 80% of people go to virtual exhibits, museums, and classes for further education [21]. The online event industry, which is a subcategory of virtual tourism, is identified to grow by $2.3 trillion in 2026 [22].

Many websites and portals on the Internet provide virtual tours to various destinations, such as museums [9], architectures [23] [26], [28], underwater explores [24], archaeological sites [25] [27], monuments [29], Holy sites [30] [31], and remote locations [32] [33]. From the couch of your living room, you can party in Ibiza, be alone in the clouds in a hot air balloon, dive with fishes in the Georgia Aquarium, fly over the skies of Paris; and much more [34].

However, most of these sites are using VT for marketing purposes only. As the boom of VT expands, more and more people will prefer virtual tourism over traditional tourism [16]. For making the system financially feasible, a content distribution system is required, which provides all the tours at a single portal, to allow authentication, authorization, accountability, and payment facilities along with security and trust. Such a content distribution system for VT is missing; however, research has already been done for permission, access, and usage control.

In [35], the authors have reviewed various access control systems and identified the challenges and opportunities. They have recommended the use of blockchain for developing an access control system with decentralized trust. Kishigami et al. have proposed a decentralized blockchain-based digital content distribution system along with a prototype application [36]. It uses a digital currency-based payment system; however, a usage control system is missing. Maesa et al. have proposed an access control system, which uses blockchain to publish the policies expressing the right to access a resource and to allow the distributed transfer of such rights among users. The system provides distributed auditing and prevents fraudulent transactions. It is deployed on the Bitcoin blockchain [37]. In [38], the same team has proposed the use of blockchain technologies to implement the traditional access control system to achieve better auditability.

Khan et al. have proposed an extended usage control model known as DistU (Distributed Usage Control), which considers all possible access control models required by a business for permissioned blockchain frameworks [39]. The detailed mechanism of DistU is provided in [40]. Khan et al. also presented an extension to the UCON model in [42].   Their model is an extension of the UCON model developed by Park

and Sandhu [41]. UCON is the most generalized usability control system with work based on pre-defined attributes, conditions, and obligations. It uses a set of policies for maintaining usage before, during, and after access is granted to a user.

The main contribution of this work is a comprehensive and purpose-built permission and usage control for virtual tourism cases, in which the customers can register into the system, subscribe to and visit the virtual tours, and pay for services in a trusted environment. At the same time, the service provider can financially benefit from the virtual tours.

The proposed system is also a derivation of the UCON model, in which the model has been modified according to the requirements based on the customer as a consumer of the service, virtual tours as the service and tourist cites as the service providers. Along with the usage control of UCON, the access and permission control systems have been embedded for the completeness of the solution. A working system on top of a consortium blockchain using the Hyperledger Fabric as a proof of concept has also been proposed. The next section provides the detail of the proposed system.

III. BLOCKCHAIN-BASED PERMISSION AND USAGE CONTROL

To enable content streaming for virtual tourism, a comprehensive system has been proposed, which allows users to register, maintain their accounts and subscriptions, and handle bills. For tourism agencies and service providers, it supports content provision, billing itineraries, and service management. It also enables secure and trusted permission, rights, and usage control using blockchain technologies. Fig. 3 shows the modules available in the proposed system and the following sections discuss the details of each module.

A. Membership

The system is developed on a permission blockchain, therefore only the members are allowed to enter the system. A new user must register into the system by creating an account through a valid email address, contact, and financial details, which enable the system to charge or/and remit the user for the services. There are mainly two types of users: customers and service providers. On registering into the system, a user is created in the blockchain's couch DB, which serves as a distributed database, while the certification authority creates security credentials for authentication and maintaining the confidentiality of the data. In this section, the various participants and entities involved in the system, and their interactions are defined.
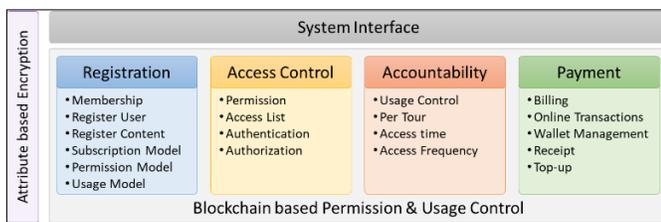
Fig. 3. Modular Diagram of the Proposed Blockchain-based Permission and Usage Control System.

Definition 1: A customer is defined in the system as an active member (ui), such that ui ∈ U, while U specifies the set of customers with attributes: identity (id), credentials (ci), set of subscribed tours (Ti), wallet (wi), and states Σ and A.

$$u_i \in U(id, c, T, w, \Sigma, A) \qquad (1)$$

Where, Σ is the state of user (customer or service provider) as signed-in, signed-out, granted, or denied. A is a set of states for the customer against the set of subscribed tours. The state value can be initial, subscribed, requested, accessing, or denied.

Definition 2: A Service Provider (ri) is also a user in the system, (ri), such that ri ∈ R, while R specifies the set of service providers with attributes: identity (id), credentials (ci), set of provided tours (Ti), wallet (wi), and state Σ. A service provider can create/provide a set of Tours (Ti).

$$r_i \in R(id, c, T, w, A) \qquad (2)$$

Upon arrival into the system, a user must sign-in to the system with its id and credentials. A membership function *(Ψ)* checks a user (u_i/r_i) against corresponding credentials (c_i) to allow or deny an entry into the system by updating the state *Σ*.

$$\Sigma(u_i): \Psi(u_i, c_i) \qquad (3)$$

The membership model provides authentication in the system by defining who can access and enter into the system.

B. Content Registration

The content provided by the service provider to the customers is in the form of virtual tours. A customer can view and experience the virtual tour as per the subscription and is charged according to the subscription rate, which is paid through an associated wallet.

Definition 3: A tour is defined as a service consumed by the user (u) and provided by the service provider (r). It is denoted by t such that t ∈ T (id, r, β, ρ, τ, U).

Where, *r* is the service provider of the tour, *β* is the subscription fee, *ρ* is the usage rate per frequency of access, *τ* is the usage rate per time of access, and *U* is the set of users subscribed to consume that resource.

A service provider ri can enter a tour tj into the system if the state is signed-in. An online portal allows the service provider to manage her/his tours in terms of editing information, subscription rates, content, narration, etc. A tour has various attributes such as id, location, subscription rates, viewer's rating and reviews, historical information, content, narration, information and details, and list of attractions, etc., out of which the mandatory tuple is defined in definition 3. After a successful signed-in into the system, the service provider can add tours along with their details and media files, which are saved in a distributed database. After a service provider creates a tour, it is assigned an id and associated with that service provider. The function O(r, t) checks if the service provider ri is the owner of the tour ti, based on which the system permits the service provider to edit the tour or deny access.

C. Subscription

The subscription model defines which user and service provider can access and edit/manage which tours respectively. It enables authorization in the system.

After a customer (u_i) has successfully signed into the system, the customer can enlist the tours available in the system. The customer can search the tours based on various categories such as id, location, subscription rate, viewer's rating, etc. From the available list of tours, a customer can subscribe to the tour given that the subscription condition is fulfilled, which checks if there is enough amount in the customer's wallet to pay the subscription fee against the selected tour:

$$\Delta : w(u_i) \geq \beta(t_j) \tag{4}$$

Where u_i is the customer, t_j is the selected tour, $w(u_i)$ is the amount in the wallet of customer u_i, and β is the subscription fee for the tour t_j, and Δ is the subscription status.

- If Δ = allow (1), then tj is inserted into the set of Ti for customer ui.

$$T_i(u_i) \leftarrow t_j, \text{ or } t_j \in T_i(u_i) \tag{5}$$

At the same time, for the tour t_j, u_i will be inserted in its set of subscribed users U_j.

$$U_j(t_j) \leftarrow u_i, \text{ or } u_i \in U_j(T_j) \tag{6}$$

- If Δ = deny (0), then the customer ui is informed that his wallet (wi) does not have the required amount to subscribe to the tour tj.

Θ(ui, tj) is the subscription function that checks for a user's (ui) access to a tour (tj), which can be initial, subscribed, requested, accessing, or denied. The state is maintained in aj, where aj $\in$ Ai for ui.

D. Access and Permission Control

Permission control has two stages. The first stage is for the entry into the system for the users (authentication), while the second stage is for permitting users to access only their owned and subscribed contents (authorization). When a user logs-in to the system, the permission control checks corresponding credentials (id and password) to ensure a user is allowed in the system, based on which it permits or denies access. Ψ(ui, ci) is the membership function that checks user ui against user credentials ci to grant or deny entry into the system.

$$\Psi(\alpha i, c i) \rightarrow \Sigma(\alpha i) = \{granted, denied\}, \text{ where:} \tag{7}$$

$$Role\ (\alpha) = \begin{cases} customer, & \alpha \in U \\ service\ provider, & \alpha \in R \end{cases} \tag{8}$$

In case of access is denied, the user is advised to register into the system. When a customer registers successfully, a customer entry is created and added to the set of Customers *U*:

$$U \leftarrow u_i\ (id_i, c_i, T_i, w_i, \Sigma_i, A_i) \tag{9}$$

When a service provider registers successfully, an entry is created and added to the set of Service Providers *R*:

$$R \leftarrow r_i\ (id_i, c_i, T_i, w_i, A_i) \tag{10}$$

The second stage of permission control is used when a customer tries to view one of the tours. The permission control uses the subscription model to check if a customer is authorized to access the said tour. Similarly, when a service provider is trying to edit/manage the tour, the permission model checks the ownership.

Θ(ui, tj) is the subscription function that checks if a customer ui is subscribed to a tour tj to granted or denied access to the tour tj. The state is maintained in the set Ai. Similarly, Θ(ri, tj) checks if a service provider ri is the owner of a tour tj to granted or denied access to the tour tj. The state is maintained in the set Ai.

E. Usage Control

The usage control maintains accountability in the system by calculating the use of the service by each customer and ensures that the user does not exploit the system by excessing the usage defined for the said content. After the permission control grants access to the customer to view the tour/content, the usage control is activated. Usage control maintains a state of access for each user accessing a tour t_i. However, in the case of a service provider accessing pre-owned tours, the usage control is not activated.

Usage control used two functions to calculate the usage of the content/tour (t_j) for a customer (u_i). $\Xi(u_i, t_j)$ calculate the usage in terms of time; as the amount of time the customer u_i has accessed the tour t_j. While $\Pi(u_i, t_j)$ calculates the usage in terms of frequency, as how many times the tours t_j was accessed by user u_i.

For each tour t_j the service provider adds a rate ρ_j that is used to calculate the usage in terms of access time and rate τ_j, which determines how many times a tour is accessed. This approach is used to calculate the billed amount ($Q(u_i)$) for the customer (u_i), which is debited from the wallet of the customer u_i, and credited into the service provider r_k, who owns the tour t_j. These two usage parameters are designed to give flexibility to the service provider to charge its customer. If the service provider wants to charge the customers based on how many times the user accesses the system only, then a non-zero value is provided to τ_j and a zero value is provided to ρ_j. On the other hand, when the charge is based on the time of access only, then a zero value is provided to τ_j and a non-zero value is provided to ρ_j.

When the customer is allowed to access the tour and allowed by the permission control, the state of the tour for that customer is changed to 'requested.' After that, the usage control first checks the balance in the wallet of the user and if the value of τj is non-zero, then the user is allowed access only if the user has a balance of more than τj. In that case, the state for the tour is changed to accessing. If there is not enough balance, then the state is changed to 'denied' and the user is informed about the lack of balance. The change of state is visualized in Fig. 4. In case the value for τj is zero, then the condition in step 3C1 in Algorithm 1 is true and the customer would only be charged based on λ(ui), which is the amount of time in seconds.
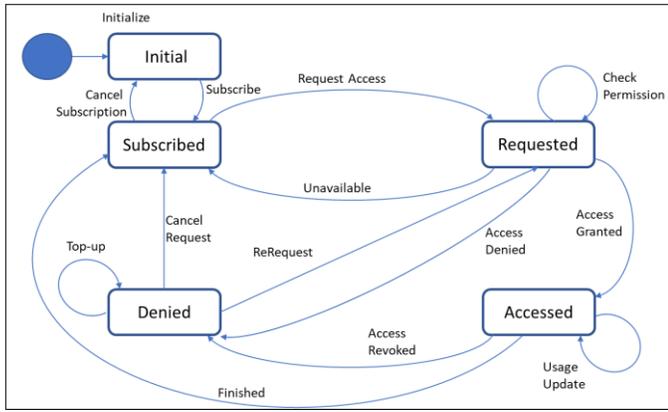
Fig. 4. The State-Model Diagram of the Usage Control Model, which Monitors the usage of a Tour by a Customer and Ensures that the user has enough Balance to Keep viewing the Tour.

Similarly, the usage control monitors and calculates the usage based on the time using equation (11) and deducts the amount from the wallet using equations (12), (13), and (14), respectively.

$$\Xi(u_i. t_j) = \rho_j \times \lambda(u_i) \tag{11}$$

$$\Pi(u_i. t_j) = \tau_j \times \nu(u_i), \tag{12}$$

$$Q(u_i) = \Xi(u_i. t_j) + \Pi(u_i. t_j) \tag{13}$$

$$w(r_k) = w(r_k) + Q(u_i) \tag{14}$$

Where, $\nu(u_i)$ is the number of times a tour (t_j) is accessed by the customer (u_i) and $\lambda(u_i)$ is the time (how long) for which the tour (t_j) is accessed by the customer (u_i).

Once the amount is lower than the threshold (which is zero in most cases), then the usage control will revoke the access of tj for the user (ui) and change the state to denied.

$$\Delta: w(u_i) = w(u_i) - \Xi(u_i. t_j), \text{ where } \Delta: w(u_i) \geq 0 \tag{15}$$

- If Δ = allow (1), then ui keeps accessing the tour tj and aj = {accessing}

- If Δ = deny (0), then tj access for ui is revoked and aj = {denied}

The customer is informed in the system about the revoked access and given an option to top-up her/his account by crediting the balance to the account through the manage account section of the portal. Afterward, the customer can send a re-request for the content/tour, which change the state to 'requested.' The usage control checks the condition (given in equation 15), if the result is 'allow', then the state is changed back to accessing and the tour is continued for the user. At any time, when the customer selects to cancel the access to the tour, the system stops the access and changes the state back to 'subscribed.' This whole process is outlined in Algorithm 1 and the flowchart in Fig. 5 for better understanding.

Algorithm 1: Customer accessing a Tour; Permission and Usage Control

1. System enlists the available tours based on id and location,
2. Customer u_i selects a tour t_j to view
3. Permission Control checks subscription
   if $u_i.a(t_j) = $ **subscribed**
   a. *Permission = Allowed*
   b. state(t_j) = **requested**
   c. **if** $w(u_i) \geq \nu(u_i)$ //(deduct usage based on number of access)
      1. $w(u_i) = w(u_i) - \tau_j * \nu(u_i)$,
      else
      i. **inform** user "*Not enough balance to view the tour*"
      ii. state(t_j) = **subscribed**
      iii. **goto** step 1
      end if
   else (**if** not **subscribed**)
      subscribe tour t_j
   a. Check for the subscription fee
      if $w(u_i) \geq \beta_j$
      i. $u_i.a(t_j) = $ **subscribed**
      ii. $w(u_i) = w(u_i) - \beta_j$ //(deduct subscription fee)
      iii. **goto** step 3
      else
      i. **deny** request (*Permission = Not Allowed*)
      ii. **end**
      end if
   end if
4. Grant Access to the Tour, state(t_j) = **accessing**
5. **while** state(t_j) = **accessing**
   a. **if** *user.action = Cancel Tour*
      i. state(t_j) = **subscribed**
      ii. **end** //Stop Access to the tour
      end if
   b. **monitor** Usage Control, $\lambda(u_i)$ = time in seconds for how long the customer u_i is accessing the tour t_j
   c. **calculate** time usage: $\Xi(u_i. t_j) = \rho_j \times \lambda(u_i)$, // Debit the wallet and check for available balance
   d. $w(u_i) = w(u_i) - \Xi_{i,j}$
   e. **if** $w(u_i) < 0$
      i. **revoke** access
      ii. state(t_j) = **denied**
      iii. **inform** user "*Not enough balance to view the tour*"
      iv. **if** *user.action == Cancel Tour*
         a. state(t_j) = **subscribed**
         b. **goto** step 1
      v. **else if** *user.action = ReRequest*
         a. state = **requested**
         b. goto step 3.c
      vi. **else if** *user.action = Topup Account* AND $w(u_i) \geq \Xi_{i,j}$
         a. Action = *ReRequest*
         b. goto step 5.e.iv
         else
         a. **goto** step 5.e.iii
         end if
      end if
6. **end while**
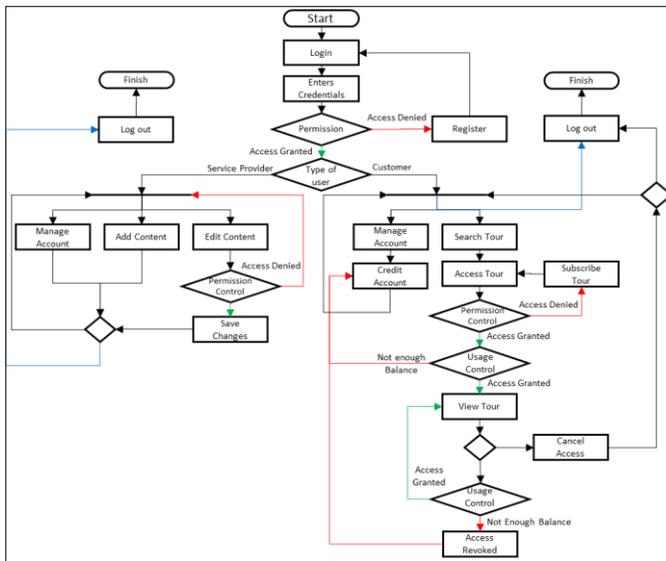
Fig. 5. The Flowchart Depiction of Algorithm 1, Showing the Permission and usage Control.

Fig. 6. Member and Content Registration Steps for the Service Provider on the Consortium Blockchain.

Fig. 7. Permission and Usage Control Model on Top of the Consortium Blockchain.

IV. IMPLEMENTATION AND EVALUATION

The proposed usage and permission control system is based on the concept of consortium blockchain, where some permanent nodes are part of the blockchain, and other nodes can register for membership and become part of the blockchain. As shown in Fig. 2, for making the tours available for the customer without needing a centralized server, the service providers can register as a member and then register their content on the blockchain. Similarly, customers can become members and can search for and access the tours. Fig. 6 shows the registration activity for the service provider on the blockchain, verified by various nodes.

Due to its inherent properties, the blockchain provides distributed trust among customers and service providers. Content distribution and financial transactions can be implemented securely on the blockchain. Using smart contracts, the usage and permission control can implement the policies and makes sure that authentication, authorization, and accountability are maintained.

Fig. 7 shows the working of the permission and usage control system on top of the blockchain network A user can register on the blockchain as a member, then search and subscribe for various tours by paying the subscription fee. Afterward, the user can view the tours. The permission control maintains a permission state against a user accessing a tour. It checks the customer's permission to access the said tour and allows or denies the customer based on permission rights.

Similarly, during the access, the usage control maintains the usage state. It monitors the frequency and time of view for the customer against a tour. The system calculates the bill and makes sure that the customer has enough balance in the associated wallet. Accordingly, the usage control can revoke (or grant) access to the customer with a notification.
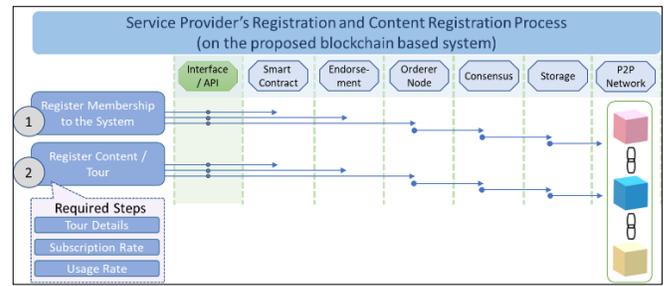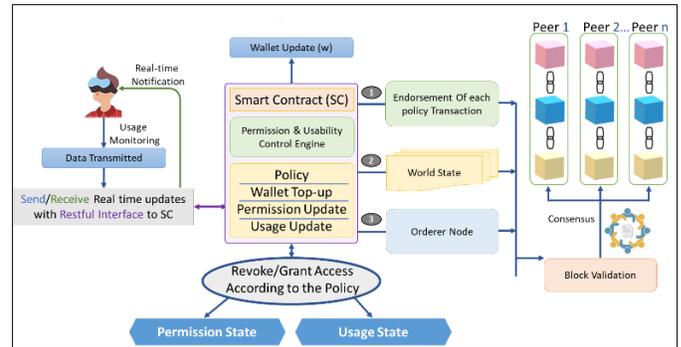
F. Implementation of Blockchain

The following text details the configuration of the testbed created to verify the permission and usage control model presented in the previous section. The proposed system is implemented on Hyperledger Fabric 1.4 designed as a consortium blockchain, with two permanent nodes and three service provider nodes, and two customer nodes. In the end, the performance evaluation of the consortium blockchain is presented in terms of the throughput and the latency of the system. However, first, a small introduction to blockchain technologies is given, along with the motivation for using the Hyperledger Fabric.

1) Blockchain technologies: The earlier blockchains, such as bitcoin, etc., face challenges with energy consumption and speed due to their public nature. On the other hand, modern blockchains have been developed as private or consortium concepts, to help overcome these limitations and provide practical value for other business uses and applications. The top modern blockchain frameworks are R3 Corda, Hyperledger, and Ethereum. Some of the modern blockchain frameworks are discussed below:

2) IBM Blockchain: IBM Blockchain is a private, decentralized blockchain network that has been the most successful with enterprise clients as it links into enterprise cloud and legacy technologies more seamlessly than is possible in other decentralized networks. The developer tool was designed to be flexible, functional, and customizable with a user-friendly interface to simplify critical tasks, such as setting up, testing, and rapidly deploying smart contracts [43].

3) R3 Corda: R3 Corda [44] is a distributed ledger platform that uses a novel consensus mechanism in which transactions are cryptographically linked but does not periodically batch multiple transactions into a block. It processes all transactions in real-time, which makes it superior in performance. However, R3 declared it as a blockchain and not a blockchain at the same time, as some large companies have observed implementation constraints and is deemed not suitable for consortium blockchains.

4) Ethereum: Ethereum [48] is one of the oldest and most established blockchain platforms. It provides a truly decentralized blockchain that is comparable to the Bitcoin blockchain network. It enables true decentralization with support for smart contracts. Its key weaknesses include slow processing times and higher transaction processing costs compared to other platforms.

5) HydraChain: Hydrachain blockchain platform [45] is an extension to the Ethereum platform that supports the development of a permissioned distributed ledger. Therefore, the tools for developing smart contracts and decentralized apps are same as the Ethereum; however, the issues of Ethereum are also present.

6) Quorum: Quorum [46] is a customized version of Ethereum developed by the financial services company JPMorgan. It has been optimized to support high-speed transactions between institutions such as banks and insurance companies on a private blockchain. It also adds various privacy enhancements to Ethereum to improve support for regulations. It supports confidentiality and privacy for all transactions using smart contracts. Quorum has lower throughput at higher load and experiment results suggest an increase in transaction latency with regard to reading and writing operations.

7) IOTA: IOTA [47] is a blockchain platform designed specifically for IoT blockchain applications, which uses a directed acyclic graph to store transactions on its ledger, motivated by potentially higher scalability over blockchain-based distributed ledgers. IOTA achieves consensus through a coordinator node, operated by the IOTA Foundation. It is not a strict blockchain solution due to the coordinator serving as a central control feature, which goes against the idea of blockchain as "distributed ledger technology".

8) Hyperledger Fabric [49]: Hyperledger Fabric provides a collection of tools to create permissioned blockchain applications. It is engineered with distributed ledger as the core concept. It supports components that can be plugged into a modular architecture, which works well in private and permissioned blockchain deployments, which improves security and speed. An open smart contract model that can support various data models is also provided.

After the comparative analysis of these modern blockchain frameworks, it is concluded that Hyperledger Fabric, due to its prominent features, was inspired to be selected as the number one candidate for the implementation of the proposed system.

Now, a brief overview of the various components of the Hyperledger is presented in the following subsection.

G. Setup of Hyperledger Components

As a proof of concept, the proposed system is implemented as a consortium blockchain using Hyperledger fabric with the essential components, such as certificate authority, working as a membership service provider, orderer node, anchor node, committer nodes, endorser nodes, etc., carrying out the transactions, creating blocks of transactions, implementing the consensus algorithm, and saving them on the shared ledger, as shown in Fig. 8. The discussion about these components, their roles, and their implementation on Hyperledger fabric are discussed in the following sub-section.

1) Certificate authority: The job of the Certification Authority (CS) is to provide secure communication between the blockchain nodes, authenticate the users, and ensure the integrity and confidentiality of the data. It effectively works as the Membership Service Provider (MSP) that provides access to the peer nodes (blockchain nodes, customers, and service providers) in the system, creates digital certificates along with pairs of private and public keys for encryption and decryption of the messages, ensure trust, and enable registration for the new users. Each peer has its own identity given by a membership services provider (MSP), which authenticates each peer to its channel peers and services.

2) Channel: Channels are used to provide communication between the peer nodes of the blockchain consortium to conduct private and confidential transactions. A channel constitutes of. Each transaction on the network is executed on a channel, which constitutes authenticated and authorized members, anchor nodes, the shared ledger, smart contracts or chaincode, and orderer nodes.

3) Smart Contract: Smart Contracts, also called chaincode in Hyperledger fabric, are responsible for processing transaction requests and determining whether transactions are valid by executing policies. These could be third-party applications connected to the blockchain network, or chaincode can be written in Go, JavaScript (node.js), and eventually other programming languages, such as Java through Restful interfaces. The chaincode implements the logic of the usage and permission control and validates the login defined through the policies.

4) Validation of system chaincode: Validated System Chain-Code (VSCC) instigates the source of transaction in the blockchain, validates each transaction signed by peer nodes, and ensures the endorsement of the transaction is performed by enough member nodes.

5) Anchor peer: The anchor peers are responsible for providing communication between the members of various organizations, such as service providers, permanent nodes of the consortium blockchain, and customer nodes. It uses the gossip protocol that identifies other available member peers, disseminates ledger data across peers on a channel, and brings new peers up to date quickly.
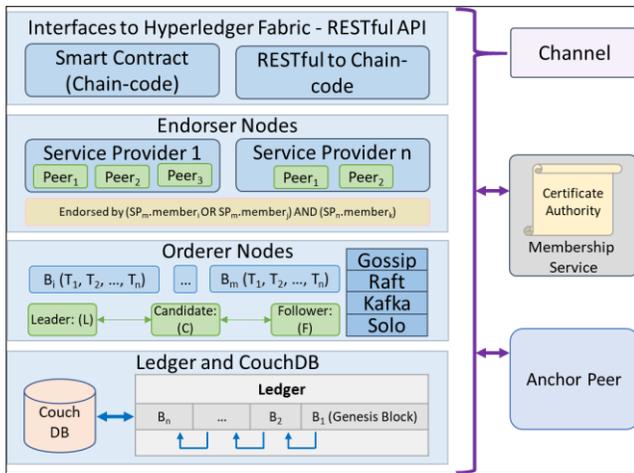
Fig. 8. The Modular Diagram of the Deployed Hyperledger with Various Internal Components.

6) Endorsement policy: The endorsement policy is implemented to ensure that before adding details of transactions into the ledger, the details are first signed by predetermined peers. Every chaincode has an endorsement policy that specifies the set of peers on the channel that must execute the chaincode and vouch for the outcome of the execution for a transaction to be considered valid. To launch different transactions from the smart contract, the multiple interface interaction is ensured through RestAPI.

7) Consensus: There are three main steps of consensus in Hyperledger Fabric. First is the endorsement, which is driven by the policy upon which the members endorse a transaction. Second is ordering, in which the endorsed transactions are arranged in an order to be committed to the ledger. Third is validation which takes a block of ordered transactions and validates the correctness of the results, including checking endorsement policy. A number of consensus mechanisms were opted for and tested among RAFT, Kafka, and Solo; however, in terms of efficiency, Raft has proven to be much better.

8) Byzantine Fault Tolerance (BFT): BFT is a simple consensus mechanism in which every node has to vote. Each node represents a blockchain ledger (Kafka divides these ledgers into clusters on each node to make searching easy). They need to reach a consensus on the current state of the network. This means a good number of the participating nodes have to reach a decision and execute an identical action to avoid failures. The algorithm is Crash Free, which means even if some nodes fail to vote, the algorithm will not crash.

The endorser node in the blockchain consortium endorses every single transaction performed related to the aforementioned activities i.e. the new members joining the system, and transactions related to funds deposited in favor of the service provider. Once the endorsement is performed, the next operation is carried out which is termed ordering. During the ordering process, all endorsed transactions are executed to form a block. That is passed through a validate node that performs validation of the created blocks. Once the blocks are validated, they are approved via a consensus mechanism. A

number of consensus mechanisms exist; however, in this implementation scenario, the default consensus mechanism of Hyperledger Fabric, Raft is used. Raft has a network of nodes that function to update the blockchain ledger. Raft was selected because it is fault-tolerant, in case of a node failure, and efficient as compared to Kafka and Solo. In addition, Raft conducts an election among the nodes for electing a leader in the blockchain network.

Once Raft opts for the leader, the endorser sends the block to the leader which actually ensures data state synchronization among members of the blockchain consortium. Ultimately, the ledger state is updated, synchronized, and trusted.

H. Performance Analysis Results

This paper, in particular, incorporates features with regard to usage control, access control, and permission control for customers associated with the Tours. In order to measure and evaluate the prototype implementation with regard to the latency and throughput of the proposed system, a testbed environment was created that only processes transactions related to usage control and permission control.

The testbed is implemented on two virtual processors of Intel Xeon server as the permanent nodes of the consortium blockchain, along with eight-gigabyte memory. The two nodes are connected through a gigabit ethernet connection. IBM chaincode's fabric-load-gen tool is used with various test cases with varying latencies and endorsement policies. The generated transactions are configured as both read and write by using load-generating scripts. The latency is calculated as the delay in completing a whole transaction, including endorsement time, broadcasting time and commit and validation time. The delay introduced by the orderer nodes is also included, while the number of transactions executed per second is used to calculate the throughput of the system.

Fig. 9 and 10 depict Hyperledger performance over different block sizes, the latencies (Fig. 9) and throughput (Fig. 10) of the transactions for different transaction arrival rates, and the effect of various endorsement policies with various transaction arrival rates. The linear increment in the throughput can be seen in Fig. 10, according to the increase in the rate of incoming transactions. It can be observed that after 150 tps, this linear increase is limited due to the saturation point. The saturation point is presented in Fig. 8. This saturation is mainly due to the three computationally intensive operations of policy validation, which are (1) x.509 certificate identity validation, which involves reconstruction of the certificate from the received serialized byte stream; (2) validation of the identity by the membership service provider and, (3) verification of the digital signature to authenticate the transaction. The size of the block gets bigger as the number of endorsements is increased with each transaction, which includes encoding and decoding of x.509 certificates.

The testbed is configured using Hyperledger Fabric 1.4 which employs Raft consensus. Raft produces more acceptable performance results as compared to other versions of the Hyperledger Fabric. These results support the idea of blockchain-based permission and usage control mechanisms for decentralized content distribution in virtual tourism.
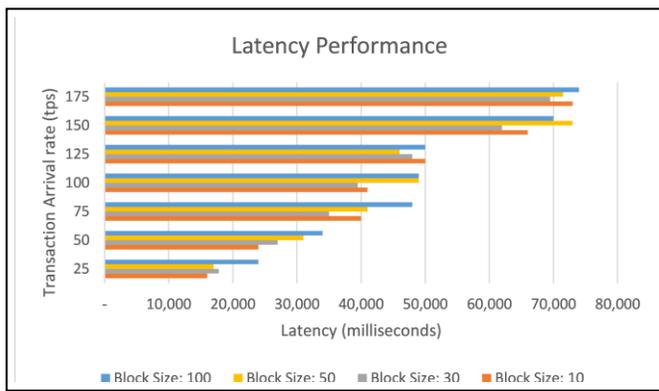
Fig. 9.   Latency Performance of the Deployed Hyperledger with Various-Sized Blocks.
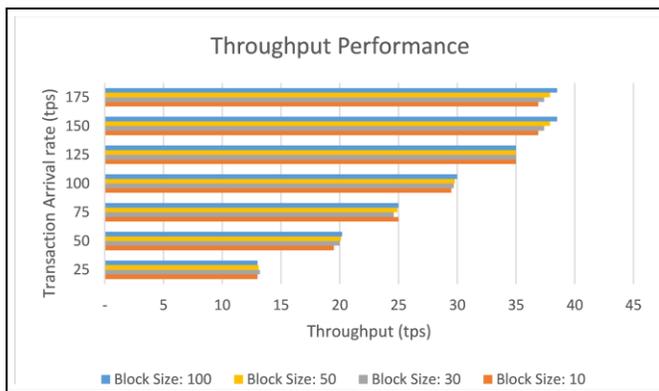
Fig. 10.   Throughput Performance of the Deployed Hyperledger with Various Sized Blocks.

V.   CONCLUSION

To facilitate the virtual tourism industry and its customers, this paper presents an online system that provides a consortium blockchain-based decentralized access to various tourism sites and content providers. Along with the permanent nodes of the blockchain, service providers and customers can become members of the blockchain, which maintains usage, permission, and payment control. It allows the tourism sites to announce their virtual tours and provides accessibility and accountability, at the same time, allows users to register, subscribe, access, and be billed according to their usage. The permission control module ensures authentication and authorization, while the usage control provides accountability and billing. The transactions are stored on the blockchain to ensure the integrity of the data and smart contracts are used to ensure automatic usage and permission control. An implementation on Hyperledger Fabric is provided as a proof of concept with performance measurements as the evaluation.

In the future, permission and rights controls can be enabled for the shared objects in a collaborated virtual reality, where various users can create their virtual objects, enter historical facts, create milestones, and frequently asked questions (FAQs) in the virtual tours. Through this, the users can control and permit who can access their virtual assets. Similarly, reviews and experiences could be added to a virtual tour with the control in the hands of the service provider.

REFERENCES

[1]   Farshid, Mana, Jeannette Paschen, Theresa Eriksson, and Jan Kietzmann. "Go boldly!: Explore augmented reality (AR), virtual reality (VR), and mixed reality (MR) for business." Business Horizons 61, no. 5 (2018): 657-663.

[2]   Alcantud, Francisco, Gerardo Herrera, Gabriel Labajo, I. Dolz, C. Gayá, V. Avila, A. Blanquer, Jose Luis Cuesta, and J. Arnáiz. "Assessing virtual reality as a tool for support imagination." In International Conference on Computers for Handicapped Persons, pp. 143-144. Springer, Berlin, Heidelberg, 2002.

[3]   Pillai, Anitha S., and Prabha Susy Mathew. "Impact of virtual reality in healthcare: a review." Virtual and augmented reality in mental health treatment (2019): 17-31.

[4]   Lányi, Cecília Sik. "Virtual reality in healthcare." In Intelligent paradigms for assistive and preventive healthcare, pp. 87-116. Springer, Berlin, Heidelberg, 2006.

[5]   Freina, Laura, and Michela Ott. "A literature review on immersive virtual reality in education: state of the art and perspectives." In The international scientific conference elearning and software for education, vol. 1, no. 133, pp. 10-1007. 2015.

[6]   Wu, Hsin-Kai, Silvia Wen-Yu Lee, Hsin-Yi Chang, and Jyh-Chong Liang. "Current status, opportunities and challenges of augmented reality in education." Computers & education 62 (2013): 41-49.

[7]   Egger, J., Masood, T.: Augmented reality in support of intelligent manufacturing–a systematic literature review. Computers & Industrial Engineering 140, 106,195 (2020).

[8]   Jamei, Elmira, Michael Mortimer, Mehdi Seyedmahmoudian, Ben Horan, and Alex Stojcevski. "Investigating the role of virtual reality in planning for sustainable smart cities." Sustainability 9, no. 11 (2017): 2006.

[9]   The 360-degree virtual museum tour of The Louvre, https://360stories.com/paris/place/louvre-museum.

[10] These 12 Famous Museums Offer Virtual Tours You Can Take on Your Couch, https://www.travelandleisure.com/attractions/museums-galleries/museums-with-virtual-tours.

[11] de los Ríos, S., Cabrera-Umpiérrez, M.F., Arredondo, M.T., Páramo, M., Baranski, B., Meis, J., Gerhard, M., Prados, B., Pérez, L., del Mar Villafranca, M.: Using augmented reality and social media in mobile applications to engage people on cultural sites. In: International Conference on Universal Access in Human-Computer Interaction, pp. 662–672. Springer (2014).

[12] Poushneh, A.: Augmented reality in retail: A trade-off between user's control of access to personal information and augmentation quality. Journal of Retailing and Consumer Services 41, 169–176 (2018).

[13] Poushneh, A., Vasquez-Parraga, A.Z.: Discernible impact of augmented reality on retail customer's experience, satisfaction and willingness to buy. Journal of Retailing and Consumer Services 34, 229–234 (2017).

[14] Verma, Sanjeev, Lekha Warrier, Brajesh Bolia, and Shraddha Mehta. "Past, present, and future of virtual tourism-a literature review." International Journal of Information Management Data Insights 2, no. 2 (2022): 100085.

[15] Global Virtual Tourism Market Research Report, Industry Analysis & Forecast (2022-2027). https://www.marketdataforecast.com/market-reports/virtual-tourism-market.

[16] Muhammad, S. S., Toqeer, A. S., Nadeem, A., Nawaz, W., & Alkhodre, A. (2022). Virtual tourism and digital heritage: An analysis of VR/AR technologies and applications. International Journal of Advanced

Computer Science and Applications, 13(7) doi:https://doi.org/10.14569/IJACSA.2022.0130739.

[17] Parry, Ross. Recoding the museum: Digital heritage and the technologies of change. Routledge, 2007.

[18] TIG Global, https://www.tourdeforce360.com/advantage_hotel.html.

[19] 360 photos, https://mailchi.mp/8996c0149181/360-stats-sheet.

[20] Panomatics, http://panomatics.net/web/statistics/.

[21] Markletic, https://www.markletic.com/blog/virtual-event-statistics/#virtual-event-preparation.

[22] Trend Exchange, Virtual Events Industry Expected to Grow from $78 Billion to $774 Billion by 2030, https://www.prnewswire.com/news-releases/virtual-events-industry-expected-to-grow-from-78-billion-to-774-billion-by-2030-301095683.html.

[23] Dubai 360 virtual tours, https://www.360virtualtour.co/dubai-360-virtual-tours/.

[24] An interactive journey to the Great Barrier Reef, https://attenboroughsreef.com/.

[25] The Great Wall of China, Google Arts and Culture, https://artsandculture.google.com/story/igVxCi6iJJ6CrA.

[26] Virtual tours: Buckingham Palace, https://www.royal.uk/virtual-tours-buckingham-palace.

[27] Yosemite Virtual Tour, https://www.virtualyosemite.org/.

[28] View Colosseum in 360 virtual tour, https://www.touristtube.com/Things-to-do-in-Rome/Colosseum-360.

[29] Statue of Liberty :: 360° VR Panorama, http://www.samrohn.com/360-panorama/statue-of-liberty-new-york/.

[30] The Holy Land Places in 360-Degree Virtual Reality Tour, https://www.p4panorama.com/gallery-item/the-holy-land/.

[31] Experience Mecca in VR, https://www.oculus.com/experiences/gear-vr/1125286047502859/.

[32] Brave the Himalayas in 'Everest VR: Journey to the Top of the World, https://www.oculus.com/blog/brave-the-himalayas-in-everest-vr-journey-to-the-top-of-the-world/.

[33] 360° VR Space Safari, https://orsted.com/en/explore/space-safari.

[34] The 35 Best Virtual Tours Online So You Can Travel From Home, https://worldwidehoneymoon.com/best-virtual-tours-online-travel-from-home/.

[35] Rouhani, Sara, and Ralph Deters. "Blockchain based access control systems: State of the art and challenges." In IEEE/WIC/ACM International Conference on Web Intelligence, pp. 423-428. 2019.

[36] Kishigami, Junichi, Shigeru Fujimura, Hiroki Watanabe, Atsushi Nakadaira, and Akihiko Akutsu. "The blockchain-based digital content

distribution system." In 2015 IEEE fifth international conference on big data and cloud computing, pp. 187-190. IEEE, 2015.

[37] Di Francesco Maesa, Damiano, Paolo Mori, and Laura Ricci. "Blockchain based access control." In IFIP international conference on distributed applications and interoperable systems, pp. 206-220. Springer, Cham, 2017.

[38] Maesa, Damiano Di Francesco, Paolo Mori, and Laura Ricci. "A blockchain based approach for the definition of auditable access control systems." Computers & Security 84 (2019): 93-119.

[39] Khan, Muhammad Yasar, Megat F. Zuhairi, Toqeer Ali, Turki Alghamdi, and Jose Antonio Marmolejo-Saucedo. "An extended access control model for permissioned blockchain frameworks." Wireless Networks 26, no. 7 (2020): 4943-4954.

[40] Marmolejo-Saucedo, José-Antonio. "An extended access control model for permissioned blockchain frameworks." OPENAIRE (2019).

[41] Park, J., & Sandhu, R. (2004). The UCON ABC usage control model. ACM Transactions on Information and System Security, (TISSEC), 7(1), 128–174.

[42] Khan, Yasar, Toqeer Ali, Megat Fariz, Fernando Moreira, Frederico Branco, José Martins, and Ramiro Gonçalves. "BlockU: Extended usage control in and for Blockchain." Expert Systems 37, no. 3 (2020): e12507.

[43] IBM, "IBM Blockchain Platform known issues," https://cloud.ibm.com/docs/services/blockchain?topic=blockchainknown-issues, 2020, [Online; accessed 30-Oct-2022].

[44] B. R.G, "Introducing r3 corda: A distributed ledger for financial services," https://gendal.me/2016/04/05/introducing-r3-cordaa-distributed-ledger-designed-for-financial-services/, 2016, [Online accessed 30-Oct-2022].

[45] T. Kumar, "What is Hydrachain technology and how it works," https://www.blockchain-council.org/blockchain/what-ishyderachain-technology-how-it-works/, 2020, [Online; accessed 30-Oct-2022].

[46] G. Scholar, "Performance evaluation of the Quorum Blochchain platform," https://arxiv.org/abs/1809.03421, 2020, [Online; accessed 30-Oct-2022].

[47] V. Skwarek, "Blockchains as security-enabler for industrial iotapplications," Asia Pacific Journal of Innovation and Entrepreneurship, vol. 11, no. 3, pp. 301–311, 2017.

[48] Wood, Gavin. "Ethereum: A secure decentralised generalised transaction ledger." Ethereum project yellow paper 151, no. 2014 (2014): 1-32.

[49] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich et al., "Hyperledger fabric a distributed operating system for permissioned blockchains," in Proceedings of the Thirteenth EuroSys Conference, 2018, pp. 1–15.

A Fast Multicore-based Window Entropy Algorithm

Suha S.A. Shokr, Hazem M. Bahig

Information and Computer Science Department, College of Computer Science and Engineering
University of Ha'il, Ha'il 81481, Saudi Arabia

Abstract—Malware analysis is a major challenge in cybersecurity due to the regular appearance of new malware and its effect in cyberspace. The existing tools for malware analysis enable reverse engineering to understand the origin, purpose, attributes, and potential consequences of malicious software. An entropy method is one of the techniques used to analyze and detect malware, which is defined as a measure of information encoded in a series of values based upon the probability of those values appearing. The window entropy algorithm is one of the methods that can be applied to calculate entropy values in an effective manner. However, it requires a significant amount of time when the size of the file is large. In this paper, we solve this problem in two ways. The first way of improvement is determining the best window size that leads to minimizing the running time of the window entropy algorithm. The second way of improvement is by parallelizing the window entropy algorithm on a multicore system. The experimental studies using artificial data show that the improved sequential algorithm can reduce the window entropy method's running time by 79% on an average. Also, the proposed parallel algorithm outperforms the modified sequential algorithm by 77% and has super-linear speed up.

Keywords—Entropy; window method; malware analysis; parallel algorithm; multicore

I. INTRODUCTION

With increased internet use, social media, and data sharing, users must better secure themselves to protect vulnerable information. One of the dangerous software that faces the user is malware.

Malware is a set of instructions that run on a computer, specifically designed to harm the user or the target system by making the system do something that an attacker wants it to do, like steal personal information, delete files, commit fraud, or steal software serial numbers. Moreover, new malware is registered periodically, posing a challenge to cybersecurity efforts. Malware remains one of the most potent threats in cyberspace, despite significant advances in cyber security mechanisms and their ongoing evolution.

Different kinds of malware (e.g., viruses, bots, rootkit, backdoors) can be distributed via various channels, transmitted, and used in various techniques to perform malicious operations [1].

The process of analyzing the malware (with and without execution) is called malware analysis. The main objectives of analyzing malware are to study malicious software's origin, purpose, attributes, and potential impact.

A large number of malware analysis techniques have been proposed. These techniques can be classified into two main categories [2]: static and dynamic. The main difference between static and dynamic analysis is the examination of the malware with or without running it. The two categories contain many malware detection techniques such as signature-based, specification-based, behavioral-based, and heuristic-based [3]. One of the techniques used in malware analysis is the running window entropy (RWE) method, where the entropy is a measure of information encoded in a series of values based upon the probability of those values appearing [1].

The entropy method has been used in many applications such that malware detection [4], lung sound classification [5], gender violence classification [6], Arabic text classification [7], image steganography [8], and sediments quality evaluation [9].

For analyzing malware using entropy algorithm, there are many research papers have been proposed such as [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28].

Lyda and Hamrock [15] analyzed malware utilizing structural entropy, which employs a skip value to ignore the entropy calculation at each index. McMillan and Garman [21] obtained the same result utilizing a statistical decision concept using malware entropy information.

Sorokin [26] proposed a method based on dividing the file into segments and then applying a discrete wavelet transform (DWT) to the structural entropy to compare the segments of the input file. Baysa et al. [11] used the same strategy to identify the malware changes. The method was implemented on binary files, and there was no need to perform any preprocessing to detect the malware.

Bat-Erdene et al. [12] used the entropy method to detect unknown packers. The method is based on representing the entropy value as a symbolic aggregate approximation (SAX) and measuring the symbol's similarity in the SAX sequence. The accuracy of this method on the tested data is 95%. Radkani et al. [25] suggested a new method to detect metamorphic malware based on entropy and dissimilarity measures. The drawback of this method is the measurement limitation, which is based on the opcode (operation code) frequency.

Jones and Wang [19] suggested a method to reduce the running time of the RWE algorithm. The method is based on removing a nested "for" loop from the old version of the algorithm. The same authors [20] used the entropy concept to extract the feature set used in machine learning algorithms.

Menéndez et al. [22] introduced a new concept called entropy time series to detect malware. The technique uses a time series to represent an entropy signature. Then, the

technique applies wavelet analysis to a file's entropy to obtain a simplified signature. The main advantage of this strategy is reducing the running time of malware analysis.

In light of the previous discussions about the use of entropy in malware analysis, the main question of the research is: how to reduce the running time of the entropy method when the size of the file is large?

This paper has focus on the RWE algorithm because the window method is one of the strategies used in different fields, such as [29, 30], to reduce the running time. This study aims to demonstrate the effect of changing the window size on the RWE algorithm in terms of running time. In addition, the paper explores if there is a common window size that reduces the running time of the RWE algorithm for all malware samples and look at ways to optimize the RWE algorithm's running time. Finally, how to parallelize the sequential algorithm to reduce the running time.

The remainder of this paper is organized as follows. Section II gives briefly the foundation concepts of the RWE algorithm. In Section III, the modified RWE algorithm in terms of execution time is given. The second improvement for RWE algorithm using a multicore system is given in Section IV. The experiments and their analysis for the proposed algorithms on artificial data are given in Section V. Finally, the conclusion and open question are given in Section VI.

II. RUNNING WINDOW ENTROPY METHOD

In this section, we briefly give the main concept of entropy in subsection A. In subsection B, we give the main idea and steps of the RWE method that are used to calculate the entropy values of the malware sample.

A. Entropy Concept

Shannon [1] introduced the concept of entropy in information theory to measure the degree of randomness in a data set. The general formula for Shannon entropy is given by

$$H(X)= -\sum_{x \in \Sigma} p(x) \log p(x) \qquad (1)$$

where

- X is a discrete random variable and has values belonging to the alphabet Σ,

- Σ is the alphabetic domain of the malware sample, and

- $p(x)$ is the probability distribution of the value x and has a value belonging to the range [0, 1].

In information theory, it is essential to recognize that more relevant information is present in the data when the entropy value is higher than when it is lower. This means that a higher entropy value denotes more randomness and useful, unpredictable information.

B. Running Window Entropy Method

Given a malware sample D of size n as $D=(d_1, d_2, d_3, \dots , d_n)$. The RWE algorithm, is a method based on dividing sample D as continuous overlap windows, each window of size w. A window of size w that starts at index i in the data D represents as follows.

$$W_i =(d_i, d_{i+1}, d_{i+2}, \dots , d_{i+w-1}), \qquad (2)$$

where $1 \leq i \leq n-w+1$.

Since the malware sample represents a sequence of bytes, each byte ranges from 0 to 255. Therefore, the entropy for W_i can be calculated as follows:

$$H_i=H(W_i)= -\sum_{0 \leq k \leq 255} p(k) \log p(k) \qquad (3)$$

where $p(k)$ is the probability of byte k within W_i and is equal to the total number of the byte k (symbol) within W_i, t_k, to the size of the window; i.e. $p(k)=t_k/w$.

In the case of the byte, k, does not appear in the window W_i, then $t_k =0$ and the method ignores the log $p(k)$ value because the logarithm of zero value is undefined.

Therefore, the malware sample after applying the RWE algorithm represents a sequence of values.

$$H =(H_1, H_2, H_3, \dots , H_{n-w+1}) \qquad (4)$$

The algorithm uses two arrays:

- H is an array of size n-w+1 and the value of H_i represents the entropy value for the window number i, $1 \leq i \leq n-w+1$.

- t is an array of size 256 and the value of $t(j)$ represents the total number of the byte j in a certain window.

The RWE algorithm consists of n-w+1 iterations, $1 \leq i \leq n-w+1$, as follows:

Step 1: Initialize the array t with zero.

Step 2: Compute $t(j)$ for each symbol in the window W_i by incrementing $t(j)$ with 1, where $0 \leq j \leq w-1$.

Step 3: Calculate the entropy, *e*, of the window W_i using (3) if $t(j) \neq 0$ and $0 < j < 256$.

Step 4: Calculate $H_i=e/K$, where K=8.

The reason for taking the value of K equal to 8 is the values of (3) between 0 and 8.

III. IMPROVED RWE METHOD

In this section, we introduce two comments on the RWE algorithm, then modify the RWE algorithm.

The first comment on the RWE algorithm is that the window size was selected experimentally from 256 to 2048 as in [19]. No experimental studies have addressed the effect of window size when w<256 or w>2048. What is the effect of these values of w on the RWE algorithm? Does the same behavior occur if we start from a window size equal to 4?

The second comment on the RWE algorithm is that when the size of window is less than 256, why do we take the summation of (3) from 0 to 255? Can we propose a fast method to calculate H_i?

For the second comment, we propose a modification on RWE algorithm as follows. If the value of w is less than 256, then we can compute (3) as follows.

$$H_i=H(W_i)= -\sum_{k \in W_i} p(k) \log p(k) \qquad (5)$$

The difference between (3) and (5) is the boundaries of the summation that are taken over the term p(k) log p(k).

The following steps computes the value of H_i as follows.

- Compute t(j) for each symbol in the window W_i by increment t(j) with 1, where $1 \leq j \leq w$.

- For each symbol (byte) in the window W_i do the following: if t(j) $\neq 0$ then update the value of the entropy, *e*, for the j-th symbol of window W_i and set t(j) equal to 0.

The reason for setting t(j)=0 after updating the entropy is that the symbol (byte) j may exist many times in the window W_i. Setting the value of t(j)=0, will remove the step of initializing the array t for each new iteration.

The steps of the modified window entropy method, MRWE, are given in Algorithm 1.

Algorithm 1: Modified Running Window Entropy (MRWE)

Input: Malware sample D of size n, window size w, K=8

1.   If w<256 then
2.   Initialize t with 0
3.   For i=1 to n-w+1 do
4.   For j=0 to w-1 do
5. $t(d_{i+j}) = t(d_{i+j})+1$
6.   End for
7.   e=0
8.   For j=0 to w-1 do
9.   If $t(d_{i+j}) \neq 0$ then
10. $e=e-(t(d_{i+j})/w*\log(t(d_{i+j})/w))$
11. $t(d_{i+j})=0$
12.   End if
13.   End for
14.   H(i)=e/K
15.   End for
16.   Else
17.   For i=1 to n-w+1 do
18.   For j=0 to w-1 do
19. $t(d_{i+j}) = t(d_{i+j})+1$
20.   End for
21.   e=0
22.   For j=0 to 256 do
23.   If t(j)$\neq 0$ then
24. $e=e-(t(j)/w*\log(t(j)/w))$
25.   End if
26.   End for
27.   H(i)=e/K
28.   End for
29.   End if

Output: Entropy values H=($H_1,H_2,...,H_{n-w+1}$)

Note that the first comment will be discussed in the experimental section.

IV. PARALLELIZING MRWE ALGORITHM

In this section, the parallelization of the MRWE algorithm (similarly RWE algorithm) on a parallel shared memory model is discussed. The parallel model consists of multi-

processors/multi-threads that can be communicate via a global memory.

The idea of parallelizing MRWE, PMRWE, is based on dividing the input array D into t subarrays. Each subarray, $D_i=(d_{(i-1)\alpha+1}, d_{(i-1)\alpha+2}, d_{(i-1)\alpha+3},..., d_{i\alpha})$, consists of α elements approximately, where $\alpha= (n-w+1)/t$ and $1 \leq i \leq n/t$. The value of numerator of α represents the number of windows in D. Then each thread, t_i, works on the subarray D_i to compute the entropy of this part. Each thread generates α entropy values, each value represents the entropy value for a window in D_i.

All the threads work on their subarray simultaneously. This means that no shared data between the threads. The complete pseudocode for PMRWE algorithm is given in Algorithm 2.

Algorithm 2: Parallel MRWE (PMRWE)

Input: Malware sample D of size n, window size w, K=8, and t threads.

1. $\alpha= (n-w+1)/t$
2.   If w<256 then
3.   Initialize t with 0
4.   For b=1 to t Do Parallel
5.   For i=1 to α do
6.   For j=0 to w-1 do
7. $t(d_{(b-1)\alpha+i+j}) = t(d_{(b-1)\alpha+i+j})+1$
8.   End for
9.   e=0
10.   For j=0 to w-1 do
11.   If $t(d_{(b-1)\alpha+i+j}) \neq 0$ then
12. $e=e-(t(d_{(b-1)\alpha+i+j})/w*\log(t(d_{(b-1)\alpha+i+j})/w))$
13. $t(d_{(b-1)\alpha+i+j})=0$
14.   End if
15.   End for
16.   H((b-1)α+i)=e/K
17.   End for
18.   End for parallel
19.   Else
20.   For b=1 to t Do Parallel
21.   For i=1 to n-w+1 do
22.   For j=0 to w-1 do
23. $t(d_{(b-1)\alpha+i+j}) = t(d_{(b-1)\alpha+i+j})+1$
24.   End for
25.   e=0
26.   For j=0 to 256 do
27.   If t(j)$\neq 0$ then
28. $e=e-(t(j)/w*\log(t(j)/w))$
29.   End if
30.   End for
31.   H(i)=e/K
32.   End for
33.   End for parallel

Output: Entropy values H=($H_1,H_2,...,H_{n-w+1}$)

V. EXPERIMENTAL STUDIES

The aim of this section is to study experimentally all proposed algorithms, sequential and parallel, and compare them with the original algorithm. The section consists of three

subsections. In the first subsection, we describe briefly the platform specifications and data set used in the implementation. The sequential and parallel comparisons are providing in the second and third subsections, respectively.

A. Platform Specifications and Data

All algorithms, RWE, MRWE, PRWE, and PMRWE, have been implemented using the programming language Python. The Python code for the RWE algorithm exist in [17]. All methods run on a machine of type Intel Core i7 with 8 logical processors speeds 3.1 GHz. The machine is able to run eight threads concurrently. The machine worked under the Windows operating system.

For sequential comparison, the running time of each sequential algorithm is measured by fixing n and w first. Then, generate a malware sample of size n and run the algorithms ten times on the malware sample. This process is repeated 20 times for different malware samples. Finally, the average time for all instances, i.e., the average of 200 runs, is calculated. The running time is measured in seconds.

The data sizes of malware samples are 1k, 2k, 4k, 8k, 16k, and 32k, whereas the window size is w=4, 8, 16, 32, …., and n/2.

For parallel comparison, the number of threads used in the comparison are 2, 4, and 8. The data sizes for malware samples are ½ MB, 1 MB, 2 MB, 4 MB, and 8 MB. The value of w is equal to w_{min}, where w_{min} is the size of window that minimize MRWE algorithm. The value of w_{min} is calculated experimentally in the second subsection.

B. Sequential Comparison

In this subsection, we study the two algorithms, RWE and MRWE, experimentally for two goals. The first goal is to answer the first comment mentioned in the previous section: what is the effect of changing the value of the window, w, on the RWE and MRWE? The second goal is: which value of w leads to minimal time for the method?

The results of running the two methods are shown in Fig. 1. From the analysis of the subfigures in Fig. 1, the behaviors of the two methods is divided into three ranges based on w as follows.

Case 1: w in the range [4,32]. In this case, the following observation were made.

- The running time for the MRWE is less time than the RWE in the case of w=4, 8, 16, and 32 for all values of n.

- The difference between the two running times for MRWE and RWE decreases with increasing the value of w.

- The improvement percentages of MRWE compared to RWE are shown in Table I. The average percentage of improvements are 79%, 66%, 45%, and 20% for w=4, 8, 16, 32, respectively.

Case 2: w in the range [64,256[. In this case, the following observation were made.

- The running time for the RWE is less than the MRWE algorithm in the case of w=64 and 128 for all values of n.

- The difference between the two methods when w=64 is very small and approximately equal to 5%.

- When w=128, the RWE algorithm improves the MRWE algorithm by 25% on average.

Case 3: w in the range [256, n]. In this case, the two methods are equal because the two methods use the same strategy.

From the observations of the three cases, we conclude that the MRWE algorithm can be achieve the minimal time when the window size is small.

TABLE I. PERCENTAGE OF IMPROVEMENTS FOR MRWE

n	w			
	4	8	16	32
1k	79.49%	66.14%	44.95%	20.05%
2k	79.19%	66.79%	45.83%	19.57%
4k	79.86%	66.45%	45.84%	20.29%
8k	79.29%	65.77%	45.77%	20.02%
16k	79.75%	66.06%	45.75%	20.57%
32k	79.06%	65.40%	45.83%	19.52%

C. Parallel Comparison

The results of running the two parallel algorithms, PRWE and PMRWE, on a multicore system using 2, 4, and 8 threads are shown in Fig. 2. The results show that the running time for PMRWE algorithm outperforms the PRWE algorithm when w_{min}=4. For example, the running times for PRWE and PMRWE algorithms are 39 and 17.6 seconds, respectively, when n=1 MB and t=2; see Fig. 2(a). From the experimental results, the percentage of improvements for the PMRWE algorithm over the PRWE algorithm using threads two, four and eight are 79.3%, 77.2%, and 77.8%, respectively, on average; see Table II.

Also, we can compute the speed up of the PMRWE algorithm by computing the ratio between the running time of MRWE algorithm and PMRWE algorithm. The results of speed up of PMRWE algorithm is given in Fig. 3. The value of speed up when t=2 is linear and this value increase to be super-linear speed up when t=4 and 8. The increasing in the speed up came from the size of file is divided into small segment when t>2.
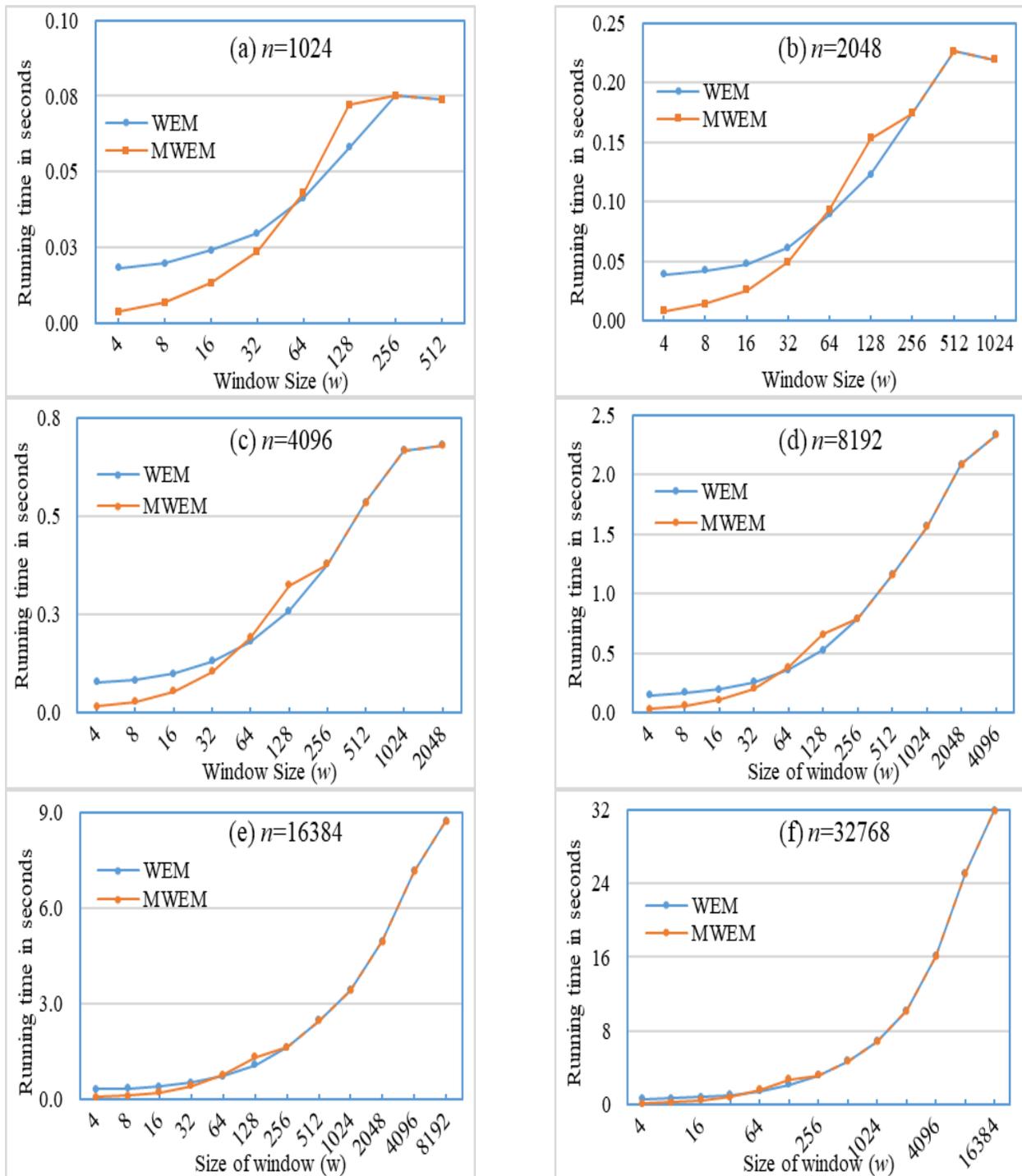
Fig. 1.   Running Time Comparison between RWE and MRWE Algorithms.

Fig. 3.   Speed up of PMRWE Algorithm.

VI. CONCLUSION

Malware analysis is important to study malicious software's functionality, purpose, origin, and potential impact. The running window entropy method is one of the methods used to find the entropy values of the malware sample. The main drawback of this method is the amount of time when the size of file is large. In this paper, two improved algorithms are given. The first improved algorithm is based on determining the best window size to minimize the running time of the window entropy method. The second improved algorithm is the parallelization of the improved sequential algorithm.

The experimental studies for different malware sizes and best window size show that the improved sequential and parallel methods outperform the original and the improved sequential methods, with 79% and 77% faster time, respectively.

There are many open questions related to improve the entropy method for malware analysis as follows: (1) How to use other high-performance systems, such as distributed system, GPU (graphic processing unit) and cloud, to speed up the execution time? (2) How to extend this work to other method for window entropy?

REFERENCES

[1] Panda Security, "2017 Cybersecurity Predictions," 2017.

[2] E. Gandotra, D. Bansal, S. Sofat, " Malware analysis and classification: a survey," J. of Inf. Security, 5, pp. 56-64, 2014.

[3] I. Kara, "A basic malware analysis method," Computer Fraud & Security, vol. 2019, Issue 6, pp. 11-19, 2019.

[4] Mafaz Mohsin Khalil Al-Anezi, "Generic Packing Detection Using Several Complexity Analysis for Accurate Malware Detection" International Journal of Advanced Computer Science and Applications, 5(1), 2014.

Fig. 2.   Time Comparisons between PRWE and PMRWE Algorithms.

TABLE II.   PERCENTAGE OF IMPROVEMENT FOR PMRWE ALGORITHM

n	Number of threads		
	2	*4*	*8*
0.5M	79.6%	76.9%	75.0%
1M	80.2%	80.0%	80.0%
2M	79.3%	74.0%	81.4%
4M	79.0%	75.8%	75.0%
8M	78.4%	79.4%	77.7%
Avg.	**79.3%**	**77.2%**	**77.8%**

[5] Achmad Rizal, Risanuri Hidayat and Hanung Adi Nugroho, "Comparison of Multilevel Wavelet Packet Entropy using Various Entropy Measurement for Lung Sound Classification" International Journal of Advanced Computer Science and Applications(IJACSA), 10(2), 2019.

[6] Abdul Azim Ismail and Marina Yusoff, "An Efficient Hybrid LSTM-CNN and CNN-LSTM with GloVe for Text Multi-class Sentiment Classification in Gender Violence" International Journal of Advanced Computer Science and Applications, 13(9), 2022.

[7] Ibrahim S Alkhazi and William J. Teahan, "Classifying and Segmenting Classical and Modern Standard Arabic using Minimum Cross-Entropy" International Journal of Advanced Computer Science and Applications(IJACSA), 8(4), 2017.

[8] Ke-Huey Ng, Siau-Chuin Liew and Ferda Ernawan, "An Improved RDWT-based Image Steganography Scheme with QR Decomposition and Double Entropy" International Journal of Advanced Computer Science and Applications(IJACSA), 11(3), 2020.

[9] Alexi Delgado, Betsy Vilchez, Fabian Chipana, Gerson Trejo, Renato Acari, Rony Camarena, Víctor Galicia and Chiara Carbajal, "Applying Grey Clustering and Shannon's Entropy to Assess Sediment Quality from a Watershed" International Journal of Advanced Computer Science and Applications(IJACSA), 12(9), 2021.

[10] D. Baysa, "Structural entropy and metamorphic malware," J. Comput. Virol. Hacking Tech., vol. 9, no. 4, pp. 179–192, 2013.

[11] M. Bat-Erdene, H. Park, H. Li, H. Lee, M. Choi, "Entropy analysis to classify unknown packing algorithms for malware detection," Int. J. Inf. Secur. 16, pp. 227–248, 2017.

[12] M. Bat-Erdene, T. Kim, H. Li, and H. Lee, "Dynamic classification of packing algorithms for inspecting executables using entropy analysis," Proc. 8th Int. Conf. Malicious Unwanted Softw. "The Am. MALWARE, pp. 19–26, 2019.

[13] A. Chakrabarti, G. Cormode, and A. Mcgregor, "A near-optimal algorithm for estimating the entropy of a stream," ACM Trans. Algorithms, vol. 6, no. 3, pp. 1–21, 2020.

[14] K. Han, J. H. Lim, and E. G. Im, "Malware analysis method using visualization of binary files," Proceedings of the Research in Adaptive and Convergent Systems. ACM, Montreal, Quebec, Canada, 2019.

[15] R. Lyda and J. Hamrock, "Using entropy analysis to find encrypted and packed malware," IEEE Security & Privacy, vol. 5, no. 2. pp. 40–45, 2021.

[16] A. Lall, V. Sekar, M. Ogihara, J. Xu, and H. Zhang, "Data streaming algorithms for estimating entropy of network traffic," SIGMETRICS Perform. Eval. Rev., vol. 34, no. 1, pp. 145–156, 2006.

[17] K. Jones, "Algorithms #1 python source code," 2017. [Online]. Available: https://github.com/keithjjones/csc705-alg1.

[18] K. Jones, "malgazer," 2017. [Online]. Available: https://github.com/keithjjones/malgazer.

[19] K. Jones , Y. Wang, "An optimized running window entropy algorithm," National Cyber Summit (NCS), 5-7 June 2018, pp. 72-77, 2018.

[20] K. Jones , Y. Wang, "Malgazer: An automated malware classifier with running window entropy and machine learning," 2020 Sixth International Conference on Mobile And Secure Services (MobiSecServ), 22-23 Feb., pp 1-6, 2020.

[21] C. Mcmillan, J. Garman, "System and method for determining data entropy to identify malware," Application PCT/US2008/051383.

[22] H. Menéndez, S. Bhattacharya, D. Clark, E. Barr, "The arms race: adversarial search defeats entropy used to detect malware," Expert Systems With Applications, 118: pp 246–260, 2019.

[23] S. Naval, V. Laxmi, M. S. Gaur, and P. Vinod, "ESCAPE: entropy score analysis of packed executable," Proc. Fifth Int. Conf. Secur. Inf. Networks, pp. 197–200, 2017.

[24] M. Paavola, "An efficient entropy estimation approach," Universtity of Oulu, 2011.

[25] E. Radkani, S. Hashemi, A. Keshavarz-Haddad, M. Haeri, "An entropy-based distance measure for analysing and detecting metamorphic malware," Appl Intell, 48, pp. 536–1546, 2018.

[26] I. Sorokin, "Comparing files using structural entropy," J. Comput. Virol., vol. 7, no. 4, pp. 259–265, 2011.

[27] S. Treadwell and M. Zhou, "A heuristic approach for detection of obfuscated malware," in 2009 IEEE International Conference on Intelligence and Security Informatics, 2009, pp. 291–299.

[28] M. Weber, M. Schmid, M. Schatz, and D. Geyer, "A toolkit for detecting and analyzing malicious software," in Computer Security Applications Conference, 2002. Proceedings. 18th Annual, 2002, pp. 423–431.

[29] H. M. Bahig, K.A. Fathy. An efficient parallel strategy for high-cost prefix operation. J Supercomput 77, pp. 5267–5288, 2021.

[30] H. M. Bahig, H. M Bahig, K. A. Fathy, "Fast and scalable algorithm for product large data on multicore system," Concurrency and Computation: Practice and Experience, vol. 33, Issue 2, 2021 Wiley.

Routing with Multi-Criteria QoS for Flying Ad-hoc Networks (FANETs)

Ch Naveen Kumar Reddy[1]

Research Scholar, Dept. of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation, AP, India

Dr. Krovi Raja Sekhar[2]

Professor, Dept. of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation, AP, India

Abstract—**Flying Ad-hoc Network (FANET) is a type of Ad-hoc network on backbone of Unmanned Aerial Vehicle (UAV). These networks are used for providing communication services in case of natural disasters. Dynamic changes in link quality and mobility distort the Quality of Service (QoS) for routing in FANETs. This work proposed a Multi Criteria QoS Optimal Routing (MCQOR) guided by prediction of link quality and three-dimensional (3D) movement of FANET nodes. The network is clustered based on prediction of movement of nodes. Over the clustered topology, routing path is selected in reactive manner with joint optimization of packet delivery ratio, delay, and network overhead. In addition, cross layer feedback is used to reduce the packet generation rate and congestion in network. Through simulation analysis, the proposed routing protocol is found to have 3.8% higher packet delivery ratio, 26% lower delay and 14% lower network overhead compared to existing works.**

Keywords—Flying ad-hoc Network; multi-criteria QoS; unmanned aerial vehicles; joint optimization

I. INTRODUCTION

FANET is a kind of ad-hoc network formed by Unmanned Aerial Vehicles (UAV). These UAVs are small sized, light weight flying nodes with dynamic speed and altitude. The network is characterized by frequency disconnections of links and partitioning of network. High-speed three-dimensional mobility introduces frequent link distortions in the network [1]. Unreliable packet delivery and high end to end latency are very common in FANET networks. Most traditional routing protocols designed for Mobile Ad-hoc Networks (MANETs) and Vehicular Ad-hoc Networks (VANETs) does not address the challenges in FANET. Challenges like flying in three dimensions, high mobility, low node density, rapid topology changes, unstable links, network partitioning and limited resources makes it difficult to achieve higher QoS in FANET [2].

The major factors to be considered in designing routing protocols for FANET are high mobility, unstable links, and low energy capacity of nodes. UAV nodes have greater mobility ranging from 30 to 460km/hour [3]. The higher mobility creates low link quality. Frequent link disconnections and network partitions increase route discovery and maintenance overhead [4]. The paths must be planned based on predicted positions of UAV to ensure higher packet delivery ratio and lower latency. Also packet collisions and congestion must be effectively controlled to ensure higher QoS. UAV nodes have limited energy and the routing

protocols must not drain the energy faster. Depleting the energy faster can create network partitions and reduce the reliability of the network. Most existing routing protocols for FANET (detailed in Section II) does not address all these factors of high mobility, unstable links, and low energy capacity of nodes in designing multi criteria QoS optimal routing paths. The joint optimization of multiple QoS criteria like delivery ratio, delay and network overhead has not been considered in most of the existing FANET routing protocols. This work addresses these problems.

This work proposes a multi criteria QoS optimal routing protocol for FANET. The network is clustered to accommodate dynamic topology and unstable links. The clustering is done based on prediction of three-dimensional mobility of nodes. Over the clustered topology, geographic routing with next relay selection based on joint optimization of multi criteria is done. Cross layer feedback of network dynamics is used to dynamically adjust the source rate of packet generation. By this way congestion is reduced in the network.

A. Following are the Novel Contributions of this Work

1) A novel clustering topology for FANET based on prediction of three-dimensional mobility of UAV nodes. Many existing works have been proposed for clustering in FANET based on current mobility. Different from it, this work proposes a clustering topology with joint consideration of density and predicted mobility. Nodes with higher stability based predicted mobility is selected as cluster heads and network is clustered based on them.

2) A novel geographic routing with next hop relay selection based on joint optimization of multi criteria QoS factors. Compared to links state based routing protocols, the proposed geographic routing has lower overhead. Relay selection in proposed work is based on QoS considering both delay and link reliability compared to distance based relay selection is existing geographic routing protocols.

3) A novel source rate control based on probability function of delay distribution to vary the packet flow on the routing path adaptive to traffic and application characteristics. By this packet loss due to congestion is reduced and reliability of the routing path is increased.

The paper is organized as follows. Section II presents the survey of exiting routing protocols for FANET. Section III presents the proposed multi criteria QoS optimal routing

protocol for FANET. Section IV presents the results of proposed routing protocol and its comparison to existing routing protocols. Section V presents the conclusion and scope of future research.

II. RELATED WORK

Lin et al. [5] proposed a shortest path routing algorithm based on grid position for FANET. Instead of Euclidean Distance, logical grid distance is used in route computation which is less sensitive to high mobility of UAV nodes. Adjacency relationship is constructed automatically, and Dijkstra's shortest path algorithm is used to construct the routing paths. Routing paths are further optimized using a region reconstruction strategy. The route selection considered link stability as the only criteria to be maximized. Xie et al. [6] proposed a new form of optimized link state routing protocol to handle dynamic topology changes in FANET. Link expiration time was predicted based on current position. The next hops were selected with higher link expiration time and higher residual energy. The network overhead to maintain the OLSR routing table is higher in this approach. Hou et al. [7] constructed OLSR routing table with trajectory prediction information of UAV. The optimal route is selected from OLSR using Q-learning. The approach considered link stability as the only criteria in route selection. Yang et al. [8] proposed a routing protocol for FANET using Q-learning and fuzzy logic. Routing path selection is based on fuzzy logic decision on both link and path performances.

Link parameters considered were transmission rate, energy state and flight status between neighbor UAV's. Path parameters considered were hop count and packet delivery time. The routing was done with objective of maximizing the lifetime. Oubbati et al. [9] proposed a route discovery mechanism for FANET. The routing paths with balanced energy consumption, link breakage prediction and higher connectivity degree are found. Packet delay and reliability were not considered in route discovery process. Sharma et al. [10] proposed a Distributed Priority Tree-based Routing protocol (DPTR) for FANET. The routing protocol addresses the problem of network partitioning and organizes the nodes in form of distributed Red-black tree. The appropriate node and channel for relaying are selected depending on application requirement. The approach considered link stability as the only parameters in routing decision. Usman et al. [11] proposed a forwarding protocol for FANET with next hop selection based on lifetime improvement.

A forwarding zone is created to spot the front relative nodes through forwarding angle and effective approximate node is selected with higher energy state. Packet reliability and delay were not considered in forwarding decision. Zafar et al. [12] proposed a multi cluster FANET topology to reduce the communication cost and optimize the network performance. Within clusters, transmission slot scheduling is done to reduce the packet collision and improve energy utilization. Authors considered only scalable network as the criteria in clustering decision and did not focus on routing protocols. Mariyappan et al. [13] proposed an energy aware routing protocol for FANET with Residual Energy Minimum (REM), Total Residual Energy (TRE), and Hop count as a

routing cost metric. The next hop is selected based the routing metric, so that routing path is energy balanced. Packet reliability, congestion and delay were not considered in the routing decision. Khan et al. [14] explored various topology-based routing protocols for FANET. Protocols were analyzed in terms of throughput, end-to-end delay, and network load. Authors found that zonal routing protocol were found to have lower delay and better throughput compared to other protocols.

But study did not consider various mobility models. Zhang et al. [15] proposed a three-dimensional Q-learning routing protocol to improve QoS and guarantee packet delivery ratio. The proposed solution has two important functionalities: link state prediction and routing decision. Three-dimensional mobile pattern is predicted to calculate link lifetime. Routing path is constructed with higher link lifetime. Energy, congestion, and packet delay etc. were not considered in the routing decision. Usman et al. [16] proposed a reliable link adaptive position-based routing protocol for FANET. The routing is based on multiple criteria of node speed, signal strength, energy, and geographic distance towards destination. The geographic distance is calculated along the forwarding angle. The relay node is optimal in terms of energy, relative speed and connectivity duration is selected and packets are forwarded. The forwarding angle is decided at source and there is no deviation in next hop. This results in sub optimal routes. Choi et al. [17] proposed a multi hop geo location-based routing protocol for FANETs. The method uses the geo-location of neighbors and finds the next hop to destination based only neighbor information. Though the method has low overhead and robustness to dynamic network topology, it does not include other factors like stability, congestion, energy availability in the path. Li et al. [18] proposed a preemptive protocol based on link stability.

Authors introduced a stability metric based on link quality, safety degree and mobility predictor factor. Multiple robust link disjoint paths are computed, and route maintenance is triggered when link is broken. Routing is done by selecting the most reliable path. The overhead for route maintenance is high and the method is suitable only for limited mobility scenarios. Hong et al. [19] proposed a topology change aware routing protocol. The topology changes are monitored continuously, and route is constructed based on the mobility mode.

Among the multiple routing modes, the best one is selected based on the predicted topology changes. The thresholds used in topology change monitoring in this approach is not generic and applies only for certain mobility models. Pu et al. [20] proposed link state routing protocol adaptive to link quality and traffic load. Link quality is estimated using statistical information of received signal strength. Traffic load is estimated using MAC layer channel contention information. The path with higher reliability and light load is selected for routing. The maintenance overhead is high in this approach. Zhang et al. [21] proposed a deep-reinforcement-learning-based geographical routing protocol based on link stability and energy prediction. The next hop selection decision was done using reinforcement learning. The reward for reinforcement learning was on maximization of packet delivery ratio, without consideration for factors like delay, overhead

reduction etc. Souza et al. [22] used fuzzy logic to improve the route discovery process in FANET. The routing decision is based on received signal strength indicator and mobility level. The factors like load on routing path, delay etc. was not considered in next hop selection in this work. He et al. [23] proposed an opportunistic routing protocol with distributed decision making. Each node exchanges aeronautical data with other nodes. From this, each node calculates the transfer probabilities to different neighbors jointly considering the position of neighbors and the destination node. Each node selects the neighbors with higher transfer probabilities as the next hop relay nodes for routing. The overhead in exchange of aeronautical data is higher in this approach. Xinchen et al. [24] used Q-learning for routing decision in FANETs. Nodes

exchange routing metric information with neighbors. Based on this information, reward function which maximizes reliability is triggered to select the next hop node. Authors considered reliability as the only metric in the routing decision.

From the survey, most of the FANET routing protocols are OLSR based with high maintenance overhead. Very few opportunistic routing protocols are available but their relay selection policies are very limited and they don't consider multi criteria QoS factors in relay selection. Also after the route path selected, there is no continuous adaptation of packet rate based on application/network dynamics. Based on these observations, this work proposes a routing protocol for FANET.
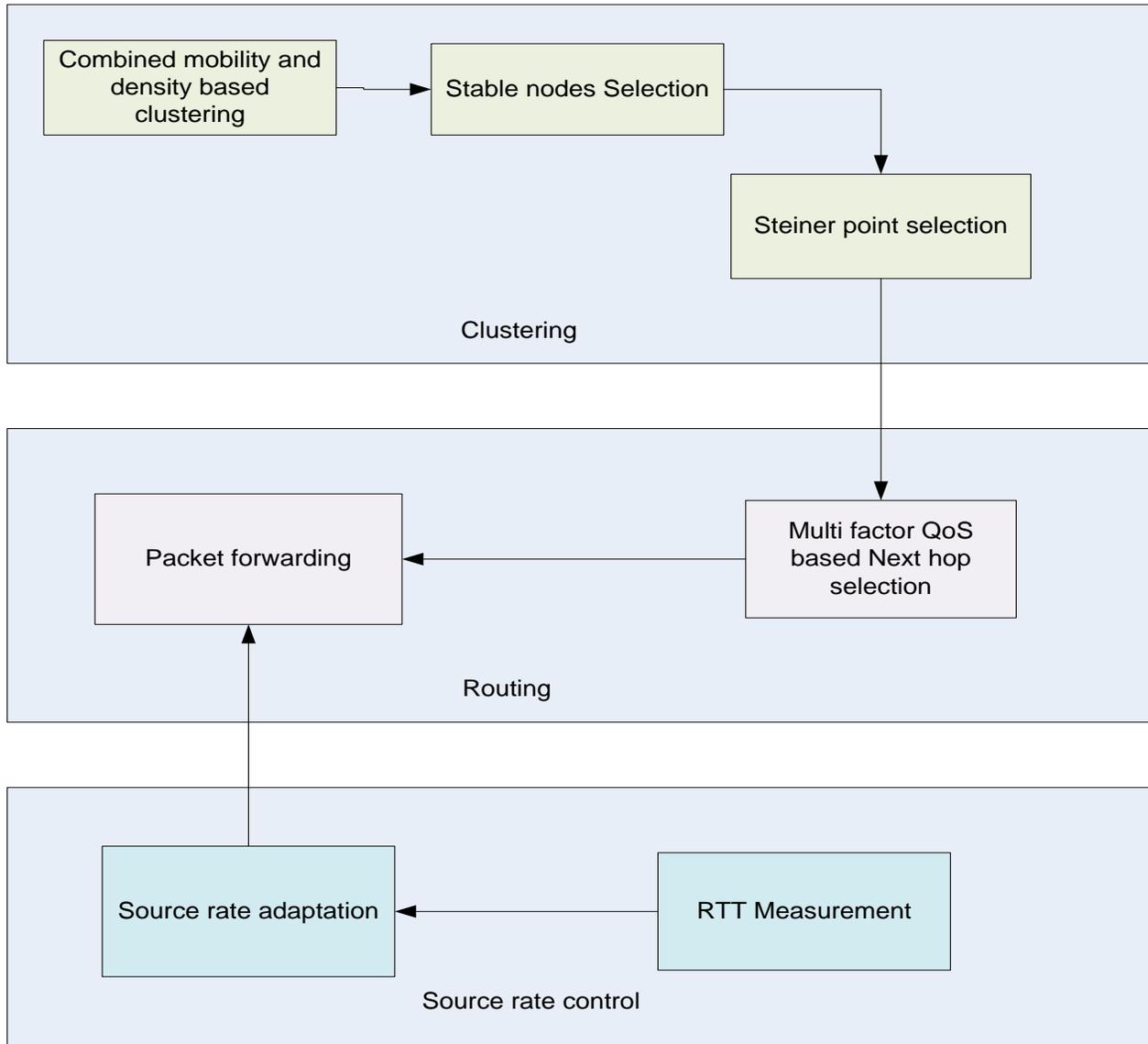
Fig. 1. Architecture of Multi Criteria QoS Optimal Routing.
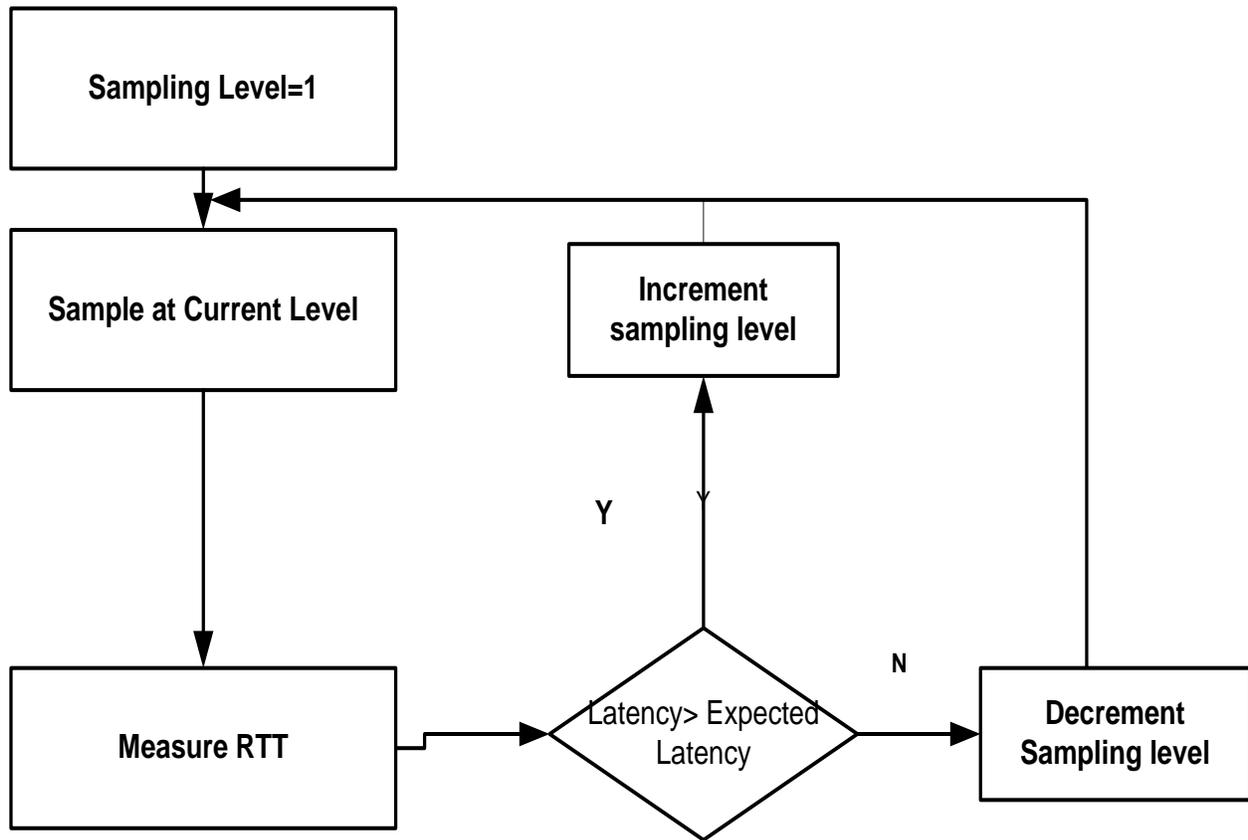
Fig. 2.   Source Rate Control.

III.   MULTI CRITERIA QOS OPTIMAL ROUTING

The proposed Multi Criteria QoS Optimal Routing (MCQOR) is constructed on clustering topology. The functionality of the proposed solution is given in Fig. 1. The proposed solution has three stages: clustering, routing, and source rate control.

The FANET network is first clustered using combined mobility and density-based clustering. The stable nodes in the clusters are selected and a routing backbone optimally connecting these stable nodes is constructed using Steiner Tree algorithm in 3D space. Geographic routing with next hop selection of nodes near the Steiner tree satisfying the multi criteria QoS factors is used for routing from source node to destination. Cross layer feedback from physical layer is collected at application layer to adopt the source data generation rate depending on traffic dynamics. Each of the stages of the proposed solution is detailed in below subsections.

A. Clustering

The FANET network is clustered by jointly considering mobility and density. For any two nodes i and , the mobility factor is calculated as:

$$M_{ij} = e^{1-\frac{V_r}{2V_{max}}}$$ (1)

Where V_r is the relative speed between nodes i and j, calculated over the period of time Δt as:

$$V_r = \frac{d_t - d_{t-\Delta t}}{\Delta t}$$ (2)

d_t and $d_{t-\Delta t}$ are the distance between the nodes observed in time between t and Δt. V_{max} is the maximum speed a node can move.

The distance between nodes is calculated in Euclidean 3D space as:

$$d_t = \sqrt{(x_i(t) - x_j(t))^2 + (y_i(t) - y(t))^2 + (z_i(t) - z(t))^2}$$ (3)

Higher the value of M_{ij}, the link between nodes i and j is highly stable.

M_{ij} is calculated between each pair of nodes in a communication radius of R and when the nodes having consistent M_{ij} are grouped as cluster. The R value is selected based on the density of nodes within the network area. It is calculated as:

$$R = T \times \frac{N_{CH}}{M}$$ (4)

Where T is total area, N_{CH} is the maximum number of clusters and M is the total number of UAV nodes. Once cluster are formed, the UAV nodes with consistent M_{ij} values to all other nodes are selected as stable nodes in the cluster.

Existing Link state-based routing protocols for FANETs have higher maintenance overhead but link state-based routing protocol provide reduced route discovery time compared to

ad-hoc search. This work adopts link state-based routing with a concept of routing backbone establishment with comparatively lower maintenance overhead. This work uses a graph theory concept called Steiner minimal tree to construct the routing backbone.

A Steiner Minimal Tree (SMT) is created with S points referred as Steiner points such that cost of N ∪ S is minimal. A weighted graph with vertex corresponding to each of N stable nodes and weight between the vertexes is set as the Received Signal Strength Index (RSSI) value between them is constructed and Steiner minimal is constructed for this graph.

A sample SMT in 2D space is shown in Fig. 3. In this, the hollow points are the vertexes, and the dotted points are the corresponding Steiner Points. The line connecting the Steiner points is the Steiner minimal tree. The tree is almost near to each of vertexes.
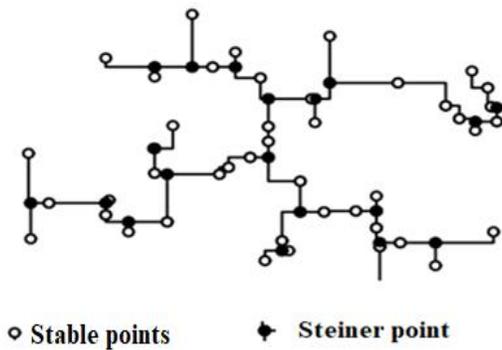
Fig. 3. Steiner Point in a Graph.

Since the stable points are in 3D space, this work adopts improved smith algorithm [25]. With stable nodes given as input, Steiner minimal tree is constructed optimally connecting the stable nodes.

B. Routing

Geographic routing with next hop selection based on multi criteria QoS factor over the Steiner minimal tree is proposed for routing the packets.

The QoS (Q) is calculated as...

$$Q = \alpha \times DF + (1 - \alpha) \times RF \tag{5}$$

Where DF is the Delay Factor. It is calculated in terms number of packets in queue (np_i) and Transmission failure Probability (P_i) and Packet transmission Time (pt_i) over the link i as

$$DF = \sum_{i=1}^{n}\{(np_i + 1).\left(\frac{1}{1-P_i}\right).pt_i\} \tag{6}$$

In terms of throughput, pt_i is calculated as

$$pt_i = \frac{L}{(1-R_i)B_i} \tag{7}$$

Where B_i is the Channel Bandwidth for link i under the influence of interference and L is the packet size. The net bandwidth usage is expressed as $(1 - R_i)$.

Reliability Factor (RF) is calculated in terms of exponential moving average of past delivery ratio (ψ) and Recent Delivery Ratio (DR).

$$RF = \alpha * \psi + (1 - \alpha) * DR \tag{8}$$

α is the weight factor with value in range of 0 to1.

Any node which wants to forward the packet to next hop selects the nodes close to Steiner tree with highest value for Q and uses that node as the relay node for packet forwarding.

C. Source Rate Control

Though routing paths with better QoS are selected, the QoS can be affected due to congestion. This work adopts source rate control based on cross layer feedback to reduce congestion on the routing path.

On routing path, congestion is measured in terms of round trip delay. It is calculated in terms of probability mass function of delay distribution between nodes A and B as

$$RTT = \begin{cases} \sum_{i=0}^{\infty} f_i(a).f_i(b) , x = 0 \\ \sum_{i=0}^{\infty} f_i(a).f_{2x+i}(b) + \sum_{i=0}^{\infty} f_i(b).f_{2x+i}(a), x > 0 \end{cases} \tag{9}$$

$f_i(a)$ is the probability mass function of delay in direction of A to B and $f_i(b)$ is the probability mass function of delay indirection of B to A.

The data sampling level is set in discrete interval from lowest to highest. The flow of source rate control is given in Fig. 2. Initially the UAV node operates in the default sampling level. RTT is measured periodically, and it is compared against expected latency. If the latency is higher than expected latency, the sampling level is decreased; this reduces the number of packets transmitted from source thereby attempting to reduce congestion. If the latency is lower than expected latency, then sampling level is increased, thereby increasing the number of packets in the network. Thus, source control rate is adaptive to traffic dynamics of the network.

IV. RESULTS

NS2 platform was used for simulation of the proposed solution. Simulation was conducted for configuration parameters given in Table I.

TABLE I. SIMULATION CONFIGURATION

Parameters	Values
Number of UAV	50 to 250
Communication range	100m
Area of simulation	1000m*1000m
Node distribution	Random distribution
Simulation time	30 minutes
Interface Queue Length	50
MAC	802.11
Number of Data collection point	1
UAV Speed	20-50 m/sec

Comparison of proposed solution's performance is done against Q-learning based QoS routing proposed by Xinchen et al. [24], link stability-based routing proposed by Zhang et al. [21], trajectory aided OLSR routing proposed by Hou et al. [7].

The simulation was conducted to measure the QoS parameters and compare the solutions in terms of their QoS performance.

Performance is measured and compared in terms of (i) packet delivery ratio (ii) delay (iii) throughput and (iv) network overhead. The QoS parameters were measured in two conditions of varying the number of UAV and varying the speed. By varying the number of UAV, QoS performance in relation to density of UAV in the network was compared. By varying the speed, QoS performance in relation to mobility of UAV was compared.

The packet delivery ratio is measured by varying the number of UAV and the result is given in Table II.

As in Fig. 4, in the proposed solution, the average packet delivery is 6.8% higher compared to Hou et al., 3.8% higher compared to Zhang et al. and 4.2% higher compared to Xinchen et al. The packet delivery ratio has increased in proposed solution due to three factors of selection of stable routing backbone, selection of multi criteria QoS based next hop and source rate control. These three factors have reduction the effective packet loss.

Varying the number of UE, the average delay is measured and given in Table III.

TABLE II.   COMPARISON OF PACKET DELIVERY RATIO

Number of UAV	Proposed	Hou et al [7]	Zhang et al [21]	Xinchen et al [24]
50	93	86	89	88
100	94	87	90	90
150	95	88	91	91
200	95	89	92	91
250	96	89	92	92
Average	94.6	87.8	90.8	90.4
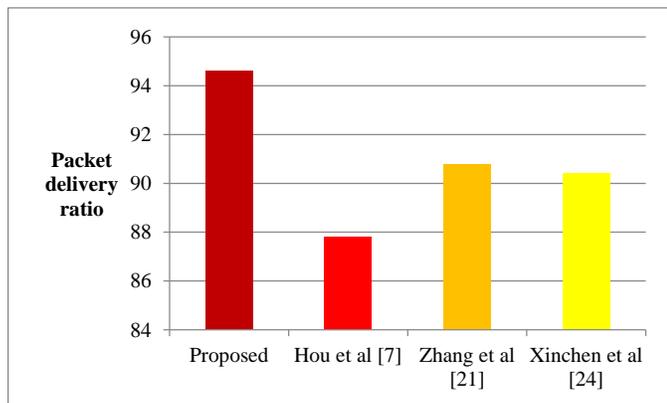
Fig. 4.   Comparison of Packet Delivery Ratio.

TABLE III.   COMPARISON OF DELAY

Number of UAV	Proposed	Hou et al [7]	Zhang et al [21]	Xinchen et al [24]
50	8	10	12	11
100	10	12	14	12
150	11	14	17	14
200	12.5	17	20	16
250	14	20	22	17
Average	11.1	14.6	17	14
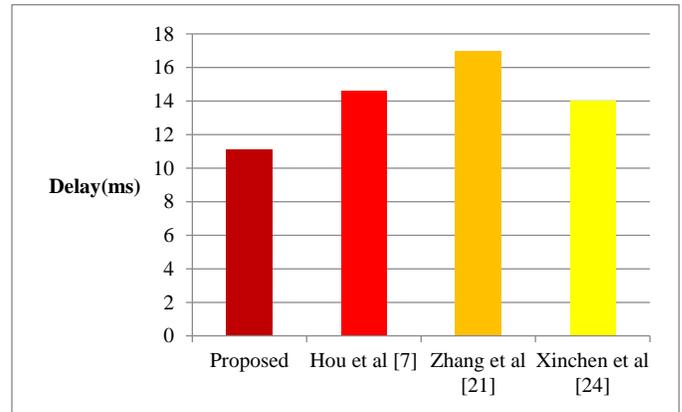
Fig. 5.   Comparison of Delay.

As shown in Fig. 5, in the proposed solution, the average delay is 31% lower compared to Hou et al., 53% lower compared to Zhang et al. and 26% lower compared to Xinchen et al. The delay in proposed solution has reduced due to selection of Steiner minimal tree constructed based on delay minimization in the proposed solution.

Varying the number of UAV, throughput is measured and given in Table IV.

As shown in Fig. 6, the throughput is almost same across the solutions. Throughput has not increased much in proposed solution due to source rate control. Source rate control effectively controls the congestion in network but due is aggressive control also reduces the throughput marginally.

Network overhead apart from data payloads packets is measured varying the number of UAV and the result is given in Table V.

TABLE IV.   COMPARISON OF THROUGHPUT

Number of UAV	Proposed	Hou et al [7]	Zhang et al [21]	Xinchen et al [24]
50	197	185	182	188
100	206	193	191	195
150	211	201	194	204
200	220	209	196	210
250	226	211	197	215
Average	212	199.8	192	202.4

Fig. 6.   Comparison of Throughput.

TABLE V.   COMPARISON OF NETWORK OVERHEAD

Number of UAV	Proposed	Hou et al [7]	Zhang et al [21]	Xinchen et al [24]
50	130	145	150	147
100	140	164	167	170
150	151	175	187	190
200	160	184	194	198
250	168	192	198	210
Average	149.8	172	179.2	183
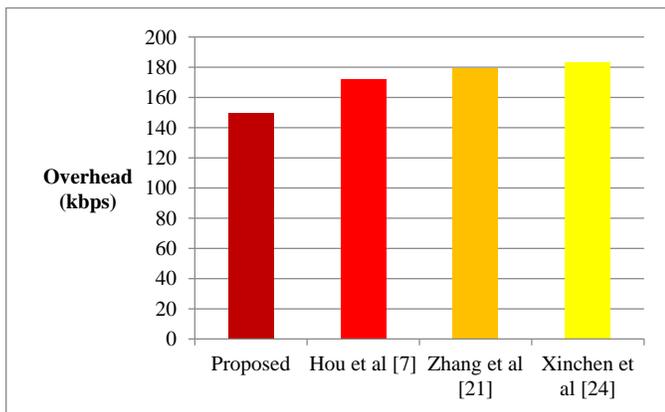
Fig. 7.   Comparison of Overhead.

As shown in Fig. 7, the network overhead in proposed solution is 14% lower compared to Hou et al., 19.6% lower compared to Zhang et al., 22% lower compared to Xinchen et al. The network overhead has reduced due to construction of routing backbone using Steiner minimal tree which involves only stable nodes.

The packet delivery ratio is measured by varying the speed of UAV and the result is given in Fig. 8.

Fig. 8.   Packet Delivery Ratio vs. Speed.

As shown in Fig. 8, packet delivery ratio drops with increase in speed of UAV. But even in drop, the packet delivery ratio is higher in proposed solution compared to existing works.

The delay is measured by varying the speed of UAV and the result is given in Fig. 9.
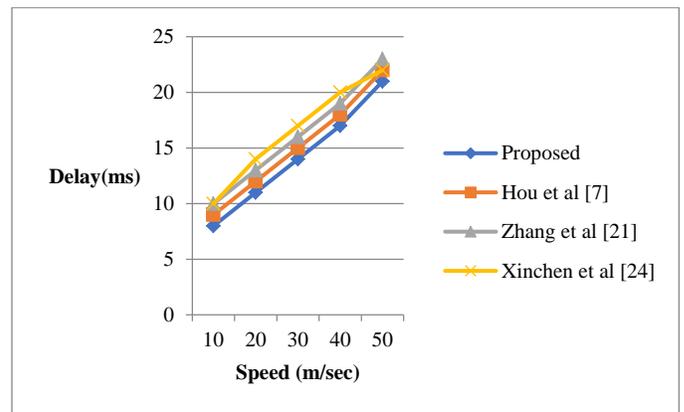
Fig. 9.   Comparison of Delay.

With increase in speed delay increases in all existing works. But in comparison to existing works, the delay increases at slower rate in proposed solution.

The delay is lower due to use of Steiner Minimal Tree-based routing in proposed solution.

Network overhead is measured by varying the speed and the result is given in Fig. 10.

The network overhead increases with speed of UAV but the overhead is lower in proposed solution compared to existing works. Overhead has reduced in proposed solution, due to use of less maintenance overhead.
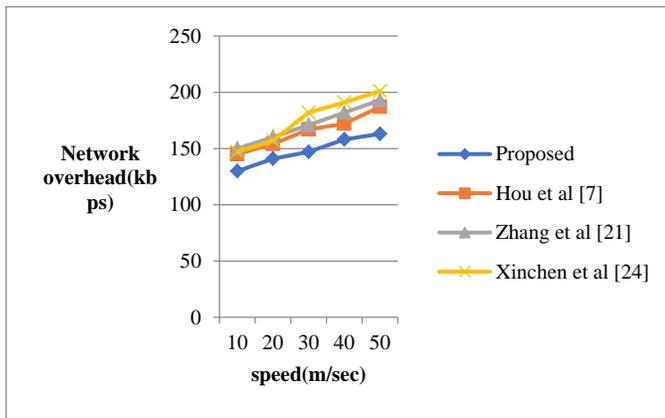
Fig. 10. Comparison of Network Overhead.

V. DISCUSSION AND RESULTS

The proposed solution performed better in all QoS metrics of packet delivery ratio, delay and throughput. Packet delivery has increased in proposed solution due to selection of more stable paths for routing and source rate control adaptable to network/application dynamics. Though Xinchen et al. used Q-learning to select the stable paths, it required high overhead to maintain the paths. Also the packet flow on the paths was not controlled based on application/network characteristics. Zhang et al. used OLSR based proactive routing protocol. Link stability was given more importance in route construction. But without packet rate control, congestion increases and affects the path reliability. Hou et al. solution was also based on OLSR. Similar to Zhang et al., the solution proposed by Hou et al. did not consider reliability after path establishment.

Delay has reduced in proposed solution due to use of geographic routing. Though Hou et al. and Zhang et al. used shorted OLSR path, due to node movement, the path has to be reconstructed. This created more packets in the network and as result, delay increased. The network overhead has increased in Hou et al. and Zhang et al. due to more route maintenance packets needed to keep OLSR link state fresh. Since the proposed solution used opportunistic routing, the overhead is lower in proposed solution.

The QoS performance was better in proposed solution compared to existing works even in higher speed. At higher speed, the link state maintenance overhead is higher in Hou et al. and Zhang et al. Due to this, number of packets increased in the network, creating congestion and reducing the reliability of the network. Proposed solution has source rate control mechanism to reduce the packet flow during congestion times, but this mechanism was lacking in the existing works.

VI. CONCLUSION

A Multi Criteria QoS Optimal Routing (MCQOR) for FANET is proposed in this work. The proposed solution has three stages of clustering, routing, and source rate control. The FANET network is clustered using combined mobility and density information. Over the clustered network, routing backbone is constructed using Steiner Minimal Tree (SMT) and Multi Criteria QoS based routing is done over the Steiner routing backbone. Due to stability consideration in clustering

and routing, the packet reliability has increased in the network. The proposed solution achieves 3.8% higher packet delivery ratio, 26% lower delay and 14% lower network overhead compared to existing works. Though QoS factors have been considered in this work, energy minimization and life time improvement has not been considered in this work. Life time maximization with joint consideration for QoS is in scope of future work.

REFERENCES

[1] Song, Q.; Jin, S.; Zheng, F.-C. Completion time and energy consumption minimization for UAV-enabled multicasting. IEEE Wirel. Commun. Lett. 2019, 8, 821–824.

[2] Munshi, A.A.; Sharma, S.; Kang, S.S. A Review on Routing Protocols for Flying Ad-Hoc Networks. In Proceedings of the 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 11–12 July 2018; pp. 1270–1274.

[3] Al-Absi, M.A.; Al-Absi, A.A.; Sain, M.; Lee, H. Moving ad hoc networks - A comparative study. Sustainability 2021, 13, 6187.

[4] Usman, Q.; Chughtai, O.; Nawaz, N.; Kaleem, Z.; Khaliq, K.A.; Nguyen, L.D. Lifetime Improvement Through Suitable Next Hop Nodes Using Forwarding Angle in FANET. In Proceedings of the 2020 4th International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom), Hanoi, Vietnam, 28–29 August 2020; pp. 50–55.

[5] Lin, Q.; Song, H.; Gui, X.; Wang, X.; Su, S. A shortest path routing algorithm for unmanned aerial systems based on grid position. J. Netw. Comput. Appl. 2018, 103, 215–224.

[6] Xie, P. An Enhanced OLSR Routing Protocol based on Node Link Expiration Time and Residual Energy in Ocean FANETS. In Proceedings of the 2018 24th Asia-Pacific Conference on Communications (APCC), Ningbo, China, 12–14 November 2018; pp. 598–603.

[7] Hou, C.; Xu, Z.; Jia, W.-K.; Cai, J.; Li, H. Improving aerial image transmission quality using trajectory-aided OLSR in flying ad hoc networks. EURASIP J. Wirel. Commun. Netw. 2020, 2020, 1–21.

[8] Yang, Q.; Jang, S.-J.; Yoo, S.-J. Q-Learning-Based Fuzzy Logic for Multi-objective Routing Algorithm in Flying Ad Hoc Networks. Wirel. Pers. Commun. 2020, 113, 115–138.

[9] Oubbati, O.S.; Mozaffari, M.; Chaib, N.; Lorenz, P.; Atiquzzaman, M.; Jamalipour, A. ECaD: Energy-efficient routing in flying ad hoc networks. Int. J. Commun. Syst. 2019, 32.

[10] Sharma, V.; Kumar, R.; Kumar, N. DPTR: Distributed priority tree-based routing protocol for FANETs. Comput. Commun. 2018, 122, 129–151.

[11] Usman, Q.; Chughtai, O.; Nawaz, N.; Kaleem, Z.; Khaliq, K.A.; Nguyen, L.D. Lifetime Improvement Through Suitable Next Hop Nodes Using Forwarding Angle in FANET. In Proceedings of the 2020 4th International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom), Hanoi, Vietnam, 28–29 August 2020; pp. 50–55.

[12] Zafar, W.; Khan, B.M. A reliable, delay bounded and less complex communication protocol for multiclusterFANETs. Digit. Commun. Netw. 2017, 3, 30–38.

[13] Mariyappan, K.; Christo, M.S.; Khilar, R. Implementation of FANET energy efficient AODV routing protocols for flying ad hoc networks [FEEAODV]. Mater. Today Proc. 2021.

[14] Khan, M.A.; Khan, I.U.; Safi, A.; Quershi, I.M. Dynamic Routing in Flying Ad-Hoc Networks Using Topology-Based Routing Protocols. Drones 2018, 2, 27.

[15] M. Zhang, C. Dong, S. Feng, X. Guan, H. Chen and Q. Wu, "Adaptive 3D routing protocol for flying ad hoc networks based on prediction-driven Q-learning," in China Communications, vol. 19, no. 5, pp. 302-317, May 2022.

[16] Usman, Q., Chughtai, O., Nawaz, N. et al. A Reliable Link-Adaptive Position-Based Routing Protocol for Flying ad hoc Network. Mobile Netw Appl 26, 1801–1820 (2021).

[17] Choi S-C, Hussen HR, Park J-H, Kim J (2018) Geolocation-based routing protocol for flying ad hoc networks (fanets). In: 2018 Tenth international conference on ubiquitous and future networks (ICUFN), pp 50–52. IEEE.

[18] Li X, Yan J (2017) Lepr: Link stability estimation-based preemptive routing protocol for flying ad hoc networks. In: 2017 IEEE symposium on computers and communications (ISCC), pp 1079–1084. IEEE.

[19] Hong J, Zhang D (2019) Tarcs: A topology change aware-based routing protocol choosing scheme of fanets. Electronics 8(3):274.

[20] Pu C (2018) Link-quality and traffic-load aware routing for UAV ad hoc networks. In: 2018 IEEE 4th International conference on collaboration and internet computing (CIC), pp 71–79. IEEE.

[21] Zhang Y, Qiu H. DDQN with Prioritized Experience Replay-Based Optimized Geographical Routing Protocol of Considering Link Stability and Energy Prediction for UANET. Sensors (Basel). 2022 Jul 3;22(13):5020.

[22] Souza, Jorge &Jailton, José & Carvalho, Tássio& Araújo, Jasmine & Frances, Carlos. (2020). QoS and QoE Aware Routing Protocol for Flying Ad-Hoc Surveillance Networks Using Fuzzy Inference Systems. Journal of Microwaves, Optoelectronics and Electromagnetic Applications. 19. 11-25. 10.1590/2179-10742020v19i11842.

[23] Y. He, X. Tang, R. Zhang, X. Du, D. Zhou and M. Guizani, "A Course-Aware Opportunistic Routing Protocol for FANETs," in IEEE Access, vol. 7, pp. 144303-144312, 2019.

[24] HUANG Xinchen, CHEN Guangzu, ZHENG Min, TAN Chong, LIU Hong. Q-learning based QoS routing for high dynamic flying Ad Hoc networks[J]. Journal of University of Chinese Academy of Sciences, 2022, 39(1): 134-143.

[25] Fonseca, R., Brazil, M., Winter, P., &Zachariasen, M. (2014). Faster Exact Algorithms for Computing Steiner Trees in Higher Dimensional Euclidean Spaces.

The Use of ICTs in the Digital Culture for Virtual Learning of University Students Applying an Artificial Neural Network Model

José Luis Morales Rocha[1], Mario Aurelio Coyla Zela[2], Nakaday Irazema Vargas Torres[3], Helen Gaite Trujillo[4]

Gestión Pública y Desarrollo Social, (Public Management and Social Development)[1, 2, 3]
Universidad Nacional de Moquegua, Moquegua, Peru[1, 2, 3]
Administración de Empresas, (Business Adminstration), Universidad Privada Domingo Savio, Tarija, Bolivia[4]

Abstract—**Artificial neural networks are mathematical models of artificial intelligence that intend to reproduce the behavior of the human brain and whose main objective is the construction of systems that are capable of demonstrating certain intelligent behavior. The purpose of the investigation is to determine the influence of the use of Information and Communication Technologies (ICTs) in the digital culture in the learning process of university students in Peru and Bolivia in the context of the Coronavirus – COVID 19 sanitary emergency, through the application of artificial neural network models. The investigation has a quantitative focus, the applied type, with a correlational level and a non-experimental design. Data was recollected by means of a digital questionnaire, applied to students of two universities. The population is composed of 3980 students of the Universidad Privada Domingo Savio (UPDS, Tarija, Bolivia) and 1506 of the Universidad Nacional de Moquegua (UNAM, Moquegua, Peru). The sample consists of 496 students. The hypothetical-deductive and the artificial intelligence methods were used. It was determined that the ability to install software and data protection programs, the use of mobile devices for academic purposes and the command of specialized software are the most influential factors in the digital culture of the students at UNAM and UPDS.**

Keywords—Artificial neural network; digital culture; ICT; virtual learning; COVID 19

I. INTRODUCTION

In [1], the complexity of the human brain is amazing, consisting of hundreds of billions of neurons with billions of connections that make the neural function of the brain very complex. In [2], the artificial neural networks are systems that simulate the properties that are observed in the biological diversity through mathematical models that are recreated by artificial mechanisms. Artificial neural networks are information processing systems that are inspired by the behavior of biological neural networks [3], [4], [5]. The artificial neural networks are mathematical models that are inspired by the biological neural systems, like the human brain, and can be considered as a data processing technique that maps, or relates, a certain type of entry information flow with an exit processing flow. There are variations of artificial neural networks to perform classification tasks, pattern recognition and prediction [6]. Artificial neural networks stand out as an automatic learning technique, due to their ease of use and comprehension compared to statistical methods, and their good performance in different automatic learning tasks [7][8].

ICTs include any communication device, such as computers, telephones, television, radio, etc., as well as services and software that work with devices, such as e-mail, software used for management, learning, videoconferencing, and so on [9]. The use of ICTs enhances student protagonism [10], keeping them up to date with technological development to maintain their sustainability in time [11].

Digital culture is the influence of technology on the behavior of people [12]. In [13], digital culture can be identified as a component of the actual digital transformation and as an epistemological obstacle in the sociological analysis of the same phenomenon. Digital culture are the abilities, knowledge, creativity and attitudes necessary to use in digital media in order to learn and comprehend in a knowledge society [14].

This article demonstrates the application of a model of an artificial neural network to explain the use of ICTs in the digital culture of students at two universities in Peru and Bolivia. The artificial neural networks that were used have an input layer of 11 units, a hidden layer of 5 units and an output layer of 1 unit, as well as training and test or validation data of the neural models. The artificial neural networks demonstrate that the dimensions that have the greatest influence in the digital culture are the ability to install software and data protection software, the command of mobile devices for academic purposes and the command of specialized software, according to the models of UNAM and UPDS.

The content of the article is organized as follows: Section II gives a short review of the state of the art in ICT, digital culture and artificial neural networks; Section III describes the methodology used in the investigation; Section IV describes and analyses the results that were obtained; Section V discusses the results; Section VI describes the final conclusions of the investigation; finally, the description of the future work that could be implemented to improve the obtained results.

II. RELATED WORK

In [14] aspects of the digital culture of student in the Faculty of Humanities and Pedagogy are described; general characteristics like cyber culture, the level of use of software, the use of technological devices and their digital abilities. [15], revises and synthesizes the expectations of employers in terms of digital abilities of graduates, the steps and measures taken by institutions of higher education to prepare students and to take advantage of their motivation to be more competitive and satisfy the needs of employability in the Fourth Industrial Revolution, supplying information about the digital alphabetization abilities that are required of young graduates, their expectations, and how digital alphabetization can be developed in institutions.

For [16], the ICTs have become a part of everyday life of people in business, entertainment, education and many other areas of human activity. Students have just begun to learn and accept new ideas, show a mature creativity, develop critical thinking and the ability to make decisions. In education, the successful integration of ICTs in learning and teaching depends on the attitudes of the teachers and their capacity to use the technology of communication not only competently but also with ability and imagination. According to [17], the access of ICTs in the world has increased, but there is still a gap between the necessity of ICT infrastructure and its availability.

According to [18], the objective of the investigation was to connect the ICTs with the distribution processes and the use of knowledge in the institutions of higher education. There was evidence that the profiles of the universities to implement knowledge management are focused on two components: the distribution and the use of knowledge. Among the more relevant results, it can be highlighted those two main profiles were found that characterize the process of knowledge management using ICTs. The strength of the first group lies in the component of investigation, which means it manages the promotion and diffusion of investigation in academic spaces. On the other hand, the second group's major strength lies in ICTs for the use of knowledge, which indicates that the universities in the second group focus their efforts on the management of ICTs as a means of knowledge management.

In the last decades, [19] artificial neural networks have gained significant attention in a part of the scientific community, as a technique that creates and combines multiple models of automatic learning to produce an optimal model that obtains the best results.

The use of neural networks is an area of growing interest, especially due to the availability of large quantities of information from multiple sources [20]. The algorithms of Artificial Intelligence (AI) have been used frequently to model complex functions for which the traditional mathematical methods no longer suffice. The artificial neural networks are one of the artificial intelligence methods that have the capacity to perform highly accurate simulations with potent learning algorithms and training capacities [21]. For [22], the neural network is an automatic learning algorithm. Neural networks help to simulate the functioning of the human brain [5].

III. METHODOLOGY

The investigation has a quantitative focus, the applied type, with a correlational level and a non-experimental design. The data were recollected through a digital questionnaire, implemented by means of the Google Forms tool, sent to students of two universities, one in Peru and the other in Bolivia. The population consisted of 3980 students of the Universidad Privada Domingo Savio in Bolivia and 1506 students of the Universidad Nacional de Moquegua in Peru. The sample consisted of 496 students (306 of the UPDS and 190 of the UNAM), for which the Stratified Random Sampling method was used. For the variables use of ICT and digital culture, questionnaires were used as the technique and a digital questionnaire as the instrument. The hypothetical-deductive method and artificial intelligence were applied.

The variables used in the investigation are the following:

1) Access to ICTs.
   a) Computer and communication resources.
   b) Command of Microsoft Office applications.
   c) Command of specialized software.
   d) Use of mobile devices for academic purposes.
   e) Use of computer accessories.
   f) Ability to install software and data protection programs.

2) Digital culture.
   a) Study habits in a virtual environment.
   b) Use of computer and communication resources in the learning process.
   c) Use of digital resources in the learning process.
   d) Command of video conferencing software.
   e) Frequency of use of computers and devices for academic and investigation purposes.
   f) Frequency of consulting virtual libraries and internet search engines.
   g) Frequency of use of social networks for academic purposes.

IV. RESULTS

A. Modeling, using an Neural Network

These are the processing results of the artificial neural networks for the Universidad Nacional de Moquegua and Universidad Privada Domingo Savio students.

The results of Table I demonstrate that the artificial neural networks used 11 units in the input layer, 5 in the hidden layer and 1 unit in the output layer. The standardized method was used as the rescaling method for the scale dependents to improve the training of the neural network. The hyperbolic tangent was used as the activation function of the hidden layers. Identity was used as the activation function of the unit of the output layer.

Fig. 1 shows the model of the artificial neural network for UNAM. The blue lines are the negative synaptic weights, and the gray lines are synaptic weights with positive values.

TABLE I.   NETWORK INFORMATION

Input Layer	Factors	1	Gender
		2	Marital status
	Variates	1	Computer and communication resources
		2	Command of Microsoft Office applications
		3	Command of specialized software
		4	Use of mobile devices for academic purposes
		5	Use of computer accessories
		6	Ability to install software and data protection programs
	Number of Units[a]		11
	Rescaling Method for Variates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		5
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Digital Culture
	Number of Units		1
	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit.
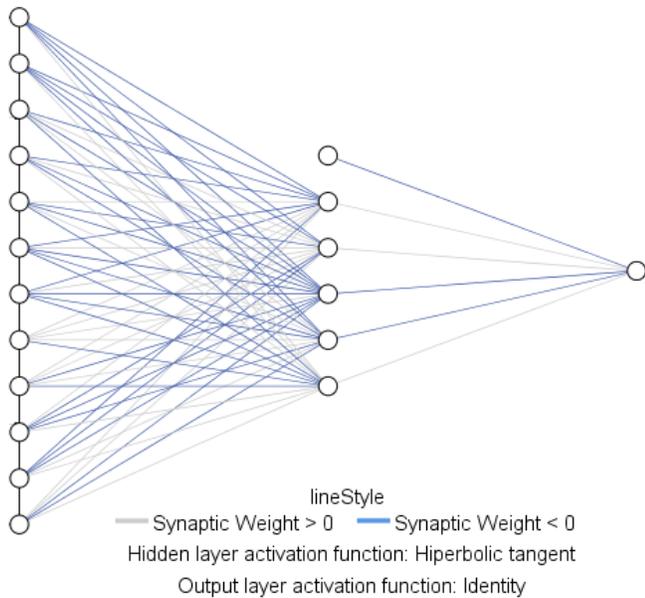
Fig. 1.   Artificial Neural Network of UNAM.

Fig. 2 shows the Artificial Neural Network model for UPDS. The blue lines are the negative synaptic weights, and the gray lines are synaptic weights with positive values.

In Table II, it is observed that the neural network model for UNAM considers 66.3% of the data for training and 33.7% for testing or validation of the model.
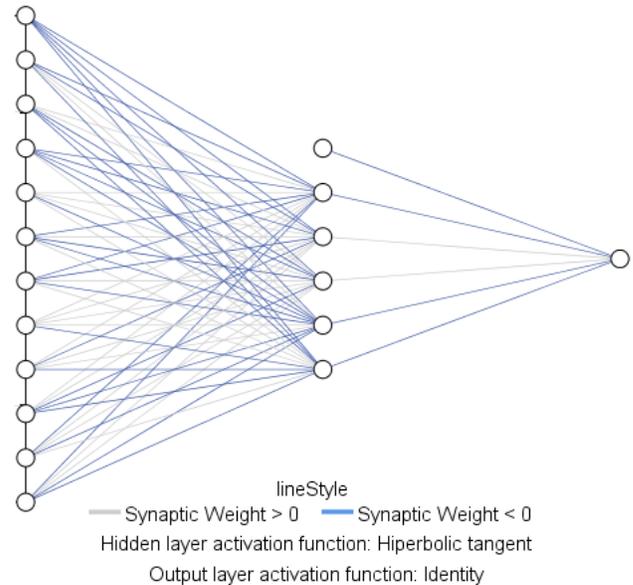
Fig. 2.   Artificial Neural Network of UPDS.

TABLE II.   CASE PROCESSING SUMMARY - UNAM

Case Processing Summary		N	Percent
Sample	Training	126	66.3%
	Testing	64	33.7%
Valid		190	100.0%
Excluded		0	
Total		190	

In Table III, the neural network model for UPDS considered 70.9% of the data for training and 29.1% for testing and validation of the model.

TABLE III.   CASE PROCESSING SUMMARY - UPDS

Case Processing Summary		N	Percent
Sample	Training	217	70.9%
	Testing	89	29.1%
Valid		306	100.0%
Excluded		0	
Total		306	

Table IV represents the results of the training and testing process of the artificial neural network of the Universidad Nacional de Moquegua in Peru.

TABLE IV.   MODEL SUMMARY - UNAM

Training	Sum of Squares Error	14.533
	Relative Error	0.233
Testing	Sum of Squares Error	7.720
	Relative Error	0.271

Dependent Variable: Digital Culture

a. Error computations are based on the testing sample.

Table V shows the results of the training and testing process of the artificial neural network of the Universidad Privada Domingo Savio in Bolivia.

TABLE V.   MODEL SUMMARY - UPDS

Training	Sum of Squares Error	18.476
	Relative Error	0.171
Testing	Sum of Squares Error	10.615
	Relative Error	0.234

Dependent Variable: Digital Culture

a. Error computations are based on the testing sample.

Table VI shows, the importance of the independent variates is observed. This measurement indicates the size of the variation of the value that is predicted by the network for different values of the independent variates. The variates are ordered by importance, from major to minor, with the variates of Ability to install software and data protection programs, use of mobile devices for academic purposes and command of specialized software as the most important ones for the network to predict the digital culture for the UNAM and UPDS models.

A comparative table (Table VII) is presented to know the realities about the use of information and communication technologies in higher level educational entities, referenced with results in other research works: Universidad Católica Andrés Bello - Venezuela, Universidad de Extremadura-España, Universidad Privada Domingo Savio - Bolivia, Universidad Nacional de Moquegua-Peru and Universidad Nacional de Cordova - Argentina [23].

TABLE VI.   IMPORTANCE OF INDEPENDENT VARIABLES IN DIGITAL CULTURE

Independent variates	UNAM		UPDS	
	Importance	Normalized Importance	Importance	Normalized Importance
E2: Gender	0.025	8.0%	0.023	6.9%
E6: Marital status	0.050	15.8%	0.041	12.6%
D1: Computer and communication resources	0.071	22.2%	0.086	26.5%
D2: Command of Microsoft Office applications	0.039	12.4%	0.055	16.8%
D3: Command of specialized software	0.169	53.3%	0.103	31.6%
D4: Use of mobile devices for academic purposes	0.259	81.7%	0.287	88.2%
D5: Use of computer accessories	0.069	21.7%	0.080	24.6%
D6: Ability to install software and data protection programs	0.317	100.0%	0.325	100.0%

TABLE VII.   COMPARATIVE ANALYSIS

Comparison of results in the use of ICTs	Universidad Católica Andrés Bello - Venezuela	Universidad de Extremadura-España	Universidad Privada Domingo Savio- Bolivia	Universidad Nacional de Moquegua-Peru	Universidad Nacional de Cordova-Argentina
Computer resources - Technological infrastructure - They have a computer		100.0%	26.5%	22.2%	100.0%
Mastery of Microsoft office applications - digital content	57.4%	50.5%	16.8%	12.4%	
Mastery of specialized software	81.2%		31.6%	53.3%	
Mastery of mobile devices for academic purposes	85.9%	77.2%	88.2%	81.7%	89.2%
Ability to install software and data protection programs - users of information technology	43.5%		100.0%	100.0%	
Internet connection at home		69.3%			89.2%
Use of virtual platform	56.2%	68.3%			98.0%
Virtual platform	Moodle	Moodle con H5P	Moodle	Plataforma SIGEUN	Classroom y Moodle
Videoconferences	Modulo 7	Zoom -Teams	Plataforma Office 365	Zoom Office 365	Zoom o Google Meet
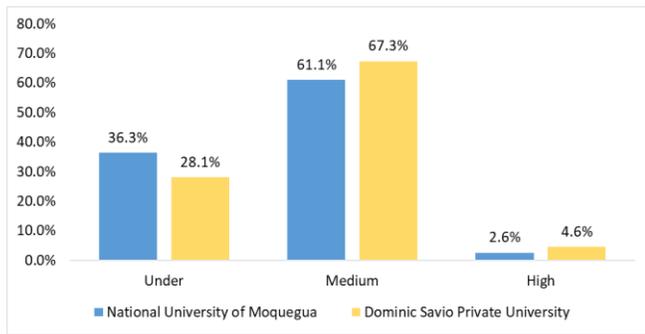
Fig. 3.   Access to ICT.

Fig. 3 shows that 36.3% of the Universidad Nacional de Moquegua students have a low level of access to ICTs, compared to 28.1% of the Universidad Privada Domingo Savio students. 67.3% of the Universidad Privada Domingo Savio students have medium access to ICTs, versus 61.1% of the Universidad Nacional de Moquegua students. 4.6% of the Universidad Privada Domingo Savio students have a high level of access to ICT, compared to 2.6% of the Universidad Nacional de Moquegua students.

Fig. 4 shows that 5.3% of the Universidad Nacional de Moquegua students have a low level of digital culture, vs 7.2% of the Universidad Privada Domingo Savio students. 80% of the Universidad Nacional de Moquegua students show a medium level of digital culture vs 68.6% of the Universidad Privada Domingo Savio students. 24.2% of the Universidad Privada Domingo Savio students have a high level of digital culture vs 14.7% of the Universidad Nacional de Moquegua students.
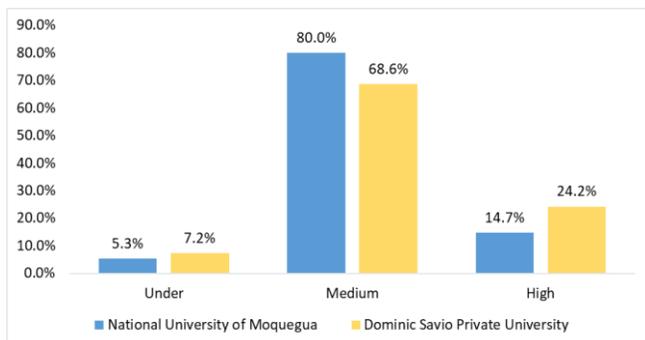
Fig. 4.   Digital Culture.

V.   DISCUSSION

The main characteristic of the study was to perform a comparative analysis of the use of ICTs and its implication on the digital culture of the students in two countries with similar social and economic realities in order to analyze the behavioral pattern of the Universidad Privada Domingo Savio students in Bolivia and Universidad Nacional de Moquegua in Peru.

The advantage of the methodology that was applied is that an artificial neural network can obtain reliable results due to its tolerance to errors and because it is permissive to the intervention of all neurons; even when a part of the system is damaged, the total functioning is not affected as it learned

from the experience of the representative sample or from the training data. In this investigation, 11 units were identified at the entry level, 5 in the hidden layer and 1 in the output layer. They required previous training through the weight adjustment of the connections in order to find an effective solution to the problems of the weight values and elements of the input variates and output variates with numeric values.

The quality of the training data represented in Tables I, II, III, IV and V warrant the model of the adaptive learning pattern, considering that the test data suffice to generate meaningful results in the time applied to adjust the weights of the entry layers of all neurons and to validate the application of the neural networks.

The variates of ability to install software and data protection programs, use of mobile devices for academic purposes and command of specialized software predict the digital culture in each of the universities involved in the investigation, which are assumed to be a product of the digital disruption due to the use of digital platforms, advanced generations of mobile networks, the Internet of Things, cloud computing, artificial intelligence and even the digital economy, that have an effect on the teaching learning processes, which requires the construction of new models of ecosystems of education and ICTs, reason why the university authorities have to anticipate this context.

The low levels of access to ICTs in both universities, 36.3% at the Universidad Nacional de Moquegua and 28.1% at the Universidad Privada Domingo Savio are determined by the social and economic conditions of the students, on top of limited connectivity. The minimal level of intermediate access to ICT of 4.6% at the Universidad Privada Domingo Savio and 2.6% at the Universidad Nacional de Moquegua are a warning signal that an intervention of the state authorities is required to focus on the necessities and inequalities in order to diminish the vulnerability of the students and to educate and promote national development through educational policies that tend to achieve equality of opportunities.

Even if the medium level of digital culture of students at 80% and 68.6% for the Universidad Nacional de Moquegua and the Universidad Privada Domingo Savio respectively is favorable, a bigger investigation is needed to determine the positive of negative incidence in academic performance, as it can be observed that in the majority of cases, students access low quality information, or that social networks are used as a source of distraction or even tend to stimulate social isolation.

A consistency is found with [14] which describes aspects of the digital culture, cyber culture, the level of use of software, the use of technological devices and their digital abilities. In [15] the expectations of employers towards digital abilities of graduates are revised and synthesized, an aspect that is not addressed in this investigation, but which has implications when it is necessary to determine the levels of employability and to project the new competencies of graduates for labor insertion, which demands a change in the curricular structure, the study plans and the need to project tendency analysis, scenarios and prospects of university education, parallel to the digital transformation of companies and the automation processes.

In [16] the successful integration of ICTs in the learning and teaching process is assumed, which leads us to believe that if new teaching tools are introduced, the teacher must also find new methods of teaching, creating and editing new contents in order to achieve active participation, even more when working with interactive young digital natives with access to an abundance of information.

The investigation is also in agreement with [17] in terms of the existence of a vast breach between the necessity of ICT infrastructure and its availability, which depends on factors in the economic, social and technological environment.

Finally, there is consistency with [18] in terms of knowledge management in universities that are focused on the distribution and use of knowledge through the application of ICT by means of promoting and diffusing investigation or the adequate management of ICTs as a means to manage knowledge. In both cases the environment determines the acquisition of knowledge, abilities and digital attitudes of the teacher and how the student assumes a critical attitude towards the use of ICT, through the selective use of information that is assessed on its quality. Only this way it is possible to achieve an efficient interrelation between the ICTs and the digital culture in the university system.

VI. CONCLUSION

The artificial neural networks made it possible to discover in an automatic way the relation between the variates Use of ICT and Digital Culture in function of the empirical data that were applied to the students of the Universidad Nacional de Moquegua - Peru and the Universidad Privada Domingo Savio - Bolivia, which information will allow for problem solving and better academic decision taking in each of the universities that were subject to the investigation.

The dimensions that have the greatest influence on digital culture as predicted by the artificial neural networks are the following: the ability to install software and data protection programs, the command of mobile devices for academic use and the command of specialized software in accordance with the models of the Universidad Nacional de Moquegua and the Universidad Privada Domingo Savio.

VII. FUTURE WORK

Future work will be related to the application of artificial intelligence algorithms to solve public management problems.

REFERENCES

[1] X. Yang, P. T. Schrader, and N. Zhang, "A deep neural network study of the ABIDE repository on autism spectrum classification," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 4, pp. 1–6, 2020, doi: 10.14569/IJACSA.2020.0110401.

[2] A. Medina-Santiago, E. M., L. C., A. Cisneros-Gómez, and N. R., "Neural Backpropagation System for the Study of Obesity in Childhood," Int. J. Adv. Res. Artif. Intell., vol. 5, no. 12, pp. 19–23, 2016, doi: 10.14569/ijarai.2016.051204.

[3] J. Morales and J. Huanca, "Regression model and neural network applied to the public spending execution," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 4, pp. 195–200, 2020, doi: 10.14569/IJACSA.2020.0110426.

[4] M. Iqbal, S. Ali, M. Abid, F. Majeed, and A. Ali, "Artificial Neural Network based Emotion Classification and Recognition from Speech," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 12, pp. 434–444, 2020, doi: 10.14569/ijacsa.2020.0111253.

[5] S. Aqab and M. U. Tariq, "Handwriting recognition using artificial intelligence neural network and image processing," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 7, pp. 137–146, 2020, doi: 10.14569/IJACSA.2020.0110719.

[6] S. N. Yahya, A. F. Ramli, M. N. Nordin, H. Basarudin, and M. A. Abu, "Comparison of Convolutional Neural Network Architectures for Face Mask Detection," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 12, pp. 667–677, 2021, doi: 10.14569/IJACSA.2021.0121283.

[7] H. Avila-George, T. Valdez-Morones, H. Pérez-Espinosa, B. Acevedo-Juárez, and W. Castrox, "Using artificial neural networks for detecting damage on tobacco leaves caused by blue mold," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 8, pp. 579–583, 2018, doi: 10.14569/ijacsa.2018.090873.

[8] Meiryani and D. L. Warganegara, "Implementation of Artificial Neural Network in Forecasting Sales Volume in Tokopedia Indonesia," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 5, pp. 416–421, 2021, doi: 10.14569/IJACSA.2021.0120551.

[9] C. Matsoga and M. Sundaram, "An Investigation on Information Communication Technology Awareness and Use in Improving Livestock Farming in Southern District, Botswana," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 2, pp. 17–23, 2017, doi: 10.14569/ijacsa.2017.080203.

[10] N. Bedregal-Alpaca, O. Sharhorodska, D. Tupacyupanqui-Jaen, and V. Corneko-Aparicio, "Problem based learning with information and communications technology support: An experience in the teaching-learning of matrix algebra," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 3, pp. 125–130, 2020, doi: 10.14569/ijacsa.2020.0110315.

[11] A. N. A. Rozmi, P. N. E. Nohuddin, A. R. A. Hadi, M. I. A. Bakar, and A. I. Nordin, "Factors affecting SME owners in adopting ICT in business using thematic analysis," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 7, pp. 208–218, 2020, doi: 10.14569/IJACSA.2020.0110727.

[12] M. Sayibu, J. Chu, T. Y. Akintunde, O. H. Rufai, T. S. Amosun, and G. George-Ufot, "Environmental conditions, mobile digital culture, mobile usability, knowledge of app in COVID-19 risk mitigation: A structural equation model analysis," Smart Heal., vol. 25, no. May, p. 100286, 2022, doi: 10.1016/j.smhl.2022.100286.

[13] J. S. Guy, "Digital technology, digital culture and the metric/nonmetric distinction," Technol. Forecast. Soc. Change, vol. 145, no. May, pp. 55–61, 2019, doi: 10.1016/j.techfore.2019.05.005.

[14] T. Ayale-Pérez and J. Joo-Nagata, "The digital culture of students of pedagogy specialising in the humanities in Santiago de Chile," Comput. Educ., vol. 133, pp. 1–12, 2019, doi: 10.1016/j.compedu.2019.01.002.

[15] K. Mn, F. Khalid, and H. Husnin, "Preparing graduates with digital literacy skills toward fulfilling employability need in 4IR Era: A review," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 6, pp. 307–316, 2020, doi: 10.14569/IJACSA.2020.0110641.

[16] S. Alshmrany and B. Wilkinson, "Factors Influencing the Adoption of ICT by Teachers in Primary Schools in Saudi Arabia," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 12, pp. 143–156, 2017, doi: 10.14569/ijacsa.2017.081218.

[17] C. Zhang, I. Khan, V. Dagar, A. Saeed, and M. W. Zafar, "Environmental impact of information and communication technology: Unveiling the role of education in developing countries," Technol. Forecast. Soc. Change, vol. 178, no. February, p. 121570, 2022, doi: 10.1016/j.techfore.2022.121570.

[18] J. H. Escorcia Guzman, R. A. Zuluaga-Ortiz, D. A. Barrios-Miranda, and E. J. Delahoz-Dominguez, "Information and Communication Technologies (ICT) in the processes of distribution and use of knowledge in Higher Education Institutions (HEIs)," Procedia Comput. Sci., vol. 198, no. 2021, pp. 644–649, 2021, doi: 10.1016/j.procs.2021.12.300.

[19] K. Dhibi, M. Mansouri, K. Bouzrara, H. Nounou, and M. Nounou, "Reduced neural network based ensemble approach for fault detection and diagnosis of wind energy converter systems," Renew. Energy, vol. 194, pp. 778–787, 2022, doi: 10.1016/j.renene.2022.05.082.

[20] M. Pereira, A. Lang, and B. Kulcsár, "Short-term traffic prediction using physics-aware neural networks," vol. 142, no. June, 2021, [Online]. Available: http://arxiv.org/abs/2109.10253.

[21] A. Shafiq, A. Batur Çolak, T. Naz Sindhu, S. Ahmad Lone, A. Alsubie, and F. Jarad, "Comparative study of artificial neural network versus parametric method in COVID-19 data analysis," Results Phys., vol. 38, no. May, p. 105613, 2022, doi: 10.1016/j.rinp.2022.105613.

[22] Y. Shi and Y. Zhang, "The neural network methods for solving Traveling Salesman Problem," Procedia Comput. Sci., vol. 199, pp. 681–686, 2021, doi: 10.1016/j.procs.2022.01.084.

[23] C. Ardini, M. B. Barroso, L. Contreras, and L. Corzo, "Study During a Pandemic," p. 41, 2020, [Online]. Available: https://bit.ly/3s3XDjZ.

A Study of Modelling IoT Security Systems with Unified Modelling Language (UML)

Hind Meziane*, Noura Ouerdi

LACSA Laboratory, Faculty of Sciences (FSO)
Mohammed First University (UMP), Oujda, Morocco

Abstract—**The Internet of Things (IoT) has emerged as a technology with the application in different areas. Hence, security is one of the major challenges that has the potential to stifle the growth of the IoT. In fact, IoT is vulnerable to several cyber attacks and needs challenging techniques to achieve its security. In this paper, the use of a UML (Unified Modelling Language) aims at modeling IoT systems in various views. The purpose of this study is to discuss the need for more modeling in terms of security. For this reason, this paper focuses on modeling security of IoT systems. The objective is to make a comparison in terms of layers by describing the IoT architecture and presenting its components. In other words, the research question is to look for the modeling of security in the IoT layers. There is no standard that takes into account the security of the IoT architecture, there are different proposals of IoT levels, which means that each author has his own vision and own proposition. Moreover, there is a lack of modelling languages for IoT security systems. The main interest of this study is to choose the layer on which we should be interested. The question then is as follows: "which is the layer whose modeling is relevant?" The obtained results were conclusive and provided the best insight into all the specifications of each layer of the IoT architecture studied.**

Keywords—Internet of things (IoT); IoT systems; IoT security; modelling; Unified Modelling Language (UML); UML extensions; IoT applications

I. INTRODUCTION

IoT or Internet of Things is a big area. The general idea of it is to make objects connected or linked together in order to communicate and exchange information via communication technologies like LoRa, LoRaWan, 4G, 3G, 2G, etc. IoT was coined by Kevin Ashton (pioneer of Radio Frequency Identification (RFID) technology (Automatic Identification Technology)) in 1999 [1]. IoT is a vast domain [1][2]. The IoT aims at enabling things to be connected at anyplace, anytime with anyone and anything using any network/path and any service. IoT is now used in multiple fields like Healthcare, Transportation, Factory, Building, City, Retail, Surveillance, Manufacturing, Agriculture, Logistics, Lifestyle, Industrial, etc.

Software engineering goals include gaining the greatest knowledge of the problem and modeling complicated systems utilizing tools such as UML (Unified Modelling Language) diagrams. The UML diagrams can be used for modelling the complex systems. In this sense, realistic IoT application scenario modeling on a wide scale with various operating conditions necessitates/requires a mix of edge devices, sensors, and actuators [3]. The goal of this paper is to review

published studies and look into the current state of IoT modelling security. The modelling of security systems for IoT is the main step, thus, security modelization is essential. In the literature, there are two types of tools: modeling languages (UML, SysML (System Modelling Language) and ThingML (Internet-of-Things Modeling Language)) and extensions to these languages (UML4IoT, IoTsec, UMLsec, SysMLsec, SysML4IoT). Various languages with their extensions have been provided to simulate and model IoT systems like IoTsec, ThingML, SysML, UML4IoT, SysML4IoT, etc.

In this context, networking and designing IoT and edge computing layers [3] is a very laborious process because of: (1) end-point networks like Bluetooth, Wi-Fi, and 4G are complex and heterogenous; (2) heterogeneity of edge and software stack and IoT hardware resources; (3) mobility of IoT devices; and (4) the complex interaction between IoT and edge layers [3]. Creating an edge computing and IoT testbed with high verisimilitude is not only difficult or complex, resource-intensive and expensive but also time-consuming. Furthermore, because of the high cost and broad domain expertise necessary to reason about their variety, usability, and scalability in actual IoT and edge computing settings, testing in reality edge computing and IoT environments is not practical [3]. Hence, in this article, we are interested in modelling IoT security systems and we also discuss the modelling challenges in the IoT environment context.

In UML 2.5, there are now fourteen diagrams that developers utilize to build systems. However, for some specific application domains, these diagrams are ambiguous. As a result, specific domain notations are covered by the UML extensions [4]. In UML, there are some extension artefacts that are well defined, these are: tags, values, stereotypes and constraints. These extensions mechanisms allow designers to develop specific models for certain areas/domains, such as modeling security concerns in systems, UML extension for Hypermedia Design [4], profiles to model Internet of Things (IoT) systems, …. [4]. However, security is one of the major issues. Three issues in designing IoT applications are addressed due to the lack of a strategy to develop/model IoT applications, the lack of a model to design security challenges in IoT, and the heterogeneity of different software and hardware devices.

The objective and originality of this paper are as follow: A lot of research work on modelling languages have been provided, but we could not find works on the following question "what is the relevant layer to model that really contributes to the security of IoT systems?" that is the strong

---

*Corresponding Author.

point of this research. In the literature, there is a lack and absence of works in the same section/point and this can be considered among the criticisms. The main idea of this work is to choose an IoT layer for which modelling is relevant. The problem to be discussed is as follows: which layer should be specified? The main objective is to look for the modeling of security in these layers, i.e., on which layer we should be interested? To do this, we compare the four layers by considering various parameters. However, there are no answers about the following question: What is/are the relevant layer(s) to model that really contributes to the security of IoT systems?

A. Contribution

The contributions of this survey are summarized as follows:

- This paper presents a state of the art of modelling IoT systems with UML, SysML and ThingML to perform a comprehensive study on modelling IoT security systems.

- This survey defines and describes the challenges that IoT systems are currently facing.

- The architecture of IoT systems is proposed as well as the role of each IoT layer is also described.

- This paper proposes a summary and taxonomy of security attacks and vulnerabilities at different layers. Then, the security issues, problems, vulnerabilities and challenges of each layer were also introduced. It also explores the security requirements for IoT layers.

- The strong point is to define the layer that needs more security in the IoT architecture. The added value is as follows: there is no survey that has been made to confirm "which is the layer whose modelling is relevant?"

B. Outline

This paper is organized as follows: The related works done in IoT security modelling are presented in Section II. Section III provides the IoT challenges. Section IV outlines the proposed methodology including architecture of IoT systems, challenges of each IoT layer, classification of attacks and vulnerabilities in IoT, IoT security requirements, OWASP IoT Project by analyzing security threats and vulnerabilities, followed by an example of modeling IoT security with UML. Results and discussion are covered in section V. Finally, Section VI concludes the paper.

II. RELATED WORKS

Currently, modeling the security of IoT systems has become highly important among researchers, each researcher uses a language (whatever its extension, e.g. IoTsec, UML4IoT, SysML-Sec...) to prove the quality of its work. In this section, we review and discuss the literature of some previous papers which are related to our theme in order to keep up with the languages and extensions used. It is extremely difficult to model a real IoT scenario due to several challenges that we will discuss later in the results and discussion section.

There are several simulation environments currently available such as IoTSim-Edge, IoTsuite, SimIoT... For instance, Jha et al., (2020) [3], aims to build a novel simulator, IoTSim-Edge, that allows users to evaluate the edge computing scenario in an easily configurable and customizable environment. Further, the authors give the general architecture of edge computing considered for modeling by IoTSim-Edge simulator. The architecture of the proposed simulator consists of multiple layers. IoTSim-Edge targets to model smart devices using low energy protocols. The authors also consider the simulation of battery power by using a predefined drainage rate. The presented architecture of IoT-Edge computing consists of two components: actuator nodes and sensing nodes. The sensing nodes collect the information of surroundings via sensors and send it for processing and storage. Whereas, the actuators will be activated according to the analysis of the information/data. The communication layer is responsible for transferring data to IoT devices, cloud, and edge devices. Different communication protocols are used for transferring data. Edge infrastructure is the next layer that consists of several types of edge devices (e.g., Raspberry Pi and Arduino). These devices can be transparently accessed through the help of various types of containerization and virtualization mechanisms. It provides an infrastructure to deploy the raw data produced by the sensing nodes. The services or application layer consists of various services which can be accessible directly to the users. These services (applications) will be accessed through a subscription model.

A. Modelling IoT Security System using UML Language

Robles-Ramirez et al., (2017) [4], mention a number of approaches that aim to model IoT systems and security considerations. They used IoTsec, which is a UML extension and another example that includes a notation for security modelling in IoT systems. Furthermore, to model IoT systems, they propose a UML/SysML extension, which attempts to encapsulate security knowledge. In particular, a new UML extension is proposed, which characterises security issues encapsulated within a nomenclature, and UML stereotypes, in order to model common actors and UML notation extensions. The objective is to move IoT development at a previous step from the implementation stage; the designing phase. This work targets security concerns in IoT development. Besides, the authors provide a simple model language to describe security actions and entities. The main aim is to facilitate the representation of security issues using a visual notation [4], even if the developers are unfamiliar to Internet security concepts. To represent an IoT system in terms of security issues/concerns, the authors, adopted the four layers architecture including sensing, network, service, and application layer.

Dhouib et al. (2016) [5], give a highlight to the current status of "Papyrus for IoT". It is a modeling environment that enable to deploy, specify and design complex IoT systems by using an IoT-A lightweight methodology. Furthermore, in order to illustrate the modelling environment, they use the example of a smart IoT-based home automation system, which consists of five steps:

- Step 1: Design the requirements and the purpose of the system by using SysML requirements diagrams and UML use case diagrams in Papyrus SysML Component.

- Step 2: Define the process specification. The IoT system's use cases are described, derived from and based on the requirements and purpose specification.

- Step 3: Define the system's functional architecture based on the IoT domain model.

- Step 4: Define the operational platform for the functional system's execution.

- Step 5: With Papyrus Designer Component, define deployment plans that include information about the allocation/assignment of functional blocks (step 3) to operational ones (step 4).

Moreover, they introduce MDE4IoT frameworks covering more than only a single MDE technique. Moreover, papyrus is an open-source Modeling Environment. In contrast to the other publications, the introduced approach already uses IoT-A in the Papyrus for IoT modelling environment.

Reggio (2018) [6] presents a method (IoTReq) for the elicitation and specification of requirements for IoT systems. The method uses UML for modelling the domain, requirements elicitation and specification of IoT requirements. This method also supports the specification of non-functional specifications. Reggio proposes a UML profile for the requirement gathering for IoT applications.

Ouchani (2018) [7] creates a formal framework for assessing the functional correctness of IoT systems. The suggested framework includes all of the major components of IoT systems, and the process is completely automated. Author describes the IoT architecture by showing its components with their interactions. The suggested IoT architecture enclosed five components which are demonstrated and analyzed subsequently in the next section. (1) Object devices, (2) User devices, (3) Computing services, (4) Social actors (are human agents) and (5) The environment. These components interact via communication protocols of various ranges (ZigBee, WiFi, Cellular, Human-machine, Bluetooth, SSH, IpSec, etc.). However, the proposed work suffers from PRISM's restrictions, and the security proprieties are not described.

(Thramboulidis and Christoulakis, 2016) [8], presented UML for IoT (UML4IoT) domain-specific modelling language to tackle the IAT (Industrial Automation Thing) domain. They presented the UML4IoT approach, a UML profile aimed at modeling cyber-physical components as their integration into IoT systems in manufacturing domain. UML4IoT is an UML profile for IoT, which have been already employed in the domain of real-time and embedded systems. However, UML4IoT extension does not support security modeling. Moreover, the authors didn't provide any reason about their choice.

B. Modelling IoT Security System using SysML Language

Ferraris et al. (2020) [9], present a model-driven approach extending UML and SysML diagrams. The aim of this work is to provide developers with a tool helping them to consider domains such as trust and security during the SDLC (System Development Life Cycle) of an IoT entity.

C. Modelling IoT Security System using ThingML Language

Harrand (2016) [10] introduced ThingML (Internet-of-Things Modeling Language), which is a modelling language that focuses on a distributed and heterogeneous systems, for generating code framework for diverse targets. ThingML is designed to support the development code generation. It can be considered as a DSML (domain specific modeling language). Generally, it has more been applied to IoT and CPS (Cyber-Physical Systems). Nonetheless, the authors do not take security into account.

However, for the existing papers in this field of modeling IoT security systems with UML, SysML, or ThingML, no survey has been made to confirm which is the layer whose modeling is relevant. In other words, there is no answer about the following questions: "Which layer whose modeling is relevant?" This means what is the relevant layer to model that really contributes to the security of IoT systems? Which calls the need for an in-depth comparison on the IoT layers including the upcoming and existing security issues. In general, as another limitation of similar works, modeling IoT security systems is still very superficial.

III. IoT CHALLENGES

The IoT infrastructures and security are still in infant phases. IoT systems have its own specific challenges. Hence, there are several obstacles to the development of the IoT. The following are some challenges and constraints in IoT systems:

A. Interconnectivity

Interconnectivity characterizes the ability of IoT systems and their constituents to communicate and use each other's services in a seamless manner [1], [2], [11]. The global information and communication infrastructure [12] allows for the interconnectivity of anything.

B. Heterogeneity

IoT devices are heterogeneous, since they are based on several networks and hardware platforms [12]. They can communicate with other service platforms or devices via various communication technologies and networks like LoRa, 2G, 4G, etc. Indeed, these IoT protocols are heterogeneous. Therefore, we need to make sure that there is a compatibility between them.

C. Interoperability

Interoperability is a basic value of the traditional Internet, the first criterion of internet connectivity is that "connected" systems/computers "speak the same language" of protocols and encodings. To support their applications, various industries nowadays use various standards. The adoption of common interfaces between these different entities becomes increasingly critical when there are various sources of data and heterogeneous devices. As a result, IoT systems must be able to deal with a high level of interoperability [1], [2],[12].

D. Scalability

In an IoT network, there are a large number of IoT devices [1] and nodes, Cisco estimated that in 2020, 26.3 billion devices were connected to the Internet [13]. Because of this large number (including the number of users and the number of participating IoT things), scalability is a crucial issue for creating effective defensive methods [13].

E. Big Data

Not only the number of smart objects [13] will be enormous, but the data generated by each object will be huge. Because each smart device is supposed to be supplied by too many sensors, each of which generates massive amounts of data over time, and also exchanges it on multiple IoT layers [1], [2], [13].

F. No Standardization

There is a lack of standardization in terms of IoT definition, vision, architecture, attacks, [1], [2]. For the IoT layers architectures, there are several proposals of IoT architecture with three, four, five, and seven layers (for the Cisco IoT model). Further, there is no robust solution that will solve most of IoT security issues.

G. Resource Constraint

Resource constraint (Limited Resources) in terms of energy, storage space and computing capacity. End devices in the IoT have limited resources like memory, storage, CPU, battery, and transmission range [13]. In other words, IoT devices are uniquely identifiable and are mostly characterized by limited processing, small memory, and low power. Furthermore, IoT devices [4] have many constraints like software based, network-based, and hardware based limitations, which depict new challenges for IoT developers. Moreover, implementing conventional security measures is impossible because it would be a very hard and complex process [14]. Since IoT devices have constrained resources, standard encryption algorithms cannot be applied directly for the IoT system. Lightweight cryptographic techniques were suggested by [14]. Therefore, a comparative study of existed lightweight cryptographic algorithms must be required. In addition, due to the rapid growth of IoT devices and the great developpement of new technologies and elements in the market, IoT systems have become vulnerable to several attacks.

H. The Lack of Encryption

The lack of encryption in the cloud which should also be considered [1]. An encryption procedure is important for IoT systems.

I. Security and Privacy

That means provide the necessary protection for data and maintain the privacy of users; this is the biggest challenge of the IoT. This is especially critical in healthcare devices. For applications in personalized medicine, such intelligent devices are becoming more popular. The information gathered is typically highly substantial and frequently includes meta-data like time, place, and context [15]. Besides, it describes the five security features, like confidentiality, availability, integrity, authentication and non-repudiation. Another serious problem is about how to get a secure and an efficient IoT platform to deliver what is required to be delivered.

J. Security Policies

We need to have policy. Policies are operational rules that must be maintained in order to keep data organized, secure, and consistent. Because security is also about how to use this flow of data. Security policies should be followed by users.

IV. METHODOLOGY

The objective of this paper is to conduct a comparative study of different layers of the IoT architecture taking into account the specifications of each layer. The layer is a factor among the factors that differentiates the modeling of IoT systems. Thus, to ensure the security of IoT systems, this contribution will compare these layers based on IoT architecture in order to outline the relevant layer to model.

The problem to be discussed is as follows: at which layer we are going to work? Do we combine two layers or we work on one layer? So, the main aim of this section is to choose which layer requires more security. To answer this research question, a systematic review on IoT architecture and security concerns could be done regarding several important axes/points. In other words, we need to do a comparison in terms of layers. By looking for layer modeling, we report the results and findings of this section. Thus, the proposed methodology is based on six steps. We gathered all specifications for each layer of IoT. For this purpose, the author is looking for a detailed description on IoT layers as well as an analysis of each layer was also done by covering the next points:

1) At first we will present the functionalities of IoT architecture to understand how IoT system basically works and underline the architecture we are working on.

2) After that, we will detail in the second step the security issues/problems of the IoT architecture which contains multiple layers, to give a detailed explanation and general understanding of different security challenges in different layers.

3) Then, we will provide a classification of security attacks and vulnerabilities in each layer of IoT systems, to conclude the most security attacks in each layer.

4) Followed by security requirements in IoT to analyze an in-depth the security services that must come with each layer.

5) In the fifth step, we will talk about the OWASP IoT Project as an example of security threats analysis.

6) And finally, we will end with, a comparison that were obtained previously to outline the extension that meet our needs regarding/in terms of security. That is why we took a UML extension as an example for modeling IoT security.

To be clearer, we followed these steps in the methodology section in order to compare the different layers by keeping the results of the six steps.

There are many papers, and researchers have presented different architectures. Some researchers present architecture composed of five layers [1], [2], some architecture composed

of three layers and in our research, we are interesting and we choose the architecture of four layers: physical layer, then, network, middleware and application layer. In this paper, the author adopts four layers architecture to present or choose the relevant layer to model that really contributes to the security of IoT systems and also to reflect the IoT architecture concerns. The proposed model of the IoT system is innovative, that take into consideration all characteristics, concerns, issues and threats of the four layers of the IoT architecture. Therefore, the best IoT architecture is that of four layers. Because, it is very general and presents in a great way the concept of the IoT. Besides, security at the middleware layer (storage/cloud/data) is not the same as security at the application layer (authentication/identification); as more than we separate the problems, we find solutions. In other words, security concepts in the cloud/middleware layer may not be integrated in the application layer. Moreover, we need cloud because we have so much data generated by many connected objects. Therefore, the middleware layer is also provided as a necessary and an integral part of the IoT model. Fig. 1 shows proposed and detailed architecture of IoT systems. It shows various devices and technologies at these layers. In this paper, author is using and describing the functionality of the following layers:

- Physical Layer
- Network Layer
- Middleware Layer
- Application Layer
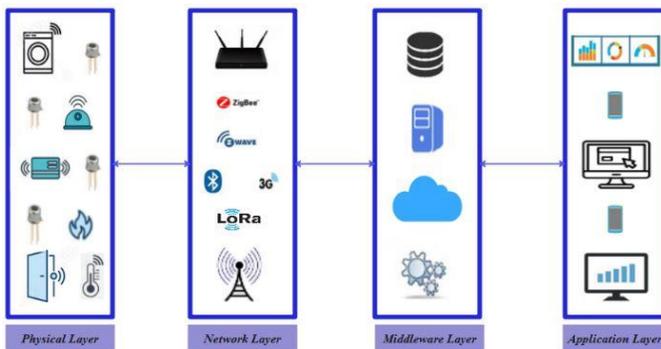
Fig. 1. IoT Architecture based on Four Layers.

A. IoT Architecture

There is not a standard for the IoT architecture, there are different proposals of IoT levels. It consists of different layers. To achieve this goal, and for simplicity, this paper focuses on four IoT layers.

1) Physical/Perception layer: Physical layer, which is also called the Perception layer. In [1] and [2], the first or base layer has been generalized by physical layer, that is not limited to actuators and sensors, it is composed of sensors, controllers, RFID tags and reader, actuators, and devices. IoT devices include mobile devices, single-board computers, and micro controller units [2]. IoT involves objects or things such as sensors, actuators, RFID tags and readers, to permit

interaction between the physical and virtual worlds. The capacity of a connected thing to optimize the probability of satisfying/achieving the user's goals may be used to determine its intelligence. Connected things intelligence spans from non-existent to perfectly rational [11]. The sensors may monitor or measure values like temperature, pressure, humidity, air quality, movement, speed, flow, and electricity…. Sensors collect information about their surroundings. This layer's main purpose is to collect useful data from the environment (like humidity, WSN, temperature, and heterogeneous devices, etc), then process, and digitize these information/data [1]. Body sensors, vehicle telematics sensors, environmental sensors, home appliance sensors, are classified according to their specific purpose. The actuator can be used for performing particular action.

The functionalities of this layer include actuating, sensing, actuating and sensing, storage, and processing [11]. Gathering and generating information by devices are the main operations of the physical layer. The collected data with the networks will be sent to the cloud. There are three types of things [11] smart things (or physical objects), sensors or/and actuators, and gateways. Finally, the first layer is "physical layer" also known as recognition, perception. The main function is to identify, generate and collect data from the physical world to perceive their environment by detecting changes using sensors.

The first layer components are: Tag, sensors/actuators, coordinator and network (LAN, PAN). In other words, LAN or PAN can be considered as a form of connectivity between sensors and sensor gateways. While WAN can be considered as a form of connectivity between sensors and servers.

Enabling technologies for the IoT can be grouped into three categories [12]:

- Technologies that enable "things" to acquire contextual information.
- Technologies that enable "thing" to process contextual information.
- Technologies to improve Security and privacy.

2) Network layer: The network layer is the connectivity layer which connects the perception layer to the cloud trough, a gateway using different technologies such as Lora, satellites, 3G, and others technologies. It is responsible for the transfer of this sensitive and massive volumes of data between different layers via network. It contains Router, Internet, Switches, Gateway [1], This layer is used for the connectivity, it defines the different protocols and communication networks or technologies such as Sigfox, 3G, 4G, 5G, WIFI, Zigbee, Bluetooth, LoRaWAN, LoRa, which represent the edge where the collected data can be processed. This layer is responsible for the connectivity of the IoT infrastructure. It also collects data from the physical layer and transmits it to the upper layer. The transmission medium can be wired or wireless.

These technologies include, but not limited to ZigBee, RFID, NFC, WSN, MANET, Wi-Fi. Furthermore, these technologies have their own security issues. Multiple networks with various technologies and access protocols are needed to work with each other in a heterogeneous configuration.

IoT communication technologies are based on two different categories [1]: Short-range and Long-range technologies. The first one include Z-Wave, Zigbee, RFID, NFC, 6LoWPan, etc. The second includes LPWAN technologies, among them LoRa/LoRaWAN, Cellular communication (2G/3G/4G/5G), NB-IoT, Sigfox, etc.

3) Middleware layer: The middleware layer contains (data analysis, data visualization, Cloud, API, Datacenter,) [1]. It stores, processes and analyses huge amounts of data, big data processing modules, employs databases, cloud computing, classification and polymerization [1]. In this layer there are different methods, tools and techniques that can be used to analyze the data collected by sensors and network layer in order to visualize it in the next layer (application layer). The stored data will be exploited by users. The main function of this layer is the complete analysis via very robust systems that contain Artificial Intelligence, other models/algorithms to do the analysis of the data. Due to its unlimited processing and storage capacity, cloud computing is crucial for the IoT [14]. It also provides a distributed architecture/infrastructure for processing and analyzing IoT data before pushing it to the application layer.

4) Application layer: The application layer visualizes the data produced by IoT computing. It is responsible for delivering personalized services to all industries and to the user [1]. It is the interface for users to communicate to their IoT things and access data. It supports protocols that can be deployed for IoT like MQTT (Message Queue Telemetry Transport)[1], XMPP (Extensible Messaging and Presence Protocol) [1], CoAP (Constrained Application Protocol) [1], SMQTT (Secure MQTT), AMQP (Advanced Messaging Queuing Protocol), etc. HTTP (Hyper Text Transfer Protocol) is used by [16]. This protocol cannot be employed in the application layer, it is not suitable/appropriate for resource-constrained since it is heavy in weight and therefore requires a large parsing overhead [17]. This layer provides personalized services based on the demands of the user. The IoT application covers "intelligent" environments/spaces in areas such as wearable devices, agriculture, transport, factory, building, health, city, lifestyle, home, commerce, vehicles, emergency, supply chain, environment and energy.

5) Summary: This subsection adopted and introduced four level model. To sum up, the IoT objects and things of the physical layer communicate with each other in order to deliver intelligent applications and services for users or human. These IoT devices collected information/data that need to be secured. Billions of devices (medical devices), sensors, actuators and IoT things are connected to the Network. The number of these connected devices is expected to grow increasingly over the coming years. Sensors, RFID and actuators are the major

components of physical layer that can be easily accessed by attackers. These devices need to be managed and secured appropriately, to avoid their significant security risks.

Moreover, IoT involves various communication technologies, which could be affected by different threats. Consequently, for each communication technology, a taxonomy of all possible attacks with their degree of severity/impact/danger must be required. In other words, the communication between these layers need to be secured. Therefore, a secure architecture is needed with some recommended technologies. The first result in this subsection illustrates that the physical and network layer contain the most components that are targeted by attackers. Moreover, because of generating and gathering information by IoT things and devices are the main operations of the physical layer. Therefore, the risk of data theft can be decreased or minimized with physical layer security. Indeed, these data can be stored and processed in local network, so communication technologies and protocols should also be secured. The next subsection will offer insights into the security issues and problems of the IoT architecture. Here, we will give some common issues and challenges of each layer of IoT architecture.

B. The Security Challenges of the IoT Architecture

In this subsection, we analyze security challenges in each layer of the proposed model. Because each IoT layer has its own challenges and security issues.

1) Security challenges and problems of physical layer: At this first layer, the main challenge is the limited resources on IoT devices: storage, memory, CPU, and energy. For instance, sensor is a small equipment with limited resource. Devices of the physical layer are often limited in terms of process and data storage resources, and the applied technologies (such as RFID, NFC, Bluetooth, ZigBee) are being limited in data transmission range and rate. For many IoT devices that are mobile and rely on embedded batteries, energy is one of the most important resources [11]. In terms of energy, the IoT devices need a significant amount of electricity because of their powerful processing capability.

The malicious attack on the identification technology and the sensor is the main challenge for the physical layer [1], which interferes with the collection of data. Things include physical objects (micro-controllers, sensor/sensor nodes, actuators, RFID tags and readers), this physical hardware are targeted by several attacks in the both layers (physical and edge computing layers).

Every IoT devices and Things is linked to the internet in order to talk and communicate to each other. So, there is the possibility of hindering the privacy. Indeed, due to a variety of security vulnerabilities, RFID, and sensors are in threat. Moreover, hardware components and IoT objects are vulnerable to several physical attacks like Object replication attacks, RF Interference on RFID, Hardware Trojan, Object jamming, Physical damage, Camouflage, Malicious node injection, Object tampering, social engineering, Side-channel attack, Malicious code injection, Tag cloning, Outage attacks

[16], False Data Injection Attack, booting vulnerabilities [17], Node Tampering, Node Jamming in WSNs [18] . For instance, the malicious node injection attack targets the physical layer since the node is physically inserted/injected into the network [18]. Besides, side channel attack is conducted at the physical layer because attacker uses side channel information to find the encryption key [18]. At sensors, an attacker can manipulate data, can also do boot attacks, and can capture a node [19].

IoT software are targeted by many attacks. Indeed, hardware components of IoT, like types of RFID tags, sensors, RFID readers, are also vulnerable [16]. The major attacks are targeting the IoT hardware components. Attackers must be located near to hardware or devices in order to launch physical attacks. The attackers may want to physically destroying the devices/hardware, endanger the communication mechanism, tampering the energy source, limiting its lifetime, etc. Also, the attacker can directly access the related attributes of the device through physical attacks, and then start further attack.

The weaknesses of the various devices are exploited in security attacks against IoT systems. IoT devices are subject to well-known vulnerabilities such as the use of unauthenticated requests to conduct/perform actions, broken authentication, sensitive data exposure, infection flaws, XSS (cross-site-scripting), CSRF (Cross Site Request Forgery), missing function-level access control [20].

2) Security challenges and problems of network layer: At the second layer, one of the main/most challenges that face this layer is the heterogeneity of data; for instance, in the network part, there is a huge heterogeneity of data in term of IoT communication technologies (Lora, WI-fi, ZigBee, 4G). This may cause compatibility problems. Compatibility is another major problem. There are some attacks specific to some IoT technologies communication. These technologies of communication include, but not limited to ZigBee, WSN, LoRa, MANET, Bluetooth, RFID, 3G, NFC, Wi-Fi, etc. Moreover, these technologies of communication have their own security concerns and issues. The following are the major security attacks that are faced at the network layer: Phishing Site Attack, Access Attack, DDoS/DoS Attack, Data Transit Attacks, Routing Attacks, unlawful attacks, common attacks [17], Traffic Analysis Attacks, RFID Spoofing, RFID Cloning, RFID Unauthorized Access, Man in the Middle Attacks, Routing Information Attacks [18], Selective Forwarding, Routing Information Attacks, RFID Unauthorized Access, RFID Spoofing, Replay Attack, Traffic Analysis Attack [21].

Additionally, the network attacks consist of manipulating the IoT network system to cause damage. Attack can be started without being close to it (network) [21]. Attacks on networks can have significant consequences, sometimes causing a total shutdown [21]. The software attacks happen when the IoT applications present some security vulnerabilities that allow the attacker to seize the opportunity and harm the system. The main attacks in the Network layer,

targeting the network protocols. An example of network and software attacks in reality is described in [1].

3) Security challenges and problems of middleware layer: Today, the most significant data attacks that are happening in the IoT world: Data Inconsistency, Data Breach and Unauthorized Access [21]. An example of data attacks was launched in reality in March 2018 [21], Cambridge Analytica had obtained access to the private data of more than 50 million Facebook users. Other major security issues at the middleware layer include cloud and data security as well as database security [17].

In the middleware layer, various possible attacks are discussed in [17] such as: Man-in-the-Middle, SQL Injection Attack, Signature Wrapping Attack, Cloud Malware Injection, Flooding Attack in Cloud.

4) Security challenges and problems of application layer: The application layer has specific security problems and issues such as privacy issues and data theft [17]. In the application layer, major attacks are discussed in [17] like Data thefts, Access Control attacks, Service Interruption attacks, Malicious Code Injection attacks, Sniffing attacks, Reprogram attacks.

5) Summary: In this subsection, we have given of the security challenges and problems being faced in each layer of IoT. In my view, the main challenges of IoT systems mainly relate to the first and second layers which are explained and analyzed previously. Since the influence stemming on IoT as IoT devices limitation (resource constraint) and heterogeneity or compatibility problem is very strong. For instance, limited resources challenge for the physical layer and the second challenge is about the heterogeneity of data for the network layer. So, we have to focus on both layers (Perception/physical, and network layer) due to their big/huge challenges and also for other reasons that will be discussed later.

C. Classification of Attacks and Vulnerabilities in Each Layer of IoT Systems

Based on [1], IoT attacks can be classified into fifteen categories: attacks based on vulnerability, on layers, on behavior, on technology of communication, on impact, on security concepts, on target, on software, on source, on devices, on encryption, etc. It is absolutely important to identify vulnerabilities and attacks against each IoT layer. The identification of IoT security attacks is crucial due to the ever-growing number of threats and vulnerabilities in the IoT domains. It is firstly necessary to identify the IoT attacks and vulnerabilities and then classify them.

1) Vulnerabilities in each layer of IoT systems: In terms of vulnerabilities of physical layer, the work [21] have highlighted the vulnerabilities against physical devices like IP cameras, Amazon Echo. For example, attacks such as device spoofing, device scanning, and brute force may control of the cameras. Attackers can get the password for a camera of any length or combination via a device spoofing attack.

Additionally, the attacker can perform a device scanning attack to discover all online cameras by enumerating all MAC addresses that could exist [21]. The vulnerability issue for the "Things" is caused by careless program design, which opens opportunities for the installation of malware or backdoors [22]. The attacker can attack an IoT system by physically weakening or tampering a node [18].

In terms of vulnerabilities of network layer, the attacker can target an IoT system from their own network by exploiting faults in the routing protocol and other network-related protocols or by employing malicious software. The network layer is highly vulnerable/susceptible to phishing site attacks [17].

In terms of vulnerabilities of middleware layer, XML signatures are utilized in the middleware's web services. By exploiting SOAP (Simple Object Access Protocol) vulnerabilities, the attacker can execute operations or alter eavesdropped messages in a signature wrapping attack, which breaks the signature algorithm [17].

At the application layer, insecure cloud interface is a vulnerability in an IoT system that can be an attack vector. Buffer overflow consists in exploiting a vulnerability of an application resulting in abnormal behavior sometimes leading to access to the system with the rights of the application. Software vulnerabilities that allow resource buffer overflows or pushing an IoT device to exhaustion state by an attacker [23]. For instance, a low battery level may cause the laptop to shut down unexpectedly. Due to the majority of the system being in "sleeping" mode, other "things" could not be interoperable [23].

2) Classification of attacks and vulnerabilities based on Layers: In this section, author proposes to summarize the attacks and vulnerabilities based only on the IoT layers, the list is endless. Moreover, each attack has a degree of severity [1], [2]. According to [1], [2], [16]–[18], [21], the security attacks and vulnerabilities at each layer in IoT system are collected and shown in Table I.

3) Summary: After establishing the taxonomy of security attacks and vulnerabilities in each layer of IoT systems. All of these attacks cause significant harm since they alter data, steal sensitive data, drop packets and encryption key, etc.[18]. According to Table I, we showed that the two first layers have been threatening by many IoT attacks and vulnerabilities compared to middleware and application layers. Most attacks on the IoT often occur in IoT objects and IoT network. In other words, most IoT attacks and vulnerabilities have resided in physical layer and network layer because of the poor security design of these connected objects as well as vulnerabilities in IoT protocols and communication technologies. In the next subsection, we will examine and summarize the security requirements for IoT.

TABLE I. CLASSIFICATION OF ATTACKS AND VULNERABILITIES BASED ON LAYERS IN IoT SYSTEMS

Layer	Attacks and vulnerabilities in IoT	Attacks description
Physical Layer	Social Engineering, Node Capture, DoS (Denial of Service) Attack, Distributed DoS Attack, spoofing attack, Fake Node, Replay Attack, Mass Node Authentication, Tag cloning, Unauthorized Access to the Tags, Denial of Sleep Attack, RF Interference on RFID, Eavesdropping, Man In the Middle, RF Jamming, Routing Threats, Object replication, Hardware Trojan, Object jamming, Camouflage, Object tampering, Sleep Deprivation Attack, Outage attacks, Wormhole and Timing attack. Malicious Node Injection, Malicious code Injection, False Data Injection Attack, booting vulnerabilities, Physical damage, Side Channel Attack, Node Tampering, Node Jamming in WSNs, Data Manipulation, Boot Attack.	Both software and hardware components of IoT are targeted by several attacks. Additionally, attackers must be located near to hardware or devices with different intent to launch physical attacks, which can also directly access the related attributes of the device through physical attacks.
Network Layer	Spoofing, MITM attack, Routing Information attack, Sinkhole attacks, Sybil attacks, DoS, Denial of Sleep Attack, Selective forwarding, Eavesdropping/sniffing, Routing attacks (Worm Hole, Hello Flood, Black Hole, Gray Hole, Sybil attack), Phishing Site Attack, Access Attack, DDoS Attack, Data Transit Attacks, Routing Attacks, Malicious code injection, RFIDs interference, unlawful attacks, common attacks, Traffic Analysis Attacks, RFID Spoofing, RFID Cloning, RFID Unauthorized Access, Man in the Middle Attacks, Replay Attack, Routing Information Attacks.	The network attacks consist of manipulating the IoT network system to cause damage. The software attacks happen when the IoT applications present some security vulnerabilities that allow the attacker to seize the opportunity and harm the system.
Middleware Layer	Flooding Attack in Cloud, Malicious Insider, Cloud Malware Injection, Unauthorized Access, Signature Wrapping Attack,Data Inconsistency, Cryptanalysis Attacks, Web Browser Attack, DoS, SQL Injection Attack, MITM, Data Breach, Data Security.	The main challenges are Cloud/data security, and database security.
Application Layer	Buffer Overflow, Botnet, Code Injection, DoS, Sleep Deprivation, Phishing Attack, Sniffing Attack, Authentication and Authorization, DDoS Malicious Scripts, Data Access and Authentication, Trojan Horse, Social engineering, Cryptanalysis Attacks, Brute Force and Cross Site Scripting, Access Control attacks, Service Interruption attacks.	Data theft; Attackers exploit the vulnerabilities of application and programs.

D. IoT Security Requirements

This subsection provides the details about the IoT security requirements. According to [1], [2], [24], the basic security services include authentication, authorization, availability, integrity, confidentiality, and non-repudiation. Therefore, the authors of [24] illustrated only three IoT layers namely Perception, Networking, and Application. Nevertheless, IoT security must also come in the Middleware layer as well as with some other security services including Privacy, Authenticity and Compatibility.

- Privacy needs to be adapted to information as well as devices, this key property must be concerned to the Network and Application layer.

- Compatibility must be concerned to the Network layer.

- The middleware layer must be integrated with confidentiality, integrity and authenticity services [25]. The middleware layer provides these three services to the data exchanged [19].

1) Security services requirements for IoT layers: Generally, security solutions consist of five main objectives as shown in Fig. 2. While, IoT security requirement must be represented by the key properties that are listed below. The IoT security requirements including integrity, availability, confidentiality, authentication, authorization, non-repudiation, privacy, compatibility and authenticity are represented at different layer of IoT as shown in Table II.

a) Confidentiality: It is the property which ensures that only authorized users, under predefined conditions, have access to the information. The IoT system cannot directly apply standard encryption algorithms due to the limited resource of IoT devices. Lightweight cryptographic algorithms are used to guarantee data protection and confidentiality [14].

b) Integrity: It is the property which ensures that information is only modified under predefined conditions. To provide data integrity, a number of cryptographic hash algorithms are utilized, such as MD5 and SH1. However, the majority of these approaches, cannot be used since IoT devices are resource constrained. Many lightweight hash functions were suggested to address this issue [14].

c) Availability: Terminology of the security environment to characterize the proper functioning of the computer system at a given time. IoT device availability is highly necessary. IoT network availability should be handled in both hardware and software. The IoT application's hardware availability refers to every device being present at all times, whereas software availability refers to the capacity to offer services anytime and anywhere[14].

d) Authentication: It is the property that ensures that only authorized entities and users have access to the system or the IoT devices. Authentication protects against identity theft. It is the procedure of validating an identity [14]. Before exchanging data, to connect a device to the network, it needs authenticate itself. Lightweight cryptographic techniques, biometric identification or physical primitives can be used to verify the authentication [14].

e) Authorization: It makes sure that entities have the necessary control permissions to carry out the operation they've requested [13].

f) Non-repudiation: It is the property which ensures that the author of an act cannot then deny having carried it out. The second idea contained in the usual notion of signature is that the signatory undertakes to honour his signature: contractual, legal commitment, he can no longer go back. It is an important element of network security [14]. It is the capacity to assure that an IoT node cannot reject/deny having

sent a message and that the recipient cannot deny having received it. Public Key Cryptography can be used to achieve it.

g) Privacy: Attacks on privacy are linked to the unauthorized collection of sensitive information about individuals. When collecting, transmitting, and storing data, data privacy must be considered. The issue of data privacy has received many practical solutions. Stream ciphers, Block ciphers, pseudo-random number generators, and anonymization are some of these methods [14].

h) Compatibility: Of the emerging and the existing IoT protocols in the network layer, the big challenge concerning this layer is the huge heterogeneity of data in term of IoT communication technologies (LoRa, Wi-fi, ZigBee, 4G, etc.).

i) Authenticity: [25] Illegal users are not permitted to access the system or obtain sensitive data.

2) Summary: According to Table II, we showed that generally, the two first layers are missing various security shields. For instance, confidentiality, integrity, authentication, availability, and authorization are the major problems/needs in the physical layer. The network layer requires integrity, availability, authentication, authorization, non-repudiation, privacy and compatibility. If these security requirements have identified in the two first layers, so the risk of unauthorized access can be minimized, this means that if IoT devices and things are secured then, the access control in the middleware and application layers can be achieved.

Fig. 2. Taxonomy of IoT Attacks based on Security Concept.

TABLE II. THE BASIC SECURITY REQUIREMENTS FOR IOT SYSTEMS

IoT Layers	Security Services/concepts								
	Confidentiality	Integrity	Availability	Authentication	Authorization	Non-repudiation	Privacy	Compatibility	Authenticity
Physical Layer	✓	✓	✓	✓	✓				
Network Layer		✓	✓	✓	✓	✓	✓	✓	
Middleware Layer	✓	✓							✓
Application Layer		✓		✓	✓		✓		

E. Security Threats and Vulnerabilities Analysis

1) Overview of OWASP Internet of Things Project: Based on an open community and a collection analysis provided by security industry professionals, the Open Web Application Security Project's or OWASP IoT Project released its Top 10 2018 [26], which publish a report that represents the top 10 security problems and issues to avoid when using, managing, creating, deploying an IoT system. This project has listed the main concerns and vulnerabilities of IoT systems based on different IoT architecture levels. The OWASP IoT Project [15] shown that a large number of IoT vulnerabilities are caused by a lack of adoption of existing/common security mechanisms including access control, authentication, role-based access control and encryption. Therefore, because of the complicated characteristics of IoT, establishing and implementing security techniques, measures, practices, and tools is not easy. The Table III represents each security concerns by a number ranged from 1 to 10.

2) Summary: After analyzing this project, author observed that the most of issues and vulnerabilities are surrounding physical layer and most of them target the IoT devices. For instance, "Lack of Physical Hardening" has been identified by this project. This means the lack of physical security measures enables potential attackers to access sensitive data that may be used in a future distant attack or to obtain local control of the device. So, "Lack of Physical Hardening" is the most important concern that tackles the physical layer. In addition, we need to ensure that only the authorized people can access the sensitive data produced by devices or physical objects. Besides, if the security will be implemented in the physical layer including IoT things, as well as in the network layer including IoT protocols and communication technologies, then, data theft will be minimized. Consequently, significant enhancements are required to make the IoT framework secure and safe. In the next subsection, we will examine an example of security modeling in IoT systems with a UML extension called IoTsec.

F. Example of IoT Security Systems Modeling with UML

Any artificial language that may be used to convey information, knowledge, or systems in a framework determined by a set of rules is referred to as a modeling language. The UML aims to standardize the various ways for describing object-oriented systems that already exist. IoT interconnects smart entities anyhow and anywhere. Since, IoT rises new issues as well as modeling IoT security systems is a field that lacks the modelling languages for representing IoT systems in several views. The main objective of this subsection is to find the most effective extension instead of a language for IoT security modeling. Based on the characteristic of each tool, we choose two important criteria: (1) specific for IoT systems and (2) System security modeling. The choice of these points depends on our goal which is modeling IoT security. According to the extensions comparison mentioned in [4], we note that UML4IoT and SysML4IoT extensions can model IoT systems, but they lack of security matters. Other extensions like UMLsec, and

SysMLsec can be specified for security modeling, but they are not specific for modeling IoT systems. In IoT systems, IoTsec aims at modeling security issues, it is a subset of SysML and UML. It combines UML, SysML, and UMLsec. All these reasons make the IoTsec an ideal example of UML extensions, because it is the only one designed to enable security modeling for IoT systems. In this section, we give an interesting example which used a new UML extension called IoTsec that involves security issues of IoT systems (see Fig. 3). Another example done by [4] shows a layer diagram with IoTsec. Fig. 3 shows a class diagram for an IoT device with IoTsec [4], where an IoTdevice called RapberryPi3. RapberryPi3 uses a relational class N to authenticate a temperature sensor, N means authentication, in this example the attributes N have not been established yet, but they might be any authentication protocol. According to this analysis, we justified the choice of IoTsec as the best UML/SysML extension compared to the existing ones.

TABLE III.   OWASP IOT TOP 10 SECURITY CONCERNS –2018 VERSION

N°	The main concerns in IoT systems	Description
1	Weak, Guessable, or Hardcoded Passwords [26]	Use of easily brute forced, publicly available, or unchangeable credentials
2	Insecure Network Services [26]	Unneeded or insecure network services running on the device itself
3	Insecure Ecosystem Interfaces [26]	Lack of authentication/authorization, lacking or weak encryption
4	Lack of Secure Update Mechanism	Lack of ability to securely update the device
5	Use of Insecure or Outdated Components [26]	Insecure software components/libraries that could allow the device to be compromised
6	Insufficient Privacy Protection [26]	All aspects of the IoT architecture that potentially expose sensitive unencrypted data must be taken into account[27].
7	Insecure Data Transfer and Storage [26]	Lack of encryption or access control of sensitive data anywhere within the ecosystem
8	Lack of Device Management [26]	Lack of security support on devices deployed in production
9	Insecure Default Settings [26]	Devices or systems shipped with insecure default settings
10	Weak, Guessable, or Hardcoded Passwords [26]	Use of easily brute forced, publicly available, or unchangeable credentials
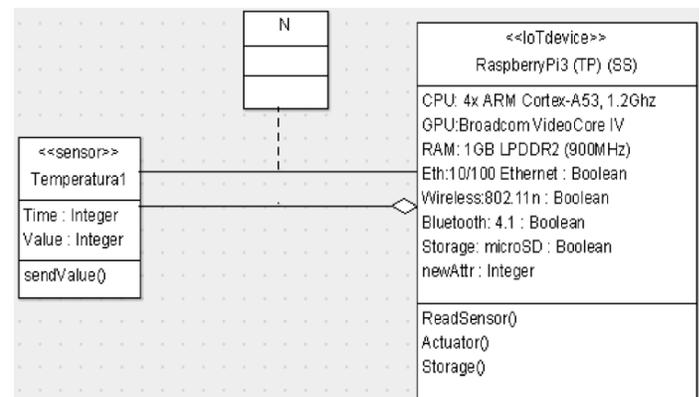
Fig. 3.   Example of a Class Diagram for an IoT Device with IoT Sec [4].

For instance, Fig. 4 shows the Ouchani's modeling. For this example, the author suggested IoT architecture enclosed five components:

- Object devices: physical objects embedded with software and sensors.

- User devices: physical objects that collect data from objects and communicate with servers.

- Computing services provided by external, internal, and cloud servers.

- Social actors: human agents that can manipulate and hold devices.

- The environment: the infrastructures which envelops the IoT entities.

These components interact via communication protocols of various ranges (ZigBee, WiFi, Cellular, Human-machine, Bluetooth, SSH, IpSec, etc.). Table IV shows results achieved and describes clearly what has been done before on the problem.
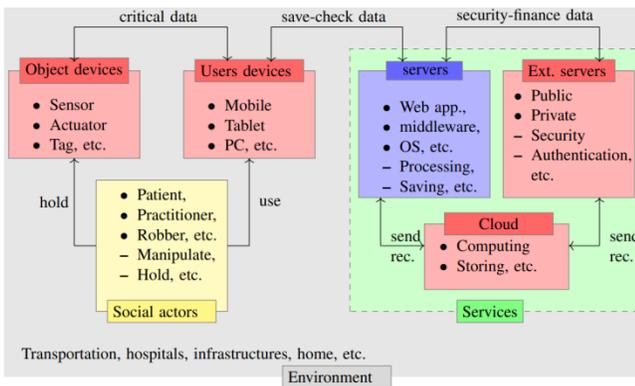
Fig. 4. IoT-SEC Components Architecture [7].

TABLE IV. COMPARATIVE STUDY RELATED TO MODELLING IOT SECURITY

Survey	Modelling IoT security
Jha et al., 2020 [3]	Authors represent the architecture of IoT-Edge computing. Authors modeled an edge infrastructure. So, they implemented and designed numerous new classes.
Robles-Ramirez et al., 2017[4]	Regarding security concerns, authors aim to represent an IoT system. They adopted the four layers architecture including sensing, network, service, and application layer. Moreover, the nomenclature in classes is also shown. Further, this may be reached by the Block Diagram of SysML.
Samir Ouchani, 2018 [7]	The author describes the IoT architecture by presenting the IoT components with their interactions. The suggested IoT architecture enclosed five components including object devices, social actors, user devices, computing services, and the environment.
Proposed	Through this paper, the four-layer IoT architecture (Physical/perception, Network, Middleware, and Application) are proposed. Further, Physical and Network layers are chosen for IoT security modelling, which really contributes to the security of IoT systems

V. RESULTS AND DISCUSSION

In this work, the main idea is to choose the layer whose modelling is relevant. Ensuring security in IoT systems is more difficult due to several challenges. There are four layers generally, so far there is no standardization. So, according to my analysis, we should specify which layer we will work on, otherwise, we model the different layers. Several IoT architectural models have been proposed in the literature. To understand better IoT systems, we choose to work with a four-layer IoT architecture, which is composed with: perception layer, network, middleware, and application layer. Based on all observations and the Tables II and III, author found that there are two layers that must be modelled in IoT security systems, these two layers are called the physical and network layers which contain great challenges, issues, concerns, major attacks and vulnerabilities.

As we discussed in methodology section that security issues and concerns were only designed for both physical and network layers and not for middleware and application layer. In other words, taking the conclusions and results into consideration from each subsection of the previous section, it is fitting that we should only center on physical and network layers, while middleware and application layers are left out. This means that both physical and network layers have specific security concerns, issues, problems, attacks, vulnerabilities, and challenges which are not present in both middleware and application layers. The main focus is on the physical layer and the network layer due to the limitations and constrains brought by the IoT devices as well as the compatibility issue.

In addition, for each layer a detailed analysis has been done in terms of four-layer architecture. The goal is to model IoT security system at the design stage, even if developers are not fully familiar with cyber security concepts. IoT system design is difficult, and UML modeling is proving to be a useful tool for overcoming it. In general, UML, SysML, and ThingML are the three modeling languages. While, IoTsec [4], UMLsec [7], SysMLsec [7], SysML4IoT [7], UML4IoT [8] are some extensions of UML and SysML. Based on extensions comparison which have done in [4], we noticed that, neither UML nor SysML languages can be used for security modeling of IoT systems. Whereas ThingML is the most used language for modeling IoT systems. However, this language only specific for IoT systems and does not model the systems security. The only UML/SysML extension to model IoT security is the "IoTsec".

There are some challenges for modelling/designing IoT applications:

- Heterogeneity: due to its different virtual and physical components with several characteristics that are embedded, forming a complex system; as well as the compatibility problem (communication technologies).

- Interoperability: in terms of infrastructure. Indeed, resources are restricted and limited in the physical layer.

- Distribution: over a large number of processing nodes.

- The absence of a model for addressing security concerns in IoT systems.

- A lack of a design paradigm for IoT applications.

- The lack of standardization in IoT architecture is another restriction.

Thus, the challenge of evolving and deploying software for the IoT is frequently underestimated [28]. IoT applications have two primary characteristics from the standpoint of software engineering. The first factor is the distribution over a large number/range of processing nodes. While, the second one is the high/significant heterogeneity of processing nodes as well as the protocols that connect them [28]. With UML resources, it is possible to represent a small IoT system. UML diagrams and extensions can be used to represent the various views of an IoT environment (security, static, behavioral, etc. [29]). Tables V and VI present the added values and the weakness of each paper to deduce what have been done before on the problem, and what is new.

Therefore, modelling IoT systems is challenging due to their heterogeneity, which is caused by the integration of physical and virtual components, resulting a complex system [29]. Indeed, the IoT is often viewed as a single-issue domain [30]. The UML is a language [29] of general usage for documenting, specifying, visualizing, and constructing artifacts of the software system [29]. Simulating a realistic scenario in IoT is very challenging [31]. The challenge of modeling such complex systems lies in the heterogeneity of these systems, due to their different virtual and physical components that are embedded. Indeed, UML resources are used to represent the various views of an IoT environment or application using its extensions and diagrams [29]. Since security is considered as one of the most crucial quality attributes in networking [32] and also in the field of IoT and software engineering, it is necessary to provide holistic protections for IoT architecture [33].

To answer the research question, author mainly based on the proposed methodology including the results of all subsections and challenges related to the security of IoT systems (IoT layers). As we discussed in the above section, the physical layer can be easily accessed by attackers as well as the main operations of this layer. The IoT communication technologies and protocols of the network layer are susceptible to security vulnerabilities. Both layers impacted differently by the security issues, problems, attacks, threats and vulnerabilities compared to two last layers. Compatibility, heterogeneity and resource constraint of the two first layers made the IoT security worse. All these reasons make security of IoT systems complex and more difficult. Consequently, we need to model the two first layers (physical and network layers).

Finally, the relevant layers to model that really contribute to the security of IoT systems are two layers namely, the physical layer and the network layer. The choice of these two layers poses at least two major and huge problems which may

slow down the whole IoT systems development. For that, we should have interested in these two layers more. To sum up, security needs to be modeled in both physical and network layers. According to Table V and Table VI we can notice that a method to design IoT systems with their security issues is required.

TABLE V. COMPARISON OF RELATED SURVEYS

	Objectives	Results
[3]	Authors aim to model realistic IoT and edge environments. To model an edge infrastructure, they implemented and designed numerous new classes. IoT & Edge computing	The findings demonstrate that IoTSim-Edge has different capabilities in terms of mobility modeling, heterogeneous protocols modeling, battery-oriented modeling, application composition and resource provisioning for IoT applications.
[4]	For IoT systems the use of the UML was proposed by the authors of [4], with also the suggestion of extensions. Besides, IoTsec was also presented by authors; The aim is to facilitate the representation of security concerns with a visual notation, even if the developers are not totally or completely familiar to Internet cybersecurity concepts.	For modeling common actors, IoTsec uses UML extensions for security encapsulated in UML nomenclature and stereotypes. They aim at detailing the activities developed in three stages.
[5]	Authors aims at modeling environment that enable to deploy, design, and specify complex IoT systems;	The approach introduced by authors uses already IoT-A in the Papyrus for IoT modelling environment.
[6]	Authors suggested extensions and used UML for IoT systems; IoTReq was the proposed method;	The IoT system present peculiar characteristics that necessitate the use of specific approaches to represent their requirements, implementing hardware, software intersection.
[7]	Authors describes the IoT architecture by showing its components with their interactions.	The suggested IoT architecture enclosed five components as shown in Fig. 4.
[8]	Authors aims at using the UML, or UML profiles, for supporting the IoT systems development.	Authors used UML4IoT an UML approach based on the use of UML profile. UML4IoT is an UML profile for IoT

TABLE VI. ANALYSIS OF THE EXISTING SURVEYS

Survey	Weakness
[3]	The authors didn't mention some different factors such as storage technology, …
[5]	This paper has not a detailed and comprehensive look;
[7]	The authors don't consider privacy, security and trust domains.
[8]	The mentioned extension does not support Security;
[10]	This paper lacks more experiments. Moreover, the proposed framework suffers from the limitations;

VI. CONCLUSION

Existing visions have resulted in a lack of knowledge of the architecture of IoT systems. Hence, there is no standardization about the IoT architecture. Moreover, in the previous works, there is no answer about the following questions: Which is the layer whose modeling is relevant? What is/are the relevant layer(s) to model that really contribute to the security of IoT systems? Through this work, several axes have been presented and detailed. Hence, a proposed IoT architecture has also been investigated. Moreover, security concerns at various layers of IoT architecture were described in this study. Then, the challenges of IoT architecture were demonstrated. The chosen or proposed IoT architecture consists of four layers (see Fig. 1), known as the perception, network, middleware and application layer.

This paper covers the modelling of IoT systems. The objective of this article is to model complex systems. Modelling of IoT application in the real environment is difficult, complex, time-consuming and not effective in terms of cost [3]. To facilitate modelling of IoT systems, several languages and extensions have been developed. As a result, this study recommends the use of IoTsec for modeling IoT security because it is the only one that enable us to model IoT systems and security that are our needs/goal. In this paper, we have studied the comparison in terms of IoT layers. Further, a comparison between previous works in terms of modelling IoT security systems have been also made in order to deduce the relevant layer to model that really contributes to the security of IoT systems.

We have studied in detail the architecture of the IoT systems in order to be able to deduce that physical and network layers have many challenges, issues, vulnerabilities, attacks and need more security. In this paper, the physical layer and the network layer are the two layers chosen for modelling. Therefore, the security of IoT systems must be taken and considered in an earlier stage during the design phase.

Our future study will focus on making a comparison between UML and SysML modeling languages to prove the efficiency of SysML with a concrete case of IoT systems which is forest fires.

REFERENCES

[1] H. Meziane, N. Ouerdi, M. A. Kasmi, and S. Mazouz, "Classifying Security Attacks in IoTUsing CTM Method," in Emerging Trends in ICT for Sustainable Development, Cham,2021, pp. 307–315. doi: 10.1007/978-3-030-53440-0_32.

[2] M. Hind, O. Noura, K. M. Amine, and M. Sanae, "Internet of Things: Classification ofattacks using CTM method," in Proceedings of the 3rd International Conference onNeworking, Information Systems & Security, New York, NY, USA, Mar. 2020, pp. 1–5. doi: 10.1145/3386723.3387876.

[3] D. N. Jha, K. Alwasel, A. Alshoshan, X. Huang, R. K. Naha, S. K. Battula, S. Garg, D. Puthal, P. James, A. Zomaya, S. Dustdar, and R. Ranjan, "IoTSim-Edge: A simulation framework for modeling the behavior of Internet of Things and edge computing environments," Software - Practice and Experience, pp. 1–19, 2020.

[4] D. A. Robles-Ramirez, P. J. Escamilla-Ambrosio, and T. Tryfonas, "IoTsec: UML Extension for Internet of Things Systems Security Modelling," in 2017 International Conference on Mechatronics,

[5] S. Dhouib et al., "Papyrus for IoT—a modeling solution for IoT," Proceedings l'Internet des Objets (IDO: Nouveaux Défis de l'Internet des Objets: Interaction Homme-Machine et Facteurs Humains. Paris, France, 2016.

[6] G. Reggio, "A UML-based proposal for IoT system requirements specification," in Proceed-ings of the 10th International Workshop on Modelling in Software Engineering - MiSE '18, Gothenburg, Sweden, 2018, pp. 9–16. doi: 10.1145/3193954.3193956.

[7] S. Ouchani, "Ensuring the Functional Correctness of IoT through Formal Modeling and Verification," in Model and Data Engineering, Cham, 2018, pp. 401–417. doi: 10.1007/978-3-030-00856-7_27.

[8] K. Thramboulidis and F. Christolakis, "UML4IoT—A UML-based approach to exploit IoT in cyber-physical manufacturing systems," Computers in Industry, vol. 82, pp. 259–272, Oct. 2016, doi: 10.1016/j.compind.2016.05.010.

[9] D. Ferraris, C. Fernandez-Gago, and J. Lopez, "A model-driven approach to ensure trust in the IoT," Hum. Cent. Comput. Inf. Sci., vol. 10, no. 1, p. 50, Dec. 2020, doi: 10.1186/s13673-020-00257-3.

[10] N. Harrand, F. Fleurey, B. Morin, and K. E. Husa, "ThingML: a language and code genera-tion framework for heterogeneous targets," in Proceedings of the ACM/IEEE 19th Inter-national Conference on Model Driven Engineering Languages and Systems, Saint-malo France, Oct. 2016, pp. 125–135. doi: 10.1145/2976767.2976812.

[11] F. Alkhabbas, R. Spalazzese, and P. Davidsson, "Characterizing Internet of Things Systems through Taxonomies: A Systematic Mapping Study," Internet of Things, vol. 7, p. 100084, Sep. 2019, doi: 10.1016/j.iot.2019.100084.

[12] K. K. Patel, S. M. Patel, and P. Scholar, "Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges," p. 10, 2016.

[13] M. Dabbagh and A. Rayes, "Internet of Things Security and Privacy," in Internet of Things From Hype to Reality, Cham: Springer International Publishing, 2017, pp. 195–223. doi: 10.1007/978-3-319-44860-2_8.

[14] Y. Harbi, Z. Aliouat, S. Harous, A. Bentaleb, and A. Refoufi, "A Review of Security in Internet of Things," Wireless Pers Commun, vol. 108, no. 1, pp. 325–344, Sep. 2019, doi: 10.1007/s11277-019-06405-y.

[15] A. Alkhalil and R. A. Ramadan, "IoT Data Provenance Implementation Challenges," Pro-cedia Computer Science, vol. 109, pp. 1134–1139, 2017, doi: 10.1016/j.procs.2017.05.436.

[16] H. Akram, D. Konstantas, and M. Mahyoub, "A Comprehensive IoT Attacks Survey based on a Building-blocked Reference Model," ijacsa, vol. 9, no. 3, 2018, doi: 10.14569/IJACSA.2018.090349.

[17] V. Hassija, V. Chamola, V. Saxena, D. Jain, P. Goyal, and B. Sikdar, "A Survey on IoT Security: Application Areas, Security Threats, and Solution Architectures," IEEE Access, vol. 7, pp. 82721–82743, 2019, doi: 10.1109/ACCESS.2019.2924045.

[18] J. Deogirikar and A. Vidhate, "Security attacks in IoT: A survey," in 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Feb. 2017, pp. 32–37. doi: 10.1109/I-SMAC.2017.8058363.

[19] M. S. A. Reshan, "IoT-based Application of Information Security Triad," International Journal of Interactive Mobile Technologies (iJIM), vol. 15, no. 24, Art. no. 24, Dec. 2021, doi: 10.3991/ijim.v15i24.27333.

[20] M. B. Barcena and C. Wueest, "Insecurity in the Internet of Things," Security response, symantec, vol. 20, 2015.

[21] J. Sengupta, S. Ruj, and S. Das Bit, "A Comprehensive Survey on Attacks, Security Issues and Blockchain Solutions for IoT and IIoT," Journal of Network and Computer Applica-tions, vol. 149, p. 102481, Jan. 2020, doi: 10.1016/j.jnca.2019.102481.

[22] Z.-K. Zhang, M. C. Y. Cho, C.-W. Wang, C.-W. Hsu, C.-K. Chen, and S. Shieh, "IoT Security: Ongoing Challenges and Research Opportunities," in 2014 IEEE 7th Interna-tional Conference on Service-Oriented Computing and Applications, Nov. 2014, pp. 230–234. doi: 10.1109/SOCA.2014.58.

[23] M. Nawir, A. Amir, N. Yaakob, and O. B. Lynn, "Internet of Things (IoT): Taxonomy of security attacks," in 2016 3rd International

Electronics and Automotive Engineering (ICMEAE), Nov. 2017, pp. 151–156. doi: 10.1109/ICMEAE.2017.20.

Conference on Electronic Design (ICED), Aug. 2016, pp. 321–326. doi: 10.1109/ICED.2016.7804660.

[24] M. Azrour, J. Mabrouki, A. Guezzaz, and A. Kanwal, "Internet of Things Security: Challenges and Key Issues," Security and Communication Networks, vol. 2021, p. e5533843, Sep. 2021, doi: 10.1155/2021/5533843.

[25] C. Ataç and S. Akleylek, "IoT Çağında Güvenlik Tehditleri ve Çözümleri Üzerine Bir Araştırma," European Journal of Science and Technology, pp. 36–42, Mar. 2019, doi: 10.31590/ejosat.494066.

[26] "OWASP Internet of Things | OWASP Foundation." https://owasp.org/www-project-internet-of-things/ (accessed Apr. 13, 2022).

[27] M. M. Anghel, P. Ianc, M. Ileana, and L. I. Modi, "The Influence of Privacy and Security on the Future of IoT," IE, vol. 24, no. 2/2020, pp. 42–53, Jun. 2020, doi: 10.24818/issn14531305/24.2.2020.04.

[28] B. Morin, N. Harrand, and F. Fleurey, "Model-Based Software Engineering to Tame the IoT Jungle," IEEE Softw., vol. 34, no. 1, pp. 30–36, Jan. 2017, doi: 10.1109/MS.2017.11.

[29] M. T. B. Geller and A. A. de M. Meneses, "Modelling IoT Systems with UML: A Case Study for Monitoring and Predicting Power Consumption," American Journal of Engi-neering and Applied Sciences, vol. 14, no. 1, pp. 81–93, Feb. 2021, doi: 10.3844/ajeassp.2021.81.93.

[30] H. Lin and N. W. Bergmann, "IoT privacy and security challenges for smart home envi-ronments," Information, vol. 7, no. 3, p. 44, 2016.

[31] G. Kecskemeti, G. Casale, D. N. Jha, J. Lyon, and R. Ranjan, "Modelling and simulation challenges in internet of things," IEEE cloud computing, vol. 4, no. 1, pp. 62–69, 2017.

[32] K. Ahmed, S. Verma, N. Kumar, and J. Shekhar, "Classification of Internet Security Attacks," in Proc. 5th Natl Comput. Nation Dev., Delhi, India, Mar. 10-11, 2011.

[33] K. Chen et al., "Internet-of-Things security and vulnerabilities: Taxonomy, challenges, and practice," Journal of Hardware and Systems Security, vol. 2, no. 2, pp. 97–110, 2018.

DevOps Enabled Agile: Combining Agile and DevOps Methodologies for Software Development

Shah Murtaza Rashid Al Masud, Md. Masnun, Mst. Afia Sultana
Anamika Sultana, Fahad Ahmed, Nasima Begum
Department of Computer Science and Engineering
University of Asia Pacific, Dhaka, Bangladesh

Abstract—The Agile and DevOps software development methodologies have made revolutionary advancements in software engineering. These methodologies vastly improve software quality and also speed up the process of developing software products. However, several limitations have been discovered in the practical implementation of Agile and DevOps, including the lack of collaboration between the development, testing and delivery sectors of different software projects and high skill requirements. This paper presents a solution to bridge the existing gaps between Agile and DevOps methodologies by integrating DevOps principles into Agile to devise a hybrid DevOps Enabled Agile for software development. This study includes the development of a small-scale, experimental pilot project to demonstrate how software development teams can combine the advantages of Agile and DevOps methodologies to fulfill the gaps and provide further improvements to the speed and quality of software development process while maintaining feasible skill requirements.

Keywords—Agile; DevOps; gaps; collaboration; skill; DevOps Enabled Agile; software development

I. INTRODUCTION

The field of software development is evolving continuously with new methodologies and techniques constantly being devised to bridge the gaps of previous ones. Our motivation is to propose and verify the DevOps Enabled Agile methodology, a methodology that will combine Agile with DevOps to fulfill the gaps of both methodologies in order to further advance the field of software development.

This study includes an in-depth analysis of Agile and DevOps methodologies, describing and comparing the different practices, requirements, challenges etc. and concluding that both methodologies have disadvantages and drawbacks that can reduce the effectiveness of individually implementing them for software development projects. On the basis of this analysis, the implementation of DevOps Enabled Agile has been justified as the solution to this problem.

For practical application of implementing DevOps Enabled Agile methodology, we propose combining standard Agile methods with the cross-functionality and collaboration principles of DevOps. We have developed an experimental pilot project using this procedure and conducted an in-depth comparison of the results gained by observing the performance of both Agile and DevOps Enabled Agile methodologies in order to demonstrate the benefits of DevOps Enabled Agile methodology.

The remaining parts of this paper are organized in the following ways; Section II presents an in-depth comparative analysis of Agile and DevOps methodologies and provides justification for combining these methodologies into DevOps Enabled Agile methodology. Section III contains the background and related works of our study in the form of a systematic literature review. Section IV describes the proposed methodology used in this study to test and verify the effectiveness of DevOps Enabled Agile for an experimental pilot project. This is followed by Section V which contains the results from the experimental pilot project and a discussion of the findings that justify the implementation of DevOps Enabled Agile. Lastly, in Section VI, the study is concluded with a brief analysis of the scope for future research.

II. ANALYSIS OF AGILE AND DEVOPS

Agile methodology is a highly efficient and effective software development methodology which is focused on four core principles [1]. These principles emphasize the importance of individuals and interactions over processes and tools, working software over comprehensive documentation, customer collaboration over contract negotiation, and responding to change over following a plan.

There are many methods and techniques to implement Agile methodology depending on project scope, knowledge domain and other factors including DAD, SAFe, XP, Scrum etc. [2] Scrum in particular is an agile framework designed for small teams which is relatively simple to use and highly adaptable to changes [3].

Agile methodology provides many different benefits over traditional software development methodologies [4]. The main benefits include lower documentation requirements, strong collaboration with customers, rapid delivery and continuous alignment with business needs and goals.

However, we have discovered that Agile methodology has a number of major gaps including the separation of the development, testing and delivery sectors and lack of cross-functionality [5]. Agile methodology does not support continuous integration of feedback during development, thus limiting the speed and efficiency of software development. Additionally, the lack of cross-functionality within developer teams may result in disconnected and uncoordinated development, testing and delivery sectors which may lead to project completion and progression not aligning with planned estimates.

However, DevOps (Development-Operations) is an extension of Agile methodology which can harmonize the development and operation processes/sectors and provide cross-functionality within teams. Thus, DevOps can mitigate some of the drawbacks of Agile methodology.

DevOps is a software development methodology focused on the collaboration of developers and operators, but it lacks concrete definition and is referred to as a skillset, a set of practices and even as a job description [6]. DevOps focuses on cross-functionality between different sectors of development, operations and quality assurance which would result in highly coordinated development, testing and delivery which would lead towards smoother project progression [7].

But, pure DevOps certainly cannot focus on fulfilling continuously changing user requirements by receiving and incorporating feedback in incremental cycles. Additionally, pure DevOps has high skill requirements that cannot be fulfilled by software engineers who lack extensive training and experience on DevOps culture and practices. Thus fully incorporating DevOps into public sectors such as mid and low level software development projects would result in difficulties and drawbacks that would be unacceptable for efficient and effective software development process [8] [9] [10].

Hence, to mitigate the existing problems, we propose combining Agile and DevOps methodologies into the hybrid DevOps Enabled Agile methodology for software development. Our research and analysis have revealed that DevOps and Agile methodologies can be combined together to produce the best results and mitigate the existing problems of both methodologies. Our method will follow the examples set by related works that have suggested combining Agile with DevOps to utilize the benefits of both methodologies [11] [12] [13] [14].

Our proposed methodology, DevOps Enabled Agile has been designed to combine the collaborative and cross-functional aspects of DevOps with the simplicity and low skill requirements of Agile to bridge the gaps of both methodologies, thus ensuring rapid progression and maximum effectiveness of the development, testing and delivery sectors of software development projects. Therefore, this analysis justifies the implementation of DevOps Enabled Agile for software development.

III. RELATED WORK

Our study includes a systematic literature review of different related works. We have screened ten existing papers in particular as the basis of our research and analyzed the problem statement, research method and limitations of these papers as presented below in order to conduct the literature review.

Marius Andersen Bjørni and Simen Haugen have conducted interviews and observations to discover the major challenges of Agile methodology which consist of high sprint workloads, lack of testing in sprints before release, poor PBI descriptions, lack of business agility, lack of documentation, inadequate PBI grooming, lack of team improvement, inefficient release process and not holding sprint reviews.

However, this study has not considered work estimation during research, leaving this factor as a research gap [5].

Koi-Akrofi et al. have also conducted a study on Agile methodology based on literature review. They have pointed out the notable advantages of Agile over traditional methodologies, which include adaptability towards changing customer requirements, simple method of updating, priority assessment, feedback integration, continuous testing and motivation for developers. They have also pointed out the notable challenges of Agile methodology which are the unpredictability of development, requirement of high dedication, time and effort, greater customer demands, lack of documentation and projects going off-track. However, this study is purely a literature review and does not use any additional method of research, thus showing that fully transitioning to DevOps also has major difficulties and drawbacks [15].

Farid et al. have conducted a study on combining DevOps with Lean Software Development. This study has presented Lean Software Development as a software development methodology focusing on waste reduction and continuous improvement while presenting DevOps as a set of practices targeted towards improvement of the overall development life cycle by integrating development and operations together in cross-functional teams. The research method involves implementing DevOps practices to bridge discovered gaps in Lean Software Development. This method was able to utilize DevOps practices to overcome issues of Lean Software Development in a number of different areas including Delays, Defects, Extra Features, Task Switching etc. The limitation of this work is that the application of DevOps principles to overcome existing issues has only been tested with Lean Software Development and the exploration of similar cases of applying DevOps principles for Scrum, XP and other methods have been left as prospects for future research [16].

Banica et al. presented DevOps as a Project Management methodology for software development projects. It utilizes a pilot project as the method of research, incorporating DevOps facilities into the development process of a website using VersionOne tool which is highly suitable for project management systems supporting DevOps methodology. The results of this project show that DevOps speeds up and improves the development process. However, due to the inexperienced teams used, some problems were faced during development, including delays and missing certain development goals, which classifies as the limitation of this research [17].

Wiedemann et al. provides an in-depth analysis on the application of DevOps methodology for IT projects. A qualitative multiple case study was conducted in eight different industries that utilize DevOps to achieve intra-IT alignment, justifying the use of DevOps in IT project management in order to orchestrate development and operations within IT functions. However, this study only proposes using DevOps methodology for this purpose, neglecting the potential of combining it with Agile methodology to gain further advantages and overcome existing disadvantages, thus leaving a gap in the research [18].

Kuusinen et al. observed and analyzed the progression of two groups of students working on software projects transitioning from Agile to DevOps methodology. This research discovered several challenges and problems in the transition process, including legacy systems, issue of rights and slow testing. Applying continuous testing and the collaboration of cross-functional teams have been suggested as solutions to solve these problems. However, a more complete solution involving the utilization of Agile methodology in coordination with DevOps methodology instead of fully transitioning to DevOps methodology has been left unexplored, which is a limitation to be addressed by future research [19].

Sikender Mohsienuddin Mohammad briefly compares Agile and DevOps Methodologies by analyzing their similarities and differences. The similarities of Agile and DevOps are that both are focused on quick and efficient iterative feedback-based development of software. They can be used together in a coordinated way to speed up and improve the software development process. On the other hand, there are also some major differences between the two methodologies. These differences include Agile only taking feedback from customers while DevOps additionally takes feedback from internal team and Agile minimizing the gap between developers and customers while DevOps minimizes the gap between developers, testers and operators. However, this paper neglects to mention the challenges of DevOps methodology, leaving it as a gap in the research [20].

Hemon et al. have proposed a maturity model for transitioning from Agile to DevOps, utilizing three stages of progression; Agile, Continuous Integration and Continuous Delivery. This model focuses on the alignment of development and operations while aiming to achieve greater smartness in IS function. However, the scope of this research does not extend towards the field of software development despite Agile and DevOps having a high degree of application in this field. Thus, it can be considered as a research gap [8].

Hemon et al. conducted an additional study on the subject of Agile to DevOps transition, focusing on individual roles, collaboration and skills. The research method applied data collected from a case study in an organization with years of experience in DevOps to form the basis of the research. The results of the research conclude that DevOps allows for balanced collaboration between individuals but also requires skilled developers and operators to be used efficiently. However, the extent of more diverse collaboration has not been fully demonstrated in this study, which could be explored in future research [10].

IV. PROPOSED METHODOLOGY

Our study has utilized the development of a small scale, experimental pilot project using Agile and DevOps Enabled Agile methodologies followed by a systematic comparison of the results in order to justify and validate our proposed methodology for software development.

The development process consists of a three-phase method followed by the developer teams:

1) The first phase consists of the identification of customer requirements before converting them into a sprint backlog.

2) The second phase consists of developing the components of the project and testing them to resolve any issues.

3) The third phase consists of delivering each component to the customer and gaining feedback to keep the project on track and ensure fulfillment of requirements.

We have selected a simple web application for the students of a university to perform course selection and manage payments as our pilot project. In the initialization process, we have formed a team consisting of three members, the developer, the tester and the customer representative.

A use case diagram illustrating the team members, their contributions to the pilot project and the information flow between them is given in Fig. 1:
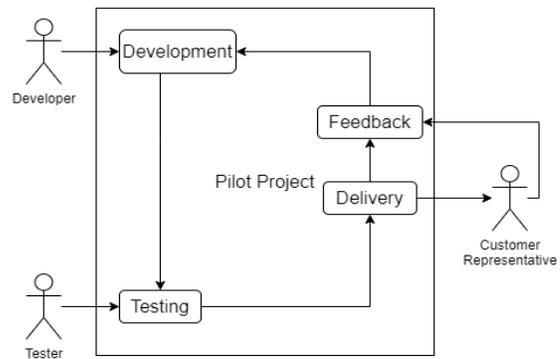
Fig. 1. Use Case Diagram of the Pilot Project.

The customer representative provided a list of requirements for the project, illustrated in Fig. 2:
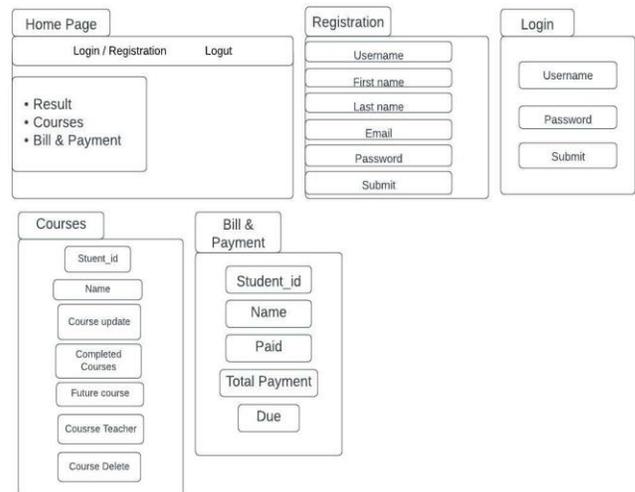
Fig. 2. Requirements Diagram of the Pilot Project.

We have analyzed these requirements describing the structure of the project and converted the results into a sprint backlog consisting of four sprints for each module of the web

application with two to three story points denoting module features adding up to a total of 10 story points, as shown in Table I.

Sprint No.	Sprint Description	Story Point No.	Story Point Description
1	Front End	1	Structure of front end system
		2	Design of front end system
		3	Navigation through front end system
2	Authentication	4	Registration feature
		5	Login feature
		6	Log out feature
3	Course Selection	7	Course Selection feature
		8	Course Viewing feature
4	Payment	9	Payment Status feature
		10	Make Payment feature

In the first iteration, we used Agile methodology by assigning the team members to work separately on the development, testing and delivery sectors over a five-day period. We have catalogued the observed data on performance and progression in the form of a series of Burndown Charts plotting actual remaining story points and estimated remaining story points against the number of days passed. The Burndown Charts of this iteration have tracked the performance of the development, testing and delivery sectors using Agile methodology.

For the next iteration we switched to DevOps Enabled Agile methodology with cross-functional team members working together over a five-day period. This led to continuous collaboration and cooperation among the team members, allowing for the development, testing and delivery sectors to become coordinated without having to fulfill the high skill requirements of fully transitioning to DevOps. We have observed the performance of this methodology by generating another series of Burndown Charts tracking the performance of the development, testing and delivery sectors using DevOps Enabled Agile methodology.

Thus, the implementation of our pilot project has been completed using both Agile methodology and DevOps Enabled Agile methodology with the Burndown Charts generated for both iterations allowing for in-depth comparisons of their performances to determine that DevOps Enabled Agile is the superior methodology in terms of project progression. The findings and discussion of our methodology have been presented and further discussed in the next section of our study.

V. FINDINGS AND DISCUSSION

The results of our pilot project have been observed and catalogued in the form of two separate series of Burndown Charts that denote the performance of Agile methodology and DevOps Enabled Agile methodology for the development and

delivery of a simple pilot project by a small team of inexperienced developers.

The Burndown Charts for the first iteration (Agile) are given below:
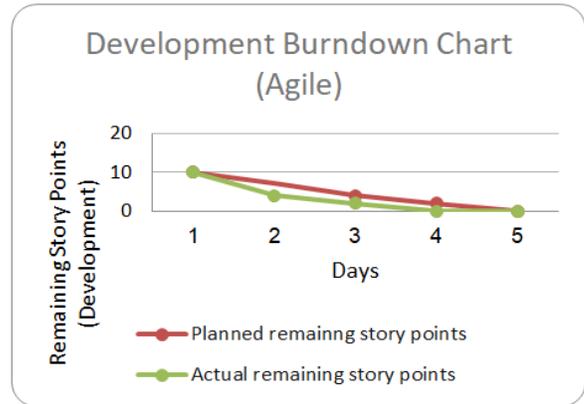
Fig. 3.   Development Burndown Chart (Agile).
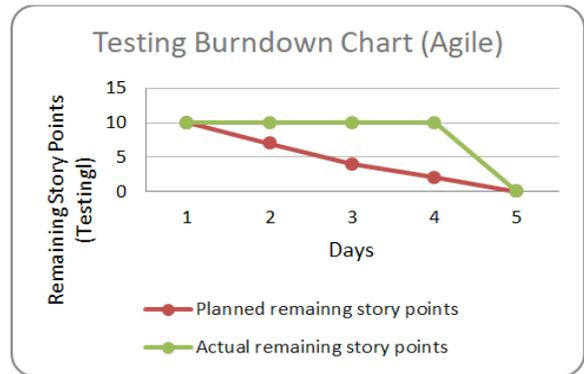
Fig. 4.   Testing Burndown Chart (Agile).
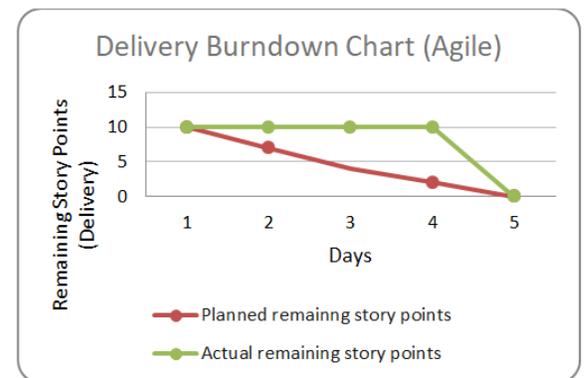
Fig. 5.   Delivery Burndown Chart (Agile).

In the first chart of this iteration shown in Fig. 3, we can see that the number of actual remaining story points consistently stayed below the number of estimated remaining story points throughout the project and development reached completion before the estimated deadline on day five, meaning that the development sector performed better than estimated, progressing faster and finishing earlier than planned.

In the second and third charts of this iteration shown in Fig. 4 and Fig. 5 respectively, we can see that the actual remaining story points for testing and delivery have not decreased at all until the last day. This means that testing and delivery sectors were unable to make any progress before the completion of the development sector on day four, which shows that the observed performance is not up to estimated standards.

Therefore, Agile methodology has not performed as well as estimated due to the lack of consistent progression in the testing and delivery sectors. Our proposed reasoning behind this is that the lack of cross-functionality or collaboration between team members lead to the testing and delivery sectors heavily depending on the completion of the development sector, resulting in overall project progression lagging behind planned estimates until the end of the project timeline.

The Burndown Charts for the next iteration of the project using DevOps Enabled Agile are illustrated below:
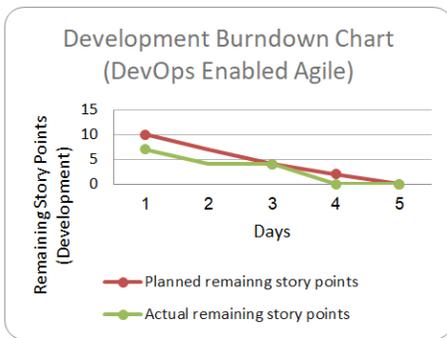
Fig. 6.   Development Burndown Chart (DevOps Enabled Agile).
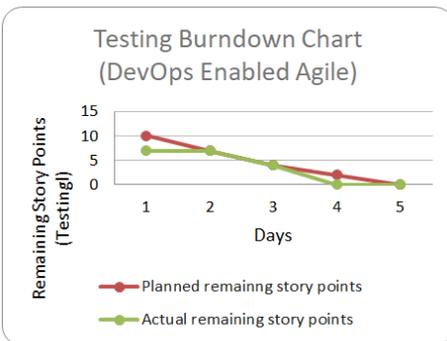
Fig. 7.   Testing Burndown Chart (DevOps Enabled Agile).
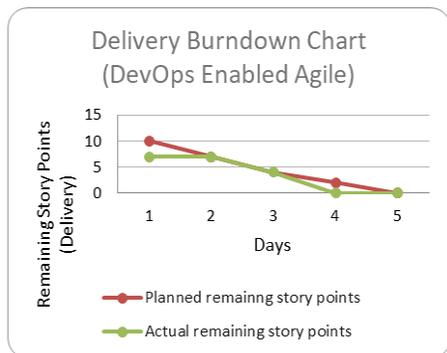
Fig. 8.   Delivery Burndown Chart (DevOps Enabled Agile).

In the first chart of this iteration shown in Fig. 6, we can see that the development sector performed better than estimated similar to the previous iteration, showing faster progression throughout the project and reaching completion before the estimated deadline.

In the second and third charts of this iteration which are shown in Fig. 7 and Fig. 8 respectively, we can see that the actual remaining story points for testing and delivery have mostly remained below the estimated remaining story points and reached completion earlier than planned. This means that the testing and delivery sectors have performed better than planned estimates, leading to consistent and efficient progression of project goals and the completion of the testing and delivery sectors before the estimated deadline.

By comparing this data with the data gained from testing Agile methodology in the previous iteration, we can conclude that our proposed methodology, DevOps Enabled Agile methodology has performed far better in terms of consistent project progression and completion of project goals. DevOps Enabled Agile methodology utilized the cross-functionality aspect of DevOps to ensure that all three sectors would progress in a coordinated way leading to all three sectors gaining significant progress throughout the project timeline as opposed to Agile where testing and delivery could not progress until development reached completion. Therefore, DevOps Enabled Agile has been validated as a superior software development methodology when compared to Agile.

VI.   CONCLUSION AND FUTURE WORK

Our research has been focused on discovering the different gaps of the existing DevOps and Agile Methodologies and combining them to devise a new hybrid methodology, DevOps Enabled Agile, designed to bridge these gaps. From our findings, we have verified the effectiveness and integrity of DevOps Enabled Agile methodology in fulfilling the gaps of Agile methodology while also avoiding the drawbacks of fully transitioning to DevOps methodology and thus we have validated the accomplishment of our research objectives.

The limitation of our research is that we have only tested our proposed DevOps Enabled Agile methodology on a small scale, experimental project. Thus, its effectiveness has not been verified for large scale industrial software development projects. Additionally, our methodology is more focused on the cross-functionality and collaboration principles of DevOps methodology without utilizing any advanced DevOps tools. Application of DevOps Enabled Agile using DevOps tools has been left unexplored as a prospect for future research.

In the near future, we are planning on further optimizing and enhancing DevOps Enabled Agile methodology to make it suitable to be used universally for software development in both small and large scale projects. Additionally, we plan to test the application of DevOps Enabled Agile using different tools and techniques that have not been used in our experiment in order to verify whether this methodology is able to effectively adapt to them. This concludes our current study with a brief description of our future plans for further research.

REFERENCES

[1] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, Jon Kern, Brian Marick, Robert C. Martin, Steve Mellor, Ken Schwaber, Jeff Sutherland and Dave Thomas, "Manifesto for Agile Software Development", Agile Alliance, 2001.

[2] M. K. Alqudah, R. Razali and M. K. Alqudah, "Agile Methods Selection Model: A Grounded Theory Study" International Journal of Advanced Computer Science and Applications (IJACSA), Volume 10, No. 7, 2019.

[3] Muhammad Asaad Subih, Babur Hayat Malik, Imran Mazhar, Izaz-ul-Hassan, Usman Sabir, Tamoor Wakeel, Wajid Ali, Amina Yousaf, Bilal-bin-Ijaz, Hadiqa Nawaz and Muhammad Suleman, "Comparison of Agile Method and Scrum Method with Software Quality Affecting Factors" International Journal of Advanced Computer Science and Applications(IJACSA), Volume 10, No. 5, 2019.

[4] R. Shankarmani, R. Pawar, S. S. Mantha and V. Babu, "Agile Methodology Adoption: Benefits and Constraints", International Journal of Computer Applications, Volume 58, 2012.

[5] M. A. Bjørni and S. Haugen, "Challenges with Agile in a System Development Department: A Case Study", IEEE Software, Volume 35, Issue 1, 2019. pp. 78 – 83.

[6] F. Erich, C. Amrit, and M. Daneva, "A Mapping Study on Cooperation between Information System Development and Operations", International Conference on Product-Focused Software Process Improvement, 2014. pp. 277–280.

[7] C. Ebert, G. Gallardo, J. Hernantes, and N. Serrano, "DevOps", IEEE Software, Volume 33, Issue 3, 2016. pp. 94 – 100.

[8] A. Hemon, B. Lyonnet, F. Rowe and B. Fitzgerald, "Conceptualizing the Transition from Agile to DevOps: A Maturity Model for a Smarter IS Function", IFIP International Federation for Information Processing, 2019. pp. 209–223.

[9] R. T. Yarlagadda, "How Public Sectors Can Adopt the DevOps Practices to Enhance the System", International Journal of Emerging Technologies and Innovative Research, ISSN, 2018. pp. 82-91.

[10] A. Hemon, B. Lyonnet, F. Rowe and B. Fitzgerald, "From Agile to DevOps: Smart Skills and Collaborations", Information Systems Frontiers, Volume 22, 2020.

[11] F. K. Kagai, "Adapting Agile DevOps for Strategic Information Systems Development", TechRxiv Preprint Server, Version 2, 2019.

[12] T. Cardoso, R. Chanin, A. R. Santos and A. Sales, "Combining Agile and DevOps to Improve Students' Tech and Non-tech Skills", CSEDU, 2021. pp. 299-306.

[13] N. Govil, M. Saurakhia, P. Agnihotri, S. Shukla and S. Agarwal, "Analyzing the Behaviour of Applying Agile Methodologies & DevOps Culture in e-Commerce Web Application", 2020 4th International Conference on Trends in Electronics and Informatics, 2020. pp. 899-902.

[14] B. Snyder and B. Curtis, "Using Analytics to Guide Improvement during an Agile–DevOps Transformation", 2017.

[15] G. Y. Koi-Akrofi, J. Koi-Akrofi and H. A. Matey, "Understanding the Characteristics, Benefits and Challenges of Agile it Project Management: A Literature Based Perspective", arXiv e-Print Archive, Volume 10, 2019. pp. 25-44.

[16] A. B. Farid, Y. M. Helmy and M/ M/ Bahloul, "Enhancing Lean Software Development by using Devops Practices" International Journal of Advanced Computer Science and Applications (IJACSA), Volume 8, No. 7, 2017.

[17] L. Banica, M. Radulescu, D. Rosca and A. Hagiu, "Is DevOps another Project Management Methodology?", Informatica Economica, Volume 21, 2017. pp. 39-51.

[18] A. Wiedemann, M. Wiesche, H. Gewald and H. Krcmar, "Understanding How DevOps Aligns Development and Operations: A Tripartite Model of Intra-IT Alignment", European Journal of Information Systems, 2020. pp. 1-16.

[19] K. Kuusinen, V. Balakumar, S. C. Jepsen, S. H. Larsen, T. A. Lemqvist, A. Muric, A. Ø. Nielsen, and O. Vestergaard, "A Large Agile Organization on its Journey towards DevOps", 2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 2018. pp. 60-63.

[20] S. M. Mohammad, "DevOps automation and Agile Methodology", International Journal of Creative Research Thoughts (IJCRT), Volume 5, Issue 3, 2017. pp. 946-949.

Issues in Requirements Specification in Malaysia's Public Sector: An Evidence from a Semi-Structured Survey and a Static Analysis

Mohd Firdaus Zahrin[1], Mohd Hafeez Osman[2]*, Alfian Abdul Halin[3], Sa'adah Hassan[4], Azlena Haron[5]
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang, Malaysia[1, 2, 3, 4]
National Administrative Institute, Public Service Department of Malaysia[5]

Abstract—**Requirement specifications (RS) are essential and fundamental artefacts in system development. RS is the primary reference in software development and is commonly written in natural language. Bad requirement quality, such as requirement smells, may lead to project delay, cost overrun, and failure. Focusing on requirement quality in the Malaysian government, this paper investigates the methods for preparing Malay RS and personnel competencies to identify the root cause of this issue. We conducted semi-structured interviews that involved 17 respondents from eight critical Malaysian public sector agencies. This study found that ambiguity, incompleteness, and inconsistency are the top three requirement smells that cause project delays and failures. Furthermore, based on our static analysis, we collected the initial Malay RS documents from various Malaysian public sector agencies; we found that 30% of the RS were ambiguous. Our analysis also found that respondents with more than 10 years of experience could manually identify the smells in RS. Most respondents chose the Public Sector Application Systems Engineering (KRISA) handbook as a guideline for preparing Malay RS documents. Respondents acknowledged a correlation between the quality of RS and project delays and failures.**

Keywords—Ambiguity; requirements engineering; requirement smell; requirement specification; semi-structured interview

I. INTRODUCTION

Software Requirement Specification (SRS) is a document that details the behaviour of a system. It describes the functionality and non-functionality of the software to meet the requirements of all relevant stakeholders (business, users, and software). RS impacts entire software development project stages [1] as it may specify the Business Requirements Specification (BRS), User Requirements Specification (URS), and Software Requirements Specification (SRS) that define the expectation of the stakeholder of a system to be developed. RS is commonly expressed in Natural language [2], which is exposed to requirement smell [3] such as ambiguity, inconsistency, incompleteness, etc. [4]. The requirement smells impact the requirements' quality which may lead to low product quality.

In Malaysia, the public sector has funded a massive amount of money to digitise its systems. A lot of systems have been developed to realise the e-government initiative. However, from our initial observation, some system development projects faced problems such as project delays

and low-quality product quality. We may assume that RS can be one of the causes of this problem. Based on a survey, 50% of requirement engineers were unaware of ambiguous Malay RS [5]. Zahid et al. [6] have highlighted those smells in RS cause roughly 70% of the project to fail. A lack of precise RS is one of the causes of project failure [7]. Unintentional ambiguity in natural language requirements leads to diverse implementations later in software development [8]. Based on Sommerville [9], "Clients of systems often struggle to translate their desires into requirements," resulting in smells and general terms in the requirements [4]. According to Iqbal et al. [10], "Ambiguous and generalised RS add significantly to project time and cost." Quality RS is determined by comprehending and managing requirements correctly [11]. Furthermore, Rios et al. [12] mentioned that based on survey results, 53% of respondents (software developers) reported that document debt is associated with requirement issues. These previous studies show that requirement smells can cause project delays or failures.

These issues motivate us to investigate the requirement smells and other problems in Malay RS. We focus on Malaysian government projects since most of the RS in the Malaysian government usually use the Malay language.

This paper aims to investigate the preparation of Malay RS and identify the issues that can result in project delays and project failures (incomplete or unfinished). This paper focus on four (4) perspectives (in four (4) research questions (RQ)): The Human factor (RQ1), the Communication factor (RQ2), the Process and Procedure (RQ3), and the Issues in RS (RQ4). We formulate the RQ (as shown in Table I) based on the Ab Aziz [13] guidelines.

This study contributes to the following:

1) We discover factors influencing the requirement smells in Malay RS.

2) We reveal the relation between requirement smells detection and software developer experience.

The remaining sections of this work are organised as follows. The Related Works section describes the relevant studies. The Methodology section reports the approach used to conduct the research. The Findings section explains the result. We discuss the findings in the Discussion section, and the Conclusion section summarises the paper.

---

*Corresponding Author.

TABLE I.   RESEARCH QUESTIONS AND RESEARCH OBJECTIVES

Problem Statement	Research Question (RQ)	Research Objective (RO)
The requirement smells in Malay RS contribute to project delays and project failures.	RQ1: How competent are Malaysian public sector requirement engineers and software developers?	RO1: To investigate the competency level of requirement engineers and software developers in the Malaysian public sector.
	RQ2: What is the most used language in RS?	RO2: To identify the most language used in RS.
	RQ3: How is RS prepared and verified?	RO3: To identify the methods used to prepare and verify the RS.
	RQ4: What are the issues in RS?	RO4: To explore and identify the RS smells.

II. RELATED WORKS

We explored credible sources and found articles that are close to our study are the following:

A. Survey on Agile Requirements Engineering (RE) Practices

Barata et al. [14] studied the agile requirements in practice that focused on how professionals view the significance of requirements in an Agile methodologies-based software development project. A survey of 46 Brazilian software development experts asked about methods for collecting and expressing Agile requirements and their features, benefits, and challenges. The authors studied respondents' experience, viewpoints on collection methods, and Agile requirements specifications. On the contrary, we focused on some factors, i.e., investigating the issues in RS, methods for preparing the RS, RS verification and validation approaches, and requirement engineers' competencies.

B. Survey on Requirement Smells Among IT Practitioners

Lenarduzzi et al. [15] surveyed requirement engineers to understand the theoretical and practical perceptions of the harmfulness of requirement smells and to compare these perceptions with one another. A precise and validated approach for reducing issues during requirement elicitation is proposed to avoid introducing requirement smells. The authors focused the survey among requirement engineers on requirement elicitation processes to prevent the requirement smells. In contrast, our survey involved requirement engineers and software developers regarding the smells commonly found in requirements documents that could delay and fail software development projects.

Rios et al. [12] surveyed to obtain feedback from a software engineer on the factors leading to documentation debt, the effects of this problem, and potential solutions. The term document debt (DD) describes issues with software project documentation, namely the search for missing, inconsistent, out-of-date, or inadequate documentation [16]. These surveys (questionnaires and an interview-based case study) involved 39 practitioners from replications in Brazil, Chile, Colombia, and the United States. The fourth research question, "Which phase of the software development life cycle is most affected by the presence of DD?" reveals a substantial correlation between DD and requirements problems. In their

examples of DD, around 53% of participants highlighted requirements issues. Examples include: "Lack of clarity and precision in the formulation of requirements" and "Needing to construct unspecified code because the requirement was not addressed in the documentation." The researchers focused on identifying the factors that led to DD based on the software development life cycle. In the way of comparison, we focused on the requirements phase. We surveyed the strategies in preparing Malay RS to determine the cause of requirement smells based on IEEE [4] quality attributes.

Within the software development life cycle context, Ahmad et al. [17] published the results of a survey that aimed to ascertain the relative importance of software requirements defects. Using questionnaires, the authors surveyed Malaysian IT professionals from diverse businesses (public, private, and software houses) and job titles (business analyst, system analyst, software engineer, etc.). According to the result, the requirement has defects, such as missing, incorrect, inconsistent, ambiguous, etc. Ahmad et al. [17] concentrated on the survey employing instruments such as questionnaires. In contrast, we conducted a semi-structured interview to investigate the smells in Malay RS documents in Malaysia's public sector agencies. Ahmad et al. [17] did not mention the language of the requirements. We assumed the authors selected the English requirements for their study. In this paper, we highlighted the most significant requirement smells.

C. Survey on Formalising System Requirements and Validation

Mokos and Katsaros [18] surveyed advancement in formalising and early validating system requirements. Several industrial research projects have provided valuable experience in pattern languages and formal property derivation. RS can use domain ontologies to identify missing information, inconsistencies, and under-specification. Our study differs because the author did not specify the survey method, whereas we used the semi-structured interview method. Mokos and Katsaros [18] also reviewed design paradigms not covered by our research.

D. Survey on Industrial IT Professionals' Awareness of Malay SRS

Haron and Abdul Ghani [5] surveyed IT professionals about their perception of ambiguity in the Malay SRS. The survey shows that IT professionals, especially in SRS, tend to overlook ambiguity. More than 50% of the respondents were unaware of the occurrence. The need for a tool to help solve the problem is undeniable. The difference with our study is; that the author sent a set of Malay SRS to IT professionals to assess their understanding of the functional specification. Meanwhile, we used Likert-scale and open-ended questions to gather software developers' understanding (semi-structured interview).

E. Survey on RE in Practice

Aguilar et al. [19] investigated the impact of RE practice in 16 small-sized software firms through a survey in Sinaloa, Mexico. The degree of relationship between each variable was determined using Pearson correlation analysis. The identified variables, i.e., company location, the scope of coverage,

number of workers, etc. The survey found no strong correlation between the seven variables analysed. The significant issues to address; are i) Ineffective client-software development team communication can harm the RS; ii) Ad-hoc RE practices that are hard to track and possibly lost (untraceable) because respondents think RE is not essential for small firms. Our study differs; Aguilar et al. [19] concentrated on issues related to RE practice in the organisation through questionnaires, while our research focused on smells and issues in RS obtained from semi-structured interviews.

Ilyas et al. [20] surveyed Pakistani software companies' current requirement process practices for identifying issues. 10% of companies are unaware of the value of standards in product development. The most critical activity in software development is gathering system requirements. This step is crucial to understanding the user's needs. Our study differs because the author used questionnaire methods, and Ilyas et al. [20] focused on organisational issues. Çamoğlu and Kandemir [21] investigated the software RE processes in Turkish firms (energy, finance, and telecommunication sector). This study identified common issues and problems in business analysis. Insufficiently defined requirements and demands are the leading cause of process failure in all sectors. There is no standardisation of RS documentation. The most commonly used RS documentation tools are Microsoft Word and Excel. The authors focused on high-level RE issues while we concentrated on RS issues. Çamoğlu and Kandemir [21] collected data from many participants via questionnaires, but we interviewed a few for more in-depth feedback.

F. Survey on Techniques Suggested by Industrial Standards

Jarzębowicz and Połocka [22] surveyed requirements documentation techniques in various software project contexts in Poland IT firms. The survey asked 42 Polish IT professionals to choose strategies for multiple projects. After the survey, two interviews with business analysts were conducted to interpret the results. Our study differs because Jarzębowicz and Połocka [22] focused on RS and documentation techniques while we concentrated on problems and requirement smells in RS. The authors used questionnaires and interviews, but we used semi-structured interviews.

G. Quality Assurance in RE

Noorin and Sirshar [23] organised a survey on quality assurance in RE. The authors focused on analysing quality parameters that ensure the requirements are met. The authors discovered that models like LaQuSo Software Product Certification Model, Neural Network, Case-Based Reasoning, and others are used to check other quality parameters like correctness and completeness. In addition, analysing quality characteristics improve requirement quality. Our study focused on RS issues gathered through semi-structured interviews, while the authors surveyed models and quality attributes to strengthen requirements.

H. Factors of Projects Failure in Terms of RE Processes

Hussain and Mkpojiogu [24] investigated the success or failure factors of software development projects related to RE processes using secondary data analysis. As can be seen, poor requirements processes lead to software project challenges and failures. However, a reasonable requirements collection and process contributes to software project success. The authors advise software project planners, engineers, and managers to include adequate RE processes in all software development projects to avoid failures and challenges. It differs from our study; the authors concentrated on causes related to RE processes that influence the software development project failure or success. In contrast, we focused on requirement smells contributing to project delays and loss.

III. METHODOLOGY

This study is conducted to understand and observe RS preparation in the Malaysian public sector. We followed Ab Aziz's [13] guidelines to perform the study through the semi-structured interview. Overall, there are nine (9) steps to conduct the study, i.e., identify respondents, develop interview questions, conduct pre-interview, send invitations to respondents, interview, collect requirement documents, analyse interview results, analyse requirement documents, and conclude the findings. Fig. 1 illustrates the research methodology.
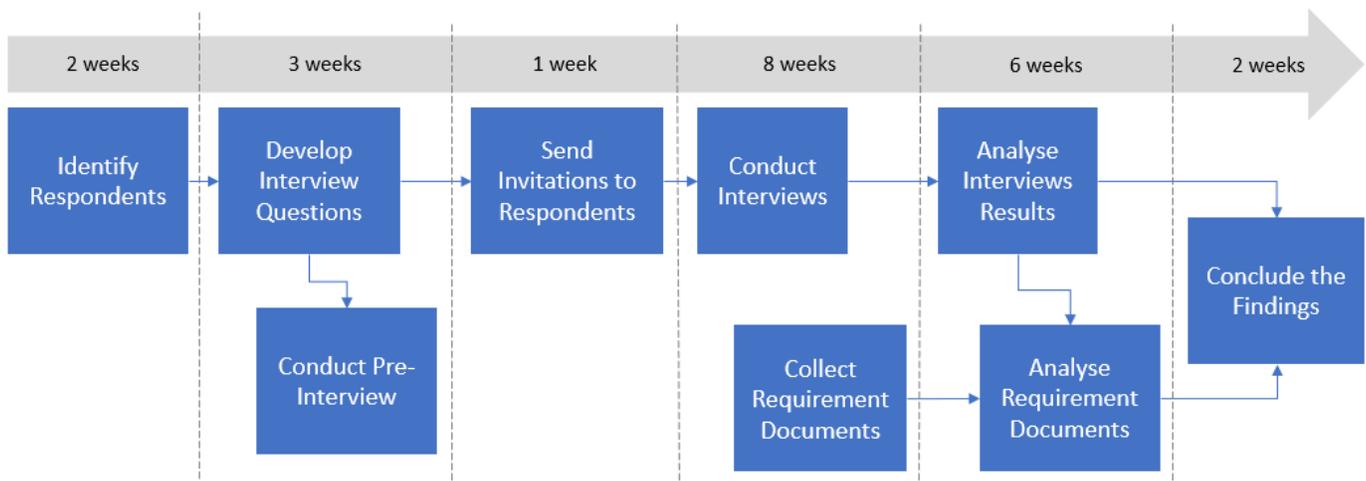
Fig. 1. Methodology of the Study.

A. Identify Respondents

First, we emailed selected agencies to nominate 15 respondents based on the criteria specified in Table II. Fortunately, we received feedback from various critical public sector agencies involving 17 respondents.

TABLE II.   RESPONDENTS' CRITERIA SELECTION

Roles:	Requirement Engineer	Software Developer
The critical domain of Government agencies	Yes	
Expertise	Attended any RS course	N/A
Experience	At least two (2) years	
Involve in RS	Preparing the RS	Developing the software based on the RS

B. Develop Interview Questions

Indeed, questions are listed as a guide to ensure the smooth running of the interview. The developed semi-structured interview questions are based on RQs in the Introduction section. There are 16 main questions; seven open-ended questions, two Likert-scale, and six follow-up questions. An expert has validated these questions. The details of the interview questions are in Table III.

TABLE III.   INTERVIEW QUESTIONS

	Question
RQ1 - Identify the competencies	
1)	How many years of service?
2)	How many years of service as a requirement engineer/ software developer?
3)	Have you attended any course related to preparing the requirements specification? Have you ever attended a professional course such as CPRE/ academic at the University (Requirement Engineering subject)?
RQ1 - Level of understanding of the requirements specification (software developer perspective)	
4)	Do you understand each of the requirement specifications? If not, why?
5)	Overall, state your understanding of the requirements specification used as a reference in developing the system. (1 - Do not understand to 5 - Understand) – Likert scales
RQ2 - The most used language in requirements specification	
6)	What are the languages used in preparing the requirements specification?
7)	How much percentage of requirements specifications are prepared in specified languages?
8)	Is there any policy for using a specific language for the requirements document?
9)	What is your preferred language for requirements specifications? And why?
RQ3 - The methods used to prepare and verify the requirements specification	
10)	What are the templates used in preparing the requirements specification? What are these templates used for?
11)	Is there any verification for the requirements specification? If not, why? If yes, how is the verification conducted? What are the verifications quality attributes?

12)	What are the tools used for requirements verification?
13)	Do you need a tool to check the quality of the requirements specification? (e.g., Checking the ambiguity, inconsistency, and incompleteness in the requirements specification) If yes, what are the specific features that you require? (e.g., syntax-based checking, template-based checking)
RQ4 - Identify the issues in requirements specification (software developer perspective)	
14)	What are the issues that frequently occur in understanding the requirements specification? How do you solve the problems?
15)	How much do you agree that the problem in the requirements causes project delays? (1 - Strongly disagree to 5 - Strongly agree) – Likert scales
16)	Choose the top three (3) smells commonly found in the RS that may contribute to project delays/ failures. ***Quality Attributes** / **Description** **Unambiguity** - The requirement is expressed so that it can be interpreted only one way. The requirement is defined and easy to understand. **Appropriate** - The specific purpose and detail of the requirement are suitable to the degree of the body to which it refers (level of abstraction appropriate to the entity's class). **Complete** - The requirement adequately briefs the necessary capability, characteristic, constraint or quality factor in meeting the entity's need without needing other information to understand the requirement. **Verifiable** - The requirement is structured and worded so that its realisation can be proven (verified) to the customer's satisfaction at the requirements level. Verifiability is enhanced when the requirement is measurable. **Necessary** - The requirement defines an essential capability, characteristic, constraint or quality factor. **Correct** - The requirement accurately represents the entity's need from which it was transformed. **Singular** - The requirement states a single capability, characteristic, constraint or quality factor. **Feasible** - The requirement can be realised within system constraints (e.g., cost, schedule, technical) with acceptable risk. * IEEE [4] quality attributes

C. Conduct Pre-Interview

Before conducting semi-structured interviews, pre-interviews were performed to validate the questions developed. Two experienced system analysts from the Malaysian public sector agency (selected based on the criteria in Table II) participated in the preliminary interviews.

D. Send Invitations to Respondents

Afterwards, invitations to respondents are made via email and attached to an invitation letter. When the respondents agree to participate, the date and time are set based on the respondents' availability.

E. Conduct Interviews

Semi-structured interviews are conducted via online video communication for one hour each. This interview involves eight critical agencies. Due to confidentiality, we could not name the agencies. Table IV lists the study's respondents.

TABLE IV.   LIST OF RESPONDENTS

Respondent (R)	Position	Grade	Working Experience	Experience with RS
R1	REng	44	11 years	9 years
R2	REng	48	18 years	3 years
R3	REng	44	14 years	10 years
R4	REng	41	9 years	6 years
R5	REng	44	14 years	6 years
R6	REng	44	20 years	10 years
R7	REng	48	17 years	12 years
R8	REng	44	11 years	6 years
R9	REng	44	13 years	10 years
R10	SD	29	8 years	8 years
R11	SD	29	12 years	10 years
R12	SD	32	13 years	13 years
R13	SD	32	15 years	13 years
R14	SD	29	11 years	10 years
R15	SD	29	6 years	2 years
R16	SD	32	18 years	18 years
R17	SD	32	13 years	2 years

Requirement Engineer (REng), Software Developer (SD)

F. Analyse Interview Results

This paper employed data visualisation and descriptive-analytic techniques by Regnell et al. [25] to analyse and synthesise the data collected following the RQs presented in the Introduction section.

G. Collect Requirement Documents

We gathered the Malay RS documents to confirm the existence of the issues raised by respondents. We officially request the Malay RS documents from 10 critical public sector agencies (11 domains). Due to confidentiality and requested by respective agencies, we could not reveal the agencies. After receiving the agencies' approval, we collected 18 documents, including the 12 BRS/URS and six SRS. These RS documents were sent to us via email in Microsoft Word and Adobe PDF format. Other than that, we received some RS documents in hard copies. Table V lists the domain of the collected Malay RS documents.

H. Analyse Requirement Documents

We identify and eliminate the non-RS. We chose and extracted the textual RS into SQL format (repository). Next, we cleaned the textual-based RS using the Rapid Miner [26]. The data cleaning process comprises identifying, addressing missing values, and manually validating Malay spelling. We manually labelled the class based on the most significant

requirement smell. Two Malay requirement specialists confirmed the class label. This paper employed the statistical descriptive-analysis method by Christopher [27] to analyse the percentage of the most significant requirement smell mentioned by respondents.

TABLE V.   DOMAIN OF COLLECTED MALAY RS

Domain	No. of RS
Business	65
Career	6
Customer Services	137
Funds/ Finance	892
Human Resources	76
General Administration	78
Miscellaneous	21
Survey	107
Trade	89
Training	49
Welfare	15
Total RS:	1535

I. Conclude the Findings

We conclude the results in the Findings section and discuss the insights in the Discussion section.

IV. FINDINGS

This section presents the key findings that answered the RQs in the Introduction section through the semi-structured interviews. Section A corresponds to the competency level of respondents (RQ1). Section B reports the languages employed by RS (RQ2). Section C corresponds to methods used to prepare and verify the RS (RQ3). Section D presents the issues in RS (RQ4).

A. Human Factor (RQ1)

1) Requirement engineers: This study investigates the competency level of requirement engineers in Malaysia's public sector. We found the respondent's minimum and maximum years of service were nine and 20. The minimum and maximum years of involvement in requirement engineering were between three and 12 years. Fig. 2 illustrates the years served as a requirement engineer. 78% of requirement engineers have attended various courses related to RS. The courses, i.e., Certified Professional Requirement Engineer (CPRE) [28], took a subject at university, boot camp, and internal courses organised by respective agencies. Hence, most of the requirement engineers in Malaysia's public sector are competent and well-equipped with training. This study also shows that we have selected trained and experienced requirement engineers.

2) Software developers: We studied the competency level of software developers in Malaysia's public sector. As a result, the respondent's minimum and maximum years of

service were six and 18 years of experience. Fig. 3 shows the years served as a software developer.

3) Level of Understanding of the RS: This study determined the understanding of the SRS provided by the software engineer from the software developer's standpoint. Therefore, we require experienced software developers with a good understanding of SRS for software development. Fig. 4 illustrates that 86% of the respondents understood the RS given to them by the requirement engineer during the software development. However, 14% of respondents sometimes did not understand RS provided by their requirement engineer. Several respondents mentioned using flowcharts and mock interfaces, not RS, as a reference for system development. Fig. 5 shows that 57% of respondents understood the overall RS used as a reference in developing the system, while 43% of the respondents sometimes did not understand. The top three typical issues found in RS were ambiguity, incomplete and incorrect. Other than that, RS was constantly changing specifically for in-house development. This condition makes it difficult for software developers to expand the system and contributes to project delays. Most of the resolution of these issues is meeting with the project team and discussing the impact analysis and feasibility of the project with requirement engineers and clients.

Fig. 4. Level Understanding Each RS.
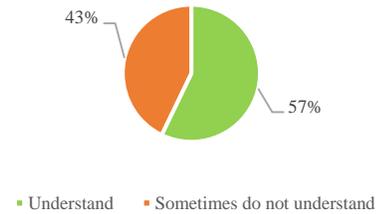
Fig. 5. Level Understanding Overall RS.

B. Communication Factor (RQ2)

This paper only focused on the language used in preparing RS Malaysia's public sector for the communication factor. We would investigate the most used prominent language. Fig. 6 shows 84% Malay and 16% English used to prepare the RS. Fig. 7 shows the policy to use a specific language for the RS. 78% did not have any policy, 11% used the internal memo, and 11% used the Service Circular No.1 / 2020 Empowerment of the National Language in the Public Service [29]. The Service Circular was to encourage writing in Malay for formal documents, including RS. Most respondents used Malay to prepare the RS, specifically for in-house projects. Most Malay RS were prepared for in-house projects because of communication among clients and requirement engineers, understanding of the RS, validation, and verification, and working culture. 44% of respondents used the English RS for outsourcing projects. Some of the outsourcing projects were prepared in English RS because the requirement engineers and developers are foreigners.

Fig. 2. Number of Years Served as Requirement Engineers.

Fig. 3. Number of Years Served as Software Developers.
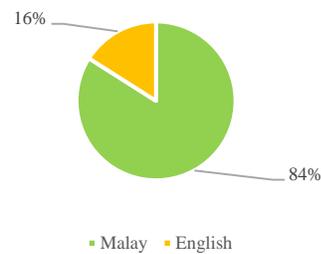
Fig. 6. Most Languages used in RS.

11%

11%

78%

- Do not have policy - Internal memo - Service Circular
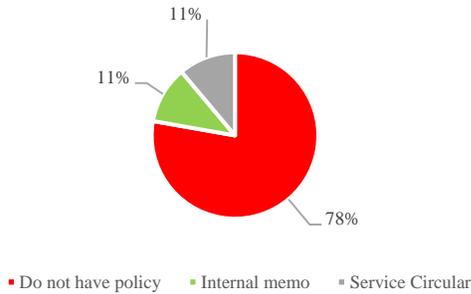
Fig. 7. Policy to use a Language for RS.

C. Process and Procedure (RQ3)

This study identified the methods and approaches used to prepare and verify the RS. Fig. 8 illustrates that 56% of the respondents used KRISA [30] to prepare the RS, and 44% did not have any template. Requirement engineers used only 48% of KRISA content as guidance. There were no formal instructions from the internal agencies or enforcement from the central agency to use KRISA documents. Most respondents took the initiative to implement KRISA as a handbook without any supervisor's orders. The KRISA document was used as a guide to preparing URS, BRS, SRS, and other documents. Fig. 9 shows that 89% of the respondents had verification methods for the RS. The respondents' verification methods, e.g., reviewing and confirming the RS through the committee meeting chaired by the Project Manager and Project Director. Most respondents mentioned that clients or subject matter experts verified the BRS and URS.

44%

56%

- KRISA - No template
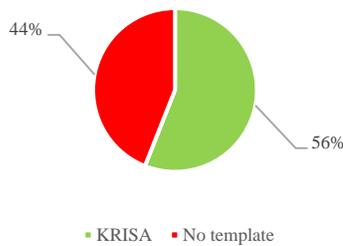
Fig. 8. Template in Preparing RS.

11%

89%

- Have verification method - No verification method
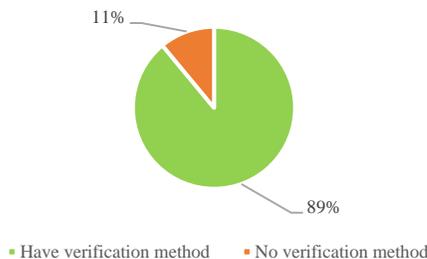
Fig. 9. Verification Method for RS.

22%

78%

- No tool - Have tools
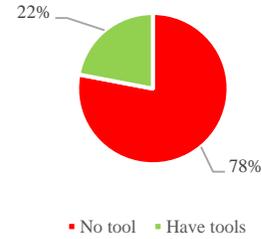
Fig. 10. Tools for RS Verification.

11%

89%

- Need tools - Not need tools
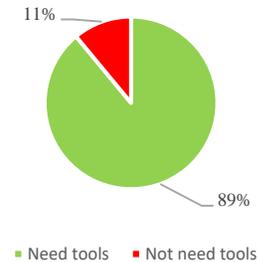
Fig. 11. Need Tools for Verifying the RS.

Meanwhile, the SRS was verified by requirement engineers. 25% of respondents found the requirement smells during verification. Fig. 10 illustrates that 78% of the respondents said no tool was used for RS verification. Verify the RS manually using Microsoft Word and Microsoft Visio tools (22%). Other tools were a mock-up interface, workflow, and flowchart. Fig. 11 depicts that 89% need tools to check RS quality, i.e., ambiguity, incomplete etc. In addition, most respondents suggested the tool could check the syntax and KRISA template-based conformance.

D. Issues in RS (RQ4)

This paper focused on exploring and identifying the smells in RS. Fig. 12 illustrates that 88% of the respondents (software developers) agreed that issues in RS could cause project delays or failures. Ambiguity, incomplete and incorrect were the top three familiar smells found in RS. Since ambiguity was the most significant requirement smell, we analysed the collected documents that may contain ambiguous RS. We found that 30% of the initial gathered Malay requirements were ambiguous.

13%

13%

75%

- Strongly agreed - Agreed - Neutral
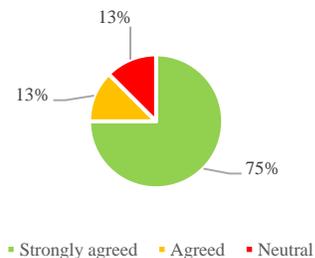
Fig. 12. Issues in RS Could Cause Project Delays or Failures.

V. DISCUSSION

In this section, the findings are further discussed. Also, this section discusses the threats to validity.

A. Software Developers' Experience Versus Requirement Smells

From the perspective of software developers' experience, Fig. 13 shows that 38% of respondents with more than 10 years of experience found the ambiguity as requirement smells commonly found in RS. Fig. 13 suggests that respondents with more than 10 years of experience can identify the requirement smells in RS.

B. Software Developers' Experience Versus Understanding the RS

From the software developers' experience perspective, Fig. 14 illustrates that 29% of respondents with six to 10 years of experience are neutral in understanding the RS. The term neutral means sometimes understanding and sometimes not understanding the RS. The factor of neutral understanding happened caused of the inexperienced requirement engineers in developing the RS. Software developers with more than six years of experience could identify the requirement smells. Therefore, experienced software developers need further clarification from requirement engineers to understand the RS. Meanwhile, 29% of software developers with less than six years of experience understand the overall RS. Software developers with less than six years of experience typically follow the RS without screening the RS's quality.
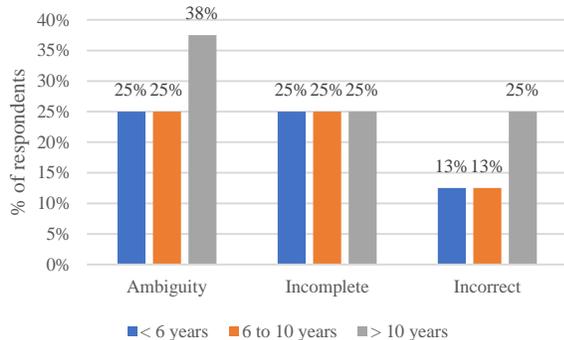
Fig. 13. Software Developers' Experience Vs Requirement Smells Commonly Found in RS.
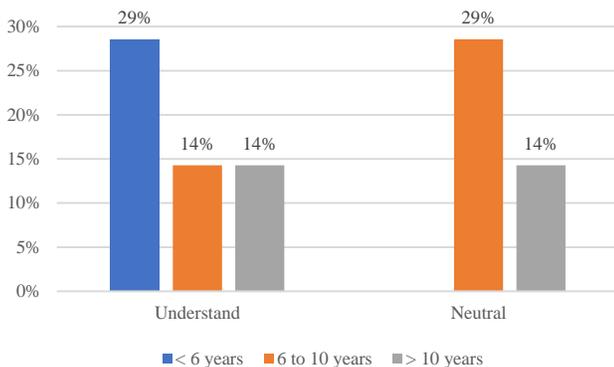
Fig. 14. Software Developers' Experience Vs Understanding the RS.

C. KRISA Handbook

No one instructed the requirement engineer to use any guide to prepare the RS. It is up to the requirement engineer to decide the template for creating the RS until completion. Most requirement engineers used the KRISA as a handbook to prepare the RS because it is written in Malay and easy to understand in their mother tongue. However, the KRISA handbook contains no clear or complete template for constructing the Malay RS.

D. Threats to Validity

During this research, there is a possibility that a few risks developed by accident due to a few circumstances brought on by various kinds of validity, such as internal, external, construct validity, and so on [31].

1) External validity: The respondents were aware of the open-ended nature of the interview questions, but they provided succinct responses nonetheless. Aside from that, respondents did not give the anticipated feedback in their responses. It may impact the low reliability of the data acquired. As a result, we questioned respondents and asked them to elaborate on their short answers.

2) Construct validity: The interview questions were vetted and tested for understandability through a preliminary interview session with some qualifying candidates before the actual interview session (illustrated in Table II).

VI. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

We conducted semi-structured interviews based on RQs in the Introduction section, i.e., RQ1) Human factor, RQ2) Communication factor, RQ3) Process and Procedure, and RQ4) Issues in RS. We found that requirement smells in RS may cause a delay in the RE processes and lead to project failure. Ambiguity is the top smell in Malay RS. Based on a static analysis, 30% of the initial collected Malay RS are ambiguous. These issues can only be identified by experienced software developers, i.e., those with more than 10 years of experience. The participants mentioned that KRISA is the primary reference for preparing RS in the Malaysian public sector in terms of a template. Nevertheless, KRISA does not contain a complete requirement specification structure template compared to ISO IEEE/ISO/IEC 29148-2018, EARS, and RUPPs.

B. Future Works

The government needs automatic tools to detect and improve ambiguous RS based on the respondents' feedback. The respondents who wish to have tools could check the RS's quality based on the KRISA handbook. The automated tools would cover the functional and non-functional RS. Therefore, we would create the aforementioned automated tools. In addition, we are ambitious to design and prepare a Malay RS template.

REFERENCES

[1] J. S. Sinpang, S. Sulaiman, and N. Idris, 'Detecting Ambiguity in Requirements Analysis Using Mamdani Fuzzy Inference,' Journal of Telecommunication, Electronic, and Computer Engineering, vol. 9, no. 3-4 Special Issue, pp. 157–162, 2017.

[2] E. Kamsties, "Understanding ambiguity in requirements engineering, Engineering and Managing Software Requirements," in Engineering and Managing Software Requirements. Berlin: Springer, 2005.

[3] H. Femmer, D. Méndez Fernández, S. Wagner, and S. Eder, 'Rapid quality assurance with Requirements Smells,' Journal of Systems and Software, vol. 123, pp. 190–213, 2015, DOI: 10.1016/j.jss.2016.02.047.

[4] IEEE, 'IEEE/ISO/IEC 29148-2018 - ISO/IEC/IEEE International Standard - Systems and software engineering -- Life cycle processes -- Requirements engineering', IEEE, 2018.

[5] H. Haron and A. A. Abdul Ghani, 'A Survey on Ambiguity Awareness towards Malay System Requirement Specification (SRS) among Industrial IT Practitioners', Procedia Comput Sci, vol. 72, pp. 261–268, 2015, DOI: 10.1016/j.procs.2015.12.139.

[6] A. H. A. Zahid, M. W. Haider, M. S. Farooq, A. Abid, and A. Ali, 'A critical analysis of software failure causes from project management perspectives,' VFAST Transactions on Software Engineering, vol. 6, no. 1, pp. 62–68, 2018.

[7] V. Rodríguez Montequín, J. Villanueva Balsera, S. M. Cousillas Fernández, and F. Ortega Fernández, 'Exploring Project Complexity through Project Failure Factors: Analysis of Cluster Patterns Using Self-Organising Maps', Complexity, vol. 2018, p. 9496731, 2018, DOI: 10.1155/2018/9496731.

[8] F. Ashfaq and I. S. Bajwa, 'Natural language ambiguity resolution by intelligent semantic annotation of software requirements,' Automated Software Engineering, vol. 28, no. 2, Nov. 2021, DOI: 10.1007/s10515-021-00291-0.

[9] I. Sommerville, Software engineering. Pearson Education, 2016.

[10] T. Iqbal, P. Elahidoost, and L. Lúcio, 'A Bird's Eye View on Requirements Engineering and Machine Learning, in 2018 25th Asia-Pacific Software Engineering Conference (APSEC), 2018, pp. 11–20. DOI: 10.1109/APSEC.2018.00015.

[11] C. R. Kavitha and S. M. Thomas, 'Requirement gathering for small projects using agile methods,' IJCA Special Issue on Computational Science-New Dimensions & Perspectives, NCCSE, 2011.

[12] N. Rios et al., 'Hearing the voice of software practitioners on causes, effects, and practices to deal with documentation debt,' in International Working Conference on Requirements Engineering: Foundation for Software Quality, 2020, pp. 55–70.

[13] A. Ab Aziz, 'Model Penerimagunaan Sistem Pengurusan Dokumen Dan Rekod Elektronik Dalam Sektor Awam Di Malaysia', Universiti Kebangsaan Malaysia, Malaysia, 2019.

[14] J. C. Barata, D. Lisboa, L. C. Bastos, and A. Neto, 'Agile requirements engineering practices: a survey in Brazilian software development companies,' arXiv preprint arXiv:2202.12956, 2022.

[15] V. Lenarduzzi, D. Fucci, and D. Mendéz, 'On the perceived harmfulness of requirement smell: An empirical study, in Joint 26th International Conference on Requirements Engineering: Foundation for Software Quality Workshops, Doctoral Symposium, Live Studies Track, and Poster Track, Pisa; Italy, 24 March 2020 through 27 March 2020, 2020, vol. 2584.

[16] C. Seaman and Y. Guo, 'Measuring and monitoring technical debt,' in Advances in Computers, vol. 82, Elsevier, 2011, pp. 25–46.

[17] S. Ahmad, S. A. Asmai, and N. A. Rosmadi, 'A Significant Study of Determining Software Requirements Defects: A Survey. WSEAS Press, 2015.

[18] K. Mokos and P. Katsaros, 'A survey on the formalisation of system requirements and their validation,' Science Direct, vol. 7, p. 100030, 2020, DOI: https://doi.org/10.1016/j.array.2020.100030.

[19] J. A. Aguilar, A. Zaldívar, C. T. Barba, R. Espinosa, S. Misra, and C. E. Zurita, 'A Survey About the Impact of Requirements Engineering Practice in Small-Sized Software Factories in Sinaloa, Mexico,' in ICCSA, 2018.

[20] F. Ilyas, K. Zahra, N. Ambreen, and W. H. Butt, 'A survey on current requirement process practices in software companies and requirement process problems,' in 2016 International Conference on Computational Science and Computational Intelligence (CSCI), 2016, pp. 1280–1285.

[21] K. Çamoğlu and R. Kandemir, 'A Survey of Software Requirements Engineering Practices in Turkey, 2019.

[22] A. Jarzębowicz and K. Połocka, 'Selecting requirements documentation techniques for software projects: A survey study, in 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), 2017, pp. 1189–1198. DOI: 10.15439/2017F387.

[23] U. Noorin and M. Sirshar, 'Quality Assurance in Requirement Engineering,' Global journal of computer science and technology, 2017.

[24] A. Hussain and E. O. C. Mkpojiogu, 'Requirements: Towards an understanding on why software projects fail,' in AIP Conference Proceedings, 2016, vol. 1761, no. 1, p. 020046.

[25] B. Regnell, M. Höst, J. Natt och Dag, P. Beremark, and T. Hjelm, 'Visualisation of agreement and satisfaction in distributed prioritisation of market requirements,' in Proceedings of 6th International Workshop on Requirements Engineering: Foundation for Software Quality, 2000.

[26] RapidMiner, 'RapidMiner tool.' 2006. [Online]. Available: https://rapidminer.com/.

[27] A. N. Christopher, Interpreting and using statistics in psychological research. SAGE Publications, 2016.

[28] K. Pohl and C. Rupp, Requirements Engineering Fundamentals: A Study Guide for the Certified Professional for Requirements Engineering Exam - Foundation Level - IREB Compliant (1st ed.). Rocky Nook, 2011.

[29] Jabatan Perkhidmatan Awam Malaysia, 'Pekeliling Perkhidmatan Awam Bilangan 1 Tahun 2020: Pemerkasaan Penggunaan Bahasa Kebangsaan Dalam Perkhidmatan Awam'. 2020.

[30] Malaysian Administrative Modernisation and Management Planning Unit (MAMPU), Public Sector Application System Engineering Handbook (KRISA) - 2nd Chapter: Business Requirements Specification. MAMPU, 2019.

[31] K. Petersen and C. Gencel, 'Worldviews, research methods, and their relationship to validity in empirical software engineering research, in 2013 joint conference of the 23rd international workshop on software measurement and the 8th international conference on software process and product measurement, 2013, pp. 81–89.

A Guideline for Designing Mobile Applications for Children with Autism within Religious Boundaries

Ajrun Azhim Zamry, Muhammad Haziq Lim Abdullah, Mohd Hafiz Zakaria
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia, Melaka

Abstract—Autism spectrum disorder is a condition related to brain development that impacts how a person perceives and socialises with others, causing problems in social interaction and communication. The disorder also includes limited and repetitive patterns of behavior. Children with autism spectrum disorder (ASD) develop at a different rate and don't necessarily develop skills in the same order as typically developing children. Nowadays, children with autism spectrum disorder (ASD) are having difficulties in gaining religious skills. This is due to the lack of schools that provide special religious education for disabled children. Many technologies have been developed to help children with autism for education. Mobile applications have extensively been used to enhance their daily learning. Researchers are extremely trailing their applications but not many applications are able to meet the requirements and needs of children with autism, especially in the religious context. The lack of religious mobile application guidelines is crucial as a reference for researchers. This paper aims to propose a guideline to design a mobile application for children with autism in religious context. A systematic review of previous literature on mobile application guidelines for autism and religious mobile application guidelines was conducted. This study resulted in two key findings: (1) elements of multimedia consist of text, images and sounds (2) features of application consist of interface, navigation, customisation and interaction. The proposed guidelines are potentially to be used by researchers who are interested in designing religious mobile applications for children with autism.

Keywords—Autism Spectrum Disorder (ASD); guidelines; mobile applications; religion; assistive technology; communication

I. INTRODUCTION

In recent years, the study of children with autism has been on the rise as many have become aware of the importance of helping this group. Mobile application has become the main platform for researchers to develop technologies that can assist them [18]. However, not all applications can work well and can be used by children with autism in improving their daily learning activities. This is due to the lack of guidelines for mobile applications, especially in religious studies. Therefore, this paper will propose a guideline for the development of mobile applications for children with autism in religious context.

Assistive technology (AT) refers to any product, equipment, software program or system that enhances learning, working, and daily living for persons with disabilities. AT has been used in the field of education to overcome barriers to learning [1]. Children with autism appear to learn differently from other individuals and frequently have difficulty with spoken and written communication. For children with autism, computer-assisted training has proven to be more effective than personal instruction [19]. Children with autism were discovered to be more interested in computers than toys, and to prefer computer education to personal instruction [4]. However, the application development for children with autism must follow the guidelines that have been established.

It is crucial to avoid building applications without proper guidance. This is because developers need to ensure children with autism feel comfortable to use applications that will assist their daily routine. The study [3] reported that there is still a lack of providing guidelines for the design of spiritual mobile applications and the need to have the certification and authentication of these digital apps must be considered especially for digital Quran. This paper focuses on two categories of guidelines which are mobile apps guidelines for autism and religious apps guidelines. The proposed guidelines have the potential to be used as a guide for researchers who want to develop mobile applications for autistic children, especially in the field of religion.

Mobile apps were frequently used as assistive technology to assist special groups such as children with autism. Mobile apps are promising to help these children in their daily learning including in religious studies which makes religious activities highly relying on mobile apps these days [6]. Religious mobile apps are becoming widely popular and changing the faith-based landscape across the world [20]. The aim of religious mobile apps is to ensure that religious practices would no longer be limited to a specific time or place. With the help of a mobile app, these religious places can build a faith in community and transfer religious teachings to a younger generation.

The next section presents the literature review, followed by the methodology of this research, findings, discussion and the last section concludes the paper with potential benefits of the proposed guidelines and the next step of this research which to evaluate the proposed guidelines.

II. LITERATURE REVIEW

A. Multimedia Elements (Apps for Autism)

It is necessary for an effective mobile application to have better features for easy to use. The main factor for effective application is multimedia elements. Multimedia elements that are often emphasised in application development are text, images, and audio. The first important element in multimedia applications is text. The study [8] mentioned that the text used

for mobile apps must be clear, simple, and short. Correspondingly, [22] stated that the font used must be clear and large and only used a maximum of two typefaces. Furthermore, [24] suggested that sans serif fonts with 14 font size are suitable for children with autism. On the other hand, [3] mentioned that developers need to avoid writing long paragraphs by using lists and heading titles in paragraphs. In addition, [3] stated that apps for children with autism spectrum disorder should avoid jargon, misspellings, metaphors, abbreviations, and acronyms, and instead utilise terminology, idioms, names, and symbols that are known to the users. This study also demonstrated that those factors are important to prevent users from feeling disturbed and losing focus while using the apps.

Images are also among the multimedia elements that need to be considered in creating apps for autistic children. Each autistic child has different preferences in using apps based on their personal interests included in the selection of images that are suitable for them [30]. It is important to have a suitable size of pictures and images in developing apps for children with autism. The authors [3] suggest allowing images magnification for better visualisation and ensure that the children with autism are able to understand when enlarged. Similarly, this statement was mentioned by [26] in her study that the mobile application must be able to allow users to adjust image size based on their preferences. Meanwhile, [13] stated that the use of photographs for concrete places and people while icons, pictograms and symbols for abstract terms is appropriate. Sharing the same opinion is [8] who said that the picture can be drawings, photographs and symbolic images. He added that the images should be easy to understand, should not fade into the background and should be in a sharp focus. The author in [21] also thinks that images must be identical to real life. For example, use a picture of real food instead of illustrated food. The research in [13] similarly reported that an application should use real images to represent people and places.

Audio also plays an important role in the creation of a successful application. The use of appropriate audio can also attract users [33]. However, audio selection for children with autism must adhere to guidelines so that it does not become a distraction for them. In [24], it was reported that mobile applications for autism should use clear, comprehensible short audio expressions. The app developer must select audio content in a way to motivate and not to distract users. In a similar vein, [3] mentioned that the application must be able to provide audio instructions and subtitles for texts but ensure that it is not the only alternate content representation. The author [5] added that the apps need to avoid the use of disturbing and explosive sounds like sirens and fireworks. In addition, [21] stated that the audio should correspond to the images and must be user-friendly. Each step can also have audio added to it, especially with the voice of someone the children is familiar with. For example, the instructor voice for the apps might be the voice of the teachers or parents.

This concludes that these multimedia elements are essential for developers to emphasise before they develop apps for autism. It should be noted that the use of multimedia elements for children with autism is not similar to normal users. The selection of text, images and even audio should be appropriate and not cause discomfort to them. These guidelines should be followed and adapted by all developers to avoid the development of less effective apps to users.

B. Features of Application (Apps for Autism)

In addition to the element of multimedia, application features for children with autism also play an important role. This is to ensure that children with autism can take full advantage of the applications developed for them. Among the following features are interface, navigation, customization and interaction. As for the interface, most researchers agree that the interface for apps for children with autism must be simple. The study [8] stated that design and structure should be simple, clear and predictable. The research [22] mentioned that an application must be designed for simplicity and few elements on screen. A simple interface only presents the features and content needed for the current task to be performed by the user [3]. In addition, [26] suggested to design an interface that is easy on the eye of users. For example, use simple icons that can be easily understood by the users. Moreover, [22] stated that a mobile application for children with autism should avoid cluttered interfaces. The arrangement of the app interface must be well-organised and easy to use. Similarly, [13] mentioned that an application for children with autism must provide an uncluttered layout with white space between the text in one column. White space helps the user to read faster and will focus on the specific paragraph [29].

The second important feature of the apps for children with autism is navigation. A simple navigation is needed to ensure children with autism can easily understand how to use the application properly. The author [21] mentioned that the application should have simple navigation and easy to use. in [22], author also stated that a mobile app for children with autism should strive for simple and clear navigation. A simplified navigation should be provided on every page [3]. Furthermore, navigation in an application should be consistent and similar in every section [8]. The position and icon used must be similar without making any changes to avoid confusion for children with autism. For example, the icon of a home was used to indicate back to the homepage and the location of the icon was located at the center bottom of each page. The study [21] describes consistent navigation as similar actions and similar buttons located in similar positions. The research in [22] suggested using standard navigation icons that are familiar to children with autism. Similarly, [3] reported that an application must present a global navigation button such as exit, back to home page and help button.

Every child with autism has different preferences and interests. It helps in making them feel comfortable when using the apps. Therefore, customization is needed in each application to assist the children with autism to choose what is best for them. Similarly, [8] and [22] mentioned that every application should allow customization for children with autism. For example, the customization of font type, font size, themes and colours. The developers must provide options to customise information visualisation with images, sound and text according to users' individual preferences [3]. For instance, the user can choose whether to use a small or large

font, loud or slow audio and small or large images. The existence of customisation in this application enables children with autism to feel more confident to use it and attract their interest in using this application.

Furthermore, the feature of application that needs to be taken into account is interaction. An interactive app is more user friendly when you allow users to interact with the app as it creates a connection between them. An interactive app provides users with the freedom to navigate from one window to another comfortably. The author [13] suggested avoiding simultaneous tasks and only one task is allowed in every section. Also, [13] stated to avoid distraction in the apps such as captchas and pop up advertisements. Moreover, [8] mentioned that the application has to make adaptive interaction with users by considering their interaction history, preferences, request, and needs. Further, [3] suggested that an application should have high sensitivity and responsiveness. Touch screen interactions should have the appropriate sensibility and prevent errors in selections and accidental touch in interface elements [25]. Lack of sensitivity will cause children with autism to feel depressed and give up using the apps again. In addition, [24] reported that it is important to provide feedback after learning activities and navigation actions. For instance, the correct text should appear for the correct answer and "try again" for incorrect answers. Prioritising in-app interaction will definitely make children with autism more focused and complete the tasks given in apps [31].

C. Multimedia Elements (Religious Apps)

Guidelines are also needed to build applications related to religion. Although there are some similarities between mobile apps guidelines and religious apps guidelines, there are still some details that need to be carefully scrutinised to distinguish between the two guidelines. In the context of elements of multimedia, emphasis is always given to elements such as text, images and even sounds. The study [1] suggested using suitable font size in developing an application for children with autism. The font size should not be too large or too small where the use of font size should be between 12 and 14 points. In a similar vein, [3] stated that an application must use a large size font type and try to make it adjustable. Medium or boldface type is encouraged. Further, [3] also added to make sure the use of text coincides with the original source. For instance, the Arabic text in the application was written based on the same number of pages in the Holy Quran. This means that if the verse in the Quran contains 10 verses, then the verse in the application must have the same number of verses which is 10 verses.

Next, the elements that need to be emphasised in multimedia are images. The author in [16] mentioned that the dress code of man and woman should be appropriate. The applications are encouraged to use formal attire that are modest to religious sentiments for any character, figure or real person. For example, the compliant driven ('awrah or clothing boundary). In addition, [3] suggested choosing appropriate images for the apps. The images used in apps should be relevant to the content and not display images that are not related to Islamic elements. The study [2] added that the application's selection buttons, icons, menus, and information

should be of excellent quality and resolution to make it easier for viewers to understand and receive clear Islamic material. Furthermore, the final multimedia element that needs to be considered in developing apps for children with autism is sounds. Also [16] suggests that the page be accompanied by an Islamic background sound. The many sorts of Islamic audio sound may help to enhance the site's Islamic spiritual sense. Moreover, [16] also suggested to use audio commonly used in Islam. For example, the sounds of Al-Quran recitation and the sounds of azan. The author [3] suggested providing good recitation audio. Pronunciation and melodious voice are important elements to attract users in learning Islam. For instance, a voice of Qari reciting Al-Quran can help the user to listen to the recitation clearly and can improve their pronunciation correctly. Finally, [3] also mentioned that an application should allow the user to adjust the volume. These customizations are required for the user in determining their preferred volume.

D. Features of Application (Religious Apps)

Apart from the elements of multimedia, features of apps also play an important role in building successful religious apps. Interface, navigation, customization and interaction are the categories that are often mentioned and discussed by previous studies. For interface, the most important thing to consider when developing a religious application is to ensure that the application shows the identity of a specific religion [23]. For example, [16] stated that the application should be based on the elements that best symbolize Islam which are widely recognized by the Islamic community. Similarly, [3] reported that an application should use features that present the Islamic identity like using the mosque icon as a homepage button for the user. Meanwhile, [2] mentioned that in terms of information and distribution, an application layout is characterised by Islamic concepts. For instance, the application needs to display an ustaz icon for male voice selection and ustazah icon for female voice.

Navigation is the second category that is important for religious application development. The research [1] suggested to ensure the user interface navigation structure is simple and straightforward. The buttons should be familiar to the user and predictable. For example, the home icon is used as a homepage button that will navigate the user to the home screen as the safe point to return. Similarly, [2] stated that the user should be able to navigate between screens in the built application at a reasonable cost and with ease. A consistent navigation button is required as it has to be arranged in the same position with the same size and same icon.

Customisation is also involved in ensuring that an application successfully meets the needs of users. The study [2] suggested that for the Islamic features, the application should provide for relevant modification based on the user's preferences. For example, the volume of the audio, the colour of text and the magnification of the screen. Furthermore, [3] stated that the application has to provide fonts, buttons and icons that can be easily resized. The users are able to choose the font size, font type and the colour of the buttons according to their preferences.

The last category to consider in developing a religious mobile application is interaction. Fast response is often mentioned by previous researchers when it comes to interaction. Further, [16] reported that the speed of the application interaction must be consistent without any delay occurring. Similarly, [2] said that the application must provide accurate and fast response to users. Interruption should be avoided while the application is running as it will distract children with autism when they are using the apps. The author [2] added that all functions in the application must work well and be reachable by the user. It is important to facilitate users in obtaining religious knowledge in the application. By using the mobile application, students can obtain information on religious studies more quickly and easily [32]. Moreover, [1] mentioned that the application should avoid using interaction timeouts to memorise Al-Quran using fast-moving objects and animation. Interaction timeout can be a distraction in memorising Al-Quran. It is encouraged to allow customization for moving objects and animation speed as the users are able to control the speed based on their preferences.

In summary, both these mobile apps guidelines are crucial as guides and references for future developers in focusing on children with autism. Each guideline has important features that the developer must adhere to. From the combination of these two categories of guidelines, a new guideline focusing on religious mobile apps for children with autism will be produced at the final stage of this study. This is in line with the goal of the study to fill in the gap where there is a lack of guidelines that focus on religion and children with autism at the same time.

III. Methodology

Systematic Literature Review (SLR) method was used in this study to identify, evaluate and summarize the related guidelines from previous literature. This study follows five steps for systematic literature review as suggested by [34]. The steps are framing questions for a review, identifying relevant work, assessing the quality of studies, summarizing the evidence and interpreting the findings. Meta analysis was conducted to take findings from several studies on the same subject and analyze this using standardized statistical procedures. In meta-analysis patterns and relationship are detected and conclusions are drawn. Meta-analysis is associated with the deductive research approach.

In order to obtain previous studies related to mobile apps guidelines for children with autism, several searching techniques were performed. The search for related information was divided into two categories which were mobile apps guidelines for children with autism and religious mobile apps guidelines. For the first category, the keywords used in the search were "mobile apps", "guidelines" and "autism spectrum disorder". As many as 12,900 publications were successfully obtained using the keyword. After filtering for published articles from the last five years (2015-2020), the number of articles decreased to 8,600 publications. Of these, only 35 articles were relevant to be analysed based on the guidelines criteria related to the study. The relevance of a paper was assessed based on its relation to searches on mobile applications guidelines for children with autism. Any paper

that writes about the guidelines used to develop mobile applications for children with autism will be selected and shortlisted for further analysis. Based on the relevant papers, only 20% of the papers were further analysed related to the study domain. As a result, this paper proposed a guideline for the development of mobile apps for children with autism.

The second searching category used the same terms of "mobile apps" and "guidelines" with additional words such as "religious" and "Islamic". A total of 32,200 publications was found based on the keywords. Last five years publications were inserted in the filters and the number of publications reduced to 16, 800 publications. Of these, only 26 publications were analysed for analysis in the first cycle. This publication was selected based on its relevance to religious mobile application guidelines for children with autism. After an extensive analysis of related guidelines criteria, only 19% publications were further analysed related to the study domain. Therefore, mobile apps guidelines for religion will be elaborated in the next section.

IV. Findings

This study found that the lack of religious mobile apps guidelines was clearly proven based on analysis that was conducted in the literature review section. Most researchers only focused on general mobile apps guidelines for children with autism and for religious guidelines researchers, most of the researchers do not focus on children with autism. From these two guidelines, a guideline focusing on religious mobile apps guidelines for children with autism was identified. This guideline will be based on the criteria that had been extracted from the literature review analysis. 19 % papers were used during the literature review analysis which combined mobile apps guidelines for children with autism and religious mobile apps guidelines. Two main categories were identified which are multimedia elements and application features.

Text, images and audio are often mentioned by previous researchers when discussing mobile application guidelines for children with autism. Table I shows the guidelines which have been classified according to the elements.

Text is a crucial component of any user interface (UI). Text was utilised to deliver immediate context-based assistance information to interactive elements, often known as hints [33]. One of the challenges that children with autism have is reading comprehension. Usability will be affected by a complex user interface with numerous text fragments [34]. It is important to organise the textual UI elements in such a way, that they do not impose reading difficulties. Real-life images aid recognition and help children learn more quickly and effectively [7]. The use of inappropriate images will cause children with autism to be confused. This will cause them to lose focus during the learning session. Appropriate sound selection helps motivate and not to distract the user [35]. The audio should be clear but not too loud as it will be disturbing for the children with autism. The audio should correspond to the images and must be user-friendly [15]. Children with ASD can benefit from visual assistance such as text, graphics, and sound to help them learn and communicate with their family, friends, and teachers [21].

The second category is the features of application. Interface, navigation, customisation and interaction are features that are required to develop applications for children with children with autism. Table II shows the guidelines which have been classified according to the features.

Making a simple interface helps children with autism to navigate easily. Complex interface with many text fragments will reduce usability [22]. The use of intuitive navigation principles will help users to navigate quickly by discovering and without additional support. Customisation is a key element for the successful user experience for children with autism because they have very different personal preferences and needs [9]. Customisation enables users to adapt documents (font typeface, font size, line spacing, colours) to their preferences [17]. It also helps children with autism to match their own reading and comprehension abilities. Providing interaction in multiple dimensions and providing them in a balance is important [36]. Considering characteristics and differences of children with autism, these interaction types should be chosen carefully and only if necessary. The study [24] reported that interactive learning software with gesture-based user interaction designed on the basis of universal design principles are highly effective learning materials.

TABLE I. MULTIMEDIA ELEMENTS GUIDELINES

Elements	Autism Apps	Religious Apps	Sources
Text	1. It should be clear, simple, and short. 2. Use a simple visual and textual language. 3. Avoid writing long paragraphs. 4. Textual style features according to individual differences and learning backgrounds of the users. 5. Amount of words must be reduced. 6. Avoid too many texts in a column 7. The typeface must be consistent and do not exceed more than two types.	1. Use suitable font size 2. Information should be well-organised. 3. Use a large size font type and try to make it adjustable. 4. Make sure the use of text coincides with the original source	[8], [22], [24], [26], [3]
Images	1. Pictures can be drawings, photographs, and symbolic images. 2. Allow images magnification for better visualisation and ensure they continue to be understandable when enlarged. 3. Able to have a suitable size of pictures, images. 4. Use photographs for concrete places and people; icons, pictograms and symbols for abstract terms. 5. Images must be identical to real life. 6. Do not overlap transparent images and text	1. The dress code of men and women should adhere to Islamic dress code. 2. All graphics should be of excellent quality and resolution to make it easier for viewers to understand and receive clear Islamic material. 3. Choose appropriate images.	[8], [2], [13] [16], [22], [24], [26], [3]
Sound	1. Provide audio instructions and subtitles for texts but ensure that this is not the only alternate content representation. 2. Avoid the use of disturbing and explosive sounds. 3. Using distinctive audio and visual effects. 4. Using non-destructive, plain and melodic effects. 5. Using clear, comprehensible short audio expressions. 6. The audio should correspond to the images and must be user-friendly.	1. The Islamic background sound accompanying the page and the types of Islamic audio sound may further stimulate Islamic spiritual sense of the site. 2. Use male voice. 3. The user can easily adjust the volume. 4. Provide good recitation audio.	[8], [3], [21] [26]

TABLE II.   FEATURES OF APPLICATION

Elements	Autism Apps	Religious Apps	Sources
Interface	1. Design and structure should be simple, clear and predictable. 2. Simple interface. 3. Using simple, objective oriented, presentative layout design. 4. Interfaces must be easy on the eye of users. 5. Provide an uncluttered layout white space between text in one column. 6. Should be user friendly. 7. Design for simplicity and few elements on screen.	1. The Islamic identity is characterised by identity identifiers. 2. Application layout is characterised by Islamic concepts in terms of information and delivery. 3. Use features that present the Islamic identity.	[8], [2], [13] [16], [22], [21] [24], [26], [3]
Navigation	1. Navigation should be consistent and similar in every page/section. 2. Provide a simplified and consistent navigation between pages on every page. 3. Employing intuitive navigation principles in which users can navigate easily by discovering and without additional support. 4. Limit the number of options in menus to reduce cognitive load. 5. Ensure the user can identify menus and navigation functions. 6. Strive for simple, clear navigation. 7. Do not use complex menus.	1. Ensure navigation structure is simple and straightforward. 2. Navigation between screens to other screens in the developed application should be affordable and easy to control by the user.	[8], [3], [13] [21] [26]
Customisation	1. Provide options to customise information visualisation with images, sound and text according to individual user's preferences. 2. Allow personalisation of apps for users. 3. Allow customization in the application.	1. On the program's Islamic features, the application should provide for relevant modification based on the user's preferences. 2. Users can easily resize the font, buttons or icons.	[8], [22], [24], [26], [3]
Interaction	1. Adapt the interaction with users. 2. Touch screen interactions should have the appropriate sensibility and prevent errors in selection and accidental touch in interface elements. 3. Providing interaction through hardware and sensors of application device. 4. Avoid captchas and simultaneous tasks.	1. Usability traits may ignite a sense of professionalism in Islam. 2. Must have good interaction with users. 3. Avoid using interaction timeouts to memorise Al-Quran. 4. Avoid using fast-moving objects and animation. 5. Users receive accurate and timely responses from the application. 6. All functions in the application work well and are fast and accurate when reached by the user. 7. Users should be able to obtain religious information through the application's interactions, which should perform smoothly and consistently	[8], [2], [13] [16], [22], [21] [24], [26], [3]

V. DISCUSSION

Based on the findings, this paper resulted in two key findings in relation to mobile application guidelines on religious context for autism. Firstly, findings revealed that elements of multimedia consist of text, images and sounds is important to encourage children with autism to engage in active learning by mentally representing the material in words and in pictures and by mentally making connections between the pictorial and verbal representations. Secondly, the features of application consist of interface, navigation, customisation and interaction makes user interfaces perceivable, operable, and understandable by people with a wide range of abilities, and people in a wide range of circumstances, environments, and conditions.

Thus, the findings extended a previous guideline to inform design framework for design with children with autism still has validity. However, this paper builds on it and adapts two new perspectives which consist of 24 guidelines under two categories as outlined in (1) and (2).

1) New perspective on multimedia elements guideline for designing apps with children with autism.

Proposed Religious Mobile Apps Guidelines for children with autism (Multimedia Elements) are as follows:

a) The text used must be clear, simple and short. (Font size of 14).

b) Use only a maximum of two typefaces.

c) Use of text coincides with the original source. (Arabic text in the application was written based on the same number of pages in the Holy Quran).

d) Provide suitable size of images.

e) Use images that are identical to real life.

f) Use formal attire that are modest to religious sentiments clothes for any character, figure or real person. For example, the compliant driven ('awrah or clothing boundary).

g) The images used in apps should be relevant to the content and not display images that are not related to Islamic elements.

h) Use clear, comprehensible short audio expressions.

i) Avoid using disturbing and explosive sounds.

j) Use familiar sounds and voices. (Voice of parents and teachers).

k) Use audio commonly used in Islam. (For example, the sounds of Al-Quran recitation and the sounds of azan).

2) New perspective on application features guideline for designing religious mobile apps with children with autism

Proposed Religious Mobile Apps Guidelines for children with autism (Features of Application) are as follows:

a) Design and structure should be simple, clear and predictable.

b) Avoid cluttered interfaces.

c) Application should be based on the elements that best symbolizes Islam which are widely recognized by the Islamic community.

d) Application should use features that present the Islamic identity.

e) Applications should have simple navigation and easy to use.

f) Navigation in an application should be consistent.

g) Use standard navigation icons that are familiar to children with autism.

h) Allow customization (Font type, font size, themes and colours)

i) On the program's Islamic features, the application should provide for relevant modification based on the user's preferences. Avoid simultaneous tasks at one time.

j) Avoid distraction and interruption in the apps.

k) Speed of the application interaction must be consistent without any delay occurring.

l) Avoid using interaction timeouts to memorise Al-Quran using fast-moving objects and animation.

This research presents 24 guidelines under two categories. The first theme was the elements of multimedia used in developing mobile applications for children with autism. The findings section highlighted three important multimedia elements which are text, images and audio.

1) Text: This element refers to the selection of appropriate and accurate text in the application. This includes font size, font type and colours. It is necessary to use clear and simple text for children with autism [28]. It is well known that people with children with autism literally interpret the text content, and have problems understanding complex sentences [10]. Some researchers suggested allowing text customization for children with autism to choose what is best according to their individual preferences. It is important to choose and organise the textual elements in such a way to ensure children with autism do not experience reading difficulties.

2) Images: This element refers to the use of proper and suitable images in the application. In a religious context, the images should be appropriate which do not expose children with autism to forbidden views such as awrah in Islam. In addition, parents, teachers and children with autism requirements, interests, and preferences should be able to be reflected in the app through personal images of genuine scenarios and activities [11].

3) Audio: This element refers to the appropriate audio selection and arrangement in the application. Familiar voices are recommended to be used in the application to ensure that the children with autism feel comfortable when using the application. Some children feel scared and less confident when hearing voices that are unfamiliar to them. For religious mobile apps, the sounds that are commonly used need to be included in the apps such as Quran recitation, sounds of Azan and Doa recitation. This will help them to better know and approach their religious studies that are in line with the practice of daily life.

The second theme is the features of applications that are required in developing apps for children with autism. The features are interface, navigation, customization and interaction.

1) Interface: This element refers to the simplicity in designing interface for the mobile application. A cluttered interface must be avoided. Cluttered applications will be frustrating to be used by children with autism. When there are too many elements on a single page, the user might miss important information. Besides, the application should use features that present Islamic identities. These requirements are less mentioned by the previous researcher but have been extended in this paper.

2) Navigation: This element refers to the use of standard and consistent navigation. The icons, buttons and symbols must be familiar to the children with autism. This will help them to easily understand the use of the button and prevent confusion [27]. The arrangement of the navigation button must also be consistent to avoid any errors while using the application. This will help children with autism to use the application with their own capability without the need for special monitoring from teachers and parents.

3) Customisation: This element refers to the enabling customization in the application. Because children with autism have highly varied personal preferences and demands, customization is a critical component of a good user experience [14]. It notes that children with autism can benefit from this as it enables them to directly ensure that the interface is personalised to their individual needs. Font size, images magnification and colours are among the customization that is needed to be inserted into the application.

4) Interaction: This element refers to the responsiveness of the application. Delay, interruption and distraction must be avoided. The application should improve the speed and accuracy to complete a task. It is also referred to give responses toward the action performed by the children with autism. The application must be able to reduce the time in performing the task. According to [12], attention deficit and concentration times are very short among children with autism with at least 10-15 minutes per session. Therefore, it is important for mobile apps to interact with children with autism to prevent them from losing focus while using the application.

There are some significant differences when comparing the previous guidelines with the proposed guidelines. Previous guidelines focus more on guidelines in general for mobile applications while the proposed guideline focuses more on

guidelines improvement in a religious context. For example, the use of text in previous guidelines focused more on font, typeface and also the readability of text in the application. For the proposed guidelines, emphasis is placed on the use of text that needs to coincide with the original source (Al-Quran). In addition, previous guidelines highlight more on the use of images, size, and quality. While the proposed guidelines have improved the existing guidelines by emphasizing the importance of using images that are compatible with religious sentiments such as clothing boundaries. Most other elements also have improvements especially in the context of religion. This clearly shows that there is a significant difference between the previous guidelines and the proposed guidelines.

The limitation of this study is that the literature review that examines mobile applications for autistic children in the context of religion is very limited. Most of the articles found are related to mobile apps for autism context. As for the context of religion, limited study do not use mobile applications as the main platform for their research. This causes difficulties in finding accurate information in line with the objectives of this study.

VI. Conclusion

In conclusion, this research found that there are guidelines that can be used for the development of religious mobile applications for children with autism. The guidelines were based on the reviewed guidelines provided in the literature. It is hoped that these findings can be used as a reference for other researchers who are interested in developing applications that focus on religious context. However, these guidelines need to be verified by experts who are experienced in developing religious mobile applications for children with autism. It is important in order to ensure that the application is effective to be used by children with autism. Next, a prototype of a mobile application that applied the proposed guidelines will be developed and given a trial. An evaluation of the application is needed to determine how children with autism react and interact with the application. Feedback from the teachers, parents, and the children with autism will be gathered as it will be useful to improve the existing guidelines.

Acknowledgment

References

[1] Abd Raof, Siti Fatimah, Nor Azzyati Hashim, and Noor Azura Zainuddin. "An Evaluation of Quran Memorization Mobile App among Middle-Aged Adults and Early Elderly." Journal of Computing Research and Innovation 4, no. 1 (2019): 1-7.

[2] Abidin, Nurul Aziera Binti Zainol, and Nurakmal Binti Ahmad Mustaffa. "Factors that Affect Attitude of Apps Developers to Comply with the Islamic Work Ethic in Developing an Islamic Mobile App. (2020)".

[3] Ahmad, Nahdatul Akma, Azaliza Zainal, Saliyah Kahar, Mohammad Ashri Abu Hassan, and Roziyani Setik. "Exploring the needs of older adult users for spiritual mobile applications." Journal of Theoretical and Applied Information Technology 88, no. 1 (2016): 154.

[4] Alghamdi, Adil. "Saudi Special Education Teachers' Perspectives on the Use of iPads to Enhance Communication Skills for Students with Autism." PhD diss., University of South Florida, 2021.

[5] Alves, Fábio Junior, Emerson Assis De Carvalho, Juliana Aguilar, Lucelmo Lacerda De Brito, and Guilherme Sousa Bastos. "Applied behavior analysis for the treatment of autism: A systematic review of assistive technologies." IEEE Access 8 (2020): 118664-118672.

[6] Bernard-Opitz, Vera, N. Sriram, and Sharul Nakhoda-Sapuan. "Enhancing social problem solving in children with autism and normal children through computer-assisted instruction." Journal of autism and developmental disorders 31, no. 4 (2001): 377-384.

[7] Buzdar, Abdul Qadir, and Muhammad Farooq. "Memorization of Quran through Mobile Application in the Era of Transformative Marketing." Federal History 12 (2020).

[8] Dattolo, Antonina, and Flaminia L. Luccio. "A review of websites and mobile applications for people with autism spectrum disorders: Towards shared guidelines." In International Conference on Smart Objects and Technologies for Social Good, pp. 264-273. Springer, Cham, 2016.

[9] Dzulkifli, Izuli. "Teaching and Learning Aids to Support the Deaf Students Studying Islamic Education." Pertanika Journal of Social Sciences & Humanities 29, no. 4 (2021).

[10] Fauzi, Ali, Yayuk Fauziyah, and Taufik Churrahman. "Analysis of Interactive Application Development as a Tahfidz Al Quran Learning Strategy." KnE Social Sciences (2022): 1-9.

[11] Gallardo-Montes, Carmen del Pilar, María Jesús Caurcel Cara, Emilio Crisol Moya, and Sonia Jarque Fernández. "Assessment of apps aimed at developing basic instrumental skills in autistic children and teenagers." Mathematics 9, no. 9 (2021): 1032.

[12] Hakiman, Hakiman, Bambang Sumardjoko, and Deddy Ramdhani. "Worship Learning for Students with Autism in Inclusive Primary School." Specialusis Ugdymas 1, no. 43 (2022): 6133-6153.

[13] Haryani, Hendriyati, Erna Astriyani, and Viola Tashya Devana. "Exploration of Islamic Religious Learning Innovation Technology with the iLearning Approach." Aptisi Transactions on Technopreneurship (ATT) 3, no. 2 (2021): 189-200.

[14] Hisham, Ahmad Khairi Hafiz Khairul, and Noraziahtulhidayu Kamarudin. "Basic SolatFor Autism-A development guidance Mobile Apps." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 13, no. 2 (2022): 978-994.

[15] Hussain, Azham, Emmanuel OC Mkpojiogu, and Pauline Chiamaka Okoroafor. "Assisting Children with Autism Spectrum Disorder with Educational Mobile Apps to Acquire Language and Communication Skills: A Review." Int. J. Interact. Mob. Technol. 15, no. 6 (2021): 161-170.

[16] Isa, Wan Abdul Rahim Wan Mohd, Nor Laila Md Noor, and Shafie Mehad. "Towards conceptualization of Islamic user interface for Islamic website: An initial investigation." the Muslim World (ICT4M) 21 (2006): 23rd.

[17] Islam, Muhammad Nazrul, Uzma Hasan, Fourkanul Islam, Shaila Tajmim Anuva, Tarannum Zaki, and AKM Najmul Islam. "IoT-Based Serious Gaming Platform for Improving Cognitive Skills of Children with Special Needs." Journal of Educational Computing Research (2022): 07356331211067725.

[18] Jafar, Nur Madihah, Nazean Jomhari, Mohd Yakub, and Zulkifli Mohd Yusoff. "Improving the Attention Performance in High-Functioning Autistics during Memorization Lesson through Neurofeedback Training Approaches." Jurnal Usuluddin 47, no. 2 (2019): 131-154.

[19] Jani, Syahrina Hayati Md, Nurjannah Salleh, Nor Aishah Mohd Ali, and Mohamed Saladin Abdul Rasool. "Adapting Public-Private Partnership as Strategic Collaboration between Government and Philanthropy-Based Autism Spectrum Disorder Centre." (2021): 117-117.

[20] Khowaja, Kamran, Bilikis Banire, Dena Al-Thani, Mohammed Tahri Sqalli, Aboubakr Aqle, Asadullah Shah, and Siti Salwah Salim. "Augmented reality for learning of children and adolescents with autism spectrum disorder (ASD): A systematic review." IEEE Access 8 (2020): 78779-78807.

[21] Muchagata, Joana, and Ana Ferreira. "Visual Schedule: A Mobile Application for Autistic Children-Preliminary Study." In ICEIS (2), pp. 452-459. 2019.

[22] Pavlov, Nikolay. "User interface for people with autism spectrum disorders." Journal of Software Engineering and Applications 2014 (2014).

[23] Rashid, Noor Mohd, Shaharuddin Md Salleh, and Norah Md Noor. "Development of Jawi Spelling Skills Mobile Applications,'Oh Jawiku'." (2019): 80-89.

[24] Sani Bozkurt, Sunagul, and Sezgin Vuran. "An analysis of the use of Social Stories in teaching social skills to children with Autism Spectrum Disorders." Educational Sciences: Theory and Practice 14, no. 5 (2014): 1875-1892.

[25] Satiakumar, Mohishaa, Nurul Ain Binti Mohd Rashid, Norliyana Binti Hut, and Fawwaz Mohd Nasir. "Autism Learning Mobile Application." Multidisciplinary Applied Research and Innovation 3, no. 1 (2022): 139-145.

[26] Sofian, Nadiah Mohamad, Ahmad Sobri Hashim, and Wan Fatimah Wan Ahmad. "A review on usability guidelines for designing mobile apps user interface for children with autism." In AIP conference proceedings, vol. 2016, no. 1, p. 020094. AIP Publishing LLC, 2018.

[27] Stathopoulou, Agathi, Dionisis Loukeris, Zoe Karabatzaki, Evangelia Politi, Yolanda Salapata, and Athanasios Drigas. "Evaluation of mobile apps effectiveness in children with autism social training via digital social stories." (2020): 4-18.

[28] Stathopoulou, Agathi, Zoe Karabatzaki, Dimosthenis Tsiros, Spiridoula Katsantoni, and Athanasios Drigas. "Mobile apps the educational solution for autistic students in secondary education." (2019): 89-101.

[29] Strickroth, Sven, Dietmar Zoerner, Tobias Moebert, Anna Morgiel, and Ulrike Lucke. "Game-based promotion of motivation and attention for socio-emotional training in Autism." i-com 19, no. 1 (2020): 17-30.

[30] Thabtah, Fadi. "An accessible and efficient autism screening method for behavioural data and predictive analyses." Health informatics journal 25, no. 4 (2019): 1739-1755.

[31] Wiliams, Alexander, Adam Faturahman, Untung Rahardja, Fitra Putri Oganda, and Anggit Panji Pangestu. "NVivo based Al-Quranic features in students for autism therapy." Aptisi Transactions on Technopreneurship (ATT) 3, no. 1 (2021): 81-90.

[32] Xanthopoulou, Maria, Gioulina Kokalia, and Athanasios Drigas. "Applications for Children with Autism in Preschool and Primary

[33] Education." Int. J. Recent Contributions Eng. Sci. IT 7, no. 2 (2019): 4-16.

[34] Xiao, Yu, and Maria Watson. "Guidance on conducting a systematic literature review." Journal of planning education and research 39, no. 1 (2019): 93-112.

[35] Zamin, Norshuhani, Norita Md Norwawi, Noreen Izza Arshad, and Dayang Rohaya Awang Rambli. "Make me speak: A mobile app for children with cerebral palsy." International Journal of Advanced Trends in Computer Science and Engineering (2019).

[36] Zulkefli, Mohd Yusof, and Norfishah Mat Rabi. "Exploring the Usage of Computer-Mediated Communication in Assisting Individual with Autism Spectrum Disorder to Communicate." Al-i'lam-Journal of Contemporary Islamic Communication and Media 1, no. 1 (2021): 126-143.

Fuzzy Support Vector Machine based Fall Detection Method for Traumatic Brain Injuries

A New Systematic Approach of Combining Fuzzy Logic with Support Vector Machine to Achieve Higher Accuracy in Fall Detection System

Mohammad Kchouri[1], Norharyati Harum[2]*, Ali Obeid[3], Hussein Hazimeh[4]

Fakulti Teknologi Maklumat Dan Komunikasi, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia[1, 2]
Psymeon CTO, Paris, France[3]
Department of Computer Science, Lebanese University, Beirut, Lebanon[4]

Abstract—**Falling is a major health issue that can lead to both physical and mental injuries. Detecting falls accurately can reduce the severe effects and improve the quality of life for disabled people. Therefore, it is critical to develop a smart fall detection system. Many approaches have been proposed in wearable-based systems. In these approaches, machine learning techniques have been conducted to provide automatic classification and to improve accuracy. One of the most commonly used algorithms is Support Vector Machine (SVM). However, classical SVM can neither use prior knowledge to process accurate classifications nor solve problems characterized by ambiguity. More specifically, some values of falls are inaccurate and similar to the features of normal activities, which can also greatly impact the performance of the learning ability of SVMs. Hence, it became necessary to look for an effective fall detection method based on a combination of Fuzzy Logic (FL) and SVM algorithms so as to reduce false positive alarms and improve accuracy. In this paper, various training data are assigned to the corresponding membership degrees. Some data points with a high chance of falling are assigned a high degree of membership, yielding a high contribution for SVM decision-making. This does not only achieve accurate fall detection, but also reduces the hesitation in labeling the outcomes and improves the heuristic transparency of the SVM. The experimental results achieved 100% specificity and precision, with an overall accuracy of 99.96%. Consequently, the experiment proved to be effective and yielded better results than the conventional approaches.**

Keywords—*Fall detection; fuzzy logic; SVM; traumatic brain injuries; wearable sensor*

I. INTRODUCTION

The main goal of this paper is to introduce and propose a new fall detection method for traumatically brain-injured people. The goal is to suggest a highly accurate method that uses the advantages of both Fuzzy Logic and SVM. The main reasons behind the idea of integrating these two latter methods are their well-known limitations as standalone techniques, with SVM considered a "black box" and Fuzzy logic being especially limited for nonlinear problem solving. Falling is one of the major life-threatening problems faced by people with physical disabilities. According to World Health Organization research, falls are the second most common cause of injury-related death worldwide. Nationally, there was

a 53% increase in the number of total deaths due to falls from 2000 to 2019 [1]. Health systems are significantly impacted by falls. The quality of life of elderly persons can also be significantly impacted out of the fear of falling.

An Ambient Intelligent (AI) environment can improve the lifestyle of the disabled by using different sensor technologies. With such technologies, environments might become sensitive, adaptive, and responsive to people's presence for the sake of supporting them to live independently in the environment they prefer. One important aim of assistive technology is to allow disabled people to stay in their homes as long as possible without changing their lifestyles. Smart fall detection can offer support to people with special needs, enabling them to live actively and independently both at home and in their communities. This improves the quality of their lives on the one hand, and reduces costs for their families and the entire society on the other hand.

Currently, available techniques that are used to design fall detection systems are classified into two main categories based on their sensor type: one is an ambient-based fall detection system whereas the other relies on a wearable-based fall detection system [2]. Ambient-based approaches use ambiance sensors, including acoustic sensors, vibration sensors, pressure sensors, and infrared sensors for detecting a fall event [3,4,5]. They also use single or multiple cameras in an indoor environment to track a person's movements and body shape while falling [6,7,8]. One main drawback of ambient sensors is that they limit falls to only those detected in a pre-set area, and this does not seem suitable for people's mobility.

Wearable sensor-based fall detection methods track the user's body motion using embedded sensors such as accelerometers. These methods can detect a fall when the person is wearing the sensor anywhere and at any time. Wearable sensors have several advantages, including low cost, low power consumption, and ease of use. As a result, they are commonly used to detect human falls. However, those with a single accelerometer are insufficient to offer a robust system and are susceptible to false positive fall detection; which reduces system accuracy [9].

*Corresponding Author.

Recently, a fusion system in which the process of combining multiple data sources can produce more robust measurement and accurate detection was developed. In this context, fall detection based on Inertial Measurement Units (IMUs) sensors, which are composed of an accelerometer, gyroscope, and magnetometer can be attached to the body. This fall detection technique is of special interest due to the fact that it is undetectable, widely available, inexpensive, and has a low power consumption [10,11]. Additionally, this technique can offer complementary information about the activity performed [12]. The IMU has been widely used for data collection and for differentiating a fall event from Activities of Daily Living (ADLs) based on two main approaches which are threshold-based or machine learning classification algorithms.

In a threshold-based approach, the system notifies that a fall happens if the real-time sensor data surpasses the given threshold values after comparing them. In particular, a low threshold value brings out false alarms whereas a high threshold value causes large fall missing issues. The authors in [13] show that machine learning algorithms can effectively enhance the system's performance in comparison to the threshold-based method. In machine learning approach, impressive results can be obtained using various classifiers [12]. An SVM is one of the most popular algorithms used in supervised machine learning for activities of recognition and classification [5]. However, labels for the dataset are needed for the employment of SVM algorithms for fall detection. In more detail, some training points of falls that are uneven and similar to being mixed with the features of ADLs can significantly impact the performance of the learning ability of SVMs. Here, it is worth mentioning that classical SVMs can never solve inaccurate and ambiguous problems. FL has the ability to mimic the human way of thinking to productively use reasoning methods that are uneven and not precise, and it has been verified to be effective in reducing the false fall detection rate [9].

SVM as a standalone algorithm, just like FL, has its own advantages and disadvantages. Merging both methods by making use of the best traits of each algorithm will certainly yield outstanding results in some applications, as in the case of fall detection. The specific contributions of this study could be summarized as follows:

- Identifying the types of disabilities that require smart fall detection as this can help in developing the desired device.

- Proposing a new fall detection method that takes advantage of both FL and classical SVM. Our proposed method uses the logical reasoning of FL by Fuzzification of all input values of the training data into fuzzy membership functions to obtain the intermediate output. Different memberships reflect various contributions to the learning of decision-making. A high degree of membership of data points, which have a significantly high chance of falling, provides a significant contribution to SVM decision-making. As a result of the functions of the fuzzy membership, an SVM output is properly produced for

modeling data and driving training. This network combination reduces false positive alarms to obtain more accurate fall detection and it achieves better generalization ability that was motivated to perform much better than conventional SVM methods on smart fall detection.

The rest of this paper is structured as follows: Section II provides a review of existing papers related to fall detection systems. Section III presents a brief review of classical SVM and Fuzzy SVM basic theory. Section IV identifies the types of disabilities that require smart fall detection. Section V describes the hardware design to collect data and highlights the proposed fall detection method with a classification process. Section VI details the experimental settings, the performance evaluation, the experimental results, and the analysis for the various scenarios. Section VII summarizes the conclusion with a comment on future work.

II. RELATED WORK

Generally, the framework of automatic fall detection systems using wearable sensors consists of the collection of sensor data, a fall detection algorithm, and an emergency alarm. Data collected by the accelerometer and gyroscope are transmitted to a microcontroller to be processed to differentiate a fall from an ADL. Recently, there has been a growing interest in identifying and detecting fall events using IMU sensors [14]. The fusion of inertial sensor-based wearable systems can be effectively used to recognize fall events by examining the impact of the body on the ground as well as the body orientation before, during, and following the fall. Nevertheless, the location of the sensor can influence the performance of the system. Thus, numerous studies related to the topic of optimal sensor placement were conducted. The waist, wrist, trunk, thigh, back, ankle, foot, neck, and head represent the most common wearing positions. The authors in [15] studied fall detection by placing accelerometers on the subject's head, waist, and wrist. They reported that the most efficient positions are the waist and the head in contrast to the wrist which is not. The authors in [16] placed sensors on the trunk and thigh and reported the trunk as a better position. The authors in [17] identified that the waist location utilizing a single sensor was a suitable placement after evaluating single IMU sensors deployed at several places in the body. In conclusion, sensor location is an important factor in developing wearable sensor-based fall detection systems. Furthermore, the waist could be the best option for a wearable sensor-based fall detection system.

The classification algorithm is applied to classify ADL and several fall events. A wearable-based fall detection algorithm can be categorized into two approaches, namely: threshold-based and machine learning. Threshold-based approaches use single or multiple threshold values that can be adjusted automatically depending on motion history to classify events [15,16,18]. Due to the low computational complexity, current fall detection studies have widely used the threshold-based method. However, a high threshold must be set in order to obtain a highly accurate system. Hence, with a high threshold, there will be some lags in the system and, subsequently, some missed falls. In contrast, with an excessively low threshold,

there will be some misjudgments and frequent false alarms. This is why a suitable threshold should be set to avoid any problems.

Machine learning is a field of Artificial Intelligence. It explores how to use computers to imitate human learning activities [19]. It uses learning algorithms to extract features from raw data in order to gain new abilities, recognize current knowledge, and constantly improve performance and achievement [20]. Data processing includes pre-processing, smoothing data, and data reduction methods where the data are acquired from the sensor and then processed. This is followed by the classification stage, which either provides a prediction or a decision at the end of the process. Learning methods can be categorized into two main groups: supervised and unsupervised learning. The algorithm in supervised learning is trained using a labeled input dataset. In the case of unsupervised algorithms, there are no explicit labels associated with the training dataset, which saves time during processing [21]. Examples of algorithms used in fall detection experiments include Hidden Markov Model (HMM), K-Nearest Neighbors (K-NN), Random Forest (RF), SVM, Decision Tree, Linear Regression, Naïve Bayes, Fuzzy Inference System (FIS), and Artificial Neural Network (ANN), which have achieved significant success in detecting falls and classifying fall events from ADLs [22].

Based on the preceding, [23] successfully differentiate falls from ADLs using six machine learning classifiers, which are the K-NN classifier, Least Squares Method (LSM), SVM, Bayesian decision making (BDM), Dynamic Time Warping (DTW), and ANNs. They achieved the greatest results with the K-NN classifier and LSM; with sensitivity, specificity, and accuracy all above 99%. The authors in [24] attained the best accuracy in fall detection using the K-NNs classifier and the highest accuracy in distinguishing various falling activities using the RF classifier. However, [25] compared the applicability of RF and SVM in the development of wearable intelligent devices. The obtained results show that SVM is more suitable for the development of wearable intelligent devices. The authors in [26] used the K-NN and SVM algorithms for classifying the fused accelerometer and gyroscope data collected from smartphone sensors. They reported a classification performance of 98.32% for SVM and 97.42% for K-NN. ANNs have been greatly improved in recent times [27]. This method regularly outperforms classic machine learning algorithms in terms of learning ability. However, the model of the Neural Network is highly dependent on the quality and quantity of the training datasets and can be affected by too much disorienting information.

Study [13] evaluated the accuracy of these two approaches, which are threshold values and five machine learning algorithms. In fact, they concluded that five machine learning algorithms' gross production was greater than the overall performance of five algorithms based on the threshold. In addition, SVM-based classification has outperformed the five machine learning in terms of sensitivity and specificity. Since they are highly accurate in comparison with threshold-based fall detection methods, machine learning based fall detection algorithms are nowadays being widely used. The authors in [28] achieved the best accuracy performance of

99%, indicating that the system's performance in comparison to the threshold-based method can be effectively enhanced. Decision-making based on machine learning algorithms, on the other hand ensures high rates of true positives.

A. SVM-Based Fall Detection

SVM is a powerful tool in machine learning for classifying data with good generalization ability that is less computationally intensive than other algorithms like artificial neural networks, decision trees, and Bayesian networks due to their high accuracy, elegance, mathematical practicability, and simple geometric interpretation. In addition, they do not require a lot of training data to prevent overfitting [29,30].

In SVM-based fall detection, distinguishing between falling activities and non-falling ADLs is possible. For instance, [31] used a hyperplane of SVM as the separating plane to replace the traditional threshold method for the detection of falling ADLs. They used the Gaussian radial basis function to construct the kernel function with the cost parameter tuned where the constant (C) is the regularization parameter in the SVM. When adjusted, a balance between margin maximization and classification is realized [32]. The results showed that the SVM method is better than the threshold-based algorithm when the parameters of falling and non-falling ADLs are very close. The authors in [33] extracted features from data collected by the Kinect sensor followed by fall recognition by using an SVM algorithm. In [3], the authors proposed a Multi-Feature Semi-Supervised SVM framework for human fall detection to specifically handle the human fall classification problem where the Radial Basis Function (RBF) classifier is selected in SVM training on extracted features from the training samples. The authors in [34] proposed a system that detects human falls by using the audio signal from a microphone. Their system was designed in a way that models each fall or noise segment by means of a Gaussian Mixture Model (GMM). Then, the SVM classifier would be employed to classify audio segments into falls and different types of noise. Study [35] collected the body's acceleration and rotational angle data in the wearable terminal to execute the SVM algorithm. As such, the hyperplane, which separates fall events from ADL events, was introduced. RBF kernel function which allows nonlinear mapping in this model was selected with a penalty parameter factor (C) that was adjustable to obtain the largest gap distance. The authors in [36] employ multiple kernels of learning to distinguish difficult fall-like events. The classification performance kept improving until it became constant at a certain point when the tuned parameter (C) was increased. Nonetheless, the number of selected kernels also expanded with (C), which raised the computational cost. Another model proposed by [10] showed that an SVM classifier provides the best performance metrics when trained on a fall dataset containing simulated falls and when cross-validated with real-world falls. The classifier appears to be appropriate for further evaluation concerning real-world applications due to robust results and high accuracy, sensitivity, and specificity. The authors in [25] found that SVM is more suitable for fall detection algorithm based on multi-sensor data fusion.

The SVM can accomplish great performance in the classification and calculation processes. This can be achieved through the use of a relatively low amount of the provided learning data. However, they cannot give a comprehensible representation of where a produced output has been attained. They are similar to black-box models due to their complex structure, numerous parameters, and excessive abstraction [37]. With classical SVMs, the experimenters can never depend on their prior knowledge to process accurate classifications or solve inaccurate and ambiguous problems, which can also greatly impact the performance of the learning ability of SVMs.

B. Fuzzy Logic-based Fall Detection

Fuzzy systems are systems in which the variables have domain fuzzy sets. Such systems allow the encoding of structured, empirical, heuristic, or linguistic knowledge in a numerical framework. As opposed to conventional logic, FL is based on the mathematical theory of fuzzy sets, which mimics human thought and tries to reflect reality while taking into account all outcomes between 0 and 1 [38], where 1 means absolute truth and 0 means absolute falsehood. The FL consists of three main parts. The first part is Fuzzification which allows the conversion of crisp values into fuzzy membership functions. The second part is the fuzzy inference aggregation which contains all the rules and if-then conditions to control the decision-making system. In this phase, the Mamdani method, which requires finding the centroid of a two-dimensional shape by integrating it across a continuously varying function, is the most commonly used technique. The third part is the defuzzification process that converts the fuzzy sets into crisp values. Despite the availability of several defuzzification methods, the centroid technique remains the most popular among all of them [39,40].

In Fuzzy Logic-based fall detection, [41] confirmed that Fuzzy is capable of detecting a fall from real-time data as it requires minimum hardware and software specifications. Reusing the existing data, balancing the load amongst FLS devices, and cost-efficiency are some of the advantages offered by the FLS architecture to introduce flexible and smooth decisions. In their proposed method, they "fuzzify" each input value as a function of fuzzy membership, where each input contains three linguistic values: low, medium, and high. Every membership is classified as a turning point with different values. To perform their experiment, the researchers created nine rules to identify whether a fall occurred or not. Finally, they transformed a fuzzy output set into a crisp value in the defuzzification phase. The authors in [42] used FL to identify the range and type of fall, which can include the position before fall, fall direction, fall velocity, and post-fall inactivity. The authors in [43] looked beyond the traditional threshold-based approaches and implemented a fuzzy inference technique for precise decision making. They fused the data from multiple sensors and generated a value between 0 and 1, which implies the chance of a fall; thus reducing the number of false alarms. The authors in [44] initiated a new FL algorithm, which is worn on wrists to detect falls and reduce the number of false alarms. The fall detection system has three major phases which are data sampling, data processing, and fuzzy classification. In the three stages, a typical FL procedure

is followed by fuzzily setting all input values into fuzzy membership functions, executing all relevant rules to calculate the fuzzy output functions, and "defuzzifying" the fuzzy output functions to get output values. They used Mamdani's minimum operation and the AND-output rule in their fall detection algorithm and a weighted average formula in defuzzification. Furthermore, since the Mamdani approach is generally accepted for the development of expert knowledge, this helps one to explain the ability in a more perceptive, human-like manner. However, this technique is not computationally effective and can be expensive due to calculating a two-dimensional form by adding it up or combining it more accurately through a function that change continuously.

III. SVM AND FUZZY SVM BASIC THEORY

In this section, we briefly review the basics of the theory of SVM in classification problems and fuzzy support vector machines, which are discussed by [45,46,47.48].

A. SVMs Theory

SVM is a robust classification method that was developed by Vladimir Vapnik and aims to construct a decision function that separates the data in the input space into different classes. The basis of this method is minimizing the structural risk method in order to reduce the error [32]. An optimal hyperplane must be found through maximizing the margin between classes in input space. Nonetheless, the samples close to the hyperplane are called support vectors [49].

In the input space, it is assumed that the patterns are drawn by the training points $\{x_1, .., x_n\}$. If this input data are linearly separable, the hyperplane that produces the separation could be described as: $w^T x + b = 0$ where x is an input vector, $w = (w_1, w_2, .., w_d)$ is a weight vector, d is the number of the input variables, and b is a bias that will determine the distance between the hyperplane and the origin. It can classify the points using the following equation:

$$\begin{cases} w^T x_i + b \geq 0 \ for \ y = +1 \\ w^T x_i + b \leq 0 \ for \ y = -1 \end{cases} \tag{1}$$

where $y_i \in \{-1, +1\}$ is the binary class label for a new point x_i, this output enables x_i to be classified as belonging to one of the two classes.

The goal of the support vector machine is to build the optimal hyperplane that maximizes the distance between the closest points of each class and the separation [50]. This is presented by solving the linear problem:

$$\begin{cases} min \ \frac{1}{2} \|w\|^2 \\ y(w^T x_i + b) \geq 1, i = 1, 2, .., N \end{cases} \tag{2}$$

To deal with data that are not separable cases, slack variables ε_i are introduced. This represents the misclassified sample of the corresponding margin hyperplane, where $i \in \{1, 2, 3, .., N\}$ is an upper bound of the number of error [48]. Thus, the optimal hyperplane in a nonlinear space can be determined by:

$$minimize\ 0.5\|w\|^2 + c \sum_{i=1}^{N} \varepsilon_i$$

subject to:

$$\begin{cases} y_i(w^T x_i + b) \geq 1 - \varepsilon_i, i = 1, 2, .., N \\ \varepsilon_i \geq 0, i = 1, 2, .., N \end{cases} \quad (3)$$

In this sense, the adjustable parameter (C) plays a major role in maximizing the margin and carefully tuning the number of misclassifications. If (C) is bigger, it makes the training of SVM fewer misclassifications and it narrows the margin. In contrast, if (C) decreases, it makes SVM disregard more training points and it widens the margin.

However, the search for a suitable hyperplane in an input space is too restrictive to be of practical use when the samples are linearly non-separable. Thus, SVM techniques utilize a group of mathematical functions known as the kernel that satisfies Mercer's theorem [51]. This can be applied to map data relevant to the feature of higher dimensions, hence seeking an optimal separating hyperplane in the feature space [29].

The main approach is to find the function that performs the mapping from the input to the feature space. By introducing the vector of Lagrange multipliers [31], the nonlinear separating hyperplane can be found as the solution of:

$$\max \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i . x_j)$$

subject to:

$$\begin{cases} \sum_{i=1}^{N} \alpha_i y_i = 0, \forall\ i \\ 0 \leq \alpha_i \leq c \end{cases} \quad (4)$$

where $k(x, x_i) = \langle \emptyset(x), \emptyset(x_i) \rangle$ is the dot products of the corresponding feature vectors into high dimensional space [51].

Four common types in kernel at the SVM algorithm are linear, polynomial, Gaussian RBF, and sigmoid kernel where each kernel function has a particular parameter that must be optimized to obtain the best result performance [49]. Three kernel types will be used in the experiments to compare their results to those of our proposed method. These kernel types are:

- Linear kernel: $k(x, x_i) = \langle x, x_i \rangle$.

- Polynomial kernel: $k(x, x_i) = (\langle x, x_i \rangle + 1)^d$ with degree d.

- RBF kernel: $k(x, x_i) = e^{\frac{-(\|\vec{x} - \vec{x_i}\|)^2}{2\sigma^2}}$ with adjustable width parameter σ.

B. Fuzzy SVMs Theory

SVM is a powerful tool for classifying data points that are assumed to belong to the one and only class [52]. However, as discussed previously, the effects of the training points are different. Especially for fall detection, classical SVM can neither use prior knowledge to process accurate classifications nor solve problems characterized by inaccuracy and ambiguity. In more specific terms, some values of falls are inaccurate and similar to the features of normal activities, which can also greatly impact the performance of the learning ability of SVMs. Furthermore, some training points no longer exactly belong to one of the two classes. For example, 80% belong to the class of falls and 20% to the class of ADLs. These points are critical and may cause false positive alarms which reduce the accuracy of the system.

To do that, the points that have a low potential for falling or normal activity will be assigned to lower membership functions. Otherwise, a high chance of falls will be assigned with a high degree of membership function. In this sense, each training point is fuzzified into a membership function. This fuzzy membership $\mu_i \in \{0, 1\}$ is considered to be the attitude of the corresponding training point toward one class in the classification problem whereas the value $(1 - \mu_i)$ is considered to be the attitude of meaninglessness. So, the idea of SVM will be expanded and combined with a fuzzy membership function to make it a Fuzzy SVM.

The term μ_i is introduced as a membership vector for each training point x_i. Thus, the optimal hyperplane problem is then regarded as the solution to:

$$min\ 0.5\|w\|^2 + c \sum_{i=1}^{N} \mu_i\ \varepsilon_i$$

subject to:

$$\begin{cases} y_i(w^T x_i + b) \geq 1 - \varepsilon_i, i = 1, 2, .., N \\ \varepsilon_i \geq 0, i = 1, 2, .., N \end{cases} \quad (5)$$

where the term $\mu_i \varepsilon_i$ is a measure of error with different weighting. Hence, it can be noticed that the effect of misclassified parameter ε_i will be reduced when the membership functions μ_i are smaller. In this case, the training point x_i is treated as less important in the training. By applying Lagrange multipliers, the above problem is reformulated as:

$$\max Q(x) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i . x_j)$$

subject to:

$$\begin{cases} \sum_{i=1}^{N} \alpha_i y_i = 0, \forall\ i \\ 0 \leq \alpha_i \leq s_i c \end{cases} \quad (6)$$

Solving problem (6), dual of (5), is the same for classical SVM with a slight difference. As a consequence, this is the basic theory of Fuzzy SVM.

IV. CLINICAL STUDY

A clinical study was conducted to identify the disabled group that needs fall detection. Patients and specialists in healthcare centers were interviewed to collect data on health and welfare. The interviewer was a physiotherapy officer in Lebanon, and the interviewees were people with special needs living in the sample households. A survey was distributed to a

group of physically disabled people who suffered traumatic brain injuries. Each case had a different kind of disability, such as Cerebral Vascular Accident (CVA), Cerebral Palsy (CP), Meningitis, or Guillain-Barre. This survey included information on the installation of sensors based on the preference of disabled people. As a result, patients with brain injuries were asked to fill out a survey with a special focus on the duration of their disability, the reason for disability, and the number of falls in the last 6–12 months.

In Fig. 1, it is illustrated that 65% of the studied cases were patients with CVA, 25% were patients with CP, 6% were patients with Meningitis, and 4% were patients with Guillain-Barre. In Fig. 2, which studies the number of falls in each case, it is revealed that CVA cases aged above 55 ran into 83 falls, which is 37%, while CVA cases aged under 55 ran into 60 falls, ranking a lower percentage, which is 27%. In general, CVAs ranked 64%, indicating the highest percentage of falls. In addition, CP cases marked 59 falls, indicating the second-highest percentage of falls, which is 26%, whereas other cases marked 24 falls, showing the lowest percentage of falls, which is 10%. Thus, it was noticed that people with disabilities related to traumatic brain injuries such as CVA, CP, Guillain-Barre, and Meningitis have a high frequency of falls. Each respondent gave a different answer regarding the position of the fall detection sensor.

Fig. 1. Percentage of Participants with Traumatic Brain Injuries.
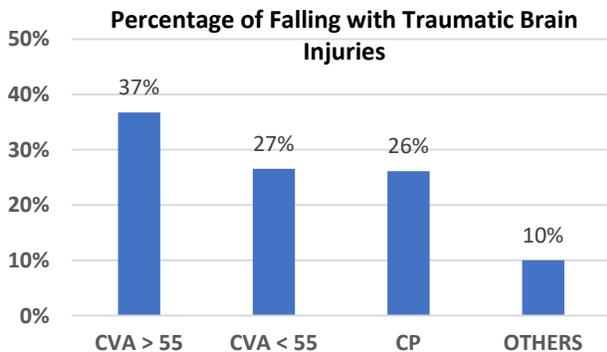
Fig. 2. Percentage of Falling for Participants with Traumatic Brain Injuries.
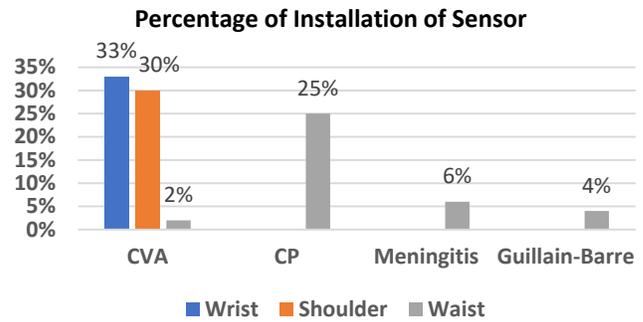
Fig. 3. Preferred Position of the Sensor.

Fig. 3 shows that 37% of the CP, Meningitis, and Guillain-Barre cases suggested placing the sensor on the waist in a way that allows it to be balanced right between the upper part of the body and its lower part. On the other hand, CVA cases had different answers, each according to their age. Those aged 55 and above (33% of the CVA cases) wanted to install the sensor on the wrist to quickly detect a fall, while people under the age of 55 (30%) wanted to install it on the shoulder.

In this paper, we are going to install a sensor for brain injury cases on the waist since it is the most fixed point of the body and is needed to maintain joint stability. Thus, it can track any movement easily.

V. PROPOSED METHOD

A prototype of the wearable device is designed using hardware in the form of a small-sized IMU sensor. The digital output of a 9-axis motion tracker by the IMU module is accessible by the I2C communication protocol (Inter-Integrated Circuit). This module is based on the MPU-9250, which assists in detecting activity changes, determining the slope of the object on which the sensor is mounted, generating acceleration, and expressing the angle and rotation about each axis in 3D space. It also achieves targets with low power consumption and robustness during the short duration of dynamic accelerations. Data gathered by the sensor are defined by an Arduino Uno microcontroller that operates at a voltage of 3.3/5 volts and is used to read the accelerometer, the gyroscope, and the magnetometer, as well as the internal temperature and the Tait Bryan angle-like pitch roll and yaw. The Baud rate is set to 9600 bits per second for serial communication between the Arduino board and the MPU-9250. The acquired data are transferred via a low-energy Bluetooth interface to the classification part. The Bluetooth module HC-05 is used for wireless communication between the Arduino Uno and MPU-9250. Bluetooth technology is a suitable choice for a lot of applications in daily life as it provides a reliable connection and low power consumption [53].

This paper suggests a new method to detect falls through the effective combination of FL and SVM (Fuzzy SVM). The input matrix consists of a 9-axis accelerometer, gyroscope, and magnetometer to collect multiple human body data points at the same time, including human body acceleration, rotational velocity, and displacement along the three

directions. The next step involves smoothing the data collected and computing the standard deviation of each of the nine axes from the IMU sensor in 3D space. After that, the magnitude of standard deviation features is extracted to indicate abnormal activity in the preprocessing phase. The building of the FL model is formed by using a trapezoidal membership function along with the input dataset to obtain the intermediate output. SVM, with selected kernels, will use the high degree of membership function of the three inputs to determine whether a fall has occurred or not and to obtain a confidential decision. Additionally, and to summarize, hyperparameters, including sensitivity, specificity, accuracy, and precision, were monitored during the methods' evaluations. As for the 9-axis sensor, all nine parameters (projections of acceleration, gyroscope, and magnetometer into three-dimensional space) were included in the model's inputs with different correlation indices.

A. Data Smoothing

A simple moving average filtering method is used for smoothing noisy raw data in order to obtain clearer data and state estimation of signals from human activities [25]. This can be obtained in (7) from the mathematical definition of a vector x:

$$y(n) = \frac{1}{windowSize} (x(n) + x(n-1) + \cdots + x(n-(windowSize-1))) \tag{7}$$

where, $y(n)$ is the current output, $x(n)$ is the current input, $x(n-1)$ is the previous input, etc.; noting that n is the length of the window size.

In Fig. 4, the signal in the original data is a pulse buried in random noise. In Fig. 5, this signal is filtered with $n = 5$ point moving average filters. The noise level becomes lower when the number of points in the filter increases. The optimal solution for this problem is the moving average filter, which gives the lowest noise possible for a given edge sharpness. The smoothing action of the moving average filter reduces the amplitude of the random noise when $N = 5$. Averaging the raw data leads to smoothing out the incidental peaks.

B. Feature Extraction

The acceleration value of human movement for specifying the changing rate of human motion can be calculated by using standard deviation. This function is sensitive to fall detection and can detect sudden tilt changes [54]. Standard deviation is useful for distinguishing static from dynamic activity and for identifying dynamic activity. Assumingly, a human fall is of high acceleration whereas walking is regarded as a low-acceleration activity [55,56].

The standard deviation is an index of how near the individual data points cluster around the mean. If we called each data point x_i, an index of dispersion, it would be represented in (8):

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{i=N} (x_i - \bar{x})^2} \tag{8}$$

where $i = 1, .., N$ is an index of data sample, N is the number of data sample, and $\bar{x}$ represents the sample mean.
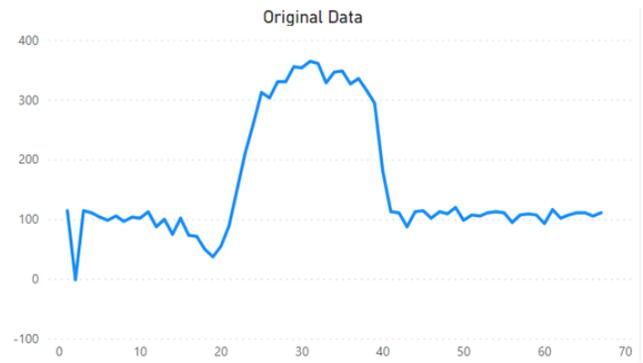
Fig. 4. Original Data by Moving Average Filter.
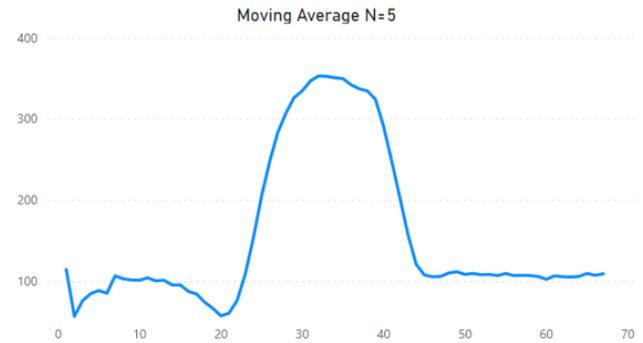
Fig. 5. Smoothing Raw Data by Moving Average Filter.

We computed the standard deviation of each of the nine axes depending on whether they were an accelerometer, a gyroscope, or a magnetometer raw data. After that, three features were extracted from the IMU sensor, which are:

The norm of a standard deviation of the acceleration $|\sigma_A|$ is calculated in (9):

$$|\sigma_A| = \sqrt{\sigma_{Ax}^2 + \sigma_{Ay}^2 + \sigma_{Az}^2} \tag{9}$$

where σ_{Ax}, σ_{Ay}, and σ_{Az} are the standard deviation along the directions of $x, y,$ and z axes of the acceleration raw data that are represented by A_x, A_y, and A_z respectively.

The norm of a standard deviation of the rotation $|\sigma_G|$ is calculated in (10):

$$|\sigma_G| = \sqrt{\sigma_{Gx}^2 + \sigma_{Gy}^2 + \sigma_{Gz}^2} \tag{10}$$

where σ_{Gx}, σ_{Gy}, and σ_{Gz} are the standard deviation along the directions of $x, y,$ and z axes of the rotation raw data that are represented by G_x, G_y, and G_z respectively.

The norm of a standard deviation of the magnetometer $|\sigma_M|$ is calculated in (11):

$$|\sigma_M| = \sqrt{\sigma_{Mx}^2 + \sigma_{My}^2 + \sigma_{Mz}^2} \tag{11}$$

where σ_{Mx}, σ_{My}, and σ_{Mz} are the standard deviation along the directions of $x, y,$ and z axes of the magnetometer raw data that are represented by M_x, M_y, and M_z respectively.

Thus, when a standard deviation method is applied, it can differentiate between an actual fall and other activities.

This method can compute a sudden change in acceleration in zero gravity. If this standard deviation value has a high changing trend, it indicates unusual activity. In Fig. 6, a fall might be shown to arise if this value has a high changing rate. As a result, the human motion after this is re-examined to check if the human body has no movement. In this way, it can be ensured that it is the actual fall.
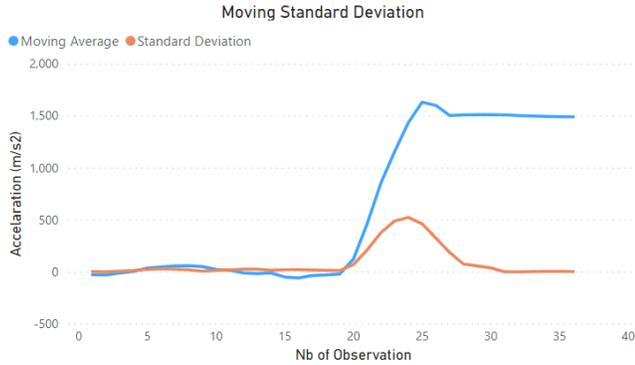
Fig. 6. Example of Standard Deviation for Accelerating Raw Data.

C. Generate Fuzzy Membership Function

A membership function is a function that is responsible for defining how each point in the input space is mapped to a membership value (or degree of membership). A fuzzy set F in X will be explained as a group of ordered pairs, as expressed below:

$$F = \{x, \mu_F(x) | x \in X\} \qquad (12)$$

In particular:

- $\mu_F(x)$ is the degree of membership of x in F.

- The boundaries of a fuzzy set F is the set of all $x \in X$ such that $0 < \mu_F(x) < 1$.

- X is the input space and its elements are denoted by x.

In this paper, an MPU–9250 sensor device in an IoT-enabled environment is used to effectively recognize fall events by examining the impact of the body on the ground in addition to the body orientation prior to, during, and following the fall. Next, the proposed method fuzzily analyses these inputs and produces an output as a crisp value between 0 and 1, which signifies the possibility of a fall. FL is used to solve such classification and decision problems without a clear threshold boundary. When a specific value does not completely belong to a certain category, the membership function is used to measure it. The process of converting the logical input value into the membership of each set (normal, medium, high) is called fuzzification.

Three input features, $|\sigma_A|$, $|\sigma_G|$, and $|\sigma_M|$, are introduced to build the suggested approach in order to get the intermediate output. These three linguistic variables for every three inputs are represented in (13):

$$f_{|\sigma_A|} = f_{|\sigma_G|} = f_{|\sigma_M|} = \{NORMAL, MEDIUM, HIGH\} \qquad (13)$$

In this paper, trapezoidal Membership Function (MF) was considered as this type is most frequently used, very flexible, and a small amount of data is needed to define it. The trapezoidal function guarantees the existence of a certainty interval in the fuzzification [40].

Each membership is configured with specific values as specification points based on our experimental test to balance sensitivity and specificity. An FL model will be built based on selectable membership functions of the input datasets to get the intermediate output.

The general form of high trapezoidal MF for the fuzzy set $|\sigma_A|$, in terms of the degree of membership, could be defined in (14):

$$\mu_{high_{|\sigma_A|}} = \begin{cases} 0, \text{for x} < 145 \\ \frac{x-145}{160-145}, \text{for } 145 \le x \le 160 \\ 1, \text{for x} \ge 160 \end{cases} \qquad (14)$$

The general form of high trapezoidal MF for the fuzzy set $|\sigma_G|$, in terms of the degree of membership, could be defined in (15):

$$\mu_{high_{|\sigma_G|}} = \begin{cases} 0, \text{for x} < 12 \\ \frac{x-12}{15-12}, \text{for } 12 \le x \le 15 \\ 1, \text{for x} \ge 15 \end{cases} \qquad (15)$$

The general form of high trapezoidal MF for the fuzzy set $|\sigma_M|$, in terms of the degree of membership, could be defined in (16):

$$\mu_{high_{|\sigma_M|}} = \begin{cases} 0, \text{for x} < 30 \\ \frac{x-30}{35-30}, \text{for } 30 \le x \le 35 \\ 1, \text{for x} \ge 35 \end{cases} \qquad (16)$$

Fig. 7 illustrates the membership of the magnitude of a standard deviation of the acceleration in a plan. The normal magnitude is assigned to 80, the medium magnitude is distributed between 95 and 145, and the maximum is defined as 160. Therefore, we put a minimum value that a fall can happen at 145 and consider it as the medium acceleration value.

Fig. 8 depicts memberships of the magnitude of a standard deviation of the rotation in a plan. The normal magnitude is assigned to 6, the medium magnitude is distributed between 9 and 12, and the maximum is defined as 15. Therefore, we put the minimum value that a fall can happen at 12 and consider it the medium angle value.

Fig. 9 represents memberships of the magnitude of a standard deviation of the displacement in a plan. The normal magnitude is assigned to 20, the medium magnitude is distributed between 25 and 30, and the maximum is defined as 35. Therefore, we put a minimum value that a fall can happen at 30 and consider it the medium magnetometer value.
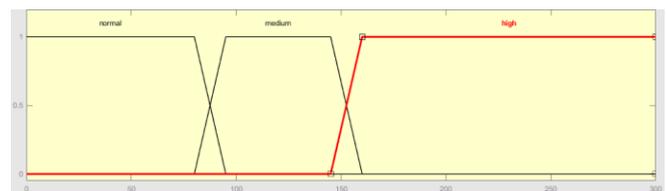
Fig. 7. Membership Function for the Input 1: $|\sigma_A|$.
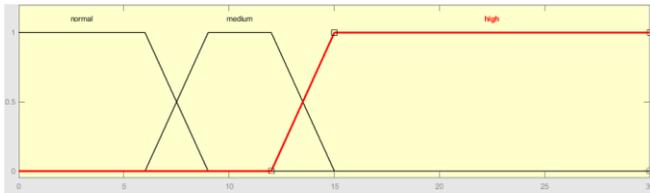
Fig. 8. Membership Function for the Input 2: $|\sigma_G|$.
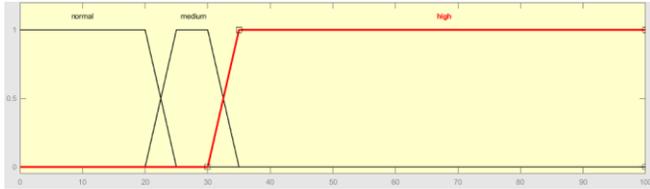
Fig. 9. Membership Function for the Input 3: $|\sigma_M|$.

The output of such FL analysis is a block formed from a high degree of membership based on these three inputs. This means that the event will be considered to have a high degree of membership if the magnitude of the standard deviation of the acceleration is greater than 145, if the magnitude of the standard deviation of the rotation is greater than 12, and if the magnitude of the standard deviation of the magnetometer is greater than 30. The intermediate output will be used as an SVM input with selectable kernels to reach the decision. This will help provide an initial decision bias on whether an event of falling likelihood is high. SVM will use the high degree of membership function of the three inputs to determine whether a fall has occurred or not. However, data points that have a potential normal or medium chance of falling are assigned lower membership degrees.

VI. Experimental Scenario

The prototype was attached at the waist to collect real-time motion data since this is the most fixed position of the body and is needed to provide joint stability and efficiently track any movement. Unlike wired systems, wireless data collection allows users to perform movements more fluently. Four volunteers were chosen to participate in a simulated falling event where they implemented the fall activities on 30 cm thick mats to prevent injuries. These volunteers have a healthy body, are aged between 25 and 35 years, weigh between 70 and 100 kg, and are 1.68 to 1.94 m tall. In this paper, different subcategories and characteristics of falls in five directions (forward, backward, left, right, and vertical falls) and normal activities including (walking, sitting, and stumbling while walking) were examined in the experiments to achieve our goal. The average duration of each trial was about 40 seconds. Each fall type was repeated more than ten times in a total of 100 trials.

The dataset was read from the CSV file and was implemented with the assistance of the LIBSVM library for MATLAB (version R2018a). Three predictors were introduced where they showed the value of each nine-axis of the trapezoidal membership function of the magnetometer, accelerometer, and gyroscope. Two classes of responses, -1 and 1, were designated to sort the non-falling and falling events, respectively. To make model predictions and estimate how accurate a predictive model will be when implemented, a

cross-validation model was used. The default option is 5-fold cross-validation to partition the data set into 5 folds. This helps protect against overfitting and examines the predictive accuracy of the fitted models. We trained an SVM classifier in the supervised machine learning model. This was done by providing a known input data set (2250 observations) in addition to known replies to the data that comprise two classes: -1 indicating data points of the non-falling type, and +1 indicating data points of the falling type. We used linear, cubic, quadratic, and RBF SVM kernel functions and standardized the features.

In this paper, the kernel function will be specified as a Medium Gaussian or RBF kernel, a Quadratic kernel, and a Cubic kernel. In order to tune the SVM classifier, the kernel scale was picked by different scales. Next, we adjusted the box constraint level with diverse values for every kernel scale to reach the self-confidential decision. The adjustable parameter (C) usually plays a vital role between cautiously tuning the number of mistakes and maximizing the margin. Therefore, the increase in the box constraint level might decrease the number of support vectors, but it can also increase the time of training, make the training of SVM have fewer misclassifications, and narrow the margin. On the other hand, it allows the SVM to discount more training points and it widens the margin. The evaluation of the Fuzzy SVM network included dividing the datasets from model training (75%) and real-time data for testing new data (25%).

A. Experimental Results Metrics

In the final phase, to show the effectiveness of the proposed method, the confusion matrix plot was calculated between the model predictions and the ground truth labels in order to check each class's performance and to know how the current classifier performed in every class [57]. Classification algorithm performance has traditionally been evaluated using a range of performance criteria, which are presented in Table I.

TABLE I. DESCRIPTIONS OF PARAMETRIC EVALUATION

	Detected Fall	Undetected Fall
Fall Occurrence	True Positive (TP)	False Negative (FN)
Fall Unoccurence	False Positive (FP)	True Negative (TN)

In this paper, we compute the following common machine learning classification metrics to assess the classifier's effectiveness in our evaluation. Sensitivity and specificity are calculated by (17) and (18), respectively. Accuracy could be defined by (19), which is the ratio of all samples that the classifier correctly classified [58]. The precision metric could be described by (20), which evaluates the number of correct positive predictions made. Hence, low precision might be an indication of a large number of FP.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{17}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{18}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad\qquad (20)$$

B. Experimental Results of Proposed Method without Fuzzy MFs

This section reports the preliminary results of the fall detection system. We used the Gaussian RBF kernel for training and adjusted the optimal regularization parameter with diverse values for every kernel scale. The box constraint level, denoted as the soft margin penalty (C), was varied with different values for each kernel scale in this experiment to provide increased flexibility.

Table II shows the experimental results of the proposed method without employing Fuzzy MFs with RBF kernel function. The percentage of accuracy ranges between 97.78 and 97.96, the percentage of sensitivity ranges between 93.39 and 95.21, whereas the percentage of specificity ranges between 98.91 and 99.45; with precision between 96.96 and 98.43. The boldfaced numbers reveal the best-obtained results in the current classifier that were used later to compare with employed Fuzzy MFs to show the improvement of the proposed method.

In Table III, the experimental results indicate that the best model was selected in the Quadratic kernel, where the percentages of accuracy and sensitivity are 97.78 and 94.71, respectively. As for specificity and precision, the best results were obtained in the cubic kernel, which reached 99.03 and 97.26, respectively.

TABLE III. EXPERIMENTAL RESULTS WITHOUT FUZZY MFS FOR LINEAR, CUBIC, AND QUADRATIC KERNEL FUNCTIONS

Kernel function	Accuracy %	Sensitivity %	Specificity %	Precision %
Linear SVM	97.47	93.88	98.78	96.6
Cubic SVM	97.64	93.88	**99.03**	**97.26**
Quadratic SVM	**97.78**	94.71	98.91	96.95

C. Experimental Results of Proposed Method with Fuzzy MFs

Table IV shows the experimental results of the proposed method by employing Fuzzy MFs with RBF kernel function. The percentage of accuracy ranges between 99.82 and 99.96, the percentage of sensitivity ranges between 99.63 and 99.81, whereas the percentage of specificity ranges between 99.88 and 100; with precision between 99.63 and 100. The boldfaced numbers reveal the best-obtained results in the current classifier by employing Fuzzy MFs to show the improvement of the proposed method.

In Table V, the experimental results of Fuzzy SVM with Cubic and Quadratic Kernel Function indicate that the percentage of accuracy is 99.96, the percentage of sensitivity is 99.81, whereas the percentage of precision and specificity is 100 for both types of the kernel function. An overall score of 100% for specificity and precision and 99.81% for sensitivity was obtained by using the new method.

TABLE II. EXPERIMENTAL RESULTS OF THE PROPOSED METHOD WITHOUT FUZZY MFS FOR GAUSSIAN RBF KERNEL FUNCTION

K=3	Accuracy %	Sensitivity %	Specificity %	Precision %
C = 1	97.82	93.39	**99.45**	**98.43**
C = 2	97.82	94.38	99.09	97.44
C = 3	97.87	94.71	99.03	97.28
C = 4	97.91	94.88	99.03	97.29
C = 5	97.91	94.88	99.03	97.29
C = 6	97.87	94.71	99.03	97.28
C = 7	97.87	94.71	99.03	97.28
C = 8	97.91	94.88	99.03	97.29
C = 9	97.82	94.88	98.91	96.96
C = 10	97.82	94.88	98.91	96.96
K=2	Accuracy %	Sensitivity %	Specificity %	Precision %
C = 1	97.78	94.38	99.03	97.27
C = 2	97.82	94.71	98.97	97.12
C = 3	97.91	95.04	98.97	97.13
C = 4	97.87	94.88	98.97	97.12
C = 5	97.78	94.71	98.91	96.95
C = 6	97.82	94.88	98.91	96.96
C = 7	97.91	95.04	98.97	97.13
C = 8	**97.96**	**95.21**	98.97	97.13
C = 9	97.91	95.21	98.91	96.97
C = 10	97.87	95.04	98.91	96.96

TABLE IV. EXPERIMENTAL RESULTS OF THE PROPOSED METHOD WITH FUZZY MFS FOR GAUSSIAN RBF KERNEL FUNCTION

K=3	Accuracy %	Sensitivity %	Specificity %	Precision %
C = 1	99.91	99.63	100	100
C = 2	99.87	99.81	99.88	99.63
C = 3	99.82	99.63	99.88	99.63
C = 4	99.82	99.63	99.88	99.63
C = 5	99.82	99.63	99.88	99.63
C = 6	99.87	99.63	99.94	99.81
C = 7	99.91	99.63	100	100
C = 8	**99.96**	**99.81**	**100**	**100**
C = 9	**99.96**	**99.81**	**100**	**100**
C = 10	**99.96**	**99.81**	**100**	**100**
K=2	Accuracy %	Sensitivity %	Specificity %	Precision %
C = 1	99.91	99.63	100	100
C = 2	99.91	99.63	100	100
C = 3	99.91	99.63	100	100
C = 4	**99.96**	**99.81**	**100**	**100**
C = 5	**99.96**	**99.81**	**100**	**100**
C = 6	**99.96**	**99.81**	**100**	**100**
C = 7	**99.96**	**99.81**	**100**	**100**
C = 8	**99.96**	**99.81**	**100**	**100**
C = 9	**99.96**	**99.81**	**100**	**100**
C = 10	**99.96**	**99.81**	**100**	**100**

TABLE V.   EXPERIMENTAL RESULTS OF THE PROPOSED METHOD WITH FUZZY MFS FOR LINEAR, CUBIC, AND QUADRATIC KERNEL FUNCTIONS

Kernel function	Accuracy %	Sensitivity %	Specificity %	Precision %
Linear SVM	99.56	99.44	99.59	89.70
Cubic SVM	99.96	99.81	100	100
Quadratic SVM	99.96	99.81	100	100

In Table VI, we compared the effectiveness of our proposed fall detection method when we employed the Fuzzy MFs. The overall performance in terms of accuracy, sensitivity, specificity, and precision was increased by 2%, 4.6%, 0.55%, and 1.57%, respectively.

A fall detection system should avoid acquiring FP and FN results to obtain more reliable results. The experimental results demonstrate that the proposed method with fuzzy membership reduced false alarms and achieved accurate fall detection with better performance than traditional SVM.

TABLE VI.   EXPERIMENTAL RESULTS OF THE PROPOSED FUZZY SVM VS. CONVENTIONAL METHOD

Proposed method	SVM WITH FUZZY MFs			
Kernel function	Accuracy %	Sensitivity %	Specificity %	Precision %
Linear SVM	99.56	99.44	99.59	89.70
Cubic SVM	99.96	99.81	100	100
Quadratic SVM	99.96	99.81	100	100
Gaussian RBF	99.96	99.81	100	100
Proposed method	SVM WITHOUT FUZZY MFs			
Kernel function	Accuracy %	Sensitivity %	Specificity %	Precision %
Linear SVM	97.47	93.88	98.78	96.6
Cubic SVM	97.64	93.88	99.03	97.26
Quadratic SVM	97.78	94.71	98.91	96.95
Gaussian RBF	97.96	95.21	99.45	98.43

VII. CONCLUSION

In this paper, we introduced a new hybrid method integrating FL and SVM as a powerful technique for fall detection. The obtained results based on the defined environment showed a significant improvement in the accuracy of detection as compared to standalone, independent methods. Moreover, this paper examined the ways of identifying, expressing, and inspecting the relationship between disabled people suffering from traumatic brain injuries and falling events. Detecting falls accurately in time can reduce the severe consequences, especially since it can improve the quality of life of people with disabilities by promoting their independence. Besides, the prototype we examined was based on the Arduino platform, with the MPU-9250 sensor forming the part that was fixed on the patient's body and that wirelessly connects over the IoT platform via a low-energy Bluetooth interface. Thus, a 9-axis of

accelerometer, gyroscope, and magnetometer data formed the input matrix. The data were then smoothed using the moving average method to obtain clearer data and reduce the amplitude of the random noise. The standard deviation of each of the nine axes from the IMU sensor, depending on whether it was an accelerometer, gyroscope, or magnetometer, raw data in the x, y, and z planes in 3D space, was then computed. After that, the magnitude of standard deviation features was extracted in a way that makes it possible to differentiate between an actual fall and other activities, and thus to indicate abnormal activity in the pre-processing phase. Our proposed fall detection method is based on the effective combination of FL and SVM algorithms. In Fuzzy SVM, a fuzzy membership is given to each data point and set by expert experience. Different memberships reflect various contributions made to the decision-making learning process. SVM uses a high degree of membership functions extracted from selected features to automatically generate a model of the data and drive training, which enhances generalization ability and makes the heuristic obviousness of the traditional SVM efficient. The overall performance of the proposed fall detection system in terms of sensitivity and accuracy were 99.81% and 99.96%, respectively. This experiment achieved the maximum specificity and precision of 100% and, accordingly, proved to be effective and yielded better results than the conventionally used approaches. In other words, there was clear evidence that combining FL and SVM detects falls more accurately and performs better in reducing the effects of false positive alarms. This type of work is common abroad, yet it is still novel in Lebanon. So this study is considered a pioneering experience in fall detection systems for disabled patients. Furthermore, knowing that fall event datasets for this type of patients are not currently available in the literature, further trials must be conducted to simulate and refine the fall detection system and evaluate it outside the workplace. The application can be tested in real-world scenarios involving various types of disability. Our continuing research includes integrating fall detection algorithms into a smartphone that would be worn around the waist. This will be very useful to improve the robustness of our proposed method.

REFERENCES

[1] Step safely: strategies for preventing and managing falls across the life-course. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.

[2] N. Pannurat, S. Thiemjarus, and E. Nantajeewarawat, "Automatic fall monitoring: A review," Sensors (Switzerland), vol. 14, no. 7, pp. 12900–12936, 2014, doi: 10.3390/s140712900.

[3] Liu, Z. Jiang, X. Su, S. Benzoni, and A. Maxwell, "Detection of human fall using floor vibration and multi-features semi-supervised SVM," Sensors (Switzerland), vol. 19, no. 17, Sep. 2019, doi: 10.3390/s19173720.

[4] L. Ren and Y. Peng, "Research of fall detection and fall prevention technologies: A systematic review," IEEE Access, vol. 7. 2019. doi: 10.1109/ACCESS.2019.2922708.

[5] Z. Chen and Y. Wang, "Infrared–ultrasonic sensor fusion for support vector machine–based fall detection," J. Intell. Mater. Syst. Struct., vol. 29, no. 9, pp. 2027–2039, 2018, doi: 10.1177/1045389X18758183.

[6] R. Espinosa, H. Ponce, S. Gutiérrez, L. Martínez-Villaseñor, J. Brieva, and E. Moya-Albor, "A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection dataset," Comput. Biol. Med., vol. 115, 2019, doi: 10.1016/j.compbiomed.2019.103520.

[7] S. Sathyanarayana, R. K. Satzoda, S. Sathyanarayana, and S. Thambipillai, "Vision-based patient monitoring: a comprehensive review of algorithms and technologies," J. Ambient Intell. Humaniz. Comput., vol. 9, no. 2, 2018, doi: 10.1007/s12652-015-0328-1.

[8] N. Harum, Z. Z. Abidin, W. Shah, and A. Hassan, "Implementation of Smart Monitoring System with Fall Dectector for Elderly using IoT Technology," Int. J. Comput., vol. 17, no. 4, pp. 243–249, 2018.

[9] P. V. Er and K. K. Tan, Wearable solution for robust fall detection. INC, 2020. doi: 10.1016/b978-0-12-818546-9.00004-x.

[10] J. Alizadeh, M. Bogdan, J. Classen, and C. Fricke, "Support vector machine classifiers show high generalizability in automatic fall detection in older adults," Sensors, vol. 21, no. 21, pp. 1–14, 2021, doi: 10.3390/s21217166.

[11] T. Özdemir, "An analysis on sensor locations of the human body for wearable fall detection devices: Principles and practice," Sensors (Switzerland), vol. 16, no. 8, 2016, doi: 10.3390/s16081161.

[12] P. Tsinganos and A. Skodras, "On the comparison of wearable sensor data fusion to a single sensor machine learning technique in fall detection," Sensors (Switzerland), vol. 18, no. 2, 2018, doi: 10.3390/s18020592.

[13] O. Aziz, M. Musngi, E. J. Park, G. Mori, and S. N. Robinovitch, "A comparison of accuracy of fall detection algorithms (threshold-based vs. machine learning) using waist-mounted tri-axial accelerometer signals from a comprehensive set of falls and non-fall trials," Med. Biol. Eng. Comput., vol. 55, no. 1, 2017, doi: 10.1007/s11517-016-1504-y.

[14] M. Safeea and P. Neto, "Minimum distance calculation using laser scanner and IMUs for safe human-robot interaction," Robot. Comput. Integr. Manuf., vol. 58, 2019, doi: 10.1016/j.rcim.2019.01.008.

[15] M. Kangas, A. Konttila, P. Lindgren, I. Winblad, and T. Jämsä, "Comparison of low-complexity fall detection algorithms for body attached accelerometers," Gait Posture, vol. 28, no. 2, 2008, doi: 10.1016/j.gaitpost.2008.01.003.

[16] K. Bourke, J. V. O'Brien, and G. M. Lyons, "Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm," Gait Posture, vol. 26, no. 2, 2007, doi: 10.1016/j.gaitpost.2006.09.012.

[17] H. Ponce, L. Martínez-Villaseñor, and J. Nuñez-Martínez, "Sensor location analysis and minimal deployment for fall detection system," IEEE Access, vol. 8, pp. 166678–166691, 2020, doi: 10.1109/ACCESS.2020.3022971.

[18] N. Otanasap, "Pre-impact fall detection based on wearable device using dynamic threshold model," Parallel Distrib. Comput. Appl. Technol. PDCAT Proc., vol. 0, pp. 362–365, 2016, doi: 10.1109/PDCAT.2016.083.

[19] H. Wang, C. Ma, and L. Zhou, "A brief review of machine learning and its application," Proc. - 2009 Int. Conf. Inf. Eng. Comput. Sci. ICIECS 2009, 2009, doi: 10.1109/ICIECS.2009.5362936.

[20] M. Rahmani et al., "Machine learning (Ml) in medicine: Review, applications, and challenges," Mathematics, vol. 9, no. 22. 2021. doi: 10.3390/math9222970.

[21] Y. Hsieh, K. C. Liu, C. N. Huang, W. C. Chu, and C. T. Chan, "Novel hierarchical fall detection algorithm using a multiphase fall model," Sensors (Switzerland), vol. 17, no. 2, 2017, doi: 10.3390/s17020307.

[22] L. Palmerini, J. Klenk, C. Becker, and L. Chiari, "Accelerometer-based fall detection using machine learning: Training and testing on real-world falls," Sensors (Switzerland), vol. 20, no. 22, pp. 1–15, Nov. 2020, doi: 10.3390/s20226479.

[23] T. Özdemir and B. Barshan, "Detecting falls with wearable sensors using machine learning techniques," Sensors (Switzerland), vol. 14, no. 6, pp. 10691–10708, 2014, doi: 10.3390/s140610691.

[24] Hussain, F. Hussain, M. Ehatisham-Ul-Haq, and M. A. Azam, "Activity-Aware Fall Detection and Recognition Based on Wearable Sensors," IEEE Sens. J., vol. 19, no. 12, 2019, doi: 10.1109/JSEN.2019.2898891.

[25] D. Pan, H. Liu, D. Qu, and Z. Zhang, "Human Falling Detection Algorithm Based on Multisensor Data Fusion with SVM," Mob. Inf. Syst., vol. 2020, 2020, doi: 10.1155/2020/8826088.

[26] Şengül, E. Ozcelik, S. Misra, R. Damaševičius, and R. Maskeliūnas, "Fusion of smartphone sensor data for classification of daily user activities," Multimed. Tools Appl., vol. 80, no. 24, pp. 33527–33546, 2021, doi: 10.1007/s11042-021-11105-6.

[27] Z. A. Kakarash, S. H. T. Karim, and M. Mohammadi, "Fall Detection Using Neural Network Based on Internet of Things Streaming Data," UHD J. Sci. Technol., vol. 4, no. 2, pp. 91–98, Oct. 2020, doi: 10.21928/uhdjst.v4n2y2020.pp91-98.

[28] T. De Quadros, A. E. Lazzaretti, and F. K. Schneider, "A Movement Decomposition and Machine Learning-Based Fall Detection System Using Wrist Wearable Device," IEEE Sens. J., vol. 18, no. 12, 2018, doi: 10.1109/JSEN.2018.2829815.

[29] Y. Zhang and L. Wu, "Classification of fruits using computer vision and a multiclass support vector machine," Sensors (Switzerland), vol. 12, no. 9, pp. 12489–12505, 2012, doi: 10.3390/s120912489.

[30] Rescio, A. Leone, and P. Siciliano, "Supervised expert system for wearable MEMS accelerometer-based fall detector," J. Sensors, vol. 2013, 2013, doi: 10.1155/2013/254629.

[31] S. H. Liu and W. C. Cheng, "Fall detection with the support vector machine during scripted and continuous unscripted activities," Sensors (Switzerland), vol. 12, no. 9, pp. 12301–12316, 2012, doi: 10.3390/s120912301.

[32] R. Amami, D. Ben Ayed, and N. Ellouze, "Practical Selection of SVM Supervised Parameters with Different Feature Representations for Vowel Recognition," vol. 7, no. May, pp. 418–424, 2015, doi: 10.4156/jdcta.vol7.issue9.50.

[33] W. Min, L. Yao, Z. Lin, and L. Liu, "Support vector machine approach to fall recognition based on simplified expression of human skeleton action and fast detection of start key frame using torso angle," IET Comput. Vis., vol. 12, no. 8, 2018, doi: 10.1049/iet-cvi.2018.5324.

[34] X. Zhuang, X. Zhou, M. A. Hasegawa-Johnson, and T. S. Huang, "Real-world acoustic event detection," Pattern Recognit. Lett., vol. 31, no. 12, 2010, doi: 10.1016/j.patrec.2010.02.005.

[35] T. Zhen, L. Mao, J. Wang, and Q. Gao, "Wearable preimpact fall detector using SVM," 2016. doi: 10.1109/ICSensT.2016.7796223.

[36] Shahzad and K. Kim, "FallDroid: An Automated Smart-Phone-Based Fall Detection System Using Multiple Kernel Learning," IEEE Trans. Ind. Informatics, vol. 15, no. 1, pp. 35–44, 2019, doi: 10.1109/TII.2018.2839749.

[37] Lee and S. Kim, "Black-Box Classifier Interpretation Using Decision Tree and Fuzzy Logic-Based Classifier Implementation," Int. J. Fuzzy Log. Intell. Syst., vol. 16, no. 1, pp. 27–35, 2016, doi: 10.5391/ijfis.2016.16.1.27.

[38] M. H. Azam, M. H. Hasan, S. Jadid, A. Kadir, and S. Hassan, "Prediction of Sunspots using Fuzzy Logic : A Triangular Membership Function-based Fuzzy C-Means Approach," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 2, pp. 357–362, 2021.

[39] B. Pękala, T. Mroczek, D. Gil, and M. Kepski, "Application of Fuzzy and Rough Logic to Posture Recognition in Fall Detection System," Sensors, vol. 22, no. 4, Feb. 2022, doi: 10.3390/s22041602.

[40] M. Casal-Guisande, A. Comesaña-Campos, I. Dutra, J. Cerqueiro-Pequeño, and J. B. Bouza-Rodríguez, "Design and Development of an Intelligent Clinical Decision Support System Applied to the Evaluation of Breast Cancer Risk," J. Pers. Med., vol. 12, no. 2, 2022, doi: 10.3390/jpm12020169.

[41] B. Pandya, A. Pourabdollah, and A. Lotfi, "Comparative Analysis of Real-Time Fall Detection Using Fuzzy Logic Web Services and Machine Learning," Technologies, vol. 8, no. 4, p. 74, 2020, doi: 10.3390/technologies8040074.

[42] Fernández-Caballero, M. V. Sokolova, and J. Serrano-Cuerda, "Lateral inhibition in accumulative computation and fuzzy sets for human fall pattern recognition in colour and infrared imagery," Sci. World J., vol. 2013, 2013, doi: 10.1155/2013/935026.

[43] S. Moulik and S. Majumdar, "FallSense: An Automatic Fall Detection and Alarm Generation System in IoT-Enabled Environment," IEEE Sens. J., vol. 19, no. 19, 2019, doi: 10.1109/JSEN.2018.2880739.

[44] Zhang, M. Alrifaai, K. Zhou, and H. Hu, "A novel fuzzy logic algorithm for accurate fall detection of smart wristband," Trans. Inst. Meas. Control, vol. 42, no. 4, pp. 786–794, Feb. 2020, doi: 10.1177/0142331219881578.

[45] C. F. Lin and S. De Wang, "Fuzzy support vector machines," IEEE Trans. Neural Networks, vol. 13, no. 2, pp. 464–471, 2002, doi: 10.1109/72.991432.

[46] E. Spyrou, G. Stamou, Y. Avrithis, and S. Kollias, "Fuzzy support vector machines for image classification fusing MPEG-7 visual descriptors," IET Semin. Dig., vol. 2005, no. 11099, pp. 23–30, 2005, doi: 10.1049/ic.2005.0706.

[47] H. Tang, Y. Liao, F. Sun, and H. Xie, "Fuzzy support vector machine with a new fuzzy membership function," Hsi-An Chiao Tung Ta Hsueh/Journal Xi'an Jiaotong Univ., vol. 43, no. 7, pp. 40–43, 2009.

[48] C. Yang, S. K. Oh, B. Yang, W. Pedrycz, and Z. W. Fu, "Fuzzy quasi-linear SVM classifier: Design and analysis," Fuzzy Sets Syst., vol. 413, pp. 42–63, 2021, doi: 10.1016/j.fss.2020.05.010.

[49] M. A. Nanda, K. B. Seminar, D. Nandika, and A. Maddu, "A comparison study of kernel functions in the support vector machine and its application for termite detection," Inf., vol. 9, no. 1, 2018, doi: 10.3390/info9010005.

[50] Fleury, M. Vacher, and N. Noury, "SVM-based multimodal classification of activities of daily living in health smart homes: Sensors, algorithms, and first experimental results," IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 2, pp. 274–283, 2010, doi: 10.1109/TITB.2009.2037317.

[51] E. A. Zanaty and A. Afifi, "Support Vector Machines (SVMs) with Universal Kernels," Appl. Artif. Intell., vol. 9514, 2011, doi: 10.1080/08839514.2011.595280.

[52] Sulistiana and M. A. Muslim, "Support Vector Machine (SVM) Optimization Using Grid Search and Unigram to Improve E-Commerce Review Accuracy," J. Soft Comput. Explor., vol. 1, no. 1, pp. 8–15, 2020.

[53] Akshay and M. Sumanth, "Bluetooth based home automation by using arduino.," Grad. Res. Eng. Technol., vol. 7, no. 02, pp. 37–42, 2021, doi: 10.47893/gret.2021.1116.

[54] M. Tolkiehn, L. Atallah, B. Lo, and G. Z. Yang, "Direction sensitive fall detection using a triaxial accelerometer and a barometric pressure sensor," Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, no. August 2011, pp. 369–372, 2011, doi: 10.1109/IEMBS.2011.6090120.

[55] Q. Li and J. A. Stankovic, "Detection for Falls with Different Activity Levels," Wirel. Heal. 11, 2011.

[56] H. Gjoreski, M. Luštrek, and M. Gams, "Accelerometer placement for posture recognition and fall detection," Proc. - 2011 7th Int. Conf. Intell. Environ. IE 2011, no. August, pp. 47–54, 2011, doi: 10.1109/IE.2011.11.

[57] Zhang and A. S. Zhang, "Real-Time Wildfire Detection and Alerting with a Novel Machine Learning Approach: A New Systematic Approach of Using Convolutional Neural Network (CNN) to Achieve Higher Accuracy in Automation," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 8, pp. 1–6, 2022, doi: 10.14569/IJACSA.2022.0130801.

[58] Saha, H. K. Tripathy, F. Masmoudi, and T. Gaber, "A Machine Learning Model for Personalized Tariff Plan b ased on Customer ' s Behavior in the Telecom Industry," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 10, pp. 171–184, 2022.

An Effective Ensemble-based Framework for Outlier Detection in Evolving Data Streams

Asmaa F. Hassan, Sherif Barakat, Amira Rezk

Department of Information Systems
Faculty of Computers and Information
Mansoura University
Cairo, Egypt

Abstract—**In the last few years, data streams have drawn lots of researchers' attention due to their various applications, such as healthcare monitoring systems, fraud and intrusion detection, the internet of things (IoT), and financial market applications. A data stream is an unbounded sequence of data continually generated over time and is prone to evolution. Outliers in streaming data are the elements that significantly deviate from the majority of elements and then have to be detected as they may be error values or events of interest. Detection of outliers is a challenging issue in streaming data and is one of the most crucial tasks in data stream mining. Existing outlier detection methods for static data are unsuitable for use in data stream settings due to the unique characteristics of streaming data such as unpredictability, uncertainty, high-dimensionality, and changes in data distribution. Thus, in this paper, a novel ensemble learning framework called** <u>E</u>**nsemble-based** <u>S</u>**treaming** <u>O</u>**utlier** <u>D</u>**etection (ESOD) is presented to perfectly detect outliers over streaming data using a sliding window technique that is updated in response to the incoming events from the data streaming environment to overcome the concept evolution nature of streaming data. The proposed framework has three phases, namely the training phase, testing/offline phase, and outlier detection/online phase. A detection weighted vote technique is used to determine the final decisions for potential outliers. In the extensive experimental study, which was conducted on 11 real-world benchmark datasets, the proposed framework was assessed using many accuracy metrics. The experiment results showed that the proposed framework beats many other state-of-the-art methods.**

Keywords—Outlier detection; data streams; data stream mining; ensemble learning; concept evolution

I. INTRODUCTION

Due to the recent advances in both software and hardware, many applications are generating streaming data, such as sensor networks, financial markets, real-time video monitoring, internet traffic, and medical data. The term "data stream" refers to a collection of temporally ordered, massive, usually arriving at a high rate, and potentially infinite data objects. A data stream can be formalized as $DS_t = \{x_{1,t}, x_{2,t}, x_{3,t}, ..., x_{N,t}\}$, where $x_{i,t}$ is the element number i at time t, and it has a set of high-dimensionality attributes or features. Through its large volume, it is therefore difficult to fully store data streams in memory and scan them several times [1]. Data streams differ from static data where they have some unique characteristics such as *concept-evolution*, *concept-drift*, and *feature-evolution*. In particular, concept-evolution happens

when new classes emerge in streams, concept-drift occurs when the distribution of data points shifts over time, and feature-evolution occurs when the feature set of data streams changes over time [2]. Data Stream Mining (DSM) is a new approach for extracting important information from data streams [3]. Hence, the traditional data mining techniques are not applicable to processing data streams because of the special characteristics of streaming data as shown in Table I.

Data streams, like traditional data, may have outliers, or data points that are considerably different from the bulk of data points [4]. They can be noise data points or interesting instances and have to be detected in many cases to achieve better performance and accuracy. Outlier detection is one of the most important data mining tasks for detecting unusual and anomalous data points or sequences hidden in a dataset [5]. However, outlier detection over data stream datasets completely differs from traditional data ones because it must be performed under only one pass, the available memory is limited, real-time response, and the concept-evolution nature of streaming data. In real life, outlier detection has a variety of essential applications, such as detecting credit card fraud; intrusion detection in computer networks or cybersecurity; system fault diagnosis in industry; early disease detection in the health care sector, etc. [6]. Outlier detection techniques may be based on one of the following learning methods: unsupervised, semi-supervised, supervised, or ensemble learning. The unsupervised learning technique does not need training data to build the model, while the semi-supervised learning technique combines a small collection of labelled data with a large dataset of unlabeled training data. In the supervised learning approach, it requires labelled training data availability [4]. On the other hand, the ensemble learning approach requires a group of multiple trained classifiers learning algorithms to detect outliers in order to improve the detection accuracy. In particular, a number of different base learners are used in an ensemble model, which is normally much stronger than all standalone base learners because it has the ability to improve the performance of weak learners [5]. For this purpose, this paper presents an effective outlier detection approach based on the ensemble learning technique where the iForestASD, decision tree, and Adaptive Random Forest (ARF) classifiers are used as the base learners to build an ensemble-based model over streaming data using the sliding window fashion with the objective of improving detection performance while decreasing detection time consumption. In the data streams environment, the sliding

window is a time-based streaming model which is generally used to effectively process the flow of streaming data through dividing these streams into several windows, as shown in Fig. 1, where each window has an equal pre-defined size *(w)* in time *(t)*. The size of a window may be specified in terms of time points or the number of recent objects. Hence, each window maintains only recent objects, while older ones are discarded, and all objects within the active sliding window have the same importance.

TABLE I. COMPARISON BETWEEN STREAMING DATA AND STATIC DATA

Characteristic	Streaming Data	Static Data
Volume	Infinite	Finite
Type of Data	Heterogeneous	Homogeneous
Scanning Time	Single pass	Multiple passes
Data Processing	Real-time processing	Offline processing
Data Storage	Aggregated and summarized data only	Raw data
Concept	Evolving	Static
Type of Result	Approximate result	Accurate result
Temporal and Spatial Contexts	Important aspects	May be considered for certain applications
Space and Time Complexity	Strict	Not strict
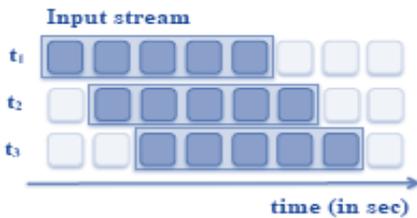
Fig. 1.   The Sliding Window Time-based Model.

To detect outlier values, a combination of different base detectors; iForestASD, decision tree, and adaptive random forest algorithms; were chosen to develop an ensemble-based learning framework, which has never been explored before in the literature. The iForestASD model can be retrained more quickly than models trained with higher complexity like SVM. IForestASD allows you to retrain the model with a sliding window so that detection performance does not degrade even with concept evolution. While the decision tree technique does not need data normalization or scaling, and it has a shorter training period. Furthermore, random forest has high learning performance with good data visualization for high dimensional data and does not require hyper-parameter tuning or tree pruning. Moreover, it enables rapid detection model training even with limited computational resources. This is the main motivation for choosing this combination of base learners to construct our ensemble-based learning model for streaming outlier detection. In streaming data, the proposed framework (ESOD) attempts to detect both outliers and new class concepts. ESOD provides a more effective solution to this problem by integrating active learning in a supervised approach to detect any novel concepts and outliers in the

streaming environment. The main research contributions and novelty are summarized in the following points:

- Proposing an effective Ensemble-based Streaming Outlier Detection (ESOD) framework for accurately detecting outliers in the context of streaming data adapting to the concept drift nature of streaming data.

- Building an ensemble-based model for the proposed framework using some heterogeneous machine learning algorithms as base learners, such as iForestASD, decision tree, and adaptive random forest, and then selecting the best combination.

- Using the sliding window time-based technique in the ESOD framework to process the flow of streaming data.

- Reducing computation time by allowing parameter tuning and comparing the outcomes of multiple algorithms in a single trial.

- Providing a detection weighted vote technique to report the potential outliers that outperforms the standalone base learners and state-of-the-art techniques in terms of accuracy, precision and other evaluation metrics.

The remainder of this paper is structured in the following way: related work on outlier detection techniques in streaming data is introduced section II. Section III explains the preliminaries of the used base learners. Section IV introduces the proposed framework. Section V discusses the experimental evaluations and the model development details. Finally, Section VI summarizes the proposed study and the next directions are given.

II.   RELATED WORK

Many attempts in the literature have been devoted to the outlier detection problem over streaming data. Nevertheless, many of these detection approaches work in batch mode, in which all data points are stored and many passes can be made over the data [7-9]. Several batch outlier detection algorithms have been adapted for data stream outlier detection, according to a review in [10]. These methods, while applicable to streaming data, are inefficient because they do not take into consideration the special features and characteristics of data streams. The outlier detection approaches in the streaming data context may be divided into three main categories, which are statistical-based, clustering-based, and classification-based methods [11, 12]. The first attempts at outlier detection were statistical-based approaches based on defining a model to represent the normal behavior of the data [13]. If an incoming data point does not fit the model or has an extremely low probability of fitting the model, it is termed an outlier. In [14], The authors proposed an algorithm called UKOF for top-n local outlier detection based on the kernel density estimation (KDE) model over large-scale high-volume data streams, in which they defined a KDE-based outlier factor (KOF) to measure the local outlierness score, as well as upper bounds of the KOF and an upper-bound-based pruning strategy to reduce the search space. Although this method had a low computational cost, it assumed a stable distribution of data, which is inappropriate in a data stream setting. In recent

research [15], a maximal weighted frequent-pattern-based outlier detection approach (MWFP-Outlier) was proposed to detect potential outliers from uncertain data streams. This approach had two phases, namely pattern mining phase and an outlier detection phase. The MWFP-Mine approach was proposed in the pattern mining phase to mine maximal weighted frequent patterns using the list structure, tree structure, and pruning strategies by fully considering the existential probabilities and weights for each pattern. During the outlier detection phase, four deviation indices were created to measure the degree of deviation, and then the top k ranked transactions were reported as potential outliers. It is unclear, however, how it dealt with outliers and fast processing.

The clustering-based methods are one of the common unsupervised methods used in outlier detection problems, which creates data clusters that represent the underlying data distribution [16]. Clustering-based methods divide the data into some clusters according to the similarity between the data points. Outliers are observations that are remote from clusters or clusters with significantly fewer data points. The clustering methods are categorized into distance-based and density-based. In 2019 [17], Tran et al. proposed three distance-based algorithms using the micro-clusters concept, and they claimed that the proposed algorithms had a reasonable processing time and a low space cost. Regarding the density-based methods, there are some state-of-the-art methods developed to detect outliers in data streams, such as [18–20]. Another notable attempt was presented in [21], a method called LiCS was introduced to detect outlier that classifies the samples using K-nearest neighbors of each node. Most recently, [22] proposed an incremental local density and cluster-based outlier factor method for detecting outliers in streaming data called iLDCBOF. The proposed method combined density-based spatial clustering of applications with noise (DBSCAN) and incremental versions of the local outlier factor (LOF), but it suffered from excessive computations. On the other hand, classification-based methods may either be incremental single model or ensemble classifier techniques, where the classification output is a function of the predictions from different classifiers. In addition, streaming ensemble-based methods have been developed to deal with the high dimensionality of data streams, and these techniques have great success in the outlier detection and prediction domains because of their accurate results and higher efficiency in both performance and resource consumption. Ensemble-based techniques integrate the results of many base models to create a more robust model that can detect outliers more effectively. Masud et al. [23] proposed a hyper outlier detection model for streaming data. They utilized the k-NN classifier first, followed by the SVM polynomial kernel classifier, and a data point is reported as an outlier if it is identified as an outlier by at least one of the classifiers. Then, to distinguish between novel and outlier data points, a neighborhood silhouette coefficient is applied. This approach addresses the novelty detection in multi-class underlying concepts, yet it requires all data chunks to be labelled to define the new concept. Wang et al. [24] proposed a model for detecting anomalies based on a matrix of uncompressed data against a matrix of compressed data, and this model utilized the original uncompressed data while considerably reducing computing costs. In [25], a new

ensemble approach called RED-PSO was presented, which is suited to the drift notion of a data stream in the classification of non-stationary data streams. In [26], the authors proposed a sliding window-based ensemble approach for streaming outlier detection where a combination of clustering algorithms were utilized to construct clusters, which were later used in a one-class classification to identify outlier data. Togbe et al. [12] afforded iForestASD, i.e., an alternative isolation forest for streaming data implementation. They built the iForestASD on the scikit-multiflow framework; it is a machine learning framework for streaming data that is free and open source. In 2021 [27], they modified their implementation of iForestASD to address the idea of data stream drift, proposing three algorithms for drift detection: ADWIN, KSWIN, and finally extending KSWIN to deal with n-dimensional data streams. Another recent study was presented in [28], where the authors proposed an ensemble-based outlier detection framework for high-dimensional data named Average Selection and Ensemble of Candidates for Outlier Detection (ASEC-OD). The proposed framework selected the most effective base outlier detectors, which have the highest performance. To summarize, none of these approaches can detect outliers accurately and efficiently in streaming data environments while maintaining high performance rates. Our approach differs from the aforementioned outlier methods in two ways. First, the proposed framework integrates outlier detection and the concept evolution phenomenon of streaming data by considering concept evolution for new outlier or inlier concepts that appear in incoming streams. It can sequentially update the outlier detection model in the case of concept evolution, resulting in a more robust outlier detection system that is matched with a streaming scenario. Second, it takes advantage of each heterogeneous base classifier in building an ensemble model to achieve the highest rates of performance metrics.

III. BASE LEARNERS

The outlier ensemble learning method will be very effective when heterogeneous base classifiers of diverse types are used. Thus, the distinct features of data can be identified or learned due to the variation between classifiers [29]. To boost efficiency, the proposed ensemble model constructs a set of models from three heterogeneous base learners. The chosen base learners (isolation forest, its variant iForestASD, decision tree, random forest, and its variant adaptive random forest) of the proposed framework are briefly introduced in this section.

A. Isolation Forest (iForest) and IForestASD

Isolation Forest (iForest) is a famous tree-based approach to outlier detection to isolate outlier instances [30]. The random selection of an attribute iteratively creates an isolation tree *(itree)* until each data point is isolated. Isolation forest is based on the notion that it is simple to isolate an outlier but more complex to characterize an inlier data point. Accordingly, the isolation forest consists of all the isolation trees constructed from its training set subsampling. The outlier score $s(x, n)$, equation (1), is calculated by calculating the path length $h(x)$ from the root to an instance x in an isolation tree constructed from n data instances [31].

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{1}$$

where $E(h(x))$; which is defined by equation (2); is the average path length $h(x)$ of x over an isolation trees collection, i.e., *itrees*.

$$E(h(x)) = \frac{\sum_{i=1}^{t} h_i(x)}{t} \tag{2}$$

where $c(n)$ represents the average path length of the binary search tree's unsuccessful search process. Finally, equation (3) calculates $c(n)$, which is used to normalize an outlier score.

$$(n) = 2H(n-1) - (2(n-1)/n) \tag{3}$$

where $H(i)$ is formulated by equation (4).

$$H(i) = \ln(i) + \gamma \text{ , } \gamma \text{ is the Euler's constant} \tag{4}$$

Therefore, x is considered as outlier data if the outlier score is close to 1, but it is considered as normal data if the score is less than 0.5.

On the other hand, the iForestASD [32] is an alternative version of the iForest method and is suitable for outlier detection over streaming data. The phrase "iForestASD" is an abbreviation for Isolation Forest Algorithm for Stream Data, where it uses the sliding window mechanism to process streaming data. In addition, iForestASD implemented the standard isolation forest technique to build the random forest. The IForestASD technique was developed to deal with the concept drift of the data stream by keeping a pre-defined anomaly rate threshold *(u)*. Fig. 2 displays the workflow of the iForestASD approach to updating the model. If the anomaly rate in the active window is higher than u, then concept drift has occurred. The iForestASD then deletes the current detector and constructs another detector using all the data in the active window. Otherwise, there was no concept drift. The trained anomaly detector was then would not change.
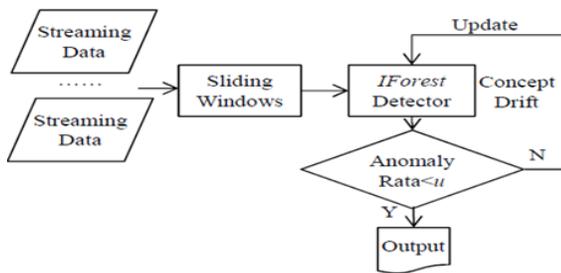
Fig. 2. The iForestASD Workflow. The Image is Reproduced from [32].

B. Decision Tree (DT)

Decision trees are popular tree-based classification methods where each non-leaf node represents an attribute, whereas the edge links between its child nodes represent the attribute values and the leaf node represents a class label. In general, a decision tree is built in three main steps:

- Select the best features of the training dataset to build the root node.
- Split the training dataset into nodes using the Gini index.

- Loop until no further splitting can be done.

Information gain, gain ratio, and Gini index are used to partition the training dataset [33]. In this research, the Gini index is used as the DT splitting criterion, which is calculated by equation (5).

$$G(X) = 1 - \sum_{i=1}^{n} p_i^2 \tag{5}$$

where p_i is the proportion of the ith class samples of the feature X.

C. Random Forest and Adaptive Random Forest

Firstly, the classical random forest (RF) [34] method is a tree-based approach that can be used for both classification and prediction tasks. The prediction of a new data point is determined by aggregating the predictions of n trees in classification tasks. Random predictors are chosen to split a node in random forests; therefore, every tree is grown based on a different random data sample. After multiple models have been generated, the overall trees of the model contribute to a weighted decision or vote to produce an overall determination. Regarding randomness splitting of the data, RF has two main sources of randomness which are bootstrap aggregating or bagging and boosting. In bootstrap aggregating, each classifier is trained on random sampling with replacement from the original dataset. On the other hand, in boosting approach, instead of evaluating all features at each tree split, only a random set of samples is used to make the eventual decision. RFs can model high dimensional data and they can handle missing values.

Adaptive Random Forest (ARF) is one variant technique that adapts the standard random forest algorithm [35]. It produces decision trees by training them on resampled varieties of the original data and by randomly picking up a small number of features that can be inspected at each node for splitting. ARF offers data visualization of high dimensional data and it has a high accuracy and resistance to overfitting.

IV. PROPOSED ENSEMBLE-BASED FRAMEWORK

In this section, a novel ensemble-based outlier detection framework for data streams called Ensemble-based Streaming Outlier Detection (ESOD) is introduced. The proposed framework utilizes three heterogeneous machine learning algorithms, which are iForestASD, decision tree (DT), and adaptive random forest (ARF) as the base learners to build the ensemble model for detecting outliers in streaming data based on a weighted voting technique. To enhance efficiency, the proposed framework employs various base learners to build an ensemble-based model. Ensemble-based Streaming Outlier Detection (ESOD) has three main phases, as shown in Fig. 3, namely the training phase, testing/offline phase, and outlier detection/online phase. These phases are detailed in depth in the subsections that follow.

A. The Training Phase

The ESOD training phase is used to prepare data and tune the hyper-parameters of each classifier for the finest outcome. It consists of a series of steps that must be completed in the following order:
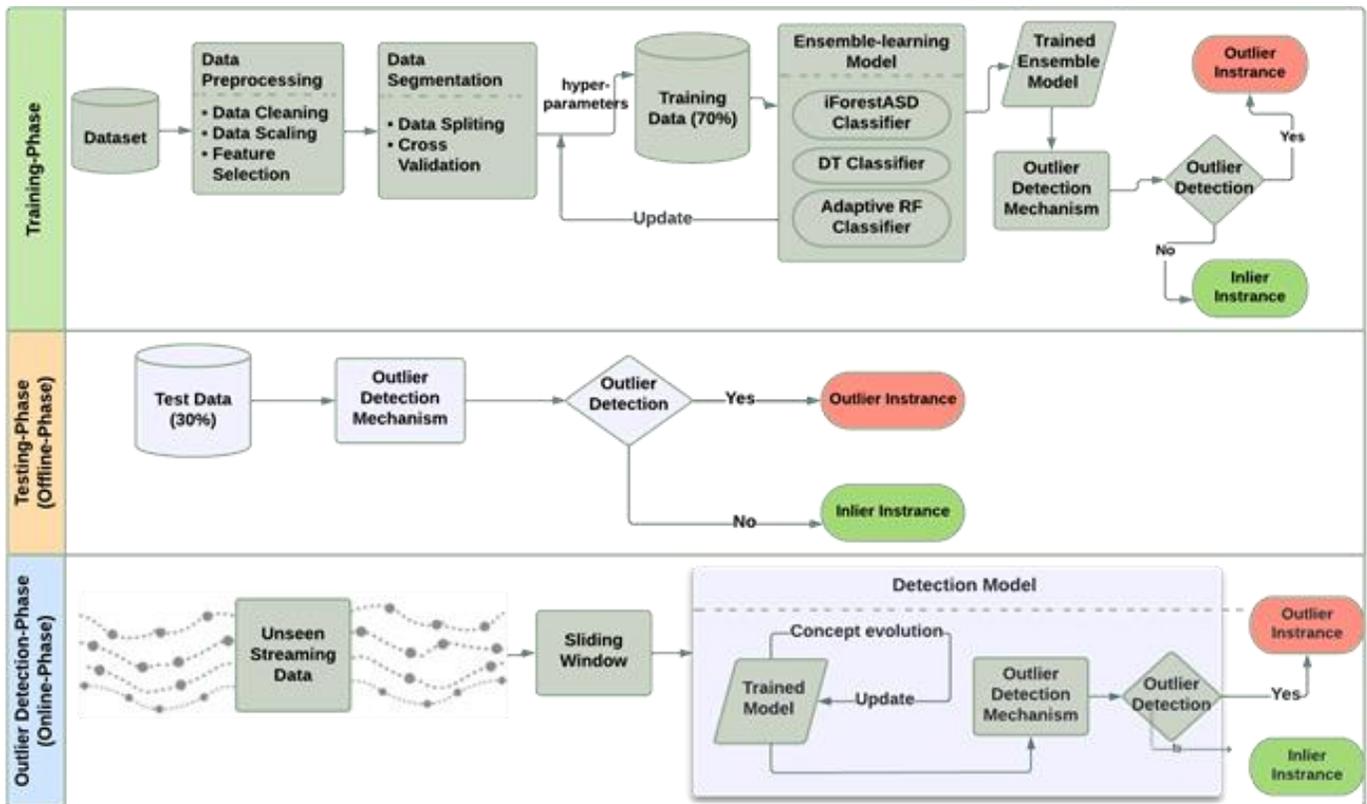
Fig. 3.   The Ensemble-based Streaming Outlier Detection (ESOD) Block Diagram.

1) Data preprocessing step: Certain data preprocessing techniques are done in order to turn raw data streams into meaningful patterns. The preprocessing stage is a critical step for the proper representation of streaming data. Hence, many different preprocessing techniques are employed, such as the removal of missing values, min-max scalar and one-hot encoding, for a more efficient classification process. There are three further phases in the ESOD's preprocessing stage, which are the *data cleaning*, *data scaling*, and *feature selection* phases. In the *data cleaning* phase, missing data is first examined and eliminated if it exists. The *data scaling* phase then employs the one-hot encoding approach on nominal or non-binary categorical data, and the min-max scaler is employed to scale values into the range of 0-1. Finally, the *feature selection* phase, where the main goal is to extract the set of relevant and non-redundant features, makes the learning model more meaningful and quicker because not all features are important for the model's learning process. In this study, the Least Absolute Shrinkage and Selection Operator (LASSO) technique is used to choose the most important and correlated features from the feature space that have a strict reflection on the target.

2) Data segmentation step: In the data segmentation step, the preprocessed data D of n selected features is divided into two groups: training set X_{train} with m data points of ratio 70% where $X_{train} \in D^{m \times n}$ and testing set X_{test} with r data points of ration 30% where $X_{test} \in D^{r \times n}$. The testing set, in contrast to the training set, should be unlabeled. Then, the K-fold cross

validation (CV) method is used, in which the training set is divided into k equal non-overlapping subsets, determined by random sampling or the bootstrap mechanism, and the model is tested on different subsets of the dataset. Fig. 4 depicts the K-fold cross validation process. If $k = 10$, for example, nine groups of the sample data are used to train the model and only one group is used to test the model at each fold, where the model parameters with the lowest mean squared error from the k training models are chosen for the final model. For each fold, the outlier detection models are trained with only data from the inlier class, so the model is trained k times. The k-fold cross validation approach is used in this study to choose the model hyper-parameters for setting each model as well as the configured models.
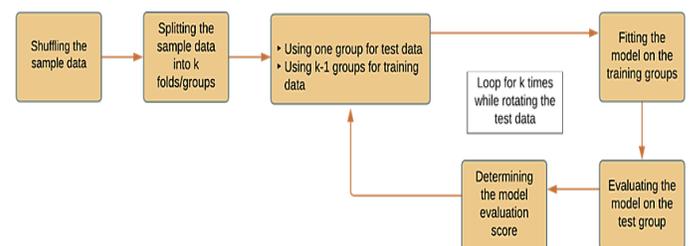
Fig. 4.   The K-Fold Cross Validation Process.

3) Ensemble-learning model step: The heterogeneous base classifiers $B = \{\beta_1, \beta_2, ..., \beta_i\}$; iForestASD, decision tree, and ARF classifiers are built in this step, employing many hyper-parameter settings that are continually updated to

acquire the optimal parameters of each classifier. All three classifiers were trained using K random data samples (s_1, s_2,… s_K) and were then used to classify outliers of the test set X_{test} in the testing phase. Specifically, the outlier detection approach ESOD uses similar insights as in the random forest approach, where all these base learners are considered as trees for the ensemble model. In more detail, the outlier detection final decision is made by the weighted voting of these trees where it is the class label that has the highest weighted vote of classifiers predicted. After constructing the base learning models, the ensemble learning model $\psi_{\beta1, \beta2,…, \beta i}$ tries to combine their outlier results to declare the final decision by estimating the weights of individual base learning models β_1, β_2, … , β_i. to aggregate their results with a weighted combination. The ESOD weighted voting approach involves assigning various weights to each classifier depending on its performance accuracy [36], calculated by equation (6), and voting on the classifiers based on this weight. After calculating the weight for each classifier, the most votes of classification results are selected as the ensemble's final detection result.

$$\omega_{\beta i} = \frac{A_{\beta i}}{\sum_i^3 A_i} \tag{6}$$

where $\omega_{\beta i}$ is the weight of classifier β_i, $A_{\beta i}$ is the classifier β_i accuracy, and A_i is the accuracy summation of the three base classifiers used in the ensemble model. The trained ensemble model has been constructed at this point and can classify a data point as an inlier or outlier.

B. The Testing/Offline Phase

The test set was made up of the remaining 30% of the whole dataset. This test data is used as a final evaluation of an unseen dataset in the offline phase to ensure that the trained model was appropriately trained and predicts correctly. Hence, any unseen test sample x is classified into inlier or outlier instances using these trained base classifiers based on the above detection voting technique.

C. The Outlier/Online Detection Phase

During the outlier detection phase, i.e., the online phase, a flow of fresh unseen data streams is generated and processed using the sliding-time window technique. As a result of the ever-changing nature of stream data, detection mechanism updates are frequently based on time intervals in the form of sliding windows, where current items are given higher priority during the detection phase than older ones. The trained model can classify the known classes on which it was trained, identify any novel classes, and finally update the model with the most recent data to include any detected outliers. The model, in particular, tries to classify incoming samples of streaming data into known classes, which are outliers or inliers, or novel patterns. Hence, it predicts class labels for instances that belong to known classes, and the novel patterns are identified as "*concept evolution*". Once the new class is detected, it is integrated into the model for detecting repeated patterns. Thus, the trained model is automatically updated to handle the concept evolution by comparing the mean difference of two adjacent sub-windows [37].

V. EXPERIMENTAL EVALUATION

This section presents the datasets, experimental setup, and parameter fine-tunning process followed by the assessment metrics and a brief discussion of the results.

A. Dataset Description

This research uses 11 publicly available outlier benchmark datasets (http, Credit-card-fraud, smtp, Annthyroid, Thyroid, Cardio, Pima Diabetes, Breast-cancer, Arrhythmia, Heart Disease, Hepatitis) from the UC Irvine Machine Learning Repository [38] for the evaluation of the base learner classifiers, state-of-the-art methods, and proposed framework. Description of these datasets is summarized in Table II by name, number of points, dimensions, outlier samples, percent of outliers, outlier class, and application domain of the datasets.

TABLE II. DESCRIPTION OF THE BENCHMARK DATASETS USED IN THE PERFORMANCE EVALUATION

Dataset	Samples (m)	Features (n)	Outliers (%)	Outlier Class	Application Domain
http	567,479	22	2,211 (0.39%)	Attack	Intrusion detection
Credit-card-fraud	284,807	30	492 (0.17%)	Fraudulent	Fraud detection
smtp	95,156	22	29 (0.03%)	Attack	Intrusion detection
Annthyroid	7,200	6	534 (7.42%)	Hyperfunction and Subnormal functioning	Disease detection
Thyroid	3,772	6	93 (2.47%)	Hyperfunction and Subnormal functioning	Disease detection
Cardio	1,831	21	176 (9.61%)	Pathologic	Disease detection
Pima Diabetes	768	9	193 (3.98%)	Diabetes	Disease detection
Breast-cancer	683	30	239 (34.99%)	Malignant	Disease detection
Arrhythmia	452	279	66 (14.60%)	The smallest classes (3, 4, 5, 7, 8, 9, 14, 15)	Disease detection
Heart Disease	297	13	137 (46%)	Class 0	Disease detection
Hepatitis	155	19	32 (20.65%)	Die	Disease detection

TABLE III.   HYPER-PARAMETER SETTINGS FOR DIFFERENT BASE LEARNER CLASSIFIERS

Classifier	Hyper-parameters Values
iForestASD	n_estimators = 100 contamination = 0.2 bootstrap = True n_jobs = -1 random_state = 42
Decision Tree (DT)	criterion = gini n_estimators = 100 Min_Samples_Split = 2 Splitter = best max_features = log2 max_depth = 6
Adaptive Random Forest (ARF)	n_models = 100 max_features = sqrt aggregation_method = mean lambda_value = 6 drift_detector = ADWIN warning_detector = ADWIN δw (warning threshold) = 0.01 δd (drift threshold) = 0.001
K-nearest neighbors (K-NN)	n_neighbors = 7 weights = distance metric = Euclidean
Logistic Regression (LR)	Penalty = l2 solver = saga dual = True tol = 0.01 random_state = 42 max_iter = 1000
SVM	Regularization parameter = 1 kernel = sigmoid gamma = scale shrinking = True coef0 = 0.01
GaussianNB (GNB)	priors = none var_smoothing = 0.001
Artificial Neural Network (ANN)	hidden_layer = 4 activation = relu optimizer = adam batch_size = 32 loss = binary_crossentropy epochs = 100
Gradient Boosting (GB)	learning_rate = 0.1 n_estimators=100 min_samples_split = 500 min_samples_leaf = 50 max_depth = 5 max_features = sqrt subsample = 0.8
Local Outlier Factor (LOF)	n_neighbors = 7 algorithm = ball_tree leaf_size = 30 metric = minkowski p = 2

B. Experimental Setup

All the developed models are implemented in the open-source Python programming language using the Anaconda distribution API. The main libraries and packages used include pandas, NumPy, SciPy, matplotlib, and seaborn. Furthermore, the scikit-learn library [39] is the most widely used open-source Python machine learning package. Scikit-learn structures various machine learning tasks such as regression, classification, and clustering algorithms and contains random forest, decision tree, K-NN classifiers, and more. Finally, to develop algorithms and conduct experiments, we used the Scikit-Multiflow framework [40], a Python-based open-source machine learning framework for evolving data streaming.

C. Hyper-parameter Tuning

The model hyper-parameters were tuned through a series of loops that iterated through changes in each relevant parameter of the given model based on the Grid Search strategy, and the model performance was assessed using stratified 10-fold cross-validation setting. Table III displays the optimal hyper-parameter values for each classifier used in the evaluation process after executing the hyper-parameter

fine-tuning method. In addition, the model performance is assessed using several ranges of sliding window size settings because of the relevance of this parameter in terms of computation time and memory use. For small datasets (Pima Diabetes, Breast-cancer, Arrhythmia, Heart Disease, Hepatitis), we used the sliding window of sizes w = 100, 120, 140, 160, 180, and 200. Whereas w = 100, 200, 300, 400, 500, 600, and 700 are the window sizes chosen for huge datasets (http, Credit-card-fraud, smtp, Annthyroid, Thyroid, Cardio).

D. Performance Metrics

To assess the effectiveness of the proposed framework, the average scores of ten separate trials were used to construct the most popular performance metrics in the outlier detection area, which are accuracy, precision, sensitivity or recall, and F1-score measures. All these metrics are calculated by the following equations, equation (7) – (10). For the binary classification, there are only four possible outcomes, which are True Negative (TN), False Negative (FN), False Positive (FP) and True Positive (TP).

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) * 100 \qquad (7)$$

$$Recall = TP/(TP + FN) * 100 \qquad (8)$$

$$Precision = TP / (TP + FP) * 100 \qquad (9)$$

$$F1 = 2*(\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \qquad (10)$$

E. Results and Discussion

The proposed framework is evaluated by comparing results with its standalone base learners and other some predictive machine learning techniques such as Logistic Regression (LR), SVM, GaussianNB, K-NN, Artificial Neural Network (ANN), Gradient Boosting, and Local Outlier Factor (LOF). In addition, ESOD is compared with many state-of-the-art methods found in [14], [15], [21], [22], [26], [27], and [28]. We individually show the performance of the comparisons on the aforementioned 11 datasets using the accuracy, precision, recall, and F1-score metrics. The performance results are the average of 30 trials of independent experiments.

In the first experiments set, the efficiency of the proposed framework is evaluated using the accuracy, precision, and recall metrics with all baselines on the datasets. Table IV compares performance results of all base learners with the proposed model. From the results, it is clear that the linear regression, K-NN, SVM, and LOF perform worse than other algorithms in most cases. Although the base learners, i.e., iForestASD, decision tree, and adaptive random forest classifiers, of the proposed model independently perform very well, but the proposed framework performs better than these three standalone base learners. On the other hand, the experimental results reveal that the ESOD performs very well on the large-scale datasets such as http, credit and smtp datasets, beside its superior performance over other small datasets. For instance, ESOD achieves 98.65% and 99.64% of

precision score on http and heart disease datasets, respectively. Another notable issue is that ESOD gives the highest recall or outlier detection rate, where it gets 99.78%, 99.05%, 99.66%, and 99.54% rates on http, credit, smtp, and hepatitis datasets, respectively.

Fig. 5 plots the performance comparison between different machine learning algorithms, and ESOD method in terms of F1-score. It should be noted that the K-NN, SVM, and LOF methods exhibited very low performance rates compared to ESOD and the others. On the http dataset, which is the biggest dataset, the K-NN, SVM and LOF models exhibit poor recall rates of 77.90%, 70.78%, and 61.58%, respectively, demonstrating their inability to learn enough samples to accurately classify outlier data, while ESOD achieves 99.21% on the same dataset. Furthermore, all the K-NN, SVM and LOF models get the lowest F1 rates on the Breast-cancer dataset, which has the largest outlier ratio, with 77.34%, 75.25%, and 69.59%, respectively, while ESOD can achieve the highest F1-score of 89.18%. In general, the proposed model has the best F1-score rates on all datasets.

The next set of experiments is performed to compare the proposed framework with some state-of-the-art methods for streaming outlier detection presented in [14], [15], [21], [22], [26], [27] and [28] and the results are shown in Fig 6. In more detail, from Fig. 6(a)-(k), the performance of the proposed model (ESOD) is better than its competitive methods on all benchmark datasets, where it achieves higher rates than others. For instance, ESOD has attained 97.37% accuracy, 98.65% precision, a recall of 99.78% and finally F1-score of 99.21% on the http dataset, Fig. 6(a). In contrast, other methods have achieved lower rates, which are 71.87%, 70.63%, 73.22%, 88.03%, 90.16%, 93.81%, and 84.32% of the recall measure for the methods UKOF [14], MWFP-Outlier [15], LiCS [21], iLDCBOF [22], Method in [26], Method in [27], and ASEC-OD [28], respectively. One thing is notable here is that the MWFP-Outlier and LiCS methods have the lowest performance rates on all datasets in most cases. However, the model in [27] has a slightly high rates like our proposed model, where it gains 93.81%, 91.75%, 91.83%, 91.75%, 92.43%, 91.31%, 81.82%, 87.33%, 90.27%, 93.50%, 90.70% of recall on http, credit, smtp, Annthyroid, thyroid, cardio, Pima, breast-cancer, arrhythmia, heart, and hepatitis datasets, respectively. Overall, the precisions of the seven state-of-the-art methods are also lower than the ESOD precision (61.86% - 93.52%). One can observe that the performance rates of our proposed framework is more consistent with increasing number of features on the dataset as compare to the other seven state-of-the-art methods on all datasets. On Arrhythmia dataset, which has 279 features, LiCS performs slightly better than iLDCBOF, Method in [26] and ASEC-OD. Meanwhile, UKOF obtains the lowest rates on breast-cancer dataset, which has the highest outlier ratio, but method in [26] approximately performs as good as ASEC-OD [28].

TABLE IV. EVALUATION OF ACCURACY, PRECISION AND RECALL RATES ON DIFFERENT INDIVIDUAL MODELS VS. ESOD FRAMEWORK. THE BEST AVERAGE SCORES PER EACH DATASET (COLUMN) AND MODEL (ROW) ARE SHOWN IN BOLD

	http	Credit	smtp	Annthyroid	Thyroid	Cardio	Pima	Breast-cancer	Arrhythmia	Heart Disease	Hepatitis
Accuracy											
iForestASD	**95.68**	93.05	95.17	94.15	95.66	94.33	86.62	74.89	83.64	**88.60**	87.54
DT	92.24	91.22	**92.36**	91.06	91.50	90.74	86.51	75.31	84.39	86.71	88.34
ARF	93.66	92.05	92.70	92.98	93.84	**94.50**	85.65	74.29	83.44	86.57	86.53
LR	71.07	85.15	85.66	85.70	86.01	**86.30**	84.28	70.99	83.16	85.72	86.11
K-NN	80.08	80.11	80.34	81.00	**81.37**	81.22	80.76	69.60	80.23	80.77	80.31
SVM	70.48	70.50	71.00	69.83	70.00	70.53	67.54	62.07	71.41	70.87	**73.04**
GNB	92.00	**92.23**	91.89	90.08	90.69	91.22	84.70	73.60	83.00	83.18	**90.60**
ANN	85.14	86.43	86.45	85.91	85.07	86.18	86.00	74.13	83.46	**86.52**	83.48
GB	90.06	90.20	89.04	**92.00**	91.25	90.24	85.16	73.36	80.33	85.15	86.02
LOF	75.09	76.00	75.35	76.48	74.08	76.63	75.62	**77.09**	76.32	76.36	76.08
ESOD	**97.37**	**94.05**	**95.67**	**95.32**	**95.70**	**96.23**	**88.18**	**79.92**	**85.45**	87.44	90.58
Precision											
iForestASD	97.20	95.18	93.03	95.21	97.43	94.64	85.97	84.24	94.28	**98.70**	96.35
DT	**94.90**	91.57	91.53	93.90	94.42	94.54	84.08	83.81	92.21	93.29	92.80
ARF	96.26	92.71	**97.36**	94.46	95.55	95.10	84.76	85.94	92.42	97.26	93.98
LR	80.54	82.50	82.45	83.67	81.65	88.92	81.36	72.67	88.22	91.24	**92.80**
K-NN	70.20	71.56	72.19	72.77	74.37	75.58	79.31	79.71	76.47	88.96	**89.34**
SVM	70.19	70.60	70.66	71.06	71.80	72.71	73.20	73.71	**74.02**	72.96	71.57
GNB	78.42	76.66	75.60	74.30	78.69	79.24	81.21	**86.62**	84.78	82.54	82.66
ANN	90.04	89.90	91.26	90.65	91.15	**93.28**	83.38	80.75	91.02	91.33	92.14
GB	89.16	90.91	91.51	**92.91**	90.72	92.73	83.88	82.41	90.78	92.17	90.49
LOF	51.25	51.36	51.48	51.79	**64.81**	53.33	54.40	62.06	58.55	63.93	62.12
ESOD	**98.65**	**96.70**	**97.92**	**96.33**	**97.46**	**98.30**	**87.57**	**88.59**	**95.06**	**99.64**	**96.54**
Recall											
iForestASD	**98.68**	98.09	94.83	95.44	94.26	92.30	89.65	86.23	95.43	97.55	98.63
DT	93.62	94.48	93.72	94.30	93.51	91.85	88.23	83.79	91.87	94.11	**96.05**
ARF	**98.16**	96.02	94.35	95.24	94.13	92.22	86.75	85.75	94.07	96.65	94.85
LR	80.25	80.51	81.76	81.78	81.80	81.94	81.88	82.33	82.43	**83.60**	83.54
K-NN	87.50	86.15	86.29	**88.57**	82.27	86.99	86.46	75.11	83.19	82.78	85.39
SVM	71.37	73.76	75.08	75.27	75.41	75.82	76.65	76.86	**78.11**	77.91	77.90
GNB	82.10	82.44	84.91	85.22	85.82	85.77	83.84	85.44	86.06	89.26	**89.83**
ANN	91.94	91.52	**93.25**	90.99	91.08	91.71	86.14	85.71	92.07	92.40	90.30
GB	81.81	82.04	86.74	82.46	**89.60**	85.26	75.06	82.23	83.94	88.31	87.39
LOF	77.14	78.55	78.16	76.99	76.55	78.71	71.09	**79.19**	73.42	76.77	72.80
ESOD	**99.78**	**99.05**	**99.66**	**98.17**	**97.14**	**95.20**	**90.92**	**89.78**	**96.86**	**98.27**	**99.54**
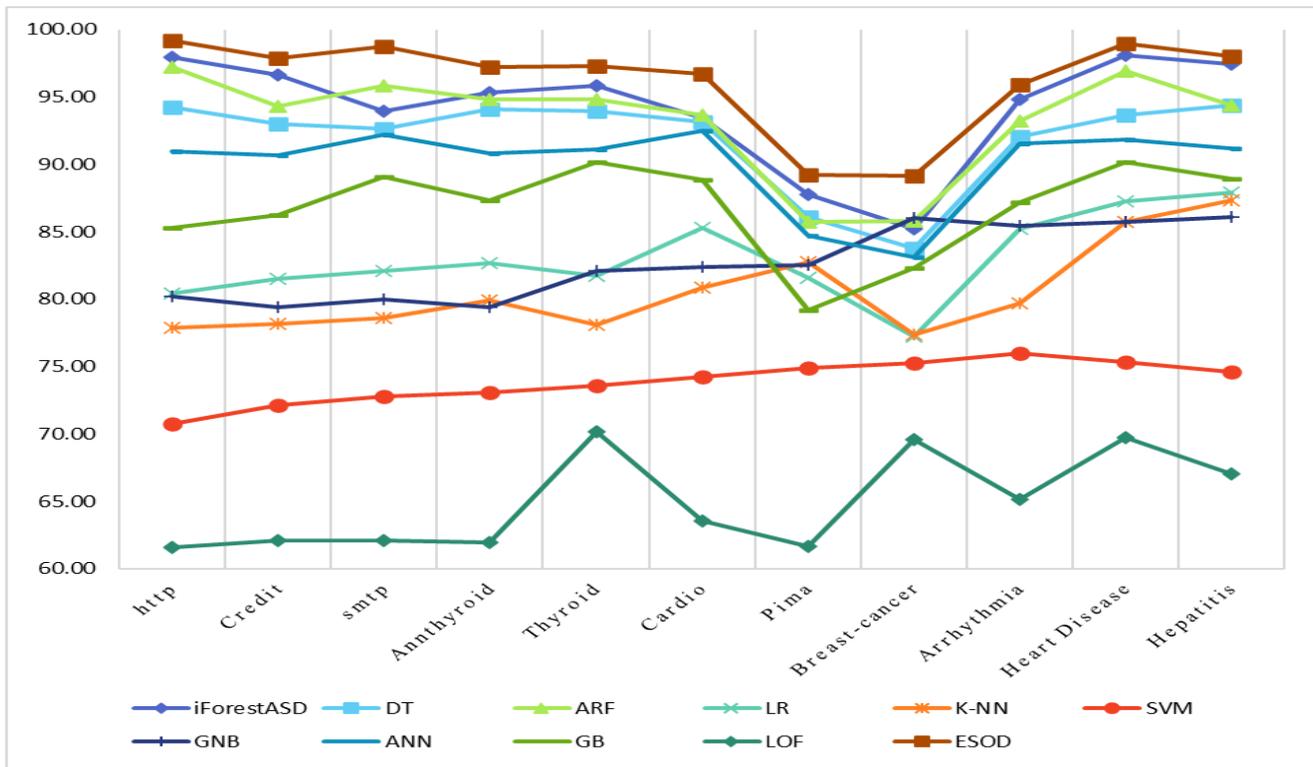
Fig. 5.   F1-score Rates of different Algorithms on different Datasets.

(a) http

(b) Credit-Card

(c) smtp
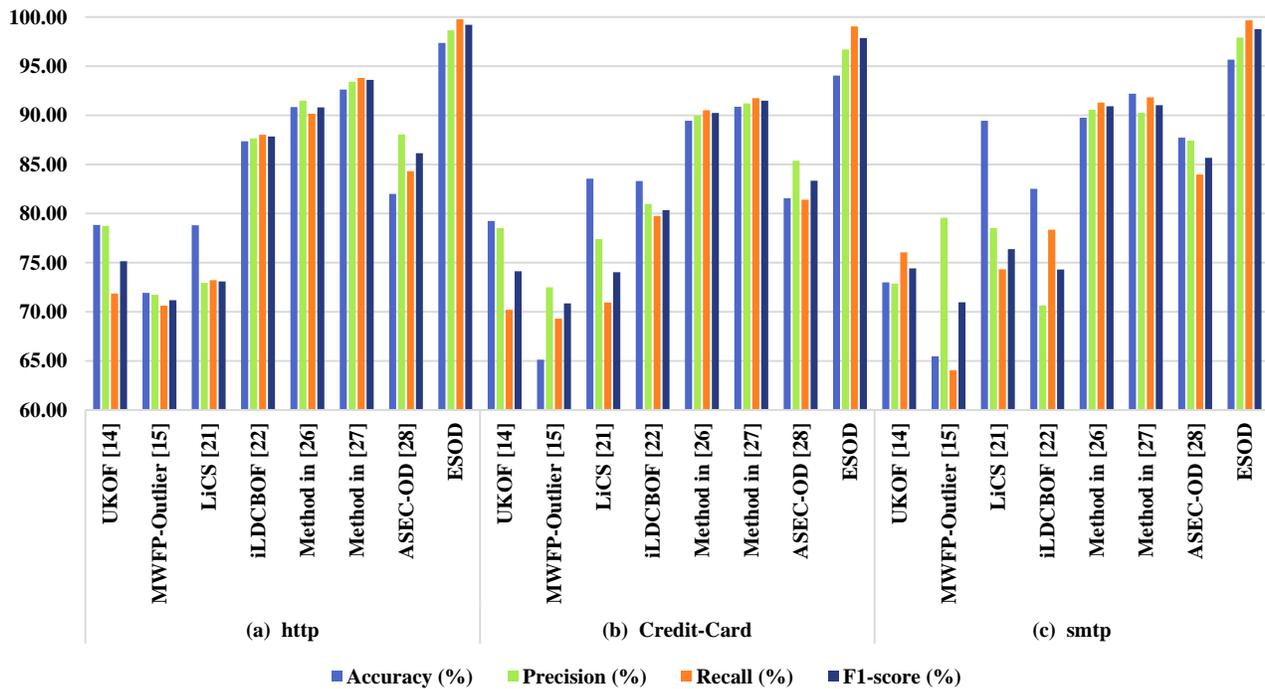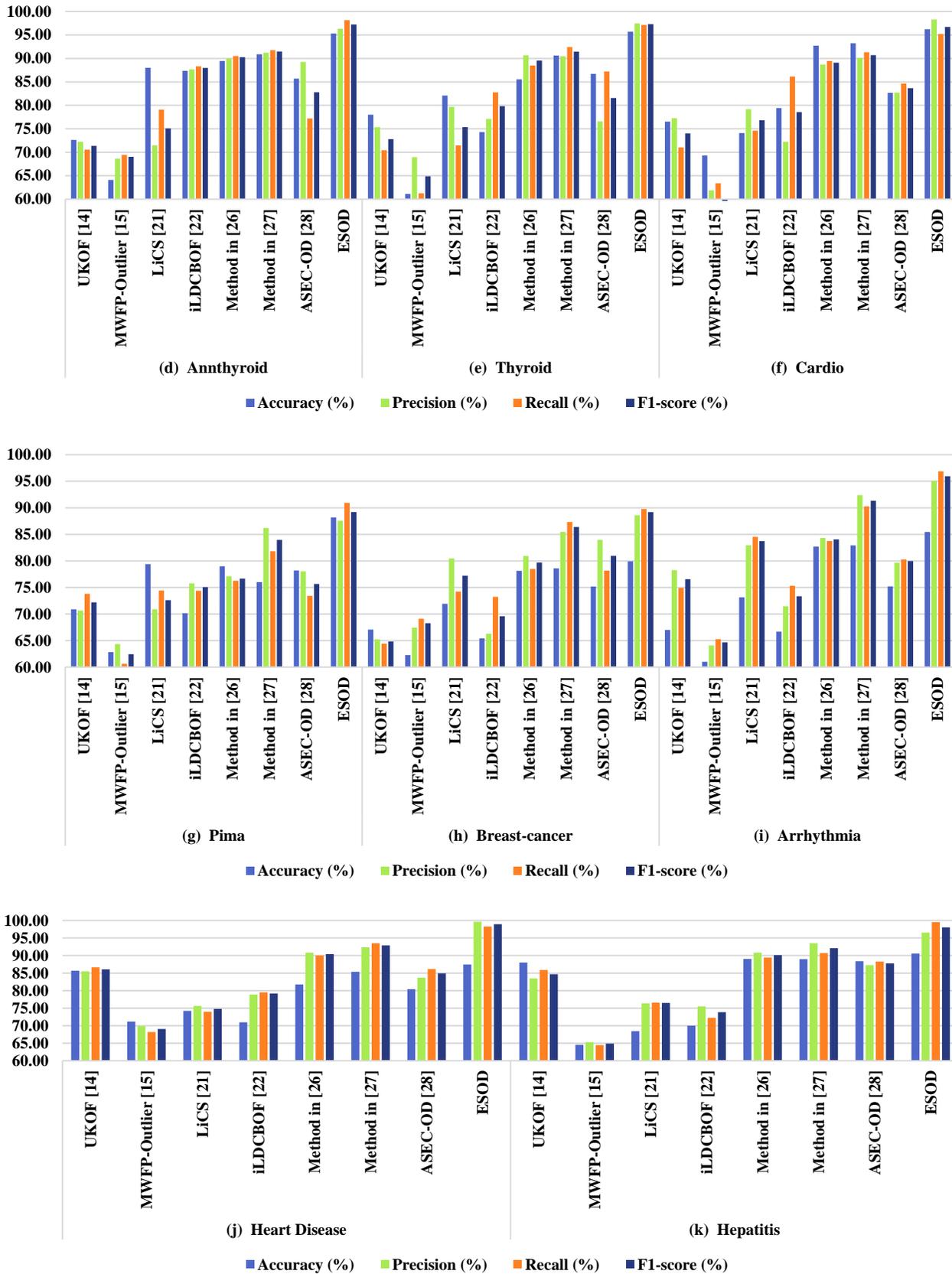
Fig. 6. Performance Comparison between State-of-the-Art Methods and the Proposed ESOD Model.

Another experiment was conducted to assess the impact of changing the window size of streaming data on the proposed framework. Table V compares the performance evaluation metrics when the window sizes were varied *(w)*. When the window size is increased to *w* = 500 for the http, credit, smtp, Annthyroid, thyroid, and cardio datasets, while *w* = 180 for the Pima, breast-cancer, arrhythmia, heart, and hepatitis datasets, the rates have improved as the F1-score grows for all datasets. However, when the window size is greater than 500 for large datasets and greater than 180 for small datasets, the F1-score falls dramatically. Furthermore, the model's performance varies significantly depending on the dataset, with ESOD earning a F1-score over 99% for the http dataset, which is the largest dataset. As a result, we expect ESOD to perform better on datasets with a high scale data.

The final experiments set is performed to evaluate the average execution time of ESOD against the other methods.

Fig. 7 shows that on the http dataset, for instance, ESOD performs with an execution time of 27.46 milliseconds compared to the LiCS, which takes about 41.56 milliseconds while the iLDCBOF method takes 37.89 milliseconds. Furthermore, LiCS takes the longest time among all the investigated methods, and the proposed method executes the http dataset at almost half the execution time of LiCS. On the other hand, comparison of execution time within the Hepatitis dataset, for instance, displays that ESOD has the shortest execution time of 11.40 milliseconds as compared to the others. In addition, within the Pima dataset, the result indicates that ESOD has an execution time of 19.64 milliseconds as against UKOF which has resulted in 26.47 milliseconds of execution time. In general, the results demonstrate that the proposed method is significantly faster than all other methods in every tested case on all datasets.

TABLE V. PERFORMANCE EVALUATION OF ESOD WHEN VARYING THE WINDOW SIZE W. THE BEST AVERAGE SCORES PER EACH DATASET ARE SHOWN IN BOLD AND HIGHLIGHTED

Dataset	Window Size (w)	Accuracy	Precision	Recall	F1-score
http	100	95.63	95.34	95.88	95.61
	200	95.72	95.72	96.12	95.92
	300	95.89	95.94	96.69	96.31
	400	96.71	96.06	96.64	96.35
	500	**97.37**	**98.65**	**99.78**	**99.21**
	600	97.11	97.41	97.83	97.62
	700	97.00	96.88	97.20	97.04
Credit	100	89.69	95.19	94.69	94.94
	200	90.61	95.27	94.21	94.74
	300	91.49	96.21	97.70	96.95
	400	92.99	96.57	97.82	97.19
	500	**94.05**	**96.70**	**99.05**	**97.86**
	600	93.79	96.66	98.42	97.53
	700	92.014	95.83	96.39	96.11
smtp	100	92.38	94.47	94.46	94.46
	200	93.47	95.35	95.24	95.29
	300	94.06	96.37	95.26	95.81
	400	95.36	96.78	98.00	97.39
	500	**95.67**	**97.92**	**99.66**	**98.78**
	600	95.28	97.50	99.58	98.53
	700	94.22	95.50	98.70	97.07
Annthyroid	100	92.34	92.65	92.04	92.34
	200	92.87	92.88	92.66	92.77
	300	93.91	93.97	92.85	93.41
	400	94.57	94.62	94.25	94.43
	500	**95.32**	**96.33**	**98.17**	**97.24**
	600	95.30	95.33	97.14	96.23

	700	95.04	95.21	96.44	95.82
Thyroid	100	92.44	95.98	94.08	95.02
	200	92.78	96.03	96.00	96.01
	300	93.65	96.14	96.20	96.17
	400	94.40	96.60	97.01	96.80
	500	**95.70**	**97.46**	**97.14**	**97.30**
	600	95.51	97.00	96.45	96.72
	700	94.43	97.00	96.19	96.59
Cardio	100	90.83	96.85	91.40	94.05
	200	91.40	96.93	92.25	94.53
	300	93.70	97.38	93.83	95.57
	400	93.63	98.02	93.89	95.91
	500	**96.23**	**98.30**	**95.20**	**96.73**
	600	94.88	97.15	95.03	96.08
	700	94.72	97.04	94.73	95.87
Pima	100	84.55	83.51	84.36	83.93
	120	84.69	84.35	83.66	84.00
	140	85.74	85.17	84.88	85.02
	160	87.01	86.58	86.9	86.74
	180	**88.18**	**87.57**	**90.92**	**89.21**
	200	88.04	85	86.81	85.90
Breast-cancer	100	74.44	75.22	75.61	75.41
	120	75.11	75.43	76.33	75.88
	140	76.53	76.7	74.94	75.81
	160	77.31	76.68	75.66	76.17
	180	**79.92**	**88.59**	**89.78**	**89.18**
	200	77.12	79.04	75.46	77.21
Arrhythmia	100	81.66	92.48	91.23	91.85
	120	81.56	92.89	92.94	92.91
	140	82.13	93.69	93.98	93.83
	160	84.78	94.44	95.48	94.96
	180	**85.45**	**95.06**	**96.86**	**95.95**
	200	83.02	94.01	93.66	93.83
Heart Disease	100	82.41	92.32	94.66	93.48
	120	83.07	93.14	96.01	94.55
	140	84.53	95.61	96.74	96.17
	160	86.18	97.85	97.37	97.61
	180	**87.44**	**99.64**	**98.27**	**98.95**
	200	87.30	98.10	98.12	98.11
Hepatitis	100	84.42	92.77	95.59	94.16
	120	85.61	92.81	95.62	94.19
	140	89.09	94.32	97.20	95.74
	160	89.35	95.05	98.20	96.60
	180	**90.58**	**96.54**	**99.54**	**98.02**
	200	86.27	96.29	97.86	97.07
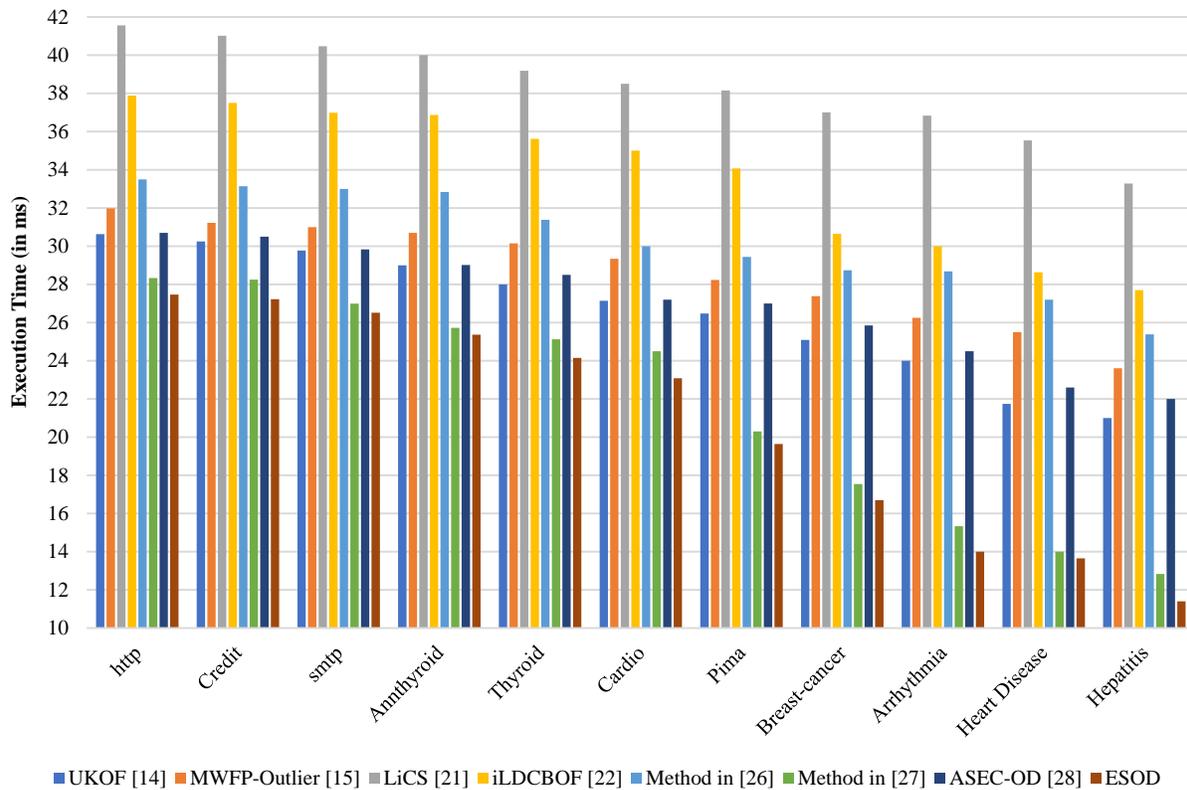
Fig. 7. Execution Time Evaluation for ESOD Compared with different Methods on different Datasets.

VI. CONCLUSION AND FUTURE WORK

In this study, we propose a novel sliding window ensemble-based framework for detecting outliers in data streams called an Ensemble-base Streaming Outlier Detection (ESOD). To improve outlier detection decisions, the proposed framework was built with three machine learning algorithms as base learners: iForestASD, decision tree, and adaptive random forest (ARF) based on a weighted voting detection technique. Furthermore, ESOD considered the concept evolution nature of streaming data. Extensive empirical evaluations on various real datasets demonstrate the performance of our framework in comparison to that of existing algorithms in the literature. The results showed that the proposed framework beat existing algorithms in terms of outlier detection rate, as well as overall performance. In the future, we will look into how to adapt the framework to the feature-evolution nature of data streams. Another option is to use more optimization techniques and different feature selection methods. Aside from improving execution performance, we are also interested in implementing various time-based techniques that may lead to improved detection and accuracy rates.

REFERENCES

[1] M. Hahsler, M. Bolaños, and J. Forrest, "Introduction to stream: An extensible framework for data stream clustering research with R," J. Stat. Softw., vol. 76, no. 1, 2017, doi: 10.18637/jss.v076.i14.

[2] S. Ramírez-Gallego, B. Krawczyk, S. García, Michał Wozniak, and F. Herrera, "A survey on data preprocessing for data stream mining: Current status and future directions," Neurocomputing, vol. 239, pp. 39–57, 2017, doi: 10.1016/j.neucom.2017.01.078.

[3] M. M. Gaber, "Advances in data stream mining," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 2, no. 1, pp. 79–85, 2012, doi: 10.1002/widm.52.

[4] V. Chandola and V. KUMAR, "Anomaly Detection: A Survey," ACM Comput. Surv., no. September, pp. 1–72, 2009.

[5] C. C. Aggarwal, Outlier Analysis, vol. 24, no. 2. 2016.

[6] T. Kim and C. H. Park, "Anomaly pattern detection for streaming data," Expert Syst. Appl., vol. 149, p. 113252, Jul. 2020, doi: 10.1016/j.eswa.2020.113252.

[7] M. Sakr, W. Atwa, and A. Keshk, "Sub-Grid Partitioning Algorithm for Distributed Outlier Detection on Big Data," 2018 13th Int. Conf. Comput. Eng. Syst., vol. IEEE, pp. 252–257, 2018.

[8] N. Paulauskas and A. Baskys, "Application of Histogram-Based Outlier Scores to Detect Computer Network Anomalies," electronics, vol. 8, pp. 1–8, 2019.

[9] M. E. Silva and I. Pereira, "Bayesian Outlier Detection in Non - Gaussian Autoregressive Time Series Bayesian outlier detection in non-Gaussian AutoRegressive time series ∗," J. Time Ser. Anal., no. December, 2018, doi: 10.1111/jtsa.12439.

[10] C. H. Park, "Outlier and anomaly pattern detection on data streams," J. Supercomput., vol. 75, no. 9, pp. 6118–6128, 2019, doi: 10.1007/s11227-018-2674-1.

[11] M. Sakr, W. Atwa, and A. Keshk, "Parallel outlier detection in real time data streams," Inf. Sci. Lett., vol. 217, no. 3, pp. 211–217, 2020.

[12] M. Togbe et al., "Anomaly Detection for Data Streams Based on Isolation Forest using Scikit-multiflow," 2020.

[13] S. Mishra and M. Chawla, "A Comparative Study of Local Outlier Factor Algorithms for Outliers Detection," in Emerging Technologies in Data Mining and Information Security, Advances in Intelligent Systems and Computing, 2019, pp. 347–356. doi: 10.1007/978-981-13-1498-8.

[14] F. Liu, Y. Yu, P. Song, Y. Fan, and X. Tong, "Scalable KDE-based top-n local outlier detection over large-scale data streams," Knowledge-Based Syst., vol. 204, p. 106186, 2020, doi: 10.1016/j.knosys.2020.106186.

[15] S. Cai et al., "MWFP-outlier: Maximal weighted frequent-pattern-based approach for detecting outliers from uncertain weighted data streams," Inf. Sci. (Ny)., vol. 591, pp. 195–225, 2022, doi: 10.1016/j.ins.2022.01.028.

[16] A. Zubaroğlu and V. Atalay, "Data stream clustering: a review," Artif. Intell. Rev., 2020, doi: 10.1007/s10462-020-09874-x.

[17] L. Tran, L. Fan, and C. Shahabi, "Fast distance-based outlier detection in data streams based on micro-clusters," ACM Int. Conf. Proceeding Ser., pp. 162–169, 2019, doi: 10.1145/3368926.3369667.

[18] A. Hassan, H. Mokhtar, and O. Hegazy, "A heuristic approach for sensor network outlier detection," Int J Res Rev Wirel Sens Netw, vol. 1, no. 4, pp. 66–72, 2011.

[19] C. HewaNadungodage, Y. Xia, and J. Lee, "Gpu-accelerated outlier detection for continuous data streams," in the 30th IEEE International Parallel and Distributed Processing Symposium (IPDPS), 2016, pp. 1133–1142.

[20] K. Yu, W. Shi, N. Santoro, and X. Ma, "Real-time outlier detection over streaming data," in 2019 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Internet of People and Smart City Innovation, SmartWorld/UIC/ATC/SCALCOM/IOP/SCI, 2019, pp. 125–132. doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00063.

[21] F. Z. Benjelloun, A. Oussous, A. Bennani, S. Belfkih, and A. Ait Lahcen, "Improving outliers detection in data streams using LiCS and voting," J. King Saud Univ. - Comput. Inf. Sci., vol. 33, no. 10, pp. 1177–1185, 2021, doi: 10.1016/j.jksuci.2019.08.003.

[22] A. Degirmenci and O. Karal, "Efficient density and cluster based incremental outlier detection in data streams," Inf. Sci. (Ny)., vol. 607, pp. 901–920, 2022, doi: 10.1016/j.ins.2022.06.013.

[23] M. M. Masud, J. Gao, J. Han, L. Khan, and B. M. Thuraising-ham, "Classification and adaptive novel class detection of feature-evolving data streams," IEEE Trans. Knowl. Data Eng., vol. 25, 2013.

[24] W. Wang, D. Wang, S. Jiang, S. Qin, and L. Xue, "Anomaly detection in big data with separable compressive sensing," in the 2015 international conference on communications, signal processing, and systems. Springer, 2016, pp. 589–94.

[25] H. Ghomeshi, M. M. Gaber, and Y. Kovalchuk, "A non-canonical hybrid metaheuristic approach to adaptive data stream classification," Futur. Gener. Comput. Syst., vol. 102, pp. 127–139, 2020, doi: 10.1016/j.future.2019.07.067.

[26] N. Iftikhar, T. Baattrup-Andersen, F. E. Nordbjerg, and K. Jeppesen, "Outlier Detection in Sensor Data using Ensemble Learning," Procedia Comput. Sci., vol. 176, pp. 1160–1169, 2020, doi: 10.1016/j.procs.2020.09.112.

[27] M. Togbe et al., "Anomalies Detection Using Isolation in Concept-Drifting Data Streams," computers, pp. 1–21, 2021, [Online]. Available: https://doi.org/10.3390/ computers10010013.

[28] N. Jayanthi, B. Vijaya Babu, and N. Sambasiva Rao, "An Ensemble Framework Based Outlier Detection System in High Dimensional Data," Mater. Today Proc., vol. 7, no. 4, pp. 1162–1175, Feb. 2021, doi: 10.1016/j.matpr.2020.11.491.

[29] S. Rayana, W. Zhong, and L. Akoglu, "Sequential ensemble learning for outlier detection: A bias-variance perspective," Proc. - IEEE Int. Conf. Data Mining, ICDM, pp. 1167–1172, 2017, doi: 10.1109/ICDM.2016.117.

[30] F. Liu, K. Ting, and Z. Zhou, "Isolation forest," in In 2008 Eighth IEEE International Conference on Data Mining, 2008, pp. 413–422.

[31] P. Karczmarek, A. Kiersztyn, W. Pedrycz, and E. Al, "K-Means-based isolation forest," Knowledge-Based Syst., vol. 195, p. 105659, 2020, doi: 10.1016/j.knosys.2020.105659.

[32] Z. Ding and M. Fei, "An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window," IFAC Proc. Vol., vol. 46, no. 20, pp. 12–17, 2013, Available: https://doi.org/https://doi.org/10.3182/20130902-3-CN-3020.00044.

[33] T. Thomas, A. P. Vijayaraghavan, and S. Emmanuel, "Applications of Decision Trees," in Machine Learning Approaches in Cyber Security Analytics, Singapore: Springer Singapore, 2020, pp. 157–184. doi: 10.1007/978-981-15-1706-8_9.

[34] L. Breiman, "Random forests," Mach Learn, vol. 45, no. 1, pp. 5–32, 2001.

[35] H. M. Gomes et al., "Adaptive random forests for evolving data stream classification," Mach. Learn., pp. 1–27, 2017, [Online]. Available: https://doi.org/10.1007/s10994-017-5642-8.

[36] V. C. Osamor and A. F. Okezie, "Enhancing the weighted voting ensemble algorithm for tuberculosis predictive diagnosis," Sci. Rep., vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1038/s41598-021-94347-6.

[37] L. Yang and A. Shami, "A Lightweight Concept Drift Detection and Adaptation Framework for IoT Data Streams," IEEE Internet Things Mag., vol. 4, no. 2, pp. 96–101, Jun. 2021, doi: 10.1109/IOTM.0001.2100012.

[38] D. Dua and C. Gra, "UCI machine learning repository," 2017. http://archive.ics.uci.edu/ml.

[39] F. Pedregosa et al., "Scikit- learn: Machine learning in Python," 12:2825–2830, Mach Learn Res, 2011. https://scikit-learn.org.

[40] J. Montiel, J. Read, A. Bifet, and T. Abdessalem, "Scikit-Multiflow: A Multi-output Streaming Framework," J. Mach. Learn. Res., vol. 19, no. 72, pp. 1–5, 2018, [Online]. Available: http://jmlr.org/papers/v19/18-251.html.

Towards a Fair Evaluation of Feature Extraction Algorithms Robustness in Structure from Motion

Dina M. Taha[1], Hala H. Zayed[2], Shady Y. El-Mashad[3]
Faculty of Engineering (at Shoubra), Benha University, Cairo, Egypt[1, 3]
Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt[2]
School of Information Technology and Computer Science, Nile University, Cairo, Egypt[2]

Abstract—**Structure from Motion is a pipeline for 3D reconstruction in which the true geometry of an object or a scene is inferred from a sequence of 2D images. As feature extraction is usually the first phase in the pipeline, the reconstruction quality depends on the accuracy of the feature extraction algorithm. Fairly evaluating the robustness of feature extraction algorithms in the absence of reconstruction ground truth is challenging due to the considerable number of parameters that affect the algorithms' sensitivity and the tradeoff between reconstruction size and error. The evaluation methodology proposed in this paper is based on two elements. The first is using constrained 3D reconstruction, in which only fixed numbers of extracted and matched features are passed to subsequent phases. The second is comparing the 3D reconstructions using size-error curves (introduced in this paper) rather than the value of reconstruction size, error, or both. The experimental results show that the proposed methodology is more transparent.**

Keywords—Feature extraction; feature matching; structure from motion; 3D reconstruction

I. INTRODUCTION

Structure from Motion (SfM) is a processing pipeline intended for reconstructing 3D models of objects from a sequence of 2D views (images). It involves several different phases in which various algorithms can be used, such as feature detection, feature description, feature matching (correspondence), triangulation, and finally 3D reconstruction. In addition, pruning or masking-out outliers may be performed between different phases. As it commonly starts with detecting features and computing descriptors, its accuracy depends mainly on the robustness of the used feature extraction algorithm. This raises a critical question; which of the available feature detectors-descriptors is best for SfM [1], [2]?

Feature detection and description is a low-level process that can be considered the cornerstone and is usually used as the starting point of many applications in computer vision. A feature (keypoint) is an interesting or important piece of information, such as points, blobs, corners, edges, junctions, etc., that is relevant to the solution of the computational task (correspondence) behind a specific application (3D reconstruction). Accordingly, a feature detector is an algorithm that can detect these interesting keypoints in a given image. On the other hand, a feature descriptor describes each detected feature in an image by assigning it a distinctive identity formed by the pixels within a certain neighborhood. This identity enables effective recognition of the

corresponding features during matching [3], [4]. After feature detection, detected features can be pruned using either a threshold for the detector response or by selecting the features with the highest response.

The simplest strategy for matching two sets of feature descriptors and finding the corresponding features is brute-force matching. In this strategy, each feature descriptor in the first set is compared with each feature descriptor in the second set. This allows for finding all possible pairs of corresponding features. However, it requires extensive computational resources. A better and faster strategy for feature matching, especially in large datasets, is to adopt a hash table or a multi-dimensional search tree as an indexing data structure for rapidly retrieving neighbors (features) from the second set that are nearest to a given feature from the first set [5].

The list of corresponding features can be pruned by comparing the similarity (distance) of each pair with a predefined threshold representing the maximum allowed distance. Only pairs that satisfy the predefined threshold are kept, while other pairs are discarded. The downside of this method is the difficulty of deciding the appropriate threshold. In addition, applying the same threshold to all pairs can be inaccurate in some circumstances, as the distance of a correct match can be greater than that of an incorrect match. A better pruning strategy in such cases is to select only the best pair (the one with the smallest distance) for every feature or adopt the Nearest Neighbor Distance Ratio (NNDR) [6]. Matching results can be further pruned by selecting the pairs with the smallest distance.

The remaining incorrect matches (outliers) can be removed or filtered out during the triangulation using many robust probabilistic methods [4], [7]. Random sampling consensus or RANSAC is one of the most widely used methods. It works by randomly picking a small subset of the seed matches, using that subset to estimate the transformation function, and then using a larger subset to check the transformation function correctness [8].

In SfM, detected features in a sequence of images are required to be accurately matched despite the image transformation (rotation and/or scaling) and illumination changes. The matched features across the sequence of images allow for predicting the camera motion by estimating the calibration matrix as well as the fundamental matrix. Accordingly, the accuracy of these estimations and consequently the 3D reconstruction is directly affected by the

robustness of the feature extraction algorithms and the accuracy of the feature matching algorithms [3]. Once feature correspondences between multiple views (images) are established, corresponding features are grouped into feature tracks. Corresponding features in each track are triangulated into a single 3D point. A sparse 3D point cloud is the result of triangulating corresponding features in all tracks [9].

More images can be incorporated into the current model using image registration. During image registration, new images are registered to the current model. This can be done by solving the Perspective-n-Point (PnP) problem. Given a set of correspondences between 3D points of an object (the current model) and their respective projections on the camera plane (of the new image), PnP problem can be solved to estimate the relative pose between the object and the camera by minimizing the reprojection error from 3D-2D point correspondences. To improve this incremental SfM pipeline, the reconstructed 3D point cloud is refined using Bundle Adjustment optimization technique resulting in a jointly optimal 3D point cloud [10].

The remaining part of this paper is organized as follows: Section II briefly covers the related work. Section III is a brief review of different feature extraction algorithms implemented for the experiment. Section IV describes the evaluation methodology proposed. Section V discusses the experimental results obtained. Section VI conclusion and future work reported.

II. RELATED WORK

Up to the best of the author's knowledge and as shown in Table. I, published research evaluates feature extraction algorithms (particularly in SfM) in terms of feature extraction time, number of extracted features, descriptor size (storage), number of matched features, point cloud size (density), and reprojection error.

TABLE I. RELATED WORK

Author	Algorithms	Factor	Winner
Govender [3]	HCD, KLY, SIFT, SURF	Error	SIFT
Urban et al. [11]	AKAZE + M-SURF, ORB, SIFT, SURF, SURF + BinBoost	Time	SURF
Chien et al. [2]	AKAZE, ORB, SIFT, SURF	#Features	SURF
		Storage	AKAZE
		Time	ORB
Pusztai et al. [5]	AGAST, AKAZE, BRISK, FAST, GFTT, KAZE, MSER, ORB, SIFT, STAR, SURF	#Features (Inliers)	SURF
Schönberger et al. [12]	ConvOpt, DeepDesc, DSP-SIFT, LIFT, SIFT, SIFT-PCA, TFeat	Density	Varies
		Error	SIFT
		Storage	SIFT
		Time	SIFT
Cao et al. [10]	BRISK, KAZE, ORB, SIFT, SURF	Error (ROS)	SURF
		Time	BRISK
Gao et al. [9]	AKAZE, DeepCompare, LF-Net, ORB, SIFT, SuperPoint, SURF	Density	SURF
		Error	AKAZE
		Time	ORB
Yusefi et al. [7]	FAST, ORB, SIFT, STAR, SURF	Error	FAST
		Time	STAR

As shown in Table II, each algorithm has one or more parameters that affect its sensitivity in most cases. Using the default values for these parameters results in a quite different average number of extracted features, as shown in Table IV. Therefore, it may be unfair to use the number of extracted features as an evaluation factor of feature extraction algorithms without trying to tune the parameters of each algorithm. In addition, the extraction time depends on the number of extracted features and the descriptor type and size. Therefore, it may be unfair to evaluate feature extraction algorithms in terms of the extraction time without taking the number of extracted features (i.e., by averaging), descriptor type, and descriptor size into consideration. Moreover, there is a trade-off between the point-cloud size and the reprojection error. Therefore, it may be unfair to evaluate feature extraction algorithms in SfM based on the value of point-cloud size only or reprojection error only.

The scope of this paper is fairly evaluating the robustness of the feature extraction algorithms in SfM in terms of point cloud size and reprojection error regardless of the space and time efficiency.

III. FEATURE DETECTORS AND DESCRIPTORS

Table II lists the fourteen feature extraction algorithms that were evaluated in this paper. A feature extraction algorithm can be designed for feature detection only, feature description only, or both feature detection and description. In the last case, feature detectors are combined with their own feature descriptors. In the first two cases, different types of feature descriptors can be paired with different kinds of feature detectors.

TABLE II. PARAMETERS OF FEATURE EXTRACTION ALGORITHMS

AKAZE	[13]	descriptor_size, descriptor_channels, threshold, nOctaves, nOctaveLayers
BEBLID*	[14]	n_bits, scale_factor
BRIEF*	[15]	bytes
BRISK	[16]	thresh, octaves
DAISY*	[17]	radius, q_radius, q_theta, q_hist, norm
FREAK*	[18]	patternScale, nOctaves
KAZE	[19]	threshold, nOctaves, nOctaveLayers
LATCH*	[20]	bytes, half_ssd_size, sigma
LUCID*	[21]	lucid_kernel, blur_kernel
ORB	[22]	nfeatures, scaleFactor, nlevels, edgeThreshold, firstLevel, WTA_K, patchSize, fastThreshold
SIFT	[23]	nfeatures, nOctaveLayers, contrastThreshold, edgeThreshold, sigma
SURF	[24]	extended, hessianThreshold, nOctaves, nOctaveLayers
TEBLID*	[25]	n_bits, scale_factor
VGG*	[26]	desc, isigma, scale_factor

Algorithms identified by * in Table II do not have a default feature detector. For such algorithms, FAST (Features from Accelerated Segment Test) feature detection algorithm is combined with them in the experimental work in Section V. FAST algorithm was proposed by Rosten et al. in 2006 for fast feature detection [27], [28]. Therefore, it is considered a viable choice for complementing those algorithms.

Feature descriptors can be categorized into two main categories: namely parameterized descriptors and binary descriptors. Each element in parameterized descriptors is a floating-point number (or any non-binary discretization of a floating-point number). On the other hand, each element in binary descriptors is a binary number (0 or 1). Typically, binary descriptors are derived using pixel-level comparisons that require minimal computational resources and result in a compact representation. Consequently, they became an attractive option for numerous contemporary applications [10]. Out of the fourteen feature extraction algorithms that were evaluated in this paper, only DAISY, KAZE, SIFT, SURF, and VGG are parameterized while the others are binary.

IV. PROPOSED EVALUATION METHODOLOGY

In this section, the proposed methodology for fairly evaluating the robustness of feature extraction algorithms in SfM is presented. The proposed methodology is expressed using Python-like pseudo-code in Fig. 1 and Fig. 2. It is also illustrated by a block diagram in Fig. 3.

```
Algorithm Constrained 3D Reconstruction (I, feat_alg)
Input : sequence of images I
         feature extraction algorithm feat_alg
Output: point cloud pc
F = {}
for i in I:
    f = feat_alg.extract(i)
    F[i] = sorted(f,
       key = lambda item:-item.response
    )[:MAX_FEAT]

M = {}
for i in I:
    for j in I:
        if I == j:
            continue
        m = feat_alg.match(F[i],F[j])
        M[i,j] = sorted(m,
           key = lambda item:item.distance
        )[:MAX_MATCH]
return colmap. mapper(I,M)
```

Fig. 1.   Proposed Algorithm for Constrained 3D Reconstruction.

Constrained 3D Reconstruction algorithm shown in Fig. 1 takes a sequence of images as well as a feature extraction algorithm and returns a point cloud. Constrained 3D reconstruction is required since the quality of the 3D reconstruction depends on not only the robustness of the detected features but also the number of detected features. In order to ensure fairness and due to the difficulty of optimizing the parameters of each algorithm to produce the same number of features, only a fixed number of features detected by each algorithm is passed to the subsequent phase, as illustrated by Constrained Feature Extraction block in Fig. 3. Selection of the detected features can be performed by sorting them in non-increasing order of their respective responses and selecting the features with the highest response. In addition, only a fixed number of corresponding features matched using each algorithm is passed to the subsequent phase, as illustrated by Constrained Feature Matching block in Fig. 3. Selection of the corresponding features can be performed by storing them in

non-decreasing order of matching distance and selecting the correspondences with the smallest matching distance.

Instead of the *ratio test* proposed by Lowe in [23] in the matching phase, brute-force matcher with cross-checking was used to return only consistent pairs of features. A pair of features (i, j) is considered consistent if j is the nearest feature to i and i is the nearest to j in the feature space. The consistent pairs of matching features are then passed to the next phase, where sparse point clouds are reconstructed.

Size-Error Curves Generation algorithm shown in Fig. 2 takes a list of point clouds generated using different feature extraction algorithms for a specific dataset and reconstructs size-error curves. Such curves visualize the average reprojection error of different 3D reconstructions as a function of the point cloud size. Moreover, the use of size-error curves allows employing methods like the elbow method [29] and finding a suitable operating point on ROC curve [30] for carefully selecting and/or pruning a reconstructed point cloud. Size-error curves are generated by first creating a list of sample sizes. The points in each point cloud are sorted in non-decreasing order of their respective reprojection error and then sampled using every sample size. The size-error curve can be visualized by plotting the average reprojection error of the points in each sample against the sample size.

```
Algorithm Size-Error Curves Generation (PC)

Input : list of point clouds PC,
         each point cloud is composed of a list of
         3D points (x, y, z, e)
Output: list of dictionaries D,
         each dictionary is a map from point cloud
         size to average reprojection error
K = []
for pc in PC:
    K.append(len(pc))
K = sorted(K)

D = []
for pc in PC:
    d = {}
    e = sorted(pc.e)
    for k in K:
        if k > len(pc):
            break
        d[k] = sum(e[:k])/k
    D.append(d)

return D
```

Fig. 2.   Proposed Algorithm for Evaluation of 3D Reconstruction.

V. EXPERIMENTAL RESULTS

A. Datasets

To evaluate the robustness of feature extraction algorithms, a well-known and publicly available collection of datasets was used as a benchmark. The collection contains six different datasets of outdoor scenes of different architectural objects. All images were captured at 3072×2028 resolution and have been corrected for radial distortion. Each dataset name, number of images, and sample image are shown in Table III. More information about the dataset can be found in [31], [32].

TABLE III.   THE SIX DATASETS USED IN THIS RESEARCH

Dataset	castle-P19	castle-P30	entry-P10	fountain-P11	Herz-Jesus-P8	Herz-Jesus-P25
Sample						
Size	19	30	10	11	8	25

B. Setup

An experiment was performed on a virtual machine with a 2 x Core Intel(R) Xeon(R) CPU @ 2.20 GHz and 13.00 GB of RAM provided by Google Collaborate Pro to evaluate the robustness of different feature extraction algorithms. OpenCV 4.6.0 was built with OPENCV_ENABLE_NONFREE as well as OPENCV_EXTRA_MODULES_PATH, while COLMAP 3.8 was built with the default configuration.
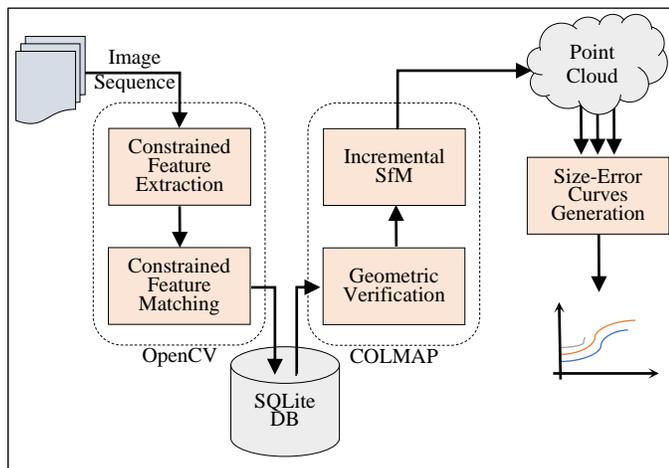
Fig. 3.   Proposed Methodology for Fair Evaluation.

Feature detection and extraction subroutines contained within OpenCV library framework [33] were used with their default parameters provided by the library for Python. When a required parameter does not have a default value, the value recommended by the algorithm author(s) is used. Each detector is only combined with its own descriptor in this experiment, and detectors with no descriptor are combined with FAST due to its speed and availability in a real-time environment. OpenCV BFMatcher was used along with crossCheck flag and the default norm of each feature extraction algorithm for feature correspondence. COLMAP SfM pipeline [34], [35] was used on the CPU for spare 3D reconstruction. The complete pipeline is illustrated by Fig. 3.

Only the best 2000 detected features are passed to the next phase. In addition, only the best 500 corresponding matches are passed to the next phase. The two numbers were chosen empirically to suit almost all the feature extraction algorithms.

In order to connect OpenCV to COLMAP, an SQLite database is used for storing the following:

- Camera parameters.
- Images filenames.
- Coordinates of each extracted feature for each image.
- Indexes of each pair of corresponding features.

In order to reconstruct the corresponding sparse point clouds, a total of 14 x 6 SQLite databases were created and passed to COLAMP as command-line parameters.

C. Results and Discussion

During this experiment, 84 different point clouds were created. As shown in Section II, the projection error is the most common and obvious way to measure how good a point cloud is. V and the radar chart in Fig. 4 illustrate the reprojection error of each point cloud. By comparing the reprojection error resulting from different feature extraction algorithms on each dataset, the feature extraction algorithm with the smallest reprojection error can be considered the winner. Based on the result of this experiment, it is clear that FREAK resulted in the smallest reprojection error on most datasets (four out of six) and on average. Therefore, it can be considered the winner. However, the quality of the resulting point cloud is defined not only by the reprojection error but also by the point cloud size (density).

According to the results shown in VI and illustrated by the radar chart in Fig. 5, it is clear that the density of the point clouds generated using FREAK feature extraction algorithm is disappointing. It is almost the smallest for most of the datasets. This means that FREAK's low reprojection error comes at a cost (the sparsity of the point cloud).

Obviously, neither Fig. 4 nor Fig. 5 (alone or side-by-side) provide an accurate measure or visualization of the relative performance of different feature extraction algorithms in SfM. This raises the need for a measure or visualization that takes care of the quantity-quality trade-off.

On the other hand, using the size-error curves shown in Fig. 6, a true winner can be easily identified by observing the area under the curve (AUC). The feature extraction algorithm corresponding to the smallest AUC and adequate density can be considered a winner regardless of point cloud size and the average reprojection error. For instance, it is clear from Fig. 6 that SIFT is the winner in Herz-Jesus-P8 dataset even though it did not generate the point cloud with the largest size or the smallest average reprojection error.

TABLE IV.   THE AVERAGE NUMBER OF EXTRACTED FEATURES PER DATASET FOR EACH FEATURE EXTRACTION ALGORITHM

AKAZE	BEBLID	BRIEF	BRISK	DAISY	FREAK	KAZE	LATCH	LUCID	ORB	SIFT	SURF	TEBLID	VGG
099888	163893	158462	081135	163893	159357	098562	158662	163893	257609	095820	359358	163893	163893

TABLE V.   REPROJECTION ERROR USING DIFFERENT FEATURE EXTRACTION ALGORITHMS

	castle-P19	castle-P30	entry-P10	fountain-P11	Herz-Jesus-P25	Herz-Jesus-P8	Average
AKAZE	0.6175	0.6171	0.7233	0.6473	0.8661	0.6831	0.6924
BEBLID	**0.4798**	0.5159	0.5705	0.5404	0.6353	0.6204	0.5604
BRIEF	0.5139	0.5075	0.6058	0.5947	0.6840	0.6364	0.5904
BRISK	0.5619	0.7170	0.7034	0.7701	0.9119	0.8535	0.7530
DAISY	0.5291	0.5687	0.6357	0.5628	0.6956	0.5994	0.5986
FREAK	0.5850	**0.4956**	**0.4923**	0.4446	**0.4891**	**0.5331**	**0.5066**
KAZE	0.6271	0.6521	0.8231	0.6488	0.8392	0.7331	0.7206
LATCH	0.5643	0.5717	0.6984	0.7157	0.8138	0.7649	0.6881
LUCID							
ORB	0.7029	0.7575	0.7366	0.8767	0.8932	0.8264	0.7989
SIFT	0.5730	0.5777	0.5336	**0.3497**	0.6307	0.5575	0.5370
SURF	0.6112	0.5870	0.6773	0.4715	0.7324	0.6789	0.6264
TEBLID	0.5133	0.5203	0.5927	0.5431	0.6507	0.6049	0.5709
VGG	0.4966	0.5372	0.6016	0.5283	0.6463	0.5915	0.5669
Average	0.5674	0.5866	0.6457	0.5918	0.7299	0.6679	0.6315

TABLE VI.   POINT CLOUD SIZE USING DIFFERENT FEATURE EXTRACTION ALGORITHMS

	castle-P19	castle-P30	entry-P10	fountain-P11	Herz-Jesus-P25	Herz-Jesus-P8	Maximum
AKAZE	2160	5869	1497	1375	3326	1212	5869
BEBLID	1559	4664	1393	1525	3036	1090	4664
BRIEF	1795	4694	1429	1473	3001	1087	4694
BRISK	1170	1229	1193	1275	2947	1088	2947
DAISY	2065	5656	**1664**	1642	3074	1145	5656
FREAK	1054	2460	1026	1228	2524	810	2524
KAZE	2151	**6177**	1591	1494	3609	**1229**	6177
LATCH	1825	5327	1412	1460	2984	1085	5327
LUCID							
ORB	1238	2246	1015	0877	2304	1046	2304
SIFT	**2240**	6118	1428	**1702**	**3614**	1197	6118
SURF	1338	4744	0934	1323	3350	1050	4744
TEBLID	1797	4528	1475	1549	3167	1140	4528
VGG	1740	4876	1468	1595	3068	1141	4876
Maximum	2240	6177	1664	1702	3614	1229	6177
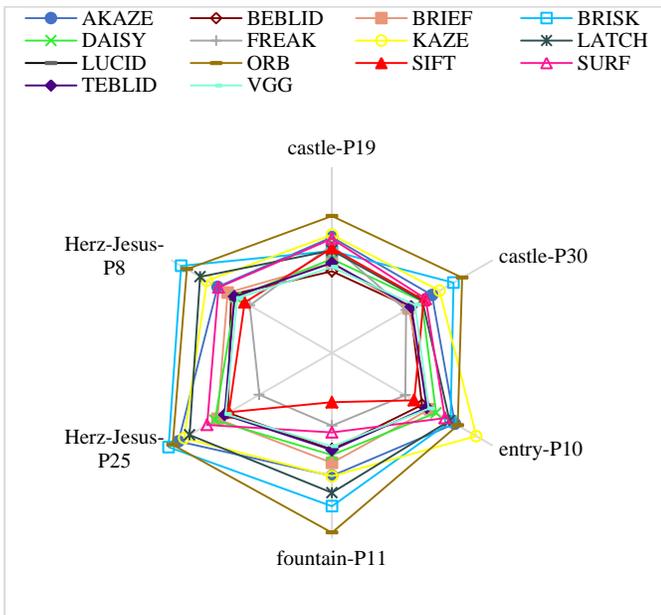
Fig. 4. Reprojection Error using different Feature Extraction Algorithms.
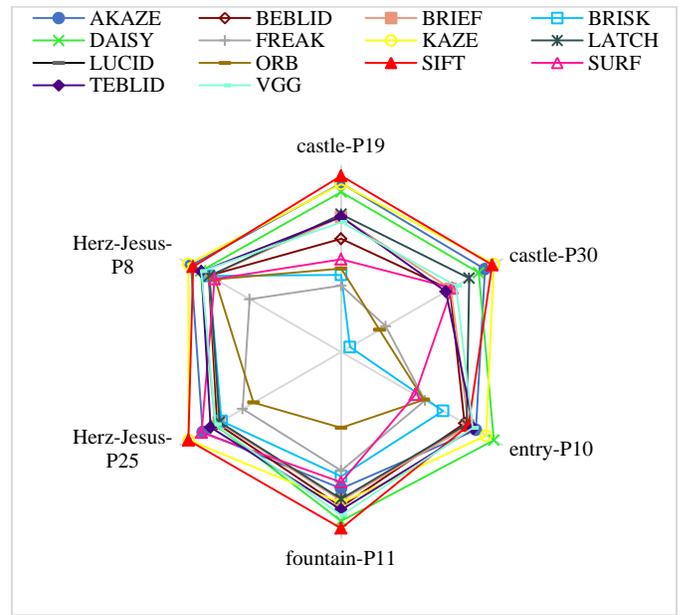
Fig. 5. Point cloud Size using different Feature Extraction Algorithms.
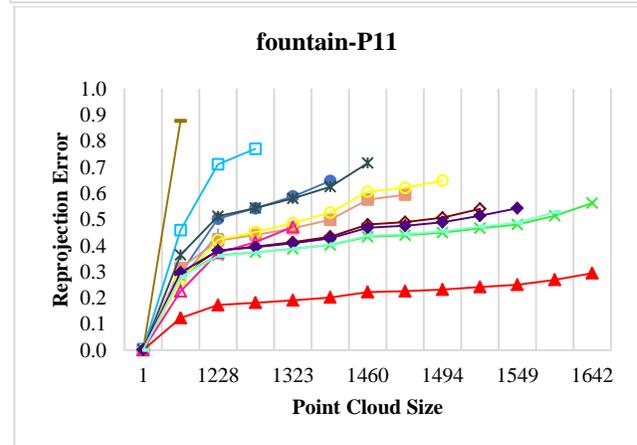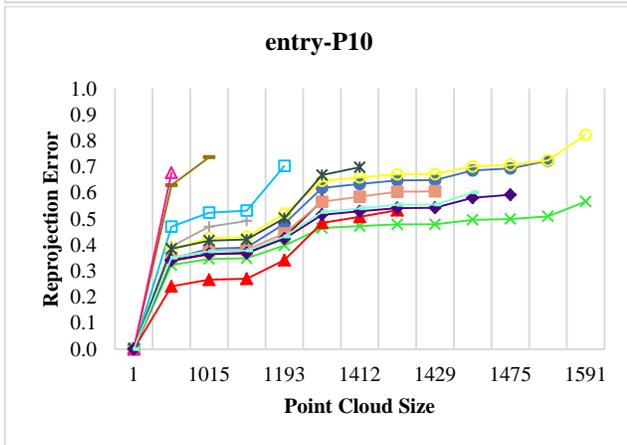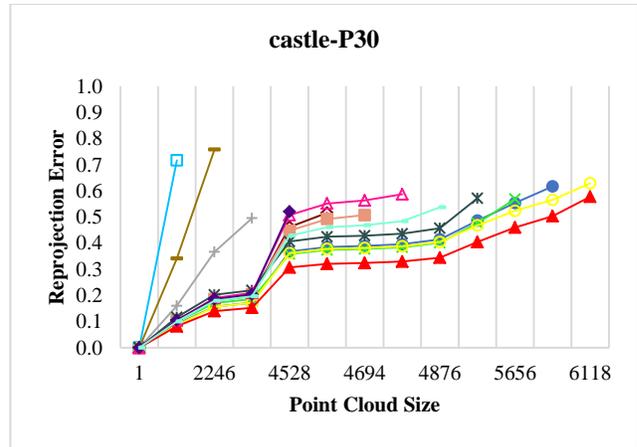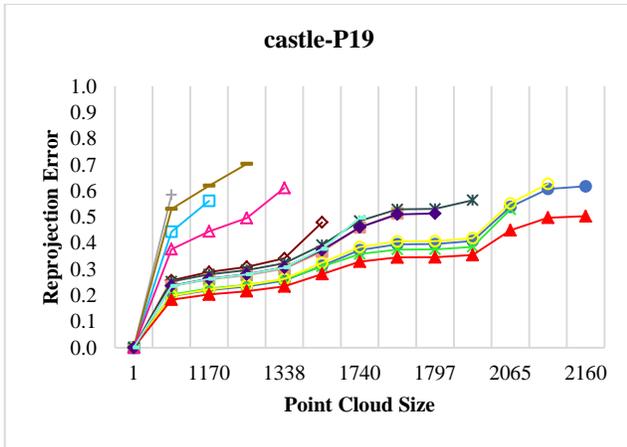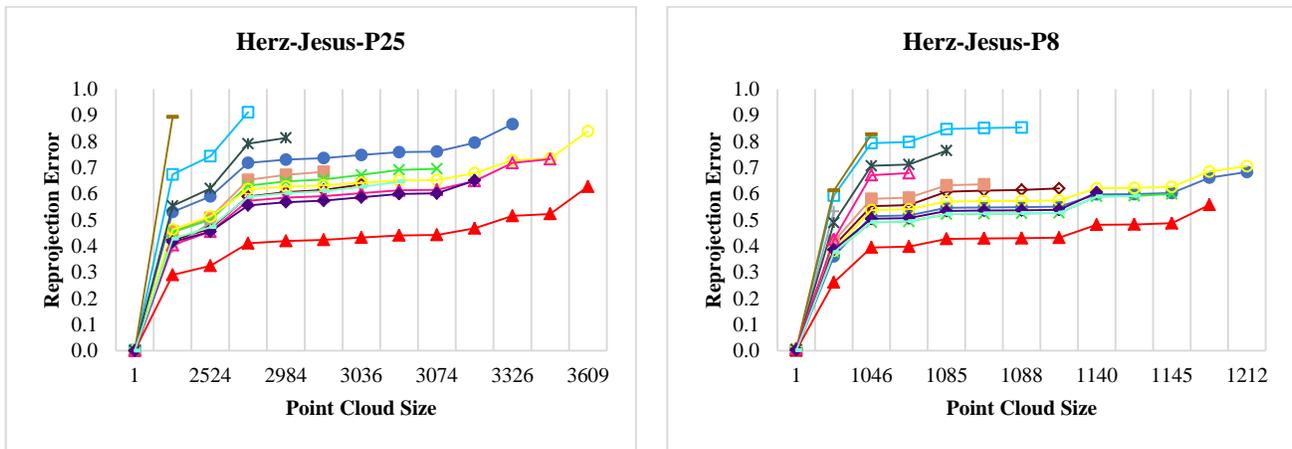
Fig. 6.   Reprojection Error as a Function of the Point cloud Size for the Six Datasets.

VI. CONCLUSION AND FUTURE WORK

In this research, fourteen different OpenCV feature extraction algorithms were evaluated in COLMAP SfM pipeline on six public datasets. Based on the experimental results reported in Section V, it can be concluded that the best feature extraction algorithm does not necessarily generate a point cloud with the minimum reprojection error nor the maximum number of points in SfM. So, the size-error curves proposed in this paper should be used for comparing different 3D reconstructions of a given dataset, as they are more transparent than using the value of point cloud size, reprojection error, or both.

The size-error curves demonstrate that SIFT feature extraction algorithm outperforms other algorithms in almost all six datasets. Our future work is to consider combining feature detectors other than FAST with descriptors that do not have a default detector and those that do. By applying the fair evaluation method proposed in this paper to the different combinations, a better combination of feature detector and feature descriptor for SfM may be found.

REFERENCES

[1] S. A. K. Tareen and Z. Saleem, "A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK," in 2018 International Conference on Computing, Mathematics and Engineering Technologies: Invent, Innovate and Integrate for Socioeconomic Development, iCoMET 2018 - Proceedings, 2018, vol. 2018-January. doi: 10.1109/ICOMET.2018.8346440.

[2] H. J. Chien, C. C. Chuang, C. Y. Chen, and R. Klette, "When to use what feature? SIFT, SURF, ORB, or A-KAZE features for monocular visual odometry," in International Conference Image and Vision Computing New Zealand, 2016, vol. 0. doi: 10.1109/IVCNZ.2016.7804434.

[3] N. Govender, "Evaluation of Feature Detection Algorithms for Structure from Motion," Csir, 2009.

[4] S. Bianco, G. Ciocca, and D. Marelli, "Evaluating the performance of structure from motion pipelines," J Imaging, vol. 4, no. 8, 2018, doi: 10.3390/jimaging4080098.

[5] Z. Pusztai and L. Hajder, "Quantitative Comparison of Feature Matchers Implemented in OpenCV3," in 21 st. Computer Vision Winter Workshop, Rimske Toplice, 2016.

[6] R. Szeliski, "Computer Vision : Algorithms and Applications 2nd Edition," Springer, 2021.

[7] A. YUSEFI, A. DURDU, and C. SUNGUR, "Performance and Trade-off Evaluation of SIFT, SURF, FAST, STAR and ORB feature detection

algorithms in Visual Odometry," European Journal of Science and Technology, 2020, doi: 10.31590/ejosat.819735.

[8] M. A. Fischler and R. C. Bolles, "Random sample consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Commun ACM, vol. 24, no. 6, 1981, doi: 10.1145/358669.358692.

[9] K. Gao et al., "Local feature performance evaluation for structure-from-motion and multi-view stereo using simulated city-scale aerial imagery," IEEE Sens J, vol. 21, no. 10, 2021, doi: 10.1109/JSEN.2020.3042810.

[10] M. Cao et al., "Evaluation of local features for structure from motion," Multimed Tools Appl, vol. 77, no. 9, 2018, doi: 10.1007/s11042-018-5864-1.

[11] S. Urban and M. Weinmann, "Finding a good feature detector-descriptor combination for the 2d keypoint-based registration of tls point clouds," in ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2015, vol. 2, no. 3W5. doi: 10.5194/isprsannals-II-3-W5-121-2015.

[12] J. L. Schönberger, H. Hardmeier, T. Sattler, and M. Pollefeys, "Comparative evaluation of hand-crafted and learned local features," in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, vol. 2017-January. doi: 10.1109/CVPR.2017.736.

[13] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in BMVC 2013 - Electronic Proceedings of the British Machine Vision Conference 2013, 2013. doi: 10.5244/C.27.13.

[14] I. Suárez, G. Sfeir, J. M. Buenaposada, and L. Baumela, "BEBLID: Boosted efficient binary local image descriptor," Pattern Recognit Lett, vol. 133, 2020, doi: 10.1016/j.patrec.2020.04.005.

[15] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2010, vol. 6314 LNCS, no. PART 4. doi: 10.1007/978-3-642-15561-1_56.

[16] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary Robust invariant scalable keypoints," in Proceedings of the IEEE International Conference on Computer Vision, 2011. doi: 10.1109/ICCV.2011.6126542.

[17] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," IEEE Trans Pattern Anal Mach Intell, vol. 32, no. 5, 2010, doi: 10.1109/TPAMI.2009.77.

[18] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012. doi: 10.1109/CVPR.2012.6247715.

[19] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012, vol. 7577 LNCS, no. PART 6. doi: 10.1007/978-3-642-33783-3_16.

[20] G. Levi and T. Hassner, "LATCH: Learned arrangements of three patch codes," in 2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016, 2016. doi: 10.1109/WACV.2016.7477723.

[21] A. Ziegler, E. Christiansen, D. Kriegman, and S. Belongie, "Locally uniform comparison image descriptor," in Advances in Neural Information Processing Systems, 2012, vol. 1.

[22] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in Proceedings of the IEEE International Conference on Computer Vision, 2011. doi: 10.1109/ICCV.2011.6126544.

[23] D. G. Low, "Distinctive image features from scale-invariant keypoints," Int J Comput Vis, 2004.

[24] H. Bay, T. Tuytelaars, and L. van Gool, "SURF: Speeded up robust features," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2006, vol. 3951 LNCS. doi: 10.1007/11744023_32.

[25] I. Suarez, J. M. Buenaposada, and L. Baumela, "Revisiting Binary Local Image Description for Resource Limited Devices," IEEE Robot Autom Lett, vol. 6, no. 4, 2021, doi: 10.1109/LRA.2021.3107024.

[26] K. Simonyan, A. Vedaldi, and A. Zisserman, "Learning local feature descriptors using convex optimisation," IEEE Trans Pattern Anal Mach Intell, vol. 36, no. 8, 2014, doi: 10.1109/TPAMI.2014.2301163.

[27] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2006, vol. 3951 LNCS. doi: 10.1007/11744023_34.

[28] S. Y. El-Mashad and A. Shoukry, "Evaluating the robustness of feature correspondence using different feature extractors," in 2014 19th International Conference on Methods and Models in Automation and Robotics, MMAR 2014, 2014. doi: 10.1109/MMAR.2014.6957371.

[29] R. L. Thorndike, "Who belongs in the family?," Psychometrika, vol. 18, no. 4, 1953, doi: 10.1007/BF02289263.

[30] C. E. Metz, "Basic principles of ROC analysis," Semin Nucl Med, vol. 8, no. 4, 1978, doi: 10.1016/S0001-2998(78)80014-2.

[31] C. Strecha, W. von Hansen, L. van Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2008. doi: 10.1109/CVPR.2008.4587706.

[32] A. Ley, R. Hänsch, and O. Hellwich, "Syb3r: A realistic synthetic benchmark for 3D reconstruction from images," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016, vol. 9911 LNCS. doi: 10.1007/978-3-319-46478-7_15.

[33] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.

[34] J. L. Schönberger, E. Zheng, J. M. Frahm, and M. Pollefeys, "Pixelwise view selection for unstructured multi-view stereo," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016, vol. 9907 LNCS. doi: 10.1007/978-3-319-46487-9_31.

[35] J. L. Schonberger and J. M. Frahm, "Structure-from-Motion Revisited," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, vol. 2016-December. doi: 10.1109/CVPR.2016.445.

Ensemble Tree Classifier based Analysis of Water Quality for Layer Poultry Farm: A Study on Cauvery River

Deepika[1], Dr. Nagarathna[2]
Department of CS & E
P.E.S College of Engineering Mandya
Karnataka, India

Dr. Channegowda[3]
Vice President Technical, ZEUS
Biotech Pvt. Ltd.
Mysore, India

Abstract—**Indian poultry industry has evolved from a simple backyard occupation to a large commercial agri-based enterprise. Chicken dominates poultry production in India, accounting for almost 95% of total egg production. Several factors affect the egg production such as feeding material, drinking water, environmental factors etc. Analyzing the water quality is one of the important tasks. Cauvery River is considered as the study area because of its importance in several states of South India which have significant contribution in poultry farming. The aim of the proposed study is to develop an automated approach of water quality analysis and present a novel machine learning approach which considers an improved feature ranking method and ensemble tree classifier with majority voting. The experimental result shows that proposed approach performs better with an accuracy of 95.12%.**

Keywords—Water quality; poultry; machine learning; Cauvery river; feature ranking; ensemble tree classifier; accuracy

I. INTRODUCTION

Agriculture and farming has become the most important source of income and it plays an important role in Indian economy. The livestock with poultry farming is one of the major parts of Indian agricultural population and considered as backbone of the Indian Farmers. According to a study presented by *Statista*, the demand of chicken meat and eggs has increased drastically worldwide [1].

Animal husbandry plays a major role in Indian economy under which taming birds such as chicken ducks, turkeys for production of various goods such as egg, meat, feathers, is called poultry farming. It has evolved from a backyard hobby to a full-fledged technological-commercial industry [2]. Chicken farming is prevalent in places like as Andhra Pradesh, Tamil Nadu, West Bengal, Maharashtra, Orissa, Bihar, Kerala, and Karnataka, among others [3].

The recent 20th Livestock Census has reported that total numbers of 851.8 million poultry birds are present in India and 30% (250 million) of this population is covered by backyard, small or marginal farmers [4].

Below given Table I present the data of 20th Livestock census which shows the livestock population and egg production in South India [4].

There is a significant variation in poultry industry in different regions of India. The South Indian states like – Karnataka, Tamil Nadu, Kerala and Andhra Pradesh produce for about 45 percent of the country's egg production, with a per capita consumption of 57 eggs and 0.5 kg of broiler meat [5].

Based on the above facts and data, we can say that the poultry is considered as the expediting segment of the agricultural sector. Moreover, current technological advancements, up gradations have paved the way for multi-fold growth in the poultry and its allied sectors. This growth has led to high yielding layers, broilers, with the help of appropriate nutrition, and management, the egg and meat production also have grown during last decades.

On the other hand, poultry industry suffers from various challenges which degrade the production of eggs. Some of the challenges are feeding material, raw materials, disease outbreak, water quality, and temperature etc. Water plays an important role in this context of poultry rearing and management. For chickens, water is one of the most vital nutrients for the following reasons:

- To maintain maximum egg production and egg shell quality, and to achieve good body weight and growth, good quality of water is needed.

- To properly manage bodily functions water is essential.

Some of the water related factors which affect the performance of poultry are as follows:

- Water source: The water source is an important parameter to be considered because the acidity or alkalinity of water depends on the depth of well from where water is extracted. Water found at 30-meter depth is generally acidic in nature whereas water level of more than 100 meters is more alkaline nature [6].

TABLE I. LIVESTOCK POPULATION EGG PRODUCTION IN INDIA [4]

State	Livestock population (million)	Year (2019-2020)
Karnataka	59m	66511 lakh
Tamil Nadu	120m	200216 lakh
Kerala	29m	21845 lakh

Consumption of acidic water is prone to parasitic infestation whereas consumption of alkaline water leads to deficiency of required minerals such as calcium, phosphorus, magnesium and potassium.

- Watering system: The water consumption through the trough based system is greater when compared with the nipple system. Layers which are using the nipple drinker system have reflected in feed conversion, less mortality when compared with channel based watering systems. Table II shows the comparison of nipple and trough watering system. Moreover, the nipple based drinker system are enclosed and more sanitary can be ensured by disinfecting and cleaning the system which prevents the disease problems [7].

TABLE II.   COMPARATIVE PERFORMANCE OF BROILER REARED ON NIPPLE OR TROUGH WATERING SYSTEM [7]

	Trough	Nipple
Water consumption	166	245
Food conversion	1.91	1.92
Bodyweight (kg)	1.95	1.96
Mortality in first 14 days (%)	1.20	1.30
Condemnations (%)	1.17	1.40

- Water temperature: For four weeks, researchers tested the effects of cold water (8°C) and regular water (29.5°C) on the production characteristics of chickens. The effects of heat stress were eased by using cold water, and the birds were able to acquire more weight and had improved feed efficiency as a result. When given cold water, birds were more sensitive to vitamin C supplementation (500mg/liter of water), resulting in enhanced survival and carcass quality, with a focus on breast meat output [8].

When exposed to different environmental temperatures the consumption of water and feed intake varies as shown in the following Table III.

TABLE III.   CONSUMPTION OF WATER AND FEED INTAKE AT DIFFERENT TEMPERATURES [9]

Characteristics	Temperature		
	18°C	29°C	35°C
Age (days)	245	200	180
Feed (lbs/hen/day) (g)	0.23	0.19	0.14
Water(gal/100hens/day)	5.5	5.8	7.9
Water to Feed Ratio	2.03	2.60	4.67

For the adequate performance of laying hens and other poultry, good quality of water is essential. The foreign substances present in water may affect its' drinkability and well-being of the hens. Drinking water composition will differ in various regions across the country. Water may get polluted from the practices that occur naturally or from run-off sources. Water quality standards include many factors that affect the quality of water used in the poultry. These factors include items like pH, bacterial load, color, hardness, total solids, and dissolved oxygen, biochemical oxygen on demand, conductivity, water temperature and mineral content [9].

Water quality plays important role in the poultry farming. The source of water should be good in poultry farms otherwise it may lead to water borne diseases which affect the egg quality and egg production. To characterize water quality several parameters are used. Below given Table IV illustrates the standard water parameters for poultry farming.

The manual inspection of water sample is not feasible solution due to continuous variations in several quality parameters. Our paper aims to analyze and predict quality of water based on water quality index computation and machine learning algorithms. However, the performance of machine learning techniques depends on the nature of attribute, dimensionality of attribute and classifier model. Thus, the main objectives of the proposed work are as follows:

- To use the weighted arithmetic model for dataset generation.

- To present the feature ranking and selection model.

- To present the tree based ensemble classifier to improve the prediction accuracy.

Rest of the paper is organized as follows: Section II presents the brief literature review about the recent existing techniques of water quality prediction, Section III provides the proposed solution for WQI prediction by using machine learning, Section IV presents results and the comparative analysis of proposed approach and finally, Section V describes the concluding remarks of the proposed approach.

II.   LITERATURE SURVEY

This section presents the brief literature review about existing schemes of water quality prediction based on chemical parameters. The previous section has described the advantages, growth of poultry farming in India. However, the review study briefs about various challenges faced in poultry farming such as manoeuvrability in broiler prices, availability of trained workers, shortage of raw material and many more. However, water quality has a significant impact on the egg and meat production.

Pujar et al. [10] presented a study for Krishna river water quality analysis by using an IoT system. The IoT system was utilized to gather data for several water quality indicators such as pH, turbidity, DO, BOD, NO3, temperature, and conductivity from selected stations, resulting in a data collection that was used to monitor water quality. Using one-way ANOVA, the acquired data was effectively used to analyze the quality of water of Krishna River. One-way ANOVA is used to analyze a specific parameter and predicts the quality of water based on the value obtained.

The study of two parameters as a separate entity as well as a combination of two parameters was done using two-way ANOVA [10].

Nayak et al. [11] developed a fuzzy logic based model by using triangular and trapezoidal membership function. This model used centroid, bisector and mean of maxima (MOM) methods for defuzzification.

TABLE IV. WATER QUALITY STANDARDS FOR PROPOSED STUDY

Parameter	Average level	Maximum Acceptable Level	Remarks
pH	6.8-7.5	8	pH of less than 6.0 is not desirable. Levels below 6.3 may degrade performance
Total Coliforms(TC)	0CFU/ml	0CFU/ml	Presences of Coliforms affect growth of poults.
Dissolved Oxygen(DO)	4.0-6.0mg/L	6mg/l	Affects growth and performance of chickens.
Biochemical Oxygen Demand(BOD)	1.0-5.0mg/L	5.0mg/L	Affects growth and performance of chickens.
Temperature(Temp)	25°C	-	Affects feed conversion ratio and growth
Conductivity(CND)	300μS	-	Performance of chickens get reduced

Khullar et al. [12] focused on anticipating the water quality of the Yamuna River in India by using a deep learning based Bi-LSTM model. Generally, the existing schemes do not focus on missing value an imputation which degrades the reliability of the system. To overcome these issues, author demonstrated novel scheme that includes missing value imputation in the first phase, feature maps generated from the given input data in the second phase, Bi-LSTM architecture in the third phase to improve the learning process, and finally, an optimised loss function to reduce the training error in the fourth phase.

Kim et al. [13] developed a water forecasting method to improve the water quality analysis. The first phase is developed by using ensemble of empirical mode decomposition with Bidirectional LSTM model. Once the predictions are obtained, a novel error correction module is incorporated which uses variation mode decomposition and BLSTM neural network. Thus, this combination helps to achieve the less accuracy during forecasting.

Prasad et al. [14] reported that handling the complexity of existing schemes is cumbersome process. Thus, authors developed artificial intelligence based machine learning approach.

Choi et al. [15] adopted three standalone and a hybrid deep learning scheme for water quality analysis based on time series analysis. These methods include univariate dataset with single dependent variable, multivariate dataset with single and multiple dependent variables and identifying the other parameters which has impact on the performance.

Dimri et al. [16] presented water quality analysis for Ganga River by using multivariate statistical techniques such as Pearson correlation, principal component analysis and cluster analysis.

Several works have considered underground water for water quality analysis and presented machine learning based techniques such as Panneerselvam et al. [17] identified that the rock-water interaction and ion exchange are also considered as important factors which affect the water quality. To identify their impact authors developed a Hierarchical Cluster Analysis (HCA) and K-mean cluster analysis based approach.

Vijay et al. [18] presented artificial neural network based approach for water quality prediction. The preliminary steps include pre-processing, and feature extraction. The ANN classifier uses three different activation functions such as Tanh, rectifier, and Maxout to process these features.

Wang et al. [19] used an AI-based LSTM (long short-term memory network) to identify water pollutant features as well as identifying the industrial pollution source. In line to this, a correlation based map is used to extract the relation among attributes. The correlation map is used to trace the fluctuations and identify the location and causing agents.

In our previous work [20], experimental study was conducted to test the water quality of different sources used in poultry farms using sensors and laboratory methods.

III. PROPOSED MODEL

According to literature survey, several water quality parameters are analysed to characterise the water quality. In this work also, we consider various parameters as in Table to predict the water quality whether it is drinkable or not and to provide the recommendation for poultry farm. In order to perform this task, we have collected the samples of Cauvery River.

A. Study Area: Cauvery River

In proposed study, we considered the Cauvery River which is a great river of India flowing in Southern region is considered as Dakshin Ganga. The origin point of this river is Talakaveri which is in the Coorg district of Karnataka. The catchment area of this river includes Karnataka, Tamil Nadu, Pondicherry, and Kerala before merging in Bay of Bengal. These southern states have agriculture as its main occupation. These states have significant livestock population where poultry farming is also considered as an important segment and share 50% of total production value. In this regard, it is important to consider the quality of Cauvery River water flowing in these regions as it is primary source of water. The water quality can have impact on the egg yield and meat production thus analyzing water quality is most important task.

The Fig.1 shows the Cauvery River basin and its tributaries flowing from Karnataka to Tamil Nadu.

We have obtained the data from various locations [21] and these samples are processed through the water quality analyser where indicator parameters are extracted.

B. Architecture Diagram for the Proposed Model

The proposed model architecture is shown in Fig. 2 and each component is explained:

1) Feature ranking method: In this work, we adopt the Relief algorithm based feature selection method to identify the relevance of features. This approach computes the proxy statistics known as feature weights $W[A]$ for each feature to identify its relevance which is ranging from -1 to +1 i.e. worst to best. Generally, the feature relevance is assigned based on the difference of probabilities. The pseudo code of complete process of this method is presented Algorithm 1. This approach estimates the nearest instance of the target where one is with same class is known as nearest hit H and the opposite class is nearest miss M. In algorithm 1 Step 9 updates the weights of feature.

Fig. 1. Cauvery River Basin.
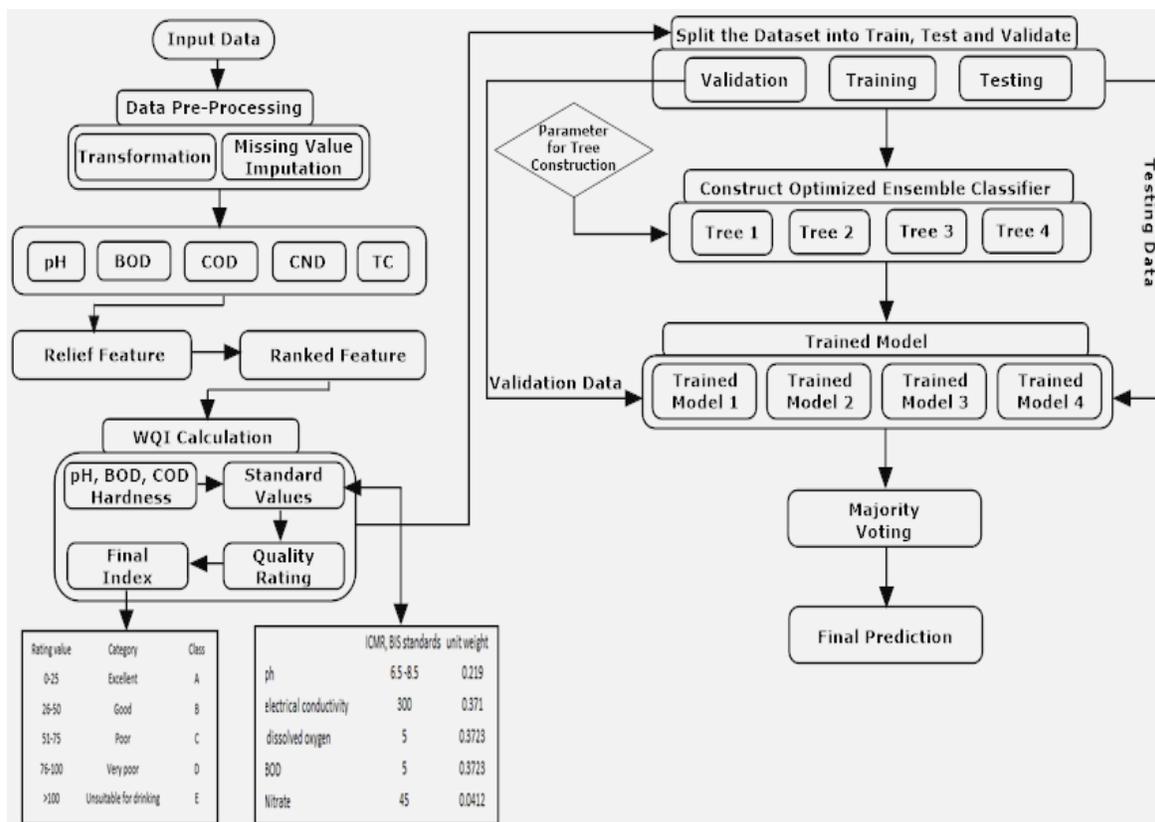
Fig. 2. Architecture of the Proposed Model.

Algorithm 1:Relief Algorithm

Step 1: Input: each training instance with feature values and corresponding class values.

Step 2: $n \leftarrow$ number of training instances

Step 3: $a \leftarrow$ number of attributes

Step 4: Output: m $\leftarrow$ pick random training from n and update the weights W

Step 5: for $i \coloneqq 1$ to m do

Step 6 : initialize the selection of random 'target' from R_i

Step 7: identify the nearest hit and miss as H and M

Step 8: For $A \coloneqq 1$ to a do

Step9: $W[A] = W[A] - diff\ (A, R_iH)/m + diff\ (A, R_iM)/m$

Step 10: End for

Step 11: End for

Step 12: Return updated weights to identify the quality of attributes

To perform weight updates, calculate the difference in value for feature A between two instances I_1 and I_2, where $I_1 = R_i$ and I_2 is either H or M, the diff function in Algorithm 2 is computed as:

$$diff(A, I_1, I_2) = \begin{cases} 0, if\ value\ (A, I_1) = value\ (A, I_2) \\ 1\ if\ otherwise \end{cases} \quad (1)$$

In order to further, improve the weight update process, we define a Bayes rule based on the conditional probability of prediction differences and probability of different class. The Bayes rule can be expressed as:

$$W[F] = \frac{P_{diffC|diffF}P_{diffF}}{P_{diffC}} - \frac{(1 - P_{diffC|diffF})P_{diffF}}{1 - P_{diffC}} \quad (2)$$

In order to calculate the difference between feature spaces, we used a weighted Euclidean distance measure which is calculated as:

$$diff(S_i, S_j) = \sqrt{\sum_k w(c_k)(L_{i,k} - L_{j,k})^2} \quad (3)$$

Where S_i and S_j represents the target description for random instance R_i and target instance I_j, respectively. $L_{i,j}$ and $L_{j,k}$ denotes their corresponding binary representation. The updated algorithm of k feature ranking and selection is given in Algorithm 2. Initialize random instance R_i and find the k nearest instances I_j. Using these instances, approximate the relevance $W[F]$ from Equation 2 for each feature by calculating $N_{diff\ C}$, $N_{diff\ F}[F]$ and $N_{diffC\ \&\ diffF}[F]$ described in lines 6, 8 and 9 of Algorithm 2. The values of these estimations are based on the calculation of distance in the feature space, $diff(F, R_i, I_j)$ as specified in lines 8 and 9 of Algorithm 2.

Algorithm 2: Pseudo code for Feature ranking and selection

Input: training instances and corresponding class values
Output: weight vector W to estimate relevance of features.

Step 1: initialize $N_{diff\ C}$, $N_{diff\ F}$, $W[F]$ to 0

Step 2: for $i = 1\ to\ m$ do

Step 3: initialize with randomly selected instances R_i

Step 4: select k instances from I_j as nearest instances to R_i

Step 5: for j=1 to m do

Step 6: $N_{diffc} = N_{diffc} + diff(S_i, S_j).d(i,j)$

Step 7: for $F = 1\ to\ f$ do

Step 8: $d_{diffF}[F] = N_{diffF}[F] + diff\ (S_i, S_j).d(i,j)$

Step9: $N_{diffC\ \&\ diffF}[F] = N_{diffC\ \&\ diffF}[F] + diff(S_i, S_j).diff(F, R_i, I_j).d(i,j)$

Step 10: end for

Step 11: end for

Step 12: end for

Step 13: for $F = 1$ to f do

Step14: $W[F] = N_{diffc}|diffF\ [F]/N_{diffc} - (N_{diffF}[F] - N_{diffc|diffF}[F])/(m - N_{diffc})$

Step 15: end for

2) Water quality index calculation: In this section, we consider the water quality parameters such as conductivity, hardness, nitrates, BOD, pH, and dissolved oxygen and presented a method to obtain the water quality index which can be used to classify the potability of water. In this work, we have used weighted arithmetic method to compute WQI which is expressed as:

$$WQI = \frac{\sum Q_n W_n}{\sum W_n}$$

□□□

Where, Q_n is the quality rating and W_n is the weight for each quality parameter. The quality rating can be calculated based on the observed value and ideal values. It can be computed as follows:

$$Q_n = 100 \left[\frac{V_n - V_0}{S_n - V_0}\right] \quad (5)$$

Where V_n is the observed value, V_0 is the ideal value (V_0 value is 7 for pH and 14.6 for DO and 0 for others). Similarly, W_n is the unit weight vector which is computed as:

$$W_n = \frac{1}{S_n} \quad (6)$$

Where S_n is the recommended value such as WHO temperature standard value is 28, then $W_n = \frac{1}{28} = 0.036$, DO recommended value is 6 then $W_n = \frac{1}{6} = 0.166$.

By substituting the values of (5) and (6) in (4), we obtain the WQI parameter and based on this value the water grading can be presented. Below given Table V shows the quality index parameters.

TABLE V. WQI WATER QUALITY GRADING

WQI Value	Comment	Grading	Labels
0-25	Excellent	A	1
26-50	Good	B	2
51-75	Poor	C	3
76-100	Very Poor	D	4
Above 100	Unsuitable	E	5

A sample dataset used for training the network is obtained as presented in Table VI which considers pH, DO(Dissolved oxygen), TC(total colifroms), BOD(biochemical oxygen demand), CND(conductivity) and temperature to measure the WQI value.

TABLE VI. WATER QUALITY PARAMETERS

pH	DO	TC	BOD	CND	Temp	WQI
7.4	10.2	6800	1	445	24.5	1
8.2	8.4	110000	2	514	25	2
8.3	9.1	1300	1	449	26	1
7.8	13.1	1100	6	374	27.5	2
8.3	7.7	7800	4	730	27.5	2
8.3	7.7	7800	4	730	28	2
7.3	6.1	9200	2	236	27	1
7.3	6.1	9200	2	236	26	1
7.8	6.4	17000	1	300	27	2
7.7	7.4	3300	1	401	26.5	1
8.5	9.5	1700	2	460	27	1
7.7	11.5	3300	4	710	27.5	1

3) Classifier model: In this work, we present an ensemble classifier where multiple instances of random forest model are constructed which produces the preliminary classification results. Later, these predicted outcomes are processed through the majority voting where final prediction is computed. Fig. 3 depicts the ensemble of tree classifier.

The Random Forest is a promising technique which is a collective learning method used for classification, regression, and other machine learning tasks. It imitates with cluster formation of decision trees and generates a class as mean of individual tree. This approach follows specific these steps which are mentioned below:

- First stage: Random Forest Tree construction.

In order to construct the tree, this approach performs four operations as:

- Selection of k random features from the m set of features.
- Determining d as the best point for split

- These nodes are further distributed as daughter nodes
- This process is repeated until l numbers of nodes are obtained via tree construction.
- Second stage: Classification

In next stage, the constructed trees are considered as inputs and classification is performed by following the steps mentioned below:

- Rules are assigned for each tree and a decision tree is formulated based on test features.
- Votes are calculated according to the target value
- The target which obtains the highest votes is considered as final result.
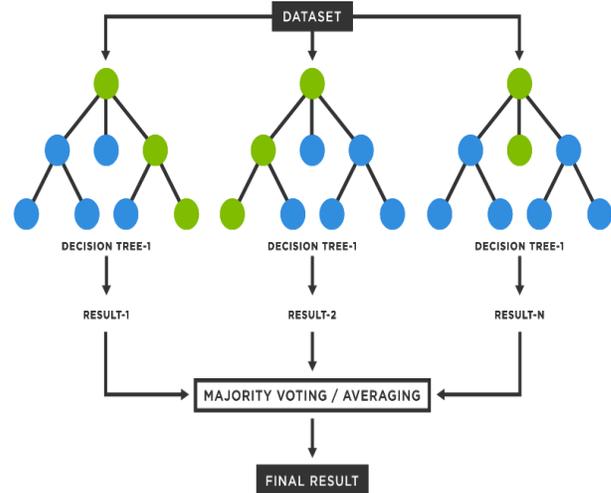
Fig. 3. Random Forest Tree Ensemble Classifier.

IV. RESULTS AND DISCUSSION

This section presents the complete outcome of proposed approach. The proposed approach uses WQI dataset generation as the first step and later classifiers are trained to obtain the predication performance. In this work, we have considered pH, electrical conductivity, dissolved oxygen, BOD, and nitrate. For these parameters, ICMR has given some standard values for identifying the water quality. These standard values are presented in Table VII.

Based on these standards, we compute the water quality parameters. Samples of generated and their calculation example for WQI data is presented in below given Table VIII.

TABLE VII. STANDARDS OF ICMR FOR WATER QUALITY ANALYSIS

	ICMR, BIS standards	unit weight
pH	6.5 -8.5	0.219
electrical conductivity	300	0.371
dissolved oxygen	5	0.3723
BOD	5	0.3723
Nitrate	45	0.0412

TABLE VIII.   WQI DATASET GENERATION

Parameter Name	Standard value	Observed value	Unit Weight	Quality rating	$W_i Q_i$
pH	6.5-8.5	7.7	0.2190	46.67	10.22
Conductivity	250 $\mu S/cm$	$96.3 \mu S/cm$	0.3710	38.52	14.29
Dissolved oxygen	$500 mg/L$	$3.2 mg/L$	0.3723	118.75	44.21
BOD	$5 mg/L$	$1.1 mg/L$	0.3723	22	8.19
Nitrate	$1.70 mg/L$	$1.7 mg/L$	0.0412	3.40	0.14
Temperature	28	27.04	0.036	96.57	3.476
			1.4118		80.52

Based on these values, we compute the WQI parameter as:

$$WQI = \frac{80.52}{1.4118} = 57.033$$

Similarly, complete dataset is generated and trained.

1) Performance measurement: The classification performance is measured using performance metrics such as accuracy, specificity, sensitivity, precision, recall and F-score. These values can be computed with the help of confusion matrix which is given in Fig. 4.

Accuracy is defined as a rate of correct classification and it is denoted by Accuracy. It is calculated by taking the ratio of correct prediction and total number of prediction [22]. It can be expressed as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (7)$$

Sensitivity is another parameter used to do the sensitivity analysis of the model. It is used to measure true positive rate. This can be expressed as:

$$Sensitivity = \frac{TP}{TP+FN} \qquad (8)$$

Fig. 4.   Confusion Matrix.

Then, the proposed approach precision can be computed by taking the ratio of True Positive and (True and False) False positives.

$$P = \frac{TP}{TP+FP} \qquad (9)$$

Similarly, recall is computed based on TP, and FN, as given below:

$$Recall = \frac{TP}{TP+FN} \qquad (10)$$

Finally, F-measure is calculated using the mean of precision and sensitivity performance. It is expressed as:

$$F = \frac{2*P*Sensitivity}{P+Sensitivity} \qquad (11)$$

2) Comparative analysis: This section presents the comparative analysis by considering different classifiers such as K-Nearest Neighbour (KNN), Naïve Bayes, Decision Tree and proposed ensemble approach. First of all, we present an experiment, where feature selection algorithm is not combined with the classification module. The classification performance in terms of percentage is presented in below given Table IX.

TABLE IX.   COMPARATIVE PERFORMANCE WITHOUT FEATURE SELECTION

	KNN	Naïve Bayes	Decision Tree	Ensemble Tree Classifier
Accuracy	92.11	93.51	94.5	95.12
Sensitivity	87.5	91.2	93.56	94.8
Specificity	91.2	93.5	95.2	96.1
Precision	92.2	93.8	95.6	96.5
F-score	93.5	94.25	95.05	97.5

Below given Fig. 5 shows the graphical representation of different classifiers where feature selection approach is not incorporated.

Further, we extend this experiment and incorporated feature selection approach with different classifiers. The obtained performance in terms of percentage is presented in below given Table X.
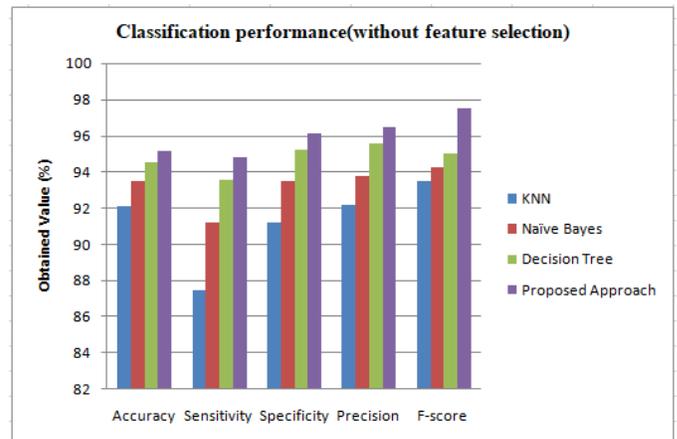
Fig. 5.   Classifier Performance without Feature Selection.

TABLE X.   COMPARATIVE PERFORMANCE WITH FEATURE SELECTION

	KNN	Naïve Bayes	Decision Tree	Ensemble Tree Classifier
Accuracy	93.52	94.63	95.86	96.55
Sensitivity	88.12	93.1	94.52	94.8
Specificity	92.5	94.5	96.12	97.22
Precision	93.42	94.68	97.1	97.3
F-score	94.6	95.2	96.3	98.21

Similarly, the overall performance with feature selection strategy is depicted in Fig. 6.
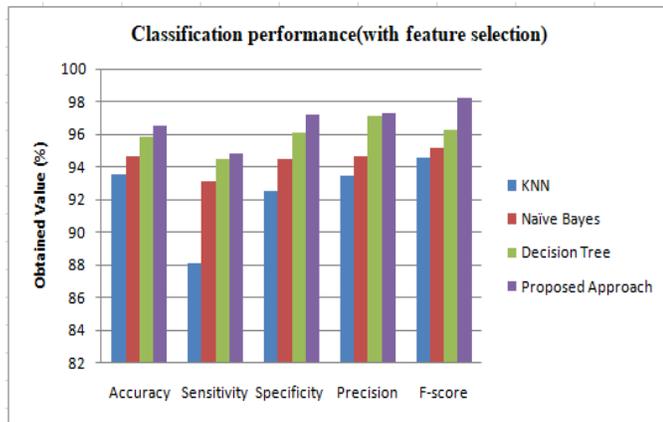
Fig. 6.   Classifier Performance with Feature Selection.

These analysis results shows that the proposed feature selection and classification approach improve the classification accuracy when compared with the conventional methods.

This model can be used by the poultry farmers to analyze the quality of water used in the farms in order to reduce the effect of water borne diseases which lead to decrease in egg quality and egg production in layer poultry farms.

V. CONCLUSION

In our work, the suitability of Cauvery River water for layer poultry farm is carried out. Several methods are present for predicting the water quality analysis but accuracy and reliability become a challenging task.

The proposed machine learning model introduces a feature ranking model along with the ensemble tree based classifier for water quality analysis and prediction. The comparative analysis shows the significant improvement in prediction performance when compared with the conventional machine learning algorithms. This automated approach can be used by the poultry farmers to analyze and predict the potability of water used in the farms.

REFERENCES

[1] https://www.statista.com/statistics/237597/leading-10-countries-worldwide-in-poultry-meat-production-in-2007.

[2] Grakh, K., Mittal, D., Prakash, A., & Jindal, N. (2022). Characterization and antimicrobial susceptibility of bio film-producing Avian Pathogenic Escherichia coli from broiler chickens and their environment in India. Veterinary Research Communications, 1-12.

[3] www.krishijagran.com.

[4] https://vikaspedia.in/agriculture/agri-directory/reports-and-policy-briefs/20th-livestock-census.

[5] https://www.lkouniv.ac.in/site/writereaddata/siteContent/202004221610 299468katiyar_bot_Poultry_Industry_in_India.pdf.

[6] Wali, S. U., Umar, K. J., Abubakar, S. D., Ifabiyi, I. P., Dankani, I. M., Shera, I. M., & Yauri, S. G. (2019). Hydro chemical characterization of shallow and deep groundwater in Basement Complex areas of southern Kebbi State, Sokoto Basin, Nigeria. Applied Water Science, 9(8), 1-36.

[7] Van Limbergen, T., Sarrazin, S., Chantziaras, I., Dewulf, J., Ducatelle, R., Kyriazakis, I., ... & Maes, D. (2020). Risk factors for poor health and performance in European broiler production systems. BMC veterinary research, 16(1), 1-13.

[8] Ranjan, A., Sinha, R., Devi, I., Rahim, A., & Tiwari, S. (2019). Effect of heat stress on poultry production and their managemental approaches. Int. J. Curr. Microbiol. Appl. Sci, 8, 1548-1555.

[9] https://conservancy.umn.edu/bitstream/handle/11299/204342/SF95_M6 58a-73 2012_magr56317.pdf.

[10] Pujar, P. M., Kenchannavar, H. H., Kulkarni, R. M., & Kulkarni, U. P. (2020). Real-time water quality monitoring through Internet of Things and ANOVA-based analysis: a case study on river Krishna. Applied Water Science, 10(1), 1-16.

[11] Nayak, J. G., Patil, L. G., & Patki, V. K. (2020). Development of water quality index for Godavari River (India) based on fuzzy inference system. Groundwater for Sustainable Development, 10, 100350.

[12] Khullar, S., & Singh, N. (2022). Water quality assessment of a river using deep learning Bi-LSTM methodology: forecasting and validation. Environmental Science and Pollution Research, 29(9), 12875-12889.

[13] Kim, J., Yu, J., Kang, C., Ryang, G., Wei, Y., & Wang, X. (2022). A novel hybrid water quality forecast model based on real-time data decomposition and error correction. Process Safety and Environmental Protection.

[14] Prasad, D. V. V., Venkataramana, L. Y., Kumar, P. S., Prasannamedha, G., Harshana, S., Srividya, S. J., ... & Indraganti, S. (2022). Analysis and prediction of water quality using deep learning and auto deep learning techniques. Science of the Total Environment, 153311.

[15] Choi, H., Suh, S. I., Kim, S. H., Han, E. J., & Ki, S. J. (2021). Assessing the Performance of Deep Learning Algorithms for Short-Term Surface Water Quality Prediction. Sustainability, 13(19), 10690.

[16] Dimri, D., Daverey, A., Kumar, A., & Sharma, A. (2021). Monitoring water quality of River Ganga using multivariate techniques and WQI (Water Quality Index) in Western Himalayan region of Uttarakhand, India. Environmental Nanotechnology, Monitoring & Management, 15, 100375.

[17] Panneerselvam, B., Muniraj, K., Pande, C., & Ravichandran, N. (2021). Prediction and evaluation of groundwater characteristics using the radial basic model in Semi-arid region, India. International Journal of Environmental Analytical Chemistry, 1-17.

[18] Vijay, S., & Kamaraj, K. (2021). Prediction of water quality index in drinking water distribution system using activation functions based Ann. Water Resources Management, 35(2), 535-553.

[19] Wang P, Yao J, Wang G, Hao F, Shrestha S, Xue B, Xie G, Peng Y (2019) Exploring the application of artificial intelligence technology for identification of water pollution characteristics and tracing the source of water quality pollutants. Sci Total Environ 693:133440.

[20] Deepika, Nagarathna, Shivshankar, Channegowda, (2019). Experimental Investigations on Quality of Water Used in Poultry Farm Using Sensors. 10.1007/978-981-13-5802-9_40.

[21] https://www.cpcb.nic.in.

[22] C. Manjula, Lilly Florence. "Deep neural network based hybrid approach for software defect prediction using software metrics" Cluster Computing, 2018.

Detection of Abnormal Human Behavior in Video Images based on a Hybrid Approach

BAI Ya-meng*, WANG Yang, WU Shen-shen

School of Information Engineering, Jiaozuo University, Jiaozuo 412000, China

Abstract—The analysis of human movement has attracted the attention of many scholars of various disciplines today. The purpose of such systems is to perceive human behavior from a sequence of video images. They monitor the population to find common properties among pedestrians on the scene. In video surveillance, the main purpose of detecting specific or malicious events is to help security personnel. Different methods have been used to detect human behavior from images. This paper has used an efficient computational algorithm for detecting anomalies in video images based on the combined approach of the differential evolution algorithm and cellular neural network. In this method, the input image's gray-level image is first generated. Because it may be possible to identify several large areas in the image after the threshold, the largest white area is selected as the target area. The images are then used to remove noise, smooth the image, and fade the morphology. The results showed that the proposed method has higher speed and accuracy than other methods. The advantage of the algorithm is that it has a runtime of three seconds on a home computer, and the average sensitivity criterion is 98.6% (97.2%).

Keywords—*Cellular neural network; detection of abnormalities; differential evolution algorithm; video images*

I. INTRODUCTION

The analysis and investigation of human movements have attracted the attention of many researchers in different fields. Such systems aim to understand human behavior from a sequence of video images [1, 2]. Security cameras are installed in many organizations that are important for security and can monitor and record the situation of a place at all times. Human behavior detection systems are also one of the most important needs of surveillance systems in the sequence of video images captured by surveillance cameras [3, 4, 5]. Usually, all automatic detection or monitoring systems limit the detection mechanisms of the desired objects in the field of view to the tracking and activities of objects in the same image areas. Crowd monitoring is used to identify general characteristics among passers-by in a scene [6, 7, 8, 9]. These properties include population size, density, growth rate, traffic patterns, and the detection of abnormal events. In video surveillance, the main goal is to detect specific or abnormal events to help security personnel. Surveillance systems in use today rely on the performance of a human operator. They are expected to view many screens recorded by different cameras, often simultaneously. Generally, classification is one of the most important and difficult steps in image processing [10, 11]. Various methods have been used to recognize human behavior from images, but some of these methods do not show optimal performance in some situations, and in some cases,

the methods used are either not able to detect people under certain conditions or do not have high accuracy [12, 13, 14, 15, 16]. Another problem with several methods is that they require a powerful processing unit to achieve the desired detection speed. Among the significant problems, when there is a need to model busy and complex scenes due to occlusion, the use of background reduction to classify motion areas and the trajectory method for feature extraction fail [17, 18, 19]. Some of these problems can be seen in Fig. 1. As seen in Fig. 1, people have much variety in the shape and color of their clothes. Also, different lighting conditions can be seen in the images. There is also a partial or total overlap of the person's image with other objects and people [20, 21, 22, 23]. The superiority of cellular neural networks in low-level image processing over common digital image processing systems is due to two features of this network. These features include parallel processing of the analog image signal and cross-sectional or local connections among network cells, making it easy to build cellular neural networks. In this research, abnormalities will be detected from video images. For this, the cellular neural network method will be used. Since this network needs optimized parameters, the differential evolution algorithm will be used to optimize the features for training. This algorithm is similar to the genetic algorithm, except that changes have been made in the operations of mutation and combination. The differential evolution algorithm will use the training data set to calculate the fitness function for the produced chromosome. Based on this, the template generated by the differential evolution algorithm is coded. Then, one of the two main images from the database is randomly selected. After producing a suitable template for the cellular neural network, all the images in the database will be given to the network for segmentation. Recognizing moving objects and people in the dark or in environments with insufficient light is one of the limitations of this research.

A. Motivation and Contributions

Detecting and tracking moving people (such as the abnormality of a person in several consecutive images of a scene) has many challenges, including the great variety of people in terms of shape, color, and clothing. It is also possible that, according to the angle of placement of people in front of the camera, the person's appearance may change, or a person may overlap with another person or other objects in the scene. For this reason, it isn't easy to detect and recognize people in crowded environments. Therefore, local motion information should be extracted at the pixel level using spatial and temporal detectors and the optical flow characteristics of shape, color, and clothing to extract features in such scenes. It

---

*Corresponding Author.

is also possible that, according to the angle of placement of people in front of the camera, the person's appearance may change, or a person may overlap with another person or other objects in the scene.

For this reason, it isn't easy to detect and recognize people in crowded environments. Therefore, local motion information should be extracted at the pixel level using spatial and temporal detectors and optical flow characteristics to extract features in such scenes. It will be possible to view the video image. The contributions of the authors in the article are as follows:

- Using a cellular neural network and a differential evolution algorithm to improve the accuracy of finding things that aren't normal.

- Increasing classification sensitivity in abnormality detection based on video images.

The continuation of the paper is configured as follows: The second section discusses previous work in detecting abnormal human behavior in video images. In the third part, the proposed method is given in detail. In the fourth part, the results of the proposed method are compared with similar methods. Finally, general conclusions and suggestions for future work are stated in the fifth part.

Fig. 1.   Complications in Identifying People [1].

II.   RELATED WORKS

This section introduces some methods to detect human behavior abnormalities in video images. It is presented a method for detecting local and series anomalies in crowded scenes using k=24 bucket representation with sparse (MHOF). In this method, the multi-comparison histogram of the optical flow on three scales is used to accurately store the movement direction information and the movement energy information to express the characteristics [24]. Another support vector regression machine (SVR) method has been used for pedestrian detection. In this paper, to extract the feature, the features of the Harr wavelet and directional histogram have been used, and in the classification section, in addition to using SVR, the classifier K-nearest neighbor (KNN) has also been used. The results of the experiments have shown the high accuracy of the algorithm presented in this paper [25]. Researchers used fuzzy logic to detect anomalies in human behavior. In this paper, to determine which rules are suitable, the effectiveness of each rule was evaluated based on the speed and environmental conditions of the pedestrian. This method is very time-consuming due to the type of search that requires the search window to be moved pixel-by-pixel in both horizontal and vertical directions and is not suitable for real-time or live applications [26]. Also, researchers have used graph formulation to diagnose human behavioral abnormalities. In this method, a model of the appearance and dynamics of the scene is created, and the features of the pixel level are used. Each designated location defines the center of the video cell, from which the edges and vertices of the graph are extracted. A tree is trained during the training phase. The probability of each of the obtained spatio-temporal volumes and the representation of its salient points are generated. An anomaly is where the probability is less than a threshold. The results showed that using graphs provided complete coverage of the problem space. However, it takes a long time to process

and calculate [27]. In this research, optical flow and gradient have been used to detect anomalies. As a result, among these methods, a bag of words is a suitable method for distinguishing abnormal combinations from normal events, especially in complex scenes that include different classes of objects and events. Like anomalies in traffic scenes, a method for rapid anomaly detection for surveillance systems, with real-time capability and reliable detection and localization, is proposed. Each event includes a set of spatio-temporal volumes. This method creates a model of normative behavior by initializing the algorithm with a small number of video sequences, then seeks to detect anomalies and gradually updates the model [28]. Researchers tried to speed up pedestrian detection by using a neural network. In this method, they were recognized based on the rescaling method in deep learning and by sharing features across the model. The results of this method have been better than the normal neural network in terms of processing speed [29]. In order to build an advanced driver assistance system (ADAS), researchers introduced a method based on fuzzy clustering and convolutional neural networks. In this paper, the candidate areas are revealed first. Then a quick method is used to extract the features. In order to extract the candidate areas, the fuzzy clustering method is used. After the features are extracted, to check whether there are pedestrians in those candidate areas or not, neural networks have been used. This type of neural network is made by imposing restrictions on weights and how to connect the neurons of a standard neural network and creates a network suitable for processing data with temporal or spatial distribution [30]. In another study, pedestrians were recognized based on gradient and texture features. This method used principal component analysis (PCA) to reduce the dimensions of textural and rotational features. Experiments on two different data sets showed the better efficiency of this approach compared to similar methods [31]. A robust method for human detection in underground image sequences that

works based on learning systems was presented. An SVM support vector machine classifier is used for detection, so that after the preprocessing process, the patterns of the image in which human presence is likely are extracted and fed into the SVM. The OSU pedestrian database has been used to learn and test the proposed algorithm. The results of the implementation of the presented algorithm on this database show the efficiency and appropriateness of the algorithm [32]. They presented another method to remove the background in the images of the outside environment that does not have a fixed background. There are things like the movement of leaves and moving flags in them. Their main idea is that the pixels of the neighboring blocks, which are assumed to be the background, have similar changes over time and that the same patterns can be extracted from them. Although this argument is correct for the blocks extracted from an object in the background, it does not seem correct for the blocks that fall on the boundary of the objects separately due to different changes on both sides of the boundary. They are used to support pedestrian detection based on underground images. This method was performed on a dataset with 18 images, and the detection accuracy of this method was 98.11% [33]. The study's [34] objective is to train the detector to identify a human action even when only a partial action example is present. To achieve this, a hybrid approach based on the fuzzy Bandler subproduct and the Kohut triangle is proposed in this work that combines the benefits of computer vision and fuzzy set theory. Building a frame-by-frame membership function for each potential motion type is novel. When a predetermined threshold is appropriately reached, detection is started. The benefits and efficacy of the suggested method are demonstrated by experimental results on a dataset that is publicly accessible. Source [35] offers a novel method for identifying and detecting human behavior. This work's major goal is to show how to identify abnormal behavior in a constrained domain. The suggested method automatically detects people in regular films, or video surveillance feeds. When a person is considered in a frame, the action is recognized, and the human position is approximated. After that, the behavior is categorized as normal or abnormal. The models included in the proposed work include SSD MOBILE NET for human action identification and the FAST R-CNN model for human detection in a video frame or image. Everyone needs to feel safe and secure at all times in our century. Many nations have purchased and implemented surveillance systems to secure their environment using high-definition CCTV cameras. As a result, automated CCTV surveillance systems can operate as security providers by spotting suspicious human behavior or intruder activity using CCTV footage. This is a difficult task. Models of suspicious behavior that only partially employ machine learning techniques can be used to do this. It has been discovered that several of the prior methods used techniques from deep learning, machine learning, IoT, and fuzzy logic. The proposed suspicious behavior detection model (SHBDM) efficiently extracts picture features using a CNN model pre-trained on the ImageNet dataset known as Inception V3 (VGG-16). Python was used to create this system using an open-source framework. Compared to VGG-16 + LSTM and a straightforward CNN model, the precision accuracy of the

system using Inception V3 + LSTM was enhanced by 88.8% [36]. Table I is an overview of the detection of abnormal human behavior in video images, whose problems and solutions have been studied in previous works.

TABLE I. OVERVIEW OF DETECTION OF ABNORMAL HUMAN BEHAVIOR IN VIDEO IMAGES

Work	Method	Solution
[37]	large numbers of people who never stop moving. The vertical arms and legs reaching out from the person distort the blobs' alignment.	Extrinsic appearance (geometry) and intrinsic appearance are used to parameterize the current mark
[38]	High volume video data, for example, has intrinsic non-linearity, spatial localization of patterns, and noise, all of which must be dealt with by the complex interplay between Multiobject and the uncontrolled scene.	Demonstrate a subspace detector that learns non-linearly
[39]	Use of Deep Metric Learning to Take Advantage of Label Correlation in a Transferred-Mediface Sample	The current bottom-up method focuses on a person's short-term behavior and trajectory.
[40]	It's challenging to make direct comparisons between motion and appearance representations, which are usually adapted to a certain scene domain.	Show a unified (temporal and spatial) model of form and movement (Mixture of dynamic texture)
[41]	Detection in a crowd scene may be impacted by factors like occlusion and fluctuating lighting.	The foreground was extracted using a double filtering technique, and the noise was diminished using a media filter.
[42]	Crowds make it hard to identify and follow individual people. When faced with unexpected situations, learning, adaptive, and incremental approaches are not effective.	Exhibit a tiered strategy encompassing low, medium, and high
[43]	Due to the intricate temporal dynamics and interactions among the behavior of numerous objects, anomalies tend to be subtle.	Implementing a Staggering Arrangement of Evolving Bayesian Networks
[44]	Analysis of crowd behavior is challenging. Complex behaviors such as line formation, laminar and turbulent flow, arching and clogging at exists, jamming around barriers, and panic have been seen in the movements of crowds.	A grid of particles was placed on top of the image, and the particles were advected using the average space-time optical flow.
[45]	Normal behavior in a video scenario is difficult to predict, necessitating intensive labeling and training. Due to the high number of objects in a crowd scene, it is challenging to precisely monitor individual objects.	Using a patch-based local motion representation, introduce the novel idea of contextual anomaly into crowd analysis.
[46]	In occlusions, the mobility of the people and things in the scene is exceedingly erratic. In a film with that many people, dissecting the behavior of each individual is an arduous process.	Exhibit a 3D Gaussian gradient in space and time that is rich, non-uniform, and localized.

III. SUGGESTED APPROACH

The aim of this research is to increase the accuracy of abnormality detection using a cellular neural network based on video images. Observations show that, when there is a need to model busy and complex scenes, due to the presence of occlusion, the use of background subtraction for zoning motion regions and the trajectory method for feature extraction fail. So, to get features out of these kinds of scenes, local motion information should be extracted at the pixel level with the help of spatial and temporal detectors and optical flow features. The proposed detection system provides an improved method for feature selection using the differential evolution algorithm in the image using a cellular neural network. Fig. 2 shows the block diagram of the proposed approach.

B. Reading the Image

The first step in the proposed approach is to pre-process the video image to obtain the anomaly region. Since the changes in the intensity of the gray levels in the images prevent the accurate detection of the area related to the abnormality, therefore, in the proposed approach, the lighting improvement technique is used to linearly map the intensity values in the image to new values that are normally distributed in the range of 0 to 255. are used. Also, to reduce existing noise and improve the image, a common median filter has been used.

C. Noise Reduction

The histogram is balanced by increasing the intensity values with the highest frequency. This method is useful in images where both background and foreground are dark, or both are light. In particular, this method can lead to better details in images produced under or overexposed to X-rays. In this paper, a median filter is used to reduce its noise. This filter is a non-linear digital image filtering method that considers each pixel and examines its nearest neighbors to determine a median value. Then the value of the pixels is replaced with the determined median value.

D. Increase Contrast

The first step in pre-processing is to use Gaussian function to increase the contrast in the original image. It is used before identifying the edges. Gaussian function in two-dimensional space is shown as Eq. (1) [4].

$$G_\sigma = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{i^2+j^2}{2\pi\sigma^2}\right)} \quad \text{،} i = 1\text{،}2\text{،}...\text{،}n\text{،} \; j = 1\text{،}2\text{،}...\text{،}m \qquad (1)$$

Gaussian filter plays an important role in many feature extraction methods in such a way that the implementation of its function is equivalent to filtering the image using the Gaussian kernel and non-uniform sampling of the image.

E. Feature Extraction

The shape of objects is one of the important features in detecting anomalies in video images. Usually, anomalies with similar shapes belong to several classes. Therefore, in this research, most of the extracted features are related to the shape of the abnormality.
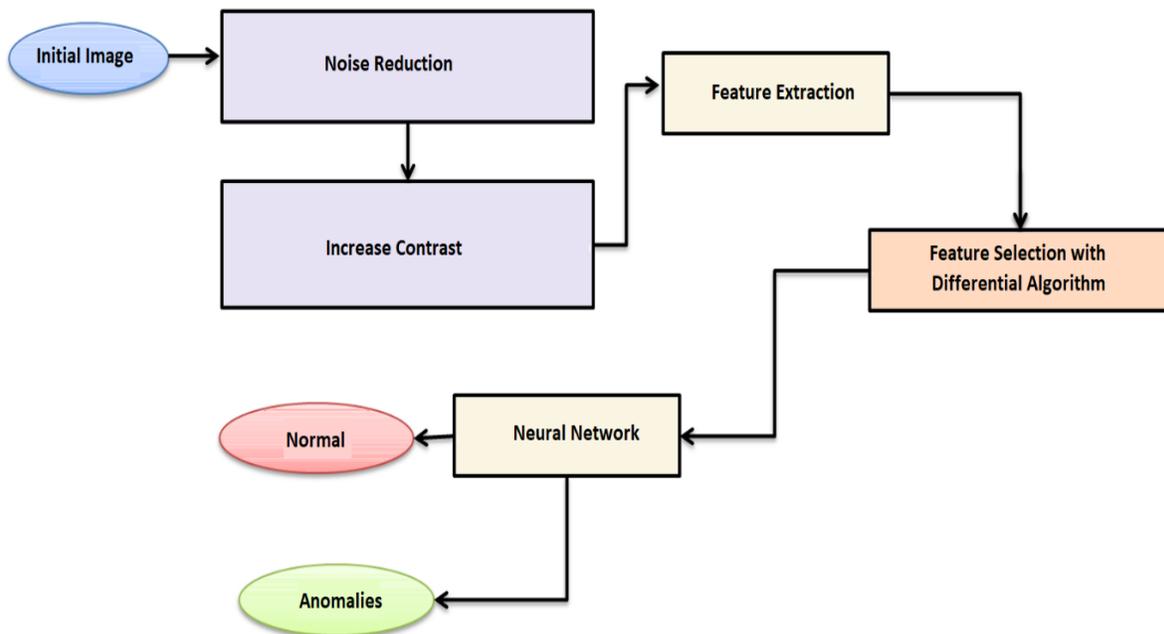
Fig. 2. Steps of the Proposed Method.
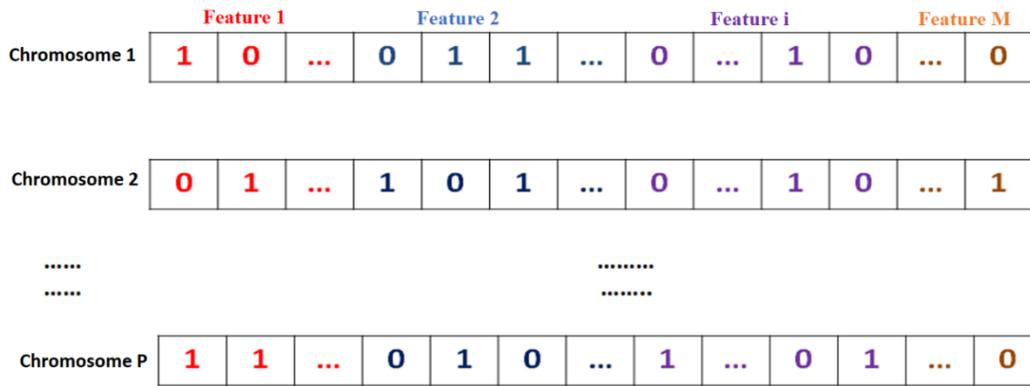
Fig. 3. Structure of Chromosomes in the Proposed Approach.

F. Feature Selection

Feature extraction is a process in which its salient and defining features are determined by performing an operation on the data. Feature extraction aims to make the raw data more usable for further statistical processing, classification, and clustering. Due to high convergence, the differential evolution algorithm was used at this stage to select basic features from among the extracted features. The dimensions of the images are 480 x 640. Due to the high and variable number of features in the images, this step on the image is necessary thing in order to be able to recognize the abnormality from the extracted features. In the differential evolution algorithm, a measure called mutual information according to Eq. (2) is used to evaluate (fitness calculation) of each chromosome produced in the algorithm:

$$MI(x; y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log(\frac{p(x,y)}{p(x)p(y)}) \qquad (2)$$

where, generally, p(t) is the probability of variable t.

In order to further investigate and evaluate the performance of the differential evolution algorithm in the feature selection stage, in addition to using the mutual information merit function, the Merit criterion according to Eq. (3) was also used as the merit function in the proposed method (main fitness function).

$$Merit = \frac{k\overline{R_{cf}}}{\sqrt{k+k(k-1)R_{ff}}} \qquad (3)$$

The differential evolution algorithm uses the generated training set to calculate the fitness function for the generated chromosome. Therefore, the evaluation of the chromosome to be in the fitness function includes the following three steps:

- A differential evolution algorithm codes the generated template.

- The selection of one of the two main images was randomly selected from the database and used by the differential evolution algorithm. In the next step, the evaluation begins.

- The output image produced by CNN to evaluate the chromosome fit value.

After generating a suitable template for CNN, all the images in two different databases were given to the CNN algorithm for classification. Also, the structure of chromosomes in the M-dimensional space of features is shown in Fig. 3. The features are coded as chromosomes in Fig. 3 to search in the space of the differential evolution algorithm. A unique code is considered for each feature in all P chromosomes.

G. Neural Network

In this research, the classification and diagnosis of the type of abnormality are based on the cellular neural network. For image processing purposes, the most common architecture is a two-dimensional network in which each processing unit interacts only with the M × N pixel cells that are its neighbors in the cell's permeation domain. Having a cellular network is a set of cells that has the property expressed in relation (4) C_{ij} for the cell $r \geq 0$ with radius $S_{ij}(r)$ neighbors:

$$S_{ij}(r) = \{C_{kl}: max(|k-i|, |l-j|) \leq r ، 1 \leq k \leq M ، 1 \leq l \leq N\} \qquad (4)$$

That the stimulus $u_{ij}(t)$ which is generally not visible, the input $x_{ij}(t)$ using the mode c_{ij} of each cell that shows a measurable and observable value, and the input $y_{ij}(t)$ is external that enters the cell, the output is M × N using the coupled M × N differential equation called bias. The neural network dynamics describe the evolution of each cell's state and its interaction with its neighbors. For image processing purposes, the most common architecture is a two-dimensional network in which each processing unit interacts only with the M × N pixel cells that are its neighbors in the cell's permeation domain. Having a cellular network is a set of cells that has the property expressed in relation (5) C_{ij} for the cell r≥0 with radius $S_{ij}(r)$ neighbors:

$$S_{ij}(r) = \{C_{kl}: max(|k-i|, |l-j|) \leq r , 1 \leq k \leq M , 1 \leq l \leq N\} \qquad (5)$$

The infiltration area with radius r = 1 corresponds to the neighborhood of 3x3, r = 2 to the neighborhood of 5x5 (Fig. 4).

That the stimulus $u_{ij}(t)$ which is generally not observable, the input $x_{ij}(t)$ using 4-time variables of the state c_{ij} of each cell that shows a measurable and observable value, and the

input $y_{ij}(t)$ is external to the cell is input, the output is M × N using the coupled M × N differential equation called bias. The neural network dynamics $z_{ij}(t)$ is described as M × N redundancy, which describes the evolution of the state of each cell and its interaction with its neighbors. It is used to process an image with pixels. In order to classify humans automatically in M × N pixel images, the intensity value of each pixel is given to humans from a cellular neural network containing the suspicious area as the input of the corresponding cell in the cellular neural network. The neural network used in this method is a network with two neural layers and three neural layers, which has the function of activating linear hidden layers and sigmoid output. Table II shows the parameters used for the neural network algorithm.

The neural network used in this research is considered in Fig. 5. For this model, relation (6) holds true:

$$\hat{y}_j(t+1) = \sum_{i=1}^{l} G_i w_{ij} = \sum_{i=1}^{l} w_{ij} \exp\left(-\frac{\|x - m_i\|^2}{2\sigma_i^2}\right) \quad (6)$$

In this regard, $x(t) = [x_1(t) \dots x_m(t)]^T$ indicates the input vector, $\hat{y}_j(t+1)$, indicates the jth output, w_{ij}, indicates the synaptic weight between neurons i-th hidden and j-th output neuron, G_i, the Gaussian function in the i-th neuron of the hidden layer, m_i and σ_i^2 respectively represent the center and width of the Gaussian function and l, the number of the Gaussian function, or in other words, equal to the number of hidden layer nodes.

Fig. 4. Neighborhood of Cell c(i,j) r = 1, r = 2 Respectively.

TABLE II. PARAMETERS USED FOR NEURAL NETWORK ALGORITHM

arrangement	Variable name
Two layers and three layers	Network type
Variable	Number of hidden layers
Variable	the neurons
Linear and sigmoid	Stimulator function
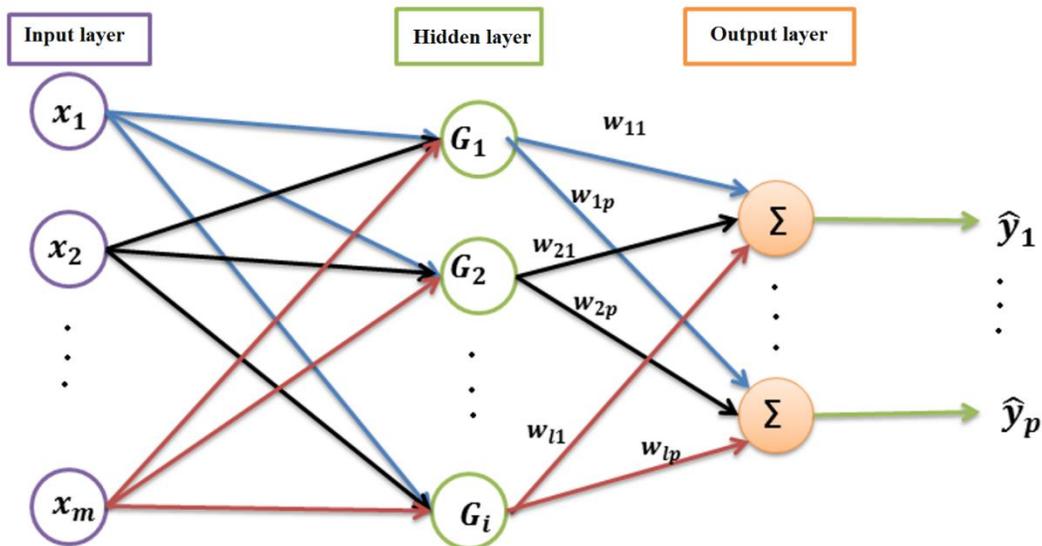
Fig. 5. Proposed Neural Network Architecture.

IV. EVALUATION AND EFFICIENCY

In this section, the numerical results of the proposed methods are presented in comparison with similar algorithms. This comparison has been made for samples in different conditions and challenges. All the experiments of this research have been done with MATLAB version 2017b software in Windows 10 operating system and on a Dell laptop with a 2-core 2 GHz central processor and 2 GB internal memory.

A. Numerical Results Obtained

This research will analyze and select images from the UCSD Anomaly Detection Dataset database. Also, the presented method will be implemented using MATLAB software, and the results presented in this thesis will be compared with those presented in previous papers. In some cases, where there are both color and abnormal size lesions, thresholding can play an important role because it has a uniform illumination intensity. In contrast, the anomaly does not have uniform illumination dimensions. As a result, thresholding determines the approximate location of the abnormality better. It should be noted that small objects and objects that exist after thresholding can be removed by calculating their area. Fig. 6 shows the implementation of the proposed method on different images from the video.

The comparison of anomaly detection in training data as seen in Fig. 7 is shown for the accuracy of detection at different levels. And as the accuracy of diagnosis has increased, the amount of training data has taken a downward trend. The more training data there are, the better the model created using the suggested algorithm covers the images, resulting in a similar system with the arrival of new test data and the ability to detect anomalies. This is the reason for improving the accuracy of detection by increasing the number of training data.
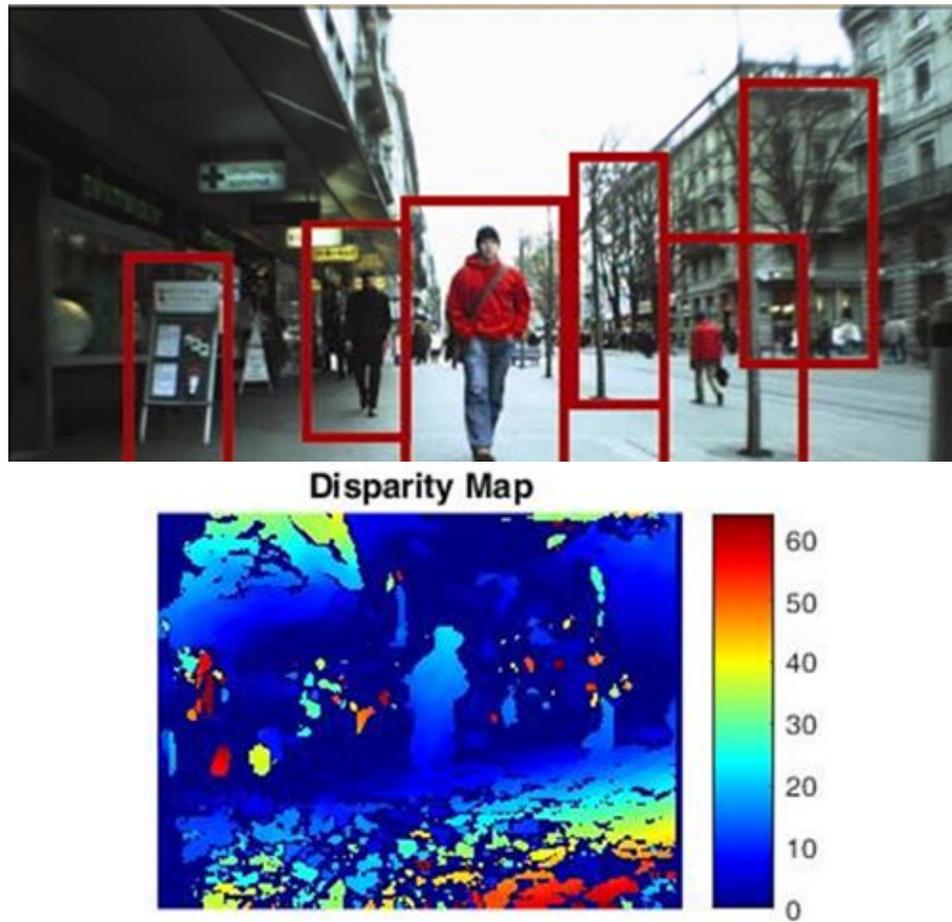
Fig. 6. Anomaly Detection using the Method in different Images from the Video.
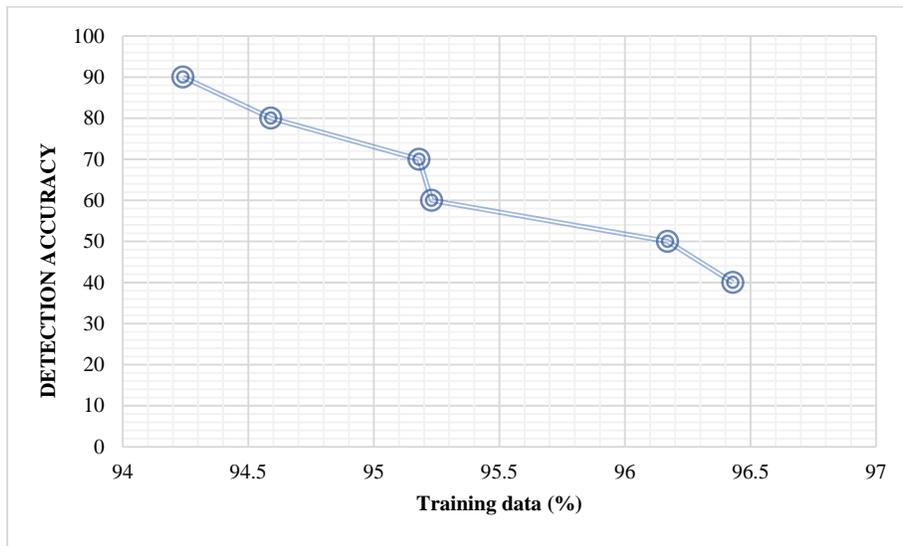
Fig. 7.   Comparison of Anomaly Detection on Training Data.

TABLE III.   ALGORITHM PARAMETERS

parameter	value
Initial population size	50
number of iterations (generations)	400
The number of features of each variable	20
Cutting rate	0/9
Composition rate	0/7
beta	0/08

Table III shows the parameters of the proposed difference evolution algorithm and their values. It should be noted that the values of the operators were selected by trial and error.

To ensure the correctness of the abnormality diagnosis, the final results and images must be reviewed and approved by an expert. However, to evaluate and check the efficiency of the proposed algorithm, the following criteria such as sensitivity, accuracy, and precision can be checked to estimate the correctness percentage of the proposed methods.

B. Comparison of Anomaly Estimates

Table IV shows the results of the proposed method. In our experiments, 40% to 90% of the training data were used, and we evaluated the anomaly detection results of the training data.

TABLE IV.   ANOMALY ESTIMATION RESULTS WITH DIFFERENT NUMBER OF TRAINING DATA

Educational data (%)	Our Method
40	96.43
50	96.17
60	95.23
70	95.18
80	94.59
90	94.24

The decrease in detection accuracy with the increase in training data is because the more training data is, the more the model built with the proposed algorithm will have a higher computational load and perform anomaly detection with less accuracy. However, as it can be seen, the presented method is also effective in a small amount of training data. It is concluded from this figure that with fewer input data, we will have better training and better results will be obtained, and the prediction error will be minimized. It can be seen that even though all the parameters of this network are the same and only their initial weights and biases are different, the number of periods required to reach the desired error limit is very different. The interesting point is that, although after training, all the networks have reached the desired error limit and stopped at that limit, they will give different answers compared to the test samples.

C. ROC Curve

It has been used to evaluate and compare the tests presented in order to detect abnormalities with other methods in the same conditions. One of the ways to detect the strength of a classifier is to use the ROC chart and the area under that chart in such a way that if the area under the curve is low, it means a weak classification, and the more it is, it means a strong classification.

The difference between the ANN method and PROPOSED is in the combination of evolutionary algorithm and ANN, which parameters were optimized with differential evolution algorithm before neural network training. In Fig. 8, three classifications are compared based on the ROC curve. As can be seen, the sub-level of the proposed method is higher compared to the other two methods, and this means that the learning method has a better performance in classifying and diagnosing people's abnormalities.

D. Comparison of the Results of Several Methods for Abnormality Diagnosis

Accuracy and time criteria have been used to evaluate the presented algorithms to classify anomalies with other methods

in the same conditions. Fig. 9 and 10, as well as Table V, compare the evaluation results of several methods for abnormality detection.

As seen in Fig. 9 and 10, the proposed solution was able to perform the anomaly detection operation in much less time and with relatively higher accuracy, which shows that the solution is optimal. This level of efficiency and optimality has been made possible with the help of a combined approach and the use of the differential evolution algorithm and the cellular neural network, which has found the solution to detect abnormalities in much less time and with proper accuracy. In the following, in the forms of 11 to 13 percent, the sensitivity, accuracy, and amount of correct prediction of the solutions are compared. The reason for checking these parameters is to evaluate the solution in the correct diagnosis with relatively high sensitivity. In diagnosing abnormal human behavior using video images, the sensitivity and correctness of the predictions must be high because the lower this level, the more reliable it is. It will also come down to the solution.
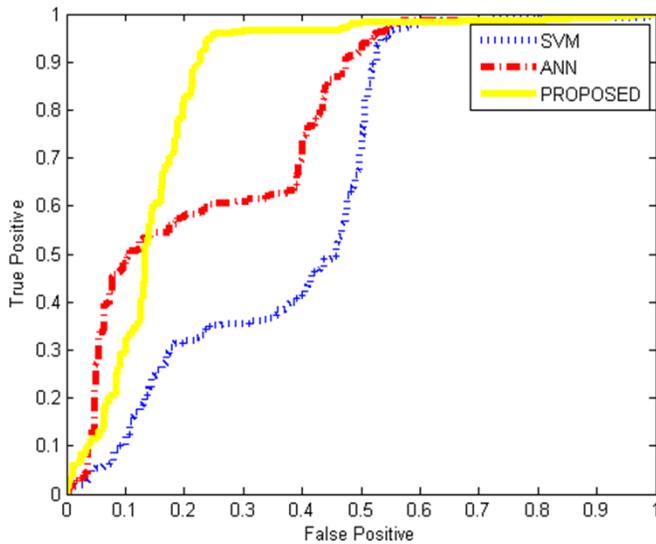
Fig. 8. Comparison of Methods.
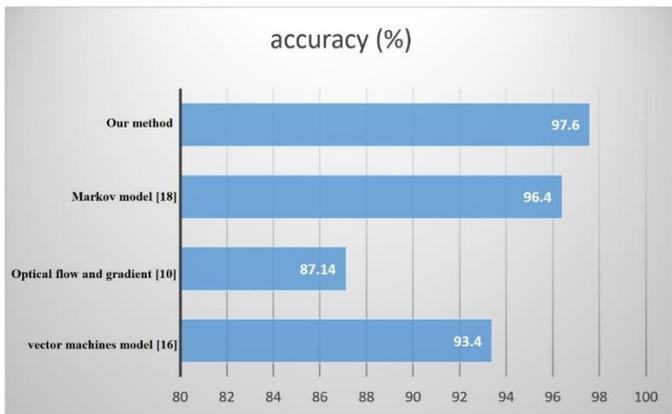
Fig. 9. Accuracy of Anomaly Detection in the Evaluated Solutions.
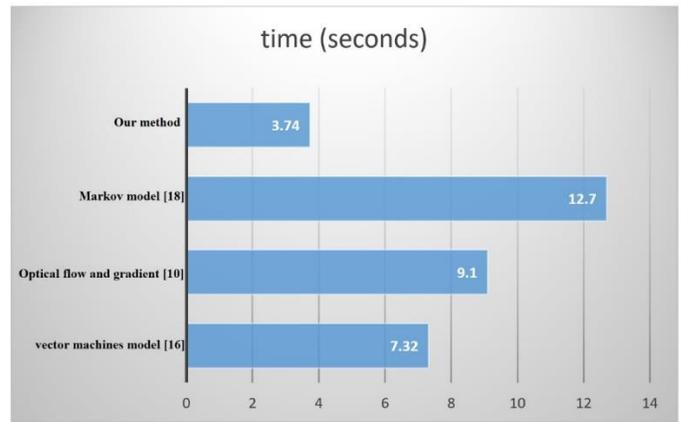
Fig. 10. Abnormality Detection Time in the Evaluated Solutions.

TABLE V. COMPARISON OF THE RESULTS OF SEVERAL METHODS FOR ABNORMALITY DIAGNOSIS

Method	accuracy (%)	time (s)
Vector machine model [16]	93.4	7.32
Optical flow and gradient [10]	87.14	9.1
Markov model [18]	96.4	12.7
Our method	**97.6**	3.74
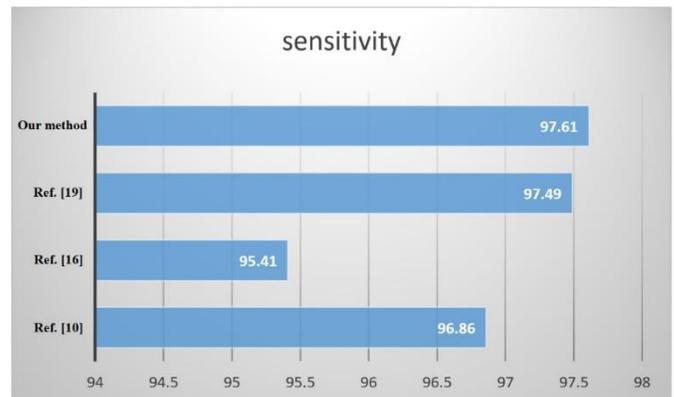
Fig. 11. The Level of Sensitivity in the Evaluated Solutions.

As seen in Fig. 11, the proposed solution has been able to perform better than all three solutions in terms of sensitivity. This level of optimization is possible due to the combined solution presented in the previous chapter.

In Fig. 12, the proposed solution has performed better than the two solutions presented in [19] and [10], but with a slight difference, it has performed weaker than the solution presented in [16], but considering that the method presented in this research In terms of other important parameters such as percentage of correct prediction or sensitivity, it has performed much better, this small weakness can be ignored.

Fig. 13 shows the percentage of correct predictions. As can be seen, the proposed solution was higher than the other evaluated solutions by a large margin. This level of optimality can compensate for the relative weakness of the solution in terms of accuracy. Table VI shows the numerical values of the proposed method in abnormality detection.
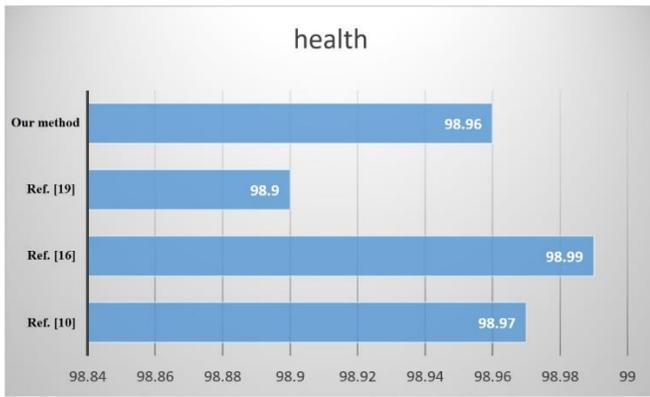
Fig. 12. The Degree of Accuracy in the Evaluated Solutions.
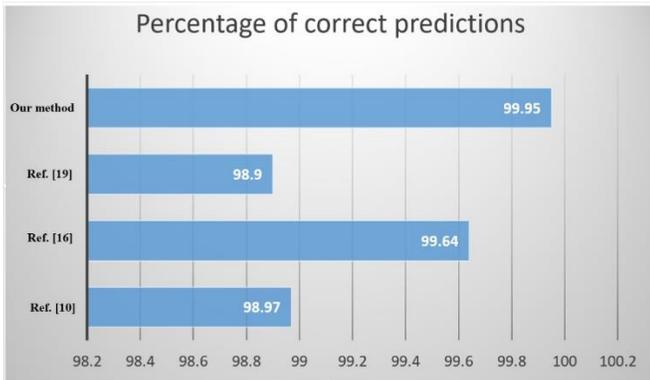
Fig. 13. Percentage of Correct Prediction in the Evaluated Solutions.

TABLE VI. NUMERICAL VALUES OF THE PROPOSED METHOD IN ABNORMALITY DETECTION

Ref.	health	sensitivity	Correct prediction
[10]	98/97	96/86	95/88
[16]	99/64	95/41	96/83
[19]	98/90	97/49	95/71
Our Method	98/96	97/61	96/54

The UCSD Anomaly Detection Dataset database was used to analyze the images.

http://www.svcl.ucsd.edu/projects/anomaly/dataset.html.

V. CONCLUSION

In this research, a computationally efficient algorithm for anomaly detection on video images with high resolution is presented. The proposed method has higher speed and accuracy compared to other methods. The advantage of the algorithm is that it has an execution time of three seconds on a home computer. The proposed method has higher accuracy and precision in comparison with other methods. The advantage of the algorithm is that it has an accuracy of 96.5% on a home computer, and the average sensitivity criterion is 98.6%/97.2%.

VI. SUGGESTIONS

It is suggested that in the future, the following things should be investigated to improve the methods presented in

this paper. The proposed method has a good potential for classifying video objects such as cars in color images. Improving the proposed methods to reduce the time of abnormality detection, which can be used to improve the abnormality detection using the combined methods of the neural network, which requires a supercomputer.

REFERENCES

[1] Vats, E., & Chan, C. S. (2016). Early detection of human actions—A hybrid approach. Applied Soft Computing, 46, 953-966.

[2] Narang, V., & Solanki, A. (2023). An Efficient Algorithm for Human Abnormal Behaviour Detection Using Object Detection and Pose Estimation. In The Fifth International Conference on Safety and Security with IoT (pp. 47-64). Springer, Cham.

[3] Y. Cong, J. Yuan, and Y. Tang, "Video anomaly search in crowded scenes via spatio-temporal motion context," IEEE Trans. Inf. forensics Secur., vol. 8, no. 10, pp. 1590–1599, 2013.

[4] Guha, A., & Samanta, D. (2021). Hybrid approach to document anomaly detection: an application to facilitate RPA in title insurance. International Journal of Automation and Computing, 18(1), 55-72.

[5] V. Kaltsa, A. Briassouli, I. Kompatsiaris, L. J. Hadjileontiadis, M. G. Strintzis, Swarm intelligence for detecting interesting events in crowded environments, IEEE Trans.ImageProcess.24(7)(2015)2153–2166.

[6] Y. Shenn, W. Hu, M. Yang, J. Liu, and C. T. Chou, "Efficient background subtraction for tracking in embedded camera networks," in Information Processing in Sensor Networks (IPSN), 2012 ACM/IEEE 11th International Conference on, 2016, pp. 103–104.

[7] J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, "A survey of video datasets for human action and activity recognition," Comput. Vis. Image Underst., vol. 117, no. 6, pp. 633–659, 2013.

[8] Sun, J., Zhang, Y., & Trik, M. (2022). PBPHS: A Profile-Based Predictive Handover Strategy for 5G Networks. Cybernetics and Systems, 1-22.

[9] Chapnevis, A., Güvenç, I., & Bulut, E. (2020, November). Traffic Shifting based Resource Optimization in Aggregated IoT Communication. In 2020 IEEE 45th Conference on Local Computer Networks (LCN) (pp. 233-243). IEEE.

[10] A. Oliver, J. Freixenet, J. Marti, E. Pérez, J. Pont, E. R. E. Denton, and R. Zwiggelaar, "A review of automatic mass detection and segmentation in mammographic images," Med. Image Anal., vol. 14, no. 2, pp. 87–110, 2017.

[11] Ali, M. M., Noorain, S., & Qaseem, M. S. (2023). Suspicious Human Behaviour Detection Focusing on Campus Sites. In Emerging IT/ICT and AI Technologies Affecting Society (pp. 171-183). Springer, Singapore.

[12] Rafiee, P., & Mirjalily, G. (2020). Distributed Network Coding-Aware Routing Protocol Incorporating Fuzzy-Logic-BasedForwarders in Wireless Ad hoc Networks. Journal of Network and Systems Management, 28(4), 1279-1315.

[13] Y. Xu, X. Cao, and H. Qiao, "An efficient tree classifier ensemble-based approach for pedestrian detection," Syst. Man, Cybern. Part B Cybern. IEEE Trans., vol. 41, no. 1, pp. 107–117, 2011.

[14] Sundaramoorthy, S. U. R. I. Y. A. (2021). Deep Learning Based Hybrid Approach For Facial Emotion Detection.

[15] Trik, M., Mozaffari, S. P., & Bidgoli, A. M. (2021). Providing an adaptive routing along with a hybrid selection strategy to increase efficiency in NoC-based neuromorphic systems. Computational Intelligence and Neuroscience, 2021.

[16] H. Mousavi, S. Mohammadi, A. Perina, R. Chellali, V. Murino, Analyzing tracklets for the detection of abnormal crowd behavior,in:IEEE Winter Conference Applications of Computer Vision(WACV) ,IEEE, Hawaii, 2015,pp. 148–155.

[17] S. Zhang, R. Benenson, and B. Schiele, "Filtered channel features for pedestrian detection," in Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, 2015, pp. 1751–1760.

[18] Lotfi, F., Semiari, O., & Saad, W. (2021). Semantic-Aware Collaborative Deep Reinforcement Learning Over Wireless Cellular Networks. arXiv preprint arXiv:2111.12064.

[19] Hashemzehi, R., Mahdavi, S. J. S., Kheirabadi, M., & Kamel, S. R. (2020). Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE. biocybernetics and biomedical engineering, 40(3), 1225-1232.

[20] A. Wiliem, V. Madasu, W. Boles, and P. Yarlagadda, "A suspicious behaviour detection using a context space model for smart surveillance systems," Comput. Vis. Image Underst., vol. 116, no. 2, pp. 194–209, 2012.

[21] Chenarlogh, V. A., Razzazi, F., & Mohammadyahya, N. (2019, December). A multi-view human action recognition system in limited data case using multi-stream CNN. In 2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS) (pp. 1-11). IEEE.

[22] Trik, M., Molk, A. M. N. G., Ghasemi, F., & Pouryeganeh, P. (2022). A Hybrid Selection Strategy Based on Traffic Analysis for Improving Performance in Networks on Chip. Journal of Sensors, 2022.

[23] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 1, pp. 18–32, 2014.

[24] Golrou, A., Sheikhani, A., Nasrabadi, A. M., & Saebipour, M. R. (2018). Enhancement of sleep quality and stability using acoustic stimulation during slow wave sleep. International Clinical Neuroscience Journal, 5(4), 126.

[25] M. J. Roshtkhari and M. D. Levine, "An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions," Comput. Vis. Image Underst., vol. 117, no. 10, pp. 1436–1452, 2013.

[26] O. P. Popoola and K. Wang, "Video-based abnormal human behavior recognition—a review," IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev., vol. 42, no. 6, pp. 865–878, 2014.

[27] Trik, M., Pour Mozafari, S., & Bidgoli, A. M. (2021). An adaptive routing strategy to reduce energy consumption in network on chip. Journal of Advances in Computer Research, 12(3), 13-26.

[28] H. Nallaivarothayan, D. Ryan, S. Denman, S. Sridharan, and C. Fookes, "An evaluation of different features and learning models for anomalous event detection," in Digital Image Computing: Techniques and Applications (DICTA), 2013 International Conference on, 2013, pp. 1–8.

[29] X. Zhu, J. Liu, J. Wang, C. Li, and H. Lu, "Sparse representation for robust abnormality detection in crowded scenes," Pattern Recognit., vol. 47, no. 5, pp. 1791–1799, 2014.

[30] Kunkel, G., Madani, M., White, S. J., Verardi, P. H., & Tarakanova, A. (2021). Modeling coronavirus spike protein dynamics: implications for immunogenicity and immune escape. Biophysical Journal, 120(24), 5592-5618.

[31] Shifu Zhou, Wei. Shen, Dan. Zeng, Mei. Fang, Yuanwang. Wei,Zhijiang. Zhang, Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes, Signal Processing: Image Communication 47 (2016), pp. 358–368.

[32] X. Liu, G. Zhao, J. Yao, and C. Qi, "Background subtraction based on low-rank and structured sparse decomposition," Image Process. IEEE Trans., vol. 24, no. 8, pp. 2502–2514, 2015.

[33] Hosseini, S. (2022). The role of pertomix approaches in early detection of cancer. Proteomics, 9(01), 177-186.

[34] A. A. Sodemann, M. P. Ross, and B. J. Borghetti, "A review of anomaly detection in automated surveillance," IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev., vol. 42, no. 6, pp. 1257–1272, 2016.

[35] Rezaee, K., Rezakhani, S. M., Khosravi, M. R., & Moghimi, M. K. (2021). A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance. Personal and Ubiquitous Computing, 1-17.

[36] Charandabi, S., & Ghanadiof, O. (2022). Evaluation of Online Markets Considering Trust and Resilience: A Framework for Predicting Customer Behavior in E-Commerce. Journal of Business and Management Studies, 4(1), 23-33.

[37] M. Errami and M. Rziza, "Improving Pedestrian Detection Using Support Vector Regression," in 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV), 2016, pp. 156–160.

[38] Usman, I., & Albesher, A. A. (2021). Abnormal Crowd Behavior Detection Using Heuristic Search and Motion Awareness. International Journal of Computer Science & Network Security, 21(4), 131-139.

[39] Sabha, A., & Selwal, A. (2021, September). HAVS: Human action-based video summarization, Taxonomy, Challenges, and Future Perspectives. In 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES) (pp. 1-9). IEEE.

[40] Mu, H., Sun, R., Yuan, G., & Wang, Y. (2021). Abnormal Human Behavior Detection in Videos: A Review. Information Technology and Control, 50(3), 522-545.

[41] Radhoush, S., Shabaninia, F., & Lin, J. (2018, February). Distribution system state estimation with measurement data using different compression methods. In 2018 IEEE Texas Power and Energy Conference (TPEC) (pp. 1-6). IEEE.

[42] Gill, R., Singh, J., & Modgill, A. (2022, March). A Retrospective CNN-LSVM Hybrid Approach for Multimodal Emotion Recognition. In 2022 International Conference on Decision Aid Sciences and Applications (DASA) (pp. 1281-1285). IEEE.

[43] HassanVandi, B., Kurdi, R., & Trik, M. (2021). Applying a modified triple modular redundancy mechanism to enhance the reliability in software-defined network. International Journal of Electrical and Computer Sciences (IJECS), 3(1), 10-16.

[44] Ramprasadi, N., Shah, P., & Vyas, D. (2020, May). Hybrid Approach For Real Time Crowd Activity Identification Using Segmentation. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 390-394). IEEE.

[45] Trik, Mohammad, Amir Massoud Bidgoli, Hossein Vashani, and Saadat Pour Mozaffari. "A new adaptive selection strategy for reducing latency in networks on chip." Integration (2022).

[46] Vahidi Farashah, M., Etebarian, A., Azmi, R., & Ebrahimzadeh Dastjerdi, R. (2021). A hybrid recommender system based-on link prediction for movie baskets analysis. Journal of Big Data, 8(1), 1-24.

Transformer-based Neural Network for Electrocardiogram Classification

Mohammed A. Atiea, Mark Adel

Computer Science Department, Faculty of Computers and Information
Suez University, Suez, Egypt

Abstract—**A transformer neural network is a powerful method that is used for sequence modeling and classification. In this paper, the transformer neural network was combined with a convolutional neural network (CNN) that is used for feature embedding to provide the transformer inputs. The proposed model accepts the raw electrocardiogram (ECG) signals side by side with extracted morphological ECG features to boost the classification performance. The raw ECG signal and the morphological features of the ECG signal experience two independent paths with the same model architecture where the output of each transformer decoder is concatenated to go through the final linear classifier to give the predicted class. The experiments and results on the PTB-XL dataset with 7-fold cross-validation have shown that the proposed model achieves high accuracy and F-score, with an average of 99.86% and 99.85% respectively, which shows and proves the robustness of the model and its feasibility to be applied in industrial applications.**

Keywords—Electrocardiogram classification; transformer neural network; convolutional neural network

I. Introduction

Cardiovascular diseases (CVDs) happen due to malfunctions in the heart as well as blood vessels. CVDs are one of the main causes of global deaths as issued by the World Health Organization (WHO) [1], CVDs share 32% of the global death causes in 2019. ECG is an essential tool for CVDs diagnoses and treatment, also it is required for continuous heart monitoring. As WHO reported, 85% of deaths by CVDs located in developing countries where there is a shortage of professional doctors who are required to interpret the ECG for going through the proper medication and treatment [2]. CVDs unleash the potential for the need to automate the ECG interpretation process to overcome the aforementioned challenges.

An extensive amount of research was done to introduce an efficient and sophisticated method for ECG classification. Cognitive algorithms are more suitable to process the ECG signals since the ECG parameters are not standard for all people, also it requires special types of models that are capable to handle sequential data efficiently due to the nature of the ECG signals [3]. Using classical signal processing techniques, machine-learning-based classification methods are introduced to operate on manually extracted ECG features. Especially, deep neural networks present an outstanding performance due to the introduction of various new, well-structured datasets.

This paper introduces a novel method for ECG beat classification based on transformer networks and convolution operation for feature embedding to prepare model inputs. Two instances of the proposed model were trained independently by the raw ECG signals as well as the morphological feature R-R Interval (RRI), as illustrated in Fig. 1, where the output of both models is concatenated with the other to give the final prediction. The experiments in this paper were held based on the PTB-XL dataset [4] where the resampling methods were done to overcome the imbalanced data distribution of the used dataset.

The remaining sections of this paper are organized as follows: Section II involves the state-of-art previous related work, Section III introduces a detailed description of the proposed model, Section IV gives the experiments setup and the results, and Section V gives the conclusion about the paper main points and the future vision.
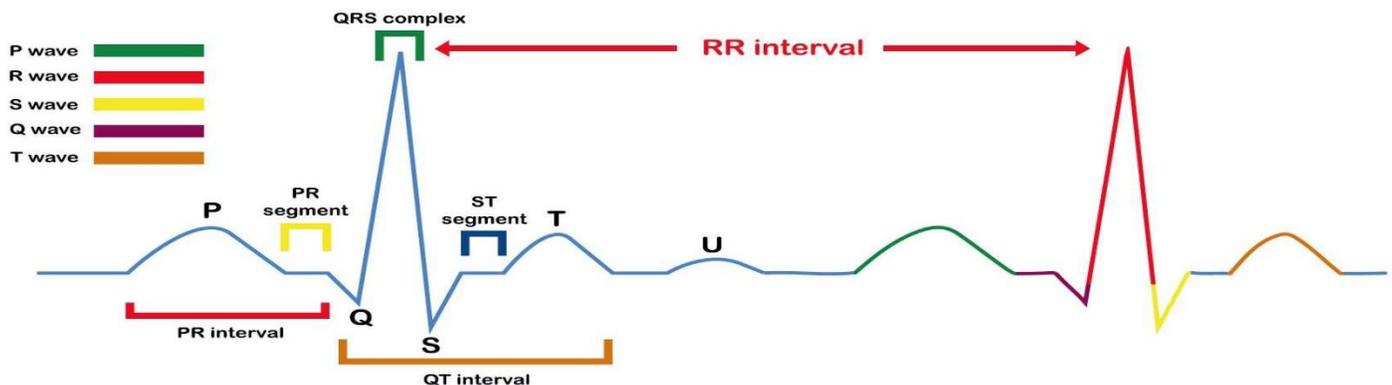
Fig. 1. An ECG Signal Sinus Rhythm.

II. RELATED WORK

ECG beats classification is considered frequently in recent scientific literature. There are two main approaches to solving an ECG signal classification problem, one of them is to use classical machine learning approaches such as random forests [5], support vector machine (SVM) [6], ensemble SVM [7], etc.

Usually, classical machine learning algorithms are preceded by many phases to achieve an acceptable performance. These phases include the ECG noise elimination process which is mainly encountered by digital signal processing techniques such as low- and high-pass filters.

Also, one of the most essential phases, that has a considerable effect on classifier performance, is feature extraction which can be based on signal domain transformations like different variants of Fourier transform [8] and Wavelet transform such as tunable Q-wavelet transform [9], the maximal overlap wavelet packet transform (MOWPT) [10], and continuous wavelet transform (CWT) [11]; features extraction can also be done based on some statistical measurements as skewness and kurtosis [9]. Another way to do feature extraction is to depend on the morphological characteristics of the ECG signal; such methods acquire features mainly from the QRS complex component of the ECG.

On the other side of classical machine learning algorithms, the deep-natural-networks-based models provide exceptional performance over other machine learning algorithms, especially when a large amount of data is fed to the models.

Convolutional neural networks (CNN) were used in two ways of manner: 1-D CNN that accepts 1-D ECG signal as its input also 2-D CNN can be adapted after applying higher signal domain transformation to generate a representative visual representation of ECG signal as a spectrogram and then feed these spectrograms to the 2-D CNN which requires a considerable amount of computational effort.

Recurrent neural networks (RNN) also can be accommodated because of the sequential nature of the ECG data [12]. Also, RNN variants such as gated recurrent units (GRU) [13] and long-short-term memory [14] are introduced to solve the ECG classification problem [15]; since these architectures solve gradient exploding and vanishing problems in the backpropagation algorithm in the network training process [14]. The main disadvantages of the RNN and its variants are that RNN cannot handle long dependencies in the sequential data also because of its sequential behavior RNN does not benefit from parallel hardware accelerators [16] such as graphical processing unit (GPU) and field-programmable gate array (FPGA). A hybrid architecture can be combined of CNN and RNN or its variants can also present a sophisticated accuracy in ECG classification [17].

III. METHOD

A. ECG Morphological Features Extraction

Morphological features of the ECG signals introduced an outstanding performance with different classifiers [18] since these features present critical information that is required to recognize different types of ECG signals. R-R interval (RRI) was chosen to be fed to the proposed model besides the raw ECG signals to optimize the model performance. The RRI was extracted by Pan–Tompkins algorithm [19] as it represents sophisticated performance measurements. The RRI and raw ECG examine two different independent paths of the same model architecture where the output of each model in concatenated before entering the classification head. The proposed model takes advantage of both the CNN and the transformer neural network [20]. The proposed classification model consists of 2 paths each with 4 main stages as illustrated in Fig. 2. Each main stage is explained in detail in the following subsections and the internal structure of each main stage is illustrated in Fig. 3.
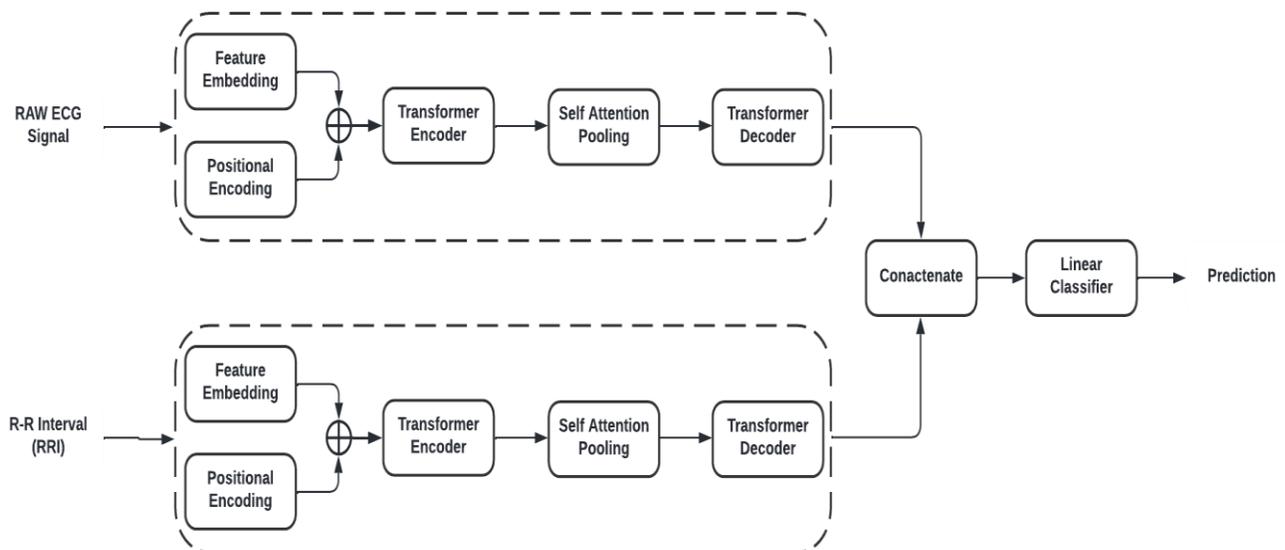
Fig. 2. Abstract Model Architecture.
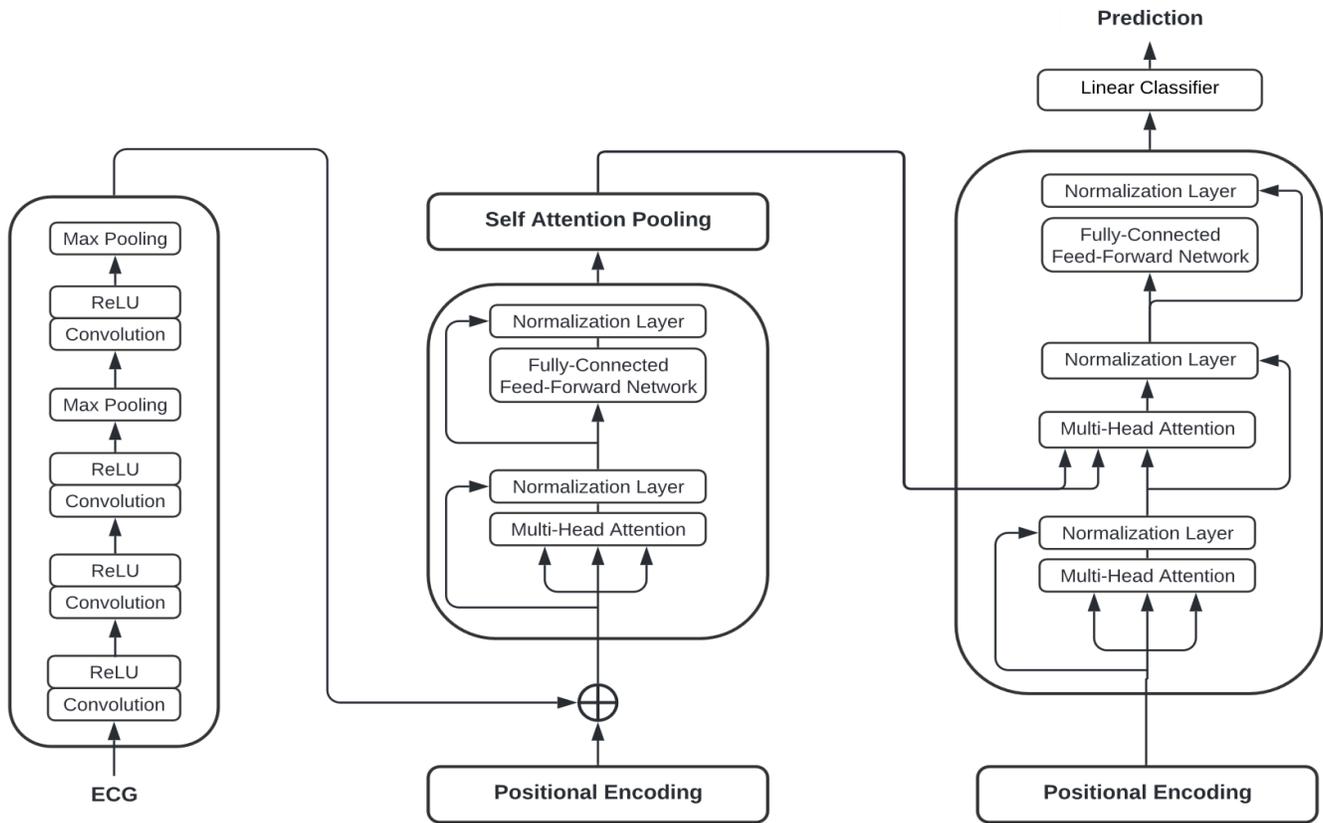
Fig. 3. Detailed Structure of the Embedding Module and Transformer Neural Network.

B. Feature Embedding

The feature embedding module consists of four convolutional layers as illustrated in Table I to generate a compact and concrete ECG features map, each of the two convolutional layers is followed by the Rectified Linear Unit (ReLU) activation function to provide the non-linearity in the ECG signals, while each of the last two layers is followed by maximum pooling layers which output a feature map with the highlighted features of the current ECG signal. Hence the fed data is a sequence of one-dimensional ECG signals, the output of the discrete convolution operation is given as

$$S(t) = (x * w)(t) = \sum_n x(n)w(t-n) \qquad (1)$$

Where x is the input signal, and w is the sliding kernel window. Also, the applied ReLU function can be computed with the following

$$y(x) = max(0, x) \qquad (2)$$

The size of the output of the feature embedding layer will be the same in the original transformer literature [20], which equals 512 (d_{model}). The size of the feature embedding output has to be the same as the output of the positional encoding, which is illustrated in the next subsection to provide the summation of both to the transformer encoder.

C. Positional Encoding

Hence the proposed model has no recurrence relation, it is a necessity to provide information about the absolute and relative positions of different timestamps in the ECG signals.

In order to introduce this information, the output of a positional encoding layer is summed to the output of the feature embedding module to be provided as an input to the transformer encoder and decoder [20].

Any periodic function is sufficient to implement the positional encoder, but in this work, the positional encoder is implemented by different frequencies of the sine and cosine functions because of their linear properties, which are feasible to be learned by the model [20]. Positional encoding can be modeled by the following

$$PE_{(pos,2i)} = \sin(pos / 10000^{2i/d_{model}})$$

$$PE_{(pos,2i+1)} = Cos(pos / 10000^{2i/d_{model}}) \qquad (3)$$

where *pos* represents the position of the sequence token, and *i* represents the spatial location of the ECG feature.

TABLE I. FEATURE EMBEDDING MODULE LAYERS

Layers	Layer Name	Kernel Size
1	Convolution	3 X 1
2	Convolution	3 X 1
3	Convolution	3 X 1
4	MaxPooling	3 X 1
5	Convolution	3 X 1
6	MaxPooling	3 X 1

D. Transformer Encoder

The proposed model consists of four identical layers where each layer is subdivided into two sublayers which are a multi-head self-attention sublayer and a position-wise fully connected feed-forward neural network. Residual connections exist around each sublayer to handle gradient flow in the network during the training time. To accommodate the residual connections each sublayer and the embedding dimension have a dimension of 512. The output of the multi-head self-attention sublayer and its input via the residual connection is normalized to provide a sustainable training process to the network. Then, the data will flow through a fully-connected feed-forward network to provide a more convenient representation of the attention output.

A self-attention pooling [21] block was added to accept the output of the encoder block to reformulate the attention score tensor, from the dimension of [length of sequence, batch size, embedding dimension] to [batch size, embedding dimension], to be accepted by the multi-head attention sublayer in the decoder module where the attention function [20] can be written as

$$Attention(Q, K, V) = softmax \left(\frac{Q\,k^t}{\sqrt{d_k}} \right) V \qquad (4)$$

where Q is the query, K is the key, V is the value, and $d_k = d_{model}/h$ [20].

E. Transformer Decoder

The transformer decoder contains the same structure as the transformer encoder with the same sublayers proceeded with normalization as well as residual connection.

In addition to a multi-head self-attention sublayer and a position-wise fully connected feed-forward neural network, the decoder provides a multi-head attention sublayer to attend over the output of the transformer encoder module. Also, the multi-head attention sublayer is modified to attend only to the previous sequence tokens from the encoder output which is performed by masking the attention tensor [20].

F. Linear Classifier

At this stage, the outputs of each path of the raw ECG and RRI are concatenated into one tensor which is passed through this linear classifier to give the final prediction. The linear classifier consists of a flatten layer and two fully-connected feed-forward networks that are separated by a dropout layer for regularization [22].

IV. EXPERIMENTS RESULTS

All the experiments were carried out by Google Colaboratory where the dataset handling and the model implementation were held in Python 3.9.

A. Dataset Description

In the training process, we used the PTB-XL dataset [4]. PTB-XL is the to-date largest freely accessible clinical 12-lead ECG-waveform dataset. The dataset covers a broad range of diagnostic classes including a large fraction of healthy records. A total of 2183721837 clinical 12-lead ECG records of 10 seconds length from 18885 patients are included in the dataset. The data is gender-balanced (52 percent male, 48 percent female) and includes the ages of 0 to 95 years old (median 62 and interquartile range of 22).

The ECG statements used for annotation are conformed to the SCP-ECG standard [23] and were assigned to three non-mutually exclusive categories diag (short for diagnostic statements such as "anterior myocardial infarction"), form (related to considerable changes of particular parts within the ECG such as "abnormal QRS complex") and rhythm (related to specific changes of the rhythm such as "atrial fibrillation"). There are 71 different statements in all, which are broken down into 44 diagnoses, 12 rhythms, and 19 form statements, four of which are also utilized as diagnostic ECG statements. A hierarchical classification into five coarse superclasses and 24 subclasses is also provided for diagnostic statements. As shown in Table II, We mainly classified by the 5 main classes (NORM, HYP, MI, STTC, CD) and by 14 classes which are (LVH, IVCD, ISC_, LAO/LAE, IMI, CRBBB, NST_, CLBBB, RAO/RAE, ILBBB, LMI, AMI, NORM, WPW).

Apart from its large nominal size, PTB-XL is notable for its diversity, both in terms of signal quality (with 77.01 percent of the highest signal quality) and in terms of a broad range of pathologies, many different co-occurring diseases, and a high proportion of healthy control samples, which is uncommon in clinical datasets. This variability is what makes PTB-XL such a valuable resource for training and evaluating algorithms in a real-world scenario, where machine learning (ML) algorithms must function reliably independent of recording settings or potentially low-quality data.

As you can see, the dataset still doesn't have a normal distribution, hence data augmentation was necessary to improve accuracy. By using data augmentation, we were able to improve the diversity of training data without having to acquire additional data.

B. Classification Model Metrics

Two models were created, one of them operates on ECG superclasses while the other operates on the ECG subclasses as shown in Table II.

Several statistical metrics were established to evaluate the classification performance of the proposed model where these metrics are precision, sensitivity, F1-score, and accuracy. All the mentioned metrics were computed due to the equations in Table III, where TP is a true positive, TN is a true negative, FP is a false positive, and FN is a false negative. The aVR lead was fed to the two models due to the amount of information it contains about the cardiac state of the individual as well as it is proven to present a good performance for classification and detection tasks [24].

Table IV shows all the classification performance metrics for the superclasses model while Table V presents the classification performance metrics for the subclasses model.

TABLE II. PTB-XL DATASET SUPERCLASSES AND SUBCLASSES

Superclass	Subclass	Records	Description
HYP	LVH	2137	left ventricular hypertrophy
	LAO/LAE	427	left atrial overload/enlargement
	RAO/RAE	99	right atrial overload/enlargement
CD	IVCD	789	non-specific intraventricular conduction disturbance (block)
	CRBBB	542	complete right bundle branch block
	CLBBB	536	complete left bundle branch block
	ILBBB	77	incomplete left bundle branch block
	WPW	80	Wolff-Parkinson-White syndrome
STTC	NST_	770	non-specific ST changes
	ISC_	1275	non-specific ischemic
MI	LMI	201	lateral myocardial infarction
	AMI	354	anterior myocardial infarction
	IMI	2685	inferior myocardial infarction
NORM	NORM	9528	normal ECG

TABLE III. STATISTICAL METRICS EQUATIONS

Metric	Equation
Accuracy	$\dfrac{TP + TN}{TP + FP + FN + TN}$
Precision	$\dfrac{TP}{TP + FP}$
Recall	$\dfrac{TP}{TP + FN}$
F1 Score	$2\,x\,\dfrac{Rec\ x\ Pre}{Rec + Pre}$

TABLE IV. CLASSIFICATION PERFORMANCE METRICS FOR SUPERCLASSES MODEL

Class	Precision	Recall	F1 Score	Average Acc.
HYP	99.73%	99.83%	99.78%	
CD	99.69%	99.72%	99.70%	
STTC	100%	99.86%	99.93%	
MI	99.86%	99.86%	99.86%	99.86%
NORM	100%	100%	100%	
Average	99.86%	99.85%	99.85%	

The proposed model was trained through 32 epochs for the superclasses model and 35 epochs for the subclasses model, Table VI introduces the average time required per epoch for each model. Also, the cross-entropy loss function is used to establish the loss during training and validation processes concerning the number of epochs which is illustrated in Fig. 4. Also, the accuracy of the training and validation processes is shown in Fig. 5 concerning the number of epochs. As illustrated in Table VII and Table VIII, the experimental results demonstrate that the proposed model introduced a considerable improvement over the current state-of-arts models for ECG classification.

TABLE V. CLASSIFICATION PERFORMANCE METRICS FOR SUBCLASSES MODEL

Class	Precision	Recall	F1 Score	Average Acc.
LVH	99.86%	100%	99.93%	
LAO/LAE	99.96%	100%	99.98%	
RAO/RAE	99.73%	100%	99.86%	
IVCD	99.83%	100%	99.91%	
CRBBB	99.02%	97.22%	98.11%	
CLBBB	100%	100%	100%	
ILBBB	99.96%	99.96%	99.96%	
WPW	99.93%	100%	99.96%	99.6%
NST_	99.06%	98.01%	98.53%	
ISC_	99.06%	99.29%	99.17%	
LMI	99.04%	100%	99.52%	
AMI	99.69%	100%	99.84%	
IMI	99.28%	100%	99.63%	
NORM	100%	100%	100%	
Average	99.6%	99.6%	99.6%	

TABLE VI. APPROXIMATE TRAINING TIME PER EPOCH

Num. of Classes	Training Time
5 Classes	≈ 1.9 min/epoch
14 Classes	≈ 5.1 min/epoch

(a)

(b)
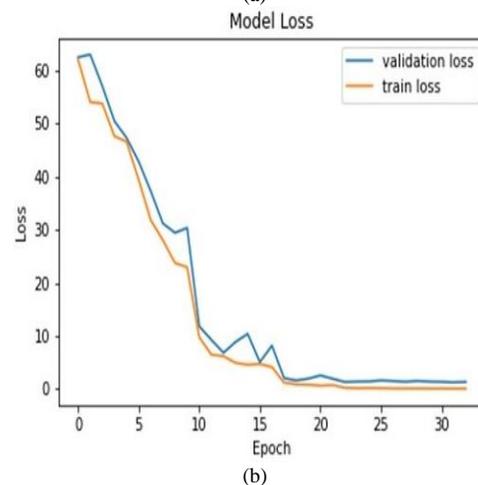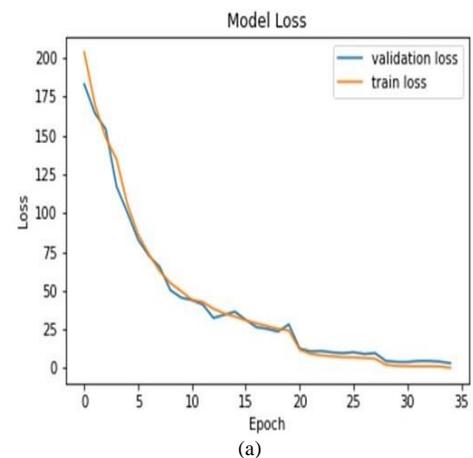
Fig. 4. (a) Subclasses Model Loss, (b) Superclasses Model Loss.

(a) (b)
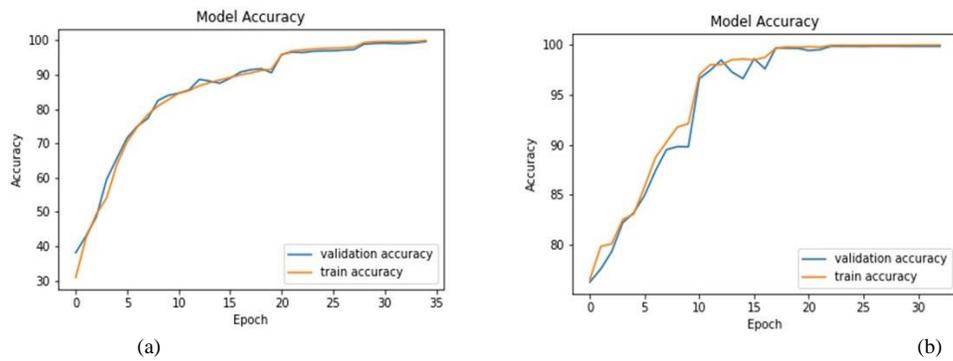
Fig. 5.   (a) Subclasses Model Accuracy, (b) Superclasses Model Accuracy.

TABLE VII.   CURRENT ADVANCED METHODS FOR ECG CLASSIFICATION VS PROPOSED MODEL

Literature	Feature Extraction	Classifier
Z. Wang et al (2019) [5]	RR Interval High Order Statistics Wavelet packet coefficient Principal Component analysis	Random Forests
Z. Ahmed et al (2021) [6]	-	CNN + SVM
S. K. Pandey et al (2020) [7]	Wavelets Transform High Order Statistics Morphological Features	ENSEMBLE SVM
S. Liu et al (2020) [8]	2D the Graph Fourier Transform (GFT)	SVM
C. K. Jha et al (2020) [9]	Tunable Q-wavelet transform	SVM
J. S. Huang et al (2020) [10]	the maximal overlap wavelet packet transform	Residual CNN
T. Wang et al (2021) [11]	Continuous Wavelet	CNN
S. Singh et al (2018) [12]	-	LSTM
H. M. Lynn et al (2019) [13]	-	Bidirectional GRU
X. Xu et al (2020) [17]	-	1D CNN + BILSTM
F. Qiao et al (2020) [25]	-	BLSTM
P. P lawiak et al (2020) [26]	Power spectral density (PSD) Discrete Fourier transform (DFT) Genetic algorithm	DGEC
Shuai Ma et al (2022) [27]	-	1D-CNN + DCGAN
The proposed model	-	1D-CNN + Transformer

TABLE VIII.   COMPARISON WITH THE MOST MODERN METHODS AND THE PROPOSED MODEL FOR PERFORMANCE MEASUREMENTS AND THE NUMBER OF CLASSES USED IN EACH MODEL

Literature	Number of classes	Precision	Recall	F1 Score	Average Acc.
Z. Wang et al (2019) [5]	5	-	37.47%	-	92.31%
Z. Ahmed et al (2021) [6]	5	94%	93%	-	96.5%
S. K. Pandey et al (2020) [7]	4	69.11%	65.26%	66.24%	94.40%
S. Liu et al (2020) [8]	4	-	52.7%	-	96.2%
C. K. Jha et al (2020) [9]	8	-	96.22%	-	99.27%
J. S. Huang et al (2020) [10]	5	99.39%	95.16%	97.23%	98.79%
T. Wang et al (2021) [11]	4	-	93.25%	94.43%	98.74%
S. Singh et al (2018) [12]	2	-	92.4%	-	88.1%
H. M. Lynn et al (2019) [13]	2	-	-	-	98.6%
X. Xu et al (2020) [17]	5	96.34%	95.90%	95.92%	95.90%
F. Qiao et al (2020) [25]	5	97.15%	97.71%	-	99.32%
P. P lawiak et al (2020) [26]	17	-	94.62%	-	99.37%
Shuai Ma et al (2022) [27]	5	98.5%	98.2%	-	98.7%
The proposed model (avg.)	14	99.6%	99.6%	99.6%	99.6%

V. CONCLUSION

This paper introduces a novel classification model for 12-lead ECG signals based on PTB-XL. The proposed model depends on the transformer neural networks for ECG sequence modeling and CNN for feature embedding. The model can be fed by the raw ECG signals as well as the RRI feature of the ECG to achieve robust classification performance. The proposed model achieved accuracy and F-score, with an average of 99.86% and 99.86% respectively where these metrics can compete with the many recent state-of-art models. Due to the model performance and since the transformers were introduced to solve the sequential models' complexity challenges, the proposed model is capable to be deployed in practical applications. In future work, this model can be integrated with wearable device technology to assist in critical cases with continuous monitoring to save more lives.

ACKNOWLEDGMENT

REFERENCES

[1] Al-Ashwal FY, Sulaiman SAS, Sheikh Ghadzi SM, Kubas MA, Halboup A. "Risk assessment of atherosclerotic cardiovascular diseases before statin therapy initiation: Knowledge, attitude, and practice of physicians in Yemen". PLoS One. (2022) May 26;17(5):e0269002. doi :10.1371/journal.pone.0269002.

[2] Mohd Nor, N.S., Chua, YA., Abdul Razak, S. et al." Identification of cardiovascular risk factors among urban and rural Malaysian youths". BMC Cardiovasc Disord 22, 70 (2022). doi :10.1186/s12872-021-02447-y.

[3] Butkevičiūtė E, Michalkovič A, Bikulčienė L. "ECG Signal Features Classification for the Mental Fatigue Recognition". Mathematics. (2022); 10(18):3395. doi :10.3390/math10183395.

[4] Wagner, P. et al. Ptb-xl, a large publicly available electrocardiography dataset. Scientific Data 2020 7:17, 1–15 (2020). doi :10.1038/s41597-020-0495-6.

[5] Wang, Z., Li, H., Han, C., Wang, S. & Shi, L. "Arrhythmia classification based on multiple features fusion and random forest using ECG". Journal of Medical Imaging and Health Informatics 9, 1645–1654 (2019). doi :10.1166/JMIHI.2019.2798.

[6] Ahmad, Z., Tabassum, A., Guan, L. & Khan, N. M. "ECG heartbeat classification using multimodal fusion" (2021). doi :10.48550/arxiv.2107.09869.

[7] Pandey, S. K., Janghel, R. R. & Vani, V. "Patient specific machine learning models for ECG signal classification". Procedia Computer Science 167, 2181–2190 (2020). doi :10.1016/J.PROCS.2020.03.269.

[8] Liu, S., Shao, J., Kong, T. & Malekian, R. "ECG arrhythmia classification using high order spectrum and 2d graph fourier transform". Applied Sciences 2020, Vol. 10, Page 4741 10, 4741 (2020). doi :10.3390/APP10144741.

[9] Jha, C. K. & Kolekar, M. H. "Cardiac arrhythmia classification using tunable q-wavelet transform based features and support vector machine classifier". Biomedical Signal Processing and Control 59 (2020). doi :10.1016/J.BSPC.2020.101875.

[10] Huang, J. S., Chen, B. Q., Zeng, N. Y., Cao, X. C. & Li, Y. "Accurate classification of ECG arrhythmia using mowpt enhanced fast compression deep learning networks". Journal of Ambient Intelligence and Humanized Computing (2020). doi :10.1007/S12652-020-02110-Y.

[11] Wang, T. et al. "Automatic ECG classification using continuous wavelet transform and convolutional neural network". Entropy 23,1–13 (2021). doi :10.3390/E23010119.

[12] Singh, S., Pandey, S. K., Pawar, U. & Janghel, R. R. "Classification of ECG arrhythmia using recurrent neural networks". Procedia Computer Science 132, 1290–1297 (2018). doi :10.1016/J.PROCS.2018.05.045.

[13] Lynn, H. M., Pan, S. B. & Kim, P. "A deep bidirectional gru network model for biometric electrocardiogram classification based on recurrent neural networks". IEEE Access 7, 145395–145405 (2019). doi :10.1109/ACCESS.2019.2939947.

[14] Hochreiter, S. & Schmidhuber, J. "Long short-term memory". Neural Computation 9, 1735–1780 (1997). doi :10.1162/NECO.1997.9.8.1735.

[15] Ozal Yildirim. "A novel wavelet sequences based on deep bidirectional lstm network model for ECG signal classification". Computers in Biology and Medicine 96, 189– 202 (2018). doi :10.1016/J.COMPBIOMED.2018.03.016.

[16] Lipton, Z. C., Berkowitz, J. & Elkan, C. "A critical review of recurrent neural networks for sequence learning" (2015). doi :10.48550/arxiv.1506.00019.

[17] Xu, X., Jeong, S. & Li, J. "Interpretation of electrocardiogram (ECG) rhythm by combined cnn and bilstm". IEEE Access 8, 125380–125388 (2020). doi :10.1109/ACCESS.2020.3006707.

[18] Berkaya, S. K. et al. "A survey on ECG analysis". Biomedical Signal Processing and Control 43, 216–235 (2018). doi :10.1016/J.BSPC.2018.03.003.

[19] Pan, J. & Tompkins, W. J. "A real-time qrs detection algorithm". IEEE Transactions on Biomedical Engineering BME-32, 230– 236 (1985). doi :10.1109/TBME.1985.325532.

[20] Vaswani, A. et al. "Attention is all you need". Advances in Neural Information Processing Systems 2017-December, 5999–6009 (2017). doi :10.48550/arxiv.1706.03762.

[21] Safari, P., India, M. & Hernando, J. "Self-attention encoding and pooling for speaker recognition". Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH 2020-October, 941–945 (2020). doi :10.48550/arxiv.2008.01077.

[22] Inoue, H. "Multi-sample dropout for accelerated training and better generalization" eprint arXiv:1905.09788 (2019). doi :10.48550/arxiv.1905.09788.

[23] ISO - iso 11073-91064:2009 - health informatics — standard communication protocol — part 91064: Computer-assisted electrocardiography. [Online] URL https://www.iso.org/standard/46493.html.

[24] Chen, T. M., Huang, C. H., Shih, E. S., Hu, Y. F. & Hwang, M. J. "Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model". iScience 23, 100886 (2020). doi :10.1016/J.ISCI.2020.100886.

[25] Qiao, F. et al. "A fast and accurate recognition of ECG signals based on elm-lrf and blstm algorithm". IEEE Access 8, 71189–71198 (2020). doi :10.1109/ACCESS.2020.2987930.

[26] P lawiak, P. & Acharya, U. R. "Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals". Neural Computing and Applications 32,11137– 11161(2020). doi :10.1007/s00521-018-03980-2.

[27] Shuai Ma, Jianfeng Cui, Chin-Ling Chen, Xuhui Chen, Ying Ma,"An effective data enhancement method for classification of ECG arrhythmia", Measurement, Volume 203, (2022), 111978, ISSN 0263-2241, doi:10.1016/j.measurement.2022.111978.

BCT-CS: Blockchain Technology Applications for Cyber Defense and Cybersecurity: A Survey and Solutions

Naresh Kshetri[1]*, Chandra Sekhar Bhushal[2], Purnendu Shekhar Pandey[3], Vasudha[4]

Assistant Professor (Cybersecurity), Dept. of Math, CS & IT, Lindenwood University, St. Charles, MO, USA[1]
School of Eng, IT & Physical Sciences, Federation University, Victoria, Australia[2]
Department of CSE, KIET Group of Institutions, Ghaziabad, UP, India[3]
Department of CSE, United Institute of Management, Prayagraj, UP, India[4]

Abstract—Blockchain technology has now emerged as a ground-breaking technology with possible solutions to applications from securing smart cities to e-voting systems. Although it started as a digital currency or cryptocurrency, bitcoin, there is no doubt that blockchain is influencing and will influence business and society more in the near future. We present a comprehensive survey of how blockchain technology is applied to provide security over the web and to counter ongoing threats as well as increasing cybercrimes and cyber-attacks. During the review, we also investigate how blockchain can affect cyber data and information over the web. Our contributions included the following: (i) summarizing the Blockchain architecture and models for cybersecurity (ii) classifying and discussing recent and relevant works for cyber countermeasures using blockchain (iii) analyzing the main challenges and obstacles of blockchain technology in response to cyber defense and cybersecurity and (iv) recommendations for improvement and future research on the integration of blockchain with cyber defense.

Keywords—*Applications; blockchain technology; blockchain solutions; countermeasures; cyber-attacks; cyber defense; cybersecurity; survey*

I. INTRODUCTION

Cyber defense is a defensive mechanism or coordinated act of resistance designed for the safety of information, system, and networks against offensive cyber operations by implementing various security procedures. Some of the major cyber defense activities include set up and maintenance of the hardware and software for security infrastructure, examination of the network's system for vulnerabilities, implementation of real-time countermeasures to stop zero-day attacks, and recovery from attacks that were either completely or partially successful [1]. It ensures the survival of any state from cyber-attacks [2]. The main focus of cyber defense is the prevention, detection, and regulation of timely response to cyber-attacks or threats to make sure that infrastructure or information are not harmed. Cyber defense aims to reduce possible attacks and understand the critical location and sensitive information by carrying out various technical analyses to find out the possible paths and areas that could be targeted by the online attackers [3].

To achieve regulatory compliance, protect assets, and compromise the assets of adversaries, information safeguards, security procedures, and IT security tools and techniques are governed, developed, managed, and applied in cybersecurity [3]. It is the actions and policies implemented to protect cyberspace by the military, public and private business sector. Cyber security has been identified as one of the crucial strategic fields of national security [4]. It is related to protecting information/data, infrastructure, and service to reduce the possibility of loss, damage, and compromise of misuse by unauthorized users. Mainly cyber security focuses on sensitive information stored electronically on computers, computer networks, mobile devices, and the internet [5].

Blockchain is a distributed database technology based on cryptography that maintains a constantly expanding list of data entries that are verified by every network node. Data records are stored in blocks, which are linked together to form the chain. Each participating node keeps a replica of the data records and transmits it to every other node in the network, improving trust and transparency, as opposed to information being routed through a central node [6]. Blockchain technology minimizes the probability of unsecured transactions because applications function decentralized and do not require intermediary authority to monitor the transactions between participants [7].

Blockchain underlying characteristics of decentralization, consensus mechanism, immutability, traceability, and privacy provide a strong foundation for cyber defense and security of data and information in cyberspace. Blockchain decentralized property with use of distributed ledger eliminates the intermediaries that are potential risk for security of the network. Also, there is no single point of failure in decentralization so there is a very low chance that an IP-based DDoS attack will disrupt the daily operations. By establishing a decentralized network using client-side encryption where data owners have complete transparency of their data, blockchain increases the security of data sharing and storing [8] [9]. Next blockchain's Consensus mechanisms govern how participants agree on a single common version of the facts when storing and verifying blocks (a shared truth). The Consensus enables nodes to trustworthily verify brand-new

---

*Corresponding Author.

blocks in the network. The Consensus enables nodes to trustworthily verify brand-new blocks in the network [9].

The immutability property of blockchain ledger provides the solid base for data integrity. Each transaction in a block is cryptographically signed by its sender, and every block in the blockchain is signed by its miner. To alter a single transaction in the blockchain, an attacker ought to modify each next block in the same manner, resolving the consensus problem for that block and all succeeding blocks, and convincing more than 50% of users on the network to adopt the updated chain. Modifying blocks in the network is nearly impossible due to the hashing features and the amount of computing and electrical effort required to achieve this goal [8] [9]. Traceability characteristics of blockchain technology makes the network more secure. Every transaction uploaded to the blockchain is digitally signed and timestamped, allowing organizations to go back to a precise time period for each transaction and identify the corresponding participant (by their public address) on the blockchain [8]. Privacy is one of the core properties of blockchain technology. The identity of those involved in a transaction is protected by cryptographic functions that ensure the anonymization of the entities participating in the blockchain, which help to gain a high degree of privacy [9].

The rest of the article is organized as related work in Section II which discusses several recent works of blockchain-based security and models. We then presented Section III of our paper as BCT for cyber defense and cybersecurity. Section IV of our study is about Countermeasures and defense initiatives with the help of blockchain technology. The advantages of blockchain technology with respect to cyber defense and cybersecurity are summarized in Section V of our findings. We have pointed out several solutions and challenges while incorporating those blockchain-based solutions in Section VI (blockchain-based solutions and challenges). In Section VII, we presented the conclusion and future scope of the study as recommendations for improvement and future research on the integration of blockchain with cyber defense.

II. RELATED WORKS

In [10], A. Razaque et al. (2021), introduced a web-based Blockchain-enabled cybersecurity awareness program (WBCA) to reduce the risk of cybercrimes. The proposed WBCA enhances user understanding of cybersecurity hygiene, best practices, current cybersecurity vulnerabilities, and trends while training users to better their security capabilities and comprehend the typical actions of cybercriminals. Blockchain technology is used by WBCA to safeguard the software against threats, and it is evaluated and tested using real-world cybersecurity issues with actual users and cybersecurity professionals. The suggested WBCA was also tested on a CentOS-based virtual private server to determine its efficacy, and the authors anticipated that the proposed program might be expanded to other areas, such national or corporate courses, to raise users' cybersecurity awareness levels. At the end, the authors also contrasted WBCA with other cutting-edge web-based cybersecurity awareness program.

In [11], A. Razaque et al. (2021), proposed a blockchain-enabled transaction scanning (BTS) approach for the discovery of anonymous actions that outlines the guidelines for outlier detection and quick currency transfers, which limits aberrant transactional behavior. The BTS method is designed to limit money laundering since using bank cards or money transfers gives terrorist groups and anyone who engage in money laundering new opportunities. The BTS method's prescribed rules determine the distinct patterns of fraudulent activity in transactions, scan the transaction history, and offer a list of entities that receive money in a suspicious manner. The performance of the proposed BTS technique was also verified using a Spring Boot application built by the authors using Java programming. Based on the outcomes of the experiments, the proposed method automates the investigation of transactions and limits the occurrences of money laundering.

In [12], H. H. Alhelou et al. (2021), proposed model using Hilbert-Huang transform and Blockchain-based ledger technology to identify fake data injection attacks (FDIA) in a micro grid (MG) system. By analyzing the voltage and current signals in smart sensors and controllers and extracting the signal features, the model is utilized to improve security in the smart DC-MGs. These networks are susceptible to numerous cyberattacks because of the concurrent growth of DC-micro grids and the employment of sophisticated control, monitoring, and operation technologies, as well as their structure. The authors also considered the outcomes of simulations on various instances in order to confirm the effectiveness of the suggested model. The findings offer that the recommended model can increase the security of data exchange in a smart DC-MG and give a more accurate and reliable detection mechanism against FDIA.

In [13], M. Sadigov et al. (2021), proposed a system-dynamic model of the business's cybersecurity system developed with blockchain technology, giving the capacity to create a computer model of a complicated cybersecurity system for more efficient design. Since 34% of cases are accounted for by system vulnerability due to user behaviors, the authors have placed a strong emphasis on reducing the threat associated with human factors. The goal of the research is to use contemporary BCT to address the problem of rising levels of cybersecurity in big businesses. A causal relationship diagram analysis, differential equations for some of the model's components, and experimental modeling for various values of some parameters at the initial level of others are the foundations of the system-dynamic model, which aims to pinpoint the system sensitivity.

In [14], S. Lee and S. Kim (2021), conducted a survey of official records, interviews, relevant news, technical reports, and research papers from 2016 to 2021, which, by methodically conducting research and analysis, helps to close the gap in blockchain for cyber security. The authors discovered that government-led program and research are both aggressively supporting blockchain, proving that it will play a significant role in cyber protection. Due to its relationship to national security, the cyber defense industry needs advanced security technology. In contrast to conventional systems, blockchain offers strong security features without a

centralized control entity, and its use in the cyber defense industry is receiving attention. With a conclusion of recommendations for future research in elements of the blockchain technology, evaluation, and survey, this work offered prospects that blockchain presents for cyber defense, research, national initiatives, restrictions, and other areas.

III. BCT FOR CYBER DEFENSE AND CYBERSECURITY

There is a rapid increase in the number of digital populations' worldwide and online businesses every year. Presently, cybercrime has pop up as the enormous security loophole and threat to the global computer information technology industry [15]. The increasing number of online connected devices and online users will ultimately give rise to online crimes, which in turn cost billions of losses to the tech field and companies globally. The digitalization of data and information offers numerous merits but also creates several loopholes and security warnings. The mode of cybercrime is slowly shifting its operating ground from home computers/laptops to mobile phones, tablets, wearable sensors, digital watches, etc. In order to explore the remedies of such threats and provide counter methods to detect or slow down those threats, we have gone through several background studies in the section above.

TABLE I. SUMMARY OF BLOCKCHAIN-BASED SECURITY AND PROPOSED MODELS FOR CYBER DEFENSE AND CYBERSECURITY AS INTRODUCED IN BACKGROUND STUDY AND LITERATURE

Ref.	BC model proposed	Author (Year)	Published Paper (Journal/Conference)
[7]	Model for Blockchain-based Agribusiness	N. Kshetri et. al. (2021)	BCT-AA: A survey of Blockchain Technology-Based Applications in context of Agribusiness (SSRN Electronic Journal)
[10]	WBCA (Web-based Blockchain-enabled Cyber Awareness program)	A. Razaque et. al. (2021)	Avoidance of Cybersecurity Threats with the Deployment of a Web-Based Blockchain Enabled Cybersecurity Awareness System (Applied Sciences)
[11]	BTS (Blockchain-enabled Transaction Scanning Method)	A. Razaque et. al. (2021)	Blockchain-Enabled Transaction Scanning Method for Money Laundering Detection (Electronics)
[12]	Model to detect FDIA in smart DC-MGs	H. H. Alhelou et. al. (2021)	Cyber-attack detection and Cyber-security enhancement in Smart DC-micro grid based on Blockchain Technology and Hilbert Huang Transform (IEEE Access)
[13]	Blockchain-based System-dynamic model	M. Sadigov et. al. (2021)	Blockchain Technology based System-dynamic Simulation modeling of Enterprise's Cybersecurity System (55th ISC on E&SE)
[14]	Blockchain security for Cyber Defense	S. Lee et. al. (2021)	Blockchain as a Cyber Defense: Opportunities, Applications and Challenges (IEEE Access)

As we have already mentioned, cyber awareness and threat identification play a vital role in combating or fighting against the bad guys or cyber criminals, who are trying to steal/damage online data and secure personal information via various social engineering techniques. We have summarized the blockchain architecture and models (from the Literature Review section of the study) for cyber defense and cyber security in the Table I. The summary includes the proposed model or system with author details and a paper published in the respective conference or journal.

IV. COUNTERMEASURES WITH BLOCKCHAIN TECHNOLOGY

Based on increasing cybercrimes and cyberattacks reported in Post Covid era, various countermeasures, and security strategies are pointed out as Legal countermeasures, Phase-wise ethics, Minimum use of Autonomous Weapons Systems (AWS), Robot Weaponry (RW) & Autonomous Vehicles (AV), Detecting insider threats from social & online data, Increase cyber capabilities & awareness, and Online gaming prevention and authentication [16] (see Fig. 1). Besides developing legal countermeasures and an increase in military cyber knowledge/power, implementation of phase-wise ethics and minimal use of AWS, RW, and AV will certainly help to counter cybercrimes and cyberattacks. The proposed model by authors (EAMV model as Ethics Authentication Monitoring Verification for online data) is also solely focused on the prevention of cybercrime in post-COVID scenarios. EAMV are described as four firewalls or security processes that hackers have to go through one by one in order to access user data or information.
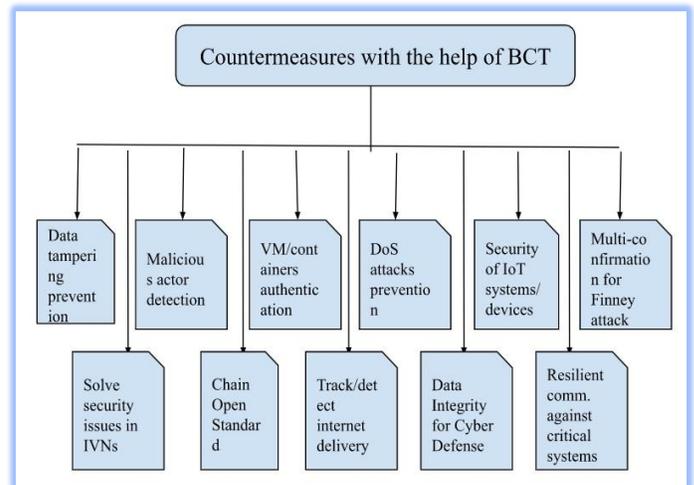
Fig. 1. Various Countermeasures and Cyber Defense Techniques with the Support of Blockchain Technology (BCT) and the Types of Blockchain used for such Countermeasures [16] - [21] [25].

One of the greatest discussed methods for safeguarding data storage, and transfer from decentralized, trustless, peer-to-peer systems is the blockchain. The preliminary keyword explorations by authors highlight blockchain as a standalone technology that carries with it an exorbitant array of possible answers for finance, logistics, healthcare, and cybersecurity [17]. Blockchain uses for cyber security have progressed and bolstered the standing efforts to deter mischievous actors and to boost security. There is an ever-increasing prerequisite to securely cope the cryptography and certification orders as the

WWW transports towards bulk adoption of HTTPS encryption and end handlers are expending it. Ethereum permits for very customizable programming of smart contracts and blockchain uses in the dialectal Solidity. Several major studies alarmed with IoT have offered their own clarifications such as the Proof-of-Possession in the IoTChain proposal.

Not only countering the attacks over the web and malicious actor detection, but a decentralized database of Blockchain on cryptographic practices is also gaining consideration to comfort the security of IoT systems [18]. The blockchain outline within an IoT system is a fascinating auxiliary for old-style database models, which has apprehensions about fulfilling the request for smart home safety. Devices such as door ringlets, light bulbs, control switches, sensor garage doors, etc. are increasing in the marketplace which has limited storage and processing power. Sensors are everywhere surrounded by the home devices via wireless connectivity for remote admission by owners to operate the devices. Design and implementation of a secure home framework model for household IoT devices on Consortium Blockchain (a refined version of blockchain) are experimented with and proposed.

The blockchain perform duties as an immutable log that allows dealings in a decentralized approach. Security threats and existing vulnerabilities to the blockchain systems because of transactions collected into bricks for processing and customary network protocol guarantees each node receives each transaction in close range real-time [19]. Some of the possible and major threats to blockchain systems or blockchain networks are Double-spending security threats, Race attacks, Finney attacks, Brute force attacks, Vector 76 attacks (One confirmation attack), Alternative history attacks, 51% (or > 50%) attack (Majority hash rate attack), Block discarding attack (Selfish mining attack), Block-withholding (BWH) attack, and Fork-after-withholding (FAW) attack. Possible countermeasures for the mentioned attacks include (i) installing observers in the web, communicating binary spending cautions among groups; (ii) waiting for multi-confirm actions for transactions; (iii) informing the merchant about an ongoing binary consume in the network; (iv) disincentive huge mining reserves, twinsCoin, PieceWork; (v) ZeroBlock technique, a timestamp-form technique such as bloomness preferred, DECOR+ protocol; (vi) involve only familiar and committed miners in the pool, cease and terminate a pool when revenue sinks from calculated or sudden change occurs; and (vii) use of cryptographic commitment schemes.

The connection amid the vehicular network and the exterior world has several security pigpens that hackers can be custom to abuse a vehicular network. Some popular protocols for in-vehicle networks (IVN) are Controller Area Network (CAN), FlexRay, and automotive Ethernet, which are not designed with security in concentration. A blockchain is a noble approach to solving standing security issues in IVNs and using hybrid blockchain, there exist several ways to improve IVN security [20]. Protocols used for IVNs have numerous vulnerabilities, such as a shortage of message authentication, a shortage of message encryption, and an ID-based arbitration appliance for conflict determination. Sophisticated attacks by hackers can be launched that may clue to the loss of property and lives using these vulnerabilities. Although several algorithms were proposed in the past for IVN intrusion detection, there are several limitations to the proposed algorithms.

Blockchain technology has cracked two major problems of the digital economy - (i) once assets are digitized, movement can be over the chain, and (ii) zero-trust cost under anonymous societies for innovative chances for the internet economy. BCT is a double-edged weapon for old-fashioned economic and financial progress that places a high demand on data handling and risk-response capabilities [21]. The ultimate goal of BC is to analyze and process information through the active integration of financial funds. BCT can benefit the financial industry to automatically and accurately detect customer credit conditions to restructure the financial shop credit system and progress the efficiency of cross-border payment. To advance the blockchain financial and economic provision quality, smart contracts can be measured in order to track criminal activities. Supervision of financial firms appealing in financial derivatives ought to be strengthened; management authorities must stipulate the bottom capital of financial firms.

V. ADVANTAGES OF BLOCKCHAIN TECHNOLOGY W.R.T. CYBER-DEFENCE AND CYBER-SECURITY

Blockchain technology is one of the popularly noted and known terms nowadays in the field of secure transmission of credential data over the internet. Each block contains user data with high secure hash code as well as the previous block hash code and forms a secured chain. Any unauthorized accessibility can be easily detected in blockchain technology. With its strong security policy, this technology uses in many areas like Markets (Billing, Marketing, etc.), Government Sector (Document digitization / Contracts, Voting, Registries etc.), IOT (Self-driving cars, personalized robots) [27] , HealthCare systems, Finance, and Accounting and many more.

As we know that in the current scenario where the number of internet users increased day by day, the number of cyber-attacks also increased comparatively [28]. The actual data of a user is threatened by the attacker and used in an unauthorized manner. So, blockchain technology with its security mechanism helps to increase the security in cyber-attacks. The use of this technology has many advantages over cyber defense [29] and cyber security [30] which are categorized in the Table II and Table III.

TABLE II.   ADVANTAGES OF BLOCKCHAIN TECHNOLOGY W.R.T. CYBER DEFENSE THAT HELPS TO INCREASE SECURITY IN CYBER-ATTACKS [27] - [31]

S.N.	Security aspects of BC	Advantages over Cyber Defense
1	Decentralized data	It reduces possible attacks. The attackers did not know where the actual data was stored.
2	Traceability	It increases the cyber defense approach by easily tracing which block gets infected and prevents it.
3	Strong Encryption	A strong encryption mechanism using hash code increases cyber defense.
4	Transparency	Transactional data can be digitally authorized and accessible by all the members of the network.
5	No Setup and maintenance of h/w, s/w	No additional security infrastructure is placed for cyber-attacks.

TABLE III.   ADVANTAGES OF BLOCKCHAIN TECHNOLOGY W.R.T. CYBER SECURITY THAT HELPS TO INCREASE SECURITY IN CYBER-ATTACKS [27] - [31]

S.N.	Security aspects of BC	Advantages over Cyber Security
1	Secured Data	User credential data is secured in blocks with high-security hash codes. It reduces cyber-attacks and increases security.
2	Unchangeable transactions	The insertion of new blocks in a chain cannot be removed or modified. Not an easy task for attackers to change the block data or address.
3	Prevention of Fraud	The consensus approach prevents it from external access and cyber-attacks.
4	No middleware security breach	Data transfer from sender to receiver without the need of mediating third parties. So, in the middle attacks get reduced.
5	No Single point failure	Any type of network or cyber-attacks will not generate any type of data loss because of multiple copies of data at different sites.

VI. BLOCKCHAIN-BASED SOLUTIONS AND CHALLENGES

Keeping transparency and trust as vital drivers, blockchain technology might be a hopeful technology. Findings illustrates that BCT has the ability to aid transparency and build citizens' trust in community deal delivery while continuing a sufficient level of privacy [32]. Transparency of evidence and procedures shows a vivacious role in gaining the citizens' trust. Blockchain has a disrupting role in handling transparency, trust, citizen satisfaction, and reducing corruption in order to develop the efficiency of civic service delivery.

Although there are several solutions proposed and used by Blockchain for Cybersecurity and Cyber Defense, there are also huge challenges while embedding blockchain solutions on the other end. To come up with exact blockchain solutions and ongoing challenges and issues, we have summarized the blockchain challenges and blockchain solutions in the Table IV with respect to cybersecurity, cyber defense, and countermeasures.

TABLE IV.   BLOCKCHAIN-BASED SOLUTIONS AND BLOCKCHAIN-BASED CHALLENGES FOR CYBER DEFENSE AND CYBERSECURITY WHILE INCORPORATING THE SOLUTIONS [17] - [26]

SN	Ref. (Year)	Solution domain(s)	BC Type	BC solutions for cybersecurity & cyber defense	BC challenges while incorporating those solutions
1.	[17] (2019)	Sidechain Security, IoT Security	Public & Private	Deployed to decipher problems related to the safety of devices, networks, and users. Cryptocurrency is preserved through a Proof-of-Work (PoW) tool.	A decentralized, trustless system cannot by this one crack all glitches in cybersecurity. Need to strongly succeed the adjoining cryptography and certification patterns.
2.	[18] (2020)	Smart Home Security	Consortium	Blockchain framework (for a secure smart home within an IoT system), a decentralized database, as a substitute to traditional centralized database models.	The complexity to instrument smart contract results possibly can rise the system fee. Public blockchain design is not suitable for smart homes owing to scalability.
3.	[19] (2018)	Privacy and Transaction Security	Public, Private & Consortium	Contains a certifiable record of each and every transaction ever finished in the system. BC is a distributed sleeve system where contestants keep replicas of files and approve on the alterations by consensus.	Despite the giant opportunities BC compromises, it undergoes from challenges and limitations such as scalability, privacy, compliance, and government disputes that have not been discovered and addressed.
4.	[20] (2021)	In-Vehicle Network (IVN) Security	Hybrid	BC is a good approach to solving existing security issues in in-vehicle networks (IVNs). A suggested way to develop IVN security is constructed on a hybrid blockchain.	Hybrid BC framework for securing in-vehicle webs uses private BC to shelter message flows between sensors and components inside the vehicle and open BC for connecting which is insecure.
5.	[21] (2018)	Finance & Economics (Cross-Border Payment Security)	Public & Private	The goal of BC in the internet era is to analyze and route data through the current integration of financial services. BC can exactly identify consumer credit situations to rearrange the market credit system &advance cross-border payment efficiency.	The weak foundation of related research, difficulties encountered, and finance applications of BC are in their infancy in countries. Governments and countries should build a general-purpose application facility platform for enterprises via BC.
6.	[22] (2020)	Electronic Health Record (EHR) Security	Private	The properties for protected EHR systems (like Data accuracy & integrity, Data privacy, Efficient data sharing, Control return of EHRs back to	Has numerous restrictions and aggressive extensions will need fundamental protocol redesign. Personal healthcare figures collected are tall in volume and at

				patients) can be accomplished using blockchain as Decentralization, Security, Pseudonymity, Immutability, and Autonomy.	reckless rate. It could lead to high network invisibility due to physical space& traffic congestion. Also, mining process may cost great, limit blockchain use.
7.	[23] (2018)	Cloud-Based Data Security	Public & Private	BC is secure by policy that provides the aptitude to achieve decentralized consensus and consistency, and resilience to planned / unplanned attacks with crucial benefits (settlement without mediator, patient's control).	Strong data integrity outcomes in immutability (data cannot be reformed or deleted). One of the ethics of privacy standard, based on data safety laws, provides the right-to-erasure to folks. Healthcare data can be huge and requires examining.
8.	[24] (2017)	Automotive Security and Privacy	Private	BC based architecture to shelter the user's privacy and to upturn vehicular ecosystem security with evolving services like dynamic vehicle insurance charges and wireless remote software modernizes.	Some of forthcoming research challenges include the Key management (each vehicle owns several communication key with users which may alter) and Data caching (vehicle must take data from cloud, that suffers overhead and interval).
9.	[25] (2016)	National Defense (Cyber-enabled defense systems)	Private	BC work independent of security and trust by conserving truths in dual ways – (i) ensure digital dealings on BC network, (ii) via consensus, events are safe in database, without modification.	Narrow awareness/knowledge of blockchain technology and need to create a line of research to confirm scalable, adaptable, and securable to back missions in air, space, and cyber domains.
10.	[26] (2018)	Electric Vehicles, Cloud and Edge Computing Security	Public & Private	Built on distributed consensus, BC-inspired data coins and energy coins are suggested where data contribution occurrence and energy contribution volume are applied to triumph the proof of work with security results for vehicular exchanges.	Proof grit via data contribution occurrence and energy contribution amounts can stance challenge and reply due to secret data transmission via sensors. EV and sensors will be charged by wired and wireless skill and trading coins can be risky.

VII. CONCLUSION AND FUTURE SCOPE

We have presented a comprehensive survey of how blockchain technology can be useful to provide security over the web/internet and can be used to counter ongoing cyber threats as well as increasing cybercrimes and cyber-attacks. We have summarized the recent blockchain models and architectures used for cyber defense and cybersecurity. Various countermeasures and cyber defense techniques from "prevention of data tampering" to "detecting internet delivery" have been analyzed and discussed. We conclude that blockchain technology can have several advantages over cybersecurity and cyber defense with various security aspects like traceability, encryption, transparency, data security, unchangeable transactions, fraud prevention, no single point failure, and no setup/maintenance of hardware/software etc. We also pointed out several challenges and solutions (blockchain-based), with particular blockchain types like public, private, hybrid, and consortium Blockchain that can act as recommendations for improvement on integrating blockchain with cyber defense and cybersecurity.

There is no doubt that Blockchain technology has already been used in many fields, especially finance, supply chain, digital advertising, IoT Security, and many more. Due to its unique properties, the scope of blockchain technology in the zone of cybersecurity and cyber defense can be revolutionary. Despite its open and public nature, data can be verified and encrypted using the most secure cryptographic technology which makes the authorization access of data or information impossible. The key strength of blockchain technology is decentralization property that highly increases the capabilities of the system in making security decisions of its own. The technology is most likely to be used in IoT and Networking Security, Secure Data Transmission, Securing DNS, Securing Data Storage, Verification of Cyber-Physical Infrastructure, and many more.

REFERENCES

[1] https://www.ironnet.com/topics/what-is-cyber defense#:~:text=Cyber%20defense%20is%20a%20coordinated,that%20occur%20within%20a%20network. Accessed on 25/09/2022.

[2] S. Lee, & S. Kim, "Blockchain as a Cyber Defense: Opportunities, Applications and Challenges", IEEE Access, 10, 2602-2618,2021.

[3] D. Galinec, D. Možnik, & B. Guberina, "Cybersecurity and cyber defense: national level strategic approach", Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije, 58(3), 273-286, 2017.

[4] C. Solar, "Cybersecurity and cyber defense in the emerging democracies", Journal of Cyber Policy, 5(3), 392-412, 2020.

[5] Y. Li, & Q. Liu, "A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments", Energy Reports, 7, 8176-8186, 2021.

[6] C. S. Bhusal, "Blockchain Technology in Agriculture: a case study of blockchain Start-up Companies", Int. J. Comput. Sci. Iinf. Technol., 13(5), 2021.

[7] N. Kshetri, C.S. Bhusal, & D. Chapagain, "BCT-AA: A survey of Blockchain Technology-based Applications in context with Agribusiness.", DOI: https://dx.doi.org/10.2139/ssrn.3834004, Available at SSRN 3834004, 2021.

[8] E. Piscini, D. Dalton, & L. Kehoe, "Blockchain & Cyber Security", https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/technology-media-telecommunications/Blockchain-and-Cyber.pdf, accessed on 02/10/2022.

[9] O. Lage, S. de Diego, B. Urkizu, E. Gómez, & I. Gutiérrez, "Blockchain applications in cybersecurity. Computer Security Threats", 73, 2019.

[10] Razaque et. al. "Avoidance of Cybersecurity Threats with the Deployment of a Web-Based Blockchain Enabled Cybersecurity Awareness System", Applied Sciences 2021, 11, 7880. Published: 26 August 2021, DOI: https://doi.org/10.3390/app11177880.

[11] Razaque et. al. "Blockchain-Enabled Transaction Scanning Method for Money Laundering Detection", Electronics 2021, 10, 1766. Published: 24 July 2021, DOI: https://doi.org/10.3390/electronics10151766.

[12] Alhelou et. al. "Cyber-attack detection and Cyber-security enhancement in Smart DC-micro grid based on Blockchain Technology and Hilbert Huang Transform", IEEE Access 2021, Vol- 9. Published: 12 February 2021, DOI: https://doi.org/10.1109/ACCESS.2021.3059042.

[13] Sadigov et. al. "Blockchain Technology based System-Dynamic Simulation Modeling of Enterprise's Cybersecurity System", 55th International Conference on Economic and Social Development - Baku, 18 - 19 June 2020, Vol. 1/4, P. 399-408. URI: https://essuir. sumdu.edu.ua/handle/123456789/85701.

[14] Lee et. al. "Blockchain as a Cyber Defense: Opportunities, Applications and Challenges", IEEE Access 2022, Volume 10. Published: 16 December 2021, DOI: https://doi.org/10.1109/ACCESS.2021.3136328.

[15] Almiani et. al. "Deep current neural network for IoT intrusion detection system. Science Direct, Simulation Modeling Practice and Theory", Volume 101, May 2020, 102031, DOI: https://doi.org/10.1016/j. simpat.2019.102031.

[16] N. Kshetri and A. Sharma, "A review and analysis of online crime in pre & post COVID scenario with respective counter measures and security strategies", Journal of Engineering, Computing & Architecture (JECA),ISSN: 1934-7197, Volume: XI, Issue: XII, December 2021, DOI: https://doi.org/17.0002.JECA.2021.V11I12.200786.7902.

[17] KKR Choo et. al (2019). "A systematic literature review of blockchain cyber security, Digital Communications and Networks", ScienceDirect, Volume: 6, Issue: 2, May 2020, Pages: 147-156, DOI: https://doi.org/10.1016/j.dcan.2019.01.005.

[18] M. Imran et. al., "Investigating smart home security: Is Blockchain the Answer? IEEE Access, Special Section on Blockchain-Enabled Trustworthy Systems", Volume: 8, Pages: 117802-117816, 2020, DOI: https://doi.org/10.1109/ACCESS.2020.3004662.

[19] N. Rathod and D. Motwani, "Security threats on Blockchain and its countermeasures", International Research Journal of Engineering and Technology (IRJET) Volume: 05, Issue: 11, e-ISSN: 2395-0056, November 2018, Pages: 1636-1642, DOI: IRJET-Security-threats-on-BC-and-its-countermeasures.

[20] SY Nam et. al., "Security issues with In-Vehicle Networks, and Enhanced Countermeasures based on Blockchain", MDPI Electronics 2021, Volume: 10 (Issue: 8), Pages: 893. Published: 8 April 2021, DOI: https://doi.org/10.3390/electronics10080893.

[21] Y. Zheng and T. Huang, "The Challenges and Countermeasures of Blockchain in Finance and Economics", Systems Research & Behavioral Science (SRBS), Wiley Online Library, Volume: 37, Issue: 4, Special Issue: Industry 4.0, July/August 2020, Pages: 691-698, DOI: https://doi.org/10.1002/sres.2710.

[22] D. He et. al., "Applications of blockchain in ensuring the security and privacy of electronic health record systems: A survey. Computers & Security", Elsevier Ltd., 2020, DOI: https://doi.org/10.1016/j.cose. 2020.101966.

[23] C. Esposito et. al., "Blockchain: A panacea for Healthcare cloud-based data security and privacy?", IEEE Cloud Computing, Jan/Feb 2018, Department: Cloud and the Law.

[24] A. Dorri et. al., "BlockChain: A distributed solution to Automotive security and privacy", IEEE Communications Magazine, December 2017, DOI: 10.1109/MCOM.2017.1700879.

[25] N. B. Barnas, "Blockchains in National Defense: Trustworthy systems in a Trustless World", Air University, Maxwell Air Force Base, Alabama, June 2016.

[26] H. Liu et. al., "Blockchain-enabled security in Electric vehicles cloud and Edge computing", IEEE Network May/June 2018, DOI: 10.1109/MNET.2018.1700344.

[27] Tareq Ahram, Arman Sargolzaei, Saman Sargolzaei, Jeff Daniels, Ben Amaba, "Blockchain technology innovations. IEEE Technology & Engineering Management", Conference (TEMSCON).

[28] Kshetri, N., "The global rise of online devices, cybercrime, and cyber defense: Enhancing ethical actions, counter measures, cyber strategy, and approaches", University of Missouri – Saint Louis, ProQuest Dissertations Publishing, 29165177, May 2022, DOI: https://dx.doi.org/10.13140/RG.2.2.33257.57446,

[29] Suhyeon Lee, Seungjoo Kim, "Blockchain as a Cyber Defense: Opportunities, Applications, and Challenges", IEEE Access, Volume: 10, 2021.

[30] A. R. Mathew, "Cyber Security through Blockchain Technology", International Journal of Engineering and Advanced Technology (IJEAT), 2019.

[31] K. Anderson, Capterra, "The benefits of Blockchain for IT, Part 2: Cybersecurity", Published: 6 Jun 2018, https://blog.capterra.com/benefits-of-blockchain-cybersecurity.

[32] Kshetri, N., "Blockchain Technology for Improving Transparency and Citizen's Trust", In: Arai, K. (eds)Advances in Information and Communication (book series AISC, Volume - 1363), FICC 2021, Springer Nature Switzerland AG 2021, Page: 716-735, DOI: https://doi.org/10.1007/978-3-030-73100-7_52.

AUTHORS' PROFILE

Dr. Naresh Kshetri (Member, IEEE) is currently an Assistant Professor of Cyber Security at Lindenwood University, USA. He completed his Master of Computer Applications (MCA) from University of Allahabad, MS (Cybersecurity) from Webster University, and PhD (CS) from the University of Missouri–St. Louis (UMSL), Missouri, USA. He also worked as a graduate teaching assistant/graduate research assistant for the computer science department, UMSL besides working as an Adjunct Instructor (CS) at Lindenwood University. With nine+ years of experience in teaching and research, he has a total of eight publications *(all as first author)* in reputed journals, conferences/book chapters. His current research interests include blockchain technology and cybersecurity. For more about Dr. Kshetri, please visit: https://sites.google.com/view/nareshkshetri.

Chandra Sekhar Bhushal is a Master graduate from Federation University, Australia majoring in Software Engineering in the year 2020. He has also completed a Master of Computer Application (MCA) from Visvesvaraya Technological University (VTU), India (2015). He has 1.5 years of experience in teaching. He has a total of three publications in the International Journal. His current research interests include information security, social engineering, cybersecurity and blockchain technology.

Dr. Purnendu Shekhar Pandey has done his Ph.D. from IIIT-A (Indian Institut e of Information Technology, Allahabad) in Information Technology. He worked in various Government and private institutions. Presently he is working as an Associate Professor in the Department of Computer Science & Engineering, KIET, Group of Institutions, Ghaziabad, U.P. His areas of research are mainly in the field of IoT, Machine Learning, Cyber-Security, Sensor Networks, Network Coding, D2D assisted Networks, etc. He has published various papers in SCI and ESCI Journals, top-notch conferences like ANTS, FRUCT, NOPE, and CCNC (h-index> 25), and reviewer in various peer-reviewed SCI journals. He has given various talks on IoT and discrete process modeling, machine learning in industrial IoT, etc. in various national and international conferences and has hosted various session chairs at various international conferences.

Vasudha is currently working as an Assistant Professor in the Computer Science Department at United Institute of Management Naini, Prayagraj, Uttar Pradesh (India). She has done MTech (Computer Science and Engineering) from Dr. A.P.J. Abdul Kalam Technical University Lucknow, Uttar Pradesh in the year 2018. S he has also completed Masters in Computer Science (MCA) and Bachelor in Computer Science (BCA) from Integral University, Lucknow (U.P.) in the year 2014 and 2011 respectively. She has published two Conference Proceedings papers conducted by IEEE. Her research area is image processing, database security, cybersecurity and blockchain technology.

Low-rate DDoS attack Detection using Deep Learning for SDN-enabled IoT Networks

Abdussalam Ahmed Alashhab[1], Mohd Soperi Mohd Zahid[2], Amgad Muneer[3], Mujaheed Abdullahi[4]

Department of Computer and Information Science, Universiti Teknologi Petronas, Seri Iskandar, Malaysia[1, 2, 3, 4]

Faculty of Information Technology, Alasmarya Islamic University, Zliten, Libya[1]

Abstract—**Software Defined Networks (SDN) can logically route traffic and utilize underutilized network resources, which has enabled the deployment of SDN-enabled Internet of Things (IoT) architecture in many industrial systems. SDN also removes bottlenecks and helps process IoT data efficiently without overloading the network. An SDN-based IoT in an evolving environment is vulnerable to various types of distributed denial of service (DDoS) attacks. Many research papers focus on high-rate DDoS attacks, while few address low-rate DDoS attacks in SDN-based IoT networks. There's a need to enhance the accuracy of LDDoS attack detection in SDN-based IoT networks and OpenFlow communication channel. In this paper, we propose LDDoS attack detection approach based on deep learning (DL) model that consists of an activation function of the Long-Short Term Memory (LSTM) to detect different types of LDDoS attacks in IoT networks by analyzing the characteristic values of different types of LDDoS attacks and natural traffic, improve the accuracy of LDDoS attack detection, and reduce the malicious traffic flow. The experiment result shows that the model achieved an accuracy of 98.88%. In addition, the model has been tested and validated using benchmark Edge IIoTset dataset which consist of cyber security attacks.**

Keywords—SDN; LDDoS attack; OpenFlow; Deep Learning; Long-Short Term Memory

I. INTRODUCTION

Communication networks have significantly evolved, allowing users to connect with each other anytime, anywhere. The proliferation of various intelligent devices and applications is increasing demand and generating unprecedented traffic [1]. The increasing number of Internet-connected objects has made the IoT an increasingly important topic in recent years, and with the introduction of fifth-generation (5G) mobile networks, data traffic in communication networks is expected to increase by 20% in the next three years. The number of IoT devices will reach five billion by 2025 [2]. The exponential rise of the Internet of Things is driving the development of new advanced services with stricter criteria, including flexible administration and low latency. Emerging technologies such as SDN and Network Function Virtualization (NFV) are fundamental to the subsequent 5G mobile networks [3]. These technologies make the network more versatile in terms of hardware management and control, as more complex algorithms can be employed to administer the network and new features can be introduced with simple software updates. Traditional networks frequently require the deployment of vendor-specific hardware and proprietary software, and closed development, which hinder

the introduction of new protocols and technologies in communication networks [4].

To achieve the integrated success that ensures the availability and high performance of advanced communications networks, advanced mechanisms must be developed to provide an adequate level of security to detect sophisticated cyberattacks. This is a significant challenge because IoT devices can handle sensitive data. Many low-cost commercial devices typically do not support robust security mechanisms, making them targets for various attacks. Although SDN can provide extensive functionality for IoT networks, network utilization efficiency has been improved. However, at the same time, SDN still faces many security challenges, such as DoS/DDoS attacks [5, 6], link failure [7, 8], switch data leakage [9], and other common attacks in traditional networks [10]. Distributed Denial of Services (DDoS) attacks is a security threat that has plagued the Internet for more than 20 years. It is becoming even more violent as the Internet evolves, such as with the development of IoT and 5G mobile networks. DDoS is a tremendously devastating attack that hackers frequently use to make numerous requests to a target and thwart the system's standard service. The DDoS attacks have developed from essential to high-rate traffic and sensible low-rate flows. As a result, a new evolution of DDoS attack called Low-rate DDoS (LDDoS) has recently emerged [11]. LDDoS attacks are described by low rapidity, persistence, and concealment, making them difficult to detect. The first problem with the current solutions is that the feature selection is not based on IoT networks, as it only relies on sampling the OpenFlow switches as features using the flow table. In this work, the traffic characteristics of IoT devices are targeted, and the features are sampled from the entire network traffic flows by the SDN controller. This improves the detection accuracy and achieves a precise detection effect. The second problem is that the current solutions have low detection accuracy if the type of network traffic varies. In this work, the LSTM activation function proposed to detect different types of LDDoS attacks by analyzing the characteristic values of different types of LDDoS attacks and natural traffic in IoT networks, so the characteristics are used to update the parameters. The sample is richer, which increases the detection rate, improves the accuracy of LDDoS attack detection, and reduces the flow of malicious traffic.

Therefore, this study proposes a DL-based method to detect LDDoS attacks for IoT network-based SDN. The method is based on the knowledge of the traffic of LDDoS

attacks from the IoT networks at the data layer of SDN architecture, including the network devices and the control plane belonging to the SDN controller. This paper makes the following contributions:

- Propose a Deep Learning-based LDDoS attack detection method for IoT networks based on SDN.

- Design and implement an LDDoS attack in an experimental SDN-based IoT environment that includes various IoT devices and OpenFlow messages managed by the SDN controller.

- Apply RNN as a supervised learning technique for the classification task and evaluate the performance of our method using various performance metrics.

The rest of this paper is organized as follows: Section II Related work. Section III presents Background on SDN, IoT, LDDoS, and Deep Learning. Section IV presents the LDDoS attack challenge. Section V describes our proposed model, analyzes and presents the results. Finally, the last section concludes this paper.

II. RELATED WORK

LDDoS attack detection is critical to communications network infrastructure, especially for advanced technologies such as SDN, IoT, and 5G mobile networks [12]. Presently, researchers in academia and industry are developing detecting algorithms to defend against LDDoS attacks. At the current, attack detection approaches are divided into two types: There are both threshold and machine learning-based detection approaches. Threshold detection methods identify one or more traffic indicators, including traffic rate, packet delay, and maximum entropy. When real-time traffic measurements surpass a predetermined threshold, an attack can occur. The burst time is commonly employed as a detection threshold for LDDoS attacks. In the machine learning-based detection methods, a classifier is used to differentiate between abnormalities and regular traffic. In general, these approaches employ machine learning algorithms to construct a classification model containing the features of the attack behavior to detect the attack. For example, W. Zhijun et al. [13], described a protective strategy based on the dynamic deletion of traffic flow rules. They used the LDDoS attack technique on the SDN data layer to improve identification accuracy and provide numerous functionalities based on the Factorization Machine (FM). In [14], L. Yang and H. Zhao, created an SDN framework based on machine learning to identify and prevent LDDoS threats. For the system, they employed two aspects (traffic detection and flow table delivery). To detect the attack traffic, they employed the SVM algorithm and traffic attributes extracted from the flow table's statistical data.

In [15], K. M. Sudar and P. Deepalakshmi propose that their detection system uses four network flow parameters, including the length of the flow, the number of packets, the relative distribution of the packet interval, and the relative distribution of matched bytes. When the module detects an LDDoS attack, it adds information about the attack flow to the blocklist table and alerts the controller to remove a particular

flow from the flow table by entering mitigation rules. However, the characteristics of the behavior of IoT networks based on SDN are different from the characteristics of the usual network based on SDN. For instance, IoT services have a stable temporal foundation and relatively consistent packet data volume, whereas SDNs separate information control and data forwarding. Therefore, attackers can leverage these new characteristics to launch LDDoS assaults that conceal themselves within normal data flows and are difficult to detect with conventional detection methods. For instance, attackers utilize compromised cameras to launch attacks, and the attack flow conceals itself among outgoing video flows. In this work, we solve the limitations of the current solutions, our method is based on sampling the traffic characteristics of IoT devices, and the characteristics are sampled from the entire network traffic flow by the SDN controller to improve the detection accuracy. In addition, we use the LSTM activation function to detect different types of LDDoS attacks by analyzing the characteristic values of different types of LDDoS attacks and natural traffic in IoT networks, so that the characteristics are used to update the parameters. The sample is richer, which increases the detection rate and improves the accuracy of LDDoS attack detection. Our proposed approach to detect LDDoS attacks using a DL-based model classifies traffic in several steps, which are explained in the Proposed Methodology Section.

III. BACKGROUND ON SDN, IOT, LDDOS, AND DEEP LEARNING

This section provides an overview of SDN model, IoT technology, LDDoS attack, and DL techniques; it also illustrates how LDDoS attacks in IoT network-based SDN cause bottlenecks.

A. Software-Defined Network

SDN is a proposed network model that avoids the limitations of existing network infrastructures, separating control data (control plane) from forwarding data (data plane) and breaking the vertical integration of network management using an SDN controller to control a network through a comprehensive view of all network devices, allowing easy control and flexibility in installing network devices from different vendors. Unlike traditional networks that are managed by multiple components, including specialized vendor software and switches that depend on vendor installation mechanisms, which complicates the management of IP networks and results in vendor solutions that often lack flexibility and scalability [16]. The SDN architecture, as shown in Fig. 1 contains three layers: The infrastructure layer, which contains the network devices, the control layer, which maintains the network, and the application layer, which executes the software on the network, are the three layers that make up the OSI model. At the infrastructure layer, network devices such as OpenFlow switches, IoT devices, gateways, etc. In the control plane, the network is logically managed centrally by the SDN controller, using the SDN data plane to make decisions that are made by the control plane in all network devices. The application layer includes network software and SDN applications that perform functions assigned to a network domain, IoT applications, and other

requirements such as cloud storage and client-server connectivity requirements [17].
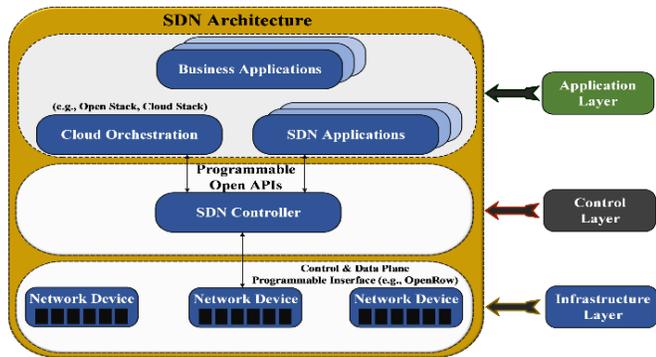
Fig. 1.   Software Defined Networking Architecture.

B. *Internet of Things (IoT)*

The IoT is a computer term that conveys the idea of numerous physical items connected to the Internet and each device's capacity to recognize other devices. It's worth mentioning that the term "thing" on the Internet of Things refers to more than just inanimate items and tiny gadgets. A person wearing a heart rate monitor, a youngster carrying a tracker, a car outfitted with sensors, lighting systems in homes and retail centers, and so on are examples of "things.". In short, the term covers everything that could come to mind, and there are billions of connected objects in the Internet of Things network, which makes controlling and managing these devices a complex task [18]. Moreover, improving security in the IoT context has become a must [19]. The future IoT architecture must be secure enough to prevent the illegal activation of devices. Moreover, since most devices have limited resources, security techniques must be lightweight. In addition, verifying that data is up to date is critical. The lack of adequate security support in IoT can shake the confidence of IoT users and lead to the failure of the technology [20].

C. *Low-Rate Distributed Denial of Service Attack*

The LDDoS attack is a variant of a DDoS attack. Unlike a traditional flooding DDoS attack which requires a huge amount of resources to launch a successful attack by sending large data flows from infected hosts to the target element in the network [21]. LDDoS attack sends a small amount of malicious traffic, representing 20% or less of network traffic. Therefore, it does not show an apparent statistical anomaly to network monitors during the attack. LDDoS attack is hidden in the normal traffic flow and can perform a covert attack through slow traffic, so there is no obvious anomaly for network monitoring. Furthermore, [22], offers a thorough analysis of LDDoS attack detection techniques in software-defined networks. Currently, the most advanced detection methods for LDDoS attacks are mainly divided into three categories: Feature detection, frequency-domain detection, and time-domain detection. In feature detection, a feature dataset must be created. The feature record contains the features of the known LDDoS attack, and once the features of

the LDDoS attack are detected, the LDDoS attack flow is evaluated. In frequency domain detection, the multifractal features of the data flow in the frequency domain are used, and the change conditions in the frequency domain are examined by methods such as spectral analysis, wavelet transform, etc., to detect the LDDoS attack. In time-domain detection, algorithms such as autocorrelation, etc., are used to determine whether the attack flow is present or not by comparing the calculated value with a static threshold [23].

D. *Deep Learning*

Deep Learning (DL) is a subfield of machine learning concerned with discovering ideas and algorithms that enable a computer to learn autonomously by stimulating human neurons. Deep Learning is a science that focuses on the development of methods to obtain a high level of abstraction by studying a huge data set including linear and nonlinear variables. Numerous disciplines, such as speech recognition, face recognition, computer vision, and natural language processing have made substantial, rapid, and applicable strides due to this field's discoveries. The system learns from massive amounts of data using multiple deep learning network architectures, including Recurrent Networks (RNNs) commonly used with text and continuous data, Convolutional Neural Networks (CNNs) guided by biological processes in the visual center, and others capable of extracting raw data features without human intervention. It can meet the high-performance rate by automatically finding the correlation of the raw data by training the model and displaying the results [24]. Moreover [25], provide a comprehensive review based on deep learning capability, approaches for IoT security. To address the current security and privacy issues such as intrusion detection systems (IDS) within IoT environment. Deep learning approaches generally contain a deep structure of hidden layers as shown in Fig. 2. It relies on abstractions and on automatic learning of features that provide facilities for modularization and transfer of learning.

The RNN method is utilized to choose the essential features and then provides the best data classification. As a result, we have used RNN as a supervised learning technique for network traffic classification to distinguish LDDoS traffic from regular traffic.
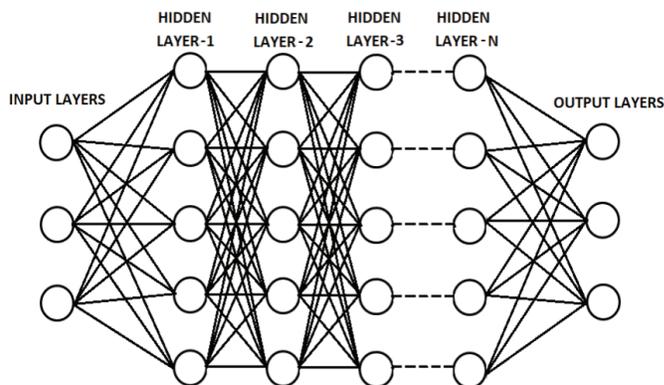
Fig. 2.   Overview of Deep Learning Architecture.

IV. PROPOSED METHODOLOGY

A. SDN-based LDDoS Attack in IoT Networks

In SDN architecture, the infrastructure layer and its communication channel to the controller are most vulnerable to LDDoS threats in an SDN-based IoT network. The infrastructure layer consists of SDN forwarding devices, such as "OpenFlow switches, IoT gateways, and access points; the SDN device can be (re)configured for multiple reasons, including traffic separation, data path modification, and device virtualization". Forwarding rules are recorded in the flow table of a switch and govern the data forwarding path. The controller is accountable for receiving data forwarding requests and giving switch flow rules. SDN apps are responsible for starting data layer configuration instructions by calling controller-integrated functionalities. IoT gateways connect IoT devices using various protocols (WiFi, Bluetooth, ZigBee, etc.) and transmit the collected data to other network devices like switches and routers. Some IoT devices, such as WiFi-based cameras and sensors, can transfer data directly to other network devices by connecting to a WiFi access point without an IoT gateway [26].

In the SDN switches, the idle timeout for the input is the maximum amount of time the flow input can be mismatched. If there is no matching packet in the flow rule within this time, it is automatically discarded. Thus, an LDDoS attack can occur if the transmission rate remains constant and the inverse of the idle time is assured, as is the presence of the input flow and OpenFlow switch. The LDDoS attack target SDN switch, and control channel are seen in Fig. 3; when the capacity of the switch's flow table is complete, it can no longer install new flow rules and will not forward new packets. The number of low-rate attack flows grows in a linear fashion. If the attacker transmits 5 malicious packets during the first idle timeout interval, 10 flow rules are formed in the flow table at the same time. The attack packet is transmitted at regular intervals in the second idle timeout interval to ensure that the flow rules in the flow table do not disappear, but 5 attack packets are added concurrently. The flow table currently contains 20 flow bases. The flow rules will gradually increase as the attack continues until the flow table is saturated.

The data generated by IoT devices can be divided into two types: small data packets generated by sensors in vehicles, houses, etc., and large data packets generated by video display devices from surveillance cameras, among others. The infrastructure layer of the SDN presents these kinds of data packets as ordinary data flows. These two types of devices allow attackers to configure diverse LDDoS attack packets. Such as, they can utilize webcams to send high-volume malicious packets to overwhelm the target and, use IoT sensors to launch malicious volume packets to overwhelm the target or use a combination of both to launch a hybrid attack. For instant, attackers can utilize webcams to send large amounts of malicious traffic or IoT sensors to generate small amounts of malicious traffic, often using a combination of the two to launch a hybrid attack that floods the target. During this work, the model was developed to handle traffic from IoT devices that generate traffic and use traffic features to train

and test the models to present the results presented in the next section.
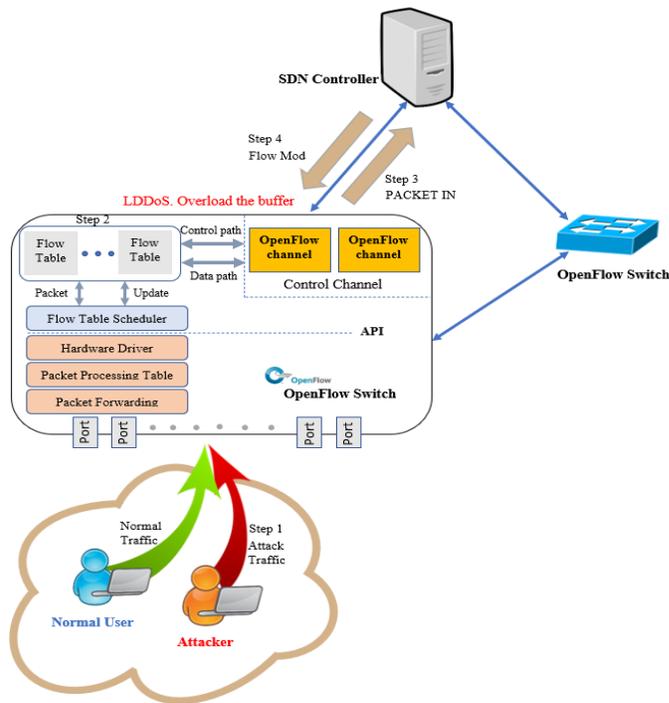
Fig. 3. LDDoS Attack Scenario Targeting SDN Switch and Control Channel.

B. LDDOS Attack Detection Method

The embodiment of the LDDoS attack detection method is based on RNN Deep Learning, which is used to effectively detect various types of LDDoS attacks for SDN-enabled IoT networks. The proposed method was tested using a comprehensive, realistic cybersecurity dataset for IoT and IIoT applications called Edge-IIoTset, which machine learning-based systems can use to detect intrusions. The Edge-IIoTset dataset is divided into several layers, including the IoT perception layer, the IIoT layer, the software-defined network layer and the network function virtualization layer. These layers leverage new emerging technologies that address the key requirements of IoT and IIoT applications, such as the ONOS SDN controller, Things Board IoT platform, and OPNFV platform. IoT data is generated from more than 10 different IoT devices, such as low-cost ultrasonic sensors, water level detection sensors, Ph sensors, soil moisture sensors, heart rate sensors, flame sensors, and digital temperature sensors for temperature and humidity, etc. In the Edge-IIoTset dataset, various attacks including DDoS attacks were identified and analyzed and extracted their features from various sources including alerts, system resources, logs and network traffic [27].

The sampling data were divided into training (70%) and testing (30%). Learning algorithms learn from current datasets and make informed decisions. The proposed detection method contains six main steps: data collection, data pre-processing, training, and testing data, model evaluation, model Prediction, and decision making. As illustrated in Fig. 4, multiple independent processes make up the proposed detection

method overall. The proposed method is based on Recurrent Neural Networks (RNN), where the model consists of an activation function of the LSTM layer to take time-series data as input and learn how to value time, each step of the method is explained as follows.

- Data collection: The datasets have been collected from Edge-IIoT and consist of cybersecurity attacks, which include DoS/DDoS. In these attack categories, the attackers tend to deny the services of legitimate users, either solely or in a distributed fashion. We look at the four most common methods: the TCP SYN Flood, the UDP Flood, the HTTP Flood, and the ICMP Flood. Also, the quality and variety of legitimate entries in a dataset are important for building a profile of how a system normally works. Additionally, malicious entries are essential for security solutions to recognize not only the precise attack patterns but also to identify new ones.

- Data pre-processing: Any machine learning approach requires exploratory data analysis and data observations, so we first create a data collection that can be fed to any classifier. The procedures involve dealing with the missing data, Colum, which was missing value from Edge-IIoT datasets. Our data reprocessing procedures include converting row data into a clean dataset and removing the missing data (column) that was missing from the Edge-IIoT datasets. Second, the phase will unify the engineering steps required to determine the data feature type among the datasets. The data set included categorical datasets as well as numerical data. It was normalized using min-max data normalization as shown in Eq. 1.

$$\mathbf{x}'_i = \frac{\mathbf{x}_i - min\,\mathbf{x}_i}{max\,\mathbf{x}_i - min\,\mathbf{x}_i} \qquad (1)$$

- LSTM model training: The datasets have been split into training 70% and testing 30%. The proposed LSTM model which has been calculated for 30 epochs using training and validation of Edge-IIoT datasets. The model parameter and calculation in percentage is an interpretable way to determine the model performance. The LSTM model has been optimized using Adam optimizer with a learning rate 0.03. The dropout method with a probability of 0.5 has been used to prevent the model from over fitting. Therefore, the proposed LSTM model attempts to sequence-dependent behaviour such as LDDoS attacks detection in network traffic known as IDS. This is performed by feeding back the output of a neural network layer at time T to the input of some layer at time T + 1. Moreover, we used deep learning solution based on LSTM for LDDoS attack detection in SDN enabled IoT networks due to its high efficiency of network data flow.

- Evaluation Metrics: accuracy (ACC) has been used for evaluation measure in this research. The classification models can be evaluated on a variety of parameters, including their accuracy. It depicts its single-class accuracy measurement. Accuracy will also be given by

the total number of predictions made by the primary performance metric used in the behaviour recognition domain that measures different values predicted by a trained model and observed values from the environment. Furthermore, the authors used accuracy for practical decisions and accuracy for model result outcomes in decision-making after training. The accuracy of the proposed LSTM model was determined using Eq. 2.

$$\text{Accuracy} = \frac{(tp + tn)}{(tp + fp + tn + fn)} \qquad (2)$$

where *tp* means true positive, *tn* is a true negative, *fp* denotes false positive, and *fn* is a false negative.

$$\text{Precision} = \frac{tp}{(tp + fp)}, \qquad (3)$$

$$\text{Recall} = \frac{tp}{(tp + fn)}, \qquad (4)$$

$$\text{F-measure} = \frac{2 \times precision \times recall}{recision + recall}, \qquad (5)$$
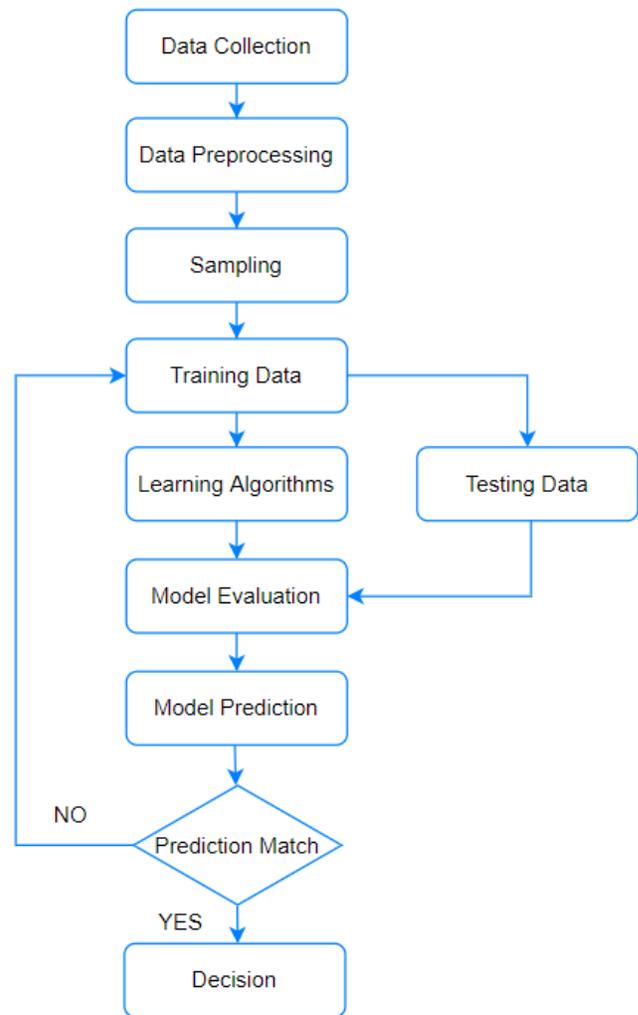
Fig. 4. Flowchart of Proposed Method.

V. RESULTS AND DISCUSSION

According to training and validation data from industrial Edge-IIoTset datasets in Fig. 5, the proposed RNN model's accuracy has been calculated over 30 iterations. The model parameter and percentage calculation are an easy-to-understand way to measure the model's effectiveness. As the result indicate training accuracy is 98.88%, this result can be used for LDDoS attacks prediction and decision, which determines that there is a significant increase after iterations with fluctuation.
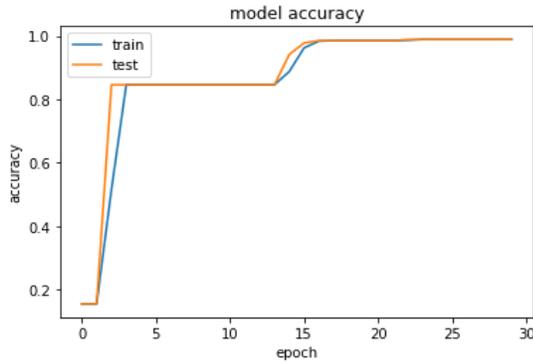
Fig. 5. Training and Validation of Accuracy Performance for 30 Epochs.
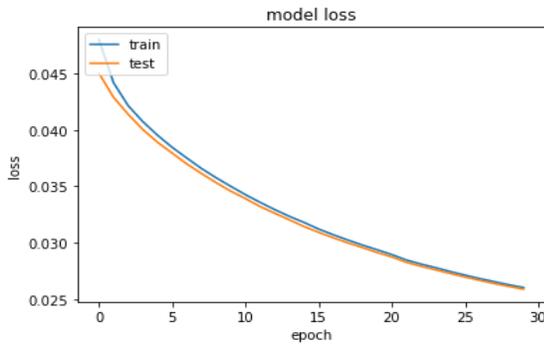
Fig. 6. Training and Validation of Loss Performance for 30 Epochs.

Fig. 6 shows the model's loss performance was also calculated for 30 epochs using training and validation datasets from the industrial Edge-IIoTset dataset. Adding to this, the authors' contribution results show better performance against LDDoS attacks with high accuracy, which was significantly improved after training and testing. Furthermore, the loss parameter calculates training and validation to interpret how the model performance of input datasets. The loss did not increase significantly after iterations but rather decreased. The final loss validation performance was 0.25.

A. Evaluation of Performance

Table I shows the model evaluation measures where the method obtains an accuracy of 98.88% and precision, recall, f1-score of 0.9746, 0.9657, 0.9691, respectively. The obtained performance of the method utilized in this research is presented in Fig. 7, where the model shows a good performance in term of the accuracy and the F1-score.

B. Evaluation Metrics

This section emphasizes o the proposed method performance comparing to the state-of-the-art methods. The proposed LSTM model performance in this work was benchmarked and compared with the related literature contributions with the available performance measures shown in Table II. Our proposed method shows an improvement in the network attack detection accuracy and surpassed other methods exist in the literature review. the proposed method performance outperformed the suggested methods by authors in [13], and authors in [14], as well as the introduced model in the work in [15]. Table II summarize the comparison with related work.

TABLE I. EVALUATION MEASURES OF THE MODEL

Accuracy	Loss rate	Precision	Recall	F1-score
98.88%	0.25	0.9746	0.9657	0.9691
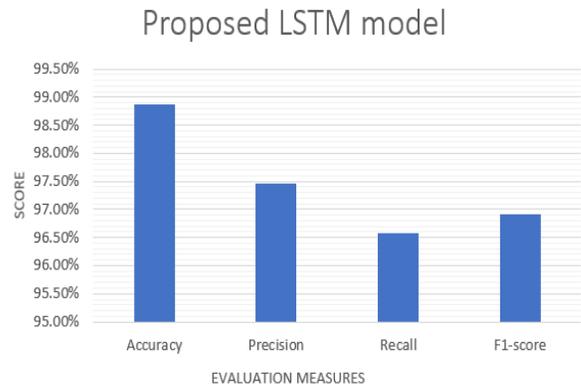
Fig. 7. Results Summary of the Attack Detection using LSTM.

TABLE II. COMPARISON OF PROPOSED METHOD PERFORMANCE WITH THE RELATED LITERATURE CONTRIBUTIONS

References	Year	Layer Location	Classifier/ Method	Dataset	SDN Controller	Network-based	Detection Results
[13]	2019	Data layer	FM	NSL-KDD, CAIDA	Ryu	Traditional wired Network	95.8%
[14]	2021	Control Layer	GBDT, GBDT-LR	Custom	Ryu	Traditional wired Network	96%
[15]	2022	Control Layer and Data Layer	SVM,DT,NB	CIC	POX	Traditional wired Network	93%
Proposed Method	2022	Control Layer and Data Layer	RNN-LSTM	Edge-IIoTset	Ryu	IoT Network	98.8%

VI. CONCLUSION

To conclude, despite the fact that SDN removes bottlenecks and helps handle IoT data efficiently without overloading the network. However, in an evolving environment, SDN-based IoT is vulnerable to various types of distributed denial of service (DDoS) attacks. Traditional DDoS attack detection methods only allow detection of high-rate DDoS attacks; there are problems with low detection accuracy and poor scalability in the case of low-rate DDoS attacks. Our proposed method based on RNN enables detection that targets various LDDoS attacks in the network and more finely divides the types of LDDoS attacks. Meanwhile, for each type of LDDoS attack, the feature types that are significantly different from the normal data flow are identified by the analysis, and the relevant features of the attack can be described using the feature set. Therefore, the proposed detection model can detect various types of LDDoS attacks with high accuracy, output classification results for detection, improve the extensibility of a detection system, and increase the rate of malicious data flow reduction. According to the analysis of the tests performed, our LSTM model has achieved better results in classifying LDDoS attacks because it achieves a fast training time and the results were measured based on accuracy. The accuracy is 98.8% after the training and validation of the model. Finally, one of the future research directions is to implement bio-inspired metaheuristic optimization techniques and investigates their significance performance in obtaining the optimal hyper-parameters and architecture of DNNs with massive-scale data. The other research direction will be exploring deep reinforcement learning in detecting low-rate DDoS attack in IoT networks.

REFERENCES

[1] A. Sivanathan et al., "Classifying IoT devices in smart environments using network traffic characteristics," IEEE Transactions on Mobile Computing, vol. 18, no. 8, pp. 1745-1759, 2018.

[2] Ericsson, "Ericsson Mobility Report," 2019. [Online]. Available: https://www.ericsson.com/assets/local/reports-papers/mobility-report/documents/2019/emr-november-2019.pdf

[3] C. Bouras, P. Ntarzanos, and A. Papazois, "Cost modeling for SDN/NFV based mobile 5G networks," in 2016 8th international congress on ultra modern telecommunications and control systems and workshops (ICUMT), 2016: IEEE, pp. 56-61.

[4] R. Amin, M. Reisslein, and N. Shah, "Hybrid SDN networks: A survey of existing approaches," IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 3259-3306, 2018.

[5] A. A. Alashhab, M. S. M. Zahid, A. A. Barka, and A. M. Albaboh, "Experimenting and evaluating the impact of DoS attacks on different SDN controllers," in 2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA, 2021: IEEE, pp. 722-727.

[6] L. F. Eliyan and R. Di Pietro, "DoS and DDoS attacks in Software Defined Networks: A survey of existing solutions and research challenges," Future Generation Computer Systems, vol. 122, pp. 149-171, 2021.

[7] M. Y. Daha, M. S. M. Zahid, B. Isyaku, and A. A. Alashhab, "Cdra: A community detection based routing algorithm for link failure recovery in software defined networks," International Journal of Advanced Computer Science and Applications, vol. 12, no. 11, 2021.

[8] M. Y. Daha, M. S. M. Zahid, A. Alashhab, and S. U. Hassan, "Comparative Analysis of Community Detection Methods for Link Failure Recovery in Software Defined Networks," in 2021 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), 2021: IEEE, pp. 157-162.

[9] A. H. Shamsan and A. R. Faridi, "Security Issues and Challenges in SDN," in International Conference on Advances in Cyber Security, 2021: Springer, pp. 515-535.

[10] A. N. Alhaj and N. Dutta, "Analysis of security attacks in SDN network: A comprehensive survey," Contemporary Issues in Communication, Cloud and Big Data Analytics, pp. 27-37, 2022.

[11] S. Sambangi, L. Gondi, and S. Aljawarneh, "A Feature Similarity Machine Learning Model for DDoS Attack Detection in Modern Network Environments for Industry 4.0," Computers and Electrical Engineering, vol. 100, p. 107955, 2022.

[12] H. Chen, C. Meng, Z. Shan, Z. Fu, and B. K. Bhargava, "A novel Low-rate Denial of Service attack detection approach in ZigBee wireless sensor network by combining Hilbert-Huang Transformation and Trust Evaluation," IEEE Access, vol. 7, pp. 32853-32866, 2019.

[13] W. Zhijun, X. Qing, W. Jingjie, Y. Meng, and L. Liang, "Low-rate DDoS attack detection based on factorization machine in software defined network," IEEE Access, vol. 8, pp. 17404-17418, 2020.

[14] L. Yang and H. Zhao, "DDoS attack identification and defense using SDN based on machine learning method," in 2018 15th international symposium on pervasive systems, algorithms and networks (I-SPAN), 2018: IEEE, pp. 174-178.

[15] K. M. Sudar and P. Deepalakshmi, "Flow-Based Detection and Mitigation of Low-Rate DDOS Attack in SDN Environment Using Machine Learning Techniques," in IoT and Analytics for Sensor Networks: Springer, 2022, pp. 193-205.

[16] D. Kumar and J. Thakur, "Handling Security Issues in Software-defined Networks (SDNs) Using Machine Learning," in Computational Vision and Bio-Inspired Computing: Springer, 2022, pp. 263-277.

[17] D. B. Rawat and S. R. Reddy, "Software defined networking architecture, security and energy efficiency: A survey," IEEE Communications Surveys & Tutorials, vol. 19, no. 1, pp. 325-346, 2016.

[18] Y. Lu and L. Da Xu, "Internet of Things (IoT) cybersecurity research: A review of current research topics," IEEE Internet of Things Journal, vol. 6, no. 2, pp. 2103-2115, 2018.

[19] I. H. Sarker, A. I. Khan, Y. B. Abushark, and F. Alsolami, "Internet of things (iot) security intelligence: a comprehensive overview, machine learning solutions and research directions," Mobile Networks and Applications, pp. 1-17, 2022.

[20] R. F. Ali, A. Muneer, P. Dominic, S. M. Taib, and E. A. Ghaleb, "Internet of Things (IoT) Security Challenges and Solutions: A Systematic Literature Review," in International Conference on Advances in Cyber Security, 2021: Springer, pp. 128-154.

[21] Z. Li, H. Jin, D. Zou, and B. Yuan, "Exploring new opportunities to defeat low-rate DDoS attack in container-based cloud environment," IEEE Transactions on Parallel and Distributed Systems, vol. 31, no. 3, pp. 695-706, 2019.

[22] A. A. Alashhab, M. S. M. Zahid, M. A. Azim, M. Y. Daha, B. Isyaku, and S. Ali, "A Survey of Low Rate DDoS Detection Techniques Based on Machine Learning in Software-Defined Networks," Symmetry, vol. 14, no. 8, p. 1563, 2022. [Online]. Available: https://www.mdpi.com/2073-8994/14/8/1563.

[23] T. A. Pascoal, I. E. Fonseca, and V. Nigam, "Slow denial-of-service attacks on software defined networks," Computer Networks, vol. 173, p. 107223, 2020.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, pp. 436-444, 2015.

[25] M. Abdullahi et al., "Detecting Cybersecurity Attacks in Internet of Things Using Artificial Intelligence Methods: A Systematic Literature Review," Electronics, vol. 11, no. 2, p. 198, 2022.

[26] D. C. Y. Vargas and C. E. P. Salvador, "Smart IoT gateway for heterogeneous devices interoperability," IEEE Latin America Transactions, vol. 14, no. 8, pp. 3900-3906, 2016.

[27] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning," IEEE Access, vol. 10, pp. 40281-40306, 2022.

Stock Price Forecasting using Convolutional Neural Networks and Optimization Techniques

Nilesh B. Korade[1]

Research Scholar, Department of Computer Science and
Engineering, Madhyanchal Professional University
Ratibad, Bhopal, Madhya Pradesh-462044, India

Dr. Mohd. Zuber[2]

Associate Professor, Department of Computer Science and
Engineering, Madhyanchal Professional University
Ratibad, Bhopal, Madhya Pradesh-462044, India

Abstract—**Forecasting the correct stock price is intriguing and difficult for investors due to its irregular, inherent dynamics, and tricky nature. Convolutional neural networks (CNN) have impressive performance in forecasting stock prices. One of the most crucial tasks when training a CNN on a stock dataset is identifying the optimal hyperparameter that increases accuracy. In this research, we propose the use of the Firefly algorithm to optimize CNN hyperparameters. The hyperparameters for CNN were tuned with the help of Random Search (RS), Particle Swarm Optimization (PSO), and Firefly (FF) algorithms on different epochs, and CNN is trained on selected hyperparameters. Different evaluation metrics are calculated for training and testing datasets. The experimental finding demonstrates that the FF method finds the ideal parameter with a minimal number of fireflies and epochs. The objective function of the optimization technique is to reduce MSE. The PSO method delivers good results with increasing particle counts, while the FF method gives good results with fewer fireflies. In comparison with PSO, the MSE of the FF approach converges with increasing epoch.**

Keywords—Convolutional neural networks; swarm intelligence; random search; particle swarm optimization; firefly

I. INTRODUCTION

The non-linear characteristics of stock market data make it challenging to guess the next movement of stock value. Exact stock forecasting can boost consumer and seller confidence in the stock market, which will attract investors to buy shares and grow the nation's economy [1]. The accuracy of neural networks and their variations in predicting stock prices is rising day by day [2, 3]. Several studies using time-series data have demonstrated that CNN is useful for forecasting issues [4]. CNN can accurately and efficiently identify the changing trend in stock value, and it may be used in other financial transactions. Choosing the best CNN parameters is one of the difficulties we encounter while constructing CNN architectures. The result may vary if we apply different parameter values to CNN architectures while solving the same problem. An optimization process is defined as determining the ideal combination of inputs to reduce or enhance the cost of the objective function without impacting training performance. Optimization is a computational problem whose aim is to extract the best solution among all possible solutions. The hyperparameters that affect CNN's architecture include filters, kernel size, stride, padding, pool size, batch size, epoch, and others. In a reasonable amount of time, we would like to identify the ideal set of hyperparameter values for a particular dataset. To address this issue, several researchers have proposed various techniques based on evolutionary computation to automatically detect the best CNN structures and improve performance [5]. It can be difficult to determine the best parameter in a high-dimensional space.

The procedure used to set the hyper-parameter values is typically a random search, including running a number of tests or making manual adjustments. Random search is the process of selecting and analyzing inputs randomly for the objective function [6]. Swarm intelligence (SI) is inspired by social behavior found in nature, such as the movement of fish and birds. Based on a group's intelligence and behavior, the SI algorithms have a significant ability to determine the optimal solution [7]. The PSO algorithm is considered one of the meta-heuristic evolutionary algorithms. Each particle in PSO has its own position, velocity, and fitness and also keeps track of its best fitness value and best fitness position. The PSO maintains a record of the global best fitness position and the global best fitness value [8]. The FF algorithm is a metaheuristic algorithm based on the attraction of fireflies towards brighter fireflies. The FF algorithm is able to identify optimal parameters for CNN that minimize the error or fitness function with fewer iterations. Each firefly in the FF algorithm has a position in the search space that corresponds to a solution, and it progresses toward more brilliant solutions. For each iteration, the FF algorithm keeps track of the best position, which will reduce the objective function cost [9].

This study evaluates optimization techniques such as RS, PSO, and FF on CNN to forecast stock prices. The Tata Motors stock dataset was taken from Yahoo Finance between January 1, 2003, and September 30, 2022. The CNN is trained on hyperparameter return using RS, PSO, and FF algorithms, and the trained model is used to forecast stock prices for the next day as well as stock prices for the entire month of September. The results show that the FF method returns the best hyperparameters with fewer fireflies, reduces MSE, and requires the fewest training epochs. The remaining paper is structured as follows: In Section II, the stock price forecasting literature is discussed, and the dataset used, the CNN architecture, and different optimization techniques are explained in Section III. The evaluation metrics used and the accuracy of CNN trained on hyperparameter returns by different optimization techniques are discussed in Section IV. We conclude the work in Section V with a summary and recommendations for additional research.

II. Literature Survey

Traditional stock value analysis is based on economics and finance, and it primarily focuses on external factors influencing stock prices, such as international relations, exchange rates, interest rates, business policy, financial institutions, political factors, etc. It's challenging to convince people of the accuracy of the traditional fundamental analysis method. The technical analysis approach primarily concentrates on the movement of the stock price, psychological expectations of investors, trading volume, historical data, etc. [10]. In comparison to other deep learning models, CNN is more accurate and can detect both uptrend and downtrend stock movements [11]. Setting hyperparameters is necessary for CNN implementation, which influences accuracy, learning time, and CNN architecture. The effectiveness of machine learning will be significantly increased if an effective hyperparameter optimization algorithm can be designed to optimize any specific machine learning method. Bayesian optimization based on the Bayesian theorem can be used to solve the hyperparameter tuning problem, as it can be considered an optimization issue [12]. Manual search and RS are the most widely used strategies for hyper-parameter optimization. Randomly searching is more efficient for hyperparameter optimization, where random combinations of hyperparameters are chosen and used to train a model. The hyperparameter combinations with the best costs are selected [13]. Finding the ideal collection of hyperparameter values in a reasonable amount of time is difficult because the values of the hyperparameters change when the dataset changes. The weighted random search approach, which locates the global optimum more quickly than RS, combines RS with a probabilistic greedy heuristic method [14]. Reinforcement learning (RL)-based optimization algorithms give good performance and highly competitive results for CNN hyperparameter tuning with fewer iterations [15].

The hyperparameters include the number of layers, the size of the kernel in each layer, the loss function, the optimizer, and many others. Integers or real numbers may be used as hyperparameters, and there is an infinite range of possible values for each, so CNN hyperparameter tuning is a challenging problem. Swarm intelligence algorithms have been used for decades to solve challenging optimization issues and give promising results [16]. The fish swarm optimization method utilizes a fish's social behavior to carry out a variety of tasks, with each fish serving as a potential solution to the optimization problem. The aquarium is the design space where the fish are found, and the food density is related to an objective function to be optimized [17]. The bee algorithm emulates honey bee foraging activities to find the optimal solution to an optimization issue. The food source or flower is considered a candidate solution, and the population of n bees searches the solution space [18]. Two important parameters of a deep learning network are the number of hidden layers and the number of neurons in each layer. The PSO method has great potential for optimizing parameters and saving precious computational resources when deep learning models are being tuned. Each particle in a PSO has three properties: a position that corresponds to a solution's; a velocity that is a moving parameter; and a fitness value that is calculated by an objective function that takes the particle's position into consideration. The position of every particle in the swarm is updated such that it will migrate toward the one with the best position [19]. One of the most effective metaheuristic algorithms for solving optimization issues is the FF algorithm, which is based on the flashing characteristic of fireflies. The unisex nature of fireflies makes them attracted to brighter lights, and as distance increases, their attractiveness and brightness will decrease [20]. Metaheuristic algorithms, particularly evolutionary and SI algorithms, are strong methods for addressing a wide range of challenging engineering issues that arise in the real world. The echolocation abilities of microbats served as an inspiration for the bat algorithm. Each bat has some position, flying randomly with velocity, fixed frequency, varying wavelength, and loudness in order to search for prey [21]. While training CNN on a stock dataset, it is required to select optimal parameters that improve training performance and reduce error. It is essential to choose the best optimization technique to boost the CNN's performance and reduce errors.

III. Methodology

A. Data Description

From January 1, 2003, until September 30, 2022, historical stock information for Tata Motors was retrieved from Yahoo Finance [22]. Table I shows features and values for Tata Motors stock data. The dataset contains features such as date, open value, low value, close value, high value, volume, and adjacent close value.

It is typical to assess the calculation of profit or loss using the closing price of a stock on a given date; hence, we considered the closing price as the target variable. The plot of the target variable against time is shown in Fig. 1.

TABLE I.   Tata Motors Stock Price Dataset

Date	High	Low	Open	Close	Volume	Adj. Close
2003-01-01	31.26	30.66	30.66	31.15	4460733.00	25.15
2003-01-02	31.41	30.76	31.29	30.84	4800428.00	24.91
2003-01-03	31.22	30.77	30.84	30.88	3939402.00	24.94
….	….	….	….	….	….	….
2022-09-28	406.60	392.85	394.90	399.10	18114880.00	399.10
2022-09-29	413.25	399.60	411.00	402.25	20725995.00	402.25
2022-09-30	408.25	392.50	398.00	404.60	20951277.00	404.60

Fig. 1.   Tata Motors Closing Price.

B. CNN

Deep learning is a subset of machine learning that contains algorithms that are designed to mimic how neural networks or the human brain function. CNN, or ConvNets, is one of the models that has significantly impacted image analysis and computer vision. Recently, there has been a rising interest among researchers in using CNNs for time-series forecasting challenges. We could use a CNN model effectively for the prediction issue if we could transform the 1D time-series sequence into an input image matrix structure. Fig. 2 shows the architecture for CNN.
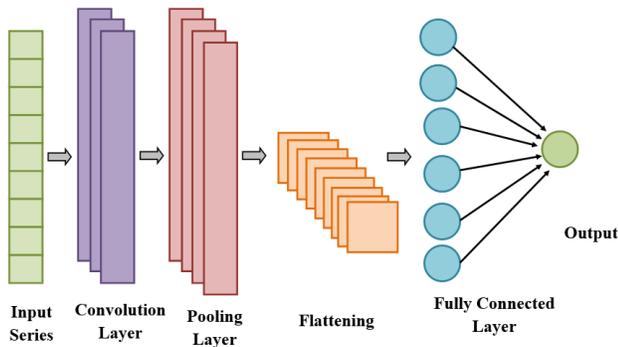
Fig. 2.   CNN Architecture.

The convolution layer, the first layer of the CNN, extracts features from the input by sliding the filter over the input. In order to extract features, CNN uses filter matrices of various sizes that are multiplied by the input matrix. The stride and padding operations influence how the convolution process takes place, resulting in an increase or decrease in the output matrix's dimensions. The size of an output matrix is regulated by CNN using padding, which specifies the number of pixels that are added to an input matrix during the convolution process. The stride, or number of pixels shifted, regulates how the filter convolves across the input matrix. The following formula can be used to calculate the convolution process output matrix size:

$$d = \left(\left(\frac{I - F + 2P}{S} \right) + 1 \right) \tag{1}$$

Where I and F stand for the input and filter matrix dimensions, respectively, S stands for stride, and P stands for padding. The pooling layer's aim is to gradually reduce the spatial dimension of the representation in order to decrease the

number of parameters and calculations in the network by multiplying the resultant matrix from the convolution layer and pooling matrix. The input units are randomly set to zero at a random frequency at each step during training by the dropout layer, which helps in preventing overfitting. The dense layer is a combination of a fully connected layer and an output layer that implements the following operation:

$$O = activation \left((I \cdot K) + bias \right) \tag{2}$$

where 'I' is the input vector, 'K' is a weight matrix generated by the layer, and "bias" is a bias vector generated by the layer. The most frequently used activation functions are tanh, sigmoid, relu, etc.

$$sigmoid\ function\ g(z) = \frac{1}{(1+e^{-z})} \tag{3}$$

$$tanh\ function\ g(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}} \tag{4}$$

$$relu\ function\ g(z) = max(0, z) \tag{5}$$

C. Random Search

Hyperparameter tuning refers to the building of the model architecture from the available space. Random search is a method for finding the optimal solution to build the model by selecting combinations of the hyperparameters at random or using probability.

$$x^{*} = \arg min_{x \in A} f(x) \tag{6}$$

where 'f' is the objective function such as MSE to be minimized on the validation set. 'x*' is the set of hyperparameters for which the objective function returns the lowest value, and any value from the domain 'A' can be assigned to 'x' [23]. The flowchart for RS-CNN is shown in Fig. 3.
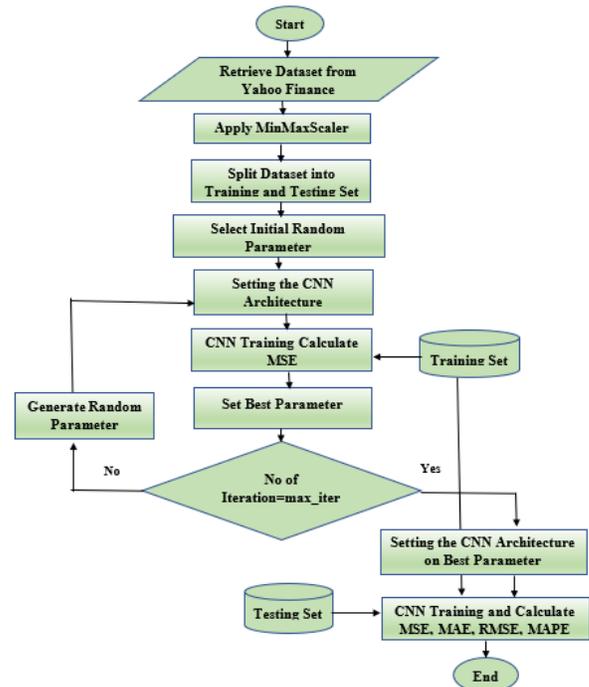
Fig. 3.   Flowchart of the RS-CNN Algorithm.

D. PSO Algorithm

The particle swarm optimization technique is inspired by the social behavior of bird swarms and was designed by Kennedy and Eberhart in 1995. PSO simulates swarm behaviors to optimize the way of finding food, and each bird in the swarms constantly changes the search pattern based on its own and other members' learning experiences [24]. The objective function in data science entails feeding a candidate solution into a model and evaluating it against the training dataset. The cost may be an error score, also known as the loss of the model, which is to be decreased, or an accuracy score, which is to be increased. The optimization process aims to increase or decrease the cost of an objective function. There may be multiple local maximums and minimums for objective functions, but only one global maximum or minimum. Each particle in a PSO has a position, which corresponds to the attributes of a solution; a velocity, which is a moving parameter; and a fitness value, which is determined by an objective function taking into account the position of the particle [25]. The particle's quality is measured by its fitness value. Each swarm particle's position is updated such that it will move closer to the one with the best position. Each particle maintains pbest, the best solution each particle independently found, and gbest, the best solution found by all particles, to update its position and velocity in each iteration [26]. The processing steps of the PSO algorithm are mentioned below [27].

Step 1: Generate random position p and velocity v in all dimensions by using the following equation:

$$x_i = l + r * (u - l) \tag{7}$$

$$v_i = l + r * (u - l) \tag{8}$$

x_i represents the position of particle i, r is a random number, l represents the lower bound, and u represents the upper bound. For velocity calculation, the lower bound is 0 and the upper bound is 1.

Step 2: Calculate objective function value $f(x_i)$ for each particle where i=1..n. The initial position of a particle is pbest for that particle, i.e., pbest=x_i.

Step 3: Find gbest i.e., best position from all particle.

Step 4:

For t=1 to max_iterations

    For i= 1 to no_of_particles

        Update particle velocity and position

        Calculate Objective function value and pbest

    End for i

    Find gbest position having less cost

End for t

Equation 9 describes how to calculate particle velocity, and Equation 10 describes how to calculate particle position.

$$V_i^{t+1} = WV_i^t + C_1 r_1^t (P_t^b - X_t) + C_2 r_2^t (P^q - X_t) \tag{9}$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \tag{10}$$

where V_i^{t+1} and X_i^{t+1} are the new calculated velocity and position, t represents the iteration number, V_i^t is the velocity of the ith particle at iteration t, X_i^t is the position of a particle. W is the inertia weight that regulates the motion of the particles. The particle may continue travelling in the same direction if W = 1, because the particle's motion is entirely determined by the preceding motion. if 0≤W<1, this influence is diminished, causing a particle to travel to different areas within the search domain. The C1, C2 are the correlation factors and if C1=C2 =0, all particles continue to move at their current speed until they collide with the search space border. The r_1 and r_2 are the random number in between 0 and 1. The P_t^b is the particle's best position for iteration t and P^q is the global best position. The updated position is used to evaluate the value for the objection function in each iteration. The particle best position (pbest) and global best position (gbest) are updated based on the objective function return value. The best parameters returned by PSO were used to train CNN, and performance was calculated using evaluation metrics. The flowchart for PSO-CNN is shown in Fig. 4.
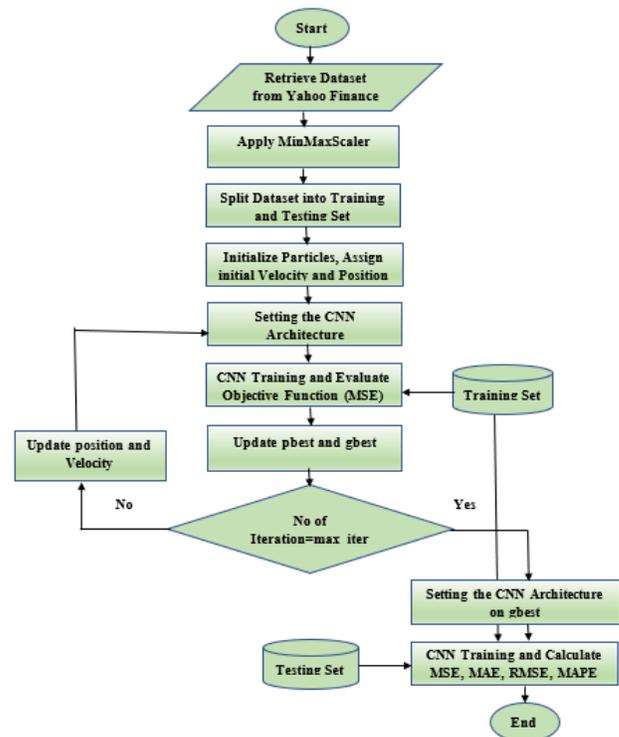
Fig. 4. Flowchart of the PSO-CNN Algorithm.

E. Firefly Algorithm

An optimization issue is one where an objective function is maximized or minimized by selecting appropriate values from a set of possible values for the variables. This research takes into consideration a minimization problem in which the objective function calculates and returns the difference between the actual and forecasted stock price. The Xin-Shi Yang-developed Firefly Algorithm is a stochastic and metaheuristic optimization algorithm inspired by nature. It was designed to mimic the social behavior of fireflies based on their flashing and attraction properties [28].

Below are the guidelines for the FF optimization algorithm:

- An objective function is used to determine the brilliance of a firefly.

- Any firefly is attracted towards more brilliant fireflies, as all fireflies are unisex, and they will move randomly if there is no brilliant firefly.

- Attractiveness is directly proportional to brilliance, and as distance grows, both attractiveness and brilliance will decrease [29].

In an optimization problem, the objective function accepts input variables and returns costs based on their calculation. By taking into account the optimal parameters, an optimization algorithm aims to reduce or enhance the cost return by the objective function [30]. The processing steps of the firefly algorithm are as below.

Step 1: Generate a random position X in all dimension using Equation 11.

$$X_i = l + r * (u - l) \tag{11}$$

X_i is position of firefly i, r is random number, l is lower bound and u is upper bound.

Step 2: Calculate cost, determined by objective function $f(X_i)$, for each firefly where i=1..n.

Step 3:

For t=1 to max_iterations

    For i= 1 to n

        For j= 1 to n

            If $f(X_i) > f(X_j)$

                Move firefly i towards j in all d dimensions

                Evaluate new position and update cost

            Else

                Move firefly i randomly

            End if

        End for j

    End for i

Find best position having less cost

End for T

The following equation describes how a firefly i will migrate when it encounters a firefly j, which is more illuminated [31].

$$X_i^{t+1} = X_i^t + \beta_0 e^{-\gamma r_{ij}^2} * (X_j^t - X_i^t) + \alpha_t \epsilon_i^t \tag{12}$$

where t is the iteration count, X_i^{t+1} is the new position of the ith firefly, X_i^t is the current position of the ith firefly. The term β_0 and α_t in the equation are the attraction and randomization parameters respectively. The recommended values used in most of the implementation are $\beta_0=1$ and $\alpha \in$ [0, 1] respectively. The term ϵ_i is a random number obtained from a Gaussian or uniform distribution. Theoretically, the light absorption coefficient (γ) varies from 0 to ∞, but in most applications, it frequently ranges from 0.01 to 100. The distance r between fireflies i and j can be calculated using the Euclidean distance formula [32].

$$r_{i,j} = \sqrt{\sum_{k=1}^{d}(X_{i,k} - X_{j,k})^2} \tag{13}$$

The flowchart for FF-CNN is shown in Fig. 5. Initially, the FF algorithm initializes a population, assigns a random position, and calculates the objective function cost. The initial position of the firefly is the best position. The cost of each firefly is compared with every other firefly in each iteration, and if the cost of firefly i is greater than the cost of firefly j, then firefly i will move towards firefly j. After n iterations, the best positions were used to train CNN, and accuracy on the training and testing datasets was calculated.
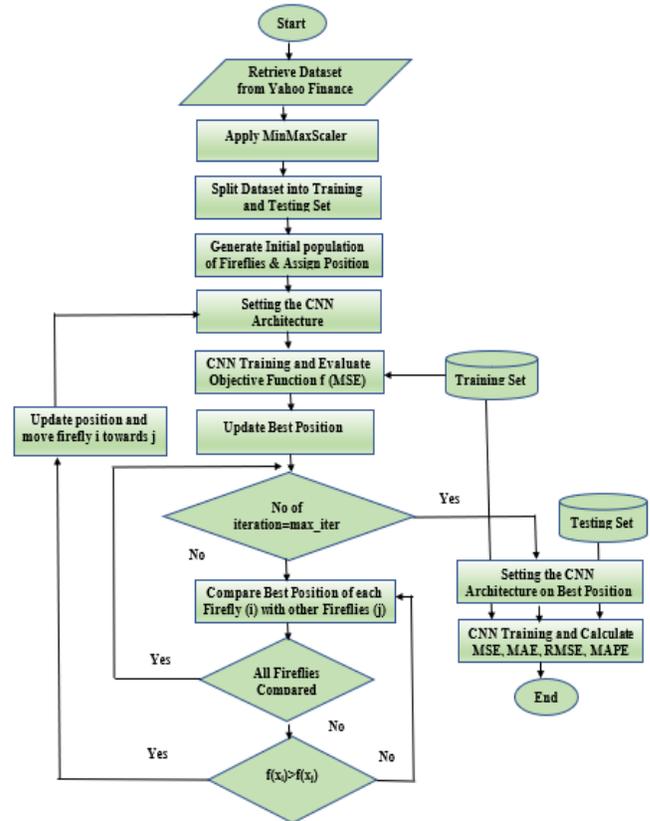
Fig. 5. Flowchart of the FF-CNN Algorithm.

IV. RESULTS

The dataset contains 4902 records split into two sets: the training set has 4882 records covering the period from 01/01/2003 to 31/08/2022; and the testing set has 22 records covering the period from 01/09/2022 to 30/09/2022. The CNN models are evaluated for their ability to predict the stock price for the following day and for the entire month of September. Based on the past 1000 samples, the stock price for the following day is computed, and the stock price for the entire month of September is predicted using the entire training set. The assessment metrics should show the difference between the actual and anticipated value in the stock forecasting problem, which is a regression problem. The performance of

the model is evaluated against MSE, MAE, RMSE, and MAPE [33]. The formula for evaluation metrics is as below.

$$Mean\ Square\ Error = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2 \qquad (14)$$

$$Mean\ Absolute\ Error = \frac{1}{N}\sum_{i=1}^{N}\left|Y_i - \hat{Y}_i\right| \qquad (15)$$

$$Root\ Mean\ Square\ Error = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2} \qquad (16)$$

$$\begin{matrix} Mean\ Absolute \\ Percentage\ Error \end{matrix} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|Y_i - \hat{Y}_i\right|}{Y_i} * 100\% \qquad (17)$$

where N is the total number of samples in the stock dataset, Y_i is the actual stock price, and $\hat{Y}_i$ is the predicted stock price by model. The evaluation metrics shows how forecasted value closer to actual and variation in between them.

The goal of an optimization problem is to identify the optimal solution from a collection of possible options. The parameters for CNN are convolution layer filters, kernel size, pool size in max pooling operation, batch size, etc. If we apply different parameter values to CNN architectures while solving the same problem, the result may vary. Selecting the right hyperparameter gives better accuracy with a minimum epoch. The hyperparameter values returned by the RS, PSO, and FF algorithms were used to train the CNN model. Table II displays the upper and lower bounds for the hyperparameters, and step size is used to increase or decrease their value. Table III provides a summary of the control parameters for the FF and PSO algorithms.

Tables IV and V compare the training and testing performances of CNN, RS-CNN, FF-CNN, and PSO-CNN. The manual hyperparameter was considered while training CNN on different epochs, and resultant metrics were calculated. The number of iterations in the RS-CNN is set to 10, and the current iteration's result is compared to the best one. The FF and PSO algorithms are executed using 5 and 10 particles for different epochs, and CNN is trained using the hyperparameter return from the FF and PSO algorithms. The evaluation metrics for the training set and the testing set are calculated. The objective function for FF and PSO is to reduce MSE. The results reveal that PSO delivers good results with increasing particle counts, while the firefly method gives good results with fewer fireflies. With a large number of particles, the PSO can find the ideal hyperparameter, and the MSE starts decreasing with increasing epoch. In comparison with PSO, the MSE of the FF approach converges with increasing epoch.

The FF-CNN method identifies the ideal parameter with fewer iterations and converges the result with subsequent iterations. This indicates that the FF-CNN strategy has the potential to be more beneficial than CNN, RS-CNN, or PSO-CNN in resolving the stock forecasting problem.

The manual hyperparameter selection method's cost decreases with increasing epoch, while the RS-CNN method's cost varies as parameters are randomly chosen. PSO-CNN requires more particles and epochs, while FF-CNN achieves

better outcomes with fewer fireflies and epochs. If we increase the firefly or epoch, the results converge in the early stages. The evaluation metrics value of a CNN trained on a hyperparameter return by various optimization approaches is shown in Fig. 6 and 7.

TABLE II.   UPPER AND LOWER BOUNDS FOR THE HYPERPARAMETERS

Parameter	Lower Bound	Upper Bound	Step Size
Convolution layer filters	16	128	16
kernel size	1	5	1
pool size	1	5	1
batch size	32	256	32

TABLE III.   SUMMARY OF THE CONTROL PARAMETERS

PSO-CNN		FF-CNN	
Inertia weight (W)	0.5	Attraction parameter (β_0)	0.97
Correlation factors (C_1, C_2)	0.5, 0.5	Randomization parameter (α_t)	1
Random number (r_1, r_2)	0≤1	Absorption coefficient (γ)	0.01
Maximum iteration number	10	Maximum iteration number	10

TABLE IV.   OPTIMIZATION METHODS AND EVALUATION METRICS FOR TRAINING DATASET

Epoch	Metrics	CNN	RS-CNN	PSO-CNN Particle=5	FF-CNN Firefly=5	PSO-CNN Particle=10	FF-CNN Firefly=10
1	MSE	2807.70	1394.60	2028.91	1521.22	1812.57	836.48
	MAE	40.61	28.16	35.05	28.80	31.90	22.19
	RMSE	52.98	37.34	45.04	39.00	42.57	28.92
	MAPE	20.27	13.22	16.14	13.07	15.88	12.95
3	MSE	1466.55	593.12	938.39	569.33	718.46	552.14
	MAE	28.86	18.28	22.99	18.18	19.93	17.73
	RMSE	38.29	24.35	30.63	23.86	26.80	23.49
	MAPE	14.86	8.57	11.06	8.24	9.30	8.22
5	MSE	863.72	1174.88	786.61	330.05	355.49	421.93
	MAE	21.68	26.18	21.20	13.70	14.02	15.58
	RMSE	29.38	34.27	28.04	18.16	18.85	20.54
	MAPE	10.05	13.40	10.27	7.80	7.78	7.09
7	MSE	573.12	953.33	444.52	300.75	335.13	328.43
	MAE	17.89	27.05	15.79	13.10	13.73	13.36
	RMSE	23.93	30.87	21.08	17.34	18.30	18.12
	MAPE	8.21	6.83	8.67	6.61	6.68	5.59
9	MSE	550.88	435.38	454.32	303.94	227.92	198.12
	MAE	17.64	15.85	16.19	12.98	11.37	10.38
	RMSE	23.47	20.86	21.31	17.43	15.09	12.56
	MAPE	8.22	6.63	8.24	6.39	6.58	5.56

TABLE V.   OPTIMIZATION METHODS AND EVALUATION METRICS FOR TESTING DATASET

Epoch	Evaluation Metrics	CNN	RS-CNN	PSO-CNN Particle=5	FF-CNN Firefly=5	PSO-CNN Particle=10	FF-CNN Firefly=10
1	MSE	3360.60	2108.53	1895.34	1457.71	2300.72	1895.70
	MAE	52.54	38.79	38.26	32.72	42.60	38.75
	RMSE	57.97	45.91	43.53	38.18	47.96	43.53
	MAPE	12.41	9.24	9.07	7.78	10.09	9.17
3	MSE	1743.05	681.92	1897.19	325.52	1458.45	381.67
	MAE	36.46	19.93	36.96	13.91	34.31	14.67
	RMSE	41.74	26.11	43.55	18.04	38.18	19.53
	MAPE	8.65	4.79	8.80	3.33	8.11	3.53
5	MSE	1687.88	1185.88	1401.91	188.75	523.64	213.27
	MAE	36.65	28.02	30.78	11.74	18.46	10.84
	RMSE	41.08	34.43	37.44	13.73	22.88	14.60
	MAPE	8.67	6.70	7.35	2.66	4.41	2.60
7	MSE	850.56	953.33	695.28	122.97	191.09	121.54
	MAE	23.45	27.05	22.52	9.55	10.19	8.92
	RMSE	29.16	30.87	26.36	11.08	13.82	11.02
	MAPE	5.61	6.41	5.35	2.17	2.43	2.09
9	MSE	509.63	717.68	336.91	104.75	171.86	102.89
	MAE	17.71	25.39	13.60	8.44	11.18	6.36
	RMSE	22.57	26.78	18.35	10.23	13.10	8.69
	MAPE	4.24	5.78	3.27	1.96	2.55	1.94

Fig. 6.   Evaluation Metrics for Training Dataset.
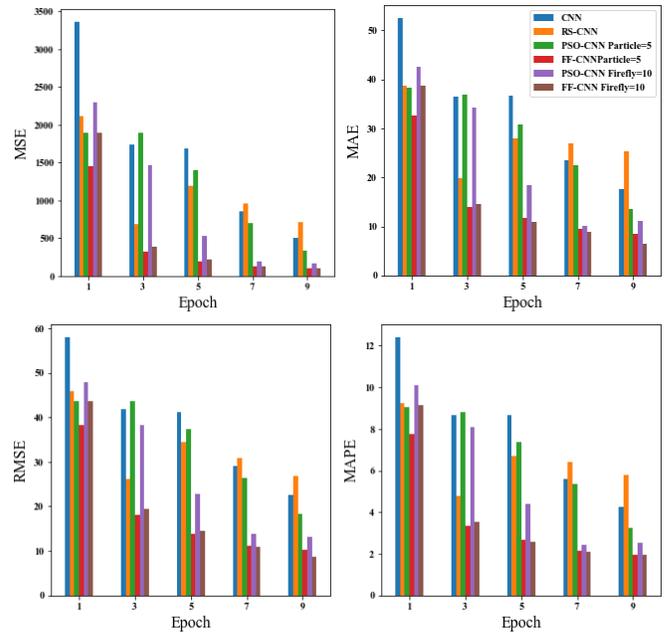
Fig. 7.   Evaluation Metrics for Testing Dataset.

The result shows that the FF-CNN technique returns the best hyperparameter for CNN that gives the best accuracy and a low value for evaluation metrics.

V. CONCLUSION

Multiple parameters, including convolution layer filters, kernel size, pool size of the max pooling operation, batch size, etc., must be determined while training a stock dataset using a CNN. The selection of hyperparameters in CNN has an impact on forecasting accuracy. When selecting a hyperparameter, it is essential to apply optimization techniques that reduce the objective function's cost. This research compares manual selection, random search, PSO, and FF optimization algorithms in order to choose the appropriate hyperparameter for training CNN on a stock dataset. The outcome demonstrates that PSO reduces costs as particle size and epoch increase, whereas FF has the ability to accomplish this with lower firefly counts and epoch. Less MSE is given for both the training and testing datasets by the CNN that has been trained on hyperparameter returns by the FF method. The experimental instance reveals that the FF-CNN framework outperformed other state-of-the-art methods in terms of both the quality of the best solutions and the efficient use of computational resources. Future improvements will focus on modifying the FF algorithm to identify the ideal parameter with fewer iterations, and multichannel CNN can boost the accuracy of the algorithm.

REFERENCES

[1] Milka Grbić, "Stock market development and economic growth: the case of the Republic of Serbia," Post-Communist Economies, 33(4), pp. 484-499, 2021, doi: 10.1080/14631377.2020.1745566.

[2] L. Sayavong, Z. Wu, and S. Chalita, "Research on Stock Price Prediction Method Based on Convolutional Neural Network," 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS),), pp. 173-176, 2019, doi: 10.1109/ICVRIS.2019.00050.
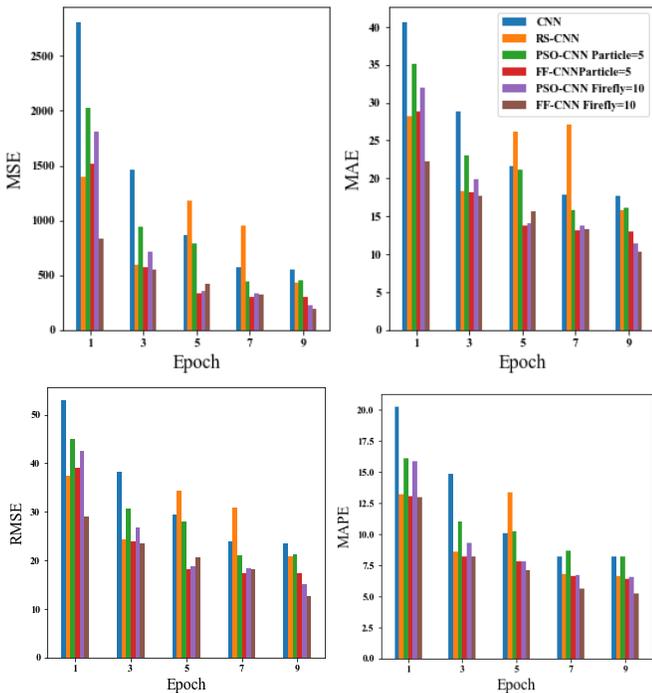
[3] W. Budiharto, "Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM)," Journal of Big Data, 8(47), 2021, doi: 10.1186/s40537-021-00430-0.

[4] R. Chauhan, H. Kaur, and B. Alankar, "Air Quality Forecast using Convolutional Neural Network for Sustainable Development in Urban Environments," Sustainable Cities and Society, 75, 2021, doi: 10.1016/j.scs.2021.103239.

[5] G. Habib, and S. Qureshi, "Optimization and acceleration of convolutional neural networks: A survey," Journal of King Saud University –Computer and Information Sciences, 34(7), pp. 4244-4268, 2022, doi: 10.1016/j.jksuci.2020.10.004.

[6] D.G. Misganu, and K. G. Arnt, "Mapping Seasonal Agricultural Land Use Types Using Deep Learning on Sentinel-2 Image Time Series," remote sensing, 13(2), pp. 289-305, 2021, doi: 10.3390/rs13020289.

[7] J. Xue, and B. Shen, "A novel swarm intelligence optimization approach: sparrow search algorithm," Systems Science & Control Engineering, 8(1), pp. 22–34, 2020, doi: 10.1080/21642583.2019.1708830.

[8] K. Shahana, S. Ghosh, and C. Jeganathan, "A Survey of Particle Swarm Optimization and Random Forest based Land Cover Classification," International Conference on Computing, Communication and Automation (ICCCA2016), pp. 241-2452016, doi: 10.1109/CCAA.2016.7813756.

[9] S. Arora, and S. Singh, "The Firefly Optimization Algorithm: Convergence Analysis and Parameter Selection," International Journal of Computer Applications, 69(3), pp. 48-52, 2013.

[10] L. Wenjie, L. Jiazheng, L. Yifan, S. Aijun, and W. Jingyang, " A CNN-LSTM-Based Model to Forecast Stock Prices," Complexity, 2020, pp. 1076-2787, 2020, doi: 10.1155/2020/6622927.

[11] S. Chen, and H. He, "Stock Prediction Using Convolutional Neural Network," IOP Conference Series: Materials Science and Engineering, 435, 2018, doi:10.1088/1757-899X/435/1/012026.

[12] J. Wu, X. Chen, H. Zhang, L. Xiong, and H. Lei, S. Deng, "Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization," Journal of Electronic Science and Technology, 7(1), pp. 26-40, 2019, doi:10.11989/JEST.1674-862X.80904120.

[13] J. Bergstra, and Y. Bengio, "Random Search for Hyper-Parameter Optimization," Journal of Machine Learning Research, 13, pp. 281-305, 2012.

[14] R. Andonie, and A.C. Florea, "Weighted Random Search for CNN Hyperparameter Optimization," International Journal of Computers Communications & Control, 5(2), 2020, doi: 10.15837/ijccc.2020.2.3868.

[15] F. M. Talaat, and S. A. Gamel, "RL based hyper-parameters optimization algorithm (ROA) for convolutional neural network," Journal of Ambient Intelligence and Humanized Computing, 2022, doi: 10.1007/s12652-022-03788-y.

[16] Y. Cai, and A. Sharma, "Swarm Intelligence Optimization: An Exploration and Application of Machine Learning Technology," Journal of Intelligent Systems, 30(1), pp. 460-469, 2021, doi: 10.1515/jisys-2020-0084.

[17] F. Lobato, and J. V. Steffen, "Fish Swarm Optimization Algorithm Applied to Engineering System Design," Latin American Journal of Solids and Structures, 11, pp. 143-156, 2014, doi: 10.1590/S1679-78252014000100009.

[18] M. Ahmad, A. A. Ikram, R. Lela, I. Wahid, and R. Ulla, "Honey bee algorithm–based efficient cluster formation and optimization scheme in mobile ad hoc networks," International Journal of Distributed Sensor Networks, 13(6), 2017, doi:10.1177/1550147717716815.

[19] Z. Fouad, M. Alfonse, M. Roushdy, A. M. Salem, "Hyper-parameter optimization of convolutional neural network based on particle swarm optimization algorithm", Bulletin of Electrical Engineering and Informatics, 10(6), 2021, pp. 3377-3384, doi: 10.11591/eei.v10i6.3257.

[20] S. Dhawan, "Optimization Performance Analysis of Firefly Algorithm using Standard Benchmark Functions," International Journal of Engineering Research & Technology, 11(2), pp.385-392, 2022, doi: 10.17577/IJERTV11IS020165.

[21] X. Yang and A. H. Gandomi, "Bat Algorithm: A Novel Approach for Global Engineering Optimization," Engineering Computations, 29(5), pp. 464-483, 2012.

[22] Yahoo Finance - Business Finance Stock Market News, <https://in.finance.yahoo.com/> [Accessed on October 01,2022].

[23] E. Elgeldawi, A. Sayed, A. R. Galal , and A. M. Zaki, "Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis," Informatics, 2021, 8(4), pp. 79-99, 2021, doi:10.3390/informatics8040079.

[24] B. Qolomany, M. Maabreh, A. Al-Fuqaha, A. Gupta, and D. Benhaddou, "Parameters optimization of deep learning models using Particle swarm optimization," 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 1285-1290, 2017, doi: 10.1109/IWCMC.2017.7986470.

[25] D. Wang, D. Tan, and L. Liu, "Particle swarm optimization algorithm: an overview," Soft Comput, 22, pp.387-408, 2018, doi: 10.1007/s00500-016-2474-6.

[26] A. G. Gad, "Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review," Archives of Computational Methods in Engineering, 29, pp. 2531–2561, 2022, doi: 10.1007/s11831-021-09694-4.

[27] J. Kennedy and R. Eberhart, "Particle swarm optimization," Proceedings of ICNN'95 - International Conference on Neural Networks, 4, pp. 1942-1948, 1995, doi: 10.1109/ICNN.1995.488968.

[28] A. Abdulhadi, "Application of the Firefly Algorithm for Optimal Production and Demand Forecasting at Selected Industrial Plant," Open Journal of Business and Management, 08, pp. 2451-2459, 2022, doi:10.4236/ojbm.2020.86151.

[29] A. Sharma, A. Zaidi, R. Singh, S. Jain, and A. Sahoo, "Optimization of SVM classifier using Firefly algorithm," 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), pp. 198-202, 2013, doi: 10.1109/ICIIP.2013.6707582.

[30] M. Farrell, K. N. Ramadhani, and S. Suyanto, "Combined Firefly Algorithm-Random Forest to Classify Autistic Spectrum Disorders," 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), pp. 505-508, 2020, doi: 10.1109/ISRITI51436.2020.9315396.

[31] [31] XS. Yang, "Firefly Algorithms for Multimodal Optimization," O. Watanabe and T. Zeugmann (Eds.): SAGA 2009, Lecture Notes in Computer Science, 5792, pp. 169–178 doi: 10.1007/978-3-642-04944-6_14.

[32] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, "Euclidean Distance Geometry and Applications" SIAM Review. 56, 2012 doi: 10.1137/120875909.

[33] A. Botchkarev, "Evaluating performance of regression machine learning models using multiple error metrics in Azure Machine Learning Studio," SSRN Electronic Journal, 2018, doi: 10.2139/ssrn.3177507.

Vision-based Human Detection by Fine-Tuned SSD Models

Tang Jin Cheng[1], Ahmad Fakhri Ab. Nasir[2]*, Mohd Azraai Mohd Razman[4]

Innovative Manufacturing, Mechatronics and Sports Laboratory[1,2,4]
Faculty of Computing[2]
Faculty of Manufacturing and Mechatronic Engineering Technology[1,4]
Universiti Malaysia Pahang, 26600 Pekan Pahang, Malaysia

Anwar P. P. Abdul Majeed[3]
School of Robotics, XJTLU Entrepreneur College (Taicang), Xi'an Jiaotong-Liverpool University
Suzhou, 215123, P. R. China

Thai Li Lim[5]
TT Vision Holdings Berhad, 11900 Plot 106, Sungai Hilir Keluang 5, Bayan Lepas FIZ.4, Bayan Lepas Pulau Pinang, Malaysia

Abstract—**Human-robot interaction (HRI) and human-robot collaboration (HRC) has become more popular as the industries are taking initiative to idealize the era of automation and digitalization. Introduction of robots are often considered as a risk due to the fact that robots do not own the intelligent as human does. However, the literature that uses deep learning technologies as the base to improve HRI safety are limited, not to mention transfer learning approach. Hence, this study intended to empirically examine the efficacy of transfer learning approach in human detection task by fine-tuning the SSD models. A custom image dataset is developed by using the surveillance system in TT Vision Holdings Berhad and annotated accordingly. Thereafter, the dataset is partitioned into the train, validation, and test set by a ratio of 70:20:10. The learning behaviour of the models was monitored throughout the fine-tuning process via total loss graph. The result reveals that the SSD fine-tuned model with MobileNetV1 achieved 87.20% test AP, which is 6.1% higher than the SSD fine-tuned model with MobileNetV2. As a trade-off, the SSD fine-tuned model with MobileNetV1 attained 46.2 ms inference time on RTX 3070, which is 9.6 ms slower as compared to SSD fine-tuned model with MobileNetV2. Taking test AP as the key metric, SSD fine-tuned model with MobileNetV1 is considered as the best fine-tuned model in this study. In conclusion, it has shown that the transfer learning approach within the deep learning domain can help to protect human from the risk by detecting human at the first place.**

Keywords—*Human detection; deep learning; transfer learning; SSD; fine-tuning; human-robot interactions*

I. Introduction

Robotics and automation systems has become a key technology that can help in creating an idealistic future [1]. According to the worldwide trend, it is reported that although the deployment of concepts related to human-robot collaboration (HRC) and human-robot interaction (HRI) has increase progressively, yet it shows that the adoption of the market on these concepts is still in the early stage [2]. As much as the HRI concepts are concerned, the introduction of robots can be a stumbling block to the industries as it possesses potential risks to the human workers when it comes to sharing of working space between the robots and the human

workers [3]. Perhaps, both human workers and robots must ensure appropriate communication to promote safe collaboration [4].

Insufficient safety devices can be the most remarkable barrier in forming the trust of HRI concept [5]. It has been found out that the occupational safety has ranked as the most important factor in bringing successful HRI applications. Enormous amount of research has been carried out to improve the safety of HRI [6]–[8]. In light of this, the major safety mechanisms can be generally categorized into four senses, which are vision, tactile, audition and distance [9]. It is noted that even though vision-based sensor seems to be the trickiest and most computationally extensive, it is still considered to have the most unique sense owing to its irreplaceable richness of the features extracted from this sense. With that in mind, vision-based sensors were incorporated in the present study.

Recent advances in deep learning and transfer learning have garnered great success in various domains [10]–[12]. In particular to HRI applications, limited number of studies has exploited on the usage of deep learning in ensuring a safe interaction and collaboration, not to mention about using transfer learning approach. Within this context, this study intended to examine the applicability of deep learning-based object detection approach via transfer learning approach to detect human at the first place and evaluate on the performance of such approach.

II. Related Works

Mohammed et al. [13] presented a novel online collision avoidance approach that only relies on inputs from Microsoft Kinect sensors. The depth images were used with the background subtraction approach to obtain the virtual model of the human operators. Distance between human operators and the robot was then computed for the collision detection based on a threshold value. Instead of solely rely on the Microsoft Kinect sensors, Magrini et al. [14] established a layered control architecture to improve the safety of HRC applications by integrating multiple sensors and controllers. Microsoft Kinects sensors were utilized to compute the

*Corresponding Author.

distance between two parties while laser scanners were deployed to ensure in which area the human worker is located. Both studies used dual-camera settings as to obtain the depth information of the workers as well as to detect the presence of the human operators within the robot cell.

A study has claimed that it is the first time a deep neural network is utilized for developing a real-time collision detection system and solid results are yielded with a highly satisfied detection speed [15]. Joint signals were used as the input of the training of the deep neural networks. The authors highlighted that the study is limited to cyclic motion due to the fact that cycle-normalization technique is applied. Another study integrated both tactile and visual perception for safety purposes [16]. 1D-CNN was used for the physical contact detection while 3D-CNN was meant for the visual perception to perform human action recognition. Joint signals and images are used as the input for the networks, respectively.

As for the advance of deep learning, transfer learning seems to be a favourable approach in different research fields. The effectiveness of transfer learning approach are investigated to extract the wink-based EEG signals that is converted by means of continuous wavelet transform (CWT) method [17]. The findings were prominent which can potentially be used for controlling the rehabilitation devices. A deep learning-based pre-trained object detection model were adopted in the research with huge amount of data to diagnose the rice leaf disease in a real-time manner [12]. The results further justify the applicability of deep learning technologies, might as well for the transfer learning approach in yielding promising results across different domain.

III. METHODOLOGY

In this section, a general description of the overall research workflow is provided. First, the image acquisition process is explained and the dataset that is used in this work is reported, followed by the image annotation stage. Next, appropriate files are generated for training the models under specific deep learning framework. The training strategies and procedures are then clarified and visualized in a flowchart for better glance of the process. Lastly, the performance metrics that is used to evaluate the performance of the models are described.

A. Image Acquisition

The surveillance cameras that are mounted in TT Vision Holdings Berhad were used to acquire the image dataset. In particular, the recorded video footages were obtained from the surveillance database to extract relevant images. Since it is always beneficial to introduce variation to the deep learning models, the location of the surveillance cameras was carefully chosen. Among all the surveillance cameras, it was observed that the people working in the production area will have higher possibility of moving around compared to the office due to their work nature. Therefore, only the surveillance cameras that are in the production area were selected.

In total, 1463 images were acquired from the recorded video footages. Example of the obtained images was depicted in Fig. 1. In compliance to the standard data splitting procedure, the dataset was separated into three portions, which are training, validation, and testing. The ratio of the data

splitting used in this study was set to be 70:20:10. In addition, data augmentation methods such as horizontal flipping and cropping were utilized to further increase the size of the image dataset and allow more variation of the dataset.

Fig. 1. Examples of Training Images.

B. Image Annotation

All the human workers that shown up in the images were annotated manually with a tight bounding box and labelled with the "Person" class name. Not to confuse the model, it is noted that in this study the occluded part of the human workers was not annotated. With respect to the image annotation procedure, a popular image annotation tool known as LabelIMG [18] was used to perform the annotation. The output annotation files with the format of PASCAL VOC were generated after the bounding box annotation.

C. TensorFlow Records (TFRecords) Generation

Throughout the study, TensorFlow is used as the deep learning framework to develop the deep learning-based object detection models. Thus, TFRecords files were required to generate because this is the only format that can be translated by TensorFlow library for loading the datasets. For instance, TFRecord files can be understood as a simple format that stores the dataset as sequence of binary strings for efficiency purposes. A script was used to iterates through all the annotations in the XML files so that the annotations can be converted into TFRecord annotation files. Since TFRecords only stores binary strings, the label class are stored in binary value and a label map is required for the annotations to have a reference on the class name. For this reason, a one-class label map was developed for mapping the class integer presented in the TFRecords file and the class name.

D. Transfer Learning: Fine-tuning

An open-source deep learning framework that is developed on top of TensorFlow known as TensorFlow Object Detection API was leveraged since it can be considered as a ready to use toolkit in developing, training and inferencing object detection models. Particularly, the Single Shot Multibox Detector (SSD) within the TensorFlow 2 Detection Model Zoo was trained to perform human detection tasks [19]. The meta-architecture of the fine-tuned SSD models is shown in Fig. 2. In this study, both MobileNetV1 and MobileNet V2 that were pretrained on

COCO 2017 dataset were used as the backbone networks [20], [21]. The feature pyramid network was deployed as the neck to extract richer semantic features [22]. Whereas the single shot convolutional prediction head was used correspond to the class prediction and the bounding box regression.

Instead of training the object detection models from scratch, transfer learning approach known as fine-tuning was applied as it is beneficial from the perspective of training time as well as the requirement of a large dataset. By mean of fine-tuning strategy, it is indicated that only the detection head was subject to the training of human detection task, while the backbone network and the neck was remained.
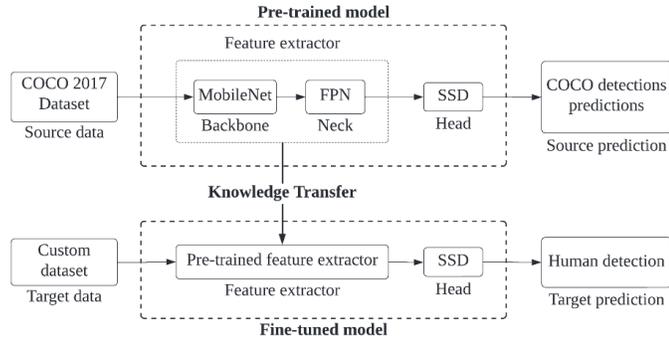
Fig. 2.   Meta-architecture of Fine-tuned SSD Models.

E. Learning Rate and Loss Inspection

Various hyperparameters can be adjusted in order to improve the learning performance of the models. Learning rate acts an important role due to the fact that it can affects the scale of how much the weights are updated within the networks. Instead of using a mere value for the learning rate, this hyperparameter was scheduled via the cosine annealing and warm restart techniques [23]. The motivation behind is to improve the learning behaviour and avoid the occurrence of exploding gradient and vanishing gradient. Both SSD models were configured to undergo a 2.5k warmup steps with learning rate of 0.001. In total, the models were fine-tuned for 100k training steps with learning rate of 0.01.

As for deep learning approaches as well as transfer learning, the learning behaviour has to be monitored throughout the training process. Despite that the scheduling techniques applied can help in improving the model learning, still it is possible for exploding gradient and vanishing gradient to take place. In this sense, loss graphs were utilized to further ensure the learning behaviour of the SSD models goes well throughout the fine-tuning process.

F. Performance Evaluation

In the field object detection, Average Precision (AP) is commonly used as a standard evaluation metric to evaluate the robustness of an object detection model. With regards to the object detection task, precision is computed based on the Intersection over Union (IoU) threshold. IoU can be defined as the ratio of the overlap area between the predicted bounding box and the ground-truth box to the union area of these two boxes. The concept of IoU is visualized in Fig. 3.
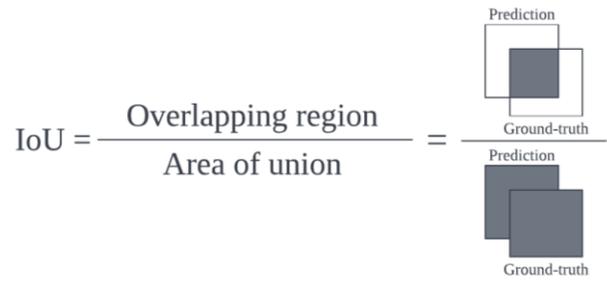
Fig. 3.   Intersection over Union (IoU).

According to the predefined IoU threshold value, the predictions were then be classified as true or false. As such, the IoU threshold value is set to 0.5 in the present study. After the calculation of precision and recall, AP was then calculated by computing the area under the precision-recall curve with the expression as below:

$$AP= \sum_n (Recall_n Recall_{n-1})Precision_n \qquad (1)$$

IV. RESULT AND DISCUSSION

In this section, the results were reported throughout the fine-tuning process. The total loss of the models over the fine-tuning process was visualized and discussed, followed by the performance evaluation of the SSD models after the fine-tuning process. Then, the performance of the SSD models, particularly the AP was reported together with several considerable performance details.

A. Loss Inspection via Loss Graphs

To ensure the fine-tuning of the object detection models are neither underfitting nor overfitting, the learning behaviour of the models were monitored throughout the fine-tuning process via the training and validation loss graph [24], [25]. The training loss and validation loss graphs were plotted as in Fig. 4 and 5. As described in the graph legends, the orange line is the training loss while the grey line is the validation loss. Each of them corresponds to the training set and the validation set of the image dataset.

From both learning curves, it can be clearly seen that the training and the validation losses greatly decrease at the beginning and end up converge at the end of the fine-tuning process. It is reasonable for the losses to be high in the initial stage because the knowledge learned from the previous domain is not specifically cater for the human detection task. However, as the training steps increase, the models had learned the important features for the human detection tasks, hence the losses decrease over the iterations. Towards the end of the fine-tuning process, the losses have become more stable suggesting that the models have converge. In addition, it is shown that the validation loss curve stays above the training loss curve for both SSD models, in turn indicates that the fine-tuned models are neither underfitting nor overfitting.
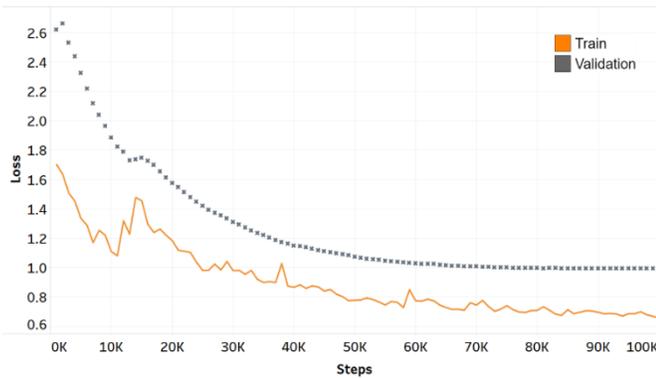
Fig. 4.   Learning Curve of Fine-tuned SSD_MobileNetV1.
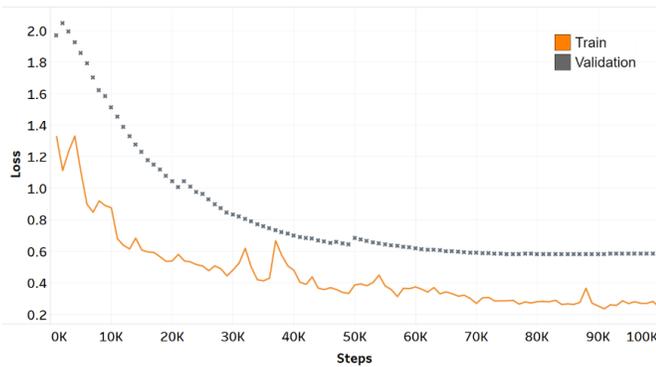
Fig. 5.   Learning Curve of Fine-tuned SSD_MobileNetV2.

B. Performance Evaluation

The AP of the fine-tuned models is depicted as in Fig. 6 and the performance details of the fine-tuned models are tabulated in Table I [26]. The SSD_MobileNetV1 is referring to the SSD model with MobileNetV1 as the base network while SSD_MobileNetV2 is referring to the SSD model with MobileNetV2 as the backbone network. In the performance evaluation stage, the SSD_MobileNetV1 achieves 94.10% train AP, 86.90% validation AP and 87.20% test AP with an inference speed of 46.2 ms. Whereas SSD_MobileNetV2 achieves 89.40% train AP, 82.40% validation AP and 81.10% test AP with an inference speed of 36.6ms. For instance, SSD_MobileNetV1 has a model size of 4.62MB and 10.89 million parameters, while SSD_MobileNetV2 has a model size of 6.46MB and 2.60 million of parameters.

Discussing with regards to the model size and number of parameters, noted that although the SSD_MobileNetV1 fine-tuned model has greater number of parameters than SSD_MobileNetV2, but it has a smaller model size than the SSD_MobileNetV2 fine-tuned model. This is because in the implementation of SSD_MobileNetV2, depthwise separable convolution technique is used for reduction of parameters. Still, the model size is not reduced given the fact the depth of SSD_MobileNetV2 is more than SSD_MobileNetV1.

Despite the comparison of model size and number of parameters, AP and inference speed are the key factors to evaluate the performance of the fine-tuned models. From Fig. 6 and Table I, it can be observed that the

SSD_MobileNetV1 has higher AP for all three train, validation, and test AP with a slower inference speed on RTX 3070 GPU as compared to SSD_MobileNetV2. In fact, the only difference between these two fine-tuned models is the backbone network, hence the differences in AP and inference speed are most probably attributed to the architecture of the base network used in the fine-tuned models [27]–[30]. In a better context, SSD_MobileNetV2 is said to be the recommended fine-tuned model if inference speed comes into consideration before AP. Since this study has considered AP to be more important than inference speed as it is concerned with the HRI safety, SSD_MobileNetV1 is a better choice among these two SSD fine-tuned models. In summary, SSD_MobileNetV1 is proposed as the best fine-tuned model in this case with respect to the human detection task.
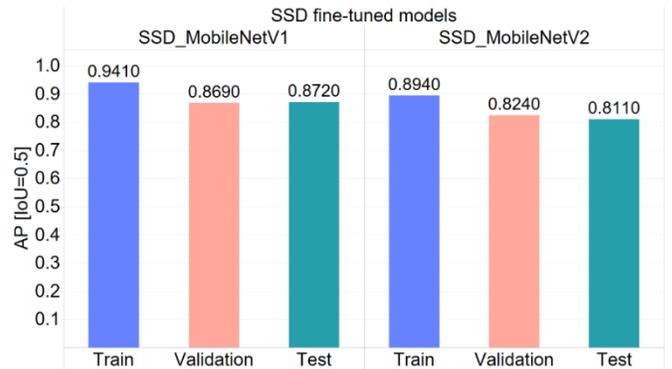
Fig. 6.   Average Precision of SSD Fine-tuned Models with Respect to Train, Validation and Test Dataset.

TABLE I.   PERFORMANCE DETAILS OF THE SSD FINE-TUNED MODELS

SSD Fine-Tuned Models	Performance Details		
	Model Size	No. of Parameters	Inference speed
SSD_MobileNetV1	4.62 MB	10.89M	46.2 ms
SSD_MobileNetV2	6.46 MB	2.60M	36.6 ms

V. CONCLUSION

In the present study, the surveillance system was used to acquire the image dataset for human detection task. The human workers presented in the images were annotated with relevant annotation tools. According to the deep learning framework that has been used in this study, the dataset was developed to specific format that suits the framework. The transfer learning strategy called fine-tuning was leveraged to decrease the training time. In particular, the pre-trained weights of the base network were restored, only the prediction head was subjected to the fine-tuning by using the custom dataset. The result has shown that the SSD_MobileNetV1 fine-tuned model has the highest AP with tolerable sacrifice of inference speed.

ACKNOWLEDGMENT

collaboration with Universiti Malaysia Pahang via UIC200816 and RDU202405.

REFERENCES

[1] L. Onnasch and E. Roesler, "A Taxonomy to Structure and Analyze Human–Robot Interaction," Int. J. Soc. Robot., vol. 13, no. 4, pp. 833–849, 2021, doi: 10.1007/s12369-020-00666-5.

[2] International Federation of Robotics, "World Robotics Report 2019," 2020. [Online]. Available: https://ifr.org/downloads/press2018/2020-09-24_IFR_press_release_WR_industrial_robots.pdf.

[3] M. Vasic and A. Billard, "Safety issues in human-robot interactions," in 2013 IEEE International Conference on Robotics and Automation, May 2013, pp. 197–204, doi: 10.1109/ICRA.2013.6630576.

[4] V. V. Unhelkar, S. Li, and J. A. Shah, "Decision-making for bidirectional communication in sequential human-robot collaborative tasks," ACM/IEEE Int. Conf. Human-Robot Interact., pp. 329–341, 2020, doi: 10.1145/3319502.3374779.

[5] T. Kopp, M. Baumgartner, and S. Kinkel, "Success factors for introducing industrial human-robot interaction in practice: an empirically driven framework," Int. J. Adv. Manuf. Technol., vol. 112, no. 3–4, pp. 685–704, Jan. 2021, doi: 10.1007/s00170-020-06398-0.

[6] Z. M. Bi, C. Luo, Z. Miao, B. Zhang, W. J. Zhang, and L. Wang, "Safety assurance mechanisms of collaborative robotic systems in manufacturing," Robot. Comput. Integr. Manuf., vol. 67, no. January 2020, p. 102022, Feb. 2021, doi: 10.1016/j.rcim.2020.102022.

[7] D. Kim et al., "Design of a sensitive balloon sensor for safe human–robot interaction," Sensors, vol. 21, no. 6, pp. 1–12, 2021, doi: 10.3390/s21062163.

[8] S. Robla-Gomez, V. M. Becerra, J. R. Llata, E. Gonzalez-Sarabia, C. Torre-Ferrero, and J. Perez-Oria, "Working Together: A Review on Safe Human-Robot Collaboration in Industrial Environments," IEEE Access, vol. 5, pp. 26754–26773, 2017, doi: 10.1109/ACCESS.2017.2773127.

[9] A. Cherubini and D. Navarro-Alarcon, "Sensor-Based Control for Collaborative Robots: Fundamentals, Challenges, and Opportunities," Front. Neurorobot., vol. 14, no. January, pp. 1–14, Jan. 2021, doi: 10.3389/fnbot.2020.576846.

[10] O. P. Toon, M. A. Zakaria, A. F. Ab. Nasir, A. P.P. Abdul Majeed, C. Y. Tan, and L. C. Y. Ng, "Autonomous Tomato Harvesting Robotic System in Greenhouses: Deep Learning Classification," MEKATRONIKA, vol. 1, no. 1, pp. 80–86, Jan. 2019, doi: 10.15282/mekatronika.v1i1.1148.

[11] J. L. Mahendra Kumar et al., "An Evaluation of Different Fast Fourier Transform - Transfer Learning Pipelines for the Classification of Wink-based EEG Signals.pdf," MEKATRONIKA, vol. 2, no. 1, pp. 1–7, 2020, doi: https://doi.org/10.15282/mekatronika.v2i1.4881.

[12] B. S. Bari et al., "A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN framework," PeerJ Comput. Sci., vol. 7, pp. 1–27, 2021, doi: 10.7717/PEERJ-CS.432.

[13] A. Mohammed, B. Schmidt, and L. Wang, "Active collision avoidance for human–robot collaboration driven by vision sensors," Int. J. Comput. Integr. Manuf., vol. 30, no. 9, pp. 970–980, Sep. 2017, doi: 10.1080/0951192X.2016.1268269.

[14] E. Magrini, F. Ferraguti, A. J. Ronga, F. Pini, A. De Luca, and F. Leali, "Human-robot coexistence and interaction in open industrial cells," Robot. Comput. Integr. Manuf., vol. 61, no. June 2018, p. 101846, Feb. 2020, doi: 10.1016/j.rcim.2019.101846.

[15] Y. J. Heo, D. Kim, W. Lee, H. Kim, J. Park, and W. K. Chung, "Collision Detection for Industrial Collaborative Robots: A Deep Learning Approach," IEEE Robot. Autom. Lett., vol. 4, no. 2, pp. 740–746, Apr. 2019, doi: 10.1109/LRA.2019.2893400.

[16] F. M. Amin, M. Rezayati, H. W. van de Venn, and H. Karimpour, "A Mixed-Perception Approach for Safe Human–Robot Collaboration in Industrial Automation," Sensors, vol. 20, no. 21, p. 6347, Nov. 2020, doi: 10.3390/s20216347.

[17] J. L. Mahendra Kumar et al., "The classification of EEG-based wink signals: A CWT-Transfer Learning pipeline," ICT Express, vol. 7, no. 4, pp. 421–425, 2021, doi: 10.1016/j.icte.2021.01.004.

[18] L. Tzu Ta, "LabelImg. Git code." 2015, Accessed: Jun. 28, 2021. [Online]. Available: https://github.com/tzutalin/labelImg.

[19] W. Liu et al., "SSD: Single Shot MultiBox Detector," in Eccv, vol. 9905, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 21–37.

[20] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv, Apr. 2017, [Online]. Available: http://arxiv.org/abs/1704.04861.

[21] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Jun. 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[22] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," Proc. - 2019 IEEE Intl Conf Parallel Distrib. Process. with Appl. Big Data Cloud Comput. Sustain. Comput. Commun. Soc. Comput. Networking, ISPA/BDCloud/SustainCom/SocialCom 2019, pp. 1500–1504, Dec. 2016, doi: 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00217.

[23] I. Loshchilov and F. Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts," Aug. 2016, doi: 10.48550/arxiv.1608.03983.

[24] C. Zheng et al., "Detecting glaucoma based on spectral domain optical coherence tomography imaging of peripapillary retinal nerve fiber layer: a comparison study between hand-crafted features and deep learning model," Graefe's Archive for Clinical and Experimental Ophthalmology, vol. 258, no. 3. pp. 577–585, 2020, doi: 10.1007/s00417-019-04543-4.

[25] S. Du et al., "The connectivity evaluation among wells in reservoir utilizing machine learning methods," IEEE Access, vol. 8, pp. 47209–47219, 2020, doi: 10.1109/ACCESS.2020.2976910.

[26] H. J. Lee, I. Ullah, W. Wan, Y. Gao, and Z. Fang, "Real-Time vehicle make and model recognition with the residual squeezenet architecture," Sensors (Switzerland), vol. 19, no. 5, 2019, doi: 10.3390/s19050982.

[27] A. F. Nurfirdausi, S. Soekirno, and S. Aminah, "Implementation of Single Shot Detector (SSD) MobileNet V2 on Disabled Patient's Hand Gesture Recognition as a Notification System," 2021 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSIS 2021, pp. 19–24, 2021, doi: 10.1109/ICACSIS53237.2021.9631333.

[28] W. Rahmaniar and A. Hernawan, "Real-time human detection using deep learning on embedded platforms: A review," J. Robot. Control, vol. 2, no. 6, pp. 462-468Y, 2021, doi: 10.18196/jrc.26123.

[29] X. Gao, J. Xu, C. Luo, J. Zhou, P. Huang, and J. Deng, "Detection of Lower Body for AGV Based on SSD Algorithm with ResNet," Sensors, vol. 22, no. 5, 2022, doi: 10.3390/s22052008.

[30] B. Mathurabai, V. P. Maddali, C. Devineni, and I. Bhukya, "Object Detetcion using SSD-MobileNet," pp. 2668–2671, 2022.

Data Warehouse Analysis and Design based on Research and Service Standards

Lasmedi Afuan, Nurul Hidayat, Dadang Iskandar, Arief Kelik Nugroho, Bangun Wijayanto, Ana Romadhona Yasifa

Informatics, Engeneering Faculty, Universitas Jenderal Soedirman, Purwokerto, Central Java, Indonesia

Abstract—Data are not easy to organize, especially if the data are big in quantity and stored manually and in a non-computerized way. Therefore, in the last few years, many organizations or companies used information systems to help their activities organize and manage the data. Universitas Jenderal Soedirman (UNSOED) is a state college that has existed for a long time and has many Study Programs and Faculties, including the Faculty of Engineering. Data organization in UNSOED is mainly performed through computerization. However, retrieving data needs to be improved because UNSOED has various information systems, and the data produced keeps increasing over time. The data have yet to be evaluated following the needs of Tri Dharma, with indicators of achievement as expressed in the Regulation of Minister of Education and Culture (PERMENDIKBUD) on the National Standard of Higher Education (SNDIKTI). Data warehouse technology can be applied to storing, collecting, and processing media within a specific time from various data sources. The data processing results in the data warehouse are later displayed using the tools Knowage which may help the executives of the Faculty of Engineering make a decision and monitor the businesses, mainly about research and service, by the society of academicians in the Faculty of Engineering regularly from time to time.

Keywords—Data warehouse; knowage; SNDIKTI; UNSOED

I. INTRODUCTION

In this digital era, any information can be easily acquired by anyone. In this era, data are essential and valuable, and any meaningful information can be retrieved for use from there, especially for organizations or companies that deal with data daily. Important information can be produced if the data in possession are processed appropriately. In addition, past data in possession can provide general and other important information that an organization or company may consider in performing their business.

The urgency of an organization or company's data storing keeps increasing, given that the longer an organization or company has been established or developing, the more data they produce. If the data produced are not well managed or stored, the organization or company will find it difficult to retrieve any critical information that it will take a long time to analyze the data, and applying a data management or storing system to the organization or company will take a long time since the data produced will be abundant and relations between data will be unclear. There will be possible duplicate data, requiring cross-checking for their validity.

UNSOED is one of the State Colleges in Indonesia established in 1963 and today has 12 Faculties, including the Faculty of Engineering consisting of five Study Programs: Electrical Engineering, Civil Engineering, Geological Engineering, Informatics, and Industrial Engineering [1]. The Faculty of Engineering took a long time to establish. This started in 2000 with only the Electrical Engineering and Civil Engineering study programs under the Undergraduate Program of Engineering, and finally, in 2014, the Faculty of Engineering was established [2].

In its implementation, Colleges, Faculties, and Study Programs must follow and comply with the existing standards and regulations issued by the Ministry of Education, Culture, Research, and Technology, including the Regulation of Minister of Education and Culture of the Republic of Indonesia Number 3 of 2020 on the National Standard of Higher Education (SNDIKTI), for Colleges to achieve the quality of learning, research and public service when they have exceeded the criteria defined in the National Standards of Higher Education [3].

Fulfilling the whole conditions or indicators in the SNDIKTI requires various data, especially related to Research and Public Service. In UNSOED, an institution manages the whole Research and Service activities performed by the society of academicians, the Research, and the Public Service Institution. In the data recording process, UNSOED has an information system, SINELITABMAS (Research and Public Service Information System) [4], in which the data are interconnected with other systems, including SIHURA (Human Resources Administration System) and SISTER (Integrated Resources Information System) that contain data of lecturers and educational workers.

The data collection, not only to satisfy SNDIKTI but also to satisfy other standardization such as accreditation or quality assurance, still needs to be done and takes a relatively long time for the UNSOED. The reason is that it has too many information systems. However, the data produced still needs to be evaluated following the needs of Tri Dharma Perguruan Tinggi (three pillars of higher education). Therefore, a working instrument is needed to process and store large-scale data from which a variety of important information can be displayed for the executives' use as an instrument and tool of evaluation in decision making.

By implementing a data warehouse and using Business Intelligent tools of Knowage, any data needed will be accessed more quickly, and they will be interconnected with each other. Besides, since the data warehouse is historical, it can be used in the long run. Previous data can be used for analysis and reporting to help executives in management decision-making.

II. RESEARCH METHOD

This research was conducted in phases by applying the Data Warehouse Life Cycle development method as follows [5]-[10]:

1) Literature study: The initial phase was conducted by learning some literatures such as textbooks and journals on data warehouse and, in addition, learning the content of Regulation of Minister of Education and Culture of the Republic of Indonesia Number 3 of 2020 on the National Standard of Higher Education for indicators to use.

2) Design: The initial process was interview with the executives of the Faculty of Engineering UNSOED to examine the kinds of data in possession and a business analysis on the process of and need for data was conducted, and from results of which, what data and where from the data were to be used were derived, were determined. Besides, database designing started from inter-table scheme, designing of dimension table and fact table along with attributes and sizes.

3) Prototype: This process was conducted using the results of designing and the need for data in the previous phase into data visualization or presentation dashboard designing. The prototype was made to adjust and describe user's needs. In case of shortcoming, developer can return to early phase and return to the prototype phase when added needs have been obtained. Second repetition in this phase can keep occurring until final prototype meets users all needs.

4) Deploy: Data presentation dashboard prototype agreed upon was then built into the environment or the tools of production used. In addition, integrated data processing was also conducted in this phase that completely processed data would be immediately stored into the data warehouse environment built.

5) Operate: In this phase, the data warehouse built was operated and tested. Test was conducted only on the developer part (Black Box Testing) and the operation in this phase could be observing suitability of information displayed, inspecting data processing performed, and other operation.

6) Enhance: In this final phase, some aspects lacking in the previous phases were modified and added covering physical components, operating process and data management, and logical scheme designing. One of the corrections included adding one dimension data with dummy data in order to produce information that was more suitable to the content of SNDIKTI.

III. RESULT AND DISCUSSION

A. Design

The design phase was conducted to examine the data needed for use. In this research, the needed data were collected through learning the document SNDIKTI No. 3 of 2020 and reconstructing the model or determining the format of data to be applied, and interview with the executives of the Faculty of Engineering UNSOED. The result of model reconstruction shown in Table I.

TABLE I. MODEL RECONSTRUCTION BASED ON SNDIKTI

Aspect in SNDIKTI	Description
Research Result	Research result is not confidential, does not disturb and/or does not harm public or national interest, must be disseminated through seminar, publication, patent, and/or other means that can be used to deliver the research results to the public.
Research Funding and Financing	College is obliged to provide fund for internal research and research management fund. Funding can also be derived from the government, cooperation with other domestic of foreign institutions, or fund from the public.

From the results of model reconstruction and interview, database was designed consisting of conceptual design, logical design, and physical design processes.

1) Conceptual design: In this phase, data were determined as needed, from process business identification and determining from which data were to acquire. The result of process business identification and some data mappings conducted can be seen on Table II.

TABLE II. RESULT OF PROCESS BUSINESS IDENTIFICATION

Process Business	Description	Function Involved
Research	Is business process related to research, from registration and submission of research proposal, selection, socialization, monitoring and evaluation, seminar of research results, and research publication.	LPPM
Public Service	Is business process related to service from registration and submission of service proposal, selection, socialization, monitoring and evaluation, seminar of public service results, and service publication.	LPPM

Research data could be acquired from some sources, such as data of lecturers from SISTER (Integrated Resources Information System) and data file analitik monev in which the data were contained in Excel file, data of study program could be generated from file of research data of PKM, research data could be acquired from file of research data of PKM and file of research data, data of types of publication and activity categories could be generated from file of publication data. The Mapping of Data Sources in Research Standards show in Fig. 1.
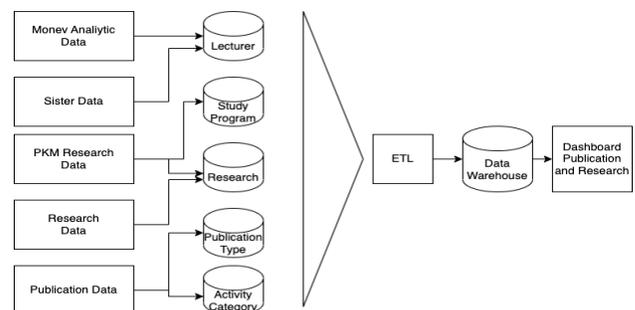
Fig. 1. Mapping of Data Sources in Research Standards.

After determining business process and data mapping, grains were determined to be fact table. Some grains that could later be taken as fact table shown in Table III.

TABLE III. GRAIN DETERMINATION

Grain	Description	Related Business Process
Research Transaction	Number of researches conducted along with publication and funding received	Research
Service Transaction	Number of public services conducted along with publication and funding received	Public service

The next phase was determining dimensions to be used pursuant to the predetermined grains. The description of dimension table to be made shown in Table IV.

TABLE IV. DIMENSION TABLE DETERMINATION

Dimension	Description	Grain
Lecturer	Data of lecturers were taken from data of SISTER in Excel file and from data of Excel file Analitik Monev	Research and Service Transactions
Researcher's Role	Information of researcher's role from Excel file	Research and Service Transactions
Time	Information of data of time including year, date, month, day, and semester generated from 2016 to present	Research and Service Transactions
Activity Category	Existing activity categories from data of research and service sources	Research and Service Transactions
Study Program	Data of study program	Research and Service Transactions
Type of SKIM	Data of the existing types of SKIM from data of research and service sources	Research and Service Transactions
Type of Publication	Data of the existing types of from data of research and service sources	Research and Service Transactions
Source of Dana	Information of source of fund of each research and service	Research and Service Transactions

After dimensions were determined, fact table was designed along with corresponding content of dimension. The fact table design produced can be seen on Table V.

2) Logical design: In this phase, dimensional modeling was made for dimension tables and facts made. The scheme used was star scheme which was a simple and easy-to-understand scheme. A dimensional modeling with star scheme designed to integrate data in data warehouse are shown in Fig. 2.

3) Physical design: In this phase, all dimension table and fact table designs were analyzed for metadata, covering information on the name of the table, detail of contents of the table such as a primary key or in data warehouse commonly known as a surrogate key, name of the field, type of field, size of the field, and source of data.

TABLE V. FACT TABLE DESIGNING

Fact	Description	Content
Fact_penelitianpublikasi	Facts of published research cover number of researches and funding obtained	Judul, biayadisetujui, dim_dosen_scd, dim_peranpeneliti, dim_waktu, dim_kategorikegiatan, dim_prodi, dim_skim, dim_jenispub, dim_sumberdana.
Fact_pengabdian	Facts of published service covers number of researches and amount of funding obtained	Judul, biayadisetujui, dim_dosen_scd, dim_peranpeneliti, dim_waktu, dim_kategorikegiatan, dim_prodi, dim_skim, dim_jenispub, dim_sumberdana.
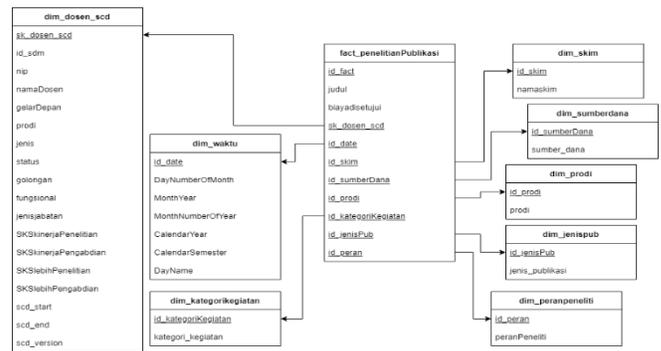
Fig. 2. Star Schema 'Fact_penelitianPublikasi'.

B. Prototype

In the prototype phase, the dashboard was designed to depict how data or information was visualized with various graphical forms as needed. This was designed using Google Spreadsheets as the tools of visualization and using some dummy data for graphical display to be easily understood by both developer and executive. Some designs of the dashboard regarding the research are shown in Fig. 3.
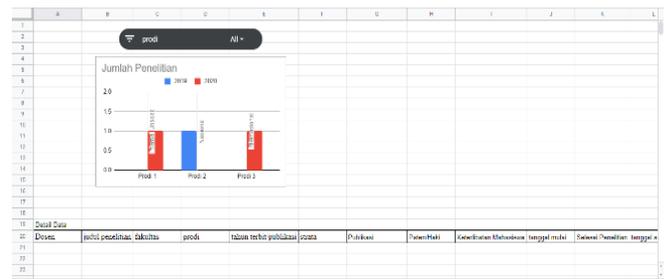
Fig. 3. Research Dashboard Design Aspects of Research Results.

Based on the model reconstruction, in the research result aspect, information of publication along with the respective number by study program in certain years could be displayed.

C. Deploy

The next phase was building data warehouse along with data visualization or presentation in the development environment used. In this research, software Talend Open Studio was used as the instrument in building ETL process

and software Knowage as the instrument in building data visualization or presentation.

ETL process started with storing data in source data that would be used in staging area and in this early phase data transformation processing was still not performed [11]. If the whole source data to be used have been stored in staging area, dimension table or fact table could be formed out of the data. Below is ETL process in building dimension table of lecturer:

1) Extract: Lecturer data stored were retrieved from data of staging area 'stg_sdm' and 'stg_dosen' in which 'stg_dosen' served as lookup while data 'stg_sdm' served as main data. Out of the two data staging, left outer join was performed on attribute of data with the same value for the data to complete each other (see Fig. 4).
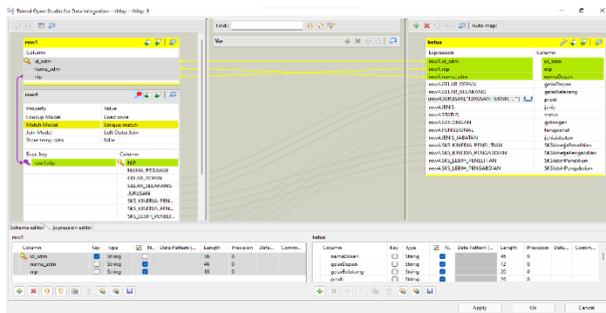
Fig. 4.   Fields or Data Attributes used in Compiling Tables 'dim_dosen_scd'.

2) Transform: The attributes of data to be stored into lecturer dimension table were selected. The attribute 'JURUSAN' on 'stg_dosen' or on *row4* would be changed for more effective writing pursuant to agreed format. For example, regarding data change, data that were previously 'JURUSAN TEKNIK INFORMATIKA' would be changed and stored into 'INFORMATIKA'. *Loading.* The data were forwarded into target database or database 'dawer' such as the configuration shown in Fig. 5.
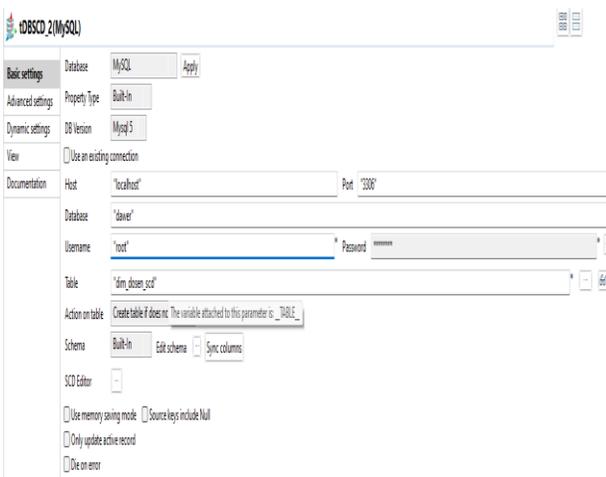
Fig. 5.   Output Database Configuration for Lecturer Dimension Table.

Since dimension table applied SCD concept, further setting was needed on each attribute pursuant to the type of SCD used.
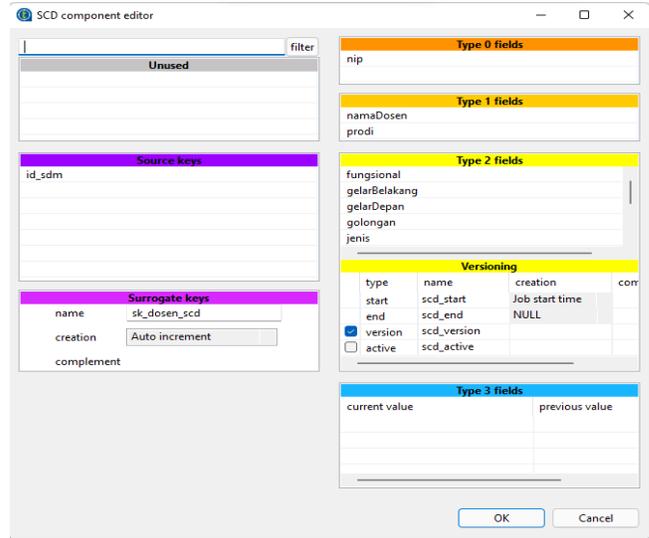
Fig. 6.   SCD Component Settings.

Based on Fig. 6 above, 'id_sdm' was set to be the key attribute on the data source, attribute 'nip' was set to type 0 so that data could not change if the data were input, 'namaDosen' and 'prodi' were set to type 1 thus data change might occur by overwriting previous data. Other attributes were set to type 2 for changes to the other attributes to form a new record in the table. In addition, a key attribute or the surrogate key was needed for each record with setting to auto-increment, and this SCD concept would form additional attributes such as start and end dates record was a valid and higher version of the record equaled the latest, and that would be used in the subsequent data processing.

After dimension table was formed, ETL would be performed for fact table [12]-[13], one of which was service fact table are shown in Fig. 7.
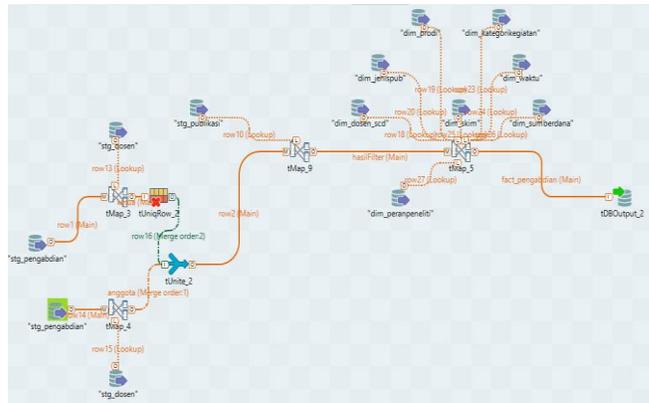
Fig. 7.   ETL Process Forming Fact Table 'Fact_pengabdian'.

3) Extract: This phase started with retrieving and selecting data from the staging area of service data, the staging area of the lecturer, and the staging area of publication. Afterward, data from the staging area selected were connected with a dimension table designed in dimensional modeling as lookup for data in the staging area by applying Left Outer Join as a join model.

4) Tranform: 'Datajudul' and 'biayadisetujui' from the staging area and all data of the surrogate key in the dimension table were selected or retrieved for use in the fact table. In addition, in this phase surrogate key was made for the fact table, and the field biayadisetujui was conditioned in case of null data. They would be changed to 0.

5) Loading: Valid data would then be stored in the table 'fact_pengabdian' in dawer database.

The results of ETL processed stored in data warehouse were then connected with Knowage to present data that were processed into the form of dashboard, OLAP, and report. Below are some outputs of data presented made using research data. The result show in Fig. 8 - 11.

Fig. 8.   Presentation of Data in Dashboard Form for Research Data Publications 1.

Fig. 9.   Presentation of Data in Dashboard Form for Research Data Publications 2.

Fig. 10. Data Presentation in OLAP Form for Lecturer Research Financing Data.

Fig. 11. Presentation of Data in the Form of Reports for Lecturer Research and Publication.

All forms of data presented above can be used by the executives to monitor the lecturer's performance of research and publication as needed.

D. Operate

After Data Warehouse and data presentation had been made, a test was performed using Black Box Testing by developer [14]-[15]. The results of black box testing shown in Table VI.

TABLE VI.   RESULTS OF BLACK BOX TESTING

Test Scenario	Test Case	Output	Test Result
Data completeness	Input or load all data from data source into data warehouse	Number of data stored in data warehouse is equal to number of data in data source	Valid
Data transformation	Transform the data and compare data resulted from transformation with data in data source	Value of data resulted from transformation remains valid pursuant to data source	Valid
Data quality	Input false or invalid data into ETL processing	Data are not input into data warehouse and data value was changed to null, 0, or no remark if the data input were empty.	Valid
	Input the same data into ETL processing	Data do not increase and the value of data remains the same	Valid
	Specifically on dimension table of lecturer, data update was performed: 1. NIP update 2. Name update 3. Title update	Data are updated pursuant to the type of SCD applied 1. Data are not updated except data before update were null 2. Data are updated in the record or in the same key 3. Data are updated in the record or in the new key	Valid
Scalability and performance	Perform the whole ETL process	Not take too long time in processing	Valid
	Run dashboard, OLAP, and report	Aggregated data displayed in data presentation are valid	Invalid

Calculation was performed based on the test results in accordance with the number of valid scenarios. The calculation using Eq. (1).

$$evaluation\ result = \frac{Number\ of\ valid\ data}{Total\ test} x\ 100\% \qquad (1)$$

The results of black box testing above shows that data warehouse made shows valid results of 86%, that it was necessary to further correct the testing part with invalidity.

E. Enhance

In this phase, ETL processing and data presentation were corrected pursuant to the test results on previous phase or based on other suggestion. The corrections and additions include:

1) Correct Aggregation Data Displayed in Data Presentation.

2) Add data of Research and Service Technology Readiness Level (TKT).

3) Present TKT data on the dashboard of research publication and service.

4) Correct data presentation in the dashboard regarding number of researches and services by lecturer.

IV. CONCLUSION

Based on the research, data warehouse designing can be performed using the Data Warehouse Life Cycle and through interviews to examine the needed data. The data will be subject to the ETL process for data integration to be presented into the dashboard, OLAP, and report to facilitate the executives in monitoring the performance of research and service by the lecturer in the Faculty of Engineering UNSOED.

ACKNOWLEDGMENT

REFERENCES

[1] Universitas Jenderal Soedirman, 2021. https://unsoed.ac.id/ (accessed Nov. 25, 2021).

[2] Tentang Fakultas – Fakultas Teknik Unsoed, 2021. http://ft.unsoed.ac.id/tentang-fakultas/ (accessed Nov. 25, 2021).

[3] Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 3 Tahun 2020 Tentang Standar Nasional Pendidikan Tinggi. 2020.

[4] Arti kata eksekutif - Kamus Besar Bahasa Indonesia (KBBI) Online. https://kbbi.web.id/eksekutif (accessed Nov. 29, 2021).

[5] W. H. Inmon, Building the Data Warehouse, 4th ed., vol. 13, no. 401. Canada: Wiley Publishing, Inc., 2005.

[6] R. Kimball and M. Ross, The Data Warehouse Toolkit, 3rd ed. Canada: John Wiley & Sons Inc., 2013.

[7] I. P. A. E. Pratama, Handbook Data Warehouse Teori dan Praktik Berbasiskan Open Source. Bandung: Informatika, 2018.

[8] T. M. Connolly and C. E. Begg, Database Systems : A Practical Approach to Design, Implementation, and Management, 4th ed. USA: Longman Inc, 2005.

[9] A. Supriyatna, "Sistem Analisis Data Mahasiswa Menggunakan Aplikasi Online Analytical Processing (OLAP) Data Warehouse," J. Pilar Nusa Mandiri, vol. 12, no. 1, pp. 62–71, Mar. 2016, Accessed: Mar. 18, 2022. [Online]. Available: http://ejournal.nusamandiri.ac.id/index.php/pilar/ article/view/260/230.

[10] M. Demarest, "Data Warehouse Prototyping: Reducing Risk, Securing Commitment and Improving Project Governance," Jan. 25, 2008. http://dssresources.com/papers/features/demarest08/demarest01252008. html (accessed Mar. 18, 2022).

[11] M. T. Rumondor and D. C. Irawati, "Designing an Online Analytical Processing (OLAP) for Project Feasibility Study in Siau Tagulandang Biaro District," Int. J. Innov. Sci. Res. Technol., vol. 4, no. 12, pp. 620–625, Dec. 2019, Accessed: Mar. 18, 2022. [Online]. Available: www.ijisrt.com620.

[12] Imelda, "BUSINESS INTELLIGENCE," Maj. Ilm. UNIKOM, vol. 11, no. 1, pp. 111–122, 2013, Accessed: Nov. 29, 2021. [Online]. Available: https://jurnal.unikom.ac.id/_s/data/jurnal/volume-11-1/09-miu-11-1-imelda.pdf/pdf/09-miu-11-1-imelda.pdf.

[13] R. Akbar, A. Soniawan, J. Adrian, R. Azim, and A. Zikri, "Implementasi Business Intelligence untuk Menganalisis Data Persalinan Anak di Klinik Ani Padang dengan Menggunakan Aplikasi Tableau Public | Zikri | Jurnal Online Informatika," J. Online Inform., vol. 2, no. 1, pp. 20–24, Jun. 2017, Accessed: Mar. 18, 2022. [Online]. Available: http://join.if.uinsgd.ac.id/index.php/join/article/view/v2i14/52.

[14] E. P. Putra and H. Juwitasary, "Evaluasi Datawareouse dan Tools untuk meningkatkan Efesiensi dan Efektifitas (Case Study PT XYZ.)."

[15] N. Leite, I. Pedrosa, and J. Bernardino, "Open Source Business Intelligence Platforms' Assessment using OSSpal Methodology," in Proceedings of the 15th International Joint Conference on e-Business and Telecommunications (ICETE 2018), 2018, pp. 190–196, Accessed: Nov. 22, 2021. [Online]. Available: https://www.scitepress.org/Papers/2018/69101/69101.pdf.

Toward an Ontological Cyberattack Framework to Secure Smart Cities with Machine Learning Support

Ola Malkawi[1], Nadim Obaid[2]
Computer Science Department
University of Jordan
Amman, Jordan

Wesam Almobaideen[3]
Electrical Engineering and Computing Sciences
Rochester Institute of Technology
Dubai, UAE

Abstract—With the emergence and the movement toward the Internet of Things (IoT), one of the most significant applications that have gained a great deal of concern is smart cities. In smart cities, IoT is leveraged to manage life and services within a minimal, or even no, human intervention. IoT paradigm has created opportunities for a wide variety of cyberattacks to threaten systems and users. Many challenges have been faced in order to encounter IoT cyberattacks, such as the diversity of attacks and the frequent appearance of new attacks. This raises the need for a general and uniform representation of cyberattacks. Ontology proposed in this paper can be used to develop a generalized framework, and to provide a comprehensive study of potential cyberattacks in a smart city system. Ontology can serve in building this intended general framework by developing a description and a knowledge base for cyberattacks as a set of concepts and relation between them. In this article we have proposed an ontology to describe cyberattacks, we have identified the benefits of such ontology, and discussed a case study to show how we can we utilize the proposed ontology to implement a simple intrusion detection system with the assistance of Machine Learning (ML). The ontology is implemented using protégé ontology editor and framework, WEKA is utilized as well to construct the inference rules of the proposed ontology. Results show that intrusion detection system developed using the ontology has shown a good performance in revealing the occurrence of different cyber-attacks, accuracy has reached 97% in detecting cyber-attacks in a smart city system.

Keywords—*Cyberattack; Internet of Things (IoT); ontology; machine learning; intrusion detection system*

I.   INTRODUCTION

The Internet of Things (IoT), defines the large number of devices that can be connected to the internet and perform different types of work. Different devices and sensors can provide our life with digital intelligence, which can serve peoples' needs with minimum or zero human intervention [1]. IoT devices are connected to each other as well as to the internet via a computer network. It is worth mentioning that wireless networks dominate the connectivity between IoT devices which may increases the opportunity for more potential attacks be launched [2].

Features of IoT facilitate the automation of a wide variety of applications and systems, such as health care, homes, traffic lights, and electricity grids to get a smart healthcare, smart home, smart traffic, and smart grids, respectively, as well as many other services and applications. Hence, the majority of recommended services of a city has been automated which creates the concept of a smart city [3]. The concept of a smart city has gained a large concern from governments and business agents as it plays a vital role in the progress and development of the new understanding of civilization in modern countries.

However, security is one of the most prominent challenges when we deal with smart life aspects such as smart cities. This is because developing intrusion detection or prevention systems to secure smart systems is not an easy task, especially with the continuous emergence of new attacks. This raises the need for a uniformed understanding of cyberattacks [4, 5]. The main goal of this uniformed understanding is to develop suitable protection tools for a certain category of attacks, which, at the same time, can protect the system against potential attacks which could appear in the future. Developing an ontology for cyberattacks can provide this formal and uniformed representation, which may provide a general base for a certain category of attacks based on predefined criteria.

Developing an ontology for cyberattacks can also help to understand needed characteristics of a certain system before selecting the protection method. Strictly speaking, different organizations may be interested in different security concepts based on the organizational type and function. A newspaper information system, for example, may consider integrity and authentication of the published news to be of great concern whereas, a healthcare system highly considers confidentiality and privacy of the exchanged patients' information to be essential [6]. By defining the security needs, potential cyberattacks and their impact on the system can be defined and characterized which could play a major role in developing and selecting the suitable protection system.

The contribution of this paper can be summarized with the following points:

1) Proposing an ontology for cyberattacks of smart cities in the context of IoT.

2) Defining the benefits of developing an ontology for cyberattacks from deferent perspectives.

3) Presenting a formal representation and implementation of the proposed ontology using Description Logic and protégé software to conduct reasoning.

4) Using ML as a tool to define the inference rules for the proposed ontology.

5) Using the proposed ontology to pick up the features needed to apply ML.

6) Integrating ML model with the implemented ontology to develop a simple knowledge base for certain attacks.

The rest of this paper is organized as follows. In Section II we present the most important related work. Section III illustrate the followed methodology in conducting this research. A case study is discussed in Section IV which is related to using machine learning to Secure Smart City which also include the conducted experiment and discussed results. We conclude the paper with Section V.

II. RELATED WORK

In cyber-security research area, developing ontologies is not new. A number of approaches have investigated ontology to develop or design a security framework [4]. In this section, we summarize some state-of-the-art research works focusing on the development of ontologies in the context of security and privacy.

Proposed ontologies of cyber-security can be categorized as follows: (1) ontologies that considered information security, (2) ontologies that considered security in IoT, and (3) ontologies that considered security in smart city. Fig. 1 presents the hierarchy of ontologies of cyber-security in state-of-the-art.

In category (3), authors of [7] and [8] have proposed ECA and OBPP ontologies, respectively. Both ontologies have considered only privacy issues with the use of cloud computing in smart cities. The other security requirements such as availability, integrity and confidentiality are not considered. In [9], authors have focused on setting guidelines to develop a secure and safe smart city system based on using ontologies. Nevertheless, no ontology is provided in the research paper. In [10] authors have proposed an ontology for cyberattacks in a smart city system. The ontology concentrated on securing smart city applications rather than cyber-attacks. Moreover, authors provided a number of use cases with mapping each use case with the proposed ontology with no inference rules to get a benefit from the proposed ontology.
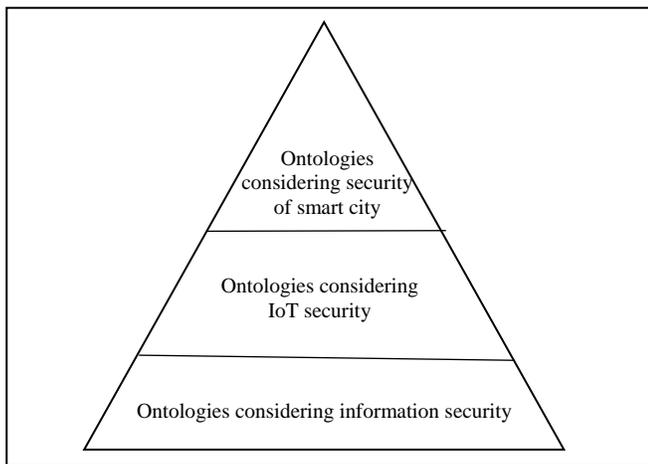
Fig. 1. Hierarchy of Ontologies of Cyber-Security in State-of-the-Art.

In category (2), authors in [11] proposed an ontology to secure an IoT system, the ontology is designed based on analyzing IoT system's vulnerabilities. This work has focused on power-IoT cloud systems which are limited to power issues. Inference rules are set based on vulnerability analysis.

In [12], an ontological analysis has been proposed to enhance security services in IoT systems. The ontology has focused on the system vulnerabilities, potential threats and security needs, it does not elaborate cyber-attacks and their specifications and symptoms.

In [4], the ontology's objective is to create a unified representation for heterogeneous data generated by IoT devices, this work has focused on cyber-attacks and their properties, however, it considers only the general aspects of cyber-attacks. In our proposed ontology, we have concentrated on the cyber-attack properties and their detailed impact on the performance of the network.

In [13] an ontology is proposed to be utilized for higher security improvement in terms of the heterogeneity in the layered cloud platform, the ontological design has focused on the IoT environment and assets such as security devices rather than attack details.

Finally, in category (1), we present a number of these works to illustrate its main structure and focus. In [14] and [15], two cybersecurity ontologies are built by expanding existing ontologies. The main focus of the proposed ontology in [14] is the environment rather than the system itself, and proposed ontology is very general and the main goal is to improve cybersecurity awareness to make suitable decision by providing safe operations rather than detection malicious behavior. In [15], the cybersecurity ontology has concentrated on finding qualification metrics to assess how much a certain system is secure.

In [16] and [17], the main concern of the proposed ontologies is to create an organized schema for cyber information. The goal of using these ontologies is data analysis. While authors of [16] have focused on virus threats and IP and DNS problems, the main focus in [17] is system assets and how to protect these assets.

In [18], authors concentrated in the proposed ontology on data sources and users and potential threats based on these two elements. Rules of ontology are set based on cybersecurity standards and concepts.

Authors of [19] proposed Unified Cybersecurity Ontology (UCO), the main concern of UCO is to identify sources of attacks. Inference is related to derive possible sources of attacks. However, the proposed ontology cannot help to improve security when mapped to IoT. In [20], a cybersecurity ontology is proposed to find guidelines for security measures to protect critical infrastructure, the main focus in this ontology is system assets. Both ontologies in [19] and [20] are customized for specific platforms and to protect simple software. Moreover, most of such ontologies are not suitable on IoT environment according to its specifications and limitation.

Most existing research works are either environment-centric, assets centric, or threats-centric which is opposed to our ontology which is attack-centric scheme. This motivates us to develop Cyber Attack Ontology (CAO) for IoT-based smart city. Moreover, the process of setting inference rules in the previous research works depends on either the analysis of the systems' vulnerabilities and threats, or it depends on user-defined rules. In our approach, we have proposed the use of ML to set inference rules based on the superiority of artificial intelligence and ML in the security during the last years [21][22][23].

III. METHODOLOGY

In this section, we present and discuss the proposed ontology, we then show the steps of utilizing the ontology to identify cyber-attacks. In order to make accurate reasoning to identify cyber-attacks, precise rules must be defined. However, and because there is no scientific base to set the inference rules for our ontology, like those in other scientific fields, we will use ML to set up the rules, definitions of attacks and other concepts included in the proposed ontology, a detailed discussion is represented in the next sections.

A. Cyber-attack Ontology (CAO)

In this section we discuss the proposed ontology which is shown in Appendix A. The ontology graph in Appendix A was depicted using protégé OWL [24]. This plugin represents a visual notation for OWL ontologies and a graphical view for the ontology's classes and relations which are joined together to shape a directed graph layout for the ontology. The proposed ontology includes three main entities, which are adversary, system and cyberattack. The adversary is the person who designs and develops a malware or acts in a malicious behavior to launch a cyberattack. There are numerous types of attackers and there are multiple goals for an attacker to launch an attack. A system is described by a number of concepts such as its functionality, components, security needs, performance aspects and system's vulnerabilities. Fig. 2 shows the classes of the proposed ontology after we have implemented this ontology using protégé OWL. Adversary, Cyber-attack and IoT System present the main classes. They are subclasses of the general built-in class Thing initiated by protégé.

Fig. 2. Classes of Cyber-Attack Ontology in Protégé Owl.

A cyber-attack which is the major entity and the subject of this research connects an adversary with the target system of the attack from multiple perspectives. Examples are that the cyber-attack violates one or more of the security principles of the system, the cyber-attack may disrupt the system's functionality, the cyber-attack may degrade the system's performance (and may not) by exploiting the system's vulnerabilities. The cyber-attack also is characterized by the network type which is corrupted with it, the channel through which it is being launched, the protocol or protocols that have vulnerabilities through which the attack could be launched, the network layer for that attack, and the countermeasure adopted against the attack.

Concepts of CAO are shown in Appendix A along with other considered concepts and features. This ontology could be thought of as a base for a wider ontology, more concepts and relations could be added and investigated.

From the discussed ontology, we notice that we can get the following suggested benefits:

1) Ontology helps to develop a taxonomy for cyber-attacks, any entity in the discussed ontology can be picked up as a classification criterion, for example by the selection of channel entity, cyber-attacks could be classified as software-based attacks, hardware-based attacks and network attacks. If we select access level entity, cyber-attacks could be classified as either passive or active attacks, and so on.

2) Ontology provides a formal and unified description and understanding for cyber-attacks and security needs for IoT environment, especially with the heterogenous nature of IoT devices and IoT systems.

3) Ontology can be utilized to define security needs and the protection method for a certain system, for example, for a smart tourism system, we care about the integrity and availability rather than confidentiality, as tourism information tend to be public [25]. By defining our security needs, we can utilize the ontology by navigating ontology structure from one concept to another to define all related needs, costs and components in order to develop a suitable protection system.

4) From an ontology, we can extract features to apply ML which has got a considerable attention during the last decades for a wide variety of applications and especially in the security field.

5) Ontology enables automated reasoning about cyber-attacks, if we build a strong ontology, reasoning could be used to develop semantic graph database for cyber-attacks and all related characteristics such as effect, cost, and suitable countermeasures.

6) Ontology is easy to be extended or changed because we can add concepts and relationships. So, the proposed ontology can evolve with the emergence of new concepts or when discovering any wrong facts with no impact on the existing systems.

In this work, the intended use of the proposed ontology is to navigate its concepts starting from a pre-specified system. With the target is to construct a model to identify cyber-

attacks of that system using ML. We have conducted a case study to build the model, the case study details along with the model designed are discussed in the next sections.

B. CAO Implementation

In this section, we present the details of Cyber Attack Ontology (CAO) including main classes, sub-classes, properties and other elements of ontology. We have implemented the proposed CAO using protégé tool. Fig. 2 presents the main classes of CAO. We have implemented main classes and sub-classes. Then, we have defined properties of our ontology, which are divided into two types: object properties and data properties. Fig. 3 and Fig. 4 present object properties and data properties, respectively. Fig. 3 shows also a representation of the hierarchy of object property "exploits" which has the domain "cyber-attack" and range "vulnerability".

Data properties relates a class to an attribute data such as "integer"," float", "string", ...etc. Fig. 4 presents a number of data properties in CAO such as traffic_sent which indicates the total number of packets sent through the network during a period of time. Traffic_sent has the domain traffic and the range integer.

Rules of CAO are not defined at this phase as they will be derived at the last step where ML will be utilized to set CAO rules.

Finally, we have shown individuals of the ontology, these individuals represent fundamental components of the ontology and they include concrete instances of ontology. In CAO, for example, sinkhole attack is an instance of a cyber-attack class. Sinkhole attack can be described by multiple properties. In this work, a cyber-attack is described by the network performance resulted by the attack occurrence. The performance is presented by network traffic measurements. For example, sinkhole attack can be described with a certain PDR, delay, overhead, etc. Rules to organize and control these properties will be set using ML at section 4.5.

Fig. 5 presents data properties assertions of class traffic. Network traffic is extracted for a period of time and analyzed, performance metrics values such as packet delivery ratio, number of sent packets, and power consumption are calculated, and the resulted values of these metrics are added as assertions for data properties shown in Fig. 4.

In CAO we have defined each cyber-attacks with a certain traffic specification, or in other words, a traffic with particular values of the aforementioned performance metrics is considered as an attack.

C. Description Logic for CAO

Description logic (DL) is one of the formal languages used for knowledge representation. DL is used in artificial intelligence applications to conduct reasoning from related concepts. In CAO, reasoning is needed to determine if there is an attack or not. In this section we present a number of concepts and relations of our proposed ontology, CAO, using DL. Fig. 6 presents DL for CAO.
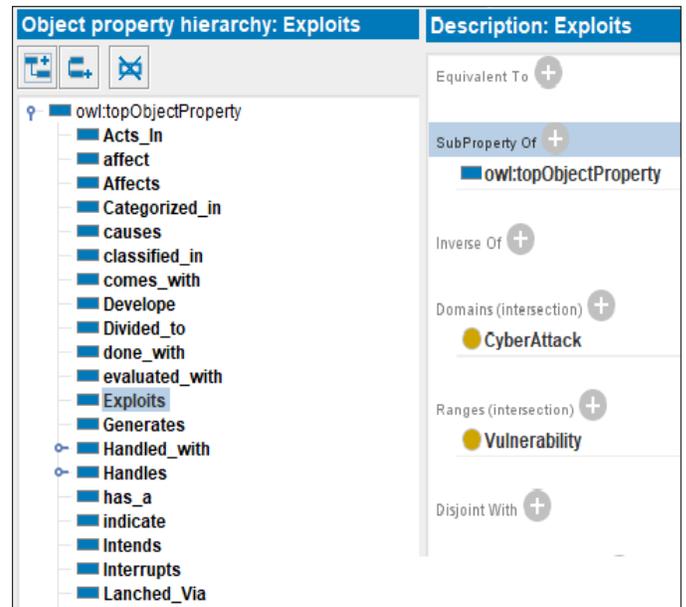
Fig. 3. Object Properties of Cyber-Attack Ontology and Object Hierarchy "Exploits".
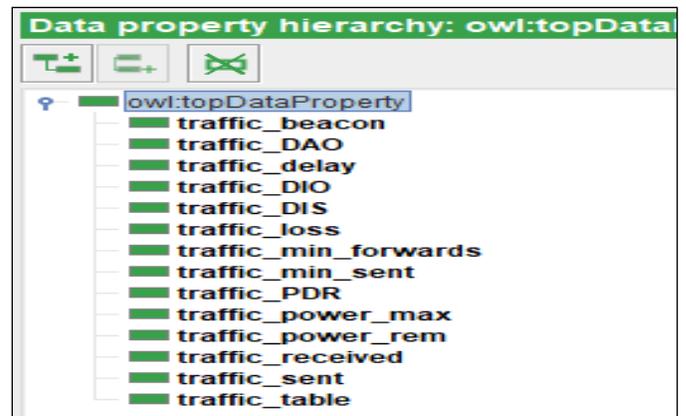
Fig. 4. Data Properties of Cyber-Attack Ontology.

Fig. 5. Data Property Assertions of a Traffic Instance (Traffic_1).

Fig. 6. Description Logic Representing CAO.

D. Ontology to Secure an IoT System

The main utility for cyber-attacks ontology in this work is to design and implement an intrusion detection system to secure a smart city. The first phase of designing any protection mechanism is to study the system to be secured and to highlight the main issues, risks and security needs. We will discuss how to go through these processes using the proposed CAO.

Before proceeding with designing and implementing a security mechanism for a certain system, we have to analyze that system and understand its needs precisely. As we mentioned before, security needs vary from one system to another depending on system's structures and goals. For instance, the most important security needs for a banking system are privacy, confidentiality and integrity, however, a newspaper or tourism systems consider integrity and availability as the most significant requirements to be maintained rather than privacy or confidentiality [25].

Thus, the first step is to determine the most important security needs for the targeted system. Then other elements can be defined in the same way, we can use and track the ontology for a specific system to determine the following proposed elements:

1) The most important security essentials needed to be maintained for the concerned system.

2) The potential domain of a cyber-attack which must be specified carefully. For instance, if computer networks are a potential domain. We have to specify network layer, network protocol and network type whether wired or wireless.

3) Vulnerabilities of the system.

4) The category of the most potential or damaging cyber-attacks that we need to handle.

5) System's components that are exposed to cyber-attacks must be specified, this component could be a computer network, file server, mail server, or any other component.

6) The functionality of each component specified in point number 5 which can be interrupted by a cyber-attack must be also determined.

7) If the performance of the system is of a considerable importance, the performance metrics should be identified.

8) If specifications of the adversary are of a significant importance, then attacker's type and benefit of a cyber-attack

are determined, the tactic of the adversary whether designing a malware or acting in a certain behavior are also specified.

9) Finally, and based on the previous points, we can decide what is the most appropriate countermeasure(s) to be designed to secure the concerned system. These may include intrusion detection system, intrusion prevention system or a mitigation method.

IV. CASE STUDY: ONTOLOGY TO SECURE SMART CITY USING MACHINE LEARNING

Wireless sensor networks (WSNs) play a significant role in smart city services. This is due to the fact that sensors comprise most of monitoring and automation systems which are the backbone of smart cities [26]. Actually, sensor networks are fundamental component in smart grids, home automation systems, traffic systems, health care applications, power system monitoring and many other smart city services. However, sensors are limited and constrained devices in terms of capabilities and resources such as power, processing and storage. This makes it harder to secure a sensor network using the common and traditional security means such as cryptography. More efficient and low-cost methods need to be investigated to secure sensor networks. ML is considered as one of the most suitable candidates in this context. ML can be employed to design an intrusion detection method by training the system to discriminate between normal traffic and traffic of a network under attack [27] [28].

To apply ML, we must have a suitable dataset, in the following subsections, we present the steps by which we use the proposed ontology to create a sample dataset with suitable features. We have opted to extract the dataset from the original network traffic, this traffic is generated from normal operations of the network. Thus, we do not add any additional packets or control message to get features of the dataset, which considers limited power of the sensors, so, we do not add any additional load onto the sensor nodes.

A. Ontology Navigation to Identify Main Entities

We have used CAO to make a decision about the features of the dataset, we have explored the ontology and specified the most significant entities. Phases of exploring the ontology that have been discussed in the previous section are applied to the selected case study. These phases are summarized in the following points:

1) The major security principle to be maintained is the availability of the sensor network.

2) Because network is the most potential domain to launch most cyber-attacks of WSNs. We have to specify network related elements including layer and network protocol through which an attack is launched. We have selected RPL routing protocol for our case study which belongs to network layer. Motives behind the selection of RPL protocol can be summarized in its convenience with the requirements of smart and its properties of self-configuration, self-restoration, and the ability to meet power consumption constraints.

3) The main vulnerabilities of WSNs in the light of sensors limitations and RPL protocol can be summarized as: lack of infrastructure, lightweight protocols which are not supported with security mechanisms, constraint devices where cryptography is not efficient, limited physical security, dynamic topologies, unreliable links, multi-hop transmission paradigm which acts as a helping factor to transmit and spread malicious messages, and finally, the distributed problem handling, which means that a malicious node can select the action of not solving a problem which can simply cause the entire network to be malfunctioning, such as local repair attack[29].

4) The most well-known attacks for sensor network regarding availability are specified. These attacks include sinkhole attack, wormhole attack, sybil attack, rank attack, flooding attack, copycat attack. These types of attacks are categorized within denial of service (DoS) attacks.

5) The system component to be secured is a sensor network which is considered as a vital part of smart city system.

6) The functionality of the sensor network which could be interrupted is the process of data transmission from all sensors to a central point. The role of each sensor is to monitor and record certain data in some location and send the sensed data to a central point (sink). We can explain the functionality as that the sensed data must be received by the sink within a certain period of time and entailed with the original sender.

7) The performance of sensor network is of a significant importance. For example, it is recommended that the transmitted packets are received within an acceptable period of time, which is referenced as delay. Moreover, all packets, or at least a satisfying portion, must be received by the intended receiver, which is commonly defined as packet delivery ratio. Both delay and packet delivery ratio can be considered as performance metrics for sensor networks. The next section describes how we have analyzed selected attacks to address the main performance metrics affected by each attack.

8) Characteristics of the adversary such as its type, benefit, or tactic are not customized in this work, the involved attacks include all these variations.

9) The countermeasure to be designed and implemented is an intrusion detection system that employs ML. The features of the dataset will be selected based on the performance of the sensor network, performance indicators or metrics will be extracted from the network traffic.

B. Impact of Studied Cyber-attacks on Performance

The targeted security principle of this work is availability, which means that the service provided by the WSN must be available when needed and with satisfying quality. Thus, we aim to detect attacks that cause the WSN to become unavailable or degrade its performance. Therefore, performance must be defined using suitable performance metrics. We have explored the ontology for the selected cyber-attacks to address performance metrics that could be affected for each attack. Performance metrics defined in this section will be considered as the features of the generated dataset:

Rank Attack: this attack interrupts the balance of routing paths distribution in a WSN uses RPL. The first result is the high congestion and interference within the attacker's zone, which causes packet loss, increases end to end delay, and decreases throughput and packet delivery ratio. Moreover, the un-optimal paths will be created and used which may increase power consumption. Furthermore, and as the rank attacker becomes the preferred router for many nodes, the length of routing table will be increased unusually in the attacker and its neighbor nodes. Therefore, the length of routing table could be considered as a feature in the generated dataset.

Sinkhole attack: this attack is similar to rank attack in its effect, however, the attacker can attract more nodes than rank attacker because it claims that it has the minimum rank. Nonetheless, it is easier to be detected. Therefore, the performance metrics affected are packet loss, end to end delay, throughput, packet delivery ratio, and power consumption.

Flooding attack: this attack is executed by broadcasting a large amount of control messages. Hence, we expect a noticeable increase in control packets. Furthermore, the generated traffic will affect end to end delay as a result of increasing interference. Nodes will spend more time in replying to the deceptive traffic which affects throughput and power consumption, Moreover, the resulted interference can increase packet loss which affects delivery ratio. Finally, the deceptive control packets may force nodes to set protocol related metrics, such as the frequency of sending hello packets, unusually.

Copycat attack: as a result of receiving old versions of control packets during this attack, nodes may adjust protocol metrics illogically, such as time interval between hello packets and routing information. Moreover, frequent and unnecessary control packet transmission can exhaust power resources.

Sybil attack: by receiving packets with fake sender's information, nodes build a faulty routing table, For example a node could be inserted in the routing table as a neighbor while it is actually far away or even it does not exist, or a node may be recorded as an optimal router while it presents a high-cost router. The result of such cases is a high packet loss and a higher cost routing paths which leads to a lower packet delivery ratio and higher power consumption and end to end delay.

Wormhole attack: the effect of wormhole attack varies based on the target of the adversary from the established tunnel. For example, if the target is to drop packets this will affect packet delivery ratio. In general, we expect longer and un-optimized paths to be created. So, the foreseeable impact will be relevant to power consumption, end to end delay and packet delivery ratio. Table I, lists the considered attacks as well as the performance metrics affected by each attack.

So far, we have specified targeted security principle that we want to preserve, which is availability, we have then defined attack category to be considered which is (DoS). Based on this category, we have specified six types of attacks from this category for our case study. Then, more navigation through the ontology has ultimately led us to the fact that these attacks affect the performance of the network, and the concerned performance metrics were defined. This has been illustrated in Table I. To set up the rules of the ontology, the most significant step is to determine what is the threshold for each metric that can be considered to decide if there is an attack or not. For instance, what is the value of PDR that

represents the lower limit of accepted PDR, and below it we decide that there is an attack? Since it is not easy to guess or evaluate these values, so we will utilize ML to set threshold values and to set rules of the ontology, and developing a model for attack detection in WSNs.

C. Dataset and Cyber Attacks

To set the rules of the proposed ontology, we have utilized ML. This is because there is no scientific rule or base for performance metrics by which we can detect cyber-attack occurrence based on performance. Thus, we applied supervised learning on labeled dataset that include the investigated attacks. The objective of using ML is not to construct the intrusion detection system (IDS), it comes as a complementary step to denote the importance of ontology in defining system needs and potential attacks, which facilitate the process of developing IDSs as well as other protection systems. To create the dataset, we have used Cooja emulator to establish a WSN, then, we have run multiple simulations with different configurations to generate both benign and malicious traffic.

TABLE I. PERFORMANCE METRICS AFFECTED BY EACH ATTACK

	Delay	Control Packets	Lost Packets	Received Packets	Sent Packets	Power Consumption	Packet Delivery Ratio	Routing Table Length	Protocol Settings
Rank attack	✓		✓	✓	✓	✓	✓	✓	
Sinkhole attack	✓		✓	✓	✓	✓	✓	✓	
Flooding attack	✓	✓	✓	✓	✓	✓	✓		✓
Copycat attack	✓	✓				✓			✓
Sybil attack	✓		✓	✓		✓	✓		✓
Wormhole attack	✓		✓	✓		✓	✓		

TABLE II. SELECTED PERFORMANCE METRICS AND DERIVED FEATURES

Performance Metric	Features derived and description
Delay	Average end to end delay (E2E)
Control Packets	Number of control packets transmitted through the network (Overhead). Presented as number of (DAO),(DIO) and (DIS) packets in RPL.
Lost Packets	The number of lost packets defined as (total sent packets – total received packets)(Lost)
Received Packets	The total number of packets received by the sink (Received), maximum number of packets received by a node (Max_received)
Sent Packets	The total number of packets sent by all nodes (Sent), minimum number of packets sent by each node (Min_sent), maximum number of packets sent by each node (Max_sent), minimum number of packets forwarded by each node (Min_forwarded), maximum number of packets forwarded by each node (Max_forwarded), total forwarding operations within the network (Forwarded).
Power Consumption	Total remaining power (Rem_power), maximum power consumption (Max_power), total power consumed by all nodes of the network (Total_power), average voltage of all sensors (Voltage)
Packet Delivery Ratio	The percentage of total received packets by the sink to the total packets sent by all nodes (PDR)
Routing Table	The maximum length of routing tables in all node (Max_length)
Protocol Settings	We have considered the beacon interval (Beacon), which is a varied period of time defines the frequency by which control packets are sent continuously by nodes in WSN.
Delay	Average end to end delay (E2E)

While benign traffic is generated by simulating a network with the original protocol, malicious traffic is generated by implementing a number of specified cyber-attacks. Then, the malicious copy of routing protocol is implemented and used to launch these attacks. The resulted traffic is collected and analyzed. Finally, features presented in Table I are extracted to form the final dataset. It is worth mentioning that the variations some features are considered instead of the mere features shown in Table I only. For instance, variations of sent packets are considered, such as minimum sent packets, and maximum sent packets for sent feature. Maximum power consumption, total power consumption and remaining power are considered instead of taking only average power consumption. Moreover, types of control packets are considered separately instead of counting the total control packets. We have included DIS, DIO, DAO control packets which are the main control packets in RPL protocol [30]. We have also considered forwarded packets as a special case of sent packets, which indicates packets received from neighbor nodes and sent again toward the intended destination. Table II presents selected performance metrics and features derived (between brackets).

D. Simulation Environment

Through this section, we present simulation that we have conducted and its related environment and tools. We have selected Cooja simulator to create and configure WSNs in a smart city. There are many reasons behind selecting Cooja simulator. Cooja simulator is designed specifically for WSNs, it implements Contiki operating system. Strictly speaking, a simulated sensor in Cooja presents an actual compiled Contiki system [31]. Moreover, Contiki, which is the operating system of sensors, is also the best candidate for IoT devices in a smart city. That is due to Contiki's design which is developed specifically for memory constrained devices with the consideration of low-power IoT. Existing employment of Contiki involves street lighting systems, radiation monitoring

systems, sound monitoring systems and alarm systems [31]. As a result, Cooja simulator with Contiki operating system is the best choice to simulate the heterogeneous WSNs in a smart city system. According to the aforementioned points, resulted traffic will be similar to a great extent with a real traffic generated from real WSN rather than being just a traffic generated from a simulation.

Table III shows simulation environment and configuration parameters selected to create the dataset. Each record of the generated dataset represents 110 seconds traffic of 50 nodes deployed within an area of 350 X 350 m^2. At each record, we have diversified network configuration, such as the distribution of nodes, nodes to sink allocation and network topology. Part of the generated dataset instances represents benign network behavior, while the remaining dataset instances represent a traffic with the discussed attacks launched. After the simulation has been conducted for the 180 instances, performance has been measured in terms of the aforementioned performance metrics which are considered as dataset features to be inputs to the classification algorithm. Table IV presents a part of the generated dataset.

TABLE III. SIMULATION ENVIRONMENT AND PARAMETERS

Simulation Parameter	Value
Network Size (Number of Sensor Nodes)	50
Routing Protocol	RPL
Transport Layer Protocol	UDP
MAC Protocol	CSMA
Sensor Type	Sky Mote
Terrain Area	350 X 350 m^2
Number of Attacking Nodes	2
Simulation Duration	110 Seconds

TABLE IV. GENERATED DATASET

Overhead	E2E	Sent	Received	PDR	Lost	Table	Max forwarded	Min forwarded	Max received	Min Sent	Max sent	Forwarded	Max power	Total power	Beacon	Rem power	Voltage	Class
268	492.65	20	15	0.75	5	6	1	0	5	1	1	22	64360	964700	121.1	5040	252	normal
355	677.882	19	14	0.73	5	12	1	0	11	0	1	35	67447	799530	115.4	4284	252	normal
402	840.211	20	15	0.75	5	11	1	0	11	1	1	49	64080	930343	113.6	4788	252	normal
395	881.947	19	15	0.78	4	9	1	0	9	0	1	44	67897	949670	113.6	4788	252	normal
346	696	20	17	0.85	3	14	1	0	14	1	1	40	60265	837921	123.66	4536	252	attack
365	931.556	19	15	0.78	4	10	1	1	10	0	1	46	64914	895441	116.33	4536	252	attack
395	881.947	19	15	0.78	4	9	1	0	9	0	1	44	67897	949670	113.63	4788	252	attack
360	741.944	19	14	0.73	5	12	1	0	11	0	1	35	67441	874014	112.66	4536	252	normal
462	933.417	20	17	0.85	3	19	1	0	19	1	1	55	65172	637302	115.91	3024	252	normal
508	1199.23	20	16	0.8	4	18	1	0	17	1	1	75	69111	684722	103.0	3276	252	attack
426	926.133	20	15	0.75	5	15	1	0	14	1	1	50	66941	832173	92	3780	252	attack
307	540.8	21	15	0.714	6	7	1	1	7	1	2	31	66780	1059204	117.8	4723	248.75	normal

E. Machine Learning

We have utilized ML as a tool to set up inference rules. We have used Weka 3.8.4 environment to apply ML on the dataset discussed in Section A [32]. Decision tree (J48) classifier has been used as it is easy to extract the model in the form of explicit rules by this classifier. Decision tree classifier is trained and tested based on 10-fold cross validation technique.

V. RESULTS AND DISCUSSIONS

By applying decision tree classifier on the discussed dataset, we have obtained 95% accuracy to classify a network traffic to either attack or normal traffic. We also have an accuracy of 81% to specify the name of the attack. The resulted rules then transferred to the ontology to carry out reasoning. Inference rules are shown in the table of Appendix B. Each rule in the table is represented in one row where the first column can be understood as a logical if statement and the second column represents the part of then statement.

A. Safety Factor

According to the previous accuracy values obtained, 95% for binary classification and 80% for attack type classification, 5% and 20% of cases are wrongly classified in both binary classification and attack type classification, respectively. However, these values cannot be considered as indicators for how much the system is secure. For instance, if the detection method wrongly classifies normal traffic as malicious, this can only add an additional cost, which is presented in taking a countermeasure reaction, but the system is still safe.

On the other hand, when the detection method classifies an attack traffic as normal, we can say that the system is not safe, because this means that we allow malicious behavior to proceed without being detected.

Fig. 7 presents confusion matrices for both binary and attack classifications. In binary classification, Fig. 7 shows that six cases out of 180 are malicious traffic that classified as normal, so, 174 cases are either correctly classified cases or they are normal traffic classified as attacks, which do not degrade the safety of a system. This means that the safety of the system in binary classification is 97%. In attack type classification, only four instances of attacks are classified as normal, which means that the safety of attack classification is 97.7%.

B. Use Cases

In this section, we introduce a simple use case to illustrate the validity of CAO. The ontology is implemented and rules obtained from ML model are added. Thus, we are ready to detect the occurrences of cyber-attacks using the proposed ontology. Different cyber-attacks are launched, traffic is analyzed to extract the defined performance metrics, then, calculated values are added and reasoning is conducted. Fig. 8 shows the description of copycat attack which has been defined as a network traffic using rules obtained from applying machine learning and in terms of performance metrics specified. Fig. 9 shows the definition of copycat attack in description logic.

We have simulated copycat attack, collected the resulted network traffic, analyzed the traffic to find out the values of performance metrics. We then defined the instance "traffic_1" of class traffic which have been assigned the obtained performance metrics as its data properties, and fed to the implemented ontology CAO. After reasoning is applied based on these inputs, an inference is done yielding that there is a copycat attack. The degree of accuracy of that inference is 95% in that there is an attack and 80% is that it is a copycat attack. The safety of using the proposed ontology along with the ML obtained rules is 97.7%.


```
      a    b    <-- classified as
     28    2  |   a = normal
      6  144  |   b = attack
```

Confusion Matrix of Binary Classification
Safety = 97%

```
      a    b    c    d    e    f    <-- classified as
     27    0    2    0    0    1  |   a = normal
      1   22    1    0    1    5  |   b = copycat
      3    0   24    0    1    2  |   c = Rank
      0    0    0   30    0    0  |   d = DISFlooding
      0    1    0    0   25    4  |   e = Sinkhole
      0    3    3    1    7   16  |   f = sybil
```

Confusion Matrix of Attack Type Classification
Safety = 98%
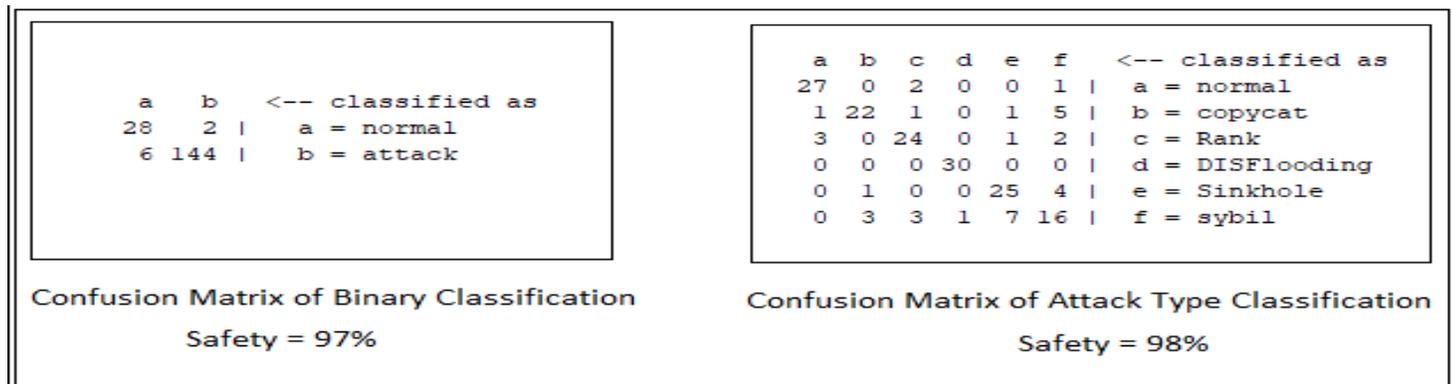
Fig. 7.   Confusion Matrices of Both Binary and Attack Classification.
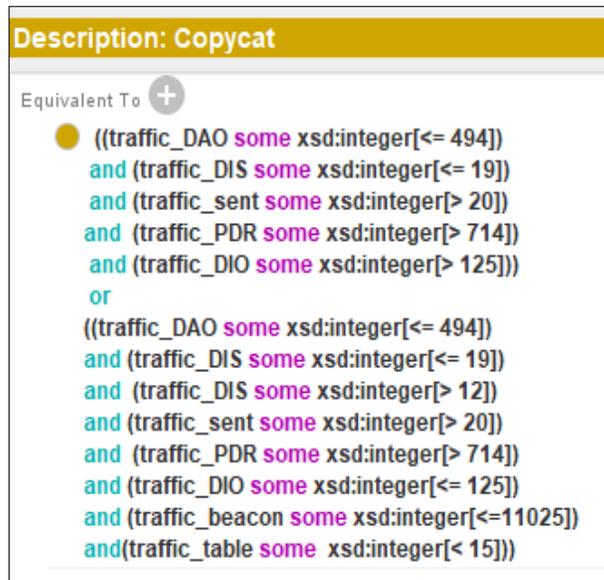
Fig. 8.   Description of Copycat Attack in Terms of Traffic Specifications and Rule Extracted from Protégé Tool.
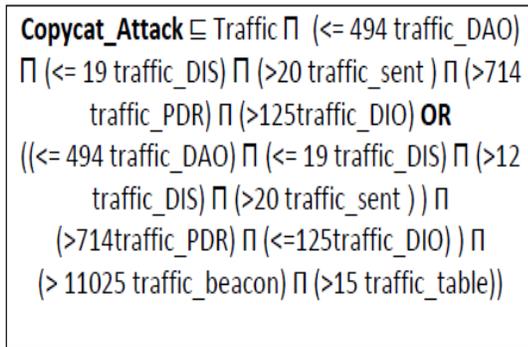
Fig. 9.   Definition of Copycat Attack in DL.

C. Comparison with Existing Works

In this section, we provide a comparison between CAO and existing research works investigating ontology and machine learning to secure smart city. Since there is no research work that have adopted the integration between ontology and ML to set inference rules of the ontology. We will compare our proposed approach in developing CAO with existing ontology-based approaches.

From the literature review in Section II, we have noticed that proposed ontologies in the domain of security are either environment-centric, assets centric, or threats-centric. However, the ontology developed in this work is cyberattack-centric ontology. Moreover, security needs in literature are considered from general view, which is not enough in IoT environment where security needs are considerably varied based on the application. In CAO, security needs are defined specifically based on the studied IoT system.

Finally, the main difference between CAO and previously proposed ontologies is the process of setting inference rules. Previous works have used the analysis of systems' vulnerabilities and threats, or they depend on user-defined rules. In CAO, we have benefit from ML to set inference rules.

VI. CONCLUSION AND FUTURE WORK

In this article, we have proposed an ontology for cyber-attacks and we have shown how this ontology can be navigated to deduce the vulnerabilities, potential attacks, and the impact of the attack on the system in order to develop a model for an IDS for a smart city based IoT system. The proposed ontology is cyber-attack centric and includes other two main entities, adversary and system.

For future researches, CAO could be thought of as a base for more comprehensive ontologies, more concepts and relations could be added to develop cyber-attack oriented knowledge. Furthermore, CAO helps to develop a taxonomy for cyber-attacks. It also provides a formal and unified description and understanding for cyber-attacks and security needs for IoT environment.

We have investigated a case study for DoS attacks, six attacks have been considered and ML is used to set the rules for the proposed ontology. We have also proposed safety factor to evaluate the effectiveness of the IDS. The proposed ontology as well as the presented case study have shown that ontology can play a significant role to secure a smart city. The development of comprehensive ontology will establish a knowledge base for cyberattacks which creates the opportunity for robust protection for existing as well as coming security threats and attacks.

The proposed CAO needs to be supported by employing the proposed approach in a diversified set of computing environments, and investigating more types of cyber-attacks. Obtained accuracy values using CAO are less than detection accuracy in existing ML learning based security methods. The accuracy can be raised by studying several types of attacks or by the generating larger dataset which is the suggested future work.

REFERENCES

[1] Wang, D., Lee, S., Zhu, Y., & Li, Y. (2017, March). A zero human-intervention provisioning for industrial IoT devices. In 2017 IEEE International Conference on Industrial Technology (ICIT) (pp. 1171-1176). IEEE.

[2] Asassfeh, M., Obeid, N., Almobaideen, W. (2020, Dec), Anonymous Authentication Protocols for IoT based-Healthcare Systems: A survey, International Journal of Communication Networks and Information Security, 6(3), pp.302-315.

[3] Haque, A. B., Bhushan, B., & Dhiman, G. (2022). Conceptualizing smart city applications: Requirements, architecture, security issues, and emerging trends. Expert Systems, 39(5), e12753.

[4] Xu, G., Cao, Y., Ren, Y., Li, X., & Feng, Z. (2017). Network security situation awareness based on semantic ontology and user-defined rules for Internet of Things. IEEE Access, 5, 21046-21056.

[5] Almobaideen, W., Jarboua, H., Sabri, K.E. (2020), Searchable encryption architectures: survey of the literature and proposing a unified architecture, International Journal of Information Privacy, Security and Integrity, 4(4), pp. 237-260.

[6] Gopalan, S., Ali Raza, A., Almobaideen, W. (2021), IoT Security in Healthcare using AI: A Survey, 2020 International Conference on Communications, Signal Processing, and their Applications (ICCSPA), pp. 1-6, IEEE.

[7] Gheisari, M., Najafabadi, H. E., Alzubi, J. A., Gao, J., Wang, G., Abbasi, A. A., & Castiglione, A. (2021). OBPP: An ontology-based framework for privacy-preserving in IoT-based smart city. Future Generation Computer Systems, 113, 1-13.

[8] Gheisari, M., Pham, Q. V., Alazab, M., Zhang, X., Fernandez-Campusano, C., & Srivastava, G. (2019). ECA: an edge computing architecture for privacy-preserving in IoT-based smart city. IEEE Access, 7, 155779-155786.

[9] Alkhammash, E. (2020). Formal modelling of owl ontologies-based requirements for the development of safe and secure smart city systems. Soft Computing, 24(15), 11095-11108.

[10] Qamar, T., & Bawany, N. Z. (2020). A Cyber Security Ontology for Smart City. International Journal on Information Technologies & Security, 11(3), 63-74.

[11] Choi, Chang, and Junho Choi. "Ontology-Based Security Context Reasoning for Power IoT-Cloud Security Service." IEEE Access 7 (2019): 110510-110517.

[12] Mozzaquatro, Bruno Augusti, et al. "An ontology-based cybersecurity framework for the internet of things." Sensors 18.9 (2018): 1-20.

[13] Tao, M., Zuo, J., Liu, Z., Castiglione, A., & Palmieri, F. (2018). Multi-layer cloud architectural model and ontology-based security service framework for IoT-based smart homes. Future Generation Computer Systems, 78, 1040-1051.

[14] Oltramari, A., Cranor, L. F., Walls, R. J., & McDaniel, P. D. (2014, November). Building an Ontology of Cyber Security. In STIDS (pp. 54-61).

[15] Doynikova, E., Fedorchenko, A., & Kotenko, I. (2019, August). Ontology of metrics for cyber security assessment. In Proceedings of the 14th International Conference on Availability, Reliability and Security (pp. 1-8).

[16] Lannacone, M., Bohn, S., Nakamura, G., Gerth, J., Huffer, K., Bridges, R. & Goodall, J. (2015, April). Developing an ontology for cyber

security knowledge graphs. In Proceedings of the 10th Annual Cyber and Information Security Research Conference (pp. 1-4).

[17] Salem, M. B., & Wacek, C. (2015). Enabling New Technologies for Cyber Security Defense with the ICAS Cyber Security Ontology. In STIDS (pp. 42-49).

[18] Obrst, L., Chase, P., & Markeloff, R. (2011, October). Developing an Ontology of the Cyber Security Domain. In STIDS (pp. 49-56).

[19] Syed, Z., Padia, A., Finin, T., Mathews, L., & Joshi, A. (2016, March). UCO: A unified cybersecurity ontology. In Workshops at the thirtieth AAAI conference on artificial intelligence.

[20] Bergner, S., & Lechner, U. (2017). Cybersecurity Ontology for Critical Infrastructures. In KEOD (pp. 80-85).

[21] Xiao, L., Wan, X., Lu, X., Zhang, Y., & Wu, D. (2018). IoT security techniques based on ML: How do IoT devices use AI to enhance security?. IEEE Signal Processing Magazine, 35(5), 41-49.

[22] Hussain, F., Hussain, R., Hassan, S. A., & Hossain, E. (2020). ML in IoT security: current solutions and future challenges. IEEE Communications Surveys & Tutorials.

[23] Alghanam, O. A., Almobaideen, W., Saadeh, M., & Adwan, O. (2023). An improved PIO feature selection algorithm for IoT network intrusion detection system based on ensemble learning. Expert Systems with Applications, 213, 118745.

[24] Musen, M. A. (2015).The protégé project: a look back and a look forward. AI matters, 1(4), 4-11.

[25] Almobaideen, W., Allan, M., Saadeh, M. (2016), Smart archaeological tourism: Contention, convenience and accessibility in the context of cloud-centric IoT, Mediterranean Archaeology & Archaeometry, 16(1).

[26] Hashim Raza Bukhari, S., Siraj, S., & Husain Rehmani, M. (2018). Wireless sensor networks in smart cities: applications of channel bonding to meet data communication requirements. Transportation and Power Grid in Smart Cities: Communication Networks and Services, 247-268.

[27] Haq, N. F., Onik, A. R., Hridoy, M. A. K., Rafni, M., Shah, F. M., & Farid, D. M. (2015). Application of machine learning approaches in intrusion detection system: a survey. IJARAI-International Journal of Advanced Research in Artificial Intelligence, 4(3), 9-18.

[28] Almseidin, M., Alzubi, M., Kovacs, S., & Alkasassbeh, M. (2017, September). Evaluation of machine learning algorithms for intrusion detection system. In 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY) (pp. 000277-000282). IEEE.

[29] Jain, A., & Jain, S. (2019). A survey on miscellaneous attacks and countermeasures for RPL routing protocol in IoT. In Emerging Technologies in Data Mining and Information Security (pp. 611-620). Springer, Singapore.

[30] Zhang, T., & Li, X. (2014, August). Evaluating and Analyzing the Performance of RPL in Contiki. In Proceedings of the first international workshop on Mobile sensing, computing and communication (pp. 19-24).

[31] Padmaja, P. L., Ramanjaneyulu, T., Narayana, I. L., & Srikanth, K. (2017). Role of COOJA simulator in IoT. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 6(2), 139-143.

[32] Srivastava, S. (2014). Weka: a tool for data preprocessing, classification, ensemble, clustering and association rule mining. International Journal of Computer Applications, 88(10).

APPENDIX A
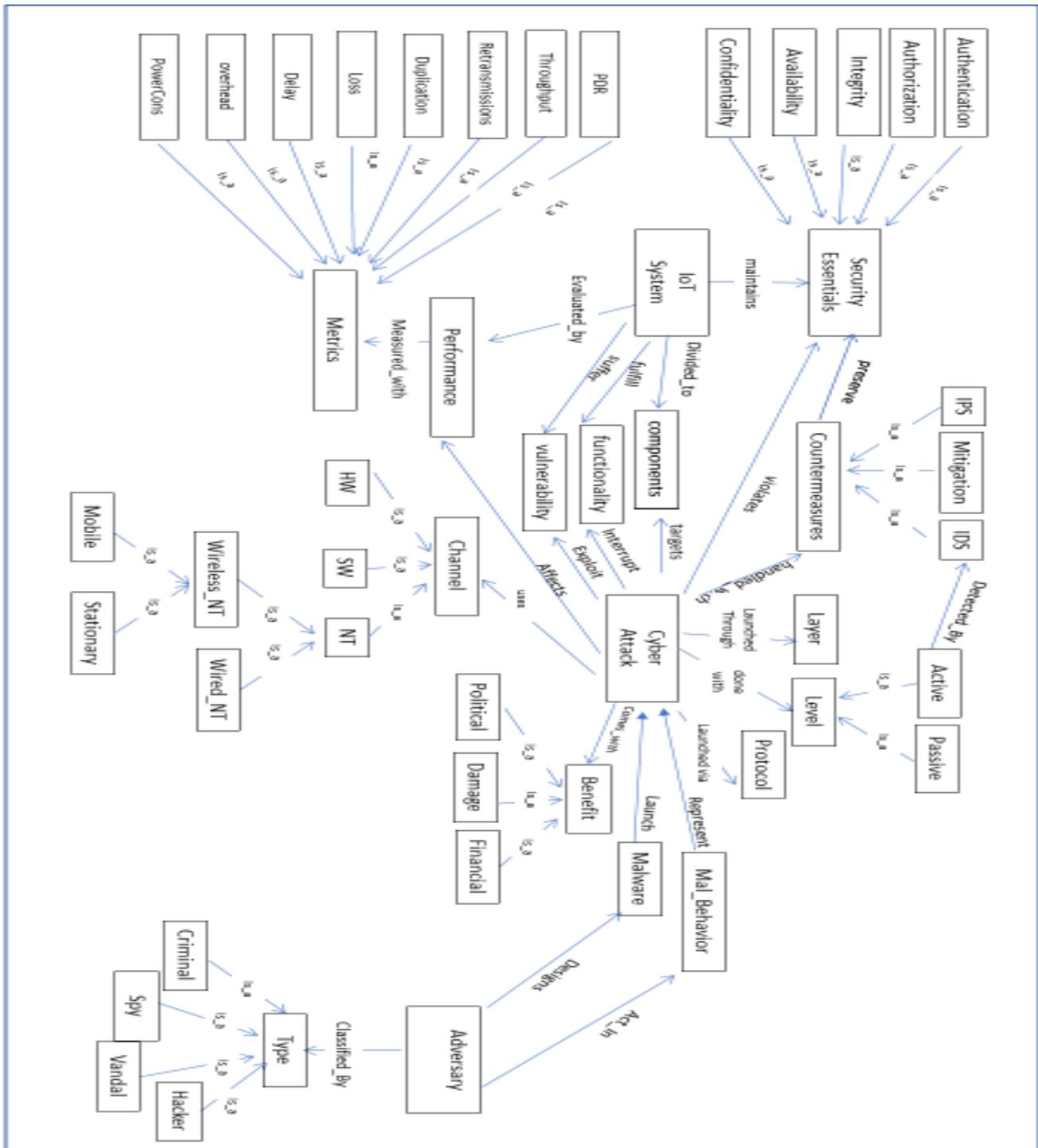
Ontology Graph.

APPENDIX B

LOGICAL INFERENCE RULES

Rule	Classification
DIS <= 19 AND sent <= 20	Normal
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR > 0.714 AND DIO > 115	Copycat Attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR > 0.714 AND DIO <= 115 AND Beacon<=110.57 AND Table <=15 AND DIS > 11	Copycat Attack
DIS > 19	Flooding attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR <= 0.714 AND Rem_power> 3724 AND Min_forwarded =0	Rank Attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR <= 0.714 AND Rem_power> 3724 AND Min_forwarded > 0 AND received<=11	Rank Attack
DIS <= 19 AND sent >20 AND DAO>494	Sinkhole attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR <= 0.714 AND Rem_power<=3724 AND Min_sent =0	Sinkhole attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR > 0.714 AND DIO <= 115 AND Beacon>110.57 AND Min_sent > 0 AND PDR <=0.77	Sinkhole attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR > 0.714 AND DIO <= 115 AND Beacon>110.57 AND Min_sent > 0 AND PDR >0.77 AND Max_power > 68598 AND DIO<=113	Sinkhole Attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR <= 0.714 AND Rem_power<=3724 AND Min_sent > 0	Sybil Attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR <= 0.714 AND Rem_power> 3724 AND Min_forwarded > 0 AND received>11	Sybil Attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR > 0.714 AND DIO <= 115 AND Beacon<110.57 AND Table >15	Sybil Attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR > 0.714 AND DIO <= 115 AND Beacon<=110.57 AND Table <=15 AND DIS <=11	Sybil Attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR > 0.714 AND DIO <= 115 AND Beacon>110.57 AND Min_sent = 0	Sybil Attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR > 0.714 AND DIO <= 115 AND Beacon>110.57 AND Min_sent > 0 AND PDR >0.77 AND Max_power <=68598	Sybil Attack
DIS <= 19 AND sent >20 AND DAO<=494 AND PDR > 0.714 AND DIO <= 115 AND Beacon>110.57 AND Min_sent > 0 AND PDR >0.77 AND Max_power > 68598 AND DIO>113	Sybil Attack

A Novel Annotation Scheme to Generate Hate Speech Corpus through Crowdsourcing and Active Learning

Nadeera Meedin, Maneesha Caldera, Suresha Perera, Indika Perera

Dept. of Computer Science and Engineering, University of Moratuwa, Katubedda, Sri Lanka

Abstract—**The number of user-generated posts is growing exponentially with social media usage growth. Promoting violence against or having the primary purpose of inciting hatred against individuals or groups based on specific attributes via social media posts is daunting. As the posts are published in multiple languages with different forms of multimedia, social media finds it challenging to moderate before reaching the audience and assessing the posts as hate speech becomes sophisticated due to subjectivity. Social media platforms lack contextual and linguistic expertise and social and cultural insights to identify hate speech accurately. Research is being carried out to detect hate speech on social media content in English using machine learning algorithms, etc., using different crowdsourcing platforms. However, these platforms' workers are unavailable from countries such as Sri Lanka. The lack of a workforce with the necessary skill set and annotation schemes symbolizes further research essentiality in low-resource language annotation. This research proposes a suitable crowdsourcing approach to label and annotates social media content to generate corpora with words and phrases to identify hate speech using machine learning algorithms in Sri Lanka. This paper summarizes the annotated Facebook posts, comments, and replies to comments from public Sri Lankan Facebook user profiles, pages and groups of 52,646 instances, unlabeled tweets based on 996 Twitter search keywords of 45,000 instances of YouTube Videos of 45,000 instances using the proposed annotation scheme. 9%, 21% and 14% of Facebook, Twitter and YouTube posts were identified as containing hate content. In addition, the posts were categorized as offensive and non-offensive, and hate targets and corpus associated with hate targets focusing on an individual or group were identified and presented in this paper. The proposed annotation scheme could be extended to other low-resource languages to identify the hate speech corpora. With the use of a well-implemented crowdsourcing platform with the proposed novel annotation scheme, it will be possible to find more subtle patterns with human judgment and filtering and take preventive measures to create a better cyberspace.**

Keywords—Annotation; crowdsourcing; hate speech detection; social media data analytics

I. INTRODUCTION

Social media users keep updating and sharing posts and comments on social media platforms at an exponential rate. Users can express themselves freely across countries and cultures in dozens of languages. People use social media to share their experiences, connect with friends and family, and build communities. However, social media platforms try to maintain their community standards so that their users feel safe using their products. One such example is the Facebook

community standard [1]. In their standards, they have specified the content not to be posted by users to prevent possible harms related to content on Facebook under categories such as violence and incitement, dangerous individuals and organization, coordinating damage and publicizing crime etc. For example, Facebook does not allow users to post content, including hate speech on Facebook to avoid creating an environment of intimidation and exclusion and which would promote real-world violence. Similarly, Twitter and YouTube have their hateful policy conduct [2] [3].

However, policies and standards exist, and social media platforms take action to remove posts with inappropriate content; users tend to spread hate speech using social media. The information shared on social media can be offensive and could lead to creating social issues. Social media research includes analyzing social media data on various topics. Some of these topics would have a direct impact on creating social issues. Therefore, it is vital to have a mechanism to identify the contents that directly impact social issues.

It has been a challenge for social media platforms to moderate billions of daily posts in more than a hundred languages. It has been found that it is impossible to maintain a balance between what is considered "hate speech" and "free speech" since social media is global. Hate speech is a broad and contested term [4], and there is no common standard definition for hate speech [5]. Multiple definitions used by different authorities and platforms are explained in the Literature Survey section. United Nations strategy and plan of action on hate speech describe hate speech as "any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language regarding a person or a group based on who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factors"[6]. Hence both "hate speech" and "free speech" are determined by region. When analyzing the speech, it is essential to consider context-specific details, social and cultural factors, etc. Determining hate speech based on user context is one of the challenges faced by social media platforms as they deal with more than a hundred languages [7] and nationalities.

By incorporating a crowdsourcing approach, it can reach a much larger audience and capture user opinions and behaviours. In social computing research, social media has provided a unique window into people's social experience; in particular, Twitter is used for assessing sensitive topics, such as discrimination[8]. Therefore, when capturing user opinions, it is required to ask the appropriate questions in the appropriate order to obtain responses effectively. Similarly,

the captured responses should be represented in such a way after ensuring their quality.

This research proposes a novel solution to address this problem: an adaptive questionnaire to identify hate speech, detect hate targets and keywords related to hate and label the social media posts written in Sinhala and Singlish. The main contributions of this paper are the following;

- A crowdsourcing framework for annotating hate speech.

- An annotating scheme to generate a hate speech corpus.

- Annotated dataset and hate speech corpus.

II. RELATED WORK

A. Use of Crowdsourcing in Annotating Hate Speech

Daniel Faggella [9] explains how crowdsourcing could be used in social media content moderation and states that moderation at a scale usually involves two elements; a trained machine learning algorithm informed by the user or outside data. In addition, several factors [10] affect the judgement of crowd moderators when deciding on the suitability of text content, such as participants often labelled unsure when they found it challenging to decide on marginal content. Considering these approaches, the factors considered in registering workers and how to direct workers to annotate, limiting the judgements are further explained in the Section III Crowdsourcing Platform.

Amazon Mechanical Turk(MTurk) is the most common crowdsourcing platform for performing labour-intensive tasks, data collection and annotation for hate speech identification [11], especially in NLP. MTurk facilitate quality management [11][12], ranking annotators[13], ensuring trustworthiness. There are no mechanisms in MTurk for detecting unfair evaluations and no metric called reputation. As a result, workers are highly susceptible to misbehaviour [14]. MTurk allows workers from only a limited number of countries to register. Hence the existing workers could not cater for the cultural, linguistic and contextual insights. Furthermore, MTurk is a general-purpose crowdsourcing platform which allows the annotation of many types of tasks and the annotation scheme adaptation for low-resource languages is crucial because of the linguistic challenges in MTurk.

B. Annotation in Hate Speech Detection

Both race and sexual orientation are among the top hate speech targets [15]. Our study examined the hate targets in the Sri Lankan context. It was different depending on the cultural, societal and religious differences. Amazon Mechanical Turk (AMT) was used in Relia et al. research for annotation after performing keyword filtering, and Support Vector Machine (SVM) and Neural Network (NN) was used for Twitter classification. As this task involved the exposure of humans to potentially sensitive content, the researchers have indicated the task was about racist Tweets as an individual Human Intelligence Task (HIT), giving workers a chance to discontinue at any point without losing payment if they felt uncomfortable. HaterNet [16] can identify and classify hate

speech in Twitter data. The Spanish national office is using it against hate crimes. HaterNet uses a combination of Long Short-Term Memory (LSTM) and Multi-layer Perceptron (MLP) neural networks.

Burnap [17] has annotated text using CrowdFlower, with a minimum of four humans, to assess whether it is likely to be offensive or antagonistic regarding race, ethnicity or religion. Scores are given based on a ternary set of classes, yes, no and undecided. Agrawal and Aweaker have used four DNN-based models for cyberbullying detection; CNN, LSTM, BLSTM, and BLSTM [18]. Here they have used a manually annotated dataset. Fernando and Asier [19] have introduced a new algorithm to detect hate speech messages. They used the Kappa coefficient to measure the degree of agreement when performing the subjective analysis. The researchers have categorized tweets into five categories: direct incitement or threat of violence, an attack on honour and human dignity, incitement to discrimination or hate and an offence to collective sensitivity.

The hate categories observed in [20] are race, behaviour, physical, sexual orientation, class, ethnicity, gender, disability, religion and others. A few of the top hate targets in the US, Canada, and the UK are black, fat, fake, stupid, gay, white, rude, ignorant, racist, old, selfish and religious, which is different in an Asian country Sri Lanka as it was found out in our study.

We followed the approach of Fernando and Asier [19] and used the five conditions given in Table IV to check if a post contains hate. In addition, the hate targets and the hate categories specified in each post were identified using the categories listed in [20]. These categories are listed in Table IV. Finally, we went one step further to identify the words, phrases and sentences which incite hatred in the post.

C. Benchmark Datasets for Hate Speech Detection

Table IX at the end of this paper summarizes existing benchmark annotated datasets used for hate speech detection in different languages. The system architecture, design and implementation of the crowdsourcing platform, information architecture of the system, the pre-selection criteria of contributors, and the proposed criterion to identify inappropriate content in social media are explained in the next section.

III. CROWDSOURCING PLATFORM

This paper presents a novel crowdsourcing platform that allows any interested participant to register by providing user profile details such as name, age, nationality, date of birth and location of contribution. The system architecture of the implemented crowdsourcing platform is given in Fig. 1.

The intrinsic rewarding process starts with the first digital badge, "Contributor", and after fulfilling the selection criteria specified in Table II, the contributors get the badge "Selected contributor" and the eligibility to earn financial rewards. The worker management process is illustrated in Fig. 2.

1) Pre-selection of contributors: Participants who were literate in Sinhala and Sri Lankan natives were selected as

contributors to the evaluation process. Those who did not qualify in the pre-selection process, failed at quality control, and failed at the trustworthiness-ensuring process were eliminated. The list of symbol definitions for pre-selection contributors is given below in Table I.
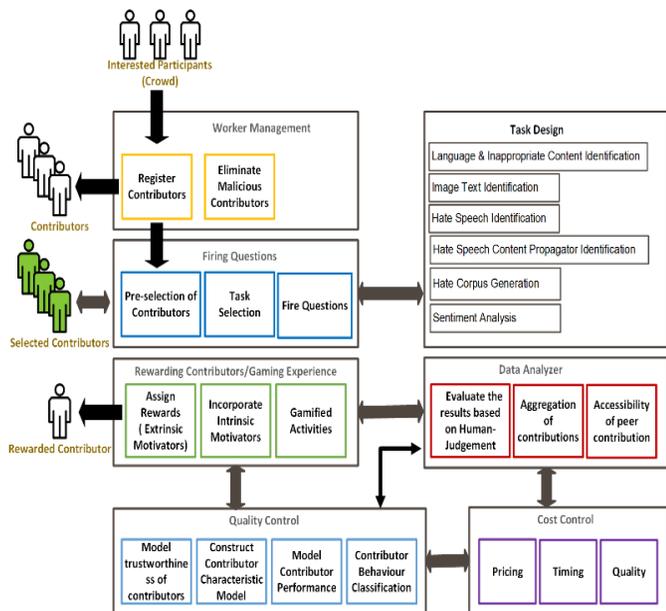
Fig. 1.   The System Architecture of the Crowdsourcing Platform.
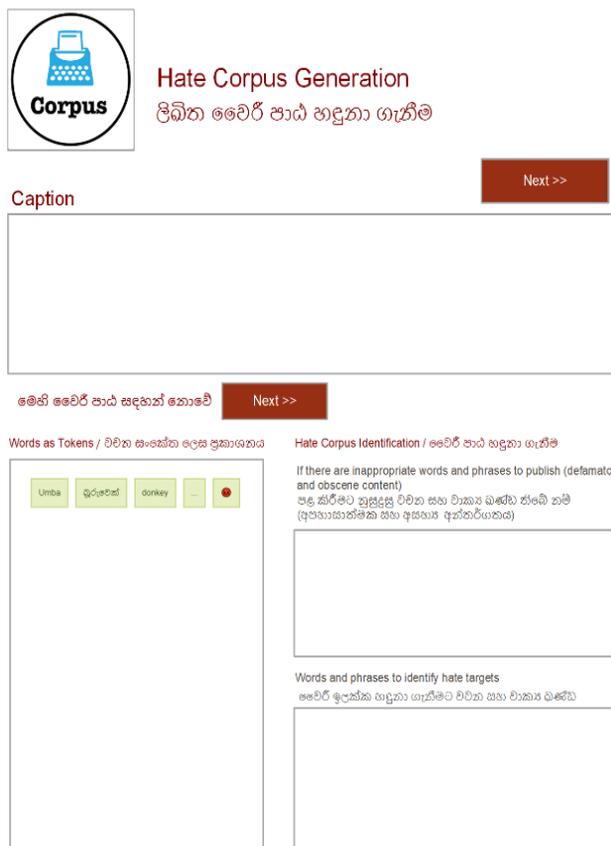
Fig. 2.   Worker Registration and Reward Process.

TABLE I.   LIST OF SYMBOL DEFINITIONS FOR PRE-SELECTION OF CONTRIBUTORS

Symbol	Definition
N	Nationality
A	Age
HS, HS_T	Knowledge level of hate speech, the Threshold value
LP, LP_T	Language proficiency(Sinhala), the Threshold value
CA, SCA_T	Comprehension & Analytical skill assessment(Sinhala), the Threshold value
L, L_T	Ability to read mixed codes(Sinhala words written in English letters), the Threshold value

TABLE II.   CRITERIA IN PRE-SELECTION OF CONTRIBUTORS

Pre-selection of contributors
if {(Nationality="Sri Lankan") and (A>=18) and (HS>=HS_T) and (LP>=LP_T) and (CA >= CA_T)and (L >=L_T)}
Badge="Selected Contributor"
else
Eliminate contributor
endif

The given threshold values were selected for the criteria in assessing prior knowledge in hate speech identification (HST=5), Sinhala language proficiency (LPT=8), ability to read Sinhala words written in English letters (LT=8), comprehension, and analytical skills (CAT=8) and the criteria in pre-selection of contributors is given in Table II.

Task Selection: Selected contributors were asked questions randomly based on a question generation mechanism helping to generate a corpus with hate speech, annotate the posts to detect and classify the intention of the hate speech, identify Sinhala texts from images, identify inappropriate words in the social media content and to sort Sinhala, Singlish and English words in a given set of social media posts. Out of the six types of task designs allowed by the crowdsourcing framework, only the results of hate speech identification and hate corpus generation are explained in the evaluation section of this paper (See Fig. 3).
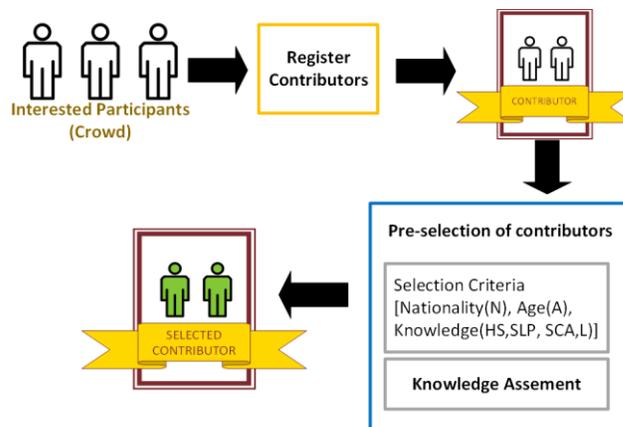
Fig. 3.   UI to Capture Inappropriate Contents and Words and Phrases Referring Hate Targets.

2) Fire questions: A rule base was used to assess the pre-selection criteria of contributors with forwarding chaining as the firing mechanism, and a dataset with JSON objects of Twitter, Facebook and Video posts was used during the task assignments.

3) Assign Rewards (Extrinsic and Intrinsic Motivators): Google crowdsource [21] is a global platform which uses the feedback of users to design products to provide a customized experience to its users. For example, they provide a gaming experience as an intrinsic reward to users instead of monetary rewards. After alterations, a similar approach has been taken in our research to motivate the Sri Lankan workers. The proposed reward system was designed after testing a few samples of workers by applying different rewarding methods.

Workers were allowed to earn monetary awards considering the completion levels, accuracy, trustworthiness of the contributor, etc. To admire the effort of the contributors to make cyberspace better monetary rewards were assigned to those who showed higher trustworthiness scores, and based on the human intelligence tasks HITs) completed, digital badges were assigned. In addition, the platform would provide a gaming experience for contributors to retain in the cause. The intrinsic and extrinsic motivators were embedded in the gaming experience, as shown in (See Fig. 4).
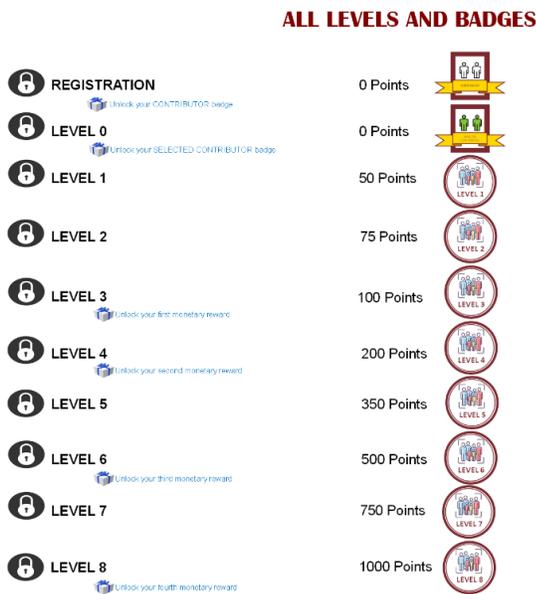
Fig. 4.   Intrinsic and Extrinsic Rewards Assignment for Selected Contributors.

4) Model the trustworthiness of a contributor: An inbuilt mechanism was built to check the trustworthiness of a particular contributor's responses and assign a badge for trustworthiness. Ten golden rules were used as the primary method of measuring trustworthiness and were compared with the predicted trustworthiness score. In addition, a higher weightage of validity to the response from a trustworthiness badge-owned contributor was considered in assessing the quality of the response.

5) Aggregation of contributions: Evaluates the results based on human judgment for each post, store the results, ignored, hate speech or not, category of the hate nature, etc., to aggregate the contributions and generate advanced questions later. The agreement scores were calculated using Cohen's kappa coefficient, Fleiss' kappa coefficient and Krippendorff's alpha.

IV. EXPERIMENTATION

The research was carried out considering three cases, the first using Facebook posts, the second using Tweets and the third using YouTube posts, as specified below in the dataset section. The initial classification step involved classifying the posts in the language in which the post was written. Here we categorized the language into five categories Sinhala, English, Singlish, Both English and Sinha-la, and None of the above. Out of the five categories, the posts written in Sinhala, Singlish and both Sinhala and English were considered under the study. Posts written in English and other languages were not considered under the study. The second step involved the task design for hate post identification, and the third in hate speech corpus generation.

V. PILOT DATASET

There is no large publicly available data set of Facebook, YouTube and Twitter written in Sinhala, Singlish and mixed code. Therefore, in this research, we selected three subsets from the three social media platforms as our cases under study, with the most significant number of Sri Lankan users as of 2022[22]. Singlish is Sinhala words written using English. Therefore, our dataset tweets were collected using Twitter API based on 996 Twitter search keywords. Web crawling was used to collect Facebook posts from public Sri Lankan Facebook user profiles, pages and groups, and YouTube posts and comments were captured from popular Sinhala YouTube channels as shown in Table III. The dataset was annotated by 20 selected annotators who passed the pre-selection test and showed a higher trustworthiness score. UPF-08 encoding errors and inconsistencies were corrected in the labelled files.

TABLE III.   UNLABELLED DATA SETS FOR HATE SPEECH IDENTIFICATION

Case	Dataset	Instances
Case 1: Facebook	Unlabeled posts, comments, and replies to comments from public Sri Lankan Facebook user profiles, pages and groups(Post image, Comments, video thumbnail)	52,646
Case 2: Twitter	Unlabeled tweets based on 996 Twitter search keywords (Tweet text, replies for each tweet, 6317 video thumbnails and pictures). The average number of comments and replies per Twitter post is 4	45,000
Case 3: Youtube	Youtube Videos(Video title, thumbnail, Comments for each video). The average number of comments and replies per video is 6	45,000

VI. ANNOTATION SCHEME TO GENERATE A HATE SPEECH CORPUS

The annotation scheme specified below in Table IV was used to get the data annotated, and the annotated datasets were published in the GitHub repository.

TABLE IV.   LIST OF SYMBOL DEFINITIONS FOR LABELLING

Symbol	Definition
L_1	Content analysis
L_2	Hate speech identification
L_3	Who does a particular post target?
L_4	Hate categories
$C1, C2, C3, C4, C5$	Direct incitement or threat of violence, An attack on honour and human dignity, Incitement to discrimination or hate, An offence to the collective sensitivity or Other.
$L_4\{1$ to $16\}$	1. Race and Ethnicity 2. Religion 3. Nationality 4. Sexual Orientation5. Disability 6. Disease 7. Immigration 8. Victims of a major violent event and their kin 9. Veteran Status/Profession10. Caste11. Political12. Regional 13. Gender 14. Economic & Business 15. A particular individual 16. Other social groups

Definition to label L_1 to L_4

Starting set of labels L_1 is:

L_1={*Offensive content, No offensive content, Cannot tell*}

Starting set of labels L_2 is:

L_2={*C1, C2 , C3, C4, C5*}

Starting set of labels L_3 is:

L_3={*Individual, Group, Cannot tell*}

Starting set of labels L_4 is:

L_4={*1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16*}

VII. RESULTS AND DISCUSSION

A. Demographic Distribution of the Annotator Profiles

The demographic distribution of the selected contributors involved with the annotation process is given in Fig. 5 to 7.
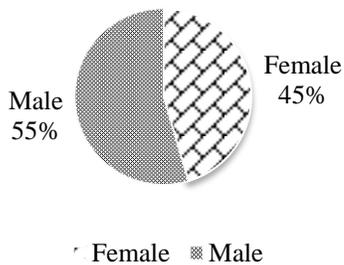
Fig. 5.   Gender Distribution of Selected Contributors.
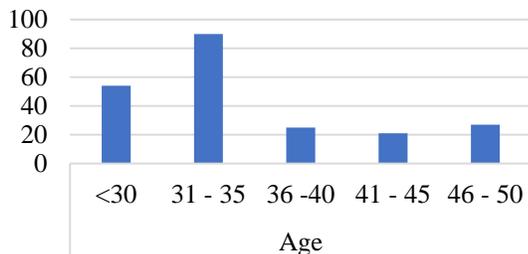
Fig. 6.   Age Distribution of Selected Contributors.
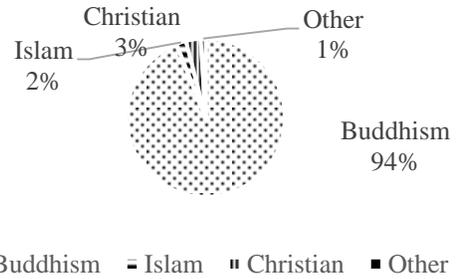
Fig. 7.   Religion-wise Distribution of Selected Contributors.

B. Inter Annotator Agreement

Table V shows the agreements obtained in terms of the average percent agreement (agr_i), average Cohen's kappa coefficient (avg k), Fleiss' kappa coefficient (Fleiss) and Krippendorff's alpha (α). The number of annotated posts/comments/tweets is also given for each batch.

If the set of items is $\{i | i \in I\}$, and is of cardinality i.

Observed agreement(A_0) over the values agr_i for all items $i \in I$.

$$A_o \frac{1}{i} \sum_{i \in I} agr_i \qquad (1)$$

$$agr_i = \begin{cases} 1, & \text{if the two coders assign i to the same category} \\ 0, & \text{if the two coders assign i to different categories} \end{cases}$$

Cohen Kappa coefficient(k) where expected agreement is A_e

$$K = \frac{A_0 - A_e}{1 - A_e} \qquad (2)$$

TABLE V.   INTER ANNOTATOR AGREEMENT ON CASE1, CASE 2 AND CASE 3 FOR L_1

Case	Instances	agr_i	k	Fleiss	α
Case 1:Facebook	52,646	90.4	0.623	0.615	0.800
Case 2:Twitter	45,000	89.4	0.624	0.613	0.740
Case 3: Youtube	45,000	88.7	0.598	0.600	0.757

C. Annotated Datasets

Hate posts/comments and tweets were identified after analyzing L_2 and L_3. If at least one choice was selected and the inter-rater agreement was more than 0.6, the posts were identified as hate posts.

If we compare Table VI and Table VII, it is evident that intuitive annotation of offensive and non-offensive content shows lesser percentages when compared with hate and non-hate classification.

TABLE VI.   DISTRIBUTION OF POSTS/COMMENTS/TWEETS ANNOTATIONS

	Facebook	Twitter	YouTube
Hate	9%	21%	14%
No Hate	88%	70%	83%
Skip	3%	9%	3%

D. Lexical Distribution

Table VIII lists the ten most frequently occurring words to refer to the hate targets with the most frequent occurrence for each class. Twitter search keywords, identified hate targets and the corpus associated with hate targets focusing on an individual or group are given in the GitHub repository. The annotated Facebook, Twitter, and YouTube datasets can be requested by emailing the authors of this paper.

TABLE VII.   DISTRIBUTION OF POSTS/COMMENTS/TWEETS OVER CATEGORIES IN THE CLEAN DATASET-FACEBOOK

	Offensive	No offensive content
Facebook		
Caption	1.68%	>98%
Video thumbnail	0.18%	>99%
Main image	0.015%	>99%
Comment text	0.023%	>99%
Comment image	0.002%	>99%
Reply text	0.78%	>99%
Group photo	0.04%	>99%
Twitter		
Tweet text	1.54%	>98%
Comment text	2.27%	>97%
Reply text	1.03%	>98%
YouTube		
Video title	0.81%	>99%
Video thumbnail	0.57%	>99%
Comment text	2.27%	>97%
Reply text	1.01	>98%

TABLE VIII.   DISTRIBUTION OF THE TEN MOST FREQUENTLY OCCURRING TERMS IN HATE SPEECH REFERRING HATE TARGETS

Word	Meaning of the word in English	Distribution
දෙමළා	Used as an insulting and contemptuous term for a person from a target community	0.65%
හම්බා	Used as an insulting and contemptuous term for a person from a target community	0.67%
තම්බියෙක්		0.65%
තම්බියා		1.9%
තම්බි	Used as an insulting and contemptuous term to refer target community	1.01%
තම්බියෝ		1.83%
තම්බිලා		0.38%
ගණයා	Used as an insulting and contemptuous term to refer to a priest from a target religion	0.25%
අන්තවදී	Used to refer to a person as an extremist	0.35%
සිංහලේ	Used to elevate an ethnicity and demean all the other ethnicities	0.65%

TABLE IX.   SUMMARY OF EXISTING BENCHMARK ANNOTATED DATASETS USED FOR HATE SPEECH DETECTION FOR DIFFERENT LANGUAGES

Dataset	Social Media Platform	Language	Criteria/Based on	Annotation question/s or Categories	Size	Intercoder-agreement score	Output
Davidson [23] Hatebase Twitter	Tweet	English	The words appearing in a given tweet and the context in which they were used.	Labelled categories: hate speech, offensive but not hate speech, or neither offensive nor hate speech	24,802 tweets	Majority Voting 92%	5% - hate speech 76% - Offensive 17% - Neither
Waseem[24]	Tweet	English	Criteria based on Critical Race Theory(CRT) : 1. uses a sexist or racial slur. 2. attacks a minority. 3. seeks to silence a minority. 4. criticizes a minority (without a well-founded argument). 5. promotes, but does not directly use, hate speech or violent crime. 6. criticizes a minority and uses a straw man argument. 7. blatantly misrepresents truth or seeks to distort views on a minority with unfounded claims. 8. shows support for problematic hashtags. E.g. "#BanIslam", "#whoriental", and "#whitegenocide" 9. negatively stereotype a minority. 10. defends xenophobia or sexism. 11. contains an offensive screen name, as per the previous criteria, the tweet is ambiguous (at best), and the tweet is on a topic that satisfies any of the above criteria	Labelled categories: Sexism, Racism, Neither	16,914 tweets	Cohen's kappa coefficient K=0.84	Racism – 12% Sexism – 20% Neither - 68
Gibert et al.[25]	Stormfront - Internet posts	English	a) deliberate attack, b) directed towards a specific group of people, c) motivated by aspects of the group's identity.	Hate, No hate, Skip	10,568 sentences	Average percent agreement 91.03 Cohen's kappa coefficient 0.614 Fleiss' kappa coefficient 0.607	Hate 1196 Skip 72 No hate 9674

Most of the phrases referring to the hate targets consisted of at least one term from the given list.

VIII. CONCLUSION

After performing this research, the following drawbacks were identified. (1) The workers found it challenging to identify if a particular comment is harmful or harmless by looking only at it. (2) based on the task, the number of annotator requirements should be identified and vary (3) eliminate the bias of the annotator response.

To make the decisions, they needed to see the original post, the replies against it, and the images associated with the comments, if any. Therefore, suggesting preventive measures would not work if we only remove a single comment. Instead, it would be required to remove a set of comments. Therefore when designing tasks for crowdsourcing platforms, it is required to redesign the tasks to include the relevant images and any context-specific data along with the post when asking questions from the crowd, as proposed in this research.

Though subjective responses were captured in labelling, it is mandatory to design the annotation scheme such that the annotator would consciously judge the comment based on the given criteria without intuitively completing the crowdsourcing tasks.

Two different techniques should be used to score the trustworthiness of users during the pre-selection of contributors to the crowdsourcing platform. Krippendorff's alpha coefficient could be used as the reliability estimation method, and the number of contributors for each type of task would be very, as shown in the results. Two contributors can be used to perform primary classifications, while in-depth analyses such as sentiment strength analysis and hate target identification should be increased by checking the reliability score.

It is essential to use equal percentages to represent each religion which is nearly impossible in this context as the majority of the Sri Lankan population is Buddhists, to avoid bias toward religion.

The beliefs of the contributor would affect their response. Therefore, it is required to randomize the worker selection to eliminate the bias and to fire different categories of posts for each worker type.

Despite the outcomes of this research, future research should aim to (1) identify multiple comments against the Facebook posts instead of a single comment or caption to remove as a preventive measure in spreading hate, (2) redesign crowdsourcing tasks to include the relevant images and any context-specific data along with the post when asking questions from the crowd (3) having a mechanism to ensure annotator would consciously judge the comment based on the given criteria without intuitively completing the crowdsourcing tasks (4) measure the biases and beliefs of workers to ensure the trustworthiness of crowd response (5) identifying clusters of worker types and fire different categories of questions to each type.

The outcomes of this research consist of a crowdsourcing framework for annotating hate speech for Sinhala, Sinhala written in English and English social media posts, an annotation scheme to generate a hate speech corpus and an annotated dataset. These outcomes can be used by NLP researchers in performing linguistic research and getting annotation done for local languages, and policymakers to take preventive measures in identifying inappropriate content, hate targets and hate categories.

The datasets, implemented project, hate targets, hate-related search key terms used with Twitter, and hate corpus could be found in the;

https://github.com/gsnadeerameedin/HateSpeechCorpus.

References

[1] 'Facebook Community Standards | Transparency Centre'. https://transparency.fb.com/en-gb/policies/community-standards/ (accessed Aug. 07, 2022).

[2] 'Twitter's policy on hateful conduct | Twitter Help'. https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy (accessed Aug. 07, 2022).

[3] 'Hate speech policy - YouTube Help'. https://support.google.com/youtube/answer/2801939?hl=en (accessed Feb. 02, 2020).

[4] I. Gagliardone, D. Gal, T. Alves, G. Martinez, and Unesco, Countering online hate speech. Paris: United Nations Educational, Scientific and Cultural Organization, 2015.

[5] S. MacAvaney, H.-R. Yao, E. Yang, K. Russell, N. Goharian, and O. Frieder, 'Hate speech detection: Challenges and solutions', PloS One, vol. 14, no. 8, p. e0221152, 2019.

[6] A. Guterres, 'What is hate speech?', UNITED NATIONS STRATEGY AND PLAN OF ACTION ON HATE SPEECH, May 2019. Accessed: Feb. 04, 2020. [Online]. Available: https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf.

[7] C. Curtis, 'Facebook's global content moderation fails to account for regional sensibilities', The Next Web, Feb. 26, 2019. https://thenextweb.com/socialmedia/2019/02/26/facebooks-global-content-moderation-fails-to-account-for-regional-sensibilities/ (accessed Feb. 02, 2020).

[8] T. Finin, W. Murnane, A. Karandikar, N. Keller, J. Martineau, and M. Dredze, 'Annotating named entities in Twitter data with crowdsourcing', p. 9.

[9] S. Juumta, 'Crowdsourced Content Moderation - How it Works and What's Possible', Emerj Artificial Intelligence Research. https://emerj.com/partner-content/crowdsourced-content-moderation-how-it-works-and-whats-possible/ (accessed Aug. 07, 2022).

[10] D. Hettiachchi and J. Goncalves, 'Towards Effective Crowd-Powered Online Content Moderation', in Proceedings of the 31st Australian Conference on Human-Computer-Interaction, Fremantle WA Australia, Dec. 2019, pp. 342–346. doi: 10.1145/3369457.3369491.

[11] P. Ipeirotis, 'Crowdsourcing using mechanical turk: quality management and scalability', in Proceedings of the 8th International Workshop on Information Integration on the Web: in conjunction with WWW 2011, 2011, p. 1.

[12] P. G. Ipeirotis, F. Provost, and J. Wang, 'Quality management on Amazon Mechanical Turk', in Proceedings of the ACM SIGKDD Workshop on Human Computation - HCOMP '10, Washington DC, 2010, p. 64. doi: 10.1145/1837885.1837906.

[13] V. C. Raykar and S. Yu, 'Ranking annotators for crowdsourced labeling tasks', in In Advances in neural information processing systems, 2011, pp. 1809–1817.

[14] M. Allahbakhsh, A. Ignjatovic, B. Benatallah, S.-M.-R. Beheshti, E. Bertino, and N. Foo, 'Reputation Management in Crowdsourcing Systems', in Proceedings of the 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, Pittsburgh, United States, 2012. doi: 10.4108/icst.collaboratecom.2012.250499.

[15] L. Silva, M. Mondal, D. Correa, F. Benevenuto, and I. Weber, 'Analyzing the Targets of Hate in Online Social Media', ArXiv160307709 Cs, Mar. 2016, Accessed: Feb. 04, 2020. [Online]. Available: http://arxiv.org/abs/1603.07709.

[16] J. C. Pereira-Kohatsu, L. Quijano-Sánchez, F. Liberatore, and M. Camacho-Collados, 'Detecting and Monitoring Hate Speech in Twitter', Sensors, vol. 19, no. 21, p. 4654, Oct. 2019, doi: 10.3390/s19214654.

[17] P. Burnap and M. L. Williams, 'Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making', Policy Internet, vol. 7, no. 2, pp. 223–242, 2015.

[18] S. Agrawal and A. Awekar, 'Deep Learning for Detecting Cyberbullying Across Multiple Social Media Platforms', in Advances in Information Retrieval, vol. 10772, G. Pasi, B. Piwowarski, L. Azzopardi, and A. Hanbury, Eds. Cham: Springer International Publishing, 2018, pp. 141–153. doi: 10.1007/978-3-319-76941-7_11.

[19] F. Miró-Llinares, A. Moneva, and M. Esteve, 'Hate is in the air! But where? Introducing an algorithm to detect hate speech in digital microenvironments', Crime Sci., vol. 7, no. 1, p. 15, Dec. 2018, doi: 10.1186/s40163-018-0089-1.

[20] M. Mondal, L. A. Silva, and F. Benevenuto, 'A Measurement Study of Hate Speech in Social Media', in Proceedings of the 28th ACM Conference on Hypertext and Social Media - HT '17, Prague, Czech Republic, 2017, pp. 85–94. doi: 10.1145/3078714.3078723.

[21] 'Crowdsource by Google'. https://crowdsource.google.com/ (accessed Nov. 21, 2022).

[22] 'Digital 2022: Sri Lanka', DataReportal – Global Digital Insights. https://datareportal.com/reports/digital-2022-sri-lanka (accessed Nov. 02, 2022).

[23] T. Davidson, D. Warmsley, M. Macy, and I. Weber, 'Automated Hate Speech Detection and the Problem of Offensive Language', ArXiv170304009 Cs, Mar. 2017, Accessed: May 31, 2021. [Online]. Available: http://arxiv.org/abs/1703.04009.

[24] Z. Waseem and D. Hovy, 'Hateful symbols or hateful people? predictive features for hate speech detection on twitter', in Proceedings of the NAACL student research workshop, 2016, pp. 88–93.

[25] O. de Gibert, N. Perez, A. García-Pablos, and M. Cuadros, 'Hate Speech Dataset from a White Supremacy Forum', in Proceedings of the 2nd Workshop on Abusive Language Online (ALW2), Brussels, Belgium, Oct. 2018, pp. 11–20. doi: 10.18653/v1/W18-5102.

Detecting Brain Diseases using Hyper Integral Segmentation Approach (HISA) and Reinforcement Learning

M. Praveena[1]

Department of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram, AP, India

M. Kameswara Rao[2]

Department of ECM
Koneru Lakshmaiah Education Foundation
Vaddeswaram, AP, India

Abstract—Medical Images are most widely done by the various image processing approaches. Image processing is used to analyze the various abnormal tissues based on given input images. Deep learning (DL) is one of the fast-growing field in the computer science and specifically in medical imaging analysis. Tumor is a mass tissue that contains abnormal cells. Normal tumor tissues may not grow in other places but if it contains the cancerous (malignant) cells these tissues may grow rapidly. It is very important to know the cause of brain tumors in humans and these should be detected in the early stages. Magnetic Resonance Imaging (MRI) images are most widely used to detect the tumors in the brain and these are also used to detect the tumors all over the body. Tumors are of various types such as noncancerous (benign) and cancerous (malignant). Sometimes tumors may convert into cancer cells based on the stage of the tumor. In this paper, a hyper integral segmentation approach (HISA) is introduced to detect cancerous tumors and non-cancerous tumors. Detecting cancerous cells in the tumors may reduce the life threat to the affected persons. The agent based reinforcement classification (ABRC) is used to classify the Alzheimer's disease (AD) and cancerous and non-cancerous cells based on the abnormalities present in the MRI images. Two publically available datasets are selected such as MRI images and AD-affected MRI images. Performance is analyzed by showing the improved metrics such as accuracy, f1-score, sensitivity, dice similarity score, and specificity.

Keywords—Benign; malignant; magnetic resonance imaging; Alzheimer's disease

I. INTRODUCTION

The brain is one of the most important organs that play a major role in the human body. This organ controls the whole body by giving instructions to the rest of the body parts. In the world, the second cause of human death is caused by brain disease if the people are not identified in the early stages [1]. From the computer point of view, many applications are developed to analyze brain tumors and other brain disorders by using MRI, CT scans, and X-rays. Very fast and timely recognition and detection of brain tumors are required for curing the tumor. This depends on the experts and professionals to treat the patients. It is a very tedious task to predict the type of the tumors present in the brain for the proper treatment [2]. Segmentation of brain tumors is mainly based on a few important factors such as removing noise, missing borders, and contrast. MRI segmentation is one of the

better strategies for recognizing brain images. Based on the density functions the parametric approaches are used for selecting the tumors [3]. Some of the basic modalities used to detect and analyze brain tumors are positron emission tomography (PET), magnetic resonance imaging (MRI), and computed tomography (CT) [4].

Generally various types of tumors are present. Tumors may convert into cancer and non-cancerous cells. After the heart disease, most of the people causing deaths are cancer patients. Among all the types of cancers, brain cancers become more complex to detect and analyze in the early stages [5]. MRI is one of the popular approaches for detecting the cancerous tissues in brain tumors [6]. Traditional segmentation of analysis is a very tedious and time-consuming process for experts also [7] [8]. Thus an automated and dynamic detection and recognition of tumors belong to cancerous or not.

Reinforcement learning (RL) [9] is used to increase the performance of the existing approaches by using agent-based learning. The RL agent mainly controls the behavior of the policy mapping from given inputs to actions. RL mainly uses the labeled training with the expected output. This also uses the unique optimal action that receives the updated sample to indicate the accurately selected action. RL is used to solve the classification tasks in DCNN. Instead of the traditional classification process, the advanced RL classification works better on same brain datasets are straightly generated by using the classifier with estimated class labels. The agent plays the major role in RL to communicate with the input and takes the required action to process the MRI brain image classification (Fig. 1).
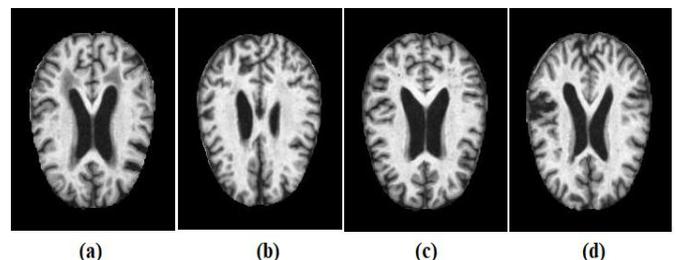
Fig. 1. Sample Images of AD (a) Mild Demented, (b) Moderate Demented, (c) Non Demented (d) Very Mild Demented.

The proposed approach is also used to detect the AD and also types of AD based on the given brain samples. This approach finds the shape and curves of the cancer cells present in the brain. It consists of multi-layered approach that contains multiple layers to get accurate results. A powerful deep learning training model is used to train the types of tumor images. Fig. 2 shows the classification of tumor cells and cancer cells present in the input image. In Fig. 2(a) is the normal MRI brain image, Fig. 2(b) is the tumor image which is represented in red color and 2(c) represents in blue color with cancer cells region.
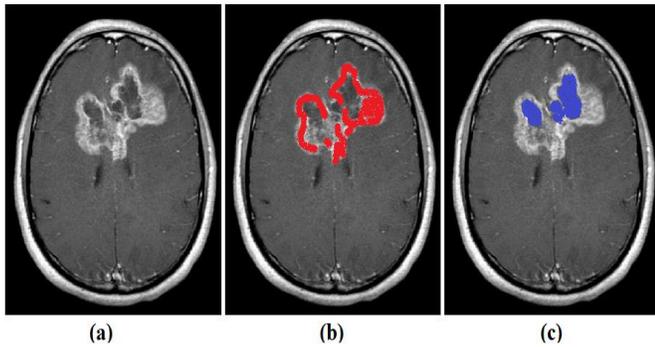
Fig. 2. (a) Tumor Image (b) Tumor Cells (c) Cancer Cells.

II. LITERATURE SURVEY

Z. Jia et al., [10] proposed the FAHS-SVM to segment the brain images. This approach is focused on detecting the cerebral venous system by using MRI images. This approach also contains one or more layers belonging to hidden nodes. M. S. Majib et al., [11] introduced the dynamic approach to classifying tumor-affected images without the interference of humans. The proposed approach VGG-SCNet's achieved a better performance for the detection of the tumor. M. N. Islam et al., [12] introduced a new approach to predicting the stage of childbirth accurately. S. Pereira et al., [13] developed the automated segmented approach by integrating CNN. Several steps are used to detect brain tumors by using the advanced segmentation approach. The normalization step is also utilized to remove the clamor from the image. Shakeel et al., [14] proposed the ML-based NN approach that analysis the tumor images by using infrared sensor imaging technology. Fractal dimension algorithm (FDA) is used for features extraction. This approach is mainly used to reduce the difficulties of tumor detection. Mallick, P. K et al., [15] proposed the improved compression approach that consists of a deep wavelet auto-encoder (DWA) for the feature reduction property. This approach also finds the size of the tumors. Li, G et al., [16] discussed various issues that are identified in the present pedestrian detection methods.

Zhang, D. et al., [17] proposed a novel approach to extract the features from the tumor images and these are called weights. This approach detects the tumor affected regions of the tumor images. Zhou, C et al., [18] proposed the lightweight model to solve several complex issues based on the detection of tumors. The post-processing method is also used to get the segmented outputs. Badrinarayanan, V., [19] presented a novel approach that deeply worked on a pixel-based segmented approach called as SegNet. The decoder network is also used to analyze the resolution of the brain images by using the feature maps. Hu, K. et al., [20] proposed a variety of segmentation approach that combines the two deep learning models such as MCCNN and CRFs. The proposed segmentation approach process the MRI images in two stages, in the first stage MCCNN is used to merge the components that take dependencies on multi-scale features segmentation. In the second stage, CRF is applied to get accurate segmentation by using spurious outputs. The proposed approach shows better results in detecting the tumors. Almahfud et al., [21] discussed several issues in detecting the brain tumors by using the optimal and nearest outliers based on the sensitive color variations. This approach is the combination of K-Means and Fuzzy C-Means. This is applied to MRI brain image datasets. This approach achieved better results compared with K-Means and FCM. K. Muhammad et al., [22] introduced the survey on various DL algorithms such as BTC models. These approaches discussed pre-processing, feature extraction, and classification. The existing approaches are applied to various benchmark datasets. M. Rizwan et al., [23] proposed the distinctive brain tumor approach that detects the tumors from the given quality MRI brain images. The proposed approach classifies the types of tumors such as pituitary, glioma, and meningioma. This approach is also focused on separating the tumor grades. A. Saleh et al., [24] aim to improve the tumor detection rate by using the DL algorithms. Various DL algorithms such as ANN, CNN, and D-CNN are applied to several approaches [35-37]. N. Noreen et al., [25] proposed the DL-based approach that contains the integrated approach to classify the tumors from MRI brain images. Several pre-trained models are used for training and feature extraction approaches are also used. All these features are combined and sent to the soft-max layer for the classification of tumors and non-tumors. The proposed approach achieved the accuracy of 99.35% with Inception-v3 and 99.56% with DensNet201. Zhou SK et al., [26] proposed the deep reinforcement learning (DRL) model that teaches the group of actions that increases the performance of the output. DRL utilizes the agent based approach that increases the accuracy of brain tumor detection. Chao Yu et al., [27] discuss several RL approaches that are applied on various medical and healthcare diagnosis. Al Walid Abdullah et al., [28] proposed the policy based redesigning the partial policies that ensures the robust and effective localization using the sub-agents by solving the several decision based approaches [29-34]. The proposed approach shows the optimal solution for the actual behavior learning framework.

III. FEATURE EXTRACTION METHODS TO REMOVE THE NOISE FROM THE MRI IMAGES

This step is mainly based on changing the pixel intensity values based on range. This approach will improve the contrast of the input image that helps image for better tumor detection. This is also called contrast stretching. This approach will transfer the n-dimensional MRI gray-scale image. The two types of MRI images such as cancer tumor image and AD based image are used to process. The following mathematical equation is used to analyze the pixels of the image.

$$\hat{X}[:,i] = \frac{X[:,i] - \min(X[:,i])}{\max(X[:,i]) - \min(X[:,i])} \quad (1)$$

For every brain image, the pixel numbers are positive and the range of normalized data is [0, 1] or [0, 255].

IV. IMAGE CLASSIFICATION USING DEEP-CONVOLUTIONAL NEURAL NETWORK (D-CNN)

Classification of brain MRI scan images is used to segment the various types of features. The MRI images have some features such as edges, intensity pixels, pixel value changes, and the size of the image. After the training and preprocessing the D-CNN is integrated with the feed-forward neural network (FFNN) and analyzes the normal and abnormal tissues based on the threshold values and processes the data in the grid-based topology. This is also called as ConvNet. Based on the types of tumors the classification is done by DCNN. Fig. 3 shows the processing of layers present in the DCNN. This is also shows the classification of AD images and normal images. Fig. 4 shows the execution flow for detecting AD and NON-AD images. This shows the functionality of every step and process the images for classification.

This consists of various hidden layers which are used to extract the data from the MRI scan images. This contains four layers such as:
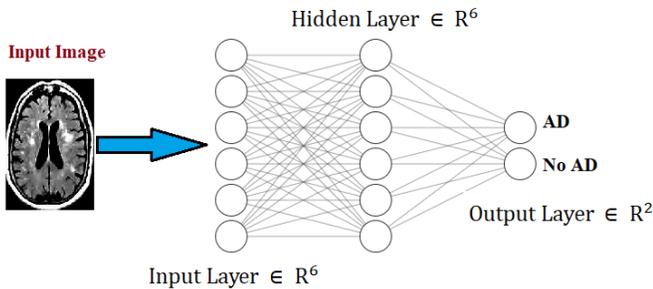
Fig. 3. D-CNN to Find the AD or Not.

A. Convolution Layer

This is one of the significant layers and this consists of the first layer and this will extract the highly valuable features from the MRI images. These layers contain the filters that use the convolution operation. Every image initializes the matrix pixel value. The range of these pixel values is 0 or 1. 3x3 filter matrix is utilized. The dot product is calculated to get the accurate feature matrix.

$$Z^l = h^{l-1} * W^l \quad (2)$$

1) ReLU layer: ReLU (Rectified Linear Unit): This is used to extract the feature maps.

This operation performs the component wise operation and the negative pixels are represented as 0. This is also used the non-linearity network and generates the output as feature map. Graph represents the ReLU function:

The MRI scan image is analyzed with no of convolutions and ReLU layers are used to locate the features. Fig. 5 shows the range values among the ReLU layers.

$$R(z) = \max(0, z) \quad (3)$$

2) Pooling layer: This layer is utilized to reduce the dimensions of feature maps. This layer creates the pooled feature map.
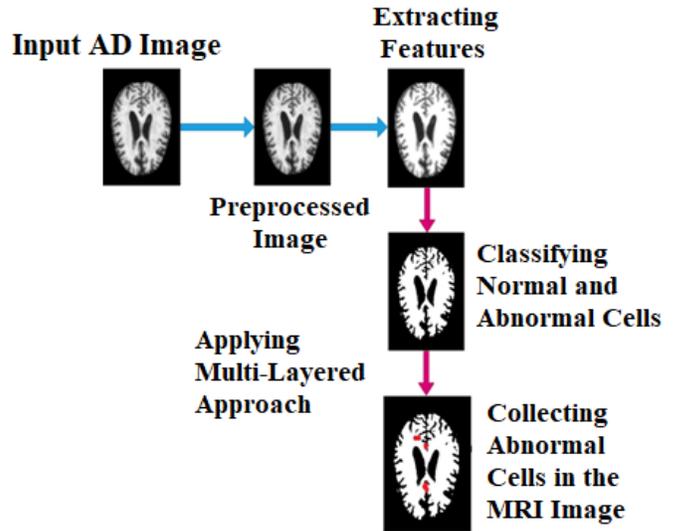
Fig. 4. AD Detection Process using Proposed Approach.
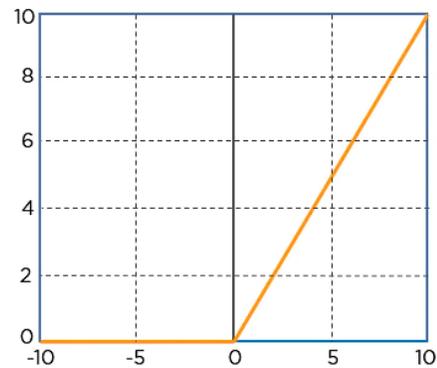
Fig. 5. Showing the Range Values 0-10.

This layer contains several filters that are used to find the significant regions such as normal and abnormal cells in the MRI scan images.

Every neuron is fully connected with all the neurons in the upper layer, the main aim of this layer is to transform the dimensions, and is also used to turn the high-dimensional features into low-dimensional every neuron in this system has activation function to get the final output.

$$h^l_{ab} = \max_{i=0,\dots,s, j=0,\dots,s} h^{l-1}_{(a+i)(b+j)} \quad (4)$$

$z^l \rightarrow$ pre $-$ activation output of layer l

$h^l \rightarrow$ activation of layer l

$* \rightarrow$ discrete convolution operator

$W, \gamma, \beta \rightarrow$ Learnable parameters

B. Steps for Agent Based Reinforcement Classification (ABRC)

The proposed classification is applied on two benchmark datasets.

Step 1: In the initial stage, the agent utilizes the Softmax exploration policy to find the strong values for original classes and low values for the wrong classes that are lies in the utilized exploration policy. The proposed approach is on-policy learning approach. Based on the changed actions the initial state is represented by.

$$P(x) = \frac{e^{\frac{P_i^x(s_t)}{\tau}}}{\sum_b e^{\frac{P_i^y(s_t)}{\tau}}} \tag{5}$$

Step 2: From the above equation the negative values shows the wrong classes and these are passed by TDlearning at the initial state. Here, 1 represents the errors (cancerous/AD affected) and 0 represents the 1 (non-cancerous/AD not effected). The testing approach calculates all the values $V_i(s0)$ for all the classes i and agents belongs to AC_i belongs to these classes. The predicted class y_p belongs to agent with high state value:

$$y_p = \text{argmax}_i V_i(s_0) \tag{6}$$

Output: The result shows that patient have affected with AD or Not and Cancerous tumors or not.

C. Segmentation

Edge-based segmentation is the technique that finds the accurate edges of the brain input images. This technique focused on detecting the edges based on the gray levels, image color, texture, etc. The gray level may change if the system moves from one region to another. If any abnormality is present then finding the edge is easy task. Different functions are used to get the image with edge segmentation and this output should not be confused with the final segmented image.

These edges are mainly connected with the "Magnitude" and "Direction". The edge detection used in this paper used both directions. This approach is represented with the following equations. To calculate the g the 7 or 8 or 9 formulas are used to calculate the g and theta. By using this strong edges are identified based on magnitude and direction.

$$g = (g_a^2 + g_b^2) \tag{7}$$

$$g = |g_a| + |g_b| \tag{8}$$

$$g = \max(|g_a| + |g_b|) \tag{9}$$

$$\theta = \arctan\frac{g_a}{g_b} \tag{10}$$

Where g_a = Magnitude of a

Where g_b = Magnitude of b

V. DATASET DESCRIPTION

For the experimental results, three datasets are used two datasets for brain tumor detection and one for AD detection. Two MRI brain tumor datasets are BraTS-2019 [12] [20] and the Kaggle dataset. Another AD dataset is collected from the Kaggle dataset and this contains 10000 MRI samples with four classes shown in Fig. 1. The training set contains 5000 images and testing contains 5000 images.

VI. PERFORMANCE METRICS

The performance is evaluated by calculating various metrics such as dice score, precision, sensitivity, specificity, accuracy. The segmentation of brain image result is analyzed by using dice score.

$$\text{Dice Score} = \frac{2TP}{FN + FP + 2TP}$$

Specificity: This parameter mainly detects the total number of TP and FP and this is used to calculate the model which is more ability to predict the background area.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Sensitivity This parameter mainly detects the total number of TP and FN and this is used to calculate the model which is more ability to predict the segmented area.

$$\text{Sensitivity or Recall} = \frac{TP}{TP + FN}$$

Accuracy is one metric that measures the reliability of the classification result. The formula is give below.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}$$

Precision: The overall percentage of the output results are defined as:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Table I and Fig. 6 show the performance of training and testing in terms of time. Among all the existing models the proposed model achieved the low training and testing time (Sec).

TABLE I. TRAINING AND TESTING TIME (SEC) FOR MRI IMAGE DATASET

Algorithms	Training Time (Sec)	Testing Time (Sec)
K-means and FCM [21]	74.67	75.78
CNN	72.21	78.89
ILPD	49.89	56.89
ISA	43.12	51.21
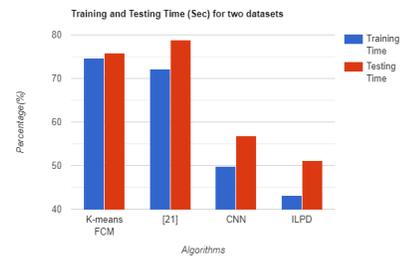
Fig. 6. Training and Testing Time (Sec) for Two Datasets.

TABLE II.   TRAINING AND TESTING TIME (SEC) FOR AD BASED MRI IMAGES

Algorithms	Training Time (Sec)	Testing Time (Sec)
Vanilla DNN	72.34	73.80
CNN	71.89	74.34
CNN-DNN	48.78	54.34
ISA	41.55	49.43
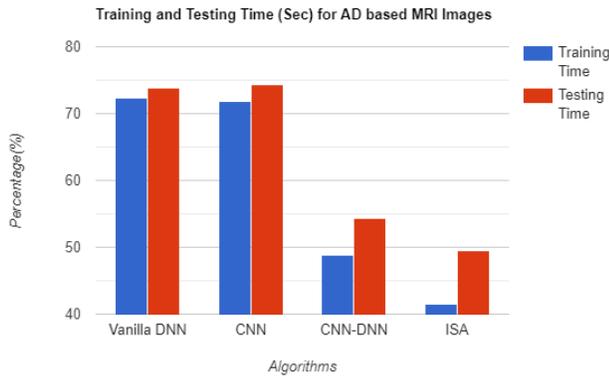
Fig. 7.   Training and Testing Time (Sec) for AD based MRI Images.

Table II and Fig. 7 show the performance of AD based MRI images based on training and testing in terms of computation time (Sec).

TABLE III.   RESULTS OF ON NAVONEEL BRAIN TUMOR IMAGES

Algorithms	Dice Score	Specificity	Sensitivity	Accuracy	Precision
K-means and FCM [29]	56.4%	53.12%	57.56%	53.12%	56.32%
CNN	82.53%	81.98%	84.12%	83.54%	85.43%
ILPD	97.87%	98.12%	97.43%	98.23%	98.67%
ISA	99.12%	99.56%	99.1%	99.67%	99.89%
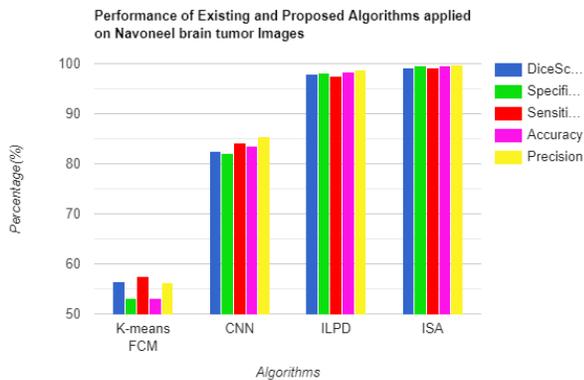
Fig. 8.   Performance of Existing and Proposed Algorithms Applied on Navoneel Brain Tumor Images.

Table III and Fig. 8 show the performance and comparison between existing and proposed algorithms by applying the confusion matrix parameters to analyze the performance. This is applied on navoneel brain tumors datasets.

TABLE IV.   RESULTS ON BRATS 2017 BRAIN TUMOR IMAGES

Algorithms	Dice Score	Specificity	Sensitivity	Accuracy	Precision
K-means and FCM [29]	56.4%	51.42%	56.34%	54.42%	57.42%
CNN	83.63%	82.56%	83.42%	85.74%	86.43%
ILPD	98.87%	99.12%	97.89%	98.45%	99.67%
ISA	99.56%	99.96%	99.78%	99.45%	100%
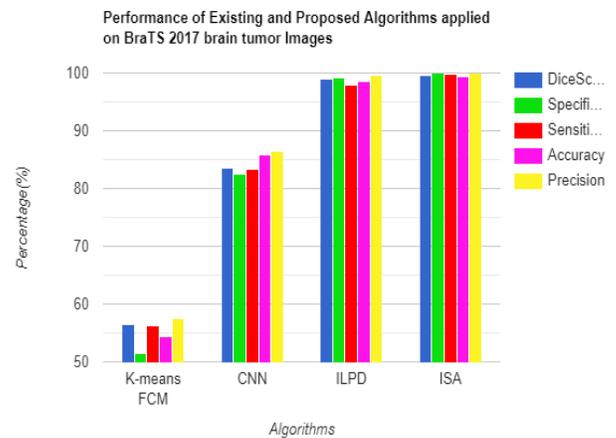
Fig. 9.   Performance of Existing and Proposed Algorithms Applied on BraTS 2017 Brain Tumor Images.

Table IV and Fig. 9 show the performance and comparison between existing and proposed algorithms by applying the confusion matrix parameters to analyze the performance. This is applied on BRATS 2017 brain tumors datasets.

Table V and Fig. 10 show the performance and comparison between existing and proposed algorithms by applying the confusion matrix parameters to analyze the performance. This is applied on AD based MRI image datasets.

TABLE V.   RESULTS ON AD BASED MRI IMAGES

Algorithms	Dice Score	Specificity	Sensitivity	Accuracy	Precision
Vanilla DNN	72.34	75.80	83.12	87.12	90.12
CNN	78.98	83.24	85.34	88.34	91.12
CNN-DNN	82.56	85.64	87.67	91.67	92.34
ISA	89.34	97.87	98.56	99.67	99.45
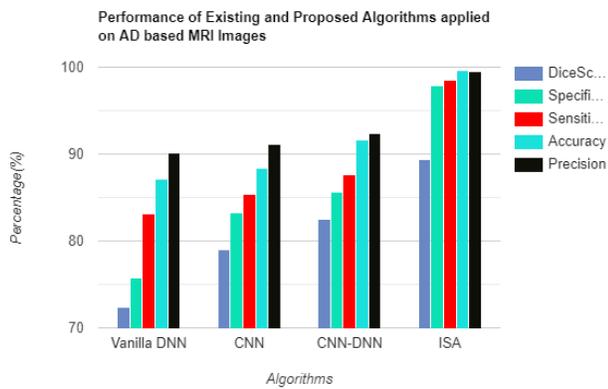
Fig. 10. Graph Representation on AD based MRI Images.

VII. Conclusion

An integrated segmentation approach (ISA) is developed to detect the cancerous cells in MRI images and also work on AD detection. The proposed approach is integrated with various advanced steps to get accurate results. Generally detecting the AD from the MRI scan images is a very difficult task. But the proposed approach becomes very easy to detect the AD and also cancer cells in the MRI scan images. Advanced training and preprocessing approaches improve the performance. The proposed approach increases the performance for the detection of MRI scan images in terms of parameters such as Dice Score-99.12%, Specificity-99.12%, Sensitivity-99.56%, Accuracy-99.1%, Precision-99.67%. AD detection performance is Dice Score-89.34%, Specificity-97.87%, Sensitivity-98.56%, Accuracy-99.67%, Precision-99.45%.

References

[1] Kapoor, L. and Thakur, S. (2017). A Survey on Brain Tumor Detection Using Image Processing Techniques. 7th International Conference on Cloud Computing, Data Science & Engineering.

[2] N. Sengupta, C. B. McNabb, N. Kasabov, and B. R. Russell, "Integrating space, time, and orientation in spiking neural networks: a case study on multimodal brain data modeling," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 11, pp. 5249–5263, 2018.

[3] Tong J., Zhao Y., Zhang P., Chen L., Jiang L. MRI brain tumor segmentation based on texture features and kernel sparse coding. Biomed. Signal Process. Control. 2018; 47:387–392. doi: 10.1016/j.bspc.2018.06.001.

[4] Logeswari T., Karnan M. An improved implementation of brain tumor detection using segmentation based on the hierarchicaself-organizing map. Int. J. Comp. Theory Eng. 2010;2:591. doi: 10.7763/IJCTE.2010.V2.207.

[5] Abdelaziz Ismael SA, Mohammed A, Hefny H. An enhanced deep learning approach for brain cancer MRI images classification using residual networks. Artif Intell Med. 2020 Jan;102:101779.

[6] Bakas, S. et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data 4(1), 1–13. https://doi.org/10.1038/sdata.2017.117 (2017).

[7] Khosravanian, A., Rahmanimanesh, M., Keshavarzi, P. & Mozaffari, S. Fast level set method for glioma brain tumor segmentation based on superpixel fuzzy clustering and lattice boltzmann method. Comput. Methods Programs Biomed. 198, 105809. https://doi.org/10.1016/j.cmpb.2020.105809 (2020).

[8] Tang, Z., Ahmad, S., Yap, P. T. & Shen, D. Multi-atlas segmentation of MR tumor brain images using low-rank based image recovery. IEEE Trans. Med. Imaging 37(10), 2224–2235. https://doi.org/10.1109/TMI.2018.2824243 (2018).

[9] M. A. Wiering, H. van Hasselt, A. Pietersma and L. Schomaker, "Reinforcement learning algorithms for solving classification problems," 2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL), 2011, pp. 91-96, doi: 10.1109/ADPRL.2011.5967372.

[10] Z. Jia and D. Chen, "Brain Tumor Identification and Classification of MRI images using deep learning techniques," in IEEE Access, doi: 10.1109/ACCESS.2020.3016319.

[11] M. S. Majib, M. M. Rahman, T. M. S. Sazzad, N. I. Khan and S. K. Dey, "VGG-SCNet: A VGG Net-Based Deep Learning Framework for Brain Tumor Detection on MRI Images," in IEEE Access, vol. 9, pp. 116942-116952, 2021, doi: 10.1109/ACCESS.2021.3105874.

[12] M. N. Islam, T. Mahmud, N. I. Khan, S. N. Mustafina and A. K. M. N. Islam, "Exploring machine learning algorithms to find the best features for predicting modes of childbirth", IEEE Access, vol. 9, pp. 1680-1692, 2021.

[13] S. Pereira, A. Pinto, V. Alves and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images", IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1240-1251, May 2016.

[14] Shakeel, P. M., Tobely, T. E. E., Al-Feel, H., Manogaran, G., & Baskar, S. (2019). Neural network-based brain tumor detection using wireless infrared imaging sensor. IEEE Access, 7, 5577-5588.

[15] Mallick, P. K., Ryu, S. H., Satapathy, S. K., Mishra, S., Nguyen, G. N., & Tiwari, P. (2019). Brain MRI image classification for cancer detection using deep wavelet autoencoder-based deep neural network. IEEE Access, 7, 46278-46287.

[16] Li, G., Yang, Y. & Qu, X. Deep learning approaches on pedestrian detection in hazy weather. IEEE Trans. Ind. Electron. 67(10), 8889–8899. https://doi.org/10.1109/TIE.2019.2945295 (2020).

[17] Zhang, D. et al. Exploring task structure for brain tumor segmentation from multi-modality MR images. IEEE Trans. IMAGE Process. https://doi.org/10.1109/TIP.2020.3023609 (2020).

[18] Zhou, C., Ding, C., Wang, X., Lu, Z. & Tao, D. One-pass multi-task networks with cross-task guided attention for brain tumor segmentation. IEEE Trans. Image Process. 29, 4516–4529. https://doi.org/10.1109/TIP.2020.2973510 (2020).

[19] Badrinarayanan, V., Kendall, A. & Cipolla, R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615 (2017).

[20] Hu, K. et al. Brain tumor segmentation using multi-cascaded convolutional neural networks and conditional random field. IEEE Access 7, 92615–92629. https://doi.org/10.1109/ACCESS.2019.2927433 (2019).

[21] Almahfud, M.A.; Setyawan, R.; Sari, C.A.; Setiadi, D.R.I.M.; Rachmawanto, E.H. An Effective MRI Brain Image Segmentation using Joint Clustering (K-Means and Fuzzy C-Means). In Proceedings of the 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 21–22 November 2018; pp. 11–16.

[22] K. Muhammad, S. Khan, J. D. Ser and V. H. C. d. Albuquerque, "Deep Learning for Multigrade Brain Tumor Classification in Smart Healthcare Systems: A Prospective Survey," in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 2, pp. 507-522, Feb. 2021, doi: 10.1109/TNNLS.2020.2995800.

[23] M. Rizwan, A. Shabbir, A. R. Javed, M. Shabbir, T. Baker and D. Al-Jumeily Obe, "Brain Tumor and Glioma Grade Classification Using Gaussian Convolutional Neural Network," in IEEE Access, vol. 10, pp. 29731-29740, 2022, doi: 10.1109/ACCESS.2022.3153108.

[24] A. Saleh, R. Sukaik and S. S. Abu-Naser, "Brain Tumor Classification Using Deep Learning," 2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech), 2020, pp. 131-136, doi: 10.1109/iCareTech49914.2020.00032.

[25] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, M. Imran and M. Shoaib, "A Deep Learning Model Based on Concatenation Approach for the Diagnosis of Brain Tumor," in IEEE Access, vol. 8, pp. 55135-55144, 2020, doi: 10.1109/ACCESS.2020.2978629.

[26] Zhou SK, Le HN, Luu K, Nguyen HV, Ayache N. Deep reinforcement learning in medical imaging: A literature review. Medical image analysis. 2021 Oct 1;73:102193.

[27] Chao Yu, Jiming Liu, Shamim Nemati, and Guosheng Yin. 2021. Reinforcement Learning in Healthcare: A Survey. ACM Comput. Surv. 55, 1, Article 5 (January 2023), 36 pages. https://doi.org/10.1145/3477600.

[28] Al Walid Abdullah and Yun Il Dong. 2019. Partial policy-based reinforcement learning for anatomical landmark localization in 3d medical images. IEEE Transactions on Medical Imaging 39, 4 (2019), 1245–1255.

[29] Hao D, Yang G, Liu F, Mo Y, Guo Y. Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In: Annual conference on medical image understanding and analysis. Springer, Cham; 2017.

[30] Mallikarjuna A. Reddy, Sudheer K. Reddy, Santhosh C.N. Kumar, Srinivasa K. Reddy, "Leveraging bio-maximum inverse rank method for iris and palm recognition", International Journal of Biometrics, 2022 Vol.14 No.3/4, pp.421 - 438, doi: 10.1504/IJBM.2022.10048978.

[31] A. M. Reddy, V. V. Krishna, L. Sumalatha and S. K. Niranjan, "Facial recognition based on straight angle fuzzy texture unit matrix," 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, 2017, pp. 366-372, doi: 10.1109/ICBDACI. 2017. 80 70865.

[32] Ilaiah Kavati, A. Mallikarjuna Reddy, E. Suresh Babu, K. Sudheer Reddy, Ramalinga Swamy Cheruku,Design of a fingerprint template protection scheme using elliptical structures,ICT Express,Volume 7, Issue 4,2021,Pages 497-500,ISSN 2405-9595,https://doi.org/ 10.1016/j.icte. 2021.04.001.

[33] Sudeepthi Govathoti, A Mallikarjuna Reddy, Deepthi Kamidi, G BalaKrishna, Sri Silpa Padmanabhuni and Pradeepini Gera, "Data Augmentation Techniques on Chilly Plants to Classify Healthy and Bacterial Blight Disease Leaves" International Journal of Advanced Computer Science and Applications (IJACSA), 13(6), 2022. http://dx.doi.org /10.14569/ IJACSA.2022.0130618.

[34] C. R. T, G. Sirisha and A. M. Reddy, "Smart Healthcare Analysis and Therapy for Voice Disorder using Cloud and Edge Computing," 2018 4th International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Mangalore, India, 2018, pp. 103-106, doi: 10.1109/iCATccT44854. 2018. 90 01280.

[35] A. M. Reddy, K. SubbaReddy and V. V. Krishna, "Classification of child and adulthood using GLCM based on diagonal LBP," 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Davangere, 2015, pp. 857-861, doi: 10.1109/ICATCCT.2015.7457003.

[36] Ayaluri MR, K. SR, Konda SR, Chidirala SR. 2021. Efficient steganalysis using convolutional to autoencoder network to ensure original image quality. PeerJ Computer Science 7:e356 https://doi.org/10.7717/peerj-cs.356.

[37] A. Mallikarjuna Reddy, K. S. Reddy, M. Jayaram, N. Venkata Maha Lakshmi, Rajanikanth Aluvalu, T. R. Mahesh, V. Vinoth Kumar, D. Stalin Alex, "An Efficient Multilevel Thresholding Scheme for Heart Image Segmentation Using a Hybrid Generalized Adversarial Network", Journal of Sensors, vol. 2022, Article ID 4093658, 11 pages, 2022. https://doi.org/10.1155/2022/4093658.

Combining AHP and Topsis to Select Eligible Social and Solidarity Economy Actors for a Call for Grants

Salma Chrit, Abdellah Azmani, Monir Azmani
Intelligent Automation Laboratory, FST of Tangier
Abdelmalek Essaadi University
Tetouan, Morocco

Abstract—The procedure for selecting projects in order to offer a grant for actors of the social and solidarity economy can be a delicate task for decision-makers (Public or Private Establishments), which is based on several eligibility and refusal criteria (economic, social and environmental); the task that can sometimes take several months before returning the results. This study proposes an integrated framework based on two multi-criteria decision methods, analytical hierarchy process (AHP) and technique for order performance by similarity to an ideal solution (TOPSIS), to select and rank viable projects to obtain a grant from the INDH (National Initiative for Human Development). Initially, the projects were randomly selected from a list of submitted projects to receive a grant. Later, AHP obtains weights of various criteria through pairwise comparison, and projects are ranked using TOPSIS. The proposed methodology is empirically applied to the social and solidarity economy sector and provides a detailed and effective decision-making tool for selecting suitable actors to obtain a grant. The results indicate that the conservation of natural resources and the rate of job creation are the essential criteria in the process of selecting projects.

Keywords—AHP; TOPSIS; project selection; decision making; multi criteria decision method

I. INTRODUCTION

To create wealth, develop the economic fabric of the country, and create new jobs, the state has implemented several initiatives to support entrepreneurship and the creation of new economic structures; these initiatives focus in particular on actors of the Social and Solidarity Economy who are present in all sectors of economic activity, from energy to culture and food.

Launched on May 18, 2005, by His Majesty King Mohammed VI, the National Initiative for Human Development aimed to fight against poverty and social exclusion [1]; during the first two phases, the INDH is focused on reducing the socio-economic deficit and especially the struggle against poverty in rural areas, exclusion in urban areas and precariousness with an envelope of 28 billion MAD between 2005 and 2018 [2], and has been classified according to the World Bank in 2015, among the top three best action programs and initiatives of general interest in the world [3].

By carrying out 43,000 human development projects, the INDH has simplified access to various resources, such as access to essential services (8,200 km of tracks and roads, 230,000 households served with drinking water, 60,000 households connected to electricity), access to care (519 health centers, 240 Dar al Oumouma, 1,150 ambulances, 560 medical caravans & health campaigns), economic inclusion (9,400 income-generating activities - 64% in rural areas, where 50% of beneficiaries are women) … [3].

Launched on September 19, 2018, the 3rd phase aims to improve the economic and social situation of the population, particularly young people between 18 and 35 in vulnerable situations, through support and help for employment or entrepreneurship, with a budget of four billion dirhams.

However, each year several projects are submitted across the country to obtain a grant to help these actors start their businesses. The selection procedure becomes more and more complex, with the number of applications, the similarity between the projects, and the degree of innovation.

This implies that the response waiting time increases in parallel with the rise in the number of files submitted, this delay in response can have significant consequences for actors in the social and solidarity economy such as cooperatives, including the activity is often based on these grants and the amount offered.

The main objective of this study is to set up a decision-making tool which allows donors, investors, and public institutions to select and classify projects according to several economic, environmental, and social criteria, and this is through the combination between the AHP method and TOPSIS.

II. INSIGHT: CLASSICAL BUSINESSES VS. SOCIAL BUSINESSES

While classical businesses have several advantages and receive great public awareness nationally and internationally, the latter has the main objective of maximizing the profit of its shareholders ([4], [5], [6]) by putting set up a service or production project with specific customers and within a competitive market. Despite being considered a business, the social enterprise has a social mission that affects all aspects of its management, whose profits are a means of achieving its missions and not an end [7].

The creation of social enterprises plays a vital role in the development and revitalization of the economic fabric [8], as it makes it possible to reduce the unemployment rate, create new jobs, and reduce the informal economy, inactivity of young people, and to promote the spirit of solidarity ([9],[10]).

To establish an entrepreneurial culture among young people, and to create new innovative projects in promising sectors, to ensure integration and social growth, the governments of countries (industrialized and developing), in particular Morocco, have ensured the implementation of various public policies and initiatives, mainly for people in a situation of vulnerability (economic and social exclusion), whose central role is to provide these young people, in a case of unemployment, underemployment or those who carry out entrepreneurial activities (formal and informal), resources, means of financing, accompaniment and support [11].

These initiatives follow several stages in order to sort out the eligible projects for funding, support, and implementation, which can sometimes take up to six months or even one year.

III. THE TRADITIONAL PROJECT SELECTION

The selection of eligible projects is made on four levels [12]:

- The launch of a call for applications: this call aims to determine the ideas for innovative projects as well as the potential beneficiaries.

- The selection of project ideas: the sorting and ranking of project ideas according to two main criteria, innovation and feasibility.

- Support and strengthening of project leaders: this stage aims to carry out the studies necessary to identify the feasibility, viability, and desirability of the project, as well as the establishment of market studies, the business model, and the business plan.

- Project selection: during this phase, to select the projects eligible and viable for financing, a committee of different actors and experts familiar with the local economic context is formed.

- The selection criteria are split into two categories:

- The eligibility criteria are economic, social, and environmental.

- The criteria for refusal include the risk to the environment, the professional status of the project leader, and obtaining public aid.

IV. THE DECISION MAKING

Roy [13] (1985) defines decision support as:

"Decision support is the activity of someone who, based on models that are clearly explained but not necessarily completely formalized, helps to obtain elements of answers to the questions posed by a participant in a decision-making process, elements to favor a behavior likely to increase the coherence between the evolution of the process, on the one hand, the objectives and the system of values in the service of which this participant finds himself placed on the other hand."

The objective of decision support is to guide and enlighten the decision-maker throughout the decision-making process. However, any decision-making activity is based on two main actors, a decision-maker and a man of study; the first is always present during the process, and the second, he is only present when the decision-maker calls on him [14].

These actors are not necessarily unique characters. The decision maker can be a board of directors, the managers of a company, a senior executive or manager, an individual, etc. Furthermore, the researcher can be made up of a team (engineers, mathematicians, economists, sociologists, doctors, etc.), and he is responsible for defining the decision model, the process, and the presentation of the results to the decision maker [15].

With each call for applications, several projects are submitted, even hundreds that span the whole country, making selecting and choosing eligible projects a mission that can take months to have a final answer.

In the case of the INDH ([11], [12]), and during the selection of projects, the first step consists in launching a call for the appointment of a service provider who will play the role of decision-maker and whose main tasks are the study and the sorting of ideas, the organization of support, reinforcement and orientation sessions for the selected projects. A committee composed of public and private actors proceeds to select viable projects; these projects must be subsequently confirmed by the PCHD (Provincial Committee for Human Development).

Several articles address the subject of project selection in the literature; however, when the research deepens in the literature, most works treat the subject either with a theoretical vision or they do not treat the subject as a whole and are interested in only one component, [16] determines the criteria that Istanbul technoparks companies take into consideration in their selection preferences and to select the most appropriate technopark based on these criteria, [17] combined the three methods (Delphi, AHP and TOPSIS) to provide decision methods to managers of projects in construction companies, [18] expose the selection of telecommunication projects by first using the Delphi method to convert the qualitative criteria into quantitative criteria and then, by applying the TOPSIS method.

Whereas [19] worked on combining Shannon Entropy to determine the weight of the criteria to apply the TOPSIS method thereafter to select eligible projects. [20], talk about the project selection, and evaluation process in Ivory Coast, the latter which follows the same process as mentioned in Section III, and this through several committees at each level to make a good selection, [21], in turn, discusses the process of selection of young project leaders, the expression decision-making committee and selection policy appear several times, and offer a grid of indicators that allow project leaders to be given a final note/score; the study [22], illustrates the project selection process by the BA from entrepreneurs' pitches and the decision to invest, and this selection in turn follows a process that spans four stages, and proposes an approach for assessment of investment criteria.

V. APPLICATION OF THE AHP AND TOPSIS METHODS

In order to be able to select the projects eligible for a grant, the study focused on three families of criteria, the economic, social and environmental criteria, taking into consideration the

eliminatory criteria at the start [12]. Table I presents the project eligibility criteria and sub-criteria, based on the three main criteria.

TABLE I.   PROJECT ELIGIBILITY CRITERIA

Criteria	Sub criterion
Economic	Creation of added values
	Funding requirement
	Annual budget
	Financial resources
	Number of partners
	Execution period
	Achievement amount
Social	Job creation
	Working conditions
	Improvement of the status of women
	Respect for the rights of the child
	Activity area
	Number of beneficiaries
Environmental	Conservation of natural resources
	Maintaining biodiversity
	Environmental risk management
	Production methods
	Environmental impact

The refusal criteria designate the criteria that lead to the elimination of the file from the start, namely:

- Environmental risk: any project that has a negative impact on the environment, either through the use of chemicals, the alteration of natural resources or which has harmful effects on health;

- The carrier's profession: any project whose tenderer is a civil servant, an agent of a public institution or a private sector employee is automatically eliminated;

- Obtaining public aid: any project whose initiator has already benefited from public funding, either individually or within the framework of a group, is not eligible.

In this study, only three sub-criteria per criteria (Table III) were selected, which are frequently found in the literature to illustrate the approach.

The distinction between the favorable and unfavorable criteria (Negative and Positive), was made based on the literature, and on the opinion of the experts, Table II, presents the criteria by nature (favorable and unfavorable), for each favorable criterion, the higher the score, the more positive the criterion, likewise for each unfavorable criterion, the higher the value, the negative criterion.

The input matrix is made up of ten projects, which were randomly selected from a list of projects submitted to obtain a grant (INDH: National Initiative for Human Development), the values found in this matrix are obtained by the evaluation

of the projects, each project has a value per sub-criterion, as shown in Table III.

Table V presents the pairwise comparison matrix (AHP), which is based on the judgment of the decision makers on the importance of each criterion, by combining the literature with the opinion of several experts to determine the pairwise comparison between different families of criteria ([25], [26], [27]).

After having built the input matrix, the article proceeds by using the AHP method, to have consistent weights of the criteria, for this, the first step is to establish the comparison matrix by pair, based on the Saaty scale ([23],[24]) present in Table IV.

TABLE II.   PROJECT ELIGIBILITY CRITERIA'S DESCRIPTION

Criteria	Sub-criterion	Coded	Nature	Comment
Economic	Funding requirement	Ec1	Unfavorable	The amount that the project leader requests in order to start his project.
	Financial resources	Ec2	Favorable	This criterion presents the own financial resources available to the project leader.
	Achievement amount	Ec3	Unfavorable	The Achievement Amount refers to the total cost of the project.
Social	Job creation	S1	Favorable	The number of new jobs that will be created following the completion of the project.
	Improvement of the status of women	S2	Favorable	Does the project take the gender-equality approach into consideration and propose solutions to help women?
	Number of beneficiaries	S3	Unfavorable	The number of people who will benefit from the project once implemented.
Environmental	Conservation of natural resources	En1	Favorable	Does the project make rational use of natural resources and aims to enhance them?
	Maintaining biodiversity	En2	Favorable	The protection of spaces and environments.
	Environmental impact	En3	Unfavorable	Does the project have a positive impact on the environment or not?

TABLE III. INPUT MATRIX

	N	P	N	P	P	N	P	P	N
	Ec1	Ec2	Ec3	So1	So2	So3	En1	En2	En3
Project 1	4	2	8	4	3	2	5	3	4
Project 2	2	3	6	5	1	5	2	1	1
Project 3	1	4	6	7	6	8	1	2	3
Project 4	5	5	5	3	2	3	6	5	2
Project 5	4	4	3	7	4	2	8	4	5
Project 6	5	8	7	5	2	3	6	5	2
Project 7	2	6	2	2	1	4	6	7	1
Project 8	3	1	4	4	5	5	5	3	6
Project 9	7	6	5	1	4	4	3	7	8
Project 10	8	3	6	8	5	8	7	5	2

$$r_{ij} = \frac{x_{ij}}{\sqrt{\Sigma_{i=1}^{n} x_{ij}^2}} \qquad (1)$$

Table VI shows the weighting of the criteria, again according to the AHP method [23].

This step consists in calculating the weight of each criterion, the following 3 steps summarize the procedure:

1) Sum of the values of each column.

2) Divide each element of the matrix by its column total.

3) Calculation of the average of the elements of each row of the matrix.

The central idea behind the TOPSIS method is to choose a solution that comes closest to the ideal solution (best on all criteria) and to move away as much as possible from the worst solution (which degrades all criteria) [28].

The normalization of the matrix is done based on the Euclidean distance, to obtain values between 0 and 1 according to equation 1, as indicated in Table VII.

The study then proceeds with the multiplication of the entries of the matrix by the weights associated with the criteria using Eq. 2, Table VIII exposes the results obtained.

$$R_{ij} = w_{ij} \times x_{ij} \qquad (2)$$

TABLE IV. SAATY'S SCALE

Verbal scale	Numerical scale
Both elements are equal	1
A little more important	3
Most important	5
much more important	7
Absolutely more important	9

TABLE V. PAIRWISE COMPARISON MATRIX

	Ec1	Ec2	Ec3	So1	So2	So3	En1	En2	En3
E1	1.00	3.00	5.00	1.00	1.00	1.00	1.00	1.00	1.00
E2	0.33	1.00	3.00	1.00	1.00	1.00	1.00	1.00	1.00
E3	0.20	0.33	1.00	1.00	1.00	1.00	1.00	1.00	1.00
S1	1.00	1.00	1.00	1.00	5.00	5.00	1.00	1.00	1.00
S2	1.00	1.00	1.00	0.20	1.00	2.00	1.00	1.00	1.00
S3	1.00	1.00	1.00	0.20	0.50	1.00	1.00	1.00	1.00
In 1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	7.00	5.00
In 2	1.00	1.00	1.00	1.00	1.00	1.00	0.14	1.00	3.00
En3	1.00	1.00	1.00	1.00	1.00	1.00	0.20	0.33	1.00
Sum	7.53	10.33	15.00	7.40	12.50	14.00	7.34	14.33	15.00

TABLE VI. WEIGHT OF CRITERIA

	Ec1	Ec2	Ec3	So1	So2	So3	En1	En2	En3	Sum	S/N
E1	0.13	0.29	0.33	0.14	0.08	0.07	0.14	0.07	0.07	1.32	0.15
E2	0.04	0.10	0.20	0.14	0.08	0.07	0.14	0.07	0.07	0.90	0.10
E3	0.03	0.03	0.07	0.14	0.08	0.07	0.14	0.07	0.07	0.68	0.08
S1	0.13	0.10	0.07	0.14	0.40	0.36	0.14	0.07	0.07	1.46	0.16
S2	0.13	0.10	0.07	0.03	0.08	0.14	0.14	0.07	0.07	0.82	0.09
S3	0.13	0.10	0.07	0.03	0.04	0.07	0.14	0.07	0.07	0.71	0.08
In 1	0.13	0.10	0.07	0.14	0.08	0.07	0.14	0.49	0.33	1.54	0.17
In 2	0.13	0.10	0.07	0.14	0.08	0.07	0.02	0.07	0.20	0.87	0.10
En3	0.13	0.10	0.07	0.14	0.08	0.07	0.03	0.02	0.07	0.70	0.08

TABLE VII.   NORMALIZED MATRIX BY CRITERION (ATTRIBUTE)

Weight	0.15	0.10	0.08	0.16	0.09	0.08	0.17	0.10	0.08
Nature	N	P	N	P	P	N	P	P	N
	Ec1	Ec2	Ec3	So1	So2	So3	En1	En2	En3
Project 1	0.27	0.14	0.46	0.25	0.26	0.13	0.30	0.21	0.31
Project 2	0.14	0.20	0.35	0.31	0.09	0.33	0.12	0.07	0.08
Project 3	0.07	0.27	0.35	0.44	0.51	0.52	0.06	0.14	0.23
Project 4	0.34	0.34	0.29	0.19	0.17	0.20	0.36	0.34	0.16
Project 5	0.27	0.27	0.17	0.44	0.34	0.13	0.47	0.27	0.39
Project 6	0.34	0.54	0.40	0.31	0.17	0.20	0.36	0.34	0.16
Project 7	0.14	0.41	0.12	0.12	0.09	0.26	0.36	0.48	0.08
Project 8	0.21	0.07	0.23	0.25	0.43	0.33	0.30	0.21	0.47
Project 9	0.48	0.41	0.29	0.06	0.34	0.26	0.18	0.48	0.62
Project 10	0.55	0.20	0.35	0.50	0.43	0.52	0.41	0.34	0.16

TABLE VIII.   NORMALIZED AND WEIGHTED MATRIX

Weight	0.15	0.10	0.08	0.16	0.09	0.08	0.17	0.10	0.08
Nature	N	P	N	P	P	N	P	P	N
	Ec1	Ec2	Ec3	So1	So2	So3	En1	En2	En3
Project 1	0.04	0.01	0.04	0.04	0.02	0.01	0.05	0.02	0.02
Project 2	0.02	0.02	0.03	0.05	0.01	0.03	0.02	0.01	0.01
Project 3	0.01	0.03	0.03	0.07	0.05	0.04	0.01	0.01	0.02
Project 4	0.05	0.03	0.02	0.03	0.02	0.02	0.06	0.03	0.01
Project 5	0.04	0.03	0.01	0.07	0.03	0.01	0.08	0.03	0.03
Project 6	0.05	0.05	0.03	0.05	0.02	0.02	0.06	0.03	0.01
Project 7	0.02	0.04	0.01	0.02	0.01	0.02	0.06	0.05	0.01
Project 8	0.03	0.01	0.02	0.04	0.04	0.03	0.05	0.02	0.04
Project 9	0.07	0.04	0.02	0.01	0.03	0.02	0.03	0.05	0.05
Project 10	0.08	0.02	0.03	0.08	0.04	0.04	0.07	0.03	0.01

For each criterion, the most favorable associated value A+ is calculated, by applying Eq. 3, depending on the nature of the criterion (favorable, unfavorable).

- Favorable criterion: the maximum value of each column.

- Unfavorable criterion: the minimum value of each column.

$$A^+ = \left\{ \max x_{ij}(i \in J^+) \,\middle|\, \min x_{ij} (i \in J^-) \right\} \quad (3)$$

Table IX and X, show the results of calculation of the ideal favorable solution A+ and the least favorable solution A-.

For each criterion, the least favorable associated value A- is calculated, by applying equation 4, depending on the nature of the criterion (favorable, unfavorable).

- Favorable criterion: the minimum value of each column.

- Unfavorable criterion: the maximum value of each column.

$$A^- = \left\{ \min x_{ij}(i \in J^+) \,\middle|\, \max x_{ij} (i \in J^-) \right\} \quad (4)$$

All the deviations are expressed by the vector E+, Each deviation is expressed as a Euclidean distance between the value of each associated criterion and the associated value of A+ , following Eeq.5.

The calculation of E- is done in the same way as E+. E- is expressed as the Euclidean distance between the value of each associated criterion and the associated value of A-, following Eq. 6, the results are shown in Table XI.

$$E^+ = \sqrt{\sum_{i=0}^{n}(A^+ - r_{ij})^2} \quad (5)$$

$$E^- = \sqrt{\sum_{i=0}^{n}(A^- - r_{ij})^2} \quad (6)$$

The last step consists in calculating the coefficient associated with each alternative, which determines its rank. Each coefficient is calculated from the components associated with the vectors E- and E+ according to the quotient expressed in Eq. 7.

$$C_p = \frac{E^-}{E^- + E^+} \quad (7)$$

TABLE IX.   CALCULATION OF THE IDEAL FAVORABLE SOLUTION A+

Weight	0.15	0.10	0.08	0.16	0.09	0.08	0.17	0.10	0.08
Nature	N	P	N	P	P	N	P	P	N
	Ec1	Ec2	Ec3	So1	So2	So3	En1	En2	En3
Project 1	0.040	0.014	0.035	0.040	0.023	0.010	0.051	0.020	0.024
Project 2	0.020	0.020	0.026	0.051	0.008	0.026	0.020	0.007	0.006
Project 3	0.010	0.027	0.026	0.071	0.047	0.041	0.010	0.013	0.018
Project 4	0.050	0.034	0.022	0.030	0.016	0.015	0.061	0.033	0.012
Project 5	0.040	0.027	0.013	0.071	0.031	0.010	0.081	0.027	0.030
Project 6	0.050	0.054	0.031	0.051	0.016	0.015	0.061	0.033	0.012
Project 7	0.020	0.041	0.009	0.020	0.008	0.020	0.061	0.047	0.006
Project 8	0.030	0.007	0.018	0.040	0.039	0.026	0.051	0.020	0.036
Project 9	0.070	0.041	0.022	0.010	0.031	0.020	0.030	0.047	0.049
Project 10	0.080	0.020	0.026	0.081	0.039	0.041	0.071	0.033	0.012
A+	0.010	0.054	0.009	0.081	0.047	0.010	0.081	0.047	0.006

TABLE X.   CALCULATION OF THE IDEAL FAVORABLE SOLUTION A-

Weight	0.15	0.10	0.08	0.16	0.09	0.08	0.17	0.10	0.08
Nature	N	P	N	P	P	N	P	P	N
	Ec1	Ec2	Ec3	So1	So2	So3	En1	En2	En3
Project 1	0.040	0.014	0.035	0.040	0.023	0.010	0.051	0.020	0.024
Project 2	0.020	0.020	0.026	0.051	0.008	0.026	0.020	0.007	0.006
Project 3	0.010	0.027	0.026	0.071	0.047	0.041	0.010	0.013	0.018
Project 4	0.050	0.034	0.022	0.030	0.016	0.015	0.061	0.033	0.012
Project 5	0.040	0.027	0.013	0.071	0.031	0.010	0.081	0.027	0.030
Project 6	0.050	0.054	0.031	0.051	0.016	0.015	0.061	0.033	0.012
Project 7	0.020	0.041	0.009	0.020	0.008	0.020	0.061	0.047	0.006
Project 8	0.030	0.007	0.018	0.040	0.039	0.026	0.051	0.020	0.036
Project 9	0.070	0.041	0.022	0.010	0.031	0.020	0.030	0.047	0.049
Project 10	0.080	0.020	0.026	0.081	0.039	0.041	0.071	0.033	0.012
A-	0.080	0.007	0.035	0.010	0.008	0.041	0.010	0.007	0.049

TABLE XI.   CALCULATION OF THE DEVIATION OF THE A+ AND A-

E+	E-
0.08607238	0.07852778
0.0976075	0.08748451
0.09157769	0.10753217
0.07979439	0.08687467
0.05473111	0.11591278
0.06800374	0.10073323
0.07738846	0.10934368
0.0850532	0.08311161
0.11704446	0.06632803
0.08775722	0.1094154

TABLE XII.   CALCULATION OF THE PROXIMITY COEFFICIENT OF THE IDEAL SOLUTION AND RANKING IN ORDER OF CHOICE

	CP	Order of choice	Distribution of coefficients
Project 5	0.68	1	0.129
Project 6	0.60	2	0.113
Project 7	0.59	3	0.111
Project 10	0.55	4	0.105
Project 3	0.54	5	0.102
Project 4	0.52	6	0.099
Project 8	0.49	7	0.094
Project 1	0.48	8	0.090
Project 2	0.47	9	0.089
Project 9	0.36	10	0.068

The proximity coefficient of each alternative, as its name suggests, measures the rate of proximity of the most unfavorable ideal solution A- compared to the most favorable ideal solution A+. It is a matter of choosing the one which is the furthest possible from the unfavorable ideal solution A- and the closest to the favorable ideal solution A+.

Note that the most favorable and worst solutions are fictitious and do not represent real alternatives. They constitute benchmarks for comparing the distances of all the alternatives.

VI. DISCUSSION

In order to have a coherent result, we opted for the use of AHP in order to calculate the weights of the criteria, after having determined the weight of the criteria using AHP, we proceeded to the normalization and the weighting of the matrix according to the TOPSIS method, then the calculation of the ideal favorable solution A+ and A-, then we calculated for each alternative, its deviation from the most favorable value, and the last step consisted in calculating the coefficient associated with each alternative, which determines its rank.

The results exposed in Table XII, the outcome obtained at the end of the application of the TOPSIS method, are ranked according to the proximity of the alternative to the ideal solution, so the projects are ranked in order of preference with project five being ranked first, project six and seven in second and third place respectively.

The combination of these two methods can have different applications, depending on the sector of activity, in this article, the emphasis has been placed on the criteria considered by the INDH, the same approach can also be applied to other set of criteria or even add more criteria to have a more specific result, or even to distribute a grant according to the order obtained at the end.

VII. CONCLUSION

The article exposed the use of the AHP method combined with the TOPSIS method in order to make the choice between 10 projects eligible for obtaining a grant. These methods, despite being simple in their use and facilitating decision-making among several alternatives, have some limits, in particular for the quantification of the parameters by the experts or the men of studies, especially at the level of the matrix of comparison of the AHP method.

In order to make this approach more solid, the consideration of combining a machine learning method, or using a method like BOCR (Benefit, Opportunity, Cost, and Risk) is in perspective, which allows the results to be more focused on the aspects of risk, benefits, opportunity and cost, which in the case of the selection of financing projects, can give a better result, similarly, in cases where we want to share the grant between the different actors according to their ranking and the importance of their project.

ACKNOWLEDGMENTS

REFERENCES

[1] A. El Bouzaidi, " Les processus participatifs de l'Initiative Nationale pour le Développement Humain (INDH) : Instruments de résistance contre la pauvreté et l'exclusion sociale ? Le cas de la ville de Salé", Revue GéoDév.ma, Volume 9 (2021).

[2] Tahir A., Moustaquim R., "l'Initiative Nationale pour le Développement Humain (INDH) : outils de lutte contre la pauvreté et l'exclusion sociale au Maroc ? Cas de la province de Béni-Mellal ", International Social Sciences & Management Journal (ISSM), 01| 2019.

[3] Royaume du Maroc, "Initiative National pour le Développement Humain, présentation de la phase III 2019 – 2023 ».

[4] Julien, P.A. & M. Marchesnay, " L'entrepreneurial, Paris, Economica", collection Gestion/Poche, 1996.

[5] Verstraete T., Saporta B., " Création d'entreprise et entrepreneuriat ", Editions de l'ADREG, janvier 2006 (http://www.adreg.net).

[6] BRUYAT C., " Création d'entreprise : contributions épistémologiques et modélisation, Thèse de doctorat en sciences de gestion ", Université Pierre-Mendès-France de Grenoble, 1993, 431 p.

[7] R. BOURJIM, M. M'HAMEDI, "Le concept de l'Entrepreneuriat Social : Essai de définition consensuelle", International Journal of Accounting, Finance, Auditing, Management and Economics, 2022, 3(4-1), 446–461. https://doi.org/10.5281/zenodo.6613893.

[8] Mattock B., Nasroun N., " Entrepreneuriat et création d'entreprises. Facteurs déterminant l'esprit d'entreprise : cas de Béjaïa ", N° 14 - Janvier-Juin 2013 • La Responsabilité sociale des entreprises et les PME, Management & Sciences Sociales, pp. 83-98.

[9] Dhekra ElHidri, " L'Economie Sociale et Solidaire : Un Levier pour une Révolution Economique ", C·A·Perspectives on Tunisia No. 03-2017.

[10] Michelacci C., Suarez J., "Business Creation and the Stock Market," Review of Economic Studies, 2014, 71, issue 2, p. 459-481.

[11] Royaume du Maroc, " Initiative National pour le Développement Humain (Aout 2005), Programme de lutte contre l'exclusion sociale en milieu urbain, Programme de lutte contre la pauvreté en milieu urbain, Programme de lutte contre la précarité ".

[12] Procédure de sélection de l'Initiative National pour le Développement Humain.

[13] Roy, B., Méthodologie multicritère d'aide à la décision, Economica, Paris, 1985.

[14] Zopounidis, C. et M. Doumpos, "Multicriteria classification and sorting methods: A literature review". European Journal of Operational Research (138), 2002, 229–246.

[15] R. Roy, Ed., "Strategic Decision Making", London: Springer London, 2004.

[16] İ. Durak, H. M. Arslan, Y. Özdemir, " Application of AHP–TOPSIS methods in technopark selection of technology companies: Turkish case", Technology Analysis & Strategic Management, 34:10, 1109-1123, 2021, DOI: 10.1080/09537325.2021.1925242.

[17] Prapawan P., "Application of the Multi Criteria Decision Making Methods for Project Selection," Universal Journal of Management 3(1) : 15-20, 2015.

[18] J. Dodangeh, M. Mojahed, R. bt Mohd Yusuff, "Best Project selection by using of Group TOPSIS Method", International Association of Computer Science and Information Technology, Spring Conference, 2009.

[19] A. K. Haddadha, A. Namazian, S. H. Yakhchali, "Project Selection Problem by Combination of Shannon Entropy and MCDM Techniques", International Conference on Literature, History, Humanities and Social Sciences (LHHSS-17) Jan. 1-2, 2017 Dubai (UAE).

[20] A. Ouattara et Y. Sangaré, "Soutenir la recherche en Côte d'Ivoire : Processus de sélection et d'évaluation des projets", Transformer l'excellence en recherche Nouvelles idées des pays du Sud Global, 2021, pp. 157.

[21] Bouarara, K., & Haddad, M , "Politiques de sélection des jeunes porteurs de projets : Cas du programme ARIEJ", International Journal of

Accounting, Finance, Auditing, Management and Economics, 2021, 2(6-1), 94-111. https://doi.org/10.5281/zenodo.5730560.

[22] Bellier, A. & Idi Cheffou, A., "Évolution des critères d'investissement des business angels : de la présélection des projets à l'investissement final", Revue internationale P.M.E., 2020, 33(3-4), 169–197. https://doi.org/10.7202/1074813ar.

[23] Saaty, T. L., "Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World", Third Revised Edition, Pittsburgh: 2012, RWS Publications.

[24] O. S. Vaidya and S. Kumar, "Analytic hierarchy process: An overview of applications", Eur. J. Oper. Res., vol. 169, no. 1, pp. 1–29, Feb. 2006.

[25] Parvaneh, F., & El-Sayegh, S. M., "Project selection using the combined approach of AHP and LP", Journal of Financial Management of Property and Construction, 2016, 21(1), 39–53. doi:10.1108/jfmpc-09-2015-0034.

[26] M. Marzouk, M. Sabbah, "AHP-TOPSIS social sustainability approach for selecting supplier in construction supply chain", Cleaner Environmental Systems, 2021, Volume 2, 100034, ISSN 2666-7894.

[27] Ö. Ekmekcioğlu, K. Koc, M. Özger, "Stakeholder perceptions in flood risk assessment: A hybrid fuzzy AHP-TOPSIS approach for Istanbul, Turkey", International Journal of Disaster Risk Reduction, Volume 60, 2021, 102327, ISSN 2212-4209.

[28] Ewa, R. (2011) Multi-criteria decision-making models by applying the TOPSIS method to crisp and interval data. In: Trzaskalik, T. and Wachowicz, T., Eds., Multiple Criteria Decision Making, The University of Economics, Katowice, pp. 200-230.

Classification of Electromyography Signal of Diabetes using Artificial Neural Networks

Muhammad Fathi Yakan Zulkifli, Noorhamizah Mohamed Nasir

Faculty of Electric and Electronic Engineering
Universiti Tun Hussein Onn Malaysia
Batu Pahat, 86400, Johor, Malaysia

Abstract—**Diabetes is one of the most chronic diseases, with an increasing number of sufferers yearly. It can lead to several serious complications, including diabetic peripheral neuropathy (DPN). DPN must be recognized early to receive appropriate treatment and prevent disease exacerbation. However, due to the rapid development of machine learning classification, like in the health science sector, it is very easy to identify DPN in the early stages. Therefore, the aim of this study is to develop a new method for detecting neuropathy based on the myoelectric signal among diabetes patients at a low cost with utilizing one of the machine learning techniques, the artificial neural network (ANN). To that aim, muscle sensor V3 is used to record the activity of the anterior tibialis muscle. Then, the representative time domain features which is mean absolute value (MAV), root mean square (RMS), variance (VAR), and standard deviation (SD) used to evaluate fatigue. During neural network training, a different number of hidden neurons were used, and it was found that using seven hidden neurons showed a high accuracy of 98.6%. Thus, this work indicates the potential of a low-cost system for classifying healthy and diabetic individuals using an ANN algorithm.**

Keywords—Electromyography; diabetic neuropathy; classification; machine learning; artificial neural networks

I. INTRODUCTION

According to the statistics, diabetes is one of the most common chronic diseases, with an increase in patients yearly [1]. It can lead to many serious complications, including diabetic peripheral neuropathy (DPN), a neuromuscular disease affecting up to half of all people with diabetes [2]. DPN can damage the nerves and blood vessels in the feet and legs, resulting in plantar foot ulcers, motor unit loss, and muscle volume reduction [3]. Furthermore, the peripheral nervous system, which transports currents to specific areas of the body to control muscle activity, will be disrupted by this disease. Early identification of DPN is important for people with diabetes to maintain a high quality of life by adopting a healthy lifestyle, such as eating healthy food, being active every day, and stopping smoking [4].

Electromyography (EMG) has been designed as the standard gold examination to diagnose DPN and determine nerve distribution and severity [5]. Some disadvantages of EMG include the complexity of data processing and the expense of the equipment [6]. Not many hospitals routinely do EMG analysis because commercially available equipment costs range from € 15.000 to € 20.000 [7].

Today, inexpensive and simple electrical sEMG (such as muscle sensor V3) can capture biological signals, resulting in affordable EMG equipment. As an example, Toro et al.[8], study to design of a low-cost sEMG system that allows for assessing when fatigue appears in a muscle. The author used low-cost sEMG sensors, an Arduino board, and a PC. The study proves the feasibility of a low-cost system to detect muscle fatigue reliably.

Meanwhile, the EMG signals generated by a healthy person differ from those generated by an individual with damaged muscle fibre or nerve groups. However, it is nearly impossible to discern the signals with the naked eye because they appear almost identical [9]. Classification techniques are one of the suitable alternative solutions for this problem in analysis and decision-making [10]. Automated EMG evaluation using artificial neural networks (ANNs) has become increasingly common in recent years due to the ability of these networks to analyze large amounts of multidimensional data and the strong capabilities of these networks for field recognition and classification [11].

An artificial neural network is a supervised classification technique replicating biological neurons' connection structure [12]. It comprises many interconnected neurons working together to solve problems [13].

In this study, the ANN algorithm used develops a new method for detecting neuropathy among diabetes patients at a low cost. The performance of this system is evaluated by taking actual signals from non-diabetic subjects (healthy) and diabetics. However, due to time constraints in finding suitable subjects, this study uses an increasingly popular way to overcome the issues of data availability is to use synthetic data [14]–[16]. The synthetic data was implemented by adding random noise to the original signals [17].

II. RELATED WORK

Researchers have compared the performance of the k-nearest neighbors algorithm (KNN) and ANN to classify and diagnose diseases [18]. Compared to KNN, ANN produces variable outcomes, but ANN has a higher accuracy of 80.86% compared to KNN accuracy is 77.24%.

In [19], an EMG signal-based authentication algorithm is proposed to compensate for personal certification methods' weaknesses. The pattern identification rates for the SVM and KNN algorithms were 90%. However, the results for personal authentication were just 64%. However, individual

authentication using ANN demonstrated a relatively high accuracy of 81.6%.

In [20] study compares a Support Vector Machine (SVM) to a hybrid of SVM and the ANN system as the best binary classification system for determining who has diabetes. The accuracy determined using an SVM was 77%, and when applying the ANN system, the accuracy was 87%. The results of this study show that this model, which is a mix of the SVM and the ANN, is more accurate than the SVM model.

Berina et al. [21] use machine learning techniques in the classification of diabetes and cardiovascular diseases (CVD) using Artificial Neural Networks (ANNs) and Bayesian Networks (BNs). ANN shows a greater chance of obtaining more precise results regarding the classification of CVD and diabetes. The author [22] employs artificial neural networks to determine a person's likelihood of having diabetes. The accuracy of predicting whether a person has diabetes was 87.3% after training the ANN model, which resulted in an average error function of the neural network of 0.01 and this value.

Previous studies have shown that ANNs perform well compared to other neural networks. In this study, ANN is trained with Scaled Conjugate Gradient (SCG) to perform better with a high accuracy percentage. One advantage of SCG is that it requires less time because it avoids computationally expensive linear searches [23].

III. METHODOLOGY

A. Subject

Twenty volunteer subjects were enrolled: ten healthy control (non-diabetes) subjects aged (mean ± SD) 61.9 ± 6.5 years and ten diabetes subjects aged (mean ± SD) 64.1 ± 8.3 years with duration of diabetes (mean ± SD) 17.1 ± 12.1 years. The data of subjects are shown in Table I.

The inclusion criteria were non-diabetics (healthy), and diabetic subjects were male and female gender and aged 18 to 65. In this study, the exclusion criteria for subjects with a bad general health state were; stroke, Parkinson's disease, legs with (ulcer, gout, and disability), peripheral nervous system history, and severe muscular atrophy in lower limbs because that could interfere with electromyographic recording activity [24], [25].

TABLE I. DEMOGRAPHIC DATA FOR THE SUBJECTS

	Healthy	Diabetics
Male/Female	5/5	5/5
Age (Years)	61.9 ± 6.5	64.1 ± 8.3
BMI (kg/m^2)	26.2	29.9
Duration of Diabetes (Years)	-	17.1 ± 12.1
HbA1c (%)	-	9.7

B. Low Cost Hardware Implementation

- Muscle Sensor V3

Fig. 1 shows the Muscle Sensor V3 that is used to record electrical potentials produced by contracting muscles. This sensor process the raw signal when records because its sequence is constructed of an instrumentation amplifier, rectifier, analog filter circuits, and an end amplifier circuit [26]. After recording the signal, the sensor amplifies and processes the complex electrical activity of a muscle before converting it to a simple analog signal that is easily read by a microcontroller Arduino) to convert it into a digital signal [27].

Fig. 1. Muscle Sensor v3.

- Electrode Pads

Gel electrodes are used with muscle sensor v3 to capture data. The electrode is replaced when to record for another subjects, to get the higher quality EMG signal, maintain cleanliness, and prevent the spreading of infectious diseases. This electrode has three colors, red, green and yellow as in Fig. 2, —the red electrode is (ground), the yellow electrode (reference) and the green color electrode is (the EMG signal electrode) [28].

Fig. 2. Ag/AgCl Disposable Electrode.

- Arduino Uno

Fig. 3 shows Arduino UNO with ATmega328P-based microcontroller board. It contains 14 input/output digital pins (of which six can be used as PWM outputs). Arduino UNO is used to convert analog to digital output due to the signal produced by muscle sensor v3. Muscle sensor v3 output was interfaced with an Arduino UNO microcontroller to record the signal on a laptop operating in battery mode for signal processing and classification [29]. Data will save and upload to Matlab workspace for extraction and classification [30].

Fig. 3.   Arduino Uno.

C. Electrode Location

The placement of muscle EMG electrodes on muscles is based on the SENIAM (Surface Electromyography for the Non-Invasive Assessment of Muscles) group guideline [32], [33]. The anterior tibialis muscle was selected to record the data because it is the most medial muscle in the lower leg and the strongest dorsiflexor movement of the leg. As in Fig. 4, the electrode is located on the anterior tibialis muscle at 1/3 of the distance between the end of the fibula and the end of the medial malleolus. Yellow electrode placed on an inactive body section, such as the bony portion, in this study is attached to the knee. Before placing electrodes, the skin's surface must be cleaned to lower its resistance.
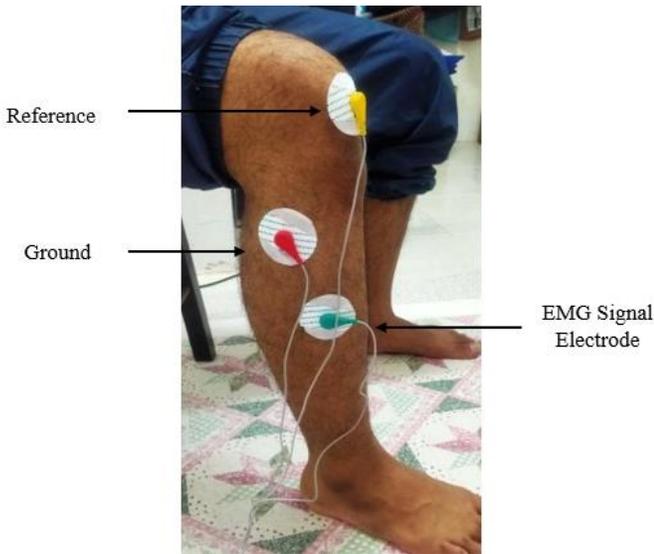
Fig. 4.   Electrode Location on Tibialis Anterior Muscle.

D. Experiment Procedure

The subject is sitting on the chair and at rest for a period of five minutes after electrode placements. After five minutes, the subject is asked to slowly lift the toes and the forefoot towards the shin, digging the heel into the floor (dorsiflexion). The toes will be raised as high as they can for one minute (hold) and then slowly lower to the floor, as in Fig. 5. This method makes TA muscle contraction while minimizing the subject's movement without requiring the subject to walk.
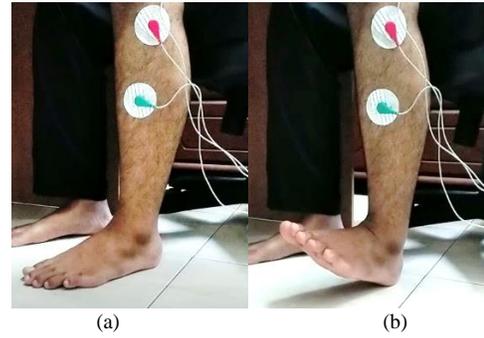
Fig. 5.   Normal Foot Position on the Floor, (b) Foot Position During Dorsiflexion.

E. Data Collection

EMG signals were obtained from 20 subjects of different gender and ages. A total of 20 recordings samples were taken from every 20 subjects. Then, the EMG data load to Matlab workspace for synthetic data process before feature extraction. The synthetic data is repeated four times for each original sample. As a result, the original 20 samples produce 80 synthetic data and make the total for the classification procedure 100 samples (20 original and 80 synthetics).

F. Feature Extraction

The most common features for the time domain are Root Mean Square (RMS), Mean Absolute Value (MAV), Standard Deviation (STD), and Variance (VAR), which were used for this study [31]–[33]. This makes 100 samples of data divided into four characteristics for classification input (4x100).

The equations are used to determine the value of the features:

Mean Absolute Value (MAV) [34]: The mean absolute value and the moving average rectified value (ARV) are probably similar. Expression is given below.

$$MAV\ (\bar{x}) = \frac{1}{N} \sum_{n=1}^{N} x_n \qquad (1)$$

Root Mean Square (RMS) [34]: This is a measurement of the square root of the input signal, expressed as:

$$RMS = \sqrt{\frac{1}{N} \sum_{n=1}^{N} x_n^2} \qquad (2)$$

Variance (VAR) [34]: If the variable signal is squared and then divided by the mean value, then the variance is identified, which is stated as:

$$VAR = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \bar{x})^2 \qquad (3)$$

Standard Deviation (SD) [34]: Using the following expression, the threshold level of muscular contraction activity is determined:

$$SD = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} (x_n - \bar{x})^2} \qquad (4)$$

N represents the data length for each feature, while x_n represents the EMG signal in a segment or the value for each feature.

G. Classification

The classification used Matlab (nprtool) to solve a pattern-recognition classification problem. In this tool, an input layer, hidden layers, and an output layer make up a feedforward neural network. Scaled backpropagation of the conjugate gradient (trainscg) used to train the network. This study used default settings and parameters bt Matlab because possible to set the parameters and settings manually except for the number of hidden neurons.

A feedforward neural network load with the sample data (4x100) as inputs and binary numbers 10 for healthy and 01 for people with diabetes as outputs (2x100). At the same time, different hidden layers and training functions were tested as affected factors were tested with (1-10) hidden neurons.

ANN requires three data sets for the inputs and outputs of the system: training, validation, and testing. It required 30% (15 % for validation and 15% for testing), leaving 70 % for training. 100 samples were divided into 70 samples for training, 15 for testing, and 15 for validations.

IV. RESULT AND DISCUSSION

This section discusses the evaluation of the results and performance of the proposed model. An ANN was created to distinguish between diabetics and healthy subjects using Matlab. After training, the network gives a confusion matrix to indicate the correctness network. The confusion matrix is constructed to compare classification during the training, testing, and validation processes.

The confusion matrix provides details about the distribution for each output class, and then it uses color coding: green for accurate classification and red for inaccurate classification. Four confusion matrices are shown, one for each stage, namely the training, validation, testing processes, and overall classification, as shown in Fig. 6.

Fig. 6. Result from the Confusion Matrix.

The performance of ANN is tested with the different number of hidden neurons (1-10) used during training. Matlab nnstart toolbox was utilized during training and development of the neural network. Based on the results, using seven hidden neurons shows the highest overall accuracy archive with 98.6% as shown in Table II.

TABLE II. CLASSIFICATION PERFORMANCE OF DIABETICS AND HEALTHY SUBJECTS

No Hidden Layer	Training (%)	Testing (%)	Validation (%)	Overall (%)
1	92.9	95.2	76.2	90.7
2	92.9	90.5	100	93.6
3	85.7	85.7	90.5	86.4
4	94.9	85.7	81.0	91.4
5	96.9	100	95.2	97.1
6	94.9	81.0	85.7	91.4
7	99.0	95.2	100	98.6
8	94.9	90.5	90.5	93.6
9	93.9	100	100	95.7
10	91.8	85.7	81.0	89.3

The result showed that the number of hidden neurons was significant in characterizing the network's performance, which should not be too large or too small [35]. If there are more hidden units than needed, the network may not work well because it will take longer to process everything, affecting backpropagation in the long run. If less than the required number of hidden units is made, some information may lose, and the prediction may not be right [36].

The obtained network will be less complex since the number of neurons in hidden layers is fixed correctly without any least significant neurons. It will be more accurate because backpropagation is used to adjust the weight after each change is made to the network [37]. With the help of the backpropagation algorithm, the network with the best training gives the best results. Besides, the selection of scaled conjugate gradient backpropagation technique also helps the best result. In [38] study results prove that scaled conjugate gradient backpropagation is the most effective choice than other backpropagation algorithms that give high accuracy.

Table III shows performance comparison between this study and other works. This comparison is based on the study that does the classification signal from the self-study data collection on diabetic or neuropathy patients.

TABLE III. A COMPARISON BETWEEN THIS STUDY AND PREVIOUS

Reference	Classifier Method	Accuracy
Ahmed et al [9]	SVM ANN	70% 85%
Yousfat et al. [39]	-	93%
Bhusari et al. [32]	KNN	70%
This Study	ANN	98.6%

V. Conclusion and Recommendation

This paper shows the performance of low-cost EMG acquisition boards using artificial neural networks (ANN) in proposing an approach for detecting diabetes neuropathy. Based on the results, the highest accuracy is 98.6% when trained with the SCG algorithm. According to the recorded EMG data, the ANN is an effective way to classify these two classes of subjects. This system can assist in the correct initial detection of diabetic neuropathy so that appropriate treatment will be given.

For future work, with more time and funds, more samples can be measured for model validation. With a larger sample size, a more accurate system is likely to be studied, thus allowing higher accuracy in classifying neuropathic pain to be carried out. High accuracy results also indicate opportunities to develop a system not only to know neuropathy pain but also to know the severity of neuropathy faced by diabetic patients.

Acknowledgment

References

[1] W. Slasko, F. Spolaor, Z. Sawacha, A. Guiotto, G. Guarnieri, and A. Avogaro, "How do different activities impact on the detection of muscle's alterations in diabetes subjects with and without neuropathy?," Gait Posture, vol. 73, pp. 418–419, 2019, doi: 10.1016/j.gaitpost.2019.07.157.

[2] A. G. Kristensen et al., "Detection of early motor involvement in diabetic polyneuropathy using a novel MUNE method – MScanFit MUNE," Clin. Neurophysiol., vol. 130, no. 10, pp. 1981–1987, 2019, doi: 10.1016/j.clinph.2019.08.003.

[3] F. Haque, M. N. I. Reaz, M. E. H. Chowdhury, F. H. Hashim, N. Arsad, and S. H. M. Ali, "Diabetic Sensorimotor Polyneuropathy Severity Calssification USing Adaptive Neuro Fuzzy Inference System," IEEE Spons. 2nd Int. Conf. Innov. Inf. Embed. Commun. Syst., 2020.

[4] Y. Shibata et al., "Validity and reliability of a point-of-care nerve conduction device in diabetes patients," J. Diabetes Investig., vol. 10, no. 5, pp. 1291–1298, 2019, doi: 10.1111/jdi.13007.

[5] Y. Zhang et al., "Relationship between hyponatremia and peripheral neuropathy in patients with diabetes," J. Diabetes Res., vol. 2021, 2021, doi: 10.1155/2021/9012887.

[6] M. Reinvee, P. Vaas, J. Ereline, and M. Pääsuke, "Applicability of Affordable sEMG in Ergonomics Practice," Procedia Manuf., vol. 3, no. Ahfe, pp. 4260–4265, 2015, doi: 10.1016/j.promfg.2015.07.412.

[7] P. Tecchio, A. Monte, and P. Zamparo, "Low-cost electromyography : validity against a commercial system depends on exercise type and intensity," vol. 31, no. 2, 2021, doi: 10.4081/ejtm.2021.9735.

[8] S. F. Del Toro, S. Santos-Cuadros, E. Olmeda, C. Álvarez-Caldas, V. Díaz, and J. L. San Román, "Is the Use of a Low-Cost sEMG Sensor Valid to Measure Muscle Fatigue?," Sensors (Basel)., vol. 19, no. 14, pp. 1–19, 2019, doi: 10.3390/s19143204.

[9] T. Ahmed and M. K. Islam, "EMG Signal Classification for Detecting Neuromuscular Disorders," J. Phys. Conf. Ser., vol. 1921, no. 1, 2021, doi: 10.1088/1742-6596/1921/1/012043.

[10] K. Lakhwani, S. Bhargava, K. K. Hiran, M. M. Bundele, and D. Somwanshi, "Prediction of the Onset of Diabetes Using Artificial Neural Network and Pima Indians Diabetes Dataset," 2020 5th IEEE Int. Conf. Recent Adv. Innov. Eng. ICRAIE 2020 - Proceeding, vol. 2020, 2020, doi: 10.1109/ICRAIE51050.2020.9358308.

[11] R. Akhundov, D. J. Saxby, S. Edwards, S. Snodgrass, P. Clausen, and L. E. Diamond, "Development of a deep neural network for automated electromyographic pattern classification," pp. 1–5, 2019, doi: 10.1242/jeb.198101.

[12] D. A. Otchere, T. Omar, A. Ganat, R. Gholami, and S. Ridha, "Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models," J. Pet. Sci. Eng., p. 108182, 2020, doi: 10.1016/j.petrol.2020.108182.

[13] M. A. Rosales, M. G. B. Palconit, A. A. Bandala, R. R. P. Vicerra, E. P. Dadios, and H. Calinao, "Prediction of total body water using scaled conjugate gradient artificial neural network," IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON, vol. 2020-Novem, pp. 218–223, 2020, doi: 10.1109/TENCON50793.2020.9293804.

[14] R. J. Chen, M. Y. Lu, T. Y. Chen, D. F. K. Williamson, and F. Mahmood, "Synthetic data in machine learning for medicine and healthcare," Nat. Biomed. Eng., vol. 5, no. 6, pp. 493–497, 2021, doi: 10.1038/s41551-021-00751-8.

[15] M. Alzantot, S. Chakraborty, and M. Srivastava, "SenseGen: A deep learning architecture for synthetic sensor data generation," 2017 IEEE Int. Conf. Pervasive Comput. Commun. Work. PerCom Work. 2017, pp. 188–193, 2017, doi: 10.1109/PERCOMW.2017.7917555.

[16] F. K. Dankar and M. Ibrahim, "Fake it till you make it: Guidelines for effective synthetic data generation," Appl. Sci., vol. 11, no. 5, pp. 1–18, 2021, doi: 10.3390/app11052158.

[17] R. Gil-Pita, P. Jarabo-Amores, M. Rosa-Zurera, and F. López-Ferreras, "Improving neural classifiers for ATR using a kernel method for generating synthetic training sets," Neural Networks Signal Process. - Proc. IEEE Work., vol. 2002-Janua, no. c, pp. 425–434, 2002, doi: 10.1109/NNSP.2002.1030054.

[18] I. S. Jasim, A. D. Duru, K. Shaker, B. M. Abed, H. M. Saleh, and I. Technology, "Diseases," pp. 11–14, 2017.

[19] S. Shin, J. Jung, and Y. T. Kim, "A study of an EMG-based authentication algorithm using an artificial neural network," Proc. IEEE Sensors, vol. 2017-Decem, pp. 1–3, 2017, doi: 10.1109/ICSENS.2017.8234158.

[20] S. Aliwadi, V. Shandila, T. Gahlawat, P. Kalra, and D. Mehrotra, "Diagnosis of diabetic nature of a person using SVM and ANN approach," 2017 6th Int. Conf. Reliab. Infocom Technol. Optim. Trends Futur. Dir. ICRITO 2017, vol. 2018-Janua, pp. 338–342, 2018, doi: 10.1109/ICRITO.2017.8342448.

[21] B. Alić, L. Gurbeta, and A. Badnjević, "Machine learning techniques for classification of diabetes and cardiovascular diseases," 2017 6th Mediterr. Conf. Embed. Comput. MECO 2017 - Incl. ECYPS 2017, Proc., no. June, pp. 17–20, 2017, doi: 10.1109/MECO.2017.7977152.

[22] S. E. Nesreen and S. S. Abu-Naser, "Diabetes prediction using artificial neural network," Deep Learn. Tech. Biomed. Heal. Informatics, vol. 121, pp. 327–339, 2020, doi: 10.1016/B978-0-12-819061-6.00014-8.

[23] P. K. Upadhyay, A. Pandita, and N. Joshi, "Scaled Conjugate Gradient Backpropagation based SLA Violation Prediction in Cloud Computing," Proc. 2019 Int. Conf. Comput. Intell. Knowl. Econ. ICCIKE 2019, pp. 203–208, 2019, doi: 10.1109/ICCIKE47802.2019.9004240.

[24] L. Coppeta et al., "Neuromuscular Functional Assessment in Low Back Pain by Surface Electromyography (SEMG)," Open Public Health J., vol. 12, no. 1, pp. 61–67, 2019, doi: 10.2174/1874944501912010061.

[25] A. Scarton et al., "Comparison of lower limb muscle strength between diabetic neuropathic and healthy subjects using OpenSim," Gait Posture, vol. 58, no. November, pp. 194–200, 2017, doi: 10.1016/j.gaitpost.2017.07.117.

[26] A. Gunadhi, F. Z. Mustofa, L. Agustine, H. Pranjoto, R. Sitepu, and Yuliati, "Microcontroller-based assistive device for training weak biceps brachii muscle," IOP Conf. Ser. Mater. Sci. Eng., vol. 673, no. 1, 2019, doi: 10.1088/1757-899X/673/1/012070.

[27] F. Ali, H. Jamaluddin, N. Abas, M. Fahmi, A. Z. Shukor, and M. A. Norizan, "Jurnal Teknologi DESIGN AND CONSTRUCTION OF RH 2000 CYBERNATICS," vol. 20, pp. 121–125, 2015.

[28] K. Samarawickrama, S. Ranasinghe, Y. Wickramasinghe, W. Mallehevidana, V. Marasinghe, and K. Wijesinghe, "Surface EMG Signal Acquisition Analysis and Classification for the Operation of a Prosthetic Limb," Int. J. Biosci. Biochem. Bioinforma., vol. 8, no. 1, pp. 32–41, 2018, doi: 10.17706/ijbbb.2018.8.1.32-41.

[29] B. Champaty, P. Dubey, S. Sahoo, S. S. Ray, and K. Pal, "Rehabilitation Devices," pp. 3–6, 2014.

[30] D. M. Dao, P. D. Phuoc, T. X. Tuy, and T. T. Le, "Research on reading muscle signals from the EMG sensor during knee flexion-Extension using the Arduino Uno controller," Int. Conf. Adv. Technol. Commun., vol. 2017-Octob, no. October 2017, pp. 270–273, 2017, doi: 10.1109/ATC.2017.8167632.

[31] M. A. Oskoei and H. Hu, "GA-based feature subset selection for myoelectric classification," 2006 IEEE Int. Conf. Robot. Biomimetics, ROBIO 2006, pp. 1465–1470, 2006, doi: 10.1109/ROBIO.2006.340145.

[32] A. Bhusari, N. Gupta, T. Kambli, and S. Kulkarni, "Comparison of SVM and kNN classifiers for palm movements using sEMG signals with different features," Proc. 3rd Int. Conf. Comput. Methodol. Commun. ICCMC 2019, no. Iccmc, pp. 881–885, 2019, doi: 10.1109/ICCMC.2019.8819727.

[33] N. S. A. Sharawardi, Y. H. Choo, S. H. Chong, A. K. Muda, and O. S. Goh, "Single channel sEMG muscle fatigue prediction: An implementation using least square support vector machine," 2014 4th World Congr. Inf. Commun. Technol. WICT 2014, pp. 320–325, 2014, doi: 10.1109/WICT.2014.7077287.

[34] A. F. Hasan, H. U. Masud, T. Anzar, and P. Hasan, "Hand Gesture Recognition Based on Surface Electromyogram Signal (sEMG) with Muscular Contraction Level and Real Time Implementation on An Artificial Prosthetic Wrist Using Artificial Neural Network (ANN),"

[35] 2020 IEEE Reg. 10 Symp. TENSYMP 2020, no. June, pp. 1648–1651, 2020, doi: 10.1109/TENSYMP50017.2020.9231038.

[35] J. Li, Y. Wu, J. Zhang, and G. Zhao, "A Novel Method to Fix Numbers of Hidden Neurons in Deep Neural Networks," pp. 8–11, 2015, doi: 10.1109/ISCID.2015.41.

[36] J. Yin, Z. Zhao, C. Lei, and S. X. Yang, "Improved optical-type measurement method of grain flow using array near-infrared photoelectric sensors," Comput. Electron. Agric., vol. 183, no. April 2020, 2021, doi: 10.1016/j.compag.2021.106075.

[37] R. Kathiroli, R. Rajakumari, and P. Gokulprasanth, "Diagnosis of Diabetes Using Cascade Correlation and Artificial Neural Network," 2018 10th Int. Conf. Adv. Comput. ICoAC 2018, pp. 299–306, 2018, doi: 10.1109/ICoAC44903.2018.8939103.

[38] C. B. Khadse, M. A. Chaudhari, and V. B. Borghate, "Comparison of seven backpropagation algorithms for three phase power quality assessment," IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON, vol. 2017-Decem, pp. 2548–2553, 2017, doi: 10.1109/TENCON.2017.8228291.

[39] K. Yousaf, H. Ather, and W. Saadeh, "Wearable Peripheral Neuropathy Detection System based on Surface Electromyography," 2019 UK/China Emerg. Technol. UCET 2019, pp. 2019–2021, 2019, doi: 10.1109/UCET.2019.8881853.

Constraints on Hyper-parameters in Deep Learning Convolutional Neural Networks

Ubaid M. Al-Saggaf[1]*, Abdelaziz Botalb[2], Muhammad Faisal[3]
Muhammad Moinuddin[4], Abdulrahman U. Alsaggaf[5], Sulhi Ali Alfakeh[6]

Electrical and Computer Engineering Department, King Abdulaziz University, Jeddah 21589, Saudi Arabia[1, 2, 4, 5]
Center of Excellence in Intelligent Engineering Systems (CEIES), King Abdulaziz University, Jeddah 21589, Saudi Arabia[1, 2, 4, 5]
Computer & Information Technology Dept., Dammam Community College
King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia[3]
Department of Internal Medicine, Child and Adolescent Psychiatrist
Faculty of Medicine, King Abdulaziz University, Jeddah 21589, Saudi Arabia[6]

Abstract—**Convolutional Neural Network (CNN), a type of Deep Learning, has a very large number of hyper-meters in contrast to the Artificial Neural Network (ANN) which makes the task of CNN training more demanding. The reason why the task of tuning parameters optimization is difficult in the CNN is the existence of a huge optimization space comprising a large number of hyper-parameters such as the number of layers, number of neurons, number of kernels, stride, padding, rows or columns truncation, parameters of the backpropagation algorithm, etc. Moreover, most of the existing techniques in the literature for the selection of these parameters are based on random practice which is developed for some specific datasets. In this work, we empirically investigated and proved that CNN performance is linked not only to choosing the right hyper-parameters but also to its implementation. More specifically, it is found that the performance is also depending on how it deals when the CNN operations require setting of hyper-parameters that do not symmetrically fit the input volume. We demonstrated two different implementations, crop or pad the input volume to make it fit. Our analysis shows that padding performs better than cropping in terms of prediction accuracy (85.58% in contrast to 82.62%) while takes lesser training time (8 minutes lesser).**

Keywords—Neural networks; convolution; pooling; hyper-parameters; CNN; deep learning; zero-padding; stride; back-propagation

I. INTRODUCTION

Convolutional Neural Networks (CNNs) have proved to be the perfect machine learning choice for a wide range of application fields, such as pattern classification and analysis of video, image, speech, and text (natural language processing). However, no doubt using CNNs requires much work compared to other machine learning solutions such as random forest, Support Vector Machines, etc. This added work is primarily due to the vast optimization space of parameters and hyper-parameters, which interact with each other in a very complex way. Furthermore, making this problem even more complex is that there is still no universal, robust theory that supports hyper-parameters optimization. That would enable us to choose the right hyper-parameters for the right problem at hand and give the best performance with less effort and time. Setting hyper-parameters without robust theory is like

working blindly, as quoted by the German philosopher Immanuel Kant:

"Experience without theory is blind, but theory without experience is mere intellectual play".

Deep convolutional neural networks often have numerous layers piled on each other and are taught to do a specific task. At the end of each layer, the network learns a variety of low, medium, and high-level features. There are several papers in the literature with different approaches to setting CNN hyper-parameters, but none of them has presented a generalized and robust systematic approach to the problem. Thus, hyper-parameters optimization is not a problem that is ever entirely solved. This work aims to investigate empirically the impact of two different strategies to deal with input volume in CNN, that is, cropping and padding.

II. THE CHALLENGE OF HYPER-PARAMETERS TUNING IN A CNN

Setting the best CNN hyper-parameters for a particular classification problem could be challenging. If the results are not making any sense or bad accuracy was achieved after the first trial, then there is no prior knowledge about what went wrong; anything of the following could be a reason:

- Activation function for every layer or set of layers.
- Learning rate.
- Momentum.
- Regularization type and factor value.
- Size of filters in a specific layer(s).
- The stride of convolution or pooling in a particular layer(s).
- Padding type (e.g., zeros, first and last values, values repeated cyclically).
- The number of convolution, pooling, and activation layers in the network.
- Order of layers e.g., conv-conv-pool; or conv-activation-pool-conv.

---

*Corresponding Author.

- Type of cost function.

- The approach to encoding the output is not appropriate.

- Backpropagation gradient implementation is incorrect (sanity check was not done or was not efficient).

- Dataset was not split correctly into proper ratios.

- Not enough datasets for training.

- Imbalanced dataset.

- Not pre-processed the dataset, e.g., not normalized at all or normalized incorrectly.

- Wrong label assignment during training could happen when a new split ratio is needed.

- The performance measure metric is not appropriate for the given problem.

- The mini-batch size is not appropriate for Stochastic Gradient Descent.

- Different approaches to weight initialization.

- Maybe CNN is not the right solution to the problem at hand.

Moreover, often the wrong choice of one hyper-parameter (e.g., sigmoid activation function) will never let the network converge no matter how all other hyper-parameters were chosen. So, it is straightforward to get frustrated and lose in the hyper-parameter space and this makes CNN debugging a challenging task. Some heuristics could be used to set the right hyper-parameters by developing a workflow that enables quick debugging. These heuristics are just rules of thumb, and they are not guaranteed to give the best possible results because the behaviour of the CNN entirely depends on the specific dataset of a particular problem. Also, this is what makes it hard to establish one universal solid theory of defining the appropriate hyper-parameters for any given problem.

A. Related Work

In contrast to the manual search for finding an optimum set of hyper-parameter values, automated search approaches were adopted by many research works, and here we highlight a small portion of them:

In [1], a grid search method and a manual search technique were investigated for neural networks and particularly for deep belief neural networks. According to their findings, the grid search method is not an optimal choice for setting the hyper-parameters of the neural networks. On the other hand, it was found that random search can be used to obtain a baseline result in order to assess the performance of the other hyper-parameter optimization methods. In [2], the issue of hyper-parameters tuning is addressed by formulating the problem as a constrained optimization task. Then, a derivative-free optimization technique is opted for tuning the parameters. As a result, a more accurate and automated technique is developed for tuning of the hyper-parameters. Moreover, it is found that this technique is not consistent to achieve the global optimum. In [3], the Taguchi method was utilized to obtain

optimal hyper-parameters which can provide faster training and enhanced classification accuracy. In [4], the authors utilized a Bayesian learning technique by employing Gaussian process-based sampling to develop a learning model for designing the tuning parameters of the algorithm. It was reported that the proposed technique for evaluating the optimum values of hyper-parameters can provide significantly faster learning compared to the learning performance of the baseline methods. In [5], a probabilistic model is developed that can deal with early learning termination in the case a bad performance is detected.

Deep convolutional neural networks often have numerous layers piled on each other and are taught to do the specific task. At the end of each layer, the network learns a variety of low, medium, and high-level features.

Different variants of deep neural networks, including CNN, have been employed on various signal processing and machine learning tasks. However, their claimed performance is specific to a certain problem and dataset. In [6], an evolutionary algorithm was developed for designing of optimal hyper-parameter for different CNNs and their performance was investigated on the MNIST dataset. In [7], the hyper-parameters for the deep neural network were optimized using Nelder-Mead and coordinate search methods. In [8], a bandit-based strategy was proposed to solve the task of hyper-parameter optimization. Datasets CIFAR-10, MNIST, and Street View House Numbers were investigated and it is found that the proposed method is faster than the existing Bayesian optimization algorithms. Another work in [9] developed a faster algorithm for the tuning of hyperparameters by employing the strategy of Boolean functions analysis. Its performance was tested using deep neural networks on the CIFAR-10 dataset, and they proved that the proposed technique has much better performance than the existing baseline methods. In [10], a non-parametric regression model-based adaptive technique was designed to calculate the optimum value of the learning hyper-parameters in a short time. The proposed idea was applied on CNN and it was shown via various experiments that the proposed algorithm performs faster than the existing state-of-the-art techniques. In [11], Bayesian optimization-based generative model was developed to analyse the validation error w.r.t. the size of the training data. Again deep neural networks were utilized to test the performance of this algorithm and it was shown that the proposed method finds better hyper-parameters in lesser time. In [12], a novel technique for optimizing the hyper-parameters was developed which was based on the combination of the input features. The NMA was used with CNN for the CIFAR-10 dataset for testing the performance of the proposed model and it was shown to achieve better classification accuracy. In [13], the Covariance Matrix Adaptation Evolution Strategy was used to develop a novel optimization technique for designing the tuning parameters for CNN and it is found to be very effective.

There are various techniques in the literature focused on optimizing the tuning parameters of the CNN. For example, a Grid Search method was proposed in [14] to optimize the CNN parameters. A Random Search based optimization of hyper-parameters of the CNN was proposed in [15]. There

exist optimization techniques that employ Bayesian optimization strategy for designing CNN parameters such as those proposed in [16], [17], [18]. A Differential Evolution based optimization algorithm was proposed in [19]. A similar task was obtained using the Harmonic Search method in [20]. Reinforcement learning-based techniques were developed in [21] and [22]. Recently, in [23], the Micro canonical Optimization algorithm was employed for the automatic selection of the hyper-parameters in CNN.

In summary, all these existing works on the selection of hyper-parameters in CNN are concerned with the development of any automated algorithm that can provide optimum or near optimum values of these hyper-parameters. The primary relevant outcome of these related works is that they always link the accuracy and performance of the neural network to the hyper-parameters optimization of the micro and macro-structural levels of the model. However, we claim that in addition to the structural levels of the CNN model another factor affects the accuracy and performance of the model. The factor is the way of implementation of certain inter-layer operations in CNN. Our preliminary results on a single dataset with lesser details were published in [24]. In this work, we provide a more detailed analysis on various datasets in this context.

B. Main Contributions

The main contributions of our work can be summarized as follows:

1) In this work, we provide a framework to empirically investigate the effect of the way of implementation of certain inter-layer operations in CNN.

2) More specifically, we compare two different inter-layer operations namely cropping and padding the input volume to make it fit for the next layer operation.

3) Our investigation is based on analyzing the CNN performance in terms of classification accuracy, processing time, and generalization by implementing two models: one using crop and the second employing padding of the input volume on Digits, MNIST, Merch, Flowers, and CIFAR-10 datasets.

4) Our work provides a foundation for future investigation of the effects of other inter-layer operations in a CNN.

C. Problem Definition

A common practice in the research of deep learning is to use built-in CNN libraries, such as Tensorflow, Caffe, and Keras. However, users do not have control over the low-level implementation of different blocks of CNN which are implemented differently in different libraries. For example, in order to deal with input volumes when the stride hyper-parameter value results in a non-integer pre-calculated output, there are two standard approaches: first is cropping the volume and second is padding the volume. To understand these approaches, consider an example where an input volume of $90 \times 70 \times 100$ which corresponds to 70 channels with mini-batch size of 100. The pooling layer with a window size of 4×4 and a stride of 4 is used with no padding. The pre-

calculated dimension for the output of both convolution and pooling is given as follows (see [25] for the formula used):

$$Feature\ Map\ Size = \frac{I-F+2P}{S} \quad \frac{90-4+(2\times0)}{4} + 1 = 22.5 \qquad (1)$$

where:

I: Row or column size of the Input volume

F: Row or column size of the filter

P: Zero-padding

S: Stride

It can be observed in the calculation provided in Eq. (1) that both the axes of the feature map gave the same measure (22.5). Otherwise, Eq. (1) can be used to calculate the dimension size separately for each case. It is to be noted that the calculation of Eq. (1) results in a float number (22.5) for this example. Thus, the pooling size used was not appropriate to fit the input volume. This problem can come across in any layer type and any layer number. Now, it is important to know how this issue is handled in different libraries. There are three possible solutions to this issue; the first is to go back to choose a different value of the hyper-parameter and check again its feasibility; the second is to crop the input volume in order to fit the volume for the next layer; and the third is to pad (usually zero-padding is employed) the volume to fit the volume. The question is: Are these approaches going to perform the same? Hence, there are Constraints on choice of hype-parameters in the design of CNNs.

D. Objectives

The issue to investigate is how both approaches will affect the network's performance. This will tell us whether non-optimal performance is purely due to the wrong choice of hyper-parameters or is it also partially due to how the library is handling that type of hyper-parameters (that gives a non-integer value for the pre-calculated output). The outcome of this work will be a reference for both CNN library users and those designing CNNs and other deep learning paradigms from scratch.

Two models are going to be implemented from scratch, the first model implementation will involve the first approach that crops the input volume, and the second model will present the implementation of the second approach that involves zero-padding the input.

III. DATASETS

The used datasets in this work are Digits, MNIST, Merch, Flowers, and CIFAR-10. The MNIST dataset contains 70,000 images of handwritten digits having a size of 28×28. For machine learning applications, this dataset is split into 60%, 25%, and 15% for training, validation, and testing, respectively. The dataset is almost balanced among different classes, so the performance measure will not be misleading, at least from this side. Also, different image size datasets were chosen with a small to a reasonably large number of images. The number of classes ranges from 5 to 10. Table I. shows some of the characteristics of the datasets used. The experiments for the rest of the datasets were performed using

the original images and no transformation such as conversion to grayscale, rotation, etc. has been performed.

TABLE I. CHARACTERISTICS OF DATASETS USED

Dataset	Image dimension	Classes	Total images
Merch	227 x 227	5	75
Flowers	224 x 224	5	3670
CIFAR-10	32 x 32	10	60,000
Digits	28 x 28	10	10,000
MNIST	28 x 28	10	70,000

IV. TERMINOLOGIES

There are terms used in the paper and what they entail for each one in the experiments. The terms are Channels, Filter, Stride, Padding, Dilation Factor, Output Size, Number of Neurons, Network Architecture, and Layer Size Calculation. Which are used from the MATLAB software [26].

We start by channels; if the number of channels is one, then the images are grayscale whereas it is three for colored images referring to the RGB spectrum. The filter convolves the input where a set of weights are applied to a specific image area. It can move either in horizontal or vertical directions or both at the same time. Stride determines the movement of the filter. The step size with which the filter moves determines the size of the output. Usually, the size of output reduces with increasing strides. The stride value is usually specified as an integer rather than a decimal. Padding refers to adding values on the border of images horizontally and/or vertically. It is used to control the output size of the layers. The dilation factor is the number of spaces used inside the filter. It expands the

filter by inserting zeros between the filter elements. The adequate size of the filter can be computed as $(Filter\ Size - 1) * Dilation\ Factor + 1$. A 3×3 filter with a dilation factor of 2 in horizontal and vertical directions is equivalent to a 5×5 filter having 0s between the elements. The output height and width of a convolutional layer can be computed using the following relation:

$$Output\ Size = \frac{I - (F-1)*DF + 1 + 2P}{S} + 1 \qquad (2)$$

where DF represents the Dilation Factor. This calculation needs to be done separately for the x- and y-direction if they are not the same. The output size needs to be an integer value and MATLAB discards the remaining part if it is not an integer.

The total number of neurons in a convolutional layer is $(h * w * c + 1) * Number\ of\ Filters$, where h and w are the height and width of the filter, c is the number of channels and 1 is the bias. A convolutional layer with five filters and a filter size of 3×3 for colored images has a total number of neurons equal to $(3 \times 3 \times 3 + 1) \times 5 = 140$. A Convolutional Neural Network (CNN) can have one or more layers depending on the size and complexity of the data. A small network with one or two layers is reasonable for small grayscale datasets whereas a complex architecture of layers is required for a dataset having millions of images. Considering the datasets used, we choose a series of 15-layer Convolutional Neural Network (CNN) architecture. It is an architecture of a deep network with all layers connected sequentially. Since we have categorical responses, softmax and classification layers are used at the end. A snapshot of this architecture when run on the Digits dataset in MATLAB is shown in Fig. 1. Also, the parameters used for such architecture are shown in Table II.
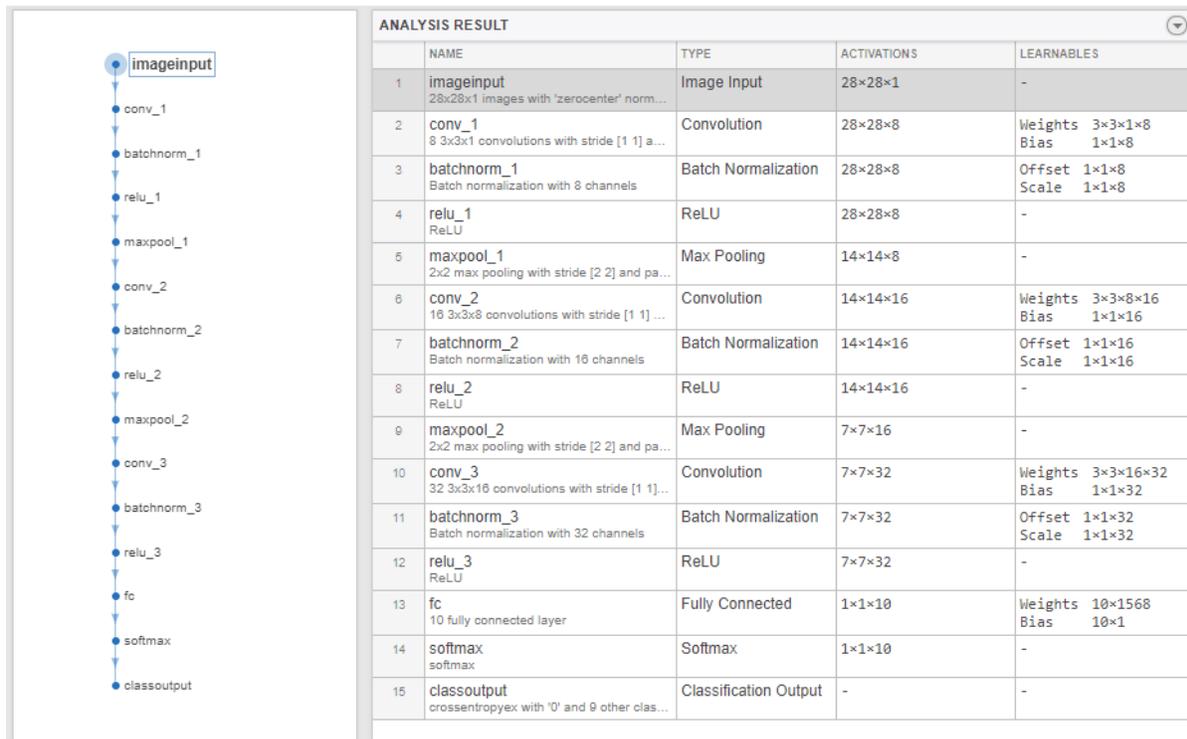
Fig. 1. 15-Layer CNN Architecture on Digits Dataset.

TABLE II.   PARAMETERS OF THE CNN USED

Parameter	Value
InitialLearnRate	0.01
MaxEpochs	6
MiniBatchSize	100

The padding and stride are considered the same in both directions in all experiments. The output size needs an integer value to cover the whole image. If it does not cover the whole image, then the MATLAB software ignores the remaining part along the right and bottom edges in the convolution. The objective of the experiments is to see the effect of this coverage on accuracies while changing the padding size and stride values. The dilation factor is set as one in all experiments. For example, suppose the image size is 32×32 and the filter of 5×5 is used in the convolution. If the padding value is two and the stride is 2 in each direction, then the output size is calculated as.

$$\left(32 - \left((5-1) * 1 + 1\right) + 2 * 2\right)/2 + 1 = (32 - 5 + 4)/2 + 1 = 16.5$$

In this case, the output size will become 16×16 and some of the zero paddings are discarded from the right and bottom of the image.

V.   IMPLEMENTATION OF MODEL 1

The initial weights are saved so that the same weights should be used for the implementation of both strategies. Also, this is to ensure that the comparison should be fair and should be affected by only cropping or padding and not by any other factors. Moreover, neither regularization nor momentum of any form was used for the same reason of fair comparison. The details of the CNN architecture employed for Model 1 in our implementation are provided in Table III.

TABLE III.   MODEL 1 ARCHITECTURE

Layer Type	Conv1	Pool1	Conv2	Pool2	FLC
Size of Filter	9×9	2×2	3×3	3×3	80×10
Depth of Filter	1	10	10	20	N/A
Number of Filters	10	N/A	20	N/A	N/A
Stride	1	2	1	3	N/A
Zero Padding	0	0	0	0	N/A
Input volume size: $28 \times 28 \times 1 \times 50$			Pooling type: mean		

To simplify the study, hyper-parameters of the last pooling layer are selected to investigate the impact of padding and cropping (this analysis can be applied to any other layer too). By using Eq. (1), the pre-calculated output sizes of every layer are integer values except for pool2 which is a float number 2.66. To see more details, let us run the CNN and see what the output size in the forward and backward pass will be and this is summarized in Table IV.

TABLE IV.   OUTPUT AND GRADIENT SIZES FOR MODEL 1

Forward Pass					
Output size	$20 \times 20 \times 10 \times 50$	$10 \times 10 \times 10 \times 50$	$8 \times 8 \times 20 \times 50$	$2 \times 2 \times 20 \times 50$	10×50
Layer x	**conv1**	**pool1**	**conv2**	**pool2**	**FLC**
$\dfrac{\partial L}{\partial x}$	$20 \times 20 \times 10 \times 50$	$10 \times 10 \times 10 \times 50$	$8 \times 8 \times 20 \times 50$	$2 \times 2 \times 20 \times 50$	10×50
Backward Pass					

The results of Table IV show the size of outputs at different layers and it can be noted that all sizes are tensors of order four except the first layer. The term $\frac{\partial L}{\partial x}$ evaluates the partial derivative of the loss function with respect to the output x which gives the expression in terms of the local error (δ_x) in that neuron (note that x here is a feature map of neurons). The local error of every neuron is multiplied by its input in order to backpropagate the error influence. This can be summarized as:

$$\frac{\partial L}{\partial W} = \delta_x * In \qquad (1)$$

For output layer $\delta_x = \delta_x \times f'(x)$ (2)

For any hidden layer $\delta_x = \delta_{x+1} * W_{x+1} \times f'(x)$ (3)

where:

$f'(.)$: the derivative of the activation function.

W_{x+1}: the weight matrix of the next layer.

δ_x: local error tensor of the current layer.

δ_{x+1}: local error tensor of next layer.

ln: is the input tensor to layer x.

$\frac{\partial L}{\partial W}$: the partial derivative of the loss with respect to the weights

Eventually, the size of local errors at layer x will be equal to the size of its output while the error is backpropagated through the layers. At this stage, we analyze what happens in the forward and backward pass only at the critical pool2 layer, as illustrated in Fig. 2.

The original sizes of tensors utilized is of order four. However, in Fig. 2, it can be noted that only the first and second dimensions of those tensors are demonstrated as these are the ones that are affected by different implementations.

1) Forward pass: During the forward pass, it can be noticed that the sliding 3×3 window with stride 3 on an input of 8×8 (which is the output size of the previous conv layer) can accommodate only the first six rows and columns. Thus, we need to truncate or crop the last two rows and columns to obtain a new input 6×6 for the pool2 layer and this matches the results of pool2 in Table IV.
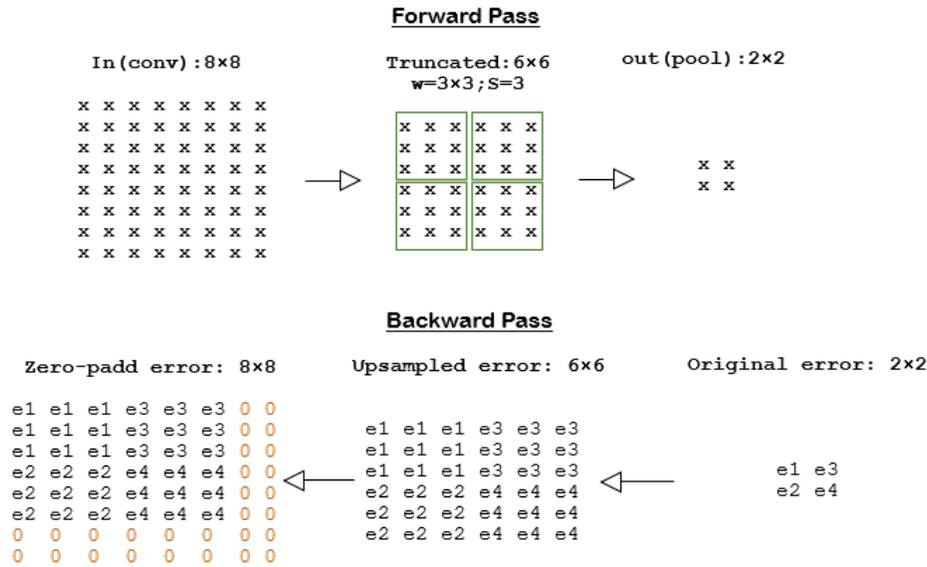
Fig. 2.   Backward and Forward Pass at Layer 4 for Model 1.

2) Backward pass: For the backward pass, using Eq.(5), it is found that the local error or the local gradient of pool2 layer has a size 2×2 as shown in Table IV. At the same layer, the local error δ_x of conv2 layer (previous layer) will be computed also using Eq. (5). The reason of doing this calculation at pool2 and not at the conv2 layer is the fact that in the forward pass from conv2 to pool2 the only process involves was down-sampling through pooling. Thus, there will be no change in weights or bias. Hence, it is easy to see that the local errors of conv2 can be obtained by up-sampling the local errors of pool2. Using Eq. (5), the local errors of conv2 is achieved by employing up-sampling on local errors of the pool2 layer which is shown in Fig. 2. Now, the *Conv2* local errors are of size 6×6. However, it must be the same size as the size of its output, i.e., 8×8. By employing the padding with zeros it is possible to accomplish the size requirement and this was done via a zero-padded error 8×8 in Fig. 2. In our case, all the weights connected to the right two columns and bottom two rows of conv2 output are not going to be updated because their local errors $\delta_x = 0$ and from Eq. (3) $\frac{\partial L}{\partial W} = 0$. This is due to the fact that these links did not contribute to the final loss as they were cropped in the forward pass. This is a major side effect of cropping inputs in the scenarios when our choice of hyper-parameter stride does not result in an appropriate output size. The next question to be investigated is how this will affect the predictive performance. The answer of this will be found in the ensuing section.

VI. IMPLEMENTATION OF MODEL 2

The architecture of this model will stay the same as the previous one with one change in the FLC weights size which will become 320×10. Table V contains the output sizes in the forward and backward pass (as before) after running CNN model 2. It can be noticed that the vectorized output of pool2

has the dimension $4 \times 4 \times 20 = 320$, which explains why the FLC layer weights were changed in this architecture. Fig. 3 illustrates how the forward and backward passes are implemented.

1) Forward pass: In this second strategy of hyper-parameters implementation, the input padding with four zeros is employed in contrast to the cropping process used in model 1 which results in an input of size 12×12. Hence, this makes the size compatible with the next layer. Eventually, pooling resulted in the output size of 4×4 as shown in Table V.

2) Backward pass: Again as was done in Model 1 implementation, the pool2 local errors are calculated using Eq. (5) which gave a size of 4×4 as shown in Table V. Next, up-sampling was employed to obtain a zero-padded error 12×12 as shown in Fig. 3. One important fact to be noted is that unlike the Model 1 implementation, in this case all weights of the neurons in conv2 are updated using Eq. (3). Hence, this can drastically impact on the predictive performance of model 2.

TABLE V.   OUTPUT AND GRADIENT SIZES FOR MODEL 2

Forward Pass					
Output size	20×20 $\times 10 \times 50$	10×10 $\times 10$ $\times 50$	8×8 $\times 20$ $\times 50$	4×4 $\times 20$ $\times 50$	10×50
Layer x	**conv1**	**pool1**	**conv2**	**pool2**	**FLC**
$\frac{\partial L}{\partial x}$	20×20 $\times 10 \times 50$	10×10 $\times 10$ $\times 50$	8×8 $\times 20$ $\times 50$	4×4 $\times 20$ $\times 50$	10×50
Backward Pass					
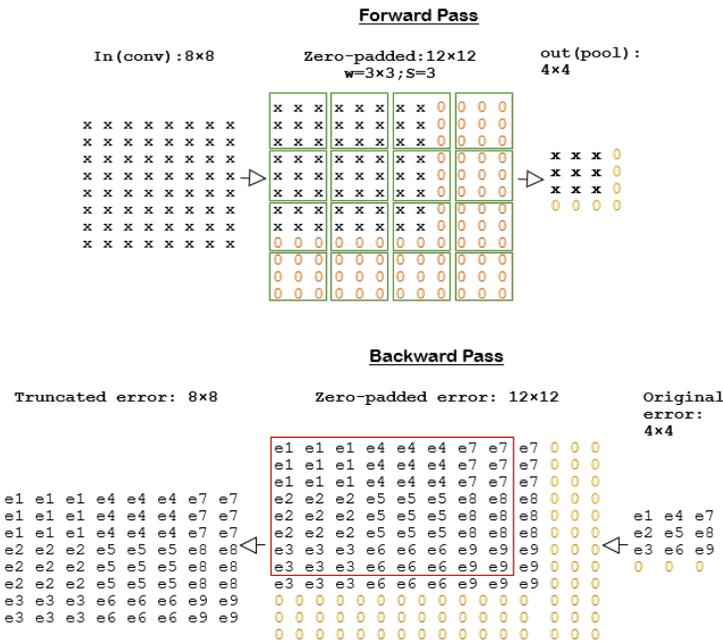
Fig. 3. Backward and Forward Pass at Layer 4 for Model 2.

VII. EXPERIMENTS AND RESULTS

In this section, the performance of the implemented Models 1 and 2 is compared in terms of loss, accuracy, and processing time. The objective of the experiments is to see the effect of padding size and stride at layer 10 which is a convolutional layer with the name 'conv3'. Filter size of three is used at the convolution layer. Table VI shows the size of the three layers used for Digits dataset. Here, the size of the convolution layer is calculated using the formula given in Eq. (2).

TABLE VI. SIZE OF LAYERS IN DIGITS DATASET

Image Dimension	Padding dimension	Stride Size	Conv. Layer Filter Size	Actual Output Size
28 x 28	0	1	5 x 5	5
		2	3 x 3	3
		3	2 x 2	2.33
		4	2 x 2	2
	1	1	7 x 7	7
		2	4 x 4	4
		3	3 x 3	3
		4	2 x 2	2.5
	2	1	9 x 9	9
		2	5 x 5	5
		3	3 x 3	3.66
		4	3 x 3	3

The experiments were performed on several datasets to observe the effect of changing padding and stride on all three types/sets of accuracies (training, validation, testing). The padding values from zero to four and stride values from one to four are used for all experiments. It is to be noted that the stride value of one is a trivial case and does not cause a problem but we keep it for comparison purposes. To achieve unbiased results, we first train the architecture with specific parameters such as the same padding size in all layers. After that, we get the weights of all layers out of which we freeze the first nine layers' weights and then see the effects of padding and stride starting from layer 10.

Results of the first experiment reported in Fig. 4 and Fig. 5 for Model 1 and Model 2, respectively, show that the algorithm is approximately converging in 20 epochs for both models. Next, Fig. 6 shows results at the final epoch for the two models. Here, it can be observed that model 1 has achieved a testing accuracy of 82.62% in contrast to 85.580% achieved by the Model 2. The same behavior can also be observed while comparing the training and validation accuracies of the two models. In Fig. 7 it is found that model 2 took almost 2 hours 30 minutes to train in contrast to the 2 hours 38 minutes of training by model 1. Thus, it is concluded that model 2 has a faster speed of convergence than model 1.
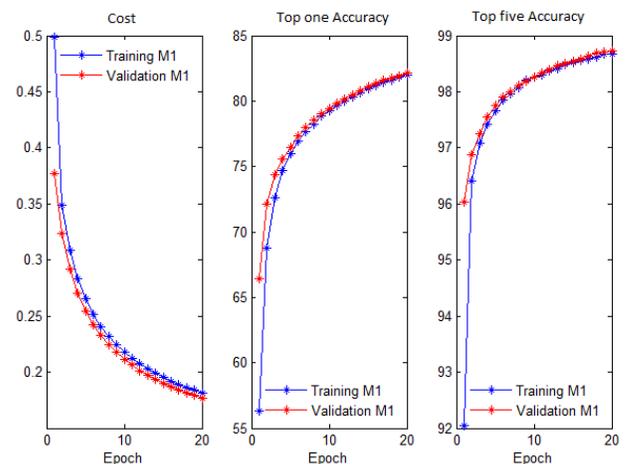
Fig. 4. Implementation of Model 1 Performance Measures.
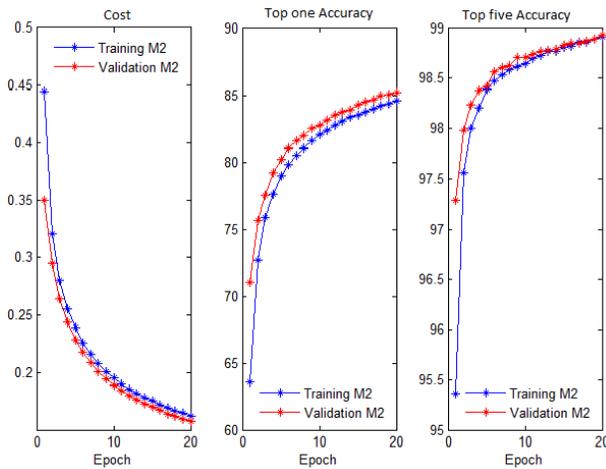
Fig. 5.   Implementation of Model 2 Performance Measures.

value is three, and it is minimum when the stride is 1. Hence, we can say that the minimum the stride, the lesser the difference between training and testing accuracy. This result is intuitive as well because more strides mean more skipping of the bits/values and less learning from the training set, which in turn reduces the testing accuracy. Similar results were observed for other datasets of Merch and CIFAR10 as can be seen in Fig. 11 to 13.

Another important observation from the results in Fig. 4 and Fig. 5 is that validation accuracy is a bit higher than training accuracy, especially in Fig. 5 showing that the trained model are more generalized in training.

In summary, it is observed that model 2 implementation has better classification accuracy, processing time, and generalization than the ones achieved by model 1. This is mainly due to the removal of the 25% conv2 neurons in the backward pass as cropping was employed. In contrast, padding with zeros used in model 2 keeps all the conv2 neurons and hence all conv2 parameters were utilized. Hence, contributions from all the neurons are included which enhances the predictive accuracy of the model.

Next, bar graphs are presented to see the effects of padding and strides in mainly the training and testing accuracies for the datasets used. The training accuracy is higher when the stride value is equal to 1, as shown in Fig. 8. In general, for all stride values, the training accuracy becomes better as the padding increases. The same is observed for testing accuracy in Fig. 9, and although not shown here, the validation accuracy also follows a similar pattern.

Fig. 7.   Processing Time of Models 1 and 2.
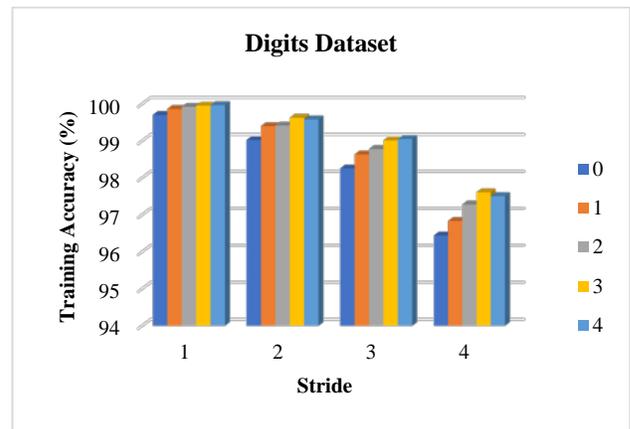
Fig. 8.   Training Accuracy Digits Dataset.

Fig. 6.   Classification Accuracy of Models 1 and 2.

By observing Fig. 10, the difference between the training and testing accuracy is more when the stride value is four for all padding values. The difference reduces when the stride
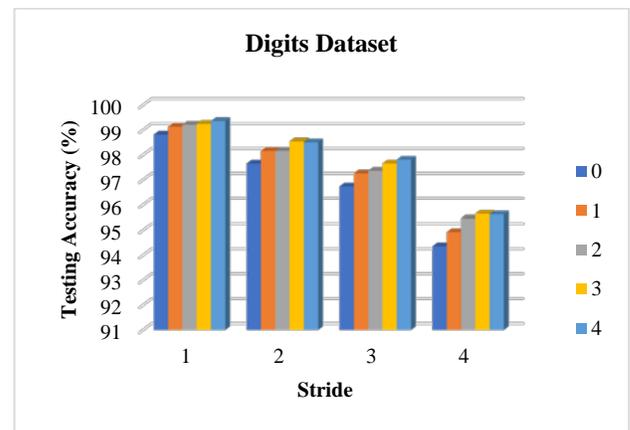
Fig. 9.   Testing Accuracy Digits Dataset.
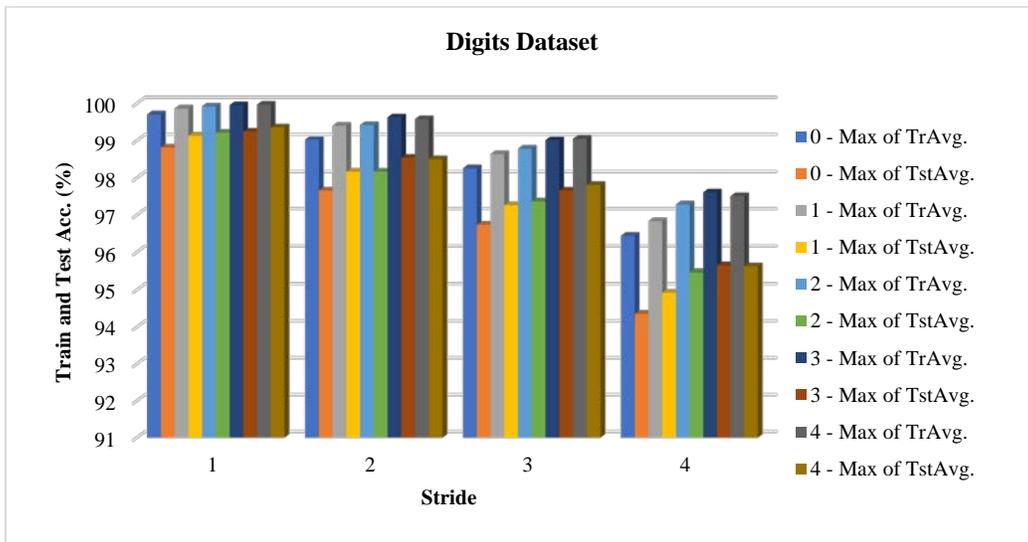
Fig. 10. Training and Testing Accuracy.
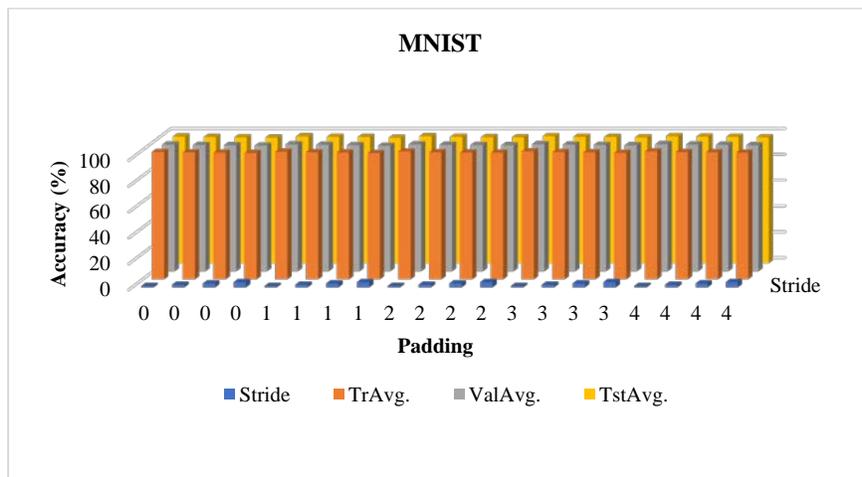
Fig. 11. FMNIST Accuracies.
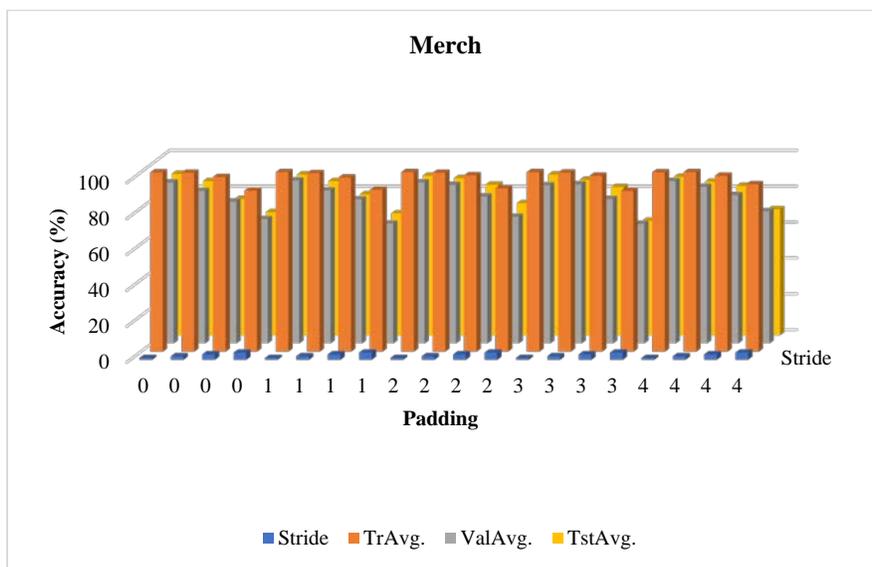
Fig. 12. Merch Accuracies.
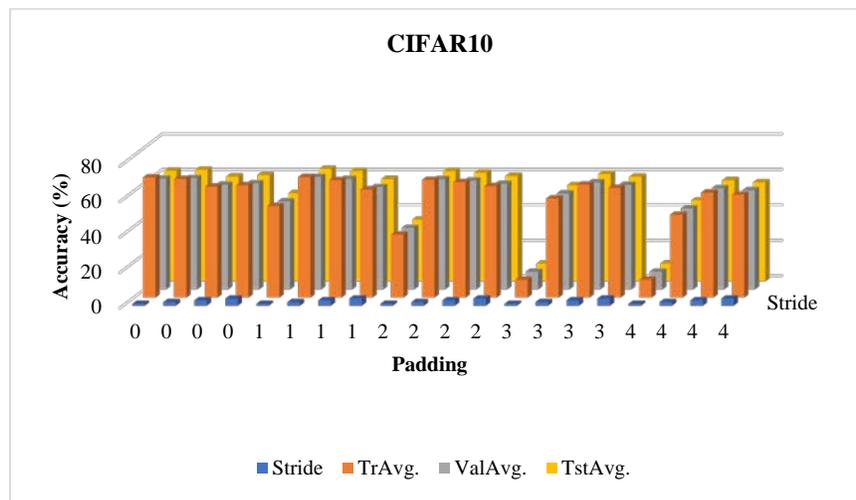
Fig. 13. CIFAR10 Accuracies.

VIII. CONCLUSION

In this work, we investigated an important issue as to how the performance of a CNN is affected by choice of hyper-parameters in the case when this choice does not fit the input volumes. It was investigated experimentally how the internal implementation of hyper-parameters particularly cropping and padding the input volumes are going to affect the performance measures of CNN. For this purpose, Digits, MNIST, Merch, Flowers, and CIFAR-10 datasets were analyzed and the two models were compared in terms of classification accuracy, processing time, and generalization. It was proved via various experiments that the model employing padding of input volume has higher accuracy with lesser training time in contrast to the model using cropping of input volume. Thus, it is concluded that the fair comparison of the performance for various CNN methods will be obtained when the hyper-parameters are set fairly.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

[1] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," Journal of Machine Learning Research, vol. 13, no. Feb, pp. 281-305,2012.

[2] G. I. Diaz, A. Fokoue-Nkoutche, G. Nannicini, and H. Samulowitz, "An effective algorithm for hyperparameter optimization of neural networks," IBM Journal of Research and Development, vol. 61, no. 4, pp. 9-1, 2017.

[3] J. F. Khaw, B. Lim, and L. E. Lim, "Optimal design of neural networks using the taguchi method," Neurocomputing, vol. 7, no. 3, pp. 225-245,1995.

[4] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in Advances in neural information processing systems, pp. 2951-2959, 2012.

[5] T. Domhan, J. T. Springenberg, and F. Hutter, "Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves.," in IJCAI, vol. 15, pp. 3460-8, 2015.

[6] E. Bochinski, T. Senst, and T. Sikora, "Hyper-parameter optimization for convolutional neural network committees based on evolutionary algorithms," in Image Processing (ICIP), 2017 IEEE International Conference on, pp. 3924-3928, IEEE, 2017.

[7] Y. Ozaki, M. Yano, and M. Onishi, "Effective hyperparameter optimization using nelder-mead method in deep learning," IPSJ Transactions on Computer Vision and Applications, vol. 9, no. 1, p. 20, 2017.

[8] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," arXiv preprint arXiv:1603.06560, 2016.

[9] E. Hazan, A. Klivans, and Y. Yuan, "Hyperparameter optimization: A spectral approach," arXiv preprint arXiv:1706.00764, 2017.

[10] A. F. Cardona-Escobar, A. F. Giraldo-Forero, A. E. Castro-Ospina, and J. A. Jaramillo-Garzon, "Effcient hyperparameter optimization in convolutional neural networks by learning curves prediction" in Iberoamerican Congress on Pattern Recognition, pp. 143-151, Springer, 2017.

[11] A. Klein, S. Falkner, S. Bartels, P. Hennig, F. Hutter, et al., "Fast bayesian hyperparameter optimization on large datasets," Electronic Journal of Statistics, vol. 11, no. 2, pp. 4945-4968, 2017.

[12] S. Albelwi and A. Mahmood, "Automated optimal architecture of deep convolutional neural networks for image recognition," in Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on, pp. 53-60, IEEE, 2016.

[13] I. Loshchilov and F. Hutter, "Cma-es for hyperparameter optimization of deep neural networks," arXiv preprint arXiv:1604.07269, 2016.

[14] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio,``An empirical evaluation of deep architectures on problems with many factors of variation," in Proc. 24th Int. Conf. Mach. Learn. (ICML), 2007, pp. 473_480.

[15] J. Bergstra and Y. Bengio, ``Random search for hyper-parameter optimization," J. Mach. Learn. Res., vol. 13, pp. 281_305, Feb. 2012.

[16] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, ``Algorithms for hyperparameter optimization,'' in Proc. Adv. Neural Inf. Process. Syst., 2011, pp. 2546_2554.

[17] F. Hutter, H. H. Hoos, and K. Leyton-Brown, ``Sequential model-based optimization for general algorithm con_guration,'' in Proc. Int. Conf. Learn. Intell. Optim. Berlin, Germany: Springer, 2011, pp. 507-523.

[18] M. W. Hoffman and B. Shahriari, ``Modular mechanisms for Bayesian optimization,'' in Proc. NIPS Workshop Bayesian Optim., 2014, pp. 1-5.

[19] B. Wang, Y. Sun, B. Xue, and M. Zhang, ``A hybrid differential evolution approach to designing deep convolutional neural networks for image classification,'' in Proc. Australas. Joint Conf. Artif. Intell. Cham, Switzerland: Springer, 2018, pp. 237-250.

[20] W.-Y. Lee, S.-M. Park, and K.-B. Sim, ``Optimal hyperparameter tuning of convolutional neural networks based on the parameter-setting-free harmony search algorithm,'' Optik, vol. 172, pp. 359_367, Nov. 2018.

[21] B. Baker, O. Gupta, N. Naik, and R. Raskar, ``Designing neural network architectures using reinforcement learning,'' 2016, arXiv:1611.02167. [Online]. Available: http://arxiv.org/abs/1611.02167.

[22] P. Neary, ``Automatic hyperparameter tuning in deep convolutional neural networks using asynchronous reinforcement learning,'' in Proc. IEEE Int. Conf. Cognit. Comput. (ICCC), Jul. 2018, pp. 73-77.

[23] A. Gülcü and Z. KUş, "Hyper-Parameter Selection in Convolutional Neural Networks Using Microcanonical Optimization Algorithm," in IEEE Access, vol. 8, pp. 52528-52540, 2020, doi: 10.1109/ACCESS.2020.2981141.

[24] U. M. Al-Saggaf, A. Botalb, M. Moinuddin, S. A. Alfakeh, S. S. A. Ali and T. T. Boon, "Either crop or pad the input volume: What is beneficial for Convolutional Neural Network?," 2020 8th International Conference on Intelligent and Advanced Systems (ICIAS), 2021, pp. 1-6, doi: 10.1109/ICIAS49414.2021.9642661.

[25] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," arXiv preprint arXiv:1603.07285, 2016.

[26] "Specify Layers of Convolutional Neural Network," Specify Layers of Convolutional Neural Network - MATLAB & Simulink, 2022. [Online]. Available: https://www.mathworks.com/help/deeplearning/ug/layers-of-a-convolutional-neural-network.html.

Design of a Speaking Training System for English Speech Education using Speech Recognition Technology

Hengheng He

School of Foreign Languages, Hankou University, Wuhan, Hubei 430212, China

Abstract—A good English speaking training system can provide an aid to the learning of English. This paper briefly introduced the English speaking training system and described the speaking training scoring and pronunciation resonance peak display modules in the system. The speaking training scoring module scored pronunciation with the Long Short-Term Memory (LSTM). The pronunciation resonance peak display module extracted the resonance peak with Fourier transform and visualized it. Finally, the speaking scoring module, the pronunciation resonance peak display module, and the effect of the whole system in improving students' speaking pronunciation was tested. The results showed that the LSTM-based speaking scoring algorithm had highest scoring accuracy than pattern matching and the recurrent neural network (RNN) algorithm, and its accuracy was 95.21% when scoring the LibriSpeech dataset and 90.12% when scoring the local English dataset. The pronunciation resonance peak display module displayed the change of mouth shape before and after training, and the pronunciation after training was closer to the standard pronunciation. The P value in the comparison of the speaking level before and after training with the system was 0.001, i.e., the difference was significant, which indicated that the students' English speaking proficiency significantly improved.

Keywords—English speech; long short-term memory; speaking training; speech recognition

I. INTRODUCTION

In order to ensure that information is communicated as accurately as possible during national exchanges, a common language is needed. English is one of the common languages for international communication [1]. Educational institutions are also paying more and more attention to English teaching, especially spoken English. The biggest role of spoken English is to communicate, and it is difficult to communicate fluently just by reading and writing. The pronunciation quality of oral English will directly affect communication efficiency [2]. In the traditional English teaching mode, classroom teaching is the main focus, and students often learn what the teacher teaches them. The lack of time for oral training in the classroom and the teacher's irregular pronunciation will affect the learning effect of students' oral language [3]. Information technology is gradually integrated into the traditional teaching mode and changes the teacher-oriented structure of the traditional teaching mode. When using information technology to assist oral English training, teachers can still teach relevant speaking knowledge in the classroom, and students can train their speaking independently through the

independent oral English learning system under information technology before or after class [4]. The oral English training system can provide the standard pronunciation of English, collect students' pronunciation, evaluate the pronunciation through the scoring module in the system, and help students adjust their pronunciation [5]. The oral English scoring module is the core of the oral training system; the higher the scoring accuracy, the higher the reference value for the oral training. This paper studied the speaking English training system that adopted the speech recognition technology in order to enhance the students' speaking English level. The LSTM was used to score the students' pronunciation in the speaking scoring module of the speaking training system. The resonance display module used the Fourier transform to extract and visualize the resonance peaks during pronunciation to assist in pronunciation training. The two modules of the speaking training system were tested. The effectiveness of the speaking training system in improving students' speaking skills was tested. The final results showed that the LSTM had higher scoring accuracy than the pattern matching and RNN algorithm, the pronunciation resonance peak display module effectively visualized the resonance peaks of the spoken pronunciation, and the students who used the speaking training system for learning had significantly improved English speaking skills. The research on the English speaking training system and the experimental results of this paper provide an effective reference for improving students' English speaking skills. The limitations of this paper are that only the scoring module and pronunciation visualization module were tested and there were few test subjects in the test of the performance of the speaking training system. The future research direction is to conduct in-depth research on the scoring module and pronunciation visualization module of the training system and to increase the scale of test subjects.

This paper is organized in the order of abstract, introduction, and literature review, introduction of English speaking training system, simulation experiments, discussion, conclusion, and references.

II. LITERATURE REVIEW

Hsieh et al. [6] used a technology acceptance model to investigate the effect of Line on speaking training in a non-native English teaching environment and found that Line had a more positive effect on English speaking instruction than traditional classroom instruction. Reitz et al. [7] embedded the English learning process into a generic 3D cooperative virtual

reality (VR) game and found through example analysis that the designed VR game effectively trained students' English communication skills. Cao et al. [8] proposed a lip movement judgment algorithm based on ultrasonic detection to aid spoken English pronunciation. They performed experiments and found that the system had a speech accuracy of 85%, i.e., it could improve the English speaking trainers to a certain extent.

III. ENGLISH SPEAKING TRAINING EVALUATION SYSTEM

A. The Basic Structure of English Speaking Training System

The basic architecture of the English speaking training system is shown in Fig. 1. The speech acquisition and playback module [9] is equivalent to the ear and mouth of the training system. The acquisition module is responsible for acquiring the audio signal of the spoken language, and the collected audio will be stored in the buffer for later processing. The playback module plays the stored English standard pronunciation to provide a reference for the students [10].

The primary function of the pronunciation resonance peak image display module is to graphically display the standard spoken audio and the student's spoken audio so that students can clearly see the difference between their pronunciation and the standard pronunciation to adjust their pronunciation [11].

The main function of the spoken pronunciation scoring module is to score the pronunciation of the spoken audio signals collected by the audio acquisition module, i.e., to rate the standard level of students' spoken English pronunciation and to make a quantitative analysis of the students' speaking training level. The scores can be used as a feedback incentive for the students' speaking training [12].

B. Speech Recognition-based Spoken Pronunciation Scoring Module

For the whole English speaking training system, the speaking pronunciation scoring module is the core, and its main function is to quantify the students' speaking pronunciation level to give a feedback incentive to the students' speaking training [13]. First, the spoken speech signal is collected using the speech acquisition module; then, features are extracted from the speech signal after preprocessing; finally, the speech is scored according to the features [14]. This paper used LSTM to score the speech collected by the speech acquisition module, and the basic flow is shown in Fig. 2.
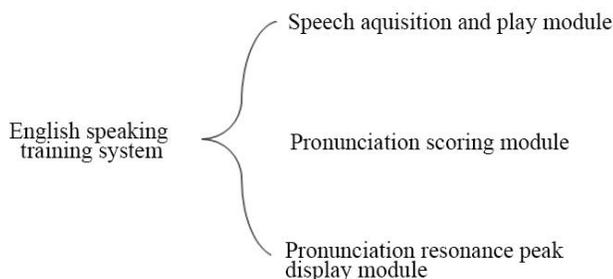
Fig. 1. The Basic Architecture Module of English Speaking Training System.
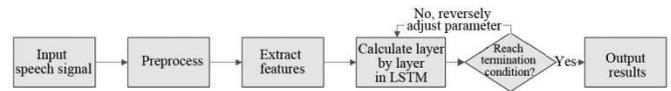
Fig. 2. The Workflow of the Pronunciation Scoring Module.

The flow of spoken pronunciation scoring based on LSTM speech recognition is shown in Fig. 2.

1) The speech signal is collected.

2) Pre-processing such as filtering, windowing, and framing [15] is performed.

3) Features are extracted from the pre-processed pronunciation signal using the Mel-frequency cepstral coefficient [16] to extract audio features.

4) The Mel-frequency cepstral coefficient features of every audio frame are input into the LSTM in order for forward calculation [17]:

$$\begin{cases} i_t = g(\omega_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(\omega_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \\ f_t = g(\omega_f \cdot [h_{t-1}, x_t] + b_f) \\ o_t = g(\omega_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \cdot \tanh(C_t) \end{cases} \quad (1)$$

where x_t represents the current input of the cell, h_{t-1} is the last hidden state, C_{t-1} is the last cell state, i_t is the weight that determines the new information to be remembered, $\tilde{C}_t$ is the cell state of the new information added [18], C_t refers to the current cell state after the new information is added, ω_i and ω_t are weights, b_i and b_C are biases [19], f_t is the weight of the information not to be forgotten in C_{t-1}, ω_f is the weight, b_f is the bias, o_t represents the weight that determines the final output information amount, and h_t refers to the final output or the next hidden state [20].

5) Whether the training of the algorithm reaches the termination condition is determined. If the termination condition is reached, the training is finished, and the parameters of the LSTM are fixed [21]. When applied to the actual pronunciation scoring, the extracted features of pronunciation are input into the LSTM in order to get the scoring results. If the termination condition is not reached, the parameters in the LSTM are adjusted reversely using the stochastic gradient descent method [22]. The termination conditions include: (1) the number of algorithm iterations reaches a preset number; (2) the error between the forward calculation result and the expected result converges to the preset threshold.

Fig. 3.   Workflow of the Pronunciation Resonance Peak Image Display Module.

C. Pronunciation Resonance Peak Image Display Module

The pronunciation scoring module in the speaking training system quantifies students' pronunciation level, but this quantification only converts the level of pronunciation into a number, which does not reflect the students' pronunciation process visually [23]. The scoring module can only give a final target, and it is difficult to guide students to correct their pronunciation directly. Therefore, there is a need for a module that can visually assist in pronunciation correction. The pronunciation resonance peak image display module is a module that can visually display the pronunciation process of students.

The human vocal tract and the oral cavity together form a resonance cavity. After the sound wave signal of the vocal cord vibration is filtered by the resonance cavity, the energy will be redistributed in different frequencies. When the mouth shape changes, the resonance cavity will also change; thus, playing a different filtering effect to change the pronunciation. The connection between the mouth shape and the resonance peaks makes it possible to guide the mouth shape and correct the pronunciation based on the changes in the resonance peaks. The workflow of the resonance peak image display module in the speaking training system is shown in Fig. 3.

1) The speech signal is input and pre-processed by filtering, windowing, and framing [24].

2) The spectrum of a single-frame speech signal is obtained by using Fast Fourier Transform (FFT).

3) The maximum resonance peak of a single-frame speech signal is calculated.

4) An image is drawn to reflect the change in students' pronunciation. Time is the horizontal coordinate axis, and the smallest unit of the horizontal coordinate axis is a frame. Frequency is the vertical coordinate axis. The line graphs are plotted in the coordinate chart in the chronological order of the speech signals and the frequency of the maximum resonance peak of every speech signal frame. The difference between students' pronunciation and standard pronunciation can be seen visually when the line graphs of students' pronunciation and standard pronunciation are placed in the same coordinate plane [25].

The combination of pronunciation scores and resonance peak comparison charts can guide students to correct their pronunciation.

IV.   SIMULATION EXPERIMENTS

A. Experimental Data

The LibriSpeech dataset (openslr.org/12/) was used for simulation experiments, which includes 1000 h of audio data. It is a dataset of audiobooks, including texts and speeches. The sampling rate of the audio data in this dataset was 16 kHz.

After subdivision and collation, every audio in the dataset was about 10 s long.

In addition to the above English speech dataset collected from the public dataset, this paper also collected English spoken pronunciation from students to construct a local English speech dataset in order to verify the actual effect of the speaking training system on students' speaking scores. Fifty students, including 25 males and 25 females, participated in the English pronunciation collection. Every student read aloud 15 randomly selected non-repeated sentences from the common oral English sentence database. The read-aloud speech was captured in a recording room using recording software with a sampling rate of 16 kHz. The sentences included:

1) How do you feel?

2) What's the weather like?

3) See you tomorrow.

B. Experimental Projects

1) Scoring module test of the speaking training system: In order to test the scoring performance of the proposed LSTM speech recognition-based speaking training system, the scoring performance of the pattern matching-based and Recurrent Neural Network (RNN) speech recognition-based speaking training system was tested in addition to the accuracy detection experiment.

The scoring algorithm of the pattern matching-based speaking training system used a dynamic time regularization algorithm to calculate the minimum matching distance of extracted features between oral pronunciation and standard pronunciation. The score was calculated based on the minimum matching distance.

The relevant parameters of the speaking training scoring algorithm based on RNN speech recognition are as follows. Thirteen input nodes, 100 hidden nodes, a sigmoid function, and one output node were used. The stochastic gradient descent was used for backward learning. The learning rate was 0.1.

The relevant parameters of the speaking training scoring algorithm based on LSTM speech recognition are as follows. The number of nodes in the LSTM input layer was set as 13. The number of node cells in the hidden layer was set as 100, and there was an input gate, forgetting gate, and output gate in every node cell. The activation function for the node cells was the sigmoid function. The number of nodes in the output layer was set as 1, and the softmax function was used. The stochastic gradient descent method is used to reverse the parameters in the node cells of the hidden layer, the learning rate was set as 0.1, and the maximum number of iterations was set as 1000.

In the process of testing the scoring accuracy of the above three scoring algorithms, the scores calculated by the algorithms were compared with the standard scores. The mean value of manual scoring by 20 experts was used as the standard score. The scoring accuracy of the algorithms is calculated as follows:

$$\begin{cases} R_i = 1 - \dfrac{|S_{i1} - S_{i2}|}{S_{i2}} \\ R = \dfrac{\sum\limits_{i=1}^{n} R_i}{n} \end{cases} \qquad (2)$$

where R_i is the score similarity of the algorithm for the i-th sample, R is the score accuracy of the algorithm, S_{i1} is the score given by the algorithm for the i-th sample, S_{i2} is the expert manual score for the i-th sample, and n is the number of test samples.

2) Testing of the pronunciation resonance peak image display module of the speaking training system: The standard pronunciation of the local English speech dataset was input in the pronunciation resonance peak image display module of the speaking system. Then, the pronunciation of the students' pronunciation of the local English speech dataset before receiving training by the speaking training system and the pronunciation after receiving training by the system were input into the module to output the pronunciation resonance peak comparison graph.

3) Testing of the improvement of students' speaking levels after training by the speaking training system: One hundred students were randomly selected from the School of Foreign Languages of Hankou University and were divided into two groups: a control group and an experimental group. The control group used the traditional teaching method for English speaking training, while the experimental group used the speaking training system in addition to the traditional teaching method for speaking training. The students in both groups were taught to speak for two weeks, and their speaking levels were scored by 20 experts before and after the teaching, using a hundred-mark system.

C. Test Results

To verify the scoring accuracy of the LSTM speech recognition-based speaking training scoring algorithm, the LibriSpeech dataset and the local English speech dataset were used for scoring accuracy testing, and it was also compared with the two speaking training scoring algorithms based on pattern matching and RNN speech recognition. The test results are shown in Fig. 4. For the LibriSpeech dataset, the accuracy of the pattern matching-based scoring algorithm was 79.69%, the accuracy of the RNN speech recognition-based scoring algorithm was 89.65%, and the accuracy of the LSTM speech recognition-based scoring algorithm was 95.21%. The corresponding accuracy for the local English speech dataset was 72.34%, 81.33 %, and 90.12%, respectively. It was seen from the comparison in Fig. 4 that the speaking training scoring algorithm based on LSTM speech recognition had the highest accuracy, followed by the speaking scoring algorithm based on RNN speech recognition, and the speaking training scoring algorithm based on pattern matching had the lowest accuracy when scoring the same speech dataset. In addition,

the accuracy of the three speaking training scoring algorithms was higher when scoring the LibriSpeech dataset.

Due to the space limitation, only a resonance peak graph for comparing a student's pronunciation of an English sentence with the standard pronunciation is shown here, as shown in Fig. 5. Fig. 5 shows the changes in the resonance peak of the students' pronunciation over time. Because of the connection between the resonance peak and the mouth shape, the graph also qualitatively reflected the changes in the student's mouth shape over time. The resonance peak line graph comparison demonstrated that the resonance peak of the student's pronunciation was larger than that of the standard pronunciation, indicating that the student's tongue position was low and the mouth was opened too wide during pronunciation. Therefore, students should raise the tongue position and reduce the mouth shape. After the training, the resonance peak broken line of the student's pronunciation almost coincided with that of the standard pronunciation, indicating that the mouth shape training was effective.

In order to verify the effect of the LSTM speech recognition-based speaking training system in improving pronunciation, 100 students were randomly selected from the Foreign Language Institute of Hankou University and then divided into two groups. The control group was taught traditional speaking, and the experimental group was trained with the speaking training system in addition to traditional speaking teaching. The speaking level of students in both groups was scored by 20 experts before and after receiving the teaching, and the results are shown in Table I.
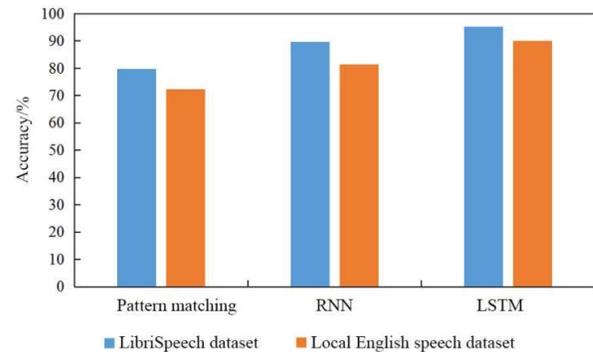
Fig. 4. Scoring Accuracy of Three Speaking Training Scoring Algorithms for Two Speech Datasets.
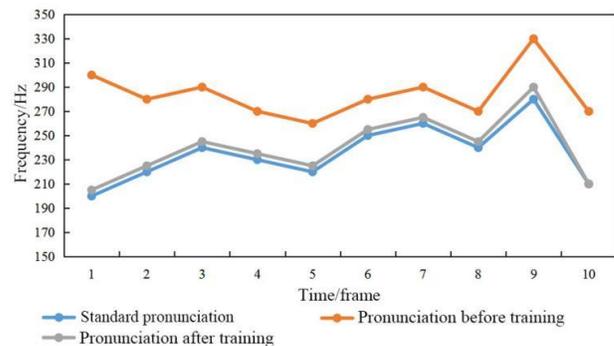
Fig. 5. Comparison of the Resonance Peak between the Student's Pronunciation before and after Training and the Standard Pronunciation.

	Control group	Experimental group	P value
Pre-teaching	75.1 ± 5.2	75.3 ± 6.1	0.126
After teaching	78.6 ± 5.7	92.1 ± 1.5	0.001
P value	0.144	0.001	

It was seen from Table I that there was no significant difference between the speaking level of the control group and the experimental group before conducting the speaking instruction. After the teaching, there was a significant difference between the speaking level of the control and experimental groups, and the speaking level of the experimental group was significantly higher than that of the control group.

It was found from the comparison of the average speaking score of the same group before and after the teaching that the P value of the difference between the average speaking score of the control group before and after the teaching was 0.144, which was greater than 0.05, i.e., the speaking level of the control group was not significantly improved after the teaching; the P value of the difference between the average speaking score of the experimental group before and after the teaching was 0.001, which was less than 0.05, i.e., the speaking level of the experimental group was significantly improved after the teaching.

V.   DISCUSSION

Students need to pronounce English successfully in addition to being able to read and write English correctly in the process of learning English. As English is a language for communication, the level of spoken English is somehow more important than the level of English reading and writing. However, in the process of learning spoken English, a standard pronunciation is needed as a reference. In the traditional teaching of spoken English, the teacher usually pronounces the words, the students follow the teacher's pronunciation, and the teacher corrects the students' pronunciation. In this teaching mode, the students' speaking training effect depends on the teacher's speaking level, and it is impossible for the teacher to teach one-on-one.

With the development of speech recognition technology, it has been gradually applied to various fields, including English speaking training. The English speaking training system built with speech recognition technology can evaluate students' pronunciation by the scoring module and visualize their pronunciation characteristics by the pronunciation resonance peak display module. With these two modules, students can evaluate their own speaking level according to the standard pronunciation given by the system and adjust their pronunciation according to the differences in pronunciation characteristics. The scoring module and pronunciation resonance peak display module were tested in the simulation experiment, and the effect of the speaking training system in improving the students' English speaking level has been studied above.

The results of the accuracy test of the scoring module showed that the accuracy when using the LSTM for

pronunciation evaluation was higher than the pattern matching and the RNN algorithm. The reason is as follows. Although the dynamic time regularization method was used in the pattern matching-based speaking training scoring algorithm to reduce the difficulty of matching due to duration randomness, some information was lost in the process of stretching or compressing the audio, and the standard speech templates in the template library were relied on, so it had better scoring accuracy for the LibriSpeech dataset containing a larger amount of data. The RNN speech recognition-based speaking training scoring algorithm used the RNN to score speech. Compared with pattern matching, the RNN did not require standard templates for scoring but directly scored the pronunciation according to the pattern obtained during training, which was more efficient. Moreover, the activation function in the RNN effectively fit the nonlinear pattern between pronunciation features and scoring, so it was more accurate than pattern matching. Compared with the RNN speech recognition-based algorithm, the LSTM speech recognition-based speaking training scoring algorithm used an activation function that can effectively fit the nonlinear pattern, but the introduced forgetting gate unit avoided the gradient explosion when facing long data, so it had better accuracy when scoring long speech.

The test results of the pronunciation resonance peak display module showed that the module could visualize the resonance peaks of students' spoken pronunciation. The resonance peaks of pronunciation before and after training were compared with those of standard pronunciation, and the results showed that the resonance peaks of pronunciation after training were closer to those of standard pronunciation. Taking the results presented in Fig. 5 as an example, the resonance peaks of the students' pronunciation before training were larger than those of the standard pronunciation, indicating that the tongue position was low and the mouth opened too wide during the pronunciation process. Thus, the tongue position needed to be raised, and the mouth shape should be smaller during the training process. The resonance peaks after training also showed the effectiveness of adjusting the pronunciation.

The results of testing the effectiveness of the speaking training system showed that the experimental group that applied the speaking training system had a significant improvement in their pronunciation after teaching compared to the control group that adopted the traditional teaching mode. The reason is as follows. In the traditional teaching mode, students adjusted their own pronunciation according to the teacher's pronunciation; the pronunciation of the whole class would be affected if the teacher's pronunciation was wrong. Moreover, different students had different pronunciation habits, so it was difficult for the teacher to provide targeted tutoring to students on a one-to-one basis, and the common tutoring would make some students have difficulty in keeping up with the learning pace. When using the speaking training system, the students adjusted their pronunciation independently with the standard pronunciation as the target based on the scoring module and pronunciation visualization module, which was considered targeted training.

VI. CONCLUSION

This paper briefly introduced the English speaking training system and described the scoring algorithm in the speaking training scoring module of the system and the pronunciation resonance peak display module in the system. The pronunciation scoring module and the pronunciation resonance peak display module in the speaking training system were tested. In addition, the speaking level of the students that were trained traditionally and trained by the speaking training system was compared before and after the teaching. The results are as follows. (1) When scoring the same speech dataset, the LSTM-based speaking scoring algorithm was the most accurate, the RNN-based scoring algorithm was the second most accurate, and the pattern matching-based scoring algorithm was the lowest. (2) The three algorithms achieved higher accuracy when scoring the LibriSpeech dataset. (3) The pronunciation resonance peak display module effectively displayed the line graph of the resonance peak of the student's pronunciation over time and visually reflected the difference of the resonance peak between students' pronunciation and standard pronunciation. (4) The difference in the speaking level between the control group and the experimental group before receiving instruction was not significant; the speaking level of the control group improved insignificantly after receiving traditional speaking instruction, and the speaking level of the experimental group improved significantly and was significantly higher than that of the control group after being trained by the speaking training system.

REFERENCES

[1]   J. Lubek, "The Cost of Training: Oral and Maxillofacial Surgery at a Crossroad," Or. Surg. Or. Med. Or. Pa., vol. 127, pp. 465-467, March 2019.

[2]   W. L. Martens, and R. Wang, "Applying adaptive recognition of the learner's vowel space to English pronunciation training of native speakers of Japanese," SHS Web Conf., vol. 102, pp. 1-8, January 2021.

[3]   J. Cai, and Y. Liu, "Research on English pronunciation training based on intelligent speech recognition," Int. J. Speech Technol., vol. 21, pp. 633-640, September 2018.

[4]   M. Suganuma, T. Yamamura, Y. Hoshino, and M. Yamada, "Proposal of the Way of English Pronunciation Training Evaluation by Lip Movement," J. Jpn. Pers. Comput. Appl. Technol. Soc., vol. 11, pp. 8-20, 2017.

[5]   T. Yoshioka, S. Karita, and T. Nakatani, "Far-field speech recognition using CNN-DNN-HMM with convolution in time," IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 4360-4364, April 2015.

[6]   J. Hsieh, Y. M. Huang, and W. Wu, "Technological acceptance of LINE in flipped EFL oral training," Comput. Hum. Behav., vol. 70, pp. 178-190, December 2016.

[7]   L. Reitz, A. Sohny, and G. Lochmann, "VR-Based Gamification of Communication Training and Oral Examination in a Second Language," Int. J. Game-Based Learn., vol. 6, pp. 46-61, April 2016.

[8]   Q. Cao, and H. Hao, "Optimization of Intelligent English Pronunciation Training System Based on Android Platform," Complexity, vol. 2021, pp. 1-11, March 2021.

[9]   Y. Wang, F. Bao, H. Zhang, and G. Gao, "Research on Mongolian Speech Recognition Based on FSMN," Natl. CCF Conf. on Natural Language Processing and Chinese Computing, pp. 243-254, January 2017.

[10] C. Agarwal, and P. Chakraborty, "A review of tools and techniques for computer aided pronunciation training (CAPT) in English," Educ. Inf. Technol., vol. 24, pp. 3731-3743, November 2019.

[11] F. G. Jonas, "Podcast-based pronunciation training: Enhancing FL learners' perception and production of fossilised segmental features," ReCALL, vol. 31, pp. 150-169, April 2019.

[12] Y. Sun, X. Jiang, "The design and application of English pronunciation training software based on Android intelligent mobile phone platform," Rev. Fac. Ing., vol. 32, pp. 756-765, January 2017.

[13] M. J. Alam, V. Gupta, P. Kenny, and P. Dumouchel, "Speech recognition in reverberant and noisy environments employing multiple feature extractors and i-vector speaker adaptation," Eurasip J. Adv. Sig. Pr., vol. 2015, pp. 1-13, December 2015.

[14] F. G. Jonas, "Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app," Lang. Learn. Technol., vol. 24, pp. 62-85, February 2020.

[15] J. Szpyra-Kozowska, and S. Stasiak, "Verifying a holistic multimodal approach to pronunciation training of intermediate Polish learners of English," Lublin Stud. Mod. Lang. Lit., vol. 40, pp. 181-198, July 2016.

[16] L. Hsu, "A Longitudinal View of Look at the Effectiveness of Elicited Imitation with Computer Assisted Pronunciation Training (CAPT)," Int. Res. Educ., vol. 4, March 2016.

[17] N. Kleynhans, W. Hartman, D. V. Niekerk, C. J. van Heerden, R. Schwartz, S. Tsakalidis et al., "Code-switched English pronunciation modeling for Swahili spoken term detection," Proc. Comput. Sci., vol. 81, pp. 128-135, December 2016.

[18] L. Tian, D. F. Wong, L. S. Chao, P. Quaresma, F. Oliveira, S. Li et al., "UM-Corpus: A Large English-Chinese Parallel Corpus for Statistical Machine Translation," Proc. of the 9th International Conference on Language Resources and Evaluation (LREC'14), January 2014.

[19] N. Hammami, M. Bedda, and F. Nadir, "The second-order derivatives of MFCC for improving spoken Arabic digits recognition using Tree distributions approximation model and HMMs," Int. Conf. on Communications and Information Technology, pp. 1-5, June 2012.

[20] M. E. Celebi, H. A. Kingravi, and P. A. Vela, "A comparative study of efficient initialization methods for the k-means clustering algorithm," Expert Syst. Appl., vol. 40, pp. 200-210, January 2013.

[21] N. Leema, H. K. Nehemiah, and A. Kannan, "Neural Network Classifier Optimization using Differential Evolution with Global Information and Back Propagation Algorithm for Clinical Datasets," Appl. Soft Comput., vol. 49, pp. 834-844, August 2016.

[22] J. Suntornsawet, "Problematic Phonological Features of Foreign Accented English Pronunciation as Threats to International Intelligibility: Thai EIL Pronunciation Core," J. Eng. Int. Lang., vol. 2019, December 2019.

[23] J. Cao, H, Cui, H. Shi, and L. Jiao, "Big Data: A Parallel Particle Swarm Optimization-Back-Propagation Neural Network Algorithm Based on MapReduce," Plos One, vol. 11, pp. 1-17, June 2016.

[24] Z. Xu, J. Liu, X. Chen, Y. Wang, and Z. Zhao, "Continuous blood pressure estimation based on multiple parameters from eletrocardiogram and photoplethysmogram by Back-propagation neural network," Comput. Ind., vol. 89, pp. 50-59, August 2017.

[25] H. Zhou, Z. Deng, Y. Xia, and M. Fu, "A new sampling method in particle filter based on Pearson correlation coefficient," Neurocomputing, vol. 216, pp. 208-215, July 2016.

Development of Underwater Pipe Crack Detection System for Low-Cost Underwater Vehicle using Raspberry Pi and Canny Edge Detection Method

Mohd Aliff[1], Nur Farah Hanisah[2], Muhammad Shafique Ashroff[3], Sallaudin Hassan[4]
Quality Engineering Research Cluster (QEREC), Universiti Kuala Lumpur, 81750, Johor, Malaysia

Siti Fairuz Nurr[5]
Faculty of Plantation and Agrotechnology
Universiti Teknologi MARA, Cawangan Melaka Kampus
Jasin, 77300 Merlimau, Melaka, Malaysia

Nor Samsiah Sani[6]
Center for Artificial Intelligence Technology
Universiti Kebangsaan Malaysia
Selangor, Malaysia

Abstract—The effective loading area decreases because of cracking, leading to a rise in stress and eventual structural failure. Monitoring for cracks is an important part of keeping any pipeline or building in excellent working order. There are several obstacles that make manual inspection and monitoring of subsea pipes challenging. The fundamental objective of this study is to create a relatively inexpensive underwater vehicle that can use an image processing technique to reliably spot cracks on the exteriors of industrial pipes. The tasks involved in this project include the planning, development, and testing of an underwater vehicle that can approach the circular pipes, take pictures, and determine whether there are fractures. In this project, we will utilize the Canny edge detection technique to identify the crack. The system could function in either an online or offline mode. Using a Raspberry Pi and a camera, the paper will discuss the procedures followed to locate the pipe cracks that activate the underwater vehicle. While Python is used for image processing to capture photographs, analyze images, and expose flaws in particular images, the underwater vehicle's movement will be controlled via a connected remote control. When the physical model has been built and tested, the results are recorded, and the system's benefits and shortcomings are discussed.

Keywords—*Crack detection; pipeline; underwater vehicle: image processing; Raspberry Pi; canny edge detection*

I. Introduction

As certain countries start an economic growth period, a subsequent increase in energy demand is anticipated. The infrastructure that supports energy production expands to meet the growing needs of the world's energy consumers. Subsea natural gas production is gaining popularity as a means of accommodating this growth. As a result, there will be an even higher need for resources like human workers, capital, and expertise. These days, robots are being utilized extensively in a wide range of fields [1-6]. Even in the oil and gas industry, robotic assistance is becoming commonplace. By decreasing the need for human intervention while simultaneously boosting operational efficiency and safety, inspection robots are being employed to carry out inspection and maintenance activities on industrial property. Technology advancements in

the field of inspection robotics have made the oil and gas industry more productive, secure, and dependable.

Extraction of petroleum from subterranean sources and transportation to the surface is a complex process that requires a wide range of structures and systems to work together. When conducting such tasks, it is essential to do so while minimizing costs and environmental damage as much as feasible. Transmission of oil and gas across great distances is essential, and subsea pipelines play a key role in this process [7].

From an economic, safety, and environmental perspective, pipelines could be regarded as the most ideal means of transporting petroleum fluids from underwater structures to floating manufacturing plants during exploration and from these sites to land oil refineries. The diameter and length of these pipelines, as well as operating pressures, subsea terrains, undersea conditions, and fluid properties, all have a significant impact on the price of subsea pipelines. The price of building a subsea pipeline varies depending on these factors, from several hundred dollars per kilometer to many millions of dollars per kilometer [7].

Several obstacles make it hard to inspect and monitor underwater pipelines. The area in which a pipeline may be situated for decades is complex and full of potential hazards, and problems can occur at any time during the pipe's lifetime. Damage to the pipeline might be caused by debris that falls from the topside. The pipeline might get caught on ship anchors or dragged by fishing gear that drags the ocean floor. Erosion of the soil beneath the pipeline by ocean currents can lead to free spanning, in which the pipeline is no longer supported anywhere but at its beginning and terminus. Corrosion and erosion can be caused by the inner fluids, which are often acidic and contain abrasive sand particles travelling at high speeds. Cracks in the pipe wall can occur because of such circumstances. Significant damage can be avoided if the crack or fracture is monitored on a regular basis, and the pipe's life can be extended with good treatment. Even seemingly tiny cracks can expand and finally lead to serious structural failure [8, 9]. Therefore, inspection and monitoring

activities must be carried out to ensure that no cracks or defects develop, as this might lead to serious implications for the industry, such as explosions and fires caused by the discharge of toxic gases and liquids.

In consideration of all these issues, it's crucial to constantly check pipelines and install monitoring equipment. Once the system is in place, it might be difficult to inspect the subsea pipeline due to the often-considerable water depths. Many useful inspection technologies cannot be delivered to the pipeline because they rely on costly and potentially risky equipment and methods. It would be impractical to strip away coating layers for routine inspection if the pipeline was multilayered and buried a few thousand feet below the ocean's surface. Pipelines can be inspected internally using inline inspection, but the benefit of doing so must be evaluated against the cost of shutting the pipeline down. In most cases, it would be preferable and more cost-effective to use a less invasive screening inspection or monitoring tool that provides a more comprehensive overview of the pipeline and recommends which regions may require additional attention.

Several studies on remotely operated vehicles (ROVs) for maritime research have recently been published [10-14]. In offshore sectors, such as oil and gas, marine structures, marine sciences, naval security, marine renewable energy, and research reasons, ROVs have been employed for underwater intervention, repair, and maintenance activities. ROVs are used in submarine applications to track mines and are programmed to do high-risk activities by executing algorithms for prediction, diagnosis, and classification. Other research initiatives have concentrated on underwater surveillance, and they are set up to manage continuous tasks with defined goals [10]. ROVs can also be used to capture underwater photographs, for which there are now open research lines [10]. Several detection techniques utilizing image processing with deep learning have previously been deployed by mounting cameras on ROVs to detect fractures in infrastructure such as walls and pipelines, but results can still vary due to ambient factors such as illumination [15]. As a result, researchers are becoming more interested in crack detection investigations to improve detection approaches [16-20]. Despite the tremendous improvements made in the various fields of use and development of ROVs, the wide range of specialties that come into play in obtaining maximum performance makes it a rich topic for research. The accurate sensing of underwater applications is currently a hot issue in academia.

The primary aim of this research is to develop a low-cost underwater vehicle that can accurately detect cracks on industrial pipe surfaces by using a Raspberry Pi as an embedded controller, which can be programmed using open-source software, is inexpensive, has multiprocessing capabilities, and can be programmed in Python. Several parameters, such as the optimal distance and angle between the camera and the object, will be examined in the experimental section. At the end of the paper, two offline and real-time online monitoring systems will be presented and discussed.

II. LITERATURE REVIEW

Multiple mode nonlinear guided waves have already been investigated and published by several academics for the purpose of detecting fatigue cracks in pipes [21]. Guan R et al. provided in-depth analyses of guided wave propagation and interaction with microcrack in a pipe structure using numerical and experimental methods. The simulation model incorporated a third-order elastic constant and a seam crack to account for material nonlinearity and CAN. Piezoelectric transducers and a nonlinear signal collection system were employed in the acquisition of wave nonlinearity experiments. Both theoretical and empirical evidence demonstrated the nonlinearity produced by a 'breathing' fracture, and the identification of several second harmonic waves was distinct from the findings of a previous investigation in a plate structure. In order to quantify nonlinearity in pipe structures, a new nonlinear index has been introduced [21].

Concerns about the stability, longevity, and functionality of a structure are often prompted by the appearance of cracks. The effective loading area decreases as fractures develop and spread, leading to a rise in stress and, finally, structural failure. Nhat-Duc Hoang proposed a model [22] for employing image processing to spot fracture faults on the exterior of structures. When dealing with digital images for fracture analysis, common problems such as low contrast, uneven illumination, and noise pollution during image processing render the conventional Otsu approach useless. Min-Max Gray Level Discrimination is an image enhancement algorithm used in the new model to enhance the Otsu method (M2GLD). The newly developed model can recognize crack objects and assess their attributes like area, perimeter, width, length, and orientation. Experiments indicate that flaws in testing images were appropriately spotted. When used in conjunction with the Otsu technique, the M2GLD can significantly improve performance [22].

Image processing techniques such as Canny edge detection can find lines suggesting fractures within the pipe and split the images to create a new image that only shows the detected fractures, allowing the fractures to be studied. The Canny edge detection method will be used to detect the fracture in this project. The Raspberry Pi will be used to control the system, which will be attached to the underwater vehicle. When compared to the slow and subjective human inspection, the image processing method for detecting cracks and flaws can offer fast and trustworthy results [23-25]. Pipe inspection has traditionally been done visually, and there are a few drawbacks that may be highlighted, including the fact that it is difficult, time-consuming, subjective to the inspector, and only provides a qualitative conclusion [26, 27].

III. METHODOLOGY

The areas of technology and science, as well as the world of computer programming, can all benefit from the use of robots. Developers and researchers are working to make robots useful in as many fields as possible [28-30]. Modern technology has led to a huge increase in the use of image processing tools in the engineering community for a variety of purposes, including measurement, robot navigation, and others. The research community and industry alike can benefit

from this technology. Robotics processing steps are shown in Fig. 1, whereas image processing steps are shown in Fig. 2. When the camera is activated, the robot will proceed through the pipe. If a camera spots a crack, it will take a picture and send it to the raspberry pi so it can be analyzed. If a fault or crack is found, it will be identified by a comparison of the two images. If the system detects a fracture, it will take a picture and record the relevant data. If no damage is found, the operation will restart with the camera. To begin the process of crack detection in Fig. 2, a camera image of the structure was first obtained. Prior to data saving, the captured image will be uploaded and processed with the Canny edge technique.
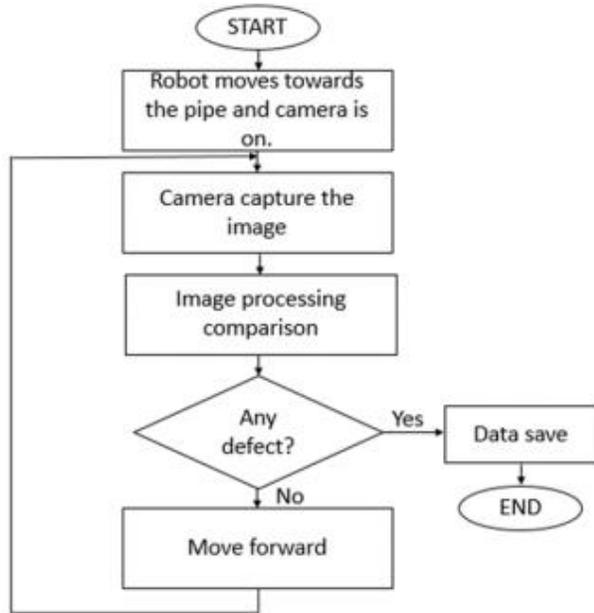
Fig. 1. Process Flow for the Robot.
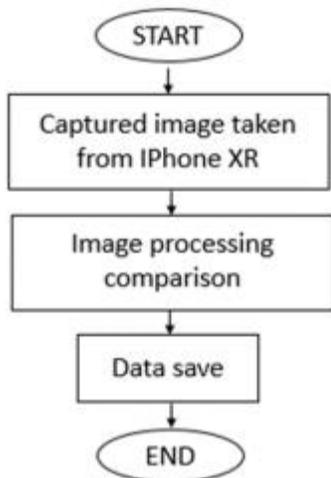
Fig. 2. Image Processing Process Flow.
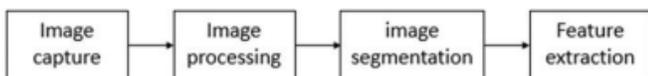
Fig. 3. Process Block Diagram.

Fig. 3 depicts the process block diagram used to identify pipe cracks. Capturing an image is the initial stage of the procedure. The online mode image is captured by the Raspberry Pi V2 NoIR camera, while the offline mode image is captured by the iPhone XR camera. The procedures will thereafter be used both online and offline. There won't be many actions taking place in the image processing phase. At first, the image will shift from its original RGB color mode to a more conventional grayscale. This brings us to our second topic: the logarithmic scale. During this procedure, the image quality improves as the number of pixels is increased by a factor of logarithm. Image smoothing comes last. The term "bilateral filter" or "blurring process" can be used to describe this technique as well. The edge of the break will be maintained during this procedure.

Moving on to the image segmentation stage, where Canny edge detection takes place. Canny edge detection typically has three goals, but the researcher here added a fourth one to boost the overall image quality. The first thing to look at is the image's intensity gradient. Each pixel's gradient vector is computed at this step. Second, the suppression that is less than maximal. To achieve a width of one pixel, the edge must be narrow. Thresholding comes next. First, the morphological operator will be reduced, and then the artificial edge. It serves to patch up the hiccups to make the main crack look more complete. To conclude, we will extract features. SIFT, SURF, and ORB are the three primary methods of feature extraction. The Oriented Fast and Rotated Brief (ORB) format was employed for this study. The object recognition relies on this.

The Raspberry Pi and ultrasonic sensor connection is shown in Fig. 4. The module can be powered by the Pi's 5V and Ground pins. In this project, the Pi's GPIO header's pins 2 and 6 were utilized. The transmission of the ultrasonic pulse is started by pressing the "trigger" input pin on the module. The general-purpose digital input/output (GPIO) uses a 3.3V signal; therefore Pin 16 is directly connected to the trigger (GPIO23). A bit more thought must be given to the module's "echo" output. The output pin is low before the module does its distance measurement (0V). It then maintains this pin high (+5V) for the same amount of time that it took for the pulse to return.

This pin's value must therefore be monitored by the script. The Pi's inputs prefer 3.3V, but the module uses a +5V level for a "high," which is too high. To make sure that the Pi is only exposed to 3.3V in this project, a straightforward voltage divider can be utilized. This is made with two resistors. If R1 and R2 are identical, the voltage is divided in half. 2.5V would be the voltage as a result. If R2 is twice as large as R1, 3.33V will result. The circuit used resistors with resistance values of 470 and 330 ohms.

- Pin 2 is directly connected to the connection for +5V.

- Pin 6 is included into the voltage divider for the Echo pin but connects directly to the GND connection.

- Pin 18 is connected to the voltage divider's center.

- Pin 16 is used to establish a direct connection to the Trig connector.
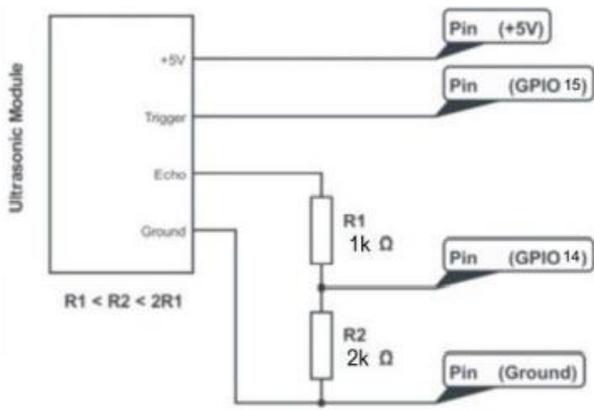
Fig. 4. Ultrasonic Sensor Connection.

For analyzing digital logic signals, the Raspberry Pi has GPIO pins for general-purpose digital input/output. The GPIO pins will receive the signal from the 'Echo' output pin. The output voltage of the Echo pin is 5V, which is unsafe for use with the Raspberry Pi's GPIO pins, which operate at 3.3V. A voltage divider is used to divide the signal between two connected resistors in a ratio equal to the resistance of the circuit, which is then fed to the Pi to generate the required voltage. As can be seen in Fig. 5, when a 5V signal is supplied from the JSN-SR04T, 1.7V is dropped across the 1k Ohm resistor, and 3.3V is dropped across the 2k Ohm resistor. Fixes and maintenance performed on the prototype of the underwater vehicle are depicted in Fig. 6.
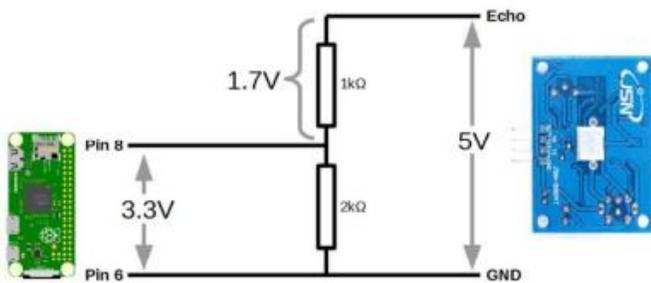
Fig. 5. Voltage Divider.

Fig. 6. Underwater Vehicle Prototype.

A. Canny Edge Detection

The basic method of edge detection has been improved using Canny edge detection. Edge tracking employing hysteresis, smoothing, gradient discovery, non-maximum suppression, and double thresholding are some of the stages that go into obtaining Canny edge detection. To remove camera noise, smoothing will inevitably damage the image. Gaussian filter application triggered the smoothing process. The fundamental idea behind the Gaussian filter was to use a standard deviation using the following equation.

$$B = \frac{1}{159} \begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix} \tag{1}$$

The standard deviation is a multiplier used to determine how many pixels in a raw image the Gaussian filter will change. This matrix filter is used to convert matrix pixels contained inside a raw image to a raw image. By looking at the region with the most change in intensity, the Canny algorithm seeks to identify the edges of the grayscale image. The gradient of each pixel was handled by the Sobel-operator. Equations following provide illustrations of the x and y gradients.

$$K_{GX} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{2}$$

$$K_{GY} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{3}$$

Finite differences $f(x + 1, y) f(x, y)$ and $f(x, y + 1) f(x, y)$ were used to simplify the finite differences used to calculate the row and column values inside the matrix, where x and y were the raw image's pixel coordinates. The gradient's magnitude is defined in (3) as the angle from Euclidean that Pythagoras' law specifies. The Manhattan distance, which is derived by comparing the distances between the starting and finishing points/blocks, simplifies this calculation (4).

$$|G| = \sqrt{G\frac{2}{x} + G\frac{2}{y}} \tag{4}$$

$$|G| = |G_x| + |G_y| \tag{5}$$

The most obvious limits are shown by the magnitudes of G_y and G_x. But, at times the margins were excessively large, making it difficult to make out exactly where the borders were. To get around this limitation, it is necessary to determine and remember the edges' orientation using the equation below.

$$\theta = \tan^{-1}\left(\frac{|G_y|}{|G_x|}\right) \tag{6}$$

The raw picture was first blurred by smoothing, and then sharpened by using non-maximum suppression. The only part of the gradient picture that was kept was the local maximum. As a second step after non-maximal suppression, we used the improved pixels to label the remaining edge pixels in a

process called double thresholding. Rough surfaces can generate noise and color shifts, which could result in a loss of pixels. We performed a double thresholding method to extract the strongest edge from the image. In this case, only the edge pixels with a value higher than the thresholding value were kept, while the rest were discarded. Weak pixels were those with a threshold value between 0 and 1. In the final stage of edge detection, the strong edge was treated as the real thing. Only edges that were found to have some sort of connection to the real edge were weak. The idea behind it is that noise and other color fluctuations render a clear advantage impossible to achieve. After this point, Canny edge detection will be finished.

IV. RESULT AND DISCUSSION

The system may function both online and offline. Consequently, the user can opt to either upload an existing image or take a new one and process it in real time. Location of the camera used in the experiment is depicted in Fig. 7. Online mode will produce three photos, one from each of the three specified perspectives and distances. The offline option also limits users to uploading and processing just a single image at a time. Fig. 8 through Fig. 16 depicts the original iPhone XR image and the final product after The Canny edge detection was applied. Fig. 17 to 19 demonstrate what happened when we applied the Canny edge detection to data, we got from a Raspberry Pi V2 NoIR camera.

A. Distance of the Camera to the Pipe (cm) and the Angle Taken to the Pipe

The camera's optimal placement and distance from the pipe were measured. In both online and offline settings, the camera's angle relative to the pipe has been set at 45 degrees, 90 degrees, and 135 degrees, as illustrated in Fig. 7. However, in both online and offline modes, the camera distance to the pipe is fixed at 18cm, 13cm, and 7cm, respectively.
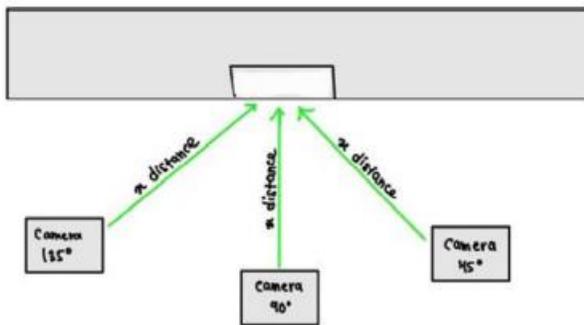
Fig. 7. Camera Angle for Testing.

TABLE I. DETECTION FOR OFFLINE MODE

Angle (°)	Distance of the camera to the pipe (cm)		
	7	13	18
45	Detect	Detect with noise	Detect with noise
90	Detect	Detect	Detect
135	Detect with noise	Detect	Detect with noise

Table I displays the results of the offline testing conducted with the various factors discussed in the introduction. Table I demonstrates that the outcome can be detected across a wide range of parameter settings. However, when testing at 7cm with a camera at an angle of 135 degrees, 13cm with an angle of 45 degrees, 18cm with an angle of 45 degrees, and 135 degrees with an angle of 45 degrees, noise from the raw image is observed. A crack can be detected without any background noise from any distance and any camera angle. Testing in offline mode at a camera distance of 18cm at an angle of 90 degrees revealed that this was the optimal configuration for detecting pipeline cracks.

TABLE II. DETECTION FOR ONLINE MODE

Angle (°)	Distance of the camera to the pipe (cm)		
	7	13	18
45	Detect with noise	Detect with noise	Detect with noise
90	Detect	Detect with noise	Detect with noise
135	Detect with noise	Detect with noise	Detect with noise

Table II displays the outcomes of the tests conducted in the online mode. Based on the testing, a camera positioned at a 90-degree angle and 7 cm away produces the best results. The alternative setting demonstrates that the crack is still detected, although with background noise. The camera on the Raspberry Pi V2 NoIR may have needed bright light to concentrate on the crack, which could explain why this happened.

B. Image Capturing with a Regular Camera for Monitoring in Offline Mode

The offline mode will enable computers input and output data quickly. As a result, relatively slow input devices are no longer required. Instead, the information is kept as files on a fast data storage system. The primary processing computer does not immediately take control of and read the data from its input devices. The data is prepared, kept on a high-speed storage device separate from the computer, and then made available as required. In this mode, the user can select the preferred camera and still apply the same algorithm. From Fig. 8 to Fig. 16, the result may be observed in detail as indicated in Table I.

Distance between the camera and the pipe (7 cm) (45°).

The outcome for a camera to pipe distance of 7 cm and a 45° angle is shown in Fig. 8. The results clearly demonstrate crack detection without any noise.
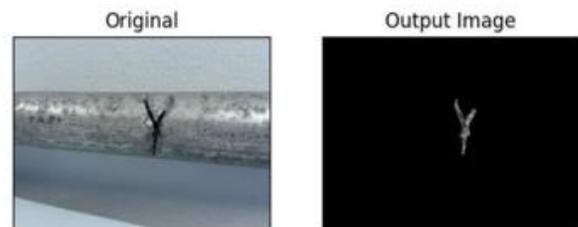
Fig. 8. Camera to Pipe Distance is 7cm and Angle is 45°.

1) Distance between the camera and the pipe (7 cm) (90°): The result is shown in Fig. 9 for a camera-to-pipe distance of 7 cm at a 90° angle. Crack detection is plainly demonstrated by the results.
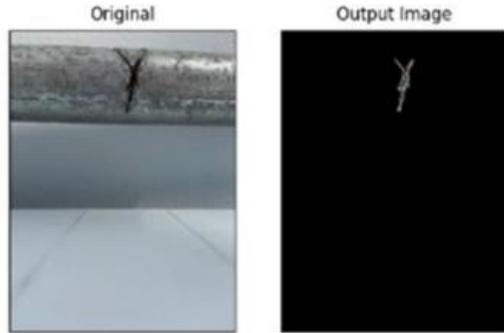
Fig. 9. Camera to Pipe Distance is 7cm and Angle is 90°.

2) Distance between the camera and the pipe (7 cm) (135°): Fig. 10 depicts the outcome with a camera-to-pipe distance of 7 cm and an angle of 135 degrees. The results indicate the detection of cracks with some background noise.

Fig. 10. Camera to Pipe Distance is 7cm and Angle is 135°.

3) Distance between the camera and the pipe (13 cm) (45°): The outcome for a 13 cm pipe to camera distance with a 45° angle is shown in Fig. 11. The results demonstrate the crack's detection with some noise.

Fig. 11. Camera to Pipe Distance is 13cm and Angle is 45°.

4) Distance between the camera and the pipe (13 cm) (90°): The outcome for a 13 cm camera to pipe distance at a 90° angle is shown in Fig. 12. The outcome demonstrates crack detection.
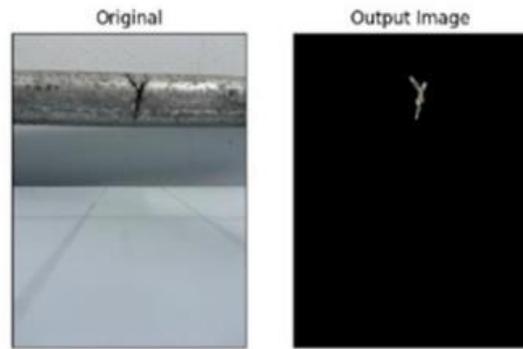
Fig. 12. Camera to Pipe Distance is 13cm and Angle is 90°.

5) Distance between the camera and the pipe (13 cm) (135°): The outcome with a camera to pipe distance of 13 cm and a 135° angle is shown in Fig. 13. The outcome demonstrates crack detection.
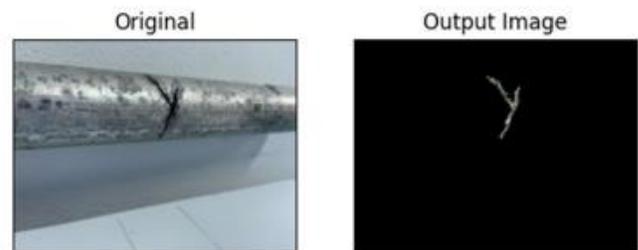
Fig. 13. Camera to Pipe Distance is 13cm and Angle is 135°.

6) Distance between the camera and the pipe (18 cm) (45°): The outcome for a camera to pipe distance of 18 cm and a 45° angle is shown in Fig. 14. The results demonstrate the crack's detection with some noise.
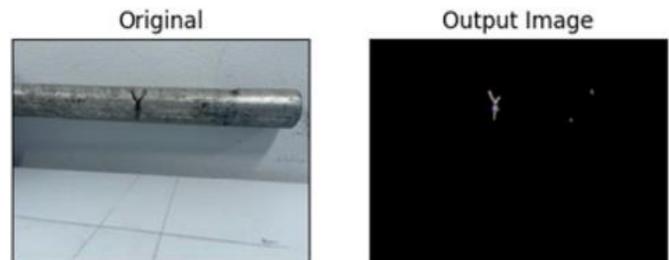
Fig. 14. Camera to Pipe Distance is 18cm and Angle is 45°.

7) Distance between the camera and the pipe (18 cm) (90°): Fig. 15 displays the outcome for an 18 cm camera to pipe distance at a 90° angle. The outcome demonstrates crack detection.
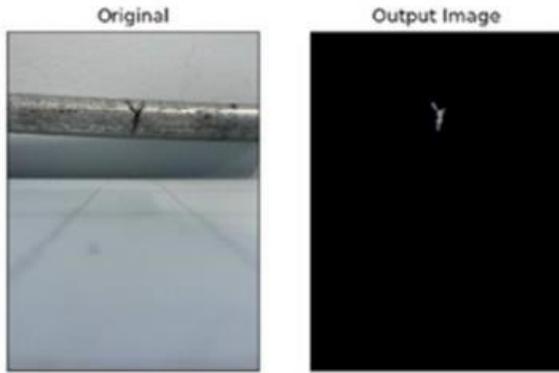
Fig. 15. Camera to Pipe Distance is 18cm and Angle is 90°.

8) Distance between the camera and the pipe (18 cm) (135°): Fig. 16 displays the outcome for an 18 cm camera to pipe distance at a 135° angle. The results demonstrate the crack's detection with some noise.
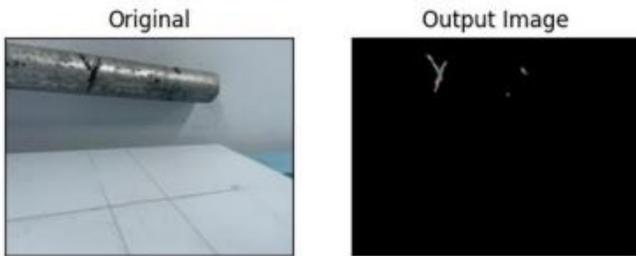
Fig. 16. Camera to Pipe Distance is 18cm and Angle is 135°.

C. Image Capturing with the Raspberry Pi V2 Module NoIR Camera for Monitoring in Real-Time Online Mode

When a system is connected to a computer and processing data files while using input, output, and storage devices, the process is referred to as real-time "online" processing. The system will take three photographs, giving each image taken a five-second pause period. With the same distance, the three photos have different angles.

The results of an online test for a 7 cm camera distance from the sample are shown in Fig. 17. When the camera is at a 90° angle, feature #2 displays the result without any noise, however testing at 45° and 135° results in some noise in the processed image.

1) Distance between the camera and the pipe (7 cm) (45°), (90°) and (135°): Fig. 18 displays the outcomes of an online test for a 13 cm camera distance from the sample. The crack is visible in the photograph after evaluating it from three different angles, but the results have significant noise. Regarding Fig. 19, the testing is done at an 18 cm camera-to-pipe distance. The crack is still detectable in the results, but there is significantly more noise than at a 13 cm distance. To compare to this chart, use the simplified data from Table II as a guide.
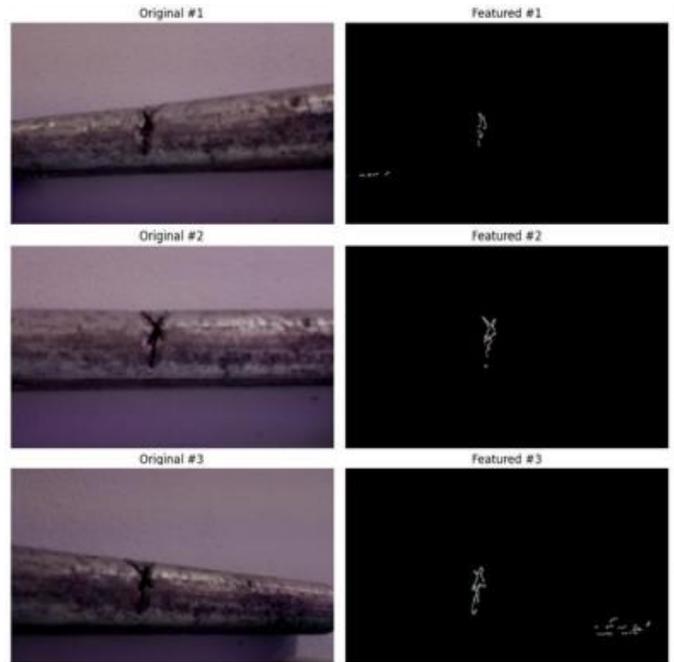
Fig. 17. Camera to Pipe Distance is 7cm and Angle #1 45°, #2 90° and #3 135°.

2) Distance between the camera and the pipe (13 cm) (45°), (90°) and (135°)
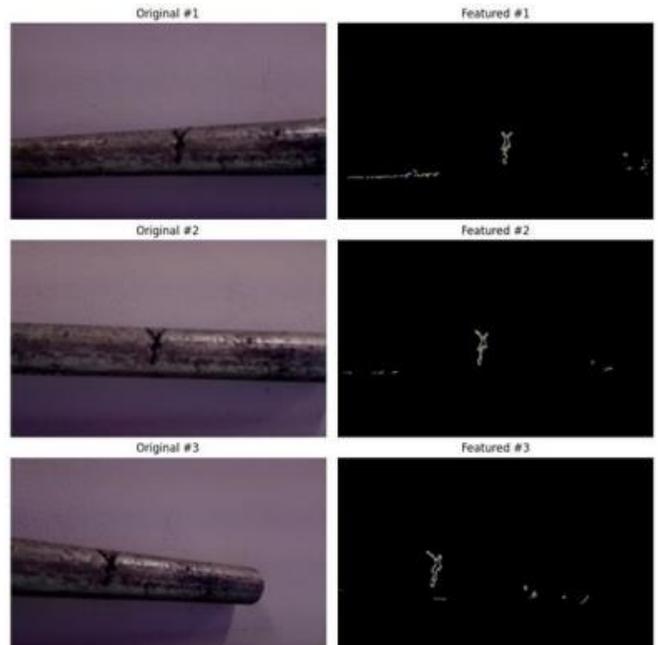
Fig. 18. Camera to Pipe Distance is 13cm and Angle #1 45°, #2 90° and #3 135°.

3) Distance between the camera and the pipe (18 cm) (45°), (90°) and (135°)
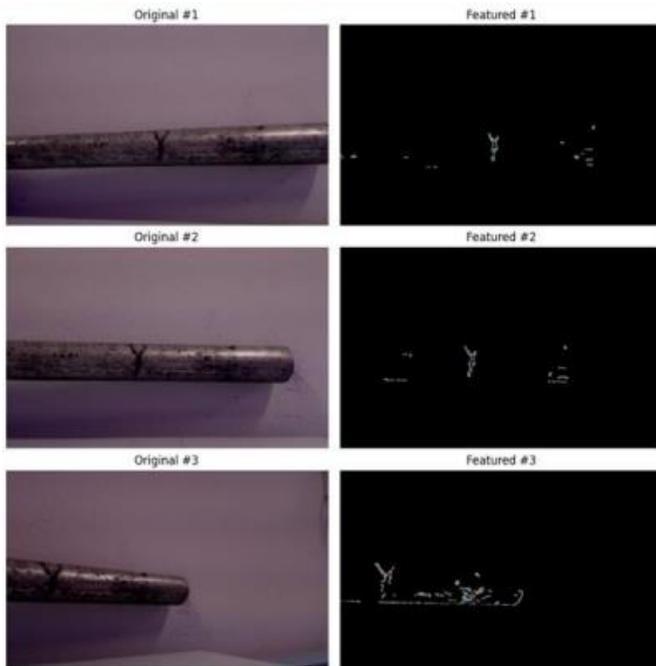
Fig. 19. Camera to Pipe Distance is 18cm and Angle #1 45°, #2 90° and #3 135°.

V. CONCLUSION

The project's purpose is to help the user detect pipeline faults under water by developing an underwater pipe crack detection system for a low-cost underwater vehicle using a Raspberry Pi and Canny edge detection. Online and offline modes are the two possible techniques. The cracks on the pipe are detected by both modes using the same method, the Canny edge detection.

The primary goal of this project is to use Python on a Raspberry Pi to implement the Canny edge detection crack technique. The algorithm Canny edge detection is employed in coding. Images that are captured online and those that are uploaded offline follow the approach depicted in Fig. 8 to 16. The picture segmentation step three includes the Canny edge detection. Canny edge detection has three fundamental steps, but for this project, one more step is included to make sure that the pipe fracture can be seen clearly. Finding the image's intensity gradient is the first step in the Canny edge detection procedure. Each pixel's gradient vector is computed throughout this process. The second process is called non-maximum suppression. The edge will become narrow thanks to this technique, giving it a one-pixel width. Thresholding is the final step. The artificial edge will be reduced in this stage, and the next step, the morphological operator, will fill in the little gaps to give the main crack more character.

The project's second goal is to create a low-cost underwater vehicle Remotely Operated Vehicle (ROV) prototype that can find cracks in industrial pipelines. Prototypes for vehicles are created with inexpensive but capable materials. The vehicle's mobility, which uses its propulsion system, is one of the precautions. The vehicle's mobility is controlled by the motor or turbine propulsion system. As a result, it is difficult for the user to control.

The project's final goal is to compare the two types of offline and online crack detection methods utilizing image processing. In the system, there are two operating modes: online mode and offline mode. Both modes use the same algorithm but a distinct concept, as is mentioned below. When a user uploads a picture into the system via the online mode, the system analyses the image as described in Fig. 17 through Fig. 19. The system will take three photographs, giving each image taken a five-second pause period. The three photos are at various distances or angles. The user can submit an image into the system for offline mode, and the system will immediately analyze the image. However, only one image may be posted at a time.

The camera utilized for the online approach to find pipe cracks must be improved due to the project's constraints. The camera cannot concentrate on the crack since the Raspberry Pi V2 NoIR camera required the proper lighting. The ultrasonic sensor is up next. It is impossible to determine the precise distance between the camera and the pipe with this ultrasonic sensor. This is due to the UT sensor's high quality. Consequently, it was necessary to purchase a high-quality UT sensor.

REFERENCES

[1] M. Aliff, S. Dohta, and T. Akagi, "Trajectory controls and its analysis for robot arm using flexible pneumatic cylinders," 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS), 2015, pp. 48–54.

[2] M. Aliff, M. A. Dinie, I. Yusof, and N. Samsiah, "Development of Smart Glove Rehabilitation Device (SGRD) for Parkinson's Disease," International Journal of Innovative Technology and Exploring Engineering, 9(2), 2019, pp. 4512–4518.

[3] K. Kusunose, T. Akagi, S. Dohta, W. Kobayashi, T. Shinohara, Y. Hane, K. Hayashi, and M. Aliff, "Development of Inchworm Type Pipe Inspection Robot using Extension Type Flexible Pneumatic Actuators," International Journal of Automotive and Mechanical Engineering, 17 (2), 2020, pp. 8019–8028.

[4] N. S. Sani, I. Shlash, M. Hassan, A. Hadi, and M. Aliff, "Enhancing Malaysia Rainfall Prediction Using Classification Techniques," Journal of Applied Environmental and Biological Sciences, 7 (2s), 2017, pp. 20–29.

[5] A. Syamim, M. Aliff, M. Ismail, S. Izwan, N. Samsiah, and M. Usairy Syafiq, "Application of Fuzzy Logic in Mobile Robots With Arduino and IoT," 2022 7th International Conference on Automation, Control and Robotics Engineering (CACRE), 2022, pp. 82–86.

[6] M. Aliff, M. Imran, S. Izwan, M. Ismail, N. Samsiah, T. Akagi, S. Dohta, W. Tian, S. Shimooka, and A. Athif, "Development of Pipe Inspection Robot using Soft Actuators, Microcontroller and LabVIEW," International Journal of Advanced Computer Science and Applications, 13(3), 2022, pp. 349–354.

[7] M. Ho, S. El-Borgi, D. Patil, and G. Song, "Inspection and monitoring systems subsea pipelines: A review paper," Structural Health Monitoring, 19(2), 2020, pp. 606–645.

[8] A. M. A. Talab, Z. Huang, F. Xi, and L. HaiMing, "Detection crack in image using Otsu method and multiple filtering in image processing techniques," Optik, 127(3), 2016, pp. 1030–1033.

[9] D. Dhital, and J. R. Lee, "A Fully Non-Contact Ultrasonic Propagation Imaging System for Closed Surface Crack Evaluation," Exp Mech, 52, 2012, pp. 1111–1122.

[10] O. A. Aguirre-Castro, E. Inzunza-González, E. E. García-Guerrero, E. Tlelo-Cuautle, O. R. López-Bonilla, J. E. Olguín-Tiznado, and J. Cárdenas-Valdez, "Design and Construction of an ROV for Underwater Exploration," Sensors, 19(24), 2019, pp. 5387.

[11] M. Aliff, N. Raihan, I. Yusof, and N. Samsiah, "Development of Remote Operated Vehicle (ROV) Control System using Twincat at Main Control Pod (MCP)," International Journal of Innovative Technology and Exploring Engineering, 8(12), 2019, pp. 5606–5610.

[12] M. Aliff, N. Firdaus, N. Rosli, M. I. Yusof, N. Samsiah, and S. Effendy, "Remotely Operated Unmanned Underwater Vehicle for Inspection," International Journal of Innovative Technology and Exploring Engineering, 9(2), 2019, pp. 4644–4649.

[13] M. S. A. M. Nor, M. Aliff, and N. Samsiah, "A Review of a Biomimicry Swimming Robot using Smart Actuator," International Journal of Advanced Computer Science and Applications, 12(11), 2021, pp. 395–405.

[14] M. Aliff, A. R. Mirza, M. Ismail, and N. Samsiah, "Development of a Low-Cost Bio-Inspired Swimming Robot (SRob) with IoT," International Journal of Advanced Computer Science and Applications, 12(7), 2021, pp. 452–457.

[15] Y. Hamishebahar, H. Guan, S. So, and J. Jo, "A Comprehensive Review of Deep Learning-Based Crack Detection Approaches," Applied Sciences, 12(3), 2022, pp. 1374.

[16] Z. Wang, M. Liu, Y. Qu, Q. Wei, Z. Zhou, Y. Tan, L. Hong, and H. Song, "The Detection of the Pipe Crack Utilizing the Operational Modal Strain Identified from Fiber Bragg Grating," Sensors, 19(11), 2019, pp. 2556.

[17] D. Rifai, A. N. Abdalla, R. Razali, K. Ali, and M. A. Faraj, "An Eddy Current Testing Platform System for Pipe Defect Inspection Based on an Optimized Eddy Current Technique Probe Design," Sensors, 17(3), 2017, pp. 579.

[18] X. Xu, M. Zhao, P. Shi, R. Ren, X. He, X. Wei, and H. Yang, "Crack Detection and Comparison Study Based on Faster R-CNN and Mask R-CNN," Sensors, 22(3), 2022, pp. 1215.

[19] Y. Zhao, Y. Han, C. Chen, and H. Seo, "Crack Detection in Frozen Soils Using Infrared Thermographic Camera," Sensors, 22(3), 2022, pp. 885.

[20] Z. Shi, X. Xu, J. Ma, D. Zhen, and H. Zhang, "Quantitative Detection of Cracks in Steel Using Eddy Current Pulsed Thermography," Sensors, 18(4), 2018, pp. 1070.

[21] R. Guan, Y. Lu, K. Wang, and Z. Su, "Fatigue crack detection in pipes with multiple mode nonlinear guided waves," Structural Health Monitoring, 18(1), 2019, pp. 180–192.

[22] N. D. Hoang, "Detection of Surface Crack in Building Structures Using Image Processing Technique with an Improved Otsu Method for Image Thresholding," Advances in Civil Engineering, 2018, pp. 1–10.

[23] N. D. Hoang and Q. L. Nguyen, "Metaheuristic Optimized Edge Detection for Recognition of Concrete Wall Cracks: A Comparative Study on the Performances of Roberts, Prewitt, Canny, and Sobel Algorithms," Advances in Civil Engineering, 2018, pp. 1–16.

[24] N. D. Hoang, Q. L. Nguyen, and D. T. Bui, "Image Processing-Based Classification of Asphalt Pavement Cracks Using Support Vector Machine Optimized by Artificial Bee Colony," Journal of Computing in Civil Engineering, 32 (5), 2018.

[25] R. S. Adhikari, O. Moselhi, and A. Bagchi, "Image-based retrieval of concrete crack properties for bridge inspection," Automation in Construction, 39, 2014, pp. 180–194.

[26] N. M. Syahrian, P. Risma, and T. Dewi, "Vision-Based Pipe Monitoring Robot for Crack Detection Using Canny Edge Detection Method as an Image Processing Technique," Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 2(4), 2017, pp. 243–250.

[27] X. Wu, Y. Jiang, K. Masaya, T. Taniguchi, and T. Yamato, "Study on the Correlation of Vibration Properties and Crack Index in the Health Assessment of Tunnel Lining," Shock and Vibration, 2017, pp. 1–9.

[28] T. Morimoto, M. Aliff, T. Akagi, and S. Dohta, "Development of Flexible Pneumatic Cylinder with Backdrivability and Its Application," International Journal of Materials Science and Engineering, 3(1), 2015, pp. 7–11.

[29] K. Hayashi, T. Akagi, S. Dohta, W. Kobayashi, T. Shinohara, K. Kusunose, and M. Aliff, "Improvement of Pipe Holding Mechanism and Inchworm Type Flexible Pipe Inspection Robot," International Journal of Mechanical Engineering and Robotics Research, 9(6), 2020, pp. 894 – 899.

[30] M. Aliff, S. Dohta, T. Akagi, T. Morimoto, "Control of Flexible Pneumatic Robot Arm Using Master Device with Pneumatic Brake Mechanism," JFPS International Journal of Fluid Power System, 8(1), 2014, pp. 38–43.

Method for 1/f Fluctuation Component Extraction from Images and Its Application to Improve Kurume Kasuri Quality Estimation

Jin Shimazoe[1], Kohei Arai[2], Mariko Oda[3], Jewon Oh[4]
Kurume Institute of Technology, AI Application Laboratory, Kurume city, Japan[1, 2, 3, 4]
Saga University, Saga City, Japan[2]

Abstract—**Method for 1/f fluctuation component extraction from images is proposed. As an application of the proposed method, Kurume Kasuri textile quality evaluation is also proposed. Frequency component analysis is used for 1/f fluctuation component extraction. Also, an attempt is conducted to discriminate the typical Kurume Kasuri textile quality, (1) Relatively smooth edge lines are included in the Kurume Kasuri textile patterns, (2) Relatively non-smooth edge lines are included in the patterns, (3) Between both of patterns (1) and (2) by using template matching method of FLANN of OpenCV. Through experiments, it is found that the proposed method does work for extraction of 1/f fluctuation component and also found that Kurume Kasuri textile quality can be done with the result of 1/f fluctuation component extraction.**

Keywords—*1/f fluctuation component extraction; Kurume Kasuri textile quality; FLANN; OpenCV*

I. INTRODUCTION

Many natural objects in the natural world are known to have fluctuations. Fluctuation is a partly random spatial and temporal fluctuation while following a certain average. 1/f fluctuations can be seen in temporal color changes such as autumn leaves, the rhythm of sea waves, and human heartbeats. Depending on the interpretation of fluctuations, the results are often highly hypothetical or oversimplified. It is considered to be of great significance.

Complex phenomena and shapes in the natural world include unpredictable subtle differences and unpredictable disturbances, and it has been reported that both these complex phenomena and irregular shapes follow the regularity of fluctuations. However, little research has been done on the relationship between color schemes and fluctuations. If there is a color scheme that shows 1/f fluctuation, it can be used in various fields such as interiors, exteriors, and urban planning.

It may say that beautifulness of drawings, textiles and the other images can be explained with 1/f fluctuations. 1/f fluctuation has the effect of suppressing the excitement of the sympathetic nerves and has the effect of making the listening side relaxed, comfortable, and sleepy. 1/f noise is a signal or process with a frequency spectrum such that the power spectral density (power per frequency interval) is inversely proportional to the frequency of the signal. In pink noise, each octave interval (halving or doubling in frequency) carries an equal amount of noise energy.

1/f noise is one of the most common signals in biological systems [1]. The name arises from the pink appearance of visible light with this power spectrum [2]. This is in contrast with white noise which has equal intensity per frequency interval. Within the scientific literature the term 1/f noise is sometimes used loosely to refer to any noise with a power spectral density of the form $S(f) \propto 1/f^{\alpha}$, where f is frequency, and $0 < \alpha < 2$, with exponent α usually close to 1. The canonical case with $\alpha = 1$ is called pink noise [3].

General $1/f^{\alpha}$-like noises occur widely in nature and are a source of considerable interest in many fields. The distinction between the noises with α near one and those with a broad range of α approximately corresponds to a much more basic distinction. The former (narrow sense) generally come from condensed-matter systems in quasi-equilibrium, as discussed below [4]. The latter (broader sense) generally correspond to a wide range of non-equilibrium driven dynamical systems.

Pink noise sources include flicker noise in electronic devices. In their study of fractional Brownian motion [5], Mandelbrot and Van Ness proposed the name fractional noise (sometimes since called fractal noise) to describe $1/f^{\alpha}$ noises for which the exponent α is not an even integer [6], or that are fractional derivatives of Brownian ($1/f^2$) noise.

As for methods of 1/f fluctuation extraction, frequency component analysis is usually used. Namely, (1) extract frequency components from signal sources, first and then, (2) α is calculated from the components. The proposed method allows extraction of frequency components from images in concern using Fast Fourier Transformation: FFT, or Discrete Cosine Transformation: DCT, and also, calculate α from the components. After that, image quality, in terms of textile of the images is evaluated.

As one of the applications of the proposed method, Kurume (Name of the city in Japan) Kasuri textile quality evaluation is conducted. Because humans feel comfortable with the 1/f fluctuation pattern. Therefore, Kurume Kasuri textile quality can be evaluated with the extracted 1/f fluctuation components. Kurume Kasuri is a traditional cotton fabric from the Chikugo region, Japan. By dyeing the cotton yarns first and then weaving them, slight shifts in the pattern occur. Such pattern shifts create a unique "texture" of the fabric, which is one of the main characteristics of Kurume Kasuri. It is also the appealing point of Kurume Kasuri.

There is a problem about the degree in the pattern shift. A moderate amount of pattern shifting brings out a good texture, but if the shifting is too large, the product will not sell, and the fabric has to be sold at a lower price. The problem is that the judgements of the degree of shifting by weavers and dealers are not always the same. In addition, it is necessary to adjust yarns at the stage of weaving Kurume Kasuri, but it is extremely difficult for inexperienced craftsmen (young people) to adjust yarns to achieve a moderate pattern shift. Furthermore, since skilled craftsmen adjust yarns based on their many years of experience and intuition, it is difficult to pass on their tacit knowledge, and there is a shortage of successors. Therefore, a system which can show the acceptable degree of shifting of patterns will help both weavers and dealers for their judgements, and will also improve quality. Furthermore, tacit knowledge can be passed on to younger weavers as explicit knowledge, thus supporting weavers who are suffering from a shortage of successors.

We will use image recognition technology to evaluate the texture (quality) of Kurume Kasuri and build a system to classify whether or not the texture is acceptable as a high-quality (whether or not the pattern shift (or misalignment) is within an acceptable range). By using Artificial Intelligence: AI technology, the team will challenge to make the advanced skills (tacit knowledge) of Kurume Kasuri craftsmen explicit on a computer screen.

In the next section, related research works are described followed by research background and theoretical background. Then, the proposed system is described at first followed by some experiments are described together with conclusion and some discussions.

II. RELATED RESEARCH WORKS

There is the following quality evaluation related papers,

Method for video data compression based on space and time domain seam carving maintaining original quality when it is replayed is proposed [7]. 3D skeleton model derived from Kinect depth sensor camera and its application to walking style quality evaluations is also proposed [8].

Quality flag of GOSAT/FTS products taking into account estimation reliability is proposed and evaluated with actual remote sensing satellite data [9]. Then report on vicarious calibration and image quality valuation of LISA/LISAT (Optical sensor onboard remote sensing satellite) is reported [10] together with methods for vicarious calibration and image quality evaluation of LISA/LISAT [11].

On the other hand, there are the following papers which deal with frequency component analysis,

Polarimetric SAR image classification with high frequency component derived from wavelet multi resolution analysis: MRA is proposed [12]. Meanwhile, hearing aid method by equalizing frequency response of phoneme extracted from human voice is proposed [13].

Incorporating frequency filtering into the tensor singular value decomposition (t-SVD) in robust tensor principal component analysis (RTPCA) is shown to improve the performance of RTPCA [14]. A sensitive frequency component analysis method for cavitation fault using Empirical Mode Decomposition (EMD) method, Fourier Transform and neural network is also proposed [15].

Meanwhile, there are the following papers which deal with noise analysis,

A method of speckle noise reduction for SAR data is proposed [16] together with a new method for SAR speckle noise reduction (Chi Square Filter: CSF) [17]. On the other hand, sensitivity analysis of Fourier Transformation Spectrometer: (FTS) against observation noise on retrievals of carbon dioxide and methane is conducted [18].

Noise suppressing edge enhancement based on Genetic Algorithm (GA) taking into account complexity of target image measured with Fractal dimension is proposed [19]. Meantime, a method for aerosol parameter estimation error analysis is proposed with a consideration of noises included in the measured solar direct and diffuse irradiance [20].

Method of noise reduction in passive remote sensing is proposed for noise and clutter rejection [21]. On the other hand, speckle noise removal of SAR images with Digital Elevation Model: DEM is proposed [22] together with a new method for SAR speckle noise reduction by Chi-Square test filter is proposed and validated with real SAR imagery data [23].

In addition, the higher-order detrending moving-average cross-correlation analysis (DMCA) is proposed to show that physical activity in daily life has a common long-range correlate with heart rate (HR) variability that shows 1/f fluctuation [24].

III. THEORETICAL BACKGROUND AND PROPOSED METHOD

We will evaluate the texture (quality) of Kurume Kasuri, build a system to classify whether it is an appropriate texture (whether the pattern shift is within the acceptable), and verify it.

A. Examples of Fluctuation Pattern of Kurume Kasuri

Fig. 1(a) shows one of the examples of the fluctuation patterns of Kurume Kasuri. As shown in Fig. 1, due to the irregular weaving timing of the loom, the Kasuri pattern does not become linear.

This non-smooth Kasuri pattern has a quaint texture and seems beautiful to the human eye. To put it the other way around, a smooth Kasuri pattern is boring and tasteless to the human eye. For instance, if the pattern edge of the current image in Fig. 1(a) is artificially smoothened, the pattern will be dull and boring as shown in Fig. 1(b).

DCT is applied to these patterns of Fig. 1(a) and (b), then Fig. 2(a) and (b) are obtained. High frequency components of the original pattern are smaller than that of the artificially smoothened edge pattern.

(a) Original Kurume Kasuri Pattern.

(b) Artificially Smoothened.

Fig. 1.   One of the Examples of Kurume Kasuri Pattern and its Artificially Smoothened Pattern.
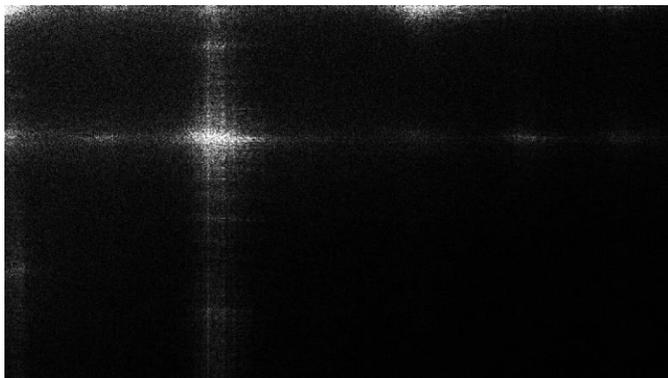
(a) Amplitude Spectrum of the Original Pattern.
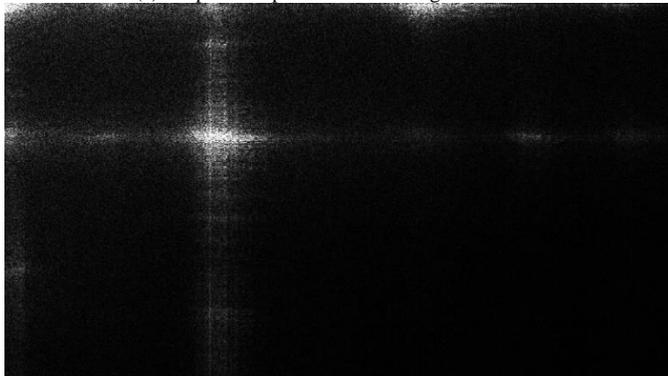
(b) Amplitude Spectrum of the Artificially Smoothened Pattern.

Fig. 2.   Amplitude Spectra of the Original and the Artificially Smoothened Patterns of Kurume Kasuri.

B. Procedure of the Proposed Method

The procedure of the proposed method is as follows:

1) Extraction and accumulation of image data for learning: Only the pattern part is cut out from the image of Kurume Kasuri (Fig. 3) by image processing by OpenCV.

2) Comparison of the number of feature points: Detect the feature points of the image obtained in 1), and compare the number of feature points with the ideal pattern and other features.

3) Comparison with 1/f fluctuation: Frequency analysis is performed on the image obtained in 1), the power spectrum is obtained, and the power spectrum is linearly approximated to obtain the proportional coefficient and the coefficient of determination.

4) Creation of image recognition model: Based on the collected image data for learning and the results obtained in 2) and 3), deep learning is performed based on the evaluation by experts, and an image recognition model is created.
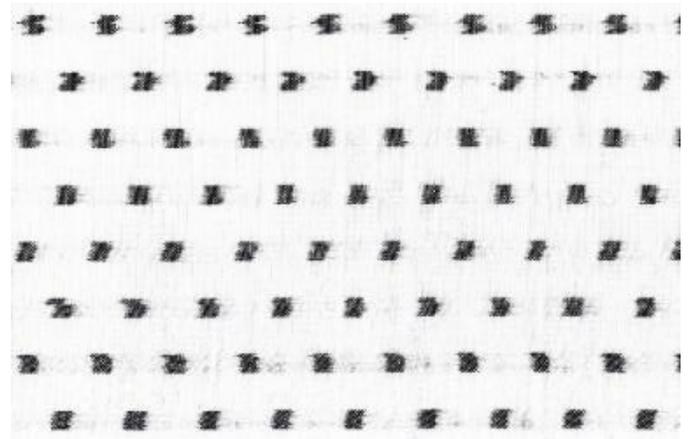
Fig. 3.   Image Example of Kurume Kasuri (Sideways).

As shown in Fig. 4, there are three types of Kurume Kasuri edge patters, (1) green colored rectangles of relatively smooth edge line patterns, (2) red colored rectangles of relatively non-smooth edge line patters and (3) yellow colored rectangle of the middle patterns between (1) and (2). The pattern type of (1) is boring and tasteless while the pattern type of (3) is quaint texture and seems beautiful. On the other hand, the pattern (2) is that the smooth line fluctuates too much, giving the impression of an unstable pattern.
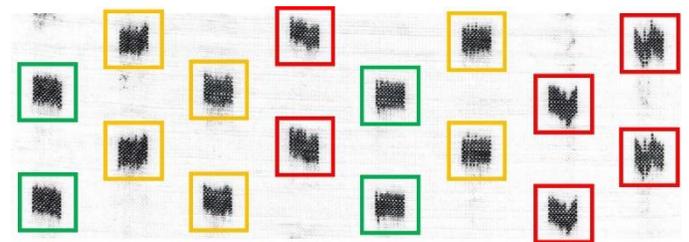
Fig. 4.   Three Types of Kurume Kasuri Edge Patters, (1) Green: Relatively Smooth Edge Line Patterns, (2) Red: Relatively Non-smooth Edge Line Patters and (3) Yellow: The Middle Patterns between (1) and (2).

In order to extract the rectangle of the outline of the Kurume Kasuri in concern, the following four steps are conducted.

1) Find the outline of the pattern

- Binarization (in black and white)

- Inversion of color (because the background is white)

- Contour extraction

2) Find the circumscribed rectangle based on the contour information

3) Find the center coordinates of the circumscribing rectangle

4) Specify the extraction range

Fig. 5 shows the procedure of the extraction of the rectangle of the outline of the Kurume Kasuri.

(a) Original.   (b) Binarize.   (c) Inverse.   (d) Contour.

(e) Rectangle and Center   (f) Extraction Range.
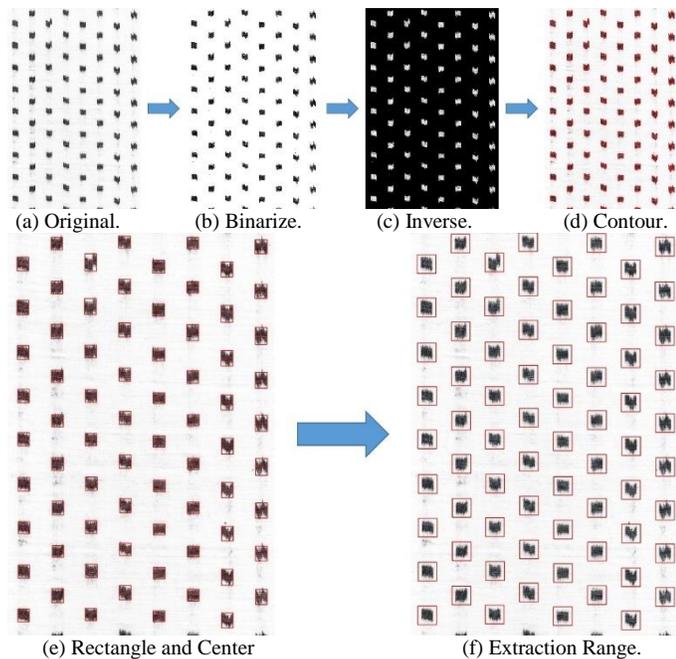
Fig. 5.   Procedure for Extracting of the Rectangle of the Outline of the Kurume Kasuri.

IV. EXPERIMENTS

A. *Pattern Comparison with Template Matching*

In order to compare the patterns in concern, the ideal pattern selected by the expert (Kurume Kasuri Producer: Orimoto in Japanese) from the obtained images was used as the template image, and the feature points of the template image and other images were compared using Fast Library for Approximate Nearest Neighbors: FLANN (high-speed nearest neighbor search) of OpenCV (Fig. 6).

Namely, the results show that the larger the fluctuation, the larger the number of feature points and the smaller the number of matching feature points. Therefore, it is possible to discriminate among three edge line patterns (1) relatively smooth edge line patterns, (2) relatively non-smooth edge line patters and (3) the middle patterns between (1) and (2).

The number of matching points
Template : 54
Input Image : 54

(a) Template & Template.

The number of matching points
Template : 54
Input Image : 71

(b) Template & Image1.

The number of matching points
Template : 54
Input Image : 98

(c) Template & Image2..

The number of matching points
Template : 54
Input Image : 130
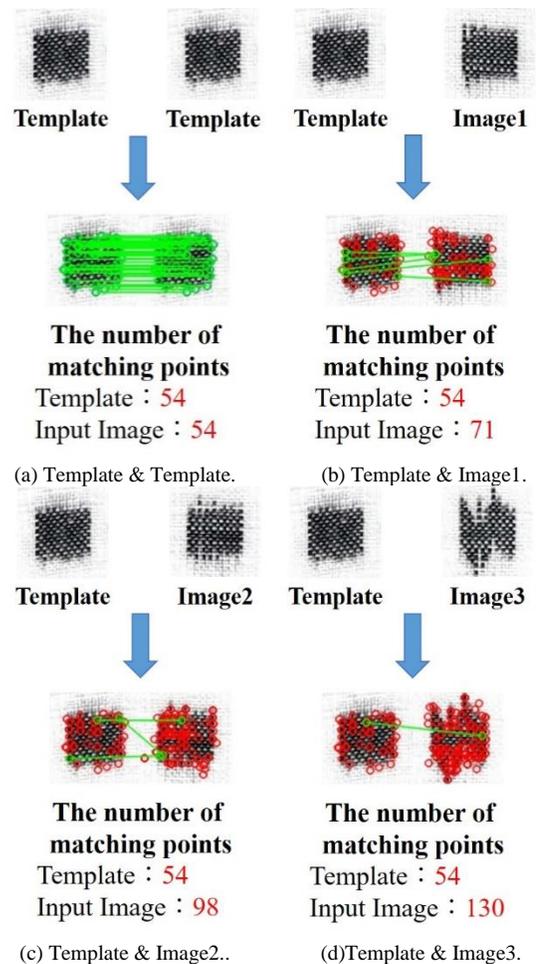
(d)Template & Image3.

Fig. 6.   The Number of Matching Points between Template and the other Images, Image1: Relatively Smooth Edge Line Patterns, Image3: Relatively Non-smooth Edge Line Patters and Image2: the Middle Patterns between Both.

B. *Frequency Component Analysis*

Fig. 7 shows aforementioned typical pattern images of three types of Kurume Kasuri patterns and their frequency components.

Spatial domain   **Frequency domain**

(a) Relatively Smooth Edge Lines.

Spatial domain   **Frequency domain**

(b) Between Both Smooth and Non-smooth Edge Lines.

Spatial domain   **Frequency domain**

(c) Relatively Non-smooth Edge Lines.
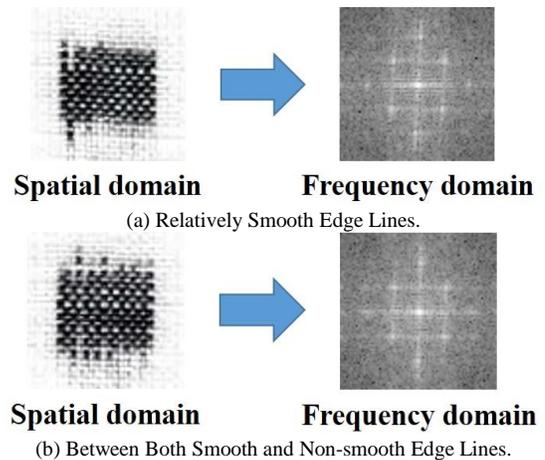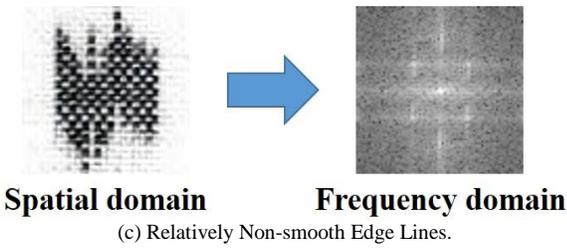
Fig. 7.   Frequency Components of the Typical Three Kurume Kasuri Patterns.

TABLE I.   PROPORTIONAL COEFFICIENTS AND R^2 VALUES OF THE TYPICAL THREE PATTERNS

Types of patterns	Proportional coefficient	R^2	Proportional coefficient	R^2
	Horizontal		Vertical	
(1) Smooth edge lines	-0.584	0.125	-1.850	0.589
(3) Between both	-1.168	0.399	-2.001	0.688
(2) Non-smooth edge lines	-1.885	0.637	-2.365	0.782

Also, Fig. 8 shows the vertical and horizontal frequency components of the pattern images of Fig. 7. The negative slope (proportional coefficient) did not become one, which is the same as the 1 / f fluctuation, but the larger the fluctuation, the better the accuracy of linear approximation (coefficient of determination), and the larger the negative slope (proportional coefficient). Obtained (Table I).

With a negative slope (proportional coefficient) of two as the threshold value, it may be possible to classify anything that exceeds it as excessive fluctuation. It is found that the fluctuation in vertical direction is greater than that of horizontal direction. It is also found that proportional coefficient of two is much comfortable than proportional coefficient of one (the 1/f fluctuation).

(a) Relatively Smooth Edge Lines.

(b) Between Both Smooth and Non-smooth Edge Lines.

(c) Relatively Non-smooth Edge Lines.

Fig. 8.   Amplitude Spectra of the Vertical and Horizontal Directions for the Typical Three Kurume Kasuri Patterns.
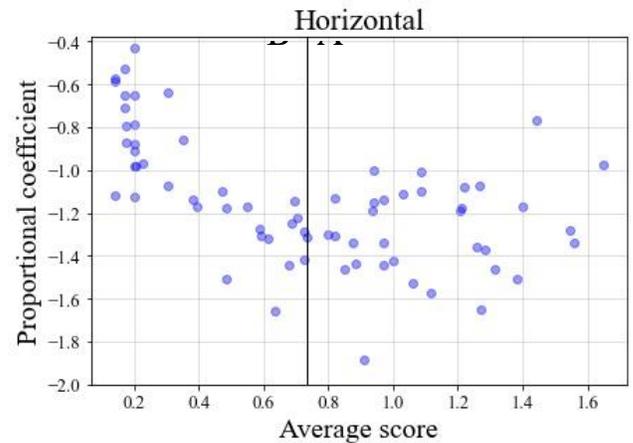
(a) Relationship between Average Scores and Horizontal Proportional Coefficients.
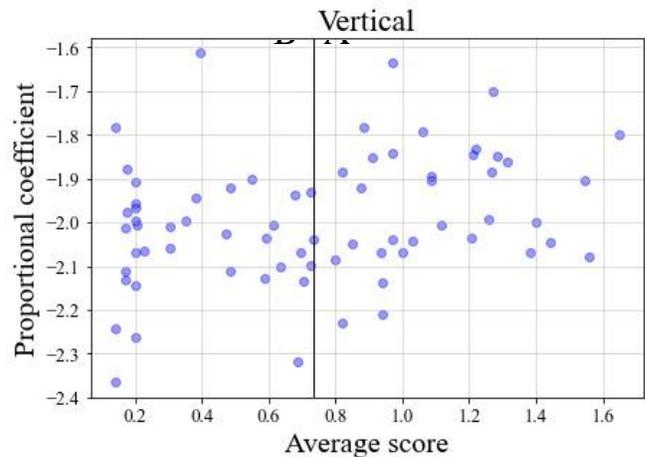
(b) Relationship between Average Scores and Vertical Proportional Coefficients.

Fig. 9.   Scatter Plots Showing the Relationship between Average Scores of Evaluations by Orimoto and Proportional Coefficients for each of 70 Patterns.

Finally, we compared proportional coefficients with results of a questionnaire conducted in advance to 35 experts (Kurume Kasuri Producer: Orimoto) in order to check if it is actually possible to classify patterns. Regarding the

questionnaire, each of 70 patterns used in this experiment was evaluated as good (2 pt), acceptable (1 pt), and unacceptable (0 pt). As a result, the average score for the overall Orimoto evaluation was 0.733. Fig. 9 shows the relationship between average scores and proportional coefficients for each pattern, which is classified into Group A and Group B using the average score (0.733) as the threshold.

As for the horizontal, most of the proportional coefficients for group A are less than -1, indicating a negative correlation. As for the vertical, most of the proportional coefficients for group A are more than -2.1, indicating a positive correlation. It is found that pattern shift can be classified by proportional coefficient because t-test was performed at significance level of 5% and found to be significantly different.

V. CONCLUSION

Method for 1/f fluctuation component extraction from images is proposed. As an application of the proposed method, Kurume Kasuri textile quality evaluation is conducted. Frequency component analysis is used for 1/f fluctuation component extraction. Also, an attempt is conducted to discriminate the typical Kurume Kasuri textile quality,

1) Relatively smooth edge lines are included in the Kurume Kasuri textile patterns.

2) Relatively non-smooth edge lines are included in the patterns.

3) Between both of patterns (1) and (2) by using template matching method of FLANN of OpenCV.

Through experiments, it is found that the proposed method does work for extraction of 1/f fluctuation component and is also found that Kurume Kasuri textile quality can be done with the result of 1/f fluctuation component extraction.

This research not only helps improve the quality of Kurume Kasuri, a local traditional craft, but also allows young people to pass on the tacit knowledge of skilled craftsmen as explicit knowledge, supporting Orimoto who is suffering from a shortage of successors.

VI. FUTURE RESEARCH WORKS

In the future, to improve the quality of Kurume Kasuri, we would like to build a model that machine-learn patterns, classifies images to determine if pattern shift is within acceptable limits, and prevents pattern shift outside of acceptable limits.

REFERENCES

[1] Szendro, P (2001). "Pink-Noise Behaviour of Biosystems". European Biophysics Journal. 30 (3): 227–231. doi:10.1007/s002490100143. PMID 11508842. S2CID 24505215, 2001.

[2] Downey, Allen (2012). Think Complexity. O'Reilly Media. p. 79. ISBN 978-1-4493-1463-7. "Visible light with this power spectrum looks pink, hence the name.", 2012.

[3] Baxandall, P. J. (November 1968). "Noise in Transistor Circuits: 1 - Mainly on fundamental noise concepts" (PDF). Wireless World. pp. 388–392. Retrieved 2019-08-08, 1968.

[4] Kogan, Shulim (1996). Electronic Noise and Fluctuations in Solids. [Cambridge University Press]. ISBN 978-0-521-46034-7, 1996.

[5] Mandelbrot, B. B.; Van Ness, J. W. (1968). "Fractional Brownian motions, fractional noises and applications". SIAM Review. 10 (4): 422–437. Bibcode:1968SIAMR..10..422M. doi:10.1137/1010093, 1968.

[6] Mandelbrot, Benoit B.; Wallis, James R. (1969). "Computer Experiments with Fractional Gaussian Noises: Part 3, Mathematical Appendix". Water Resources Research. 5 (1): 260–267, 1969.

[7] Kohei Arai, Method for video data compression based on space and time domain seam carving maintaining original quality when it is replayed, International Journal of Research and Reviews on Computer Science, 2, 4, 1063-1068, 2011.

[8] Kohei Arai, Rosa Andrie Asmara, 3D skeleton model derived from Kinect depth sensor camera and its application to walking style quality evaluations, International Journal of Advanced Research in Artificial Intelligence, 2, 7, 24-28, 2013.

[9] Kohei Arai, Takashi Higuchi, Hiroshi Okumura, Hirofumi Ohyama, Shuji Kwakami, Kei Shiomi, Quality Flag of GOSAT/FTS Products Taking into Account Estimation Reliability, International Journal of Advanced Computer Science and Applications IJACSA, 9, 9, 67-74, 2018.

[10] Kohei Arai, Report on Vicarious Calibration and Image Quality Evaluation of LISA/LISAT, LAPAN Indonesia, May, 2018.

[11] Kohei Arai, Methods for Vicarious Calibration and Image Quality Evaluation of LISA/LISAT, LAPAN Indonesia, April, 2018.

[12] Kohei Arai, Polarimetric SAR image classification with high frequency component derived from wavelet multi resolution analysis: MRA, International Journal of Advanced Computer Science and Applications, 2, 9, 37-42, 2011.

[13] Kohei Arai, Takuto Konishi, Hearing aid method by equalizing frequency response of phoneme extracted from human voice, International Journal of Advanced Computer Science and Applications IJACSA, 8, 7, 88-93, 2017.

[14] Wang, S., Liu, Y., Feng, L., & Zhu, C. Frequency-Weighted Robust Tensor Principal Component Analysis. arXiv preprint arXiv:2004.10068, 2020.

[15] Yu, J., Fu, D., Zhou, P., Li, J., Ye, F., & Shen, Y. "A neural network based method for sensitive frequency component analysis of cavitation fault." IOP Conference Series: Earth and Environmental Science. Vol. 552. No. 1. IOP Publishing, 2020.

[16] H. WakabKasurishi and Kohei Arai, A method of Speckle Noise Reduction for SAR Data, International Journal of Remote Sensing, Vol.17, No.10, pp.1837-1849, May 1995.

[17] H. WakabKasurishi and Kohei Arai, A New Method for SAR Speckle Noise Reduction(Chi Square Filter), Canadian Journal of Remote Sensing, Vol.22, No.2, pp.190-197, Jun.1995.

[18] Kohei Arai, T.Fukamachi, H.Okumura, S.Kawakami, H.Ohyama, Sensitivity analysis of Fourier Transformation Spectrometer: FTS against observation noise on retrievals of carbon dioxide and methane, International Journal of Advanced Computer Science and Applications, 3, 11, 58-64, 2012.

[19] Kohei Arai, Noise suppressing edge enhancement based on Genetic Algorithm taking into account complexity of target image measured with Fractal dimension, International Journal of Advanced Research in Artificial Intelligence, 2, 10, 7-13, 2013.

[20] Kohei Arai, Method for Aerosol Parameter Estimation Error Analysis-Consideration of Noises Included in the Measured Solar Direct and Diffuse Irradiance, International Journal of Advanced Research on Artificial Intelligence, 5, 11, 1-9, 2016.

[21] K.Tsuchiya, K.Maeda, Kohei Arai, H.Nakamura and C.Ishida, Method of noise reduction in passive remote sensing, Proc.of the International Symposium on Noise and Clutter Rejection, 1-8, 1984.

[22] H.WakabKasurishi and Kohei Arai, Speckle noise removal of SAR images with Digital Elevation Model: DEM, Proc. of the 5th ISCOPS Symposium, 1993.

[23] H.WakabKasurishi and Kohei Arai, A New Method for SAR Speckle Noise Reduction(Chi-Square Test Filter), Canadian journal of Remote Sensing, Vol.22, No.2, pp.190-197, June 1996.

[24] A. Nakata, M. Kaneko, N. Evans, T. Shigematsu, K. Kiyono, "Long-range cross-correlation between heart rate and physical activity in daily life", 11th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO), Pisa, Italy, pp1-2, 2020.

AUTHORS' PROFILE

Jin Shimazoe, He received BE degree in 2022. He also received the IEICE Kyushu Section Excellence Award. He is currently working on research that uses image processing and image recognition in Master's Program at Kurume Institute of Technology.

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January 1979 to March 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post-Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science in April 1990. He is now an Emeritus Professor of Saga University since 2014. He was a council member for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is a Science Council of Japan Special Member since 2012. He is an Adjunct Professor of University of Arizona, USA since 1998 and is an Adjunct Professor of Nishi-Kyushu University as well as Kurume Institute of Technology/AI Application Laboratory since 2021. He also is Vice Chairman of the Science Commission "A" of ICSU/COSPAR since 2008 then he is now award committee member of ICSU/COSPAR. He wrote 77 books and published 678 journal papers as well as 550conference papers. He received 66 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA and IJISA.http://teagis.ip.is.saga-u.ac.jp/index.html

Mariko Oda, She graduated from the Faculty of Engineering, Saga University in 1992, and completed her master's and doctoral studies at the Graduate School of Engineering, Saga University in 1994 and 2012, respectively. She received Ph.D(Engineering) from Saga University in 2012. She also received the IPSJ Kyushu Section Newcomer Incentive Award. In 1994, she became an assistant professor at the department of engineering in Kurume Institute of Technology; in 2001, a lecturer; from 2012 to 2014, an associate professor at the same institute; from 2014, an associate professor at Hagoromo university of International studies; from 2017 to 2020, a professor at the Department of Media studies, Hagoromo university of International studies. In 2020, she was appointed Deputy Director and Professor of the Applied of AI Research Institute at Kurume Institute of Technology. She has been in this position up to the present. She is currently working on applied AI research in the fields of education.

Jewon Oh, He received BE, ME and PhD degrees in 2012, 2015 and 2021, respectively. Appointed Assistant Professor at AI Application Laboratory, Kurume Institute of Technology in 2021. He is a lecturer at AI Application Laboratory, Kurume Institute of Technology in 2022. His research is focused on developing energy-saving technologies in the building using AI and image processing.

Image Verification and Emotion Detection using Effective Modelling Techniques

Sumana Maradithaya, Vaishnavi S

Department of Information Science and Engineering

M S Ramaiah Institute of Technology, Bangalore, India

Abstract—The feelings expressed on the face reflect the manner of thinking and provide useful insights of happenings inside the brain. Face Detection enables us to identify a face. Recognizing the facial expressions for different emotions is to familiarize the machine with human like capacity to perceive and identify human feelings, which involves classifying the given input images of the face into one of the seven classes which is achieved by building a multi class classifier. The proposed methodology is based on convolutional neural organizations and works on 48x48 pixel-based grayscale images. The proposed model is tested on various images and gives the best accuracy when compared with existing functionalities. It detects faces in images, recognizes them and identifies emotions and shows improved performance because of data augmentation. The model is experimented with varying depths and pooling layers. The best results are obtained sequential model of six layers of Convolutional Neural Network and softmax activation function applied to last layer. The approach works for real time data taken from videos or photos.

Keywords—*Face detection; face recognition; emotion detection; data augmentation*

I. INTRODUCTION

Facial emotion recognition is a part of nonverbal communication. Recognizing emotions or detecting face from input image is a challenging task. Facial emotion detection is very useful in the research field of robotics, in detecting mental disorders and in marketing division as well. Human face detection has numerous applications in computer vision domain. Facial expression detection and analysis helps in sentiment analysis of humans. The proposed methodology presents a model, which can detect, recognize and understand the emotion of any given input image by creating different pipelines. Output of face detection pipeline is the bounding box faces. Then the output will be fed to face recognition pipeline. Face recognition verifies the given two images and outputs the Boolean value. Using convolutional neural network, model is built to detect seven classes of emotions that is sad, fear, disgust, neutral, angry, surprise and happy. Model is trained using open-source dataset that is FER (Face Emotion Recognition), downloaded from Kaggle. Using image augmentation model performance is improved.

The capability to recognize and differentiate among different faces is a boon for humans. Now with the evolution of technology in machine learning and computer vision, even computers can recognize different faces and are able to distinguish between them. Using face detection and recognition access to security applications can be improved.

Example iPhone does the payments and the processes doesn't require any actual card. The model can also be used in crime detection and other healthcare related applications. Existing functionalities or models focus on analysis of complete facial features, and hence with multiple features it will be very confusing to train the model using convolutional neural network. The proposed model mainly focuses on detecting seven classes of emotions that is happy, sad, anger, disgust, fear, surprise and neutral. The proposed methodology uses face recognition library for detecting the facial locations.

The study [28] indicates the importance of identifying emotions in people, so that any physiological expressions can be detected and handled at an early stage. Facial Detection is nothing but the ability to identify the facial location of any given input image or in any video frame. Bounding box will be the output of the identified face. Face Recognition is nothing but comparing multiple faces to verify whether the face belong to same person. Face recognition is done using by differentiating among the embedding vectors of the face. Detected faces emotion is recognized using the model built based on convolutional neural network.

Face detection can be regarded as object class detection where facial locations can be identified and specify the size of the objects being detected in a specific image or images. With the advent of technology, it is now possible to identify the faces, in image or frame without any constraints like head posture, lightning, illumination, skin color. Face recognition determines whether the given images have face that is positive or negative. To achieve this motive, algorithm applied for the purpose will be trained on large datasets. Once the algorithm is trained, algorithm will be capable of analyzing the images to detect the object that is face in any images. When it comes to face detection, it is very essential for any algorithm used to know the parts of the given image to generate the face prints that are compared with the face prints stored previously to check the match.

The research [29] discusses the need of facial detection and emotion detection during the COVID -19 pandemic times. It further elaborates on the effectiveness of face recognition, face detection and emotion detection in the nonverbal communication.

Section II elaborates on the earlier works done, available methodologies and model building approaches. Section III focuses on the proposed work. Section IV discusses the obtained results. Section V` ` summarizes the proposed work done.

II. LITERATURE SURVEY

The author in [7] has presented a model, which automatically recognizes face and facial emotions and classifies the emotions into eight different classes. This model uses Support vector machine algorithm for classification and the accuracy is almost 94%. This machine learning based model is used to detect real time images along with static images, and it cannot detect multiple faces in a given frame. Initially for face detection, HAAR cascade algorithm is used. Then using the face landmarks, trained dataset along with support vector machine algorithm face detection can be performed.

In [20], the author proposed a smart vision model, which can detect human face and emotions respectively. This model uses deep convolutional network for training the model. Even after applying some of the filters, accuracy achieved in the range of 48% to 80%. The model is used for classifying seven different emotions along with face recognition. Model uses swish activation function which is more efficient in terms of performance and maximum accuracy can be reached. Filters are nothing but parameters and layers in a fully connected neural network. Basically, shallow network (single layer) and modern deep network (8 layers) results are compared and dropout mechanisms are used to avoid overfitting.

The author in [10] proposes a deep learning model for emotion detection. Two different datasets are considered and performance evaluation of the proposed methodology is carried out. Validation accuracy and loss, computationally complex and time per step, learning rate, and detection rate are the factors considered for performance evaluation. Model is trained using convolutional neural network. The accuracy achieved is 79%. In [23], Viola Jones face detection algorithm is used for face detection in the image. The proposed methodology is used for human voice detection based on Mel frequency components. KNN classifier algorithm is used for facial emotion detection. Accuracy achieved is more than 90%. Features or a component of the face is calculated and these features are stored as database. Then face detection and emotion recognition can be done using these features stored in the database.

The authors in [5] proposed face detection model based on auto assistance neural network. Earlier approaches had both patch-based auto assistance neural networks integrated with global local-based auto assistance neural network. Accuracy achieved is 90% for detecting six classes of emotion. Model can be used to detect four classes of emotions in videos. The model can be used for detecting dynamic emotion detection from human videos.

The author in [25], proposes a model which makes use of two algorithms, one is based on colour intensity estimation and the other is based on Viola-Jones algorithm. Model is used to detect emotions based on spotting the micro expressions on the face. Colour intensity-based algorithm is very useful for smaller images and faster as well on coloured images. Viola-Jones algorithm can be used for greyscale images and high-resolution images. Both algorithms considered provide efficient outcomes in recognizing the face

and eyes areas in the image. Model cannot predict entire face with higher accuracy.

The article [12], presents a model, which can be used for facial classification of emotion recognition. When considered face, there will be numerous factors that need to considered, like skin colour, textures, eye position and mouth and so on. The first step in this model is to recognize the skin using elliptical boundary. For facial features detection, an algorithm for extracting geometric and other anthropometric features from face image. Training and testing are done based on the classifiers and the accuracy achieved is 57% for detecting six different classes of emotions.

The author in [24] proposed the facial expression detection model based on Fisher face model. Facial features can be extracted using fisher face and facial expression detection is done using neural networks. The model is very sensitive in terms of illumination, noise and different backgrounds images are not suitable for this model. Accuracy is almost 86.85%.

The emotion detection of facial expression requires an image that is face of the person that needs to be recognized. Speech recognition doesn't require any image or person details. This model can be used in health department, education and customer service. Intelligence service makes use of this method to find the truthfulness of the investigation. The model uses MFCC for various frequency sound processing and it also uses Universal background models which eliminate the overfitting problems along with Gaussian mixture models for normal distribution as discussed in [1].

The authors in [8] propose a model which will capture image and detect the facial emotion based on Viola-Jones algorithm and prepares a list of song based on the emotion recognized. Viola-Jones algorithm is used for face detection and Fisher face features classifier is used for emotion detection. Model is used for detecting the face and based on the face detected, emotion recognition is performed and based on the emotion of the user, and music player is built. Accuracy is less.

One of the greatest challenges faced during face recognition is occlusion detection. This model is used for detecting the missing information in the occluded face. Viola-Jones algorithm is used for face detection and principal component analysis is used for face reconstruction and neural networks are used for face recognition. When considered other algorithms or methods, neural network yields the best result as seen in [19].

The author [22] proposed a model that is an electric wheelchair with emotion detection of the user with impairment. For capturing images, an RPI board with camera is used to record the frames captured in a video stream, along with Raspberry P12 and B model incorporated. This model can recognize the face using the images captured from the streams through sample python scripting, for face detection and sample emotion detection. Only three emotions classes were detected that is anger, fear or danger.

Study [17] elaborates on a model that is built based on hidden markov model, which can detect real time user's emotion. The model can detect six different classes of

emotions. Using hidden markov model, facial expressions can be separated and identified automatically. Model is not suitable for different positions and texture or orientation of the face with respect to images. And the accuracy is very low.

The research [3] proposes a methodology where local histograms are normalized along with their features. These features have orientations that is gradient and give better results for human detection along with densely overlapped grids. A new pedestrian database is introduced by the authors. The impact of different descriptor boundaries for fine-scale angles, fine direction binning, coarse spatial binning, nearby differentiation standardization in covering descriptor blocks is immensely significant for great execution.

The focal point of this investigation is on PC computerized impression of human feeling as discussed in [6]. The investigation was based on utilization of Fisher faces for the acknowledgment of human feeling in facial pictures. Large number of Fisher face models are trained and the overall model is considered in contrast to an autonomous test set. The results obtained are used to construct and test a compound progressive framework that endeavors to decipher human feeling continuously utilizing face discovery and calculations in combination with the look examination technique. Outcomes demonstrate that Fisher faces can be valuable in anticipating feeling dependent on content recovered from facial pictures.

Research [16] manages two particular utilizations of PCA in picture handling. The principal application comprises of picture shading that decrease, while the three shading segments are diminished into one containing a significant part of data. The second utilization of PCA exploits eigenvectors properties for assurance to choose object direction. Different strategies can be utilized for past object identification. Nature of picture division suggests to consequences of the accompanying cycle of item direction assessment dependent on PCA as well. Introduced paper momentarily presents the PCA hypothesis from the start and proceeds with its applications referenced previously. Results are archived for the chosen genuine pictures.

The study [2] features the improvement of an Android stage-based application named XBeats which goes about as a Music Player dealing with Picture Preparing essentials to catch, investigate and present music according to the feeling or temperament of the client utilizing this application. Android application was created utilizing the Android SDK programming and OpenCV programming was utilized to execute facial acknowledgment calculations and falls. The one-of-a-kind part of this task is that it centers around facial acknowledgment on the Android stage dissimilar to that on PC frameworks which utilize normally accessible programming projects for something similar. This article additionally gives correlation between utilization of different arrangement calculations utilized for facial location.

Recent FER frameworks spotlight on two significant issues: overfitting brought about by an absence of adequate preparing information and articulation inconsequential varieties, like light, head posture and personality predisposition. For the best in class in deep FER, the article [14], presents existing novel profound neural organizations and related preparing methodologies that are intended for FER dependent on both static pictures what's more, powerful picture groupings and talk about their benefits and restrictions. Several demonstrations and exploratory correlations on generally utilized benchmarks are likewise summed up. Later stretch out the overview to extra related issues and application situations. At the end, audit the leftover difficulties and comparing open doors in this field just as future headings for the plan of vigorous profound FER frameworks.

In [21], author proposes an emotion recognition framework which can be created by using the advantages of profound learning and various applications like criticism investigation, face opening and so forth can be carried out with great exactness. The primary focal point of this work is to make a Deep Convolutional Neural Network (DCNN) model that groups five unique human facial feelings. The model is prepared, tried and approved utilizing the physically gathered picture dataset.

Research [26] proposes a methodology to examine expressions of the face, dependent on both permanent and transient facial features, in an almost front facing view face picture grouping. The AFA framework perceives fine-grained changes in expressions of the face into action units (AU) of the Facial Action Coding System (FACS), rather than a couple of prototypic articulations. Multistate face and facial segment models are proposed for following and demonstrating the different facial highlights, including lips, eyes, foreheads, cheeks, and wrinkles. During following, point by point parametric depictions of the facial highlights are separated. With these boundaries as the sources of information, a gathering of activity units are perceived whether they happen alone or in blends. The framework has accomplished normal acknowledgment paces of 96.4% for upper face AU and 96.7% for lower face AU.

The authors of [4] have proven that emotional facial expressions are universal. Many other studies signify that different other culture, literate people show up the same expressions for few similar emotions. So by this fact, one can draw the conclusion that facial expressions are universal. And different cultured people have different expressions, but people in all variety of cultures are exposed to one single emotion and almost majority had the same expressions, irrespective of their facial shape, the expression was similar. Dataset was generated by considering three sets of faces using New Guinea. From the hypothesis curve, the results prove that facial expressions for certain emotions are universal.

In speech, emotion recognition is a theme on which little examination has been done to date. In [9], the authors talk about why emotions to be recognized in speech is an intriguing and relevant exploration point and present a framework for feeling acknowledgment utilizing one-class-in-one neural organizations. By utilizing an enormous data set of phonemes adjusted words, our framework is speaker and setting autonomous. Upon testing eight emotions approximately 50% of recognition rate was achieved.

Facial recognition framework is utilized in numerous applications going from HCI, reconnaissance to feelings. The

extension is wide and a great deal should be possible in this field. Highest accuracy is achieved through Neural Network (NN) as a classifier and the blend of Gabor Wavelet (GW) and Local Binary Pattern (LBP) is likewise great with a precision of 91.2 and 90%. The other techniques like Principal Component Analysis and Support Vector Machine's exhibition were low as discussed in [11].

The study [15] primarily considers face recognition with the segments by face parsing (FP). Considering the burden that various pieces of face contain distinctive measure of data for look and the weighted capacity are not the equivalent for various appearances, a thought is proposed to perceive look utilizing segments which are dynamic in demeanor exposure. The face parsing finders are prepared through deep conviction network and tuned by logistic regression. The identifiers initially recognize face, and afterward identify nose, eyes and mouth progressively. A profound engineering pretrained with stacked autoencoder is applied to face recognition with the concentrated highlights of distinguished parts. The parsing segments eliminate the excess data in articulation acknowledgment, and pictures should not be adjusted or some other fake treatment.

The authors in [13] proposed a deep convolutional neural network to arrange the 1.2 million high-goal pictures in the ImageNet challenge that is images are classified into 1000 variety of classes. Based on the testing data, the model accomplished top-1 along with top-5 blunder paces of 37.5% which is more than 17.0%, much better compared to the existing best in class. The neural organization, had 60 million boundaries and 650,000 neurons, which consists of five different convolutional layers, and some of the layers had max-pooling layers, along with three completely associated layers and last layer had 1000-way softmax activation function applied. To make processing quicker, model had non-immersing neurons along with effective GPU execution of the convolution activity. In order to reduce overfitting in the fully associated layers the model had created regularization technique called "dropout" which end up being exceptionally successful.

In the extended cohn-kanade dataset (ck+), [18] the quantity of arrangements is expanded by 22% and the quantity of subjects by 27%. The objective articulation for each succession is completely FACS coded and feeling names have been reconsidered and approved. What's more, non-presented arrangements for a few kinds of grins and their related metadata have been added. The article presents standard outcomes utilizing Active Appearance Models (AAMs) and a straight help vector machine (SVM) classifier utilizing a leave-one-out subject cross-approval for both AU and feeling location for the presented information. The feeling and AU marks, alongside the all-encompassing picture information and followed milestones will be made accessible July 2010.

Essential backpropagation, which is a straightforward strategy currently being broadly utilized in regions like example acknowledgment and deficiency finding, is inspected. The fundamental conditions for backpropagation through time, and applications to regions like example acknowledgment including dynamic frameworks, frameworks

recognizable proof, and control are examined. Further expansions of this strategy, to manage frameworks other than neural organizations, frameworks including concurrent conditions, or genuine intermittent organizations, and other pragmatic issues emerging with the technique are portrayed. Pseudocode is given to explain the calculations. The chain rule for requested subordinates the hypothesis which underlies backpropagation is momentarily talked about. The emphasis is on planning an easier form of backpropagation which can be converted into PC code and applied straight by impartial organization clients as seen in [27].

III. PROPOSED WORK

- To train the model training samples or images are collected using FER Dataset. FER is nothing but Facial Emotion or Expression Recognition. Dataset contains 48x48 pixel-based grayscale facial images. FER Data is used to identify the emotions of the unseen data into one of the seven classes of emotions. The dataset is created by browsing each emotion or expression and their synonyms.

- To detect the face and recognize from the given input images, face-recognition library is used. Given any input, image is converted into grayscale image and data is stored in the form of array using numpy array. Face-recognition library loads the given input image and then encodes the loaded image into feature vector and returns the Boolean value. To detect the emotion, faces being detected are taken as input and return the emotion type.

- Using data augmentation performance of the model is improved. The performance of the model is calculated by plotting the confusion matrix that is by considering the true positive and negative values. The accuracy is nothing but the precision, recall and f1-score.

A. Issues, Controversies, Problem

Improving the automatic facial emotion recognition will help in upgrading the social knowledge in the machines. In the field of computer vision, multiple applications make use of facial emotion recognition system and it is very challenging to identify multiple facial emotion in a given frame or image.

B. Proposed Methodology

Face detection is the key mechanism for emotion detection. Because of mental thinking and physical circumstances, humans tend to change their expression from dusk till dawn. The proposed methodology consists of input image, preprocessing, face detection, face recognition, feature extraction, emotion detection/classification. During face detection, the model automatically finds the face locations from the given input image. These face locations are the output of the face detection that is bounding box face and this achieved using face detection library. Once face is detected using face detection library face recognition is done. Once face is recognized, meaningful features are extracted and face emotions or expressions will be analyzed. Basic facial features are eyebrows, eyes, nose, cheeks and mouth. Facial

expressions are classified under seven classes, happy, sad, fear, angry, disgust, neutral and surprise.
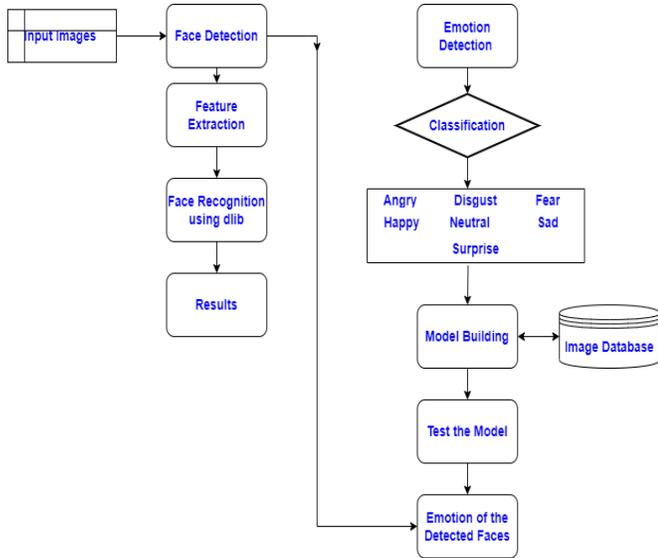
Fig. 1. Proposed Methodology.

As shown in the Fig. 1, the proposed methodology contains three pipeline stages.

1) Face Detection

2) Face Recognition

3) Emotion Detection

C. Face Detection

Detecting the faces in a given image is the first stage of the pipeline. Face recognition is the python library that is used for facial detection. This python library scans any given input image and outputs the bounding box coordinates of the faces being detected in the given input image. Face detection is special case of object detection technology in which human faces are detected from given input image digitally.

D. Face Recognition

Facial recognition is done using face-recognition library which manipulates the faces from input image. Face-recognition library is built using dlib along with deep learning. Through command line options, facial recognition can be done and that is achieved through dlib's face-recognition library. Facial detection is done using face-recognition library which return bounding box coordinates as output. Then find the required features from given input image and manipulate the features of the face from the given image. That is face recognition. Face landmarks will return person's eyes, chin, nose and mouth. And this is how any makeup related options in the camera works.

Using encoding and based on the images trained or the dataset used, who appears in the image can be recognized using face-recognition library. Using face-recognition library, one can verify any given image and recognize who appears in the image digitally. Create two folders, one with training the model with all the images required. And the other folder which contains all the images present in the training dataset

but in a different angle. Then on unseen data if the input image resembles any of the trained images, then the output will be, the model verifies the image and recognizes the face. Using face-recognition library the proposed model outputs Boolean value if the face matches, then returns true else false.

E. Emotion Detection

Human emotions can be recognized in multiple ways that is through voice, action, facial expressions, and body language and so on. Facial expressions for each emotion are innate and they are not any other part of cultural learning. But making the computer to do so is a challenging task. Facial expressions can be categorized as a form of nonverbal communication. For any given input image face detection and recognition is part of object detection and recognition from any given input images. Humans have tendency to express their emotions through facial expressions at most of the times.

The emotions can be classified into seven different classes apparently. That is happy, sad, anger, disgust, fear, neutral and surprise. The model is built using six layers of convolutional neural network as shown in the below Table I.

TABLE I.   SUMMARIZING LAYERS OF CONVOLUTIONAL NEURAL NETWORK

Layer (type)	Output Shape	Parameters
conv2d (Conv2D)	(None, 46, 46, 32)	320
conv2d_1 (Conv2D)	(None, 44, 44, 64)	18496
max_pooling2d (MaxPooling2D)	(None, 22, 22, 64)	0
conv2d_2 (Conv2D)	(None, 20, 20, 128)	73856
max_pooling2d_1 (MaxPooling2	(None, 10, 10, 128)	0
conv2d_3 (Conv2D)	(None, 8, 8, 128)	147584
max_pooling2d_2 (MaxPooling2	(None, 4, 4, 128)	0
conv2d_4 (Conv2D)	(None, 4, 4, 7)	903
conv2d_5 (Conv2D)	(None, 1, 1, 7)	791
flatten (Flatten)	(None, 7)	0
activation (Activation)	(None, 7)	0

Total params: 241,950
Trainable params: 241,950
Non-trainable params: 0

F. Convolutional Neural Network

Fully connected convolution neural network consists of four inputs that is input layer, and one hidden layer with five neurons and an output neuron in output layer. In fully connected neural network each neuron of the input layer is connected to all the neurons of hidden layer. Then all the computations are performed at hidden layer and the output is presented at the output layer. When any input image is fed into the convolutional layer, apply filters or choose proper parameters along with strides or padding required. Perform the operation that is convolutional on the input. For non-linearity apply Relu Activation function to the convolution operation. Pooling is performed to reduce the dimensionality. Until the error rate is reduced convolution layers are added. Then this flattened output is fed to fully connected convolution layer.

Then model will output based on the activation function used and then object detection and their classification are done.

Artificial Intelligence helps to view the world through computer vision as human does. It is nothing but bridging the gap between humans and computers. It helps in analyzing the world as humans do and perceiving the inputs in the similar fashion how human brain works. Convolutional neural network is a deep learning algorithm which can take input, assign the weights based on the objects in the image and identify these objects or different objects can be identified. Image pre-processing required is very lower when compared to other algorithms available for classification or recognition of objects. Using convolutional neural network, the model can identify any filters applied in the image without any training and they have the ability to learn these filters or characteristics of the image.

To obtain maximum values of the image portion covered by the kernel, max pooling is performed. Max pooling reduces the noise in the input along with the dimensionality reduction. Convolutional neural network will have these layers increased if the input dataset contains more complex images or to capture more low-level detailed information of the image. The output of the final layer can be fed as input for other neural networks, flatten the image which is transposing the row to column vector. Then backpropagation algorithm can be applied to get the network for each training iteration. Then after several series of epochs, model will learn through experience and would provide very accurate results.

G. Data Augmentation

Accuracy prediction for any supervised model largely depends on diversity of data and amount of data available during training. Success of any deep learning model depends on the relationship between deep learning models and the higher amount of data required. Data augmentation is a way of dealing with the problem of limited dataset. Image augmentation is nothing but synthesizing new data from the data available by applying various transformations.

Some of the simple geometric transformations applied on the images are flipping, cropping, scaling, rotating and color casting or pixel transformation that is color space transformations. The proposed model uses geometric transformation because of the positional biases present in the dataset. Color transformation helps in dealing with the problems related to lighting or illumination of the given input image. The proposed model uses keras image data generator augmentation.

As shown in the above Fig. 2, the workflow of the proposed methodology for any given input image, face detection and face recognition are done using face recognition library which python offers with upgraded version. For emotion detection, there are seven classes of emotions and the faces being detected from the image belongs to which class that needs to be detected. Dataset is downloaded from kaggle that is facial emotion recognition dataset. Once dataset is loaded, divide the number of classes of emotions. The data augmentation is done to improve the performance of the model.
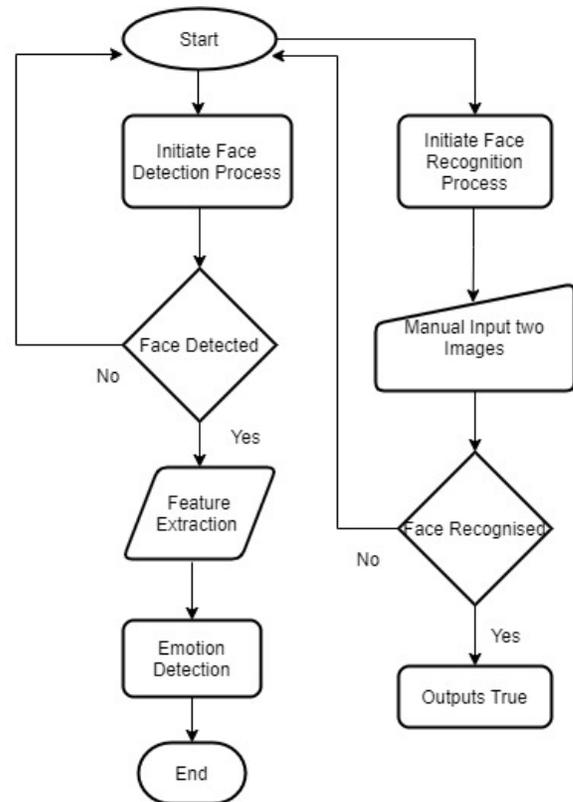
Fig. 2.   Flow Chart of the Proposed Methodology.

IV. RESULTS

Face detection is done using face-recognition library. As shown in the Fig. 3 the output of face detection using dlib's face-recognition library. Given input image has three human faces, and the output is the bounding box coordinates of all the three images present in the input image.

Fig. 3.   Face Detection.

Face recognition is done is using face-recognition library. The output of the face recognition is shown in the below Fig. 4. As shown in Fig. 4, face recognition is nothing but verifying or comparing any given two input images and if two matches then output the Boolean value true or else false. In the below example, took two input images of the same actress and the output was true.
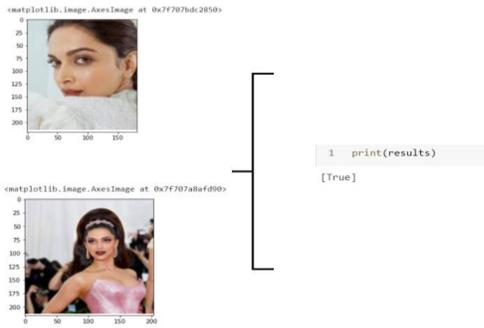
Fig. 4. Face Recognition.

Each time when validation accuracy is increased, model was saved. Calculated model loss for each epoch and plotted graph against the loss occurred vs. validation loss occurred in each epoch as shown in the Fig. 5.
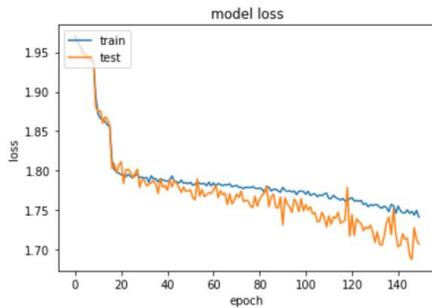
Fig. 5. Model Loss.

Model was trained using facial emotion recognition dataset. There were 28709 training samples which belong to seven different classes of emotions. And 3589 images were considered as validation samples and 3589 images were considered as testing samples. Facial emotion dataset was downloaded through kaggle. There are many open websites and from where image dataset was downloaded. Then image dataset was generated from .csv format into the specific folder. Then the dataset was used for training, validating the model and testing the model on unseen data.

As shown in the below figure, model performance was calculated by plotting confusion matrix for seven different classes of emotions considered for all validations samples considered through validation generator. The accuracy was achieved up to 85% when evaluated the model using confusion matrix as shown in the Fig. 6. Confusion matrix is actual a 2x2 matrix with actual data on one axis and predicted data on the other. Precision and recall are the important terms for performance evaluation, since positive class was considered the most and all other irrelevant or unwanted weeds are not considered while evaluating the performance of the model as shown in Fig. 7.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \qquad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \qquad (2)$$

F1-score is the mean of both precision and recall considered. And it takes both false positive and negatives and performs well on dataset which is imbalanced.

$$F1\ score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2*(Precision*Recall)}{Precision + Recall}$$

The output of the emotion detection is shown in the Fig. 8. The trained model is loaded and emotion is predicted for each face being detected.
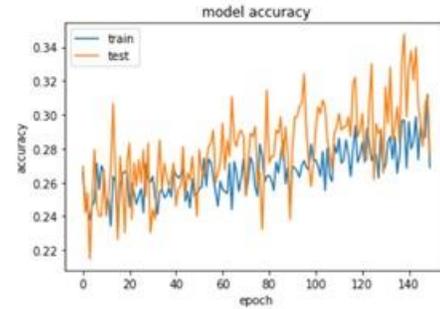
Fig. 6. Model Accuracy.

```
Confusion Matrix
[[331    1   39    8   56   52    4]
 [ 18   30    3    1    1    2    0]
 [ 88    2  243   12   59   94   30]
 [ 28    1   16  754   45   19   16]
 [114    8   79  148  113  105   59]
 [ 56    1   57   16  148  307    9]
 [ 12    0   67   19   18    6  294]]
Classification Report
              precision    recall  f1-score   support

       Angry       0.51      0.67      0.58       491
     Disgust       0.70      0.55      0.61        55
        Fear       0.48      0.46      0.47       528
       Happy       0.79      0.86      0.82       879
     Neutral       0.26      0.18      0.21       626
         Sad       0.52      0.52      0.52       594
    Surprise       0.71      0.71      0.71       416

   avg / total       0.56      0.58      0.56      3589
```

Fig. 7. Performance Evaluation of Model.
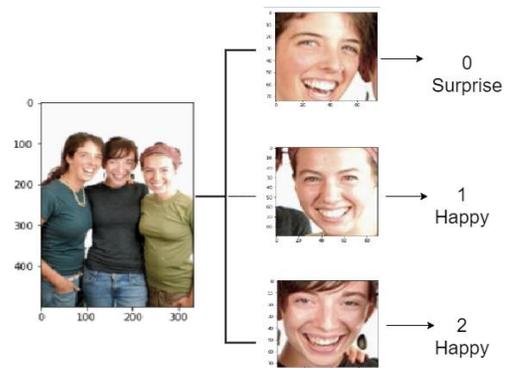
Fig. 8. Emotion Detection.

V. CONCLUSION

The proposed model is used for recognizing facial feelings using computer vision along with AI calculations which characterize the given input facial image into eight distinct feelings. On implementation using multiple algorithms and convolutional neural network, CNN outperforms with best

accuracy for image dataset. Model results suggest that computer can have human like capacity to recognize the feelings, at least for applications in which front facing perspectives can be expected utilizing the webcam. Proposed model performs well on classifying the given image into seven different classes of emotions irrespective of number of faces present in the image.

The proposed model consists of six convolutional layers attaining the accuracy of almost 85%. The model uses softmax activation function in the last layer of a fully connected convolutional neural network which makes the model more unique and more efficient in terms of performance. Visualization of the model is done based on the parameters, layers used and results obtained are also shown in terms of graphs and plotted confusion matrix as well for better understanding.

Emotion recognition will be exceptionally valuable soon in the exploration field of advanced mechanics and man-made reasoning for instance if a robot can detect the feelings of any human and that robot can act appropriately without any human aid. This model can be further extended for emotion recognition, can likewise be extended to the identification of different feelings other than these eight all-inclusive feelings.

REFERENCES

[1] Anithadevi, N., Gokul, P., Nandan, S. M., Magesh, R., & Shiddharth, S. (2020). Automated Speech Recognition System For Speaker Emotion Classification. 2020 5th International Conference on Computing, Communication and Security (ICCCS). Patna, India: IEEE.

[2] Chavan, S., Malkan, E., Bhatt, D., & Paranjpe, P. (2014). XBeats-An Emotion Based Music Player. International Journal For Advance Research In Engineering And Technology.

[3] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, CA, USA: IEEE.

[4] Ekman, P. F. (1971). Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 124–129.

[5] Engoor, S., SendhilKumar, S., Sharon, C. H., & Mahalakshmi, G. (2020). Occlusion-aware Dynamic Human Emotion Recognition Using Landmark Detection. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE.

[6] Fratesi, A. (2015). Automated Real Time Emotion Recognition Using Facial Expression Analysis. Carleton University Research Virtual Environment.

[7] Gupta, S. (2018). Facial emotion recognition in real-time and static images. 2018 2nd International Conference on Inventive Systems and Control (ICISC), DOI: 10.1109/ICISC.2018.8398861, 2018. IEEE.

[8] Iyer, A. V., Pasad, V., Sankhe, S. R., & Prajapati, K. (2017). Emotion based mood enhancing music recommendation. 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). Bangalore, India: IEEE.

[9] J. Nicholson, K. T. (2000). Emotion Recognition in Speech Using Neural Networks. Neural Computing & Applications, 290–296.

[10] Jaiswal, A. (2020). Facial Emotion Detection Using Deep Learning. 2020 International Conference for Emerging Technology (INCET), DOI: 10.1109/INCET49848.2020.9154121.

[11] Kauser, N., & Sharma, J. (2016). Automatic facial expression recognition: A survey based on feature extraction and classification techniques. International Conference on ICT in Business Industry & Government (ICTBIG). Indore , India: IEEE.

[12] Kolodziej, M., Majkowski, A., Rak, R. J., Tarnowski, P., & Pielaszkiewicz, T. (2018). Analysis of Facial Features for the Use of Emotion Recognition. 19th International Conference Computational Problems of Electrical Engineering. IEEE.

[13] Krizhevsky, A., Sutskever, I., & Hinton, G. (2017). ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACMVolume 60Issue 6, 84–90.

[14] Li, S., & Deng, W. (2020). Deep Facial Expression Recognition: A Survey. IEEE Transactions on Affective Computing, IEEE.

[15] Lv, Y., Feng, Z., & Xu, C. (2014). Facial expression recognition via deep learning. 2014 International Conference on Smart Computing. Hong Kong, China: IEEE.

[16] M. Mudrov´a, A. P. (2004). Principal Component Analysis in Image Processing . Institute of Chemical Technology, Prague, Department of Computing and Control Engineering.

[17] Mishra, P. (2018). HMM Based Emotion Detection in Games. 3rd International Conference for Convergence in Technology (I2CT) 2018. Pune, India: IEEE.

[18] P. Lucey, J. C. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE.

[19] Patel, T. B., & Patel, J. T. (2017). Occlusion Detection and Recognizing Human Face Using Neural Networks. 2017 International Conference on Intelligent Computing and Control (I2C2 2017) (pp. 904-908). Coimbatore, Indai: IEEE.

[20] Pathar, R., & Adivarekar, A. (2019). Human Emotion Recognition using Convolutional Neural Network in Real Time. 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), DOI: 10.1109/ICIICT1.2019.8741491.

[21] Pranav, E., Kamal, S., Chandran, C. S., & Supriya, M. (2020). Facial Emotion Recognition Using Deep Convolutional Neural Network. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). Coimbatore, India: IEEE.

[22] Rabhi, Y., Mrabet, M., Fnaiech, F., & Sayadi, M. (2018). A real-time emotion recognition system for disabled persons. 2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). Sousse, Tunisia: IEEE.

[23] Reney, D., & Tripathi, N. (2015). An Efficient Method to Face and Emotion Detection. 2015 Fifth International Conference on Communication Systems and Network Technologies, DOI: 10.1109/CSNT.2015.155.

[24] Saaidia, M., Zermi, N., & Ramdani, M. (2014). Facial Expression Recognition Using Neural Network Trained with Zernike Moments. 2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology. IEEE.

[25] Sergeeva, A. D., Savin, A. V., Sablina, V. A., & Melnik, O. V. (2019). Emotion Recognition from Micro-Expressions: Search for the Face and Eyes. 2019 8th Mediterranean Conference on Embedded Computing (MECO). IEEE.

[26] Tian, Y.-I., Kanade, T., & Cohn, J. (2001). Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[27] Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. Proceedings of the IEEE, 1550 - 1560.

[28] Castellano G, De Carolis B, Macchiarulo N (2021) Automatic emotion recognition from facial expressions when wearing a mask. In: CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter. CHItaly '21. Association for Computing Machinery. 10.1145/3464385.3464730.

[29] Castellano G, De Carolis B, Macchiarulo N. Automatic facial emotion recognition at the COVID-19 pandemic time(2022). Multimed Tools Appl. 2022 Oct 22:1-19. doi: 10.1007/s11042-022-14050-0. Epub ahead of print. PMID: 36313484; PMCID: PMC9589699.

Big Data Analytics Quality in Enhancing Healthcare Organizational Performance: A Conceptual Model Development

Wan Mohd Haffiz Mohd Nasir, Rusli Abdullah, Yusmadi Yah Jusoh, Salfarina Abdullah
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia (UPM), Serdang, Selangor, Malaysia

Abstract—The advancement of Big Data Analytics (BDA) has aided numerous organizations in effectively and efficiently adopting BDA as a holistic solution. However, BDA quality assessment has not yet been fully addressed, therefore it is necessary to identify essential BDA quality factors to assure the enhancement of organizational performance, particularly in the healthcare sector. Hence, the goals of this study are to recognize and analyse the determining factors of BDA quality as well as to suggest a conceptual model for enhancing the performance of healthcare organizations via BDA quality assessment. The proposed conceptual model is based on a related theoretical model and previous research on BDA quality. The essential BDA quality factors being selected as determinants consist of reliability, completeness, accuracy, timeliness, format, accessibility, usability, maintainability, and portability. The findings of this ongoing study are used to develop a conceptual model that is proposed in line with the ten-research hypothesis and may offer a better assessment quality model to improve the performance of healthcare organizations.

Keywords—Big data analytics; BDA quality factors; BDA quality assessment; organizational performance; healthcare

I. Introduction

Big Data Analytics (BDA) has been used as an end-to-end solution in most organizations, especially healthcare organizations. The use of BDA contributes a significant impact on the enhancement of organizational performance. There are various potential outcomes of BDA for healthcare organizations such as Information & Communication Technology (ICT) infrastructure, operational, managerial, strategic, and organizational benefits [1]. BDA implementation in healthcare organizations affects the efficiency of ICT infrastructure, thus reducing system redundancy and quickly transferring data among healthcare ICT systems and applications [2]. BDA potentially produces some operational outcomes such as enhancing the precision and quality of clinical judgments, processing a substantial amount of health information quickly, and having access to medical data right away for analytics [3]. Likewise, the recent pandemic of Covid-19 has proven BDA as a significant solution in healthcare organizations not just for the analytics part but also for data acquisition for better planning, pandemic, and patient monitoring [4].

The BDA implementation in the healthcare industry may face certain obstacles that have been faced when developing solutions specifically to deal with the Covid-19 pandemic [5]. Information accuracy, security, and privacy are some of the related BDA issues that may affect the performance of healthcare organizations [5][6]. Moreover, healthcare organizations must assess BDA quality to make sure that the related system and application produced by BDA implementation is portable, near real-time information, and easy to maintain to entice consumers to use it in the long run [7]. The effectiveness and efficiency of healthcare organizations can be increased with the right and essential BDA quality factors being fulfilled [8] and this is in line with the finding by Fosso Wamba [9]. The negative implications of poor BDA quality, such as incorrect decision-making, have caused considerable harm to organizations [10][11] especially in healthcare due to life or death decisions depending on having accurate information [12].

In the context of Malaysia's healthcare, Malaysia ranked 47th out of 60 countries left far behind Singapore ranked 9th for access to Universal Healthcare Coverage (UHC) in the recent The Cost of Healthcare Index report [13]. Access to the UHC is a score that represents how much of a country's population has easy access to essential health services and it covers the entire range of high-quality healthcare services, including palliative care, treatment, rehabilitation, and health promotion [14][13]. Achieving UHC is one of the targets the nation of the world set when adopting the Sustainable Development Goals (SDGs) in 2015 [14]. It is showing the urgency for the assessment of BDA quality to overcome BDA quality such as accessibility, usability, and reliability.

Various studies have been conducted on BDA quality; however, the focus is more on "data" quality and does not cater to BDA quality holistically [10][7], and the BDA quality assessment in healthcare is still an understudied research topic. Consequently, the objectives of this paper are to 1) determine the contributing factors to BDA quality; and 2) develop a BDA Quality (BDAQ) conceptual model for enhancing healthcare organizational performance.

The subsequent section will examine the theoretical context, and the third section will glance at the research methodology. The discussion and conclusions of the BDA quality factors are covered in the fourth section, which also suggests a conceptual model for the BDAQ. The study is concluded in the last section by highlighting the contribution of the research and making suggestions for future studies.

II. THEORETICAL BACKGROUND

Several studies have focused on constructing BDAQ models by adopting the Information System Success Model (ISSM) of DeLone and McLean [15] and Wang and Strong [16] while other studies are also incorporating resource-based view theory (RBV) [17]. From the RBV perspective, an organization's competency is dependent on quality to efficiently manage its vital resources and gain a competitive advantage, which can be reflected in enhanced organizational performance [9]. RBV focuses on generating exceptional organizational performance by establishing valuable, rare, unique, and irreplaceable (RIN) resources of superior quality [18]. RBV is in line with the Information System Success Model (ISSM) of DeLone and McLean [15] as both highlight the competencies of in-house information systems to influence organizational performance [9]. To improve healthcare organizational performance, ISSM has been adapted to emphasize data, information, and system quality factors related to outcomes and benefits for the healthcare organization [19].

System and information quality have both been recognized in ISSM as crucial elements of organizational effectiveness [20]. Most previous studies have used ISSM as the main theoretical framework. Wamba [21] in their findings suggests that information quality represents four essential BDA quality factors; completeness, currency, format, and accuracy have a significant positive impact on organizational performance. Chiu [22] had been used ISSM with three groups of quality factors; system quality, service quality, and information quality with a mediating effect on user satisfaction and user continuance intention revealed that the three groups of quality factors positively influenced user satisfaction, and user continuance intention.

Wang and Strong [16] as a theoretical model in another way presented quality factors into four primary categories, namely intrinsic, contextual, representational, and accessibility. The intrinsic category implies datasets having inherent quality, whereas the contextual category emphasizes the requirement of the task that quality be evaluated in context [7], [16]. The representational category highlights quality in terms of its presentation, while the accessibility category reinforces the significance of computer systems that give access to data [7], [16][23]. There are some characteristics for each category that are employed as particular indicators of quality. For instance, the intrinsic category's factors are objectivity and accuracy, while the contextual category's factors are timeliness and relevance. The representational category's factors are interpretability and understandability, while the accessibility category's factors are accessible security and simplicity of use [7]. Ghasemaghaei and Calic [24] used BDA quality factors produced by Wang and Strong [16] to examine the effect of BDA on quality factors and confirmed the critical role of BDA quality in enhancing organizational decision-making.

Current BDA quality models primarily rely on the Information System (IS) theoretical model and are largely based on data and information quality rather than a suitable and comprehensive component of BDA quality factors [7], [25]. Differing from traditional Business Intelligence (BI), BDA, as mentioned earlier, covers end-to-end solutions, a full life cycle, and is inclusive of related BDA systems and applications. Thus, it is very important to leverage broader and holistic theoretical models. To fill this theoretical gap, this study integrates the most common theoretical model in Information System (IS) research with the related Software Engineering (SE) theoretical model which are McCall[26], Boehm[27], Evan & Marciniak, Deutsch & Willis, DROMY's, FURPS+'s, SEI, and ISO/IEC25010 [28]. Table I present the occurrence of quality factor from the ten theoretical models that have been summarized for this study.

TABLE I.   QUALITY FACTORS DERIVED FROM THEORETICAL MODELS

Theoretical Models / Factors	McCall [26]	Boehm [27]	Dromey [31]	Evans & Marciniak [32]	Deutsch & Willis [33]	FURPS+'s [34]	SEI [35]	ISO/IEC 25010:2011 [36]	Wang & Strong [16]	DeLone & McLean ISSM [15]	Frequency
Reliability	×	×	×	×	×	×	×	×		×	9
Completeness	×	×		×	×			×	×	×	7
Maintainability	×	×	×	×	×		×	×			6
Accuracy	×	×		×	×				×	×	6
Portability	×	×	×	×	×			×			6
Usability	×			×	×	×		×		×	6
Efficiency	×	×	×	×	×						5
Integrity	×	×		×	×	×					5
Flexibility	×			×	×					×	4
Consistency	×	×		×							3
Understandability		×							×	×	3
Testability	×	×						×			3
Relevancy									×	×	2
Compatibility								×			1
Believability									×		1

III. METHODOLOGY

As shown in Fig. 1, several steps were taken in the creation of the conceptual model for assessing the BDA quality factors. Prior to identifying the pertinent theories, the review process continued with content analysis and was founded on the Systematic Literature Review (SLR) approach [29][30]. 23 publications on the BDA quality in various domains were investigated as part of the review activity, which divided them into two categories: empirical and case study research analysis. A research question like "What are the essential BDA quality factors to consider in constructing BDA Quality (BDAQ) model?" is formulated as the first step in the initial investigation. After further research was conducted using electronic journal databases like Scopus, Science Direct, Emerald, IEEE, Google Scholar, and the snowballing method, 23 pertinent publications were selected. After that, a matrix table was used to examine the pertinent articles (Table II).

This study selected nine factors from theoretical models and previous research to be considered in the development of the BDAQ model, with the results detailed in the subsequent section (Section IV). In conclusion, the conceptual model and related factors are discussed. Finally, in the conclusion, the BDAQ conceptual model is presented along with the essential BDA quality factors.
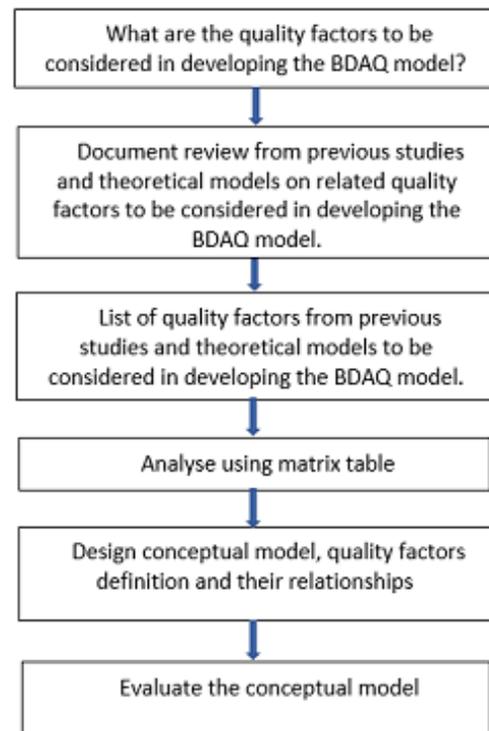
Fig. 1. The Workflow of BDAQ Conceptual Model Development.

TABLE II. BDA QUALITY FACTORS FROM PREVIOUS STUDIES

Quality Factors	Wook et al. (2021) [7]	Ji-fan Ren et al. (2017) [20]	Fosso Wamba et al., (2019) [9]	Côrte-Real et al. (2020) [10]	Aljumah et al. (2021) [37]	Trinchera, et al. (2019) [21]	Kwon et al. (2014) [38]	Ghasemaghaei (2019) [24]	Akter et al. (2017) [39]	Shariat Panahy et al. (2013) [40]	Ayyash, (2017) [41]	Ameen et al. (2020) [42]	Zheng et al. (2013) [43]	Popovič et al. (2012) [44]	Masri et al. (2020) [45]	Gürkut & Nat, (2018) [46]	Gorla et al. (2010) [47]	Alaoui & Gahi, (2019) [48]	Bahrami & Shokouhyar (2021) [49]	Haryadi et al. (2016) [50]	Verma (2017) [51]	Chiu et al. (2021) [22]	Keikhosrokiani et al. (2020) [52]	Frequency
Timeliness	×	×	×	×	×	×		×	×	×	×		×		×	×		×	×	×	×		×	18
Completeness	×	×	×	×		×	×	×	×	×	×				×	×		×		×	×	×		16
Accuracy	×	×	×	×		×		×	×	×	×				×	×		×		×	×		×	15
Reliability		×	×		×			×				×	×	×	×	×	×			×	×	×	×	14
Format		×	×	×		×		×	×				×		×		×							9
Accessibility	×	×			×			×					×	×						×	×			8
Responsiveness		×	×									×						×				×		5
Privacy		×			×			×																3
Usability																						×	×	2
Interactivity												×										×		2

IV. FINDINGS AND DISCUSSION

The relevant BDA quality factors from the SLR are presented in Table II and include the following; reliability, completeness, accuracy, timeliness, accessibility, and format that affects enhancing organizational performance. Table III presents the integration of quality factors from a related theoretical model and from the previous studies with nine quality factors that have been chosen based on their frequency and suitability in the healthcare context. The frequency of each quality factor was shown in Fig. 2 as a summary frequency of quality factors integrated from related theoretical models and previous studies.

Reliability and completeness were shown to be the two most frequent quality factors highlighted in both theoretical models and previous studies on BDA quality. This is followed by accuracy, timeliness, format, accessibility, usability, and maintainability. Maintainability and portability have been found not been discussed in BDA quality previous studies. However, these two quality factors are the most prominent quality factors found in the related theoretical model as can be seen in Table I. Based on the finding from the theoretical model matrix, this study decided to consider and choose maintainability and portability as part of the nine essential BDA quality factors for the BDAQ conceptual model. Drawing on the ten theoretical models and twenty-three previous studies from the SLR, the BDAQ model is conceptualized and determined by nine quality factors: reliability, completeness, accuracy, timeliness, format, accessibility, usability, maintainability, and portability.

A. Reliability

The definition of reliability by McCall et al. [26] is "the extent to which a program is expected to perform its intended function with the needed precision". The reliability of BDA reinforces the confidence that a BDA platform is interference and disruption-free [37]. Reliable data helps BDA to provide precise decision support and increase the effectiveness of services offered [10], [38]. An overwhelming amount of research studies discovered a relationship between improving organizational performance and reliability as a BDA quality factor [9], [20], [39]–[41]. Hence, it is encouraging to examine reliability as one of the essential BDA quality factors and its relationship with healthcare organizational performance.

Fig. 2. Summary of BDA Quality Factors.

B. Completeness

Completeness is one of the most frequent quality factors from theoretical models and previous studies as shown in Table III. Completeness in this study context refers to that all necessary data and information of BDA are provided and complete either in the stage of data acquisition, data processing, visualization, analytics, related BDA's system and application in the healthcare organization. Failure to assess the completeness of the BDA in healthcare will heavily affect not just decision-making but the confidence of the end-users for the healthcare services [53]. Numerous research demonstrated the significant impact completeness has on improving organizational performance as a primary determinant of BDA quality factors [9], [20], [42], [54]. Therefore, completeness has been selected as one of the essential BDA quality factors of the proposed BDAQ conceptual model.

C. Accuracy

According to Wixom and Todd [55], accuracy includes perceived exactness of information, conformance to truth or value in the real world, correctness, validity, and precision [16], [56]. It also checks to see if the data were entered correctly and if the values are accurate [57]. As part of the BDA quality factors, accuracy is important in healthcare because it can affect how decisions are made [58], [59] and have a strong impact on organizational performance based on the experimental case study by Alaoui & Gahi [48]. Inaccurate data may lead to clinical mistrust, the inability to properly interpret data, as well as a higher likelihood of mistakes [60].

D. Timeliness

Making decisions based on old data might result in erroneous insights, hence timeliness has been regarded as an important factor of BDA quality [61]. Timeliness or currency is defined as "the amount to which the age of data is appropriate for the task at hand" and represents the user's perception of the degree to which the information is up to date [55]. In BDA's healthcare applications such as clinical decision support, hospitals, caregivers, etc., the timeliness of data is one of the greatest issues [62]. Since doctors make decisions based on larger amounts of clean, up-to-date data, the process should be less cumbersome, faster, and more precise [62]. Most past empirical studies found timeliness was a significant BDA quality factor and had a significant positive effect in enhancing organizational performance [9], [20], [49], [63]. However, a study by Wook et al. [7] found that timeliness had no significant effect on BDA applications conducted from individual perspectives. Timeliness is one of the essential BDA factors that had been selected for the proposed BDAQ model, thus the significance of timeliness will be analyzed in the actual study and the relationship in enhancing healthcare organizational performance will be examined.

E. Format

The format "represents the user's perspective of how well the information is presented" and the BDA system's presentation of the information [9], [55]. In the area of healthcare, the healthcare information offered by BDA is well-formatted, well-organized, and presented clearly on the screen [55]. Data is obtained and acquired from numerous sources in
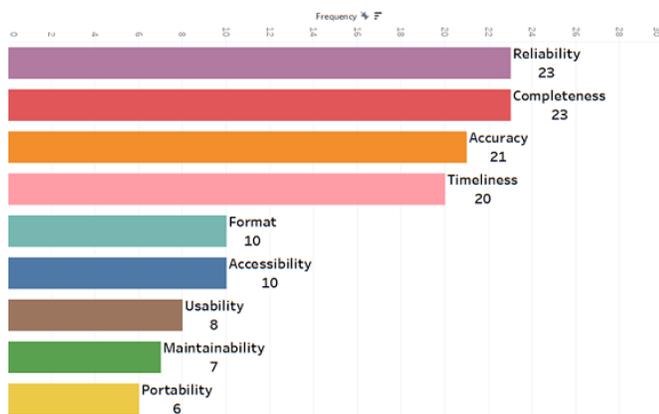
both electronic and non-electronic formats, and then it is transformed into the desired format during this essential phase [60], [64]. But data sources come in a wide range of functions, structures, and forms [65] including text, photo, audio, video, web content, GPS data, sensor data, and others, and are geographically scattered [66]. BDA in healthcare is challenging because it must deal with numerous types of unstructured data, in contrast to traditional data analysis where data types are often structured [6][67]. However, BDA is anticipated to work with a combination of unstructured and structured data [6]. Numerous studies have looked at and suggested that format was a significant determinant of BDA quality factors and significantly improved organizational performance [9], [21], [47], [63], [68].

F. Accessibility

Accessibility is the degree to which data is available, easily accessed, and quickly retrieved [16]. The system's accessibility influences the extent to which managers may access a scalable and efficient analytics platform [37]. Wang and Strong [16] emphasize that a system must be accessible without tolerance for privacy and security. Accessibility in the context of BDA in healthcare is not just about data accessibility but BDA's system and application accessibility. Wook et al. [7] selected accessibility as a BDA quality factor because the computer system must ease data access and storage in BDA applications. The efficacy of BDA implementation in healthcare is significantly determined by accessibility [69]. BDA solutions as end-to-end healthcare solutions must be able to generate unconstrained applications of diffused and accessible machine learning algorithms that support healthcare data analytics on Hadoop systems [70]. BDA applications in healthcare either web-based or mobile have now given a big impact on the public in accessing healthcare information. A finding by Wook et al. [7] demonstrated that data accessibility plays a significant role in bridging the digital divide and providing users with effective access to BDA applications. The studies revealed that accessibility has a substantial impact on BDA applications [7].

G. Usability

Usability describes the amount of work needed to understand, use, prepare input, and interpret a program's output [26]. Although modern medical BDA has numerous drawbacks, it is nonetheless a very promising resource for insurers and other academics. Not only for doctors and patients but also for everyone interested in secondary usage, improving BDA quality should be a top focus. There are a variety of steps that could be taken to improve the usability of data and the system of BDA in healthcare, and some of them are through quality assessment [71]. Ensuring high levels of usability to access pertinent data is a frequently mentioned challenge for BDA's system and application in healthcare [72]. Medical practitioners and researchers have used extensive BDA to improve patient care through data science, but their usability in the presence of BDA cannot be guaranteed to justify clinical use enough [72]. Two studies from twenty-three previous studies have usability as a quality factor as shown in Table II. User satisfaction and the success of human interaction with healthcare applications will be achieved when usability is fulfilled [52]. The testing is

typically associated with usability testing to reduce system defects and evaluate user experience [52]. Thus, this study considers usability as one of the essential BDA quality factors to be examined further.

H. Maintainability

Maintainability refers to the efforts required by users and maintenance staff to determine the cause of system failure, and the components or system that need to be fixed or restored [26]. Based on Table III, maintainability is found in seven of ten of the theoretical models. However, maintainability has been not found in any of the twenty-three previous studies on BDA quality. Maintainability has been selected as one of the determinants of BDA quality factors. Assessing maintainability will allow users to customize the setting on BDA's system and application without causing any errors or affecting the system's quality [73]. BDA's system and application in healthcare should be easy to maintain as it will very important to restore the operable condition within a specified time in case of any failure. The enormous amount of data in the healthcare system is expanding exponentially at a breakneck speed [62]. Collective decision-making will be impacted as a result of the genuine obstacles in maintaining BDA storage and processing in healthcare organizations since incomplete healthcare data will be more challenging to incorporate into an analytical platform [62].

I. Portability

The degree to which a system may be deployed or transferred from one hardware, software, or other operational or usage environment to another is referred to as portability [27], [74]. Similar to maintainability, portability has been selected to fill the theoretical gap for this study. The challenges of BDA portability include the simultaneous combination of information derived from many platforms, particularly in the healthcare context, to accommodate all healthcare information [75]. In today's digital era, everyone appears preoccupied with tracking their fitness and health data using the built-in pedometers of their portable and wearable gadgets such as smartphones, smartwatches, fitness dashboards, and tablets [76]. BDA's system and application in healthcare should be effectively adapted for a different software, in evolving hardware, and work independently with other systems [77]. Therefore, portability is selected as one of the essential BDA quality factors that could affect enhancing healthcare organizational performance.

TABLE III. SUMMARY OF BDA QUALITY FACTORS

Quality Factors	*Theoretical Models*	*Previous Studies*	*Frequency*
1. Reliability	9	14	23
2. Completeness	7	16	23
3. Accuracy	6	15	21
4. Timeliness	2	18	20
5. Format	1	9	10
6. Accessibility	2	8	10
7. Usability	6	2	8
8. Maintainability	7	0	7
9. Portability	6	0	6
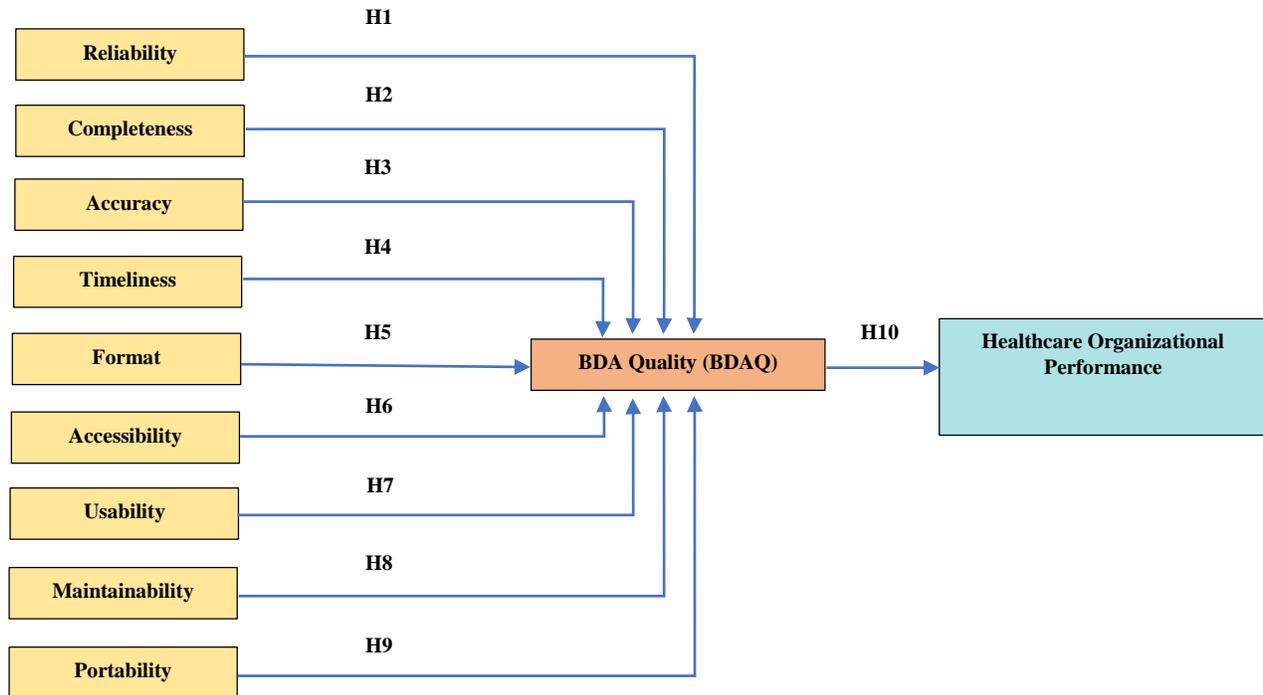
Fig. 3.   BDAQ Conceptual Model in Enhancing Healthcare Organizational Performance.

Fig. 3 illustrated the proposed BDAQ conceptual model with ten hypotheses. BDAQ will be specified as a higher-order or second-order construct that contains nine first-order formative constructs. The study will apply the repeated indicator approach to estimate the scores for the first-order BDAQ constructs [78].

H1: Reliability is significant BDAQ factor in enhancing healthcare organizational performance.

H2: Accuracy is significant BDAQ factor in enhancing healthcare organizational performance.

H3: Completeness is a significant BDAQ factor in enhancing healthcare organizational performance.

H4: Timeliness is a significant BDAQ factor in enhancing healthcare organizational performance.

H5: Format is a significant BDAQ factor in enhancing healthcare organizational performance.

H6: Accessibility is a significant BDAQ factor in enhancing healthcare organizational performance.

H7: Usability is a significant BDAQ factor in enhancing healthcare organizational performance.

H8: Maintainability is a significant BDAQ factor in enhancing healthcare organizational performance.

H9: Portability is a significant BDAQ factor in enhancing healthcare organizational performance.

H10: BDAQ has a significant positive effect on healthcare organizational performance.

V. CONCLUSION

A conceptual model was discovered to be useful in assessing BDA quality. The model may measure and analyze the relationship between reliability, completeness, accuracy, timeliness, format, accessibility, usability, maintainability, and portability as essential BDA quality factors that affect enhancing healthcare organizational performance. The subsequent research activity will involve the creation of a questionnaire-based survey instrument. The proposed conceptual model and surveys will subsequently be validated by an expert from academia, healthcare, and industry. In this regard, a pilot study will be done, which will be followed by the actual study. The model will then be tested using statistical tools, and the resulting insights will enable healthcare organizations to assess the essential BDA quality factors for ensuring that quality compliance is committed and fulfilled when implementing BDA as end-to-end solutions in healthcare organizations. Thus, will improve the efficiency and effectiveness of ICT infrastructure, operational, managerial, and strategic decision-making of healthcare organizations that affect enhance healthcare organizational performance.

REFERENCES

[1] Y. Wang, L. A. Kung, S. Gupta, and S. Ozdemir, "Leveraging Big Data Analytics to Improve Quality of Care in Healthcare Organizations: A Configurational Perspective," Br. J. Manag., vol. 30, no. 2, 2019.

[2] Y. Wang, L. A. Kung, W. Y. C. Wang, and C. G. Cegielski, "An integrated big data analytics-enabled transformation model: Application to health care," Inf. Manag., vol. 55, no. 1, pp. 64–79, 2018.

[3] Y. Wang and N. Hajli, "Exploring the path to big data analytics success in healthcare," J. Bus. Res., vol. 70, pp. 287–299, Jan. 2017.

[4] U. Awan, S. Shamim, Z. Khan, N. U. Zia, S. M. Shariq, and M. N. Khan, "Big data analytics capability and decision-making: The role of data-driven insight on circular economy performance," Technol. Forecast. Soc. Change, vol. 168, no. March, p. 120766, 2021.

[5] S. J. Alsunaidi et al., "Applications of big data analytics to control covid-19 pandemic," Sensors, vol. 21, no. 7, 2021.

[6] X. Zhang and S. Xiang, "Data Quality, Analytics, and Privacy in Big Data," in Big Data in Complex Systems, vol. 9, 2015, pp. 393–418.

[7] M. Wook et al., "Exploring big data traits and data quality dimensions for big data analytics application using partial least squares structural equation modelling," J. Big Data, vol. 8, no. 1, 2021.

[8] N. Mehta and A. Pandit, "Concurrence of big data analytics and healthcare: A systematic review," Int. J. Med. Inform., vol. 114, no. March, pp. 57–65, 2018.

[9] S. Fosso Wamba, S. Akter, and M. de Bourmont, "Quality dominant logic in big data analytics and firm performance," Bus. Process Manag. J., vol. 25, no. 3, pp. 512–532, Jun. 2019.

[10] N. Côrte-Real, P. Ruivo, and T. Oliveira, "Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value?," Inf. Manag., vol. 57, no. 1, Jan. 2020.

[11] G. A. Onyeabor and A. Ta'a, A model for addressing quality issues in big data, vol. 843. Springer International Publishing, 2019.

[12] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," Heal. Inf. Sci. Syst., vol. 2, no. 1, p. 3, Dec. 2014.

[13] Bridge, "The Cost of Healthcare Index," 2022. https://www.bridge patientportal.com/the-cost-of-healthcare.

[14] WHO, "Universal health coverage (UHC)," 2021. https://www.who.int/ news-room/fact-sheets/detail/universal-health-coverage-(uhc).

[15] W. H. DeLone and E. R. McLean, "The DeLone and McLean model of information systems success: A ten-year update," J. Manag. Inf. Syst., vol. 19, no. 4, pp. 9–30, 2003.

[16] R. Y. Wang and D. M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," J. Manag. Inf. Syst., vol. 12, no. 4, pp. 5–33, Mar. 1996.

[17] M. Wade and J. Hulland, "The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research," Source MIS Q., vol. 28, no. 1, pp. 107–142, 2004.

[18] "Barney - 1991 - Firm Resources and Sustained Competitive Advantage.pdf."

[19] C. Adrian, R. Abdullah, R. Atan, and Y. Y. Jusoh, "Conceptual Model Development of Big Data Analytics Implementation Assessment Effect on Decision-Making," Int. J. Interact. Multimed. Artif. Intell., vol. InPress, no. InPress, 2018.

[20] S. Ji-fan Ren, S. Fosso Wamba, S. Akter, R. Dubey, and S. J. Childe, "Modelling quality dynamics, business value and firm performance in a big data analytics environment," Int. J. Prod. Res., vol. 55, no. 17, pp. 5011–5026, Sep. 2017.

[21] S. Fosso Wamba, S. Akter, L. Trinchera, and M. De Bourmont, "Turning information quality into firm performance in the big data economy," Manag. Decis., vol. 57, no. 8, pp. 1756–1783, 2019.

[22] C. L. Chiu, H. C. Ho, T. Yu, Y. Liu, and Y. Mo, "Exploring information technology success of Augmented Reality Retail Applications in retail food chain," J. Retail. Consum. Serv., vol. 61, no. April, p. 102561, 2021.

[23] A. Wahyudi, G. Kuk, and M. Janssen, "A Process Pattern Model for Tackling and Improving Big Data Quality," Inf. Syst. Front., vol. 20, no. 3, 2018.

[24] M. Ghasemaghaei and G. Calic, "Can big data improve firm decision quality? The role of data quality and data diagnosticity," Decis. Support Syst., vol. 120, no. March, pp. 38–49, 2019.

[25] F. Sidi, P. H. Shariat Panahy, L. S. Affendey, M. A. Jabar, H. Ibrahim, and A. Mustapha, "Data quality: A survey of data quality dimensions," Proc. - 2012 Int. Conf. Inf. Retr. Knowl. Manag. CAMP'12, no. August, pp. 300–304, 2012.

[26] J. a. McCall, P. K. Richards, and G. F. Walters, "Factors in Software Quality," at'l Tech. Inf. Serv., vol. 1, 2 and 3, no. ADA049055, 1977.

[27] Boehm, B. W., Brown, J. R., Kaspar, H., Lipow, M., McLeod, G., and Merritt, M., Characteristics of Software Quality. New York: American Elsevier, 1978.

[28] A. Finandhita, "Development of Software Quality Assessment Model for Mobile-based Elderly Fall Detection Software," IOP Conf. Ser. Mater. Sci. Eng., vol. 879, no. 1, 2020.

[29] B. and C. S. Kitchenham, "Guidelines for performing Systematic Literature Reviews in Software Engineering," 2007.

[30] C. Okoli and K. Schabram, "A Guide to Conducting a Systematic Literature Review of Information Systems Research," SSRN Electron. J., vol. 10, no. 2010, 2010.

[31] R. G. Dromey, "A model for software product quality," IEEE Trans. Softw. Eng., vol. 21, no. 2, pp. 146–162, 1995.

[32] Evans M. W. and Marciniak J. J, Software Quality Assurance and Management. John Wiley & Sons, New York, 1987.

[33] Deutsch M. S. and Willis R. R, Software Quality Engineering: A Total Technical & Management Approach. Prentice-Hall, Englewood Cliffs, NJ, 1988.

[34] Grady R.B., Practical Software Metrics for Project Management and Process Improvement. Prentice Hall, Englewood Cliffs, USA, 1992.

[35] M. Barbacci, M. Klein, T. Logstaff, and C. Wenstock, "Technical Report CMU/SEI-95-TR-021 ESC-TR-95-021 Quality Attributes," Softw. Eng. Inst. - Carnegie Mellon Univ., no. December, 1995, [Online]. Available: http://oai.dtic.mil/oai/oai?verb=getRecord& metadataPrefix=html&identifier=ADA307888.

[36] ISO, "ISO/IEC 25010:2011 (SQuaRE) Quality Model," Geneva, Switzerland., 2011. [Online]. Available: https://www.iso.org/obp/ui/#i so:std:iso-iec:25010:ed-1:v1:en.

[37] A. I. Aljumah, M. T. Nuseir, and M. M. Alam, "Traditional marketing analytics, big data analytics and big data system quality and the success of new product development," Bus. Process Manag. J., 2021.

[38] O. Kwon, N. Lee, and B. Shin, "Data quality management, data usage experience and acquisition intention of big data analytics," Int. J. Inf. Manage., vol. 34, no. 3, pp. 387–394, Jun. 2014.

[39] S. Akter, S. Fosso Wamba, and S. Dewan, "Why PLS-SEM is suitable for complex modelling? An empirical illustration in big data analytics quality," Prod. Plan. Control, vol. 28, no. 11–12, pp. 1011–1021, 2017.

[40] P. H. Shariat Panahy, F. Sidi, L. S. Affendey, M. A. Jabar, H. Ibrahim, and A. Mustapha, "A framework to construct data quality dimensions relationships," Indian J. Sci. Technol., vol. 6, no. 5, pp. 4422–4431, 2013.

[41] M. M. Ayyash, "Scrutiny of relationship between E-banking information quality dimensions and customer satisfaction," J. Comput. Sci., vol. 13, no. 4, pp. 78–90, 2017.

[42] A. Ameen, D. Al-Ali, O. Isaac, and F. Mohammed, "Examining relationship between service quality, user satisfaction and performance impact in the context of smart government in UAE," Int. J. Electr. Comput. Eng., vol. 10, no. 6, pp. 6026–6033, 2020.

[43] Y. Zheng, K. Zhao, and A. Stylianou, "The impacts of information quality and system quality on users' continuance intention in information-exchange virtual communities: An empirical investigation," Decis. Support Syst., vol. 56, no. 1, pp. 513–524, 2013.

[44] A. Popovič, R. Hackney, P. S. Coelho, and J. Jaklič, "Towards business intelligence systems success: Effects of maturity and culture on analytical decision making," Decis. Support Syst., vol. 54, no. 1, pp. 729–739, 2012.

[45] N. W. Masri, J. J. You, A. Ruangkanjanases, S. C. Chen, and C. I. Pan, "Assessing the effects of information system quality and relationship quality on continuance intention in e-tourism," Int. J. Environ. Res. Public Health, vol. 17, no. 1, 2020.

[46] C. Gürkut and M. Nat, "Important factors affecting student information system quality and satisfaction," Eurasia J. Math. Sci. Technol. Educ., vol. 14, no. 3, pp. 923–932, 2018.

[47] N. Gorla, T. M. Somers, and B. Wong, "Organizational impact of system quality, information quality, and service quality," J. Strateg. Inf. Syst., vol. 19, no. 3, pp. 207–228, 2010.

[48] I. El Alaoui and Y. Gahi, "The impact of big data quality on sentiment analysis approaches," Procedia Comput. Sci., vol. 160, pp. 803–810, 2019.

[49] M. Bahrami and S. Shokouhyar, "The role of big data analytics capabilities in bolstering supply chain resilience and firm performance: a dynamic capability view," Inf. Technol. People, vol. ahead-of-p, no. ahead-of-print, 2021.

[50] A. F. Haryadi, J. Hulstijn, A. Wahyudi, H. Van Der Voort, and M. Janssen, "Antecedents of big data quality: An empirical examination in financial service organizations," Proc. - 2016 IEEE Int. Conf. Big Data, Big Data 2016, pp. 116–121, 2016.

[51] S. Verma, "the Adoption of Big Data Services By Manufacturing Firms: an Empirical Investigation in India," J. Inf. Syst. Technol. Manag., vol. 14, no. 1, pp. 39–68, 2017.

[52] P. Keikhosrokiani, N. Mustaffa, N. Zakaria, and R. Abdullah, "Assessment of a medical information system: the mediating role of use and user satisfaction on the success of human interaction with the mobile healthcare system (iHeart)," Cogn. Technol. Work, vol. 22, no. 2, pp. 281–305, 2020.

[53] S. S. Kamble, A. Gunasekaran, M. Goswami, and J. Manda, "A systematic perspective on the applications of big data analytics in healthcare management," Int. J. Healthc. Manag., vol. 12, no. 3, pp. 226–240, 2019.

[54] S. Akter, S. Fosso Wamba, and S. Dewan, "Why PLS-SEM is suitable for complex modelling? An empirical illustration in big data analytics quality," Prod. Plan. Control, vol. 28, no. 11–12, 2017.

[55] R. R. Nelson, P. A. Todd, and B. H. Wixom, "Antecedents of information and system quality: An empirical examination within the context of data warehousing," J. Manag. Inf. Syst., vol. 21, no. 4, pp. 199–235, 2005.

[56] D. Firmani, M. Mecella, M. Scannapieco, and C. Batini, "On the Meaningfulness of 'Big Data Quality' (Invited Paper)," Data Sci. Eng., vol. 1, no. 1, pp. 6–20, 2016.

[57] I. Taleb, M. A. Serhani, and R. Dssouli, "Big Data Quality: A Survey," Proc. - 2018 IEEE Int. Congr. Big Data, BigData Congr. 2018 - Part 2018 IEEE World Congr. Serv., pp. 166–173, 2018.

[58] M. Janssen, H. van der Voort, and A. Wahyudi, "Factors influencing big data decision-making quality," J. Bus. Res., vol. 70, pp. 338–345, 2017.

[59] J. M. Tien, "Big Data: Unleashing information," J. Syst. Sci. Syst. Eng., vol. 22, no. 2, pp. 127–151, 2013.

[60] G. Shah, A. Shah, and M. Shah, "Panacea of challenges in real-world application of big data analytics in healthcare sector," J. Data, Inf. Manag., vol. 1, no. 3–4, pp. 107–116, 2019.

[61] J. Merino, X. Xie, A. K. Parlikad, I. Lewis, and D. McFarlane, "Impact of data quality in real-time big data systems," CEUR Workshop Proc., vol. 2716, pp. 73–86, 2020.

[62] M. Ambigavathi and D. Sridharan, "Big Data Analytics in Healthcare," in 2018 Tenth International Conference on Advanced Computing (ICoAC), Dec. 2018, pp. 269–276.

[63] N. Côrte-Real, P. Ruivo, and T. Oliveira, "Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value?," Inf. Manag., vol. 57, no. 1, p. 103141, Jan. 2020.

[64] P. Kamakshi, "Importance of Big data in Healthcare System-A Survey," Int. J. Appl. Eng. Res., vol. 13, no. 15, pp. 12184–12187, 2018, [Online]. Available: http://www.ripublication.com.

[65] D. Rao, V. N. Gudivada, and V. V. Raghavan, "Data quality issues in big data," Proc. - 2015 IEEE Int. Conf. Big Data, IEEE Big Data 2015, pp. 2654–2660, 2015.

[66] M. Halaweh, "Conceptual Model for Successful Implementation of Big Data in Organizations," J. Int. Technol. Inf. Manag., vol. 24, no. 2, pp. 21–29, 2015.

[67] R. Raja, I. Mukherjee, and B. K. Sarkar, "A Systematic Review of Healthcare Big Data," Sci. Program., vol. 2020, pp. 1–15, Jul. 2020.

[68] S. Ji-fan Ren, S. Fosso Wamba, S. Akter, R. Dubey, and S. J. Childe, "Modelling quality dynamics, business value and firm performance in a big data analytics environment," Int. J. Prod. Res., vol. 55, no. 17, 2017.

[69] C. Adrian, R. Abdullah, R. Atan, and Y. Y. Jusoh, "Factors influencing to the implementation success of big data analytics: A systematic literature review," in 2017 International Conference on Research and Innovation in Information Systems (ICRIIS), Jul. 2017, pp. 1–6.

[70] A. Rehman, S. Naz, and I. Razzak, "Leveraging Big Data Analytics in Healthcare Enhancement: Trends, Challenges and Opportunities," Multimed. Syst., Apr. 2020.

[71] S. Hoffman, "Medical Big Data and Big Data Quality Problems," SSRN Electron. J., 2014.

[72] S. Imran, T. Mahmood, A. Morshed, and T. Sellis, "Big data analytics in healthcare A systematic literature review and roadmap for practical implementation," IEEE/CAA J. Autom. Sin., vol. 8, no. 1, pp. 1–22, Jan. 2021.

[73] F. L. Urera and F. F. Balahadia, "ICTeachMUPO: An Evaluation of Information E-Learning Module System for Faculty and Students An evaluation of information e-learning module system for faculty and students," Int. J. Comput. Sci. Res., vol. 3, no. 1, pp. 163–188, 2019.

[74] International Organization For Standardization Iso, "Iso/Iec 25010:2011," Softw. Process Improv. Pract., vol. 2, no. Resolution 937, 2011.

[75] A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, "Big data analytics in healthcare," Biomed Res. Int., vol. 2015, 2015.

[76] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," J. Big Data, vol. 6, no. 1, p. 54, Dec. 2019.

[77] J. Singh and N. B. Kassie, "User's Perspective of Software Quality," Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2018, no. July 2020, pp. 1958–1963, 2018.

[78] M. Wetzels, G. Odekerken-Schröder, and C. Van Oppen, "Assessing Using PLS Path Modeling Hierarchical and Empirical Construct Models : Guidelines," MIS Q., vol. 33, no. 1, pp. 177–195, 2009.

Parkinson's Disease Identification using Deep Neural Network with RESNET50

Anila M[1]

Research Scholar, Department of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram, AP, India

Pradeepini Gera[2]

Professor, Department of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram, AP, India

Abstract—**Recent Parkinson's disease (PD) research has focused on recognizing vocal defects from people's prolonged vowel phonations or running speech since 90% of Parkinson's patients demonstrate vocal dysfunction in the early stages of the illness. This research provides a hybrid analysis of time and frequency and deep learning techniques for PD signal categorization based on ResNet50. The recommended strategy eliminates manual procedures to perform feature extraction in machine learning. 2D time-frequency graphs give frequency and energy information while retaining PD morphology. The method transforms 1D PD recordings into 2D time-frequency diagrams using hybrid HT/Wigner-Ville distribution (WVD). We obtained 91.04% accuracy in five-fold cross-validation and 86.86% in testing using RESNET50. F1-score achieved 0.89186. The suggested approach is more accurate than state-of-the-art models.**

Keywords—Parkinson's disease; speech impairment; artificial intelligence; RESNET50; deep learning; ht/wigner-ville distribution; 2D time-frequency

I. INTRODUCTION

Parkinson's disease (PD) affects 1% of persons over 60 [1]. Parkinson's disease is the second most prevalent neurological condition [2]. According to European and American epidemiological studies [3], PD affects men 1.5 times more than women. Parkinson's disease causes increasing disability that affects everyday life. Dopamine neurodegeneration induces motor and non-motor symptoms in the midbrain's substantia nigra pars compacta [3].

Early on with Parkinson's, voice problems affect 90% of individuals. Recent PD telemedicine research [4-6] focused on speech problems. These studies employed speech signal processing to evaluate Parkinson's disease. The generated characteristics were used to build reliable decision support systems. Parkinson's disease causes rigidity, resting tremors, bradykinesia, and postural instability [7, 8].

Many research that compares PD and healthy persons utilize a publicly accessible dataset [9] with 195 sound measures from 23 PD patients and eight healthy controls. Another publicly accessible PD telediagnosis dataset [6] utilized in the linked research includes 20 PD and 20 healthy participants' voice recordings. In all datasets, each voice recording contains fundamental vocal frequency, amplitude variation measurements, noise-to-tone component ratio, nonlinear dynamical complexity measures, and nonlinear fundamental frequency variation measures. With these parameters achieved accuracy is less than 85%, which is too less. Using deep learning methods, it can be improved.

The remainder of this research is organized as follows: the research analyzes existing automated speech analysis techniques in Parkinson's disease patients in the section "Related Work." It discusses the 1D to 2D transformation methods and the deep learning classification model in the section "Methodology." The section "Results and Discussion" addresses the employed dataset, classification performance, and statistical analysis results, while the section "Conclusions and Future Study" propose conclusions and future work.

II. RELATED WORK

PD speech analysis research includes recording activities. Sustained vowel phonation is prevalent since it is a regular work [10]. Other research focuses on continuous speech recordings of sentences, reading texts, and spontaneous speech to study prosody [11]. Few papers discussed single-word invention. Early research focused on Parkinson's patients [12]. This study examined spectral and cepstral properties taken from 24 isolated words and five vowels spoken by Colombian Spanish speakers. Scientists used an SVM with a Gaussian kernel to classify Parkinson's patients. They compared each set of attributes individually and all coefficients together. When all utterances and attributes were combined, word and vowel accuracy was 92% and 79%. Despite positive findings, the approach lacked pre-processing.

In [13] mentioned lonely words as a speaking difficulty. The database includes Spanish (50 HC, 50 PD), German (88 PD, 88 HC), and Czech speakers (20 PD, 16 HC). Different languages and feature sets were automatically used to classify HC and PD speakers. Each corpus was modeled using four factors to identify linguistic impairment. Using a radial basis SVM, the authors claimed the latter technique was robust, with 85 to 95% classification accuracy. Both papers used PC-GITA. None studied model generalization on a single dataset. The test models in [13] were optimized, resulting in unduly optimistic findings.

In [14], to explain speech signal non-linearities, the authors proposed isolated word modeling based on Hilbert Spectrum properties. The recommended coefficients, Instantaneous Energy Deviation Coefficients (IEDCC), exceeded the traditional acoustic characteristics when dealing with isolated individual syllables, outperforming them with

accuracy ranging from 81 to 91%. The accuracy was 82%. However, the authors did not give the findings of the combined features. They used 20 Parkinson's patients and 20 healthy controls.

Little et al. [15] reported that each patient made six phonations. A soundproof audiology booth created by an industrial acoustics business caught the phonations. Human sound signals were recorded in a computerized speech laboratory. After recording phonations, they analyzed entropy and conventional and nonstandard measures. A kernel SVM classifier distinguished between healthy and Parkinson's patients. Researchers then used the Little et al. dataset to enhance their classifier. Bhattacharya et al. employed Weka and SVM [16]. Das employed neural networks, DMneural, regression, and decision trees to classify features [17]. Sakar et al. [18] employed SVM and feature selection. Polat identified Parkinson's disease patients using FCMFW and k-NN [19].

Sakar and Kursun [18] suggested an SVM model using mutual knowledge for PD feature selection. Speech characteristics were filtered and sorted using the highest relevance lowest redundancy approach (mRMR). Using the leave-one-individual-out strategy, their test on the UCI dataset obtained 81.53% accuracy and 92.75% using the bootstrap resampling method. It discovered four critical parameters associated with the voice impairment test for Parkinson's disease diagnosis, which additional studies may corroborate.

This paper proposes an approach for analyzing solitary words based on several signal processing and pattern recognition methods. We began with a pre-processing phase, followed by a 1D to 2D time-frequency conversion approach and a classification step. The PD sound series were converted into 2D time-frequency distribution diagrams to display additional information such as time, frequency, and energy. To translate 1D to 2D time-frequency distribution diagrams, a technique combining HT and WVD was presented. The hybrid analysis of the time and frequency flowchart is shown in Fig. 2. The approach first employs the HT to transform the original signal into analytic signals, followed by the analysis of time and frequency using the traditional WVD to generate the relevant time-frequency diagrams.

III. METHODOLOGY

The importance of implementation is to improve the recognition rate.

A. Method Outline

The 1D sound signal was reconstructed as 2D time-frequency diagrams in this work to gather information on the signal's time, frequency, and energy. The suggested method's flowchart is shown in Fig. 1. During data pre-processing, noises like baseline wander (BW) and 50/60 Hz power line noise are initially removed. To convert the 1D time series into 2D time and frequency diagrams, the HT-WVD technique is used. The dataset was randomly partitioned into three sets: training, validation, and testing. 70% of each category was chosen randomly for the training, 10% for the validation, and 20% for the testing.

The validation data is used to validate the performance of the learned deep learning models using a five-fold cross-validation technique. The training data is used to train multiple deep-learning models with a learning rate of 0.001. The ResNet50 was utilized for deep learning to classify the 2D time and frequency diagrams of the PD for classification tasks. The classification outcomes for the test data for the optimal neural network model are shown.

B. Data Pre-processing

1) Removing noises: Our database of PD recordings comprises familiar sounds such as 50/60 Hz powerline noise and BW. As a result, noise should be deleted from the recordings to ensure accuracy. This research used a median filter approach to remove BW from PD recordings first, followed by a wavelet transform algorithm to remove powerline noise. The PD recordings' baseline curves were retrieved using the median filtering technique. This entails running a median filtering algorithm [20], then subtracting the original PD recordings from the derived baseline maps to generate the new PD recordings without BW. This research employed the wavelet transform to remove additional noise. The Daubechies db5 wavelet is used to create a three-level wavelet decomposition.
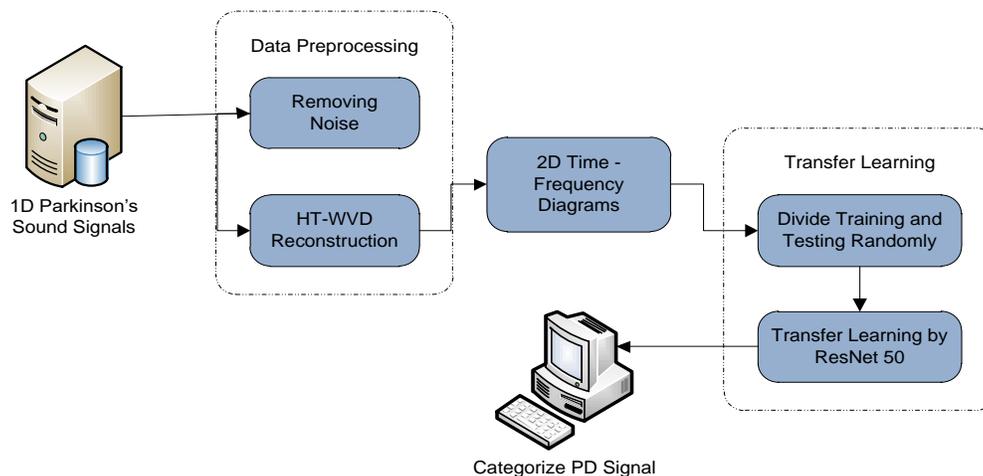
Fig. 1. The Proposed Method of Architecture.

2) Hybrid analysis of time and frequency: Because PD sound signals are 1D non-stationary time series, they cannot display the connection between time, frequency, and energy within the time series. Analysis of time and frequency techniques may convert a one-dimensional signal into a two-dimensional density function of time and frequency, revealing how many frequency components are present in the signal and how each component changes over time. Temporal frequency approaches such as the WVD and HT are widely employed for signal analysis [24].

3) One can utilize the bilinear Wigner-Ville distribution to evaluate nonlinear signals and reflect energy distributions in both the time and frequency domains. The definition of WVD is as follows:

$$W(t,f) = \int_{+\infty}^{-\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) d\tau \qquad (1)$$

4) X* is a complex conjugate, and x(t) is the signal. The study [21] explains every step of the procedure. An actual signal can be transformed into an analytical signal using the HT. The HT is described as follows for a discrete time series x(k):

$$H(k) = xH(k) = FFT\text{-}1(f(k) * h(i)) \qquad (2)$$

where the element-wise product of f and h is calculated using the symbol *. The inverse fast Fourier transform, or FFT-1, is represented by vector f, which is the FFT of the y(k), and h is defined as follows:

$$h = \begin{cases} 0 \ for \ i = (N/2) + 2, \ldots, N \\ l \ for \ i = 1, \ (N/2) + 1 \\ 2 \ for \ i = 2, 3, \ldots, (N/2) \end{cases} \qquad (3)$$

Thus, the analytic signal z(k) of the discrete time series x(k) can be represented as

$$z(k) = x(k) + jxH(k) \qquad (4)$$

5) Where j is the complex imaginary number, i.e., j = (-1)1/2, the HT is detailed in [22].

6) Because the dimensional difference of 1D signals does not affect the energy distribution of 2D time and frequency diagrams, 1D PD signals do not need to be normalized when translated into 2D time-frequency diagrams.

7) Fig. 3 to Fig. 6 shows the signal in the time domain, the signal spectrum, and the signal's spectrogram, which is the 2D time frequency of Parkinson's audio signal for vowels' a' and 'i' for the healthy and Parkinson's disease signal.
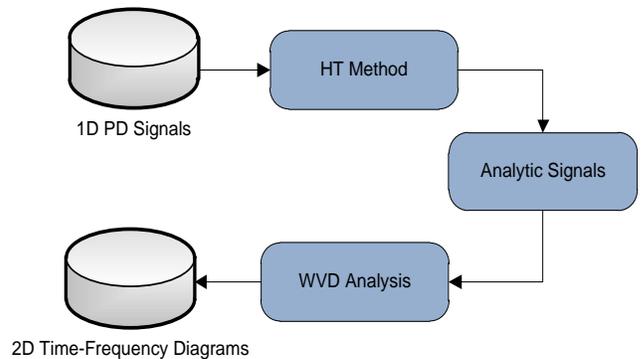
Fig. 2. The Flowchart of the Hybrid Time and Frequency Algorithm.
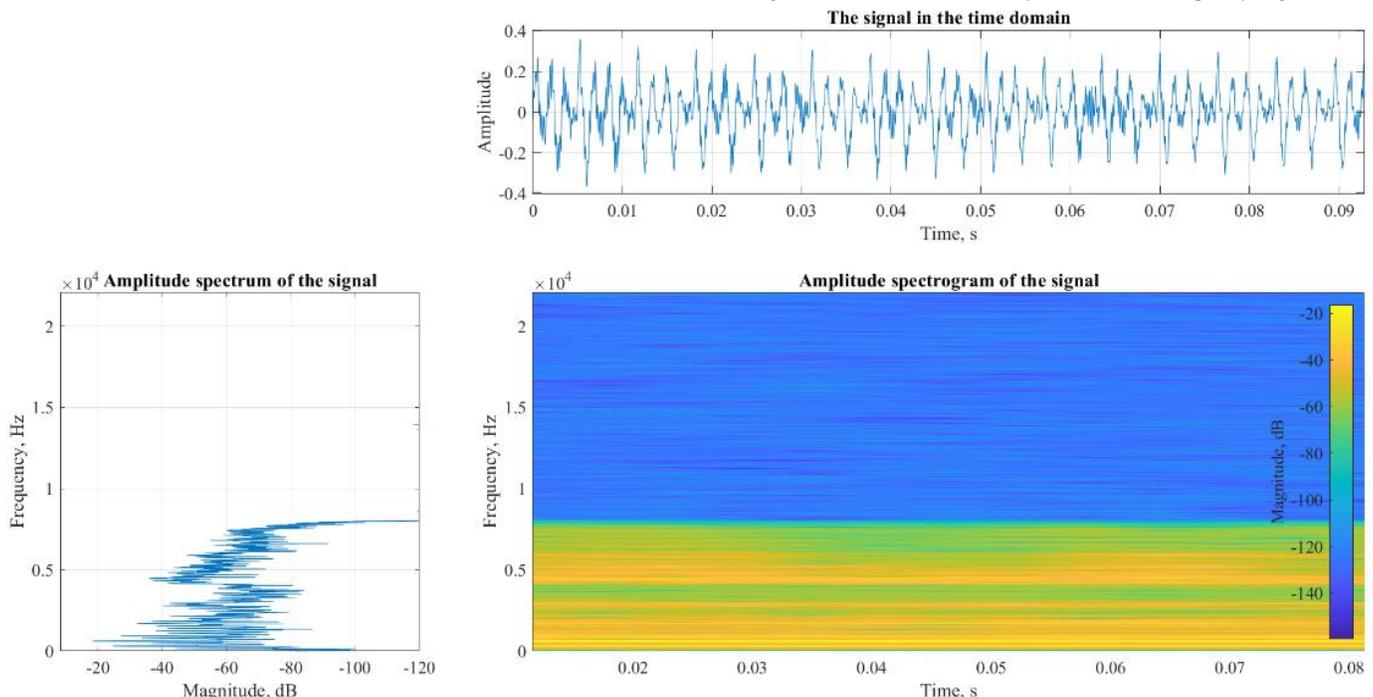
Fig. 3. 2D Time-Frequency Diagrams of PD Signals for Vowels 'A'.
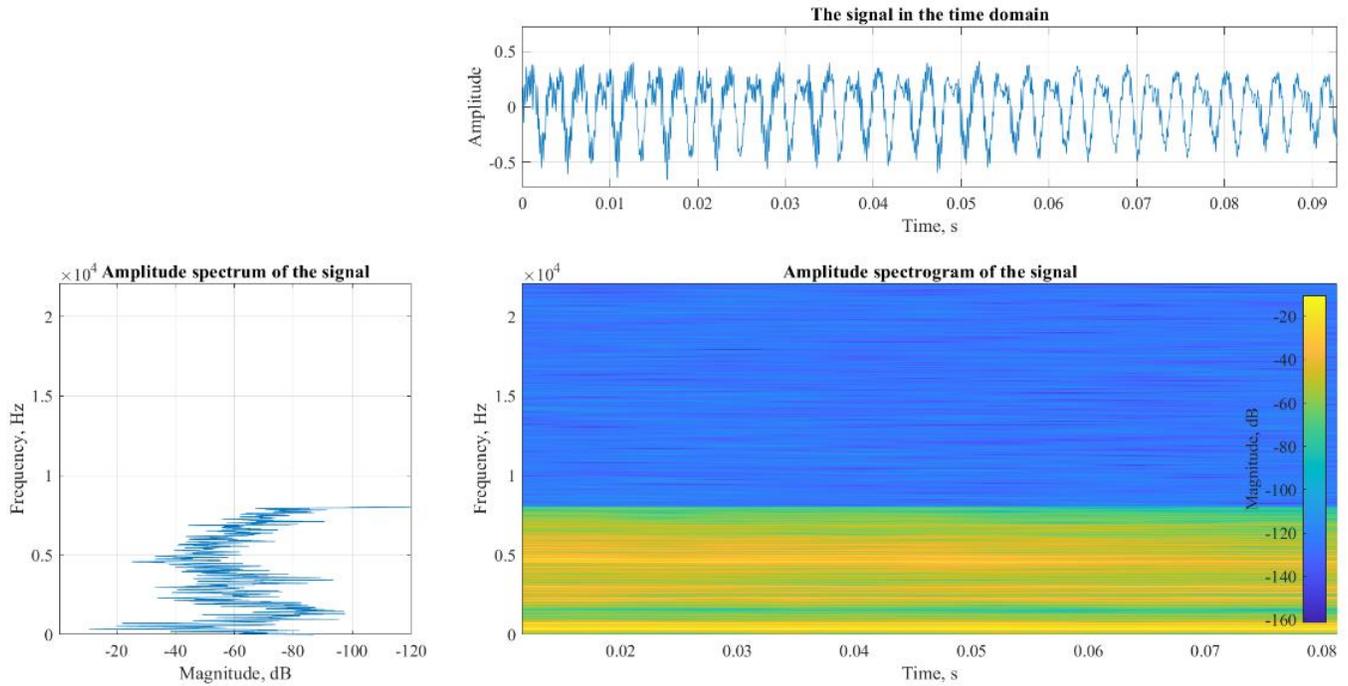
Fig. 4.   2D Time-Frequency Diagrams of PD Signals for Vowels 'I'.
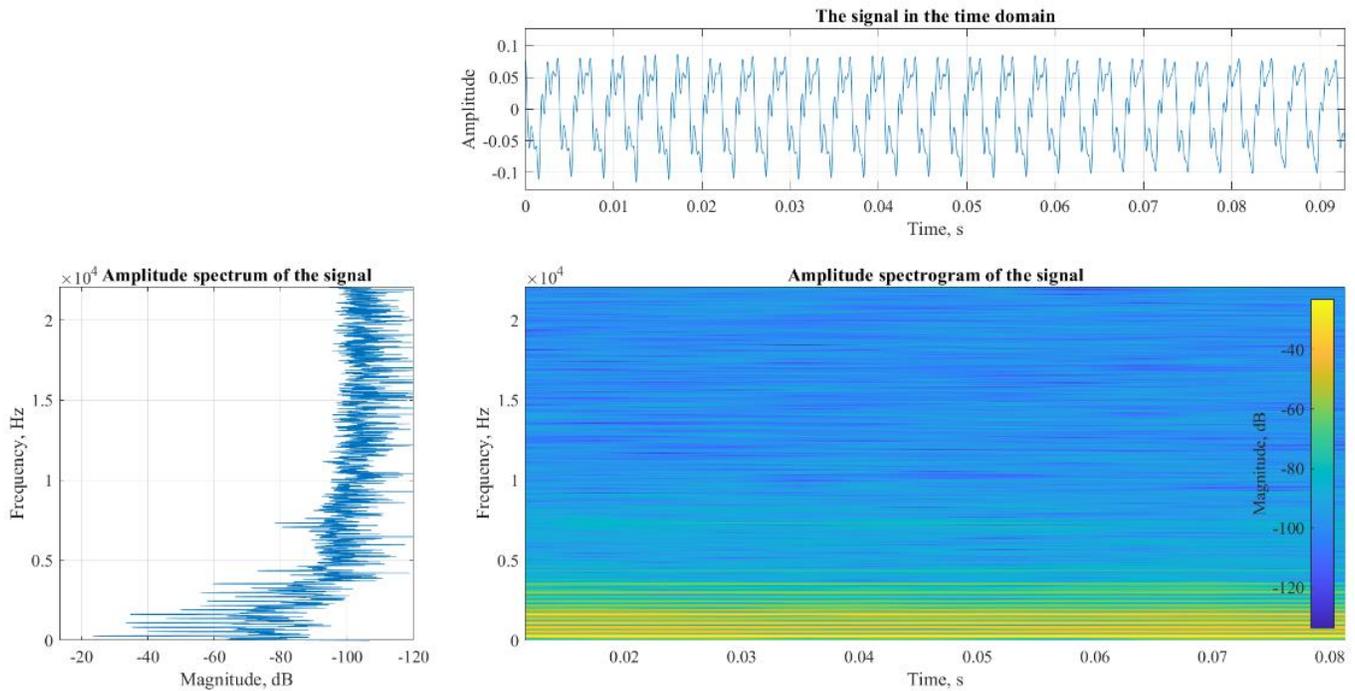
Fig. 5.   2D Time-Frequency Diagrams of HC Signals for Vowels 'A'.
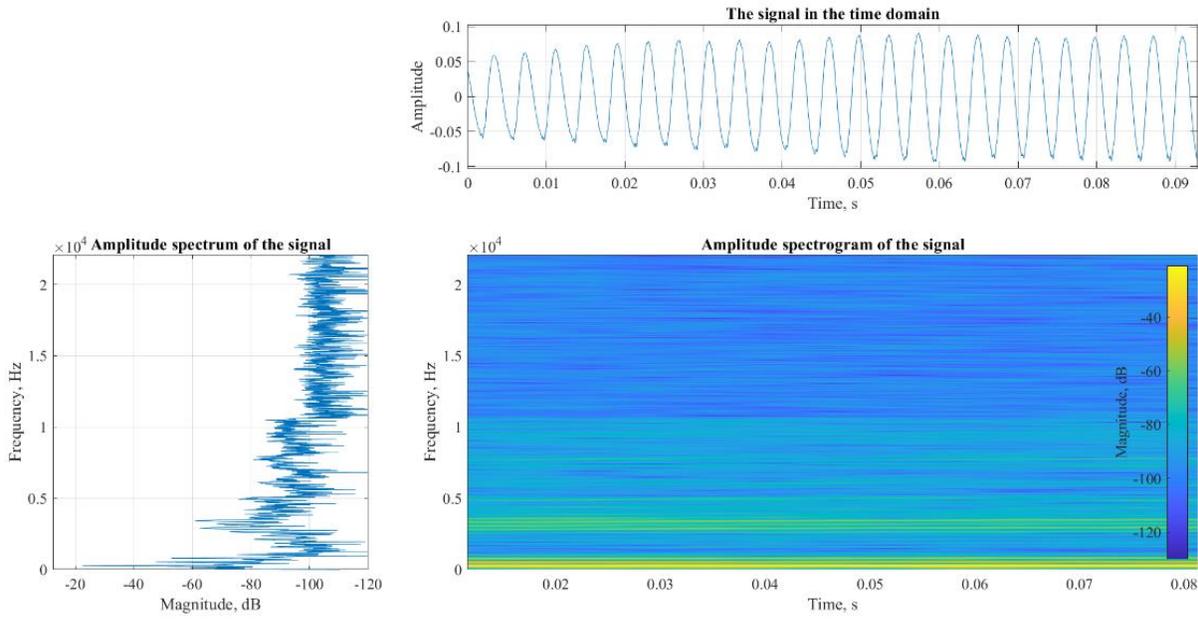
Fig. 6.   2D Time-Frequency Diagrams of HC Signals for Vowels 'I'

C. ResNet50 & Transfer Learning

He et al. [23] presented ResNet at the 2015 ILSVRC competition. ResNet includes a short connection structure that avoids gradient vanishing by immediately bypassing the input information to the output. ResNet is a network-in-network (NIN) architecture comprised of several residual modules stacked on top of one another. These residual units are employed in the construction of deep ResNet architecture. Convolution, pooling, and layering are the residual units. Instead of wholly linked layers, ResNet employs global average pooling.

The ResNet uses residual units to prevent the gradient from disappearing [25]. Let the input and output vectors for the layers be x and y, respectively, and $H(x)$ function as nonlinear stacked layers. It is possible to obtain $H(x) = F(x) + x$ by using the formula $F(x) = H(x) - x$, where F(x) and x represent stacked nonlinear layers. The residual block's typical structure is described in (5).

$$y = F(x, \{Wi\}) + x \tag{5}$$

Wi is the weight in the weight matrix, and F is the residual mapping to be learned.

The ImageNet dataset was used to train the conventional pre-trained ResNet50. 2D time and frequency diagrams made from PD recordings were the images to be analyzed in this work; these diagrams differed significantly from the photographs in ImageNet. To enable the modified ResNet50 to classify two PD classes, the last three layers of the pre-trained ResNet50 - the fully connected layer, the Softmax layer, and the classification layer - were altered. We used the same dense layer network after the convolutional layers, as shown in Fig. 7. After the convolutional layer and every block, the rectified linear unit (ReLU) is activated by doing a max-pooling operation. To make the spatial dimension of the activation map half the preceding layer, a max pooling layer with a 2x2 kernel and stride size of two is utilized. The improved ResNet50 was retrained to provide new parameters, and Table I displays the changed ResNet50's structure. The learning rate of deep learning based on ResNet50 was set at 0.001 in this research.
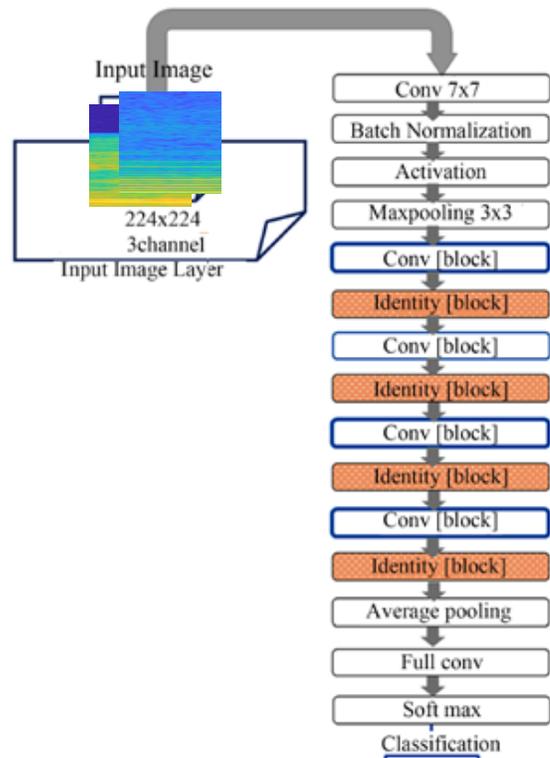
Fig. 7.   ResNet50 Architecture for Proposed Model.

TABLE I.   NETWORK PARAMETER CONFIGURATION OF RESNET50

Layer Name	Output size	Layers	
Conv1	224x224	7x7, 64, stride 2	
Conv2	112x112	3x3 Max pool, stride 2	
		1x1, 64	X3
		3x3, 64	
		1x1, 256	
Conv3	56x56	1x1, 128	X4
		3x3, 128	
		1x1, 512	
Conv4	28x28	1x1, 256	X6
		3x3, 256	
		1x1, 1024	
Conv5	14x14	1x1, 512	X3
		3x3, 512	
		1x1, 2048	
	1x1	Average Pool	
		2-d Fully Connected	
		SoftMax	

IV. RESULTS AND DISCUSSION

A. Dataset Description

This research gathered speech recordings from 20 people (10 Parkinson's patients and 10 healthy controls). The data for the research was acquired from 188 individuals with the disease (5 men and 5 women) aged 33 to 87 (65.110.9). The control group comprises 10 healthy people (5 men and 5 women) aged 41 to 82 (61.18). The microphone was adjusted to 44.1 KHz throughout the data collection procedure. After the physician's examination, the sustained phonation of the vowels/a/and/i/was gathered from each participant three times. Fig. 8 shows the sample 2D time-frequency diagrams used in the analysis.

B. Experimental Setup

A 64-bit Windows 10 with an Intel Core i7 (2.80 GHz, 2808 MHz, 4 Cores, 8 Logical processors), 16 GB of RAM, an NVIDIA GeForce GTX, and CUDA 9.0 is utilized. The experiments were carried out using the Matlab (R2019a) programming language.
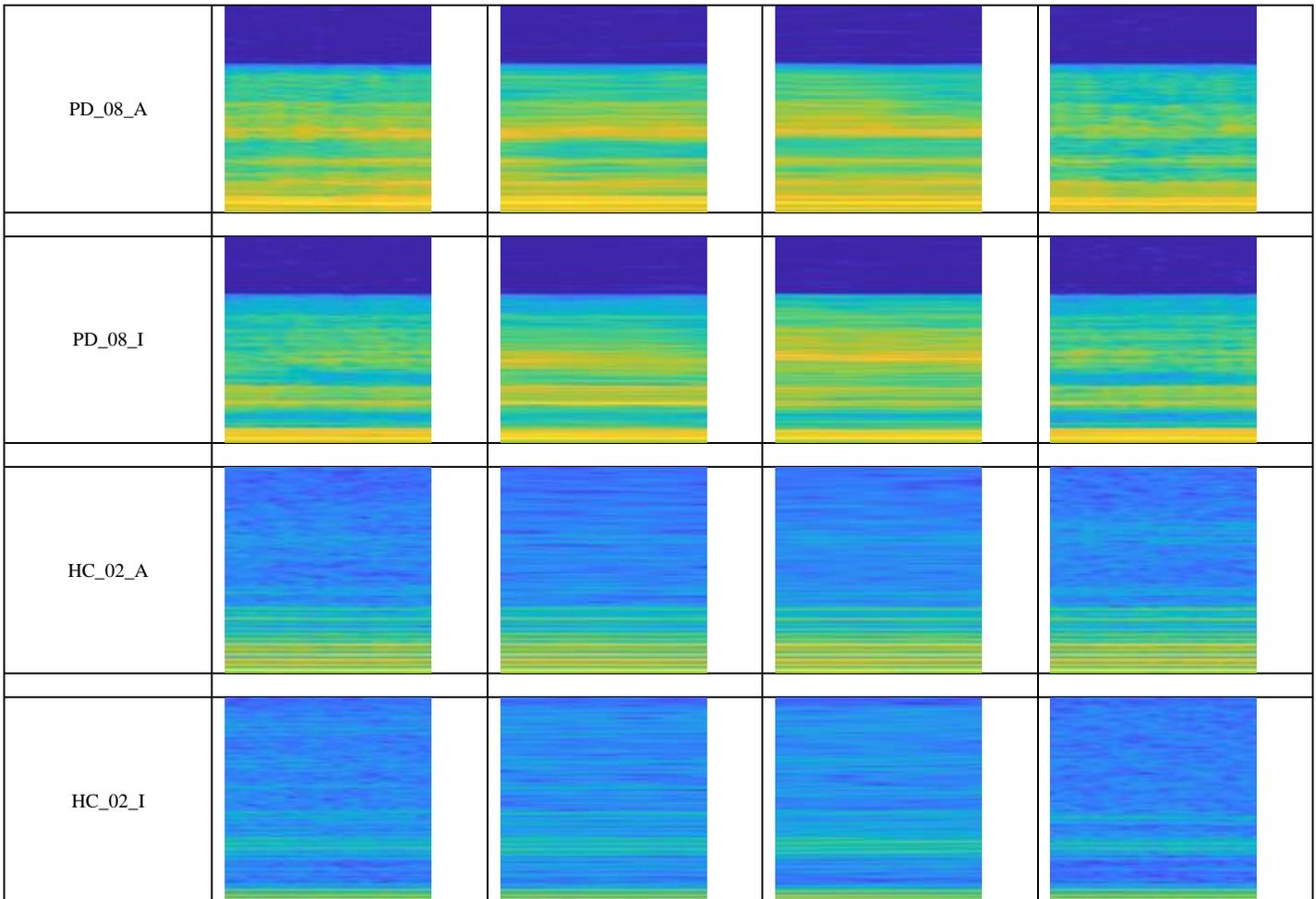
Fig. 8.   Sample 2D Time-Frequency Diagrams used for Analysis.

C. Evaluation Metrics

A confusion matrix was utilized in this research to summarize the categorization findings based on the actual and predicted categories. The construction of the confusion matrix is seen in Fig. 9. As a consequence, the numbers on the diagonal display the accurate categorization results, while the other values display the wrong findings.
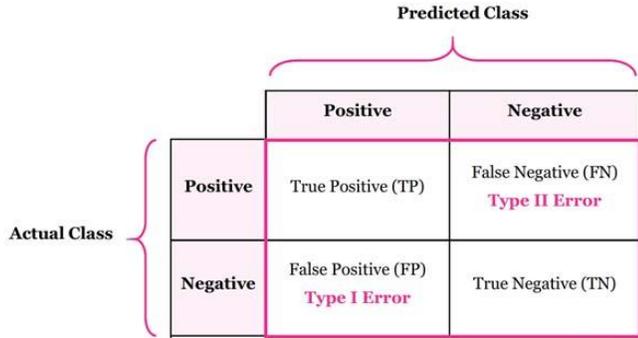
Fig. 9. Confusion Matrix.

The F1 score was also used in this research to evaluate the performance of the suggested technique. The F1 score may be viewed as a weighted average of precision and recall, with a most incredible score of 1 and the worst score of 0.

$$F1_{Score} = \frac{Tp}{Tp + \frac{1}{2}(Fp + Fn)} \qquad (6)$$

In addition, four statistical indices were produced to assess the performance of the proposed classifier in this work, namely accuracy, Precision, sensitivity, and specificity were defined as follows;

$$Accuracy = \frac{Tp + Tn}{Tp + Fp + Tn + Fn} \qquad (7)$$

$$Precision = \frac{Tp}{Tp + Fp} \qquad (8)$$

$$Sensitivity = \frac{Tp}{Tp + Fn} \qquad (9)$$

$$Specificity = \frac{Tn}{Tn + Fp} \qquad (10)$$

Where the (Tp) is the number of PD signals correctly classified as the PD category, the (Fn) is the number of signals of the PD category incorrectly classified as the HC category. The (Tn) is the number of signals of the HC categories not classified as the PD category. The (Fp) is the number of signals incorrectly classified as the PD category in the HC categories.

D. Classification Results

Table II displays the classification results for the test data using the proposed classifier trained at a learning rate of 0.001. Table II demonstrates that the classifier can get the best results: accuracy 86.86%, precision 89.19%, sensitivity 89.17%, and specificity 84.67%. Moreover, the F1-Score is 89.18%.

Fig. 10 shows the system-generated confusion matrix. Moreover, Fig. 11 shows the receiver operating curve of the system. ROC is the plot of TPR vs. FPR. The Area under the Curve for the PD and HC categories is 93.7%, which is comparatively reasonable.

Table III shows the state-of-the-art results compared with the proposed classification results. From the above comparison, we proved that the proposed method is robust in PD recognition using vowels.

TABLE II. CLASSIFICATION RESULTS

Metrics	Results
Accuracy	86.86
Precision	89.19
Sensitivity	89.17
Specificity	84.67
F1 Score	89.18
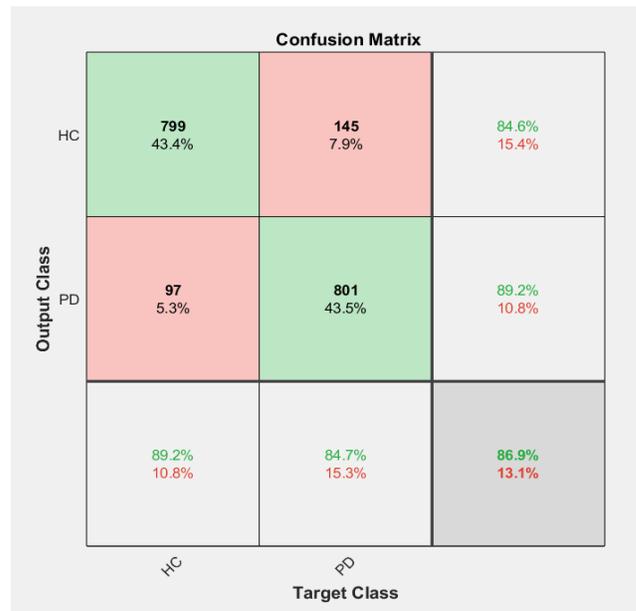
Fig. 10. Confusion Matrix shows 86.9% of Accuracy.
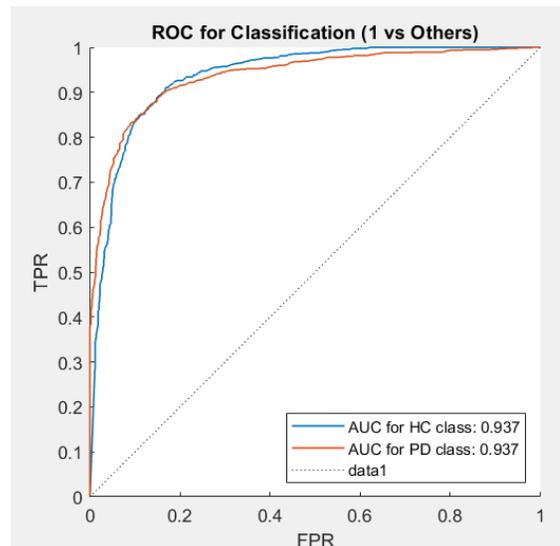
Fig. 11. ROC for PD and HC Categories.

TABLE III. COMPARISON RESULTS WITH PREVIOUS METHODS

Previous Methods	Accuracy (%)
Existing Method (Ref 12)	79.00
Existing Method (Ref 13)	85.00
Existing Method (Ref 14)	82.00
Existing Method (Ref 18)	81.53
Proposed Method	88.86

V. CONCLUSIONS AND FUTURE WORK

The HT-WVD and the deep learning based on ResNet50 are both modules that make up the recommended approach in this study. The HT-WVD removes cross terms to depict the relationship between time, frequency, and energy while keeping time-domain qualities. It directs energy towards the peak (i.e., amplitude and phase within the PD records). In addition, the depth of deep learning based on ResNet50 is adequate to leverage deep neural network performance more extensively than in past studies. Conversely, the suggested technique is more advanced than the existing 1D CNN, and the 2D ResNet model is computationally expensive.

This study proved the feasibility of a speech-based PD classification while proposing new potential approaches for vocal feature analysis. The highest, 86.86% accuracy, was achieved using 5-fold cross-validation. In a clinic, new patients with Parkinson's disease may be identified from healthy persons using the blind test interface. The suggested technique performs well in categorizing PD voice signals compared to earlier research.

REFERENCES

[1] Samii A, Nutt JG, Ransom BR. Parkinson's disease. Lancet. 2004;363:1783‑93. https:// doi. org/ 10. 1016/ S0140-6736(04) 16305-8.

[2] M. C. de Rijk, L. J. Launer, K. Berger, M. M. Breteler, J. F. Dartigues, M. Baldereschi, L. Fratiglioni, A. Lobo, J. Martinez-Lage, C. Trenkwalder, and A. Hofman, "Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts," Neurology, vol. 54, pp. 21–23, 2000.

[3] Massano J, Bhatia KP. Clinical approach to Parkinson's disease: features, diagnosis, and management principles. Cold Spring Harbor Perspect Med. 2012;2(6):8870. https:// doi. Org/ 10. 1101/ cshpe rspect. a0088 70.

[4] B. E. Sakar, M. Isenkul, C. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings," IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 4, pp. 828-834, 2013.

[5] C. O. Sakar and O. Kursun, "Telediagnosis of Parkinson's Disease Using Measurements of Dysphonia," Journal of Medical Systems, vol. 34, no. 4, pp. 591-599, 2009.

[6] A. Tsanas, M. Little, P. McSharry and L. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," Journal of the Royal Society Interface, vol. 8, no. 59, pp. 842-855, 2010.

[7] Gray P, Hildebrand K. Fall risk factors in Parkinson's disease. J Neurosci Nurs. 2000;32:222. https:// doi. org/ 10. 1097/ 01376 517-20000 8000-00006.

[8] Gunduz H. Deep learning-based Parkinson's disease classification using vocal feature sets. IEEE Access. 2019;7:115540‑51. https:// doi. org/ 10. 1109/ access. 2019. 29365 64.

[9] M. A. Little, P. Mcsharry, E. Hunter, J. Spielman, and L. Ramig, "Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease," IEEE Transactions on Biomedical Engineering, vol. 56, no. 4, pp. 1015–1022, 2009.

[10] Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng. 2009;56(4):1015‑22. https:// doi. org/ 10. 1109/ TBME. 2008. 20059 54.

[11] Hlavnika J, Cmejla R, Tykalova T, Šonka K, Ruzicka E, Rusz J. Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behavior disorder. Sci Rep. 2017;7(1):1‑13. https:// doi. org/ 10. 1038/ s41598-017-00047-5.

[12] Orozco-Arroyave JR, Honig F, Arias-Londono JD, Vargas-Bonilla JF, Noth E. Spectral and cepstral analyses for Parkinson's disease detection in Spanish vowels and words. Expert Syst. 2015;32(6):688‑97. https:// doi. org/ 10. 1111/ easy. 12106.

[13] Orozco-Arroyave JR, Honig F, Arias-Londono JD, Vargas-Bonilla JF, Daqrouq K, Skodda S, Rusz J, Noth E. Automatic detection of Parkinson's disease in running speech spoken in three different languages. J Acoust Soc Am. 2016;138:481‑500. https:// doi. org/ 10. 1121/1. 49397 39.

[14] Karan B, Sahu SS, Orozco-Arroyave JR, Mahto K. Hilbert spectrum analysis for automatic detection and evaluation of Parkinson's speech. Biomed Signal Processing Control. 2020;61:102050. https:// doi. org/ 10. 1016/j. bspc. 2020. 102018.

[15] Little, M.A.; McSharry, P.E.; Hunter, E.J.; Ramig, L.O. Suitability of dysphonia measurements for telemonitoring Parkinson's disease. IEEE Trans. Biomed. Eng. 2009, 56, 1015–1022. [CrossRef] [PubMed].

[16] Bhattacharya, I.; Bhatia, M.P.S. SVM Classification to Distinguish Parkinson Disease Patients. In Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing, Coimbatore, India, 16–17 September 2010; pp. 1–6.

[17] Das, R. A Comparison of Multiple Classification Methods for Diagnosis of Parkinson's Disease. Expert Syst. Appl. 2010, 37, 1568–1572. [CrossRef].

[18] Sakar, C.O.; Kursun, O. Telediagnosis of Parkinson's Disease Using Measurements of Dysphonia. J. Med. Syst. 2010, 34, 591–599. [CrossRef] [PubMed].

[19] Polat, K. Classification of Parkinson's Disease Using Feature Weighting Method based on Fuzzy C-Means Clustering. Int. J. Syst. Sci. 2011, 43, 597–609. [CrossRef].

[20] E. Ataman, V. Aatre, and K. Wong, "A fast method for real-time median filtering," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, no. 4, pp. 415-421, 1980.

[21] S. S. Qurraie, and R. G. Afkhami, "ECG arrhythmia classification using time-frequency distribution techniques," Biomed. Eng. Let., vol. 7, no. 4, pp. 325-332, 2017.

[22] A. Ramos, A. Lazaro, D. Girbau, R. Villarino, "Chipless Time-coded UWB RFID: Reader, Signal Processing and Tag Design," in RFID and Wireless Sensors Using Ultra-Wideband Technology, London, UK: ISTE Press-Elsevier, 2016, ch. 2, sec. 2.4, pp. 19-73.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc: Int. Conf. Computer Vision and Pattern Recognition, 2016, pp. 770-778.

[24] Yatao Zhang, Junyan Li, Shoushui Wei, Fengyu Zhou, Dong Li. "Heartbeats Classification Using Hybrid Time-Frequency Analysis and Transfer Learning Based on ResNet," IEEE Journal of Biomedical and Health Informatics, 2021.

[25] Md. Rashed-Al-Mahfuz, Mohammad Ali Moni, Shahadat Uddin, Salem A. Alyami, Matthew A.Summers, Valsamma Eapen. "A Deep Convolutional Neural Network Method to Detect Seizures and Characteristic Frequencies Using Epileptic Electroencephalogram (EEG) Data," IEEE Journal of Translational Engineering in Health and Medicine, 2021.

Design of Mobile Application Auction for Ornamental Fish to Increase Farmer Sales Transactions in Indonesia

Henry Antonius Eka Widjaja, Meyliana, Erick Fernando
Stephen Wahyudi Santoso, Surjandy, A.Raharto Condrobimo
Information Systems Department, School of Information Systems
Bina Nusantara University, Jakarta, Indonesia 11480

Abstract—This article focuses on designing a mobile application for ornamental fish auction transactions for fish cultivators in order to increase their sales. This mobile app was created using a prototyping methodology and a four step process. The first is communication, the second is Quick plan and design, the third is construction prototyping, which develops a tender application, and the last stage is development delivery and feedback. Data validation is carried out for users such as farmers, bidders, or buyers in developing the application. The results of this paper propose a mobile auction application that provides auction information and bidding by bidders and sellers. The results show that the application is validated and declared usable and feasible to conduct auctions and bids as needed. This application can increase sales and improve the economic life of ornamental fish farmers in Indonesia.

Keywords—Mobile application; auction; ornamental fish; prototyping model

I. Introduction

The Covid-19 epidemic has not prevented ornamental fish cultivators from countrywide production of ornamental fish products. Freshwater fish, sea fish, and shrimp are included in the category of decorative fish commodities. The expected harvest of this sector commodity is 450 thousand tons between April and June 2020, according to statistics from the Ministry of Maritime Affairs and Fisheries (KKP). Many places, including Aceh, Bengkulu, North Sumatra, South Sumatra, Lampung, Riau Islands, Bangka Belitung Islands, Central Java, East Java, West Java, Bali, West Nusa Tenggara, South Sulawesi, West Sulawesi, Central Sulawesi, Kalimantan, and Maluku, are involved in the harvest [1]. The production of 1.8 billion heads of ornamental fish has been set as the 2020 goal by the Government of the Republic of Indonesia through the KKP. Ornamental fish, as is well known, are one of the trustworthy sources of foreign exchange to promote national economic growth [2], [3]. As its production potential rises, Indonesia is struggling with a major issue, particularly marketing. However, because the information acquired differs from that held by actors in the ornamental fish industry, the government also encounters challenges when making decisions. A more significant sales process is required in many ways with larger manufacturing data. Online auctions are one possibility for utilizing technology to assist farmers in selling.

Online auctions have been conducted since the turn of the century on many websites, including ebay [4]. Online auctions enable participants or customers to sell and bid on goods using internet services [5], [6], [7]. Users may easily connect to one service thanks to this auction without being restricted by location or time by giving thorough descriptions of the goods[4], [8], [9], [10], [11]. The online auction system also makes sure that the information about the commodities auctioned by the auctioneer is safer. Additionally, the price paid will match the auctioneer's desired price. Additionally, it allows the bidder to decide whether to purchase by making an acceptable offer; if the offer is accepted, the bidder wins the item. Additionally, buyers might locate value products that are more appropriate, less expensive, and comfier. Additionally, consumers realize the benefits of using an online auction system to purchase things at lower costs, in a more comfortable environment, and with less financial risk [12], [13]. Because pricing will match the number of bids from customers, online auctions hugely impact selling prices. Users of the online auction system gain advantages from each of these.

This study aims to create a mobile application that functions as an online auction. This system was created using prototyping, allowing simple system adaptation to user needs. Users of the online auction system can access it anytime and anywhere. This system also makes it easy to carry out business transactions according to the agreed price, allowing buyers to easily access information about the commodity to be auctioned.

II. Literature Review

Auctions were originally introduced in ancient Greece. The auction is conducted using a declining pricing method, which starts with the highest price and goes lower to find the lowest bid. Its development is getting faster from auctions in various countries. Auctions are specifically used to sell goods with sellers setting a base price and the following rules a bidder can bid and compete with each other to be able to win bids [14].

Online auctions are services for selling or bidding on goods between auction participants or consumers and sellers via the internet [8]. This auction process is carried out with bidders able to bid, which is acceptable and increases as the

number of bidders' bids. With this online system, auctions can be carried out without geographical, time, or physical meeting restrictions [8].

Internet auctions were created to facilitate market transactions to change how sellers and buyers do business globally [8], [10], [15]. Online use with internet capabilities offers great convenience in providing information about prices, shops, and products being auctioned. Another advantage is that it offers sellers and bidders a unique way of buying and selling transactions happening in real-time.

III. METHODOLOGY

In this study, a mobile application prototype for Indonesian ornamental fish auctions was created using a prototyping methodology.

A. Prototyping Methodology

In order to construct an auction application quickly and focus on stakeholders' needs, starting with the communication stage [16]. The prototyping technique is a necessary step. Investigating all application needs and stakeholder requirements is the goal of the communication phase. The second stage, Quick plan and design, is concerned with planning and modeling quickly to reflect design elements that end users utilize, such as the application's interface or appearance. Creating a prototype that meets the needs of stakeholders and the design is the third stage of the construction prototype. The last stage of development, delivery, and feedback, explains how the prototype is provided and assessed to stakeholders to provide feedback. This model's prototyping phase is depicted in Fig. 1.
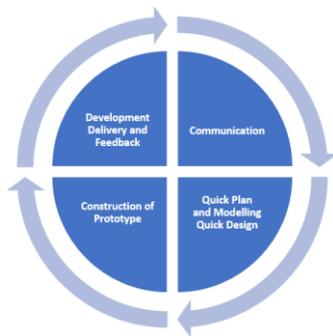
Fig. 1. Prototyping Methodology.

B. Architecture Application

This application architecture clearly explains how the application functions. This architecture illustrates how users and administrators can access the system through the application. The user can complete the auction transaction via a user interface. The database will be involved in the transaction process for the auction. The requested data and data processing, such as entering, editing, deleting, updating, and recording user bidding transactions, will be made available by the database. This Architecture application can be seen in Fig. 2.
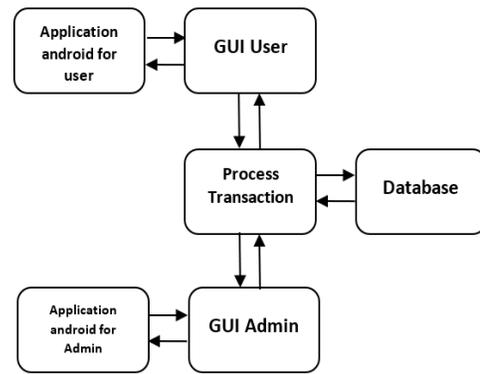
Fig. 2. Architecture Application.

C. Mobile Application Evaluation

The International Organization for Uniformity (ISO) evaluation of this mobile application system will give the same standardization between application development and industry needs in software quality measurement [17], [18], [19]. The 1991-first-developed ISO 9126 standard is being used in this investigation. This ISO standard is powerfully adaptable and can be applied to numerous systems. As shown in Fig. 3 are the six standards, and Table I describes the questions of the six standards used for evaluation.

TABLE I. EVALUATION VARIABLES AND INSTRUMENT OF ISO 9126

Functionality
Can software complete the tasks?
Has the outcome met expectations?
Will the system talk to other systems?
Can the software prevent unauthorized access?
Reliability
Have the majority of the software flaws been fixed over time?
Does the software have the ability to handle errors?
Can the software be restarted and lost data recovered after a failure?
Usability
Does the user perfectly understand how to utilize the system?
Can the user quickly learn how to utilize the system?
Can the user operate the system easily?
Is the UI appealing?
Efficiency
How rapidly does the system react?
Does the system effectively use its resources?
Maintainability
Is it simple to diagnose faults?
Can you easily modify the software?
Can updates be made without the software ceasing to function?
Easily testable software is this?
Portability
Is it possible to transfer the software to different settings?
Easy installation of the software?
Does the software follow portability requirements?
Is the program capable of readily replacing other software?
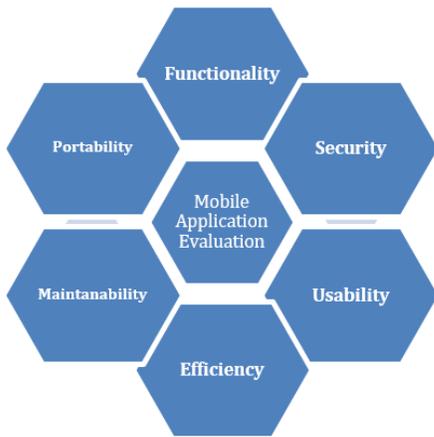
Fig. 3.   Six Characteristics of ISO 9126 as a Framework for Evaluating Software Quality.

IV. RESULT AND ANALYSIS

This section discusses the steps of prototyping methodologies for system development and the research findings. These stages include communication, rapid planning and design, prototype construction, development delivery, and feedback. The procedure for a case study on a prototype application is described in this section.

A. Communication Stage

The requirements and stories of the users are identified at this early stage. Users of the application should be made aware of the needs identification process, which includes the sales that sequentially offer information on auctions and bids and ensure that various users can enjoy the same advantages, as shown in Table II.

TABLE II.   REQUIREMENTS

Requirements
The program must contain a page requesting each user's username and password.
The app should feature a primary menu where new users may register themselves and where various jobs will be displayed.
The main menu with a list of tasks should be present in the program.
For applications to hold an auction, a bidding website is required.
A checkout page is a must for the app.
An app's history page is required.

B. Quick Plan and Design Stage

This section discusses the use case, which deals with the various temporal circumstances of each prototyping working sprint. There are two parts to this stage. The first is the Quick Plan, which assesses the requirements for developing prototypes in accordance with user needs. Second, Modeling Quick Design creates design specifications based on a system's needs analysis. The Unified Modeling Language (UML) approach is used in this work.

1) Quick plan: Interviews with relevant users, primarily ornamental fish growers and purchasers generally, are conducted in order to analyze user needs. Furthermore, the

input, process, and output needs are translated from this analysis. Data such as seller and buyer profiles, auction data, history data, payment data, and product data are needed to analyze the input requirements. While the management of the input data is described in the process data. For the information to help users comprehend, the output data is finally in the form of information from the input data processing in the program.

2) Quick design: Analysis of the needs of users is that users can register with the application, users can bid from the bid list, users can log in to use the application, and users can make payments from winning results from bidding. Users can view the history of bids that have been created. Utilizing use case diagrams in UML, this analysis performs needs modeling, as shown in Fig. 4.
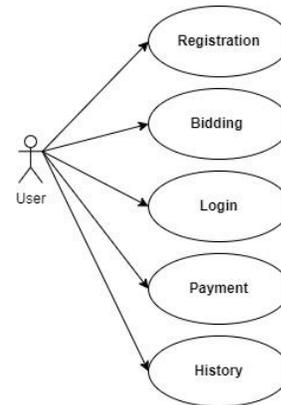
Fig. 4.   Usecase Diagram.

3) Construction of prototype: The effort to create the application starts at the prototype construction stage. In order to develop a prototype that actually meets user needs, the work is done in accordance with the information gleaned through the analysis of user needs outcomes. Execution of a design:

a) Design 1: Main Page of the application: A login page allows users to access their own accounts. When there are user users and administrator users, respectively (See Fig. 5(a)). And, to register with a new account in the application, first enter the user identity consisting of name, email address, password, phone number, country, address (see Figure 5b). This is shown in Fig. 6.

| (a) | (b) |
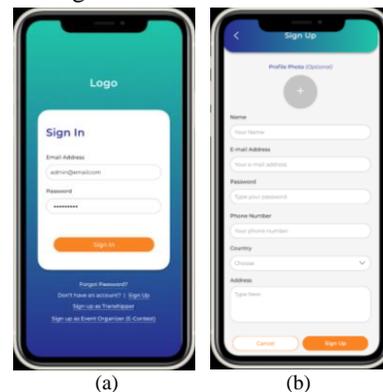
Fig. 5.   Main Page of the Application.

b) Design 2: User Main Page: Main page User user who owns store as auction maker. (a) Main Page The auction owner user in the application consists of a product menu, auction menu, delivery order menu, (b) Auction creation page, (c) View information on Auctions that have been made, (d) View list of auction winners, (e) Confirmation page to the auction winner, (f) Auction detail information list page, (g) Order history page has been made. This is shown in Fig. 6.

c) Design 3: The main page of users who bid on auctions: (a) User main page, which consists of profile menu, my bid, and history, (b) Settings page, which functions to set profile, currency, payment method, notification, and sign out, (c) My Profile page functions to update user's data, (d) My order auction page which displays auction product orders, (e) Order info page displays detailed information on unpaid auction orders, (f) Order info page displays information on auction product orders with the shipping address inputted, (g) page order info provides information that payments have not been made, (h) Checkout page displays information on choosing the payment method, (i) Order info page displays information on payment confirmation, (j) Summary page displays all auction orders that have been paid for and are ready to ship. (k) History page that displays the history of auction orders. This is shown in Fig. 7.

d) Communication stage: At this last stage, the finished application is given to the user to see whether the system is running well, and researchers get feedback from users through the evaluation of the application. This evaluation uses a forum group discussion for users. From the Mobile application evaluation results, the researcher decided to provide a solution, as seen in Table III.
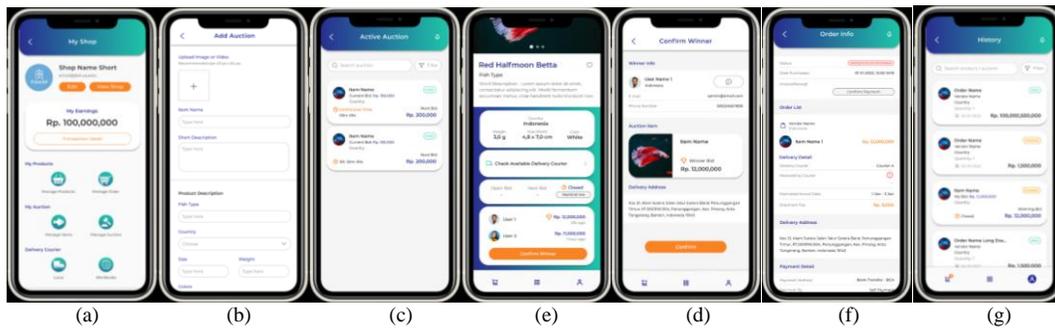
Fig. 6. The User Main Page.
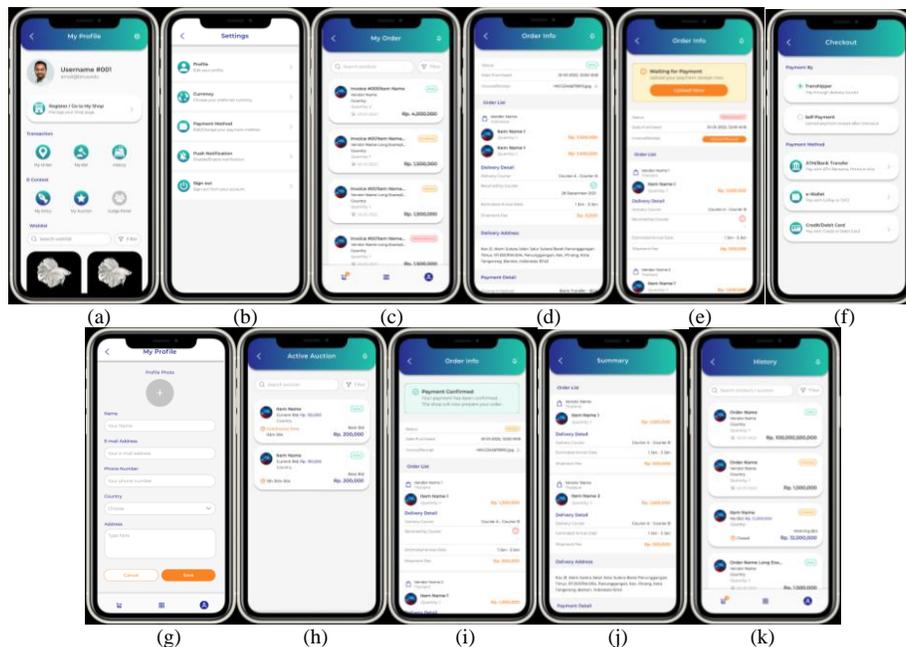
Fig. 7. The Main Page of Users who Bid on Auctions.

TABLE III.   REVIEWER COMMENTS AND SOLUTIONS FOR IMPROVING THE MOBILE APP

Comments from users	Solution
Content of Auction	
Incomplete organizing of the auction	To provide more thorough information, auction materials can be updated with fresh sources.
The instruction's scope supports the purpose	-
The auction documentation is simple to follow	-
New sources can be added to the materials used in auctions.	-
Access and Links	
The application does not provide a search feature.	The application is unable to find an object. It will be investigated more in the future.
Functionality	
The outcome is as expected.	This system can be more interactive
The interactivity access in the display is adequate.	
The program prevents illegal users from logging in.	-
Reliability	
The majority of the software's flaws may not have been fixed over time.	- The program is not set up to deal with the error
The application's ability to handle faults is uncertain.	- The application isn't built to deal with errors.
The application's ability to recover lost data after failure is uncertain.	- The software is not intended to recover lost data following a failure.
Usability	
The system's use is simple for the user to understand.	The user interface might use some improvement.
The system's user interface is simple to understand.	-
The user uses the system with little effort.	-
Efficiency	
The application has a quick response time.	The system is unable to use resources effectively.
Does the system make efficient use of its resources?	-
Maintainability	
Any signs do not show a defect application.	The software is difficult to modify because a skilled coder for multimedia was required.
It is uncertain whether the program can be changed easily.	Changes are performed, and the software is unable to continue operating.
Is it certain that the application will continue to work if changes are made?	Any signs do not show a defect application.
The program should be improved to detect errors in further study.	It is uncertain whether the program can be changed easily.
Portability	
The application cannot be loaded on an iPhone.	It will eventually be produced in an IOS version for iPhone users.
Installation of the software is simple.	
Not sure if the app can easily replace another piece of software?	

The project's completion with the provisions that were envisaged at the outset is the process's last stage. This phase explains how the functionality of the application's design and function has been correctly assessed in light of the proposal. The table will provide the traits and features from the previous stage and describe what first-time mobile app users do.

V. DISCUSSION

The development of this ornamental fish auction application has the following system: Every user who wants to participate in this auction system must register on the application. After registration, the user can log in to use the features in the application, among others, as a seller who makes auction information or as a user participating in auctions. After participating in the auction and winning, the user can see the list of winning auctions on the shopping list. Then make a payment by providing the identity of the address and selecting the sender service provided. After the payment is complete, the sale will process the delivery according to the delivery service chosen and the address of the auction winner. The auction winner can track the delivery process carried out by the seller so that it can create a sense of trust in the application or seller. After the auction recipient gets the auction product (ornamental fish) can confirm receipt on the application and provide feedback to the seller.

The development of this application has received a good response from users of the processes that occur in the system. This is because it has been by the needs required of the user.

In addition, the user provides a view of the need for continuous development from the application side, among others. First, the application can provide complete information content than the product being auctioned or other products that will be auctioned next in terms of application content. From the application access side, it is expected to provide reasonable access, for example, by providing a search feature. In terms of functionality, users hope to get a more interactive and faster system from each of the features provided. From reliability, users wish to reduce errors when accessing applications due to slow responses and can recover lost data. In terms of usability, the user is quite satisfied with the appearance but still needs to improve some application features. In terms of efficiency, the user needs efficiency from using data that is too large to reduce excessive internet usage. In terms of maintainability, the user hopes that the application user interface will not have too many defects, so it is necessary to develop an interface that is easier to use and can be accessed easily. In terms of portability, users find it very easy to install the devices used and hope that it can be developed with various devices so that it can be used more.

VI. Conclusion

The result of the application for ornamental fish auctions provides easy support for transactions for ornamental fish farmers and hobbyists. On the other hand, the application helps in processing a very wide range of transactions regardless of space and time because the application can be accessed anywhere and anytime by the user. The results received by the author are that the application platform can increase the income of ornamental fish farmers with a wide market reach and can foster consumer confidence to make transactions with clear data validation from the seller/auction maker.

In addition, the application is developed by giving priority to auctions only so that it can give focus to users to make transactions with the applicable auction procedures. This auction procedure is very easy to learn because it only bids on the auction information provided. If you win, it will be further informed to the winner of the sale that made the auction. Then, you must agree to take the next step to process the shipping address, and the payment has been completed through the auction process.

Moreover, in-depth research on the auction system, especially for ornamental fish, is very rare, so it provides opportunities for fish farmers to be more advanced in their transactions, get to know buyers, or get more income.

This research is developing by approaching the prototyping method, feeling the advantages at this stage. The stages carried out are appropriate, correct, and supportive in the development process to get applications that suit user needs. In addition, this stage can be seen clearly from application development to evaluation so that the application is appropriate in accordance with user needs. The development process can identify problems that are applied.

On the other hand, it should be noted that this prototype method has several limitations. It does not clearly explain the team members working on it, and the work is not scheduled in detail according to the originally planned time. So this note is estimated to be a top priority in development, so it needs to adjust the focus of the completion time of each application module later. Future work needs to be re-examined regarding adding features for efficient and effective payment methods with various banking services such as credit cards, various FinTechs, and enabling cryptocurrencies. And can also feature live-streaming auctions between sellers and buyers.

Reference

[1] Kementerian Kelautan dan Perikanan, "Kementerian Kelautan dan Perikanan." https://kkp.go.id/artikel/18532-pandemi-covid-19-estimasi-panen-perikanan-budidaya-capai-450-ribu-ton-sepanjang-april-hingga-juni-2020 (accessed Oct. 27, 2020).

[2] "KKP Ajak Pengusaha Ikan Hias Manfaatkan Media Sosial - Bisnis Liputan6.com." https://www.liputan6.com/bisnis/read/4383796/kkp-ajak-pengusaha-ikan-hias-manfaatkan-media-sosial (accessed Oct. 27, 2020).

[3] KKP, "Genjot devisa ekspor, tahun ini KKP menargetkan produksi ikan hias 1,8 milyar ekor," https://kkp.go.id/djpb/artikel/16865-genjot-devisa-ekspor-tahun-ini-kkp-menargetkan-produksi-ikan-hias-1-8-milyar-ekor (Diakses tanggal 16 Juni 2020), 2020. https://kkp.go.id/djpb/artikel/15803-optimalisasi-potensi-budidaya-ikan-hias-nasional (accessed Oct. 27, 2020).

[4] G. S. Black, "Consumer demographics and geographics: Determinants of retail success for online auctions," J. Targeting, Meas. Anal. Mark., vol. 15, no. 2, pp. 93–102, 2007, doi: 10.1057/palgrave.jt.5750035.

[5] Goes, Karuga, and Tripathi, "Bidding Behavior Evolution in Sequential Auctions: Characterization and Analysis," MIS Q., vol. 36, no. 4, p. 1021, 2012, doi: 10.2307/41703496.

[6] J. Castro, R. Espínola, I. Gutiérrez, and D. Gómez, "Auctions: A New Method for Selling Objects with Bimodal Density Functions," Comput. Econ., May 2022, doi: 10.1007/s10614-022-10259-1.

[7] R. H. Gunnelin, "Bidding strategies and winner's curse in auctions of non-distressed residential real estate," pp. 1–35, 2020, [Online]. Available: https://www.diva-portal.org/smash/get/diva2:1476552/FULLTEXT02.pdf

[8] A. E. Wen and H. Wong, "Bidder and Seller Strategies in Online Auctions: A Review," Proc. 2021 3rd Int. Conf. Econ. Manag. Cult. Ind. (ICEMCI 2021), vol. 203, no. Icemci, pp. 1993–1999, 2022, doi: 10.2991/assehr.k.211209.325.

[9] P. L. Lorentziadis, "Optimal bidding in auctions from a game theory perspective," Eur. J. Oper. Res., vol. 248, no. 2, pp. 347–371, 2016, doi: 10.1016/j.ejor.2015.08.012.

[10] P. Khezr and F. Menezes, "Auctions with an asking price," Int. J. Game Theory, vol. 47, no. 4, pp. 1329–1350, 2018, doi: 10.1007/s00182-018-0620-3.

[11] G. Loyola, "Effects of competition in first-price auctions," Econ. Theory, vol. 71, no. 4, pp. 1527–1567, 2021, doi: 10.1007/s00199-021-01347-8.

[12] S. Georganas and J. Kagel, "Asymmetric auctions with resale: An experimental study," J. Econ. Theory, vol. 146, no. 1, pp. 359–371, 2011, doi: 10.1016/j.jet.2010.12.001.

[13] S. Agarwal, J. Li, E. Teo, and A. Cheong, "Strategic Sequential Bidding for Government Land Auction Sales – Evidence from Singapore," J.

Real Estate Financ. Econ., vol. 57, no. 4, pp. 535–565, 2018, doi: 10.1007/s11146-017-9625-0.

[14] Y. Jiang, "The Development Process and Future Prospects of Auctions," Proc. 2021 3rd Int. Conf. Econ. Manag. Cult. Ind. (ICEMCI 2021), vol. 203, no. Icemci, pp. 176–181, 2022, doi: 10.2991/assehr.k.211209.028.

[15] P. Korgaonkar, E. Becerra, B. O'Leary, and D. Goldring, "Product classifications, consumer characteristics, and patronage preference for online auction," J. Retail. Consum. Serv., vol. 17, no. 4, pp. 270–277, 2010, doi: 10.1016/j.jretconser.2010.02.004.

[16] R. S. Pressman and B. R. Maxim, Software engineering: A Practioner's Approach. 2020.

[17] E. P. Jharko, Evaluation of the quality of a program code for high operation risk plants, vol. 19, no. 3. IFAC, 2014. doi: 10.3182/20140824-6-za-1003.02140.

[18] B. Behkamal, M. Kahani, and M. K. Akbari, "Customizing ISO 9126 quality model for evaluation of B2B applications," Inf. Softw. Technol., vol. 51, no. 3, pp. 599–609, 2009, doi: 10.1016/j.infsof.2008.08.001.

[19] H. Al-Kilidar, K. Cox, and B. Kitchenham, "The use and usefulness of the ISO/IEC 9126 quality standard," 2005 Int. Symp. Empir. Softw. Eng. ISESE 2005, pp. 126–132, 2005, doi: 10.1109/ISESE.2005.1541821.

An Automatic Adaptive Case-based Reasoning System for Depression Remedy Recommendation

Hatoon S. AlSagri[1], Abeer Abdulaziz AlSanad[4], Lulwah AlSuwaidan[5], Halah Abdulaziz Al-Alshaikh[6]
Information Systems Department. College of Computer and Information Sciences
Imam Mohammad Ibn Saud Islamic University
Riyadh, Saudi Arabia

Mourad Ykhlef[2]
Information Systems Department. College of Computer
Information Sciences
King Saud University (KSU)
Riyadh, Saudi Arabia

Mirvat Al-Qutt[3]
Computer Systems Department. Faculty of Computer and
Information Sciences
Ain Shams University
Cairo, Egypt

Abstract—**Social media data represents the fuel for advanced analytics concerning people's behaviors, physiological and health status. These analytics include identifying users' depression levels via Twitter and then recommend remedies. Remedies come in the form of suggesting some accounts to follow, displaying motivational quotes, or even recommending a visit to a psychiatrist. This paper proposes a remedy recommendation system which exploits case-based reasoning (CBR) with random forest. The system recommends the appropriate remedy for a person. The main contribution of this work is the creation of an automated, data-driven, and scalable adaptation module without human interference. The results of every stage of the system were verified by certified psychiatrist. Another contribution of this work is setting the weights in case similarity measurement by the features' importance, extracted from the depression identification system. CBR retrieval accuracy (exact hit) reached 82% while the automatic adaptation accuracy (exact remedy) reached 88%. The adaptation presented an error-tolerance advantage which enhances the overall accuracy.**

Keywords—Case-based reasoning (CBR); depression; remedy; adaptation; similarity; twitter

I. INTRODUCTION

Depression is among the most critical mental illness that put more than 300 million people lives on risk, and affect patient and his family quality of life. Depression is highly common nowadays and recorded an increased number of cases due to economic factors, global pandemics, and social isolation. Early detection, with an appropriate remedy recommendation is the key to minimize the emotional and financial burden of the depression. However, the increased usage of social network platforms has brought up the essential need to modern analysis techniques on the enormous amount of data available. Twitter, as one the most important and commonly used social network platforms, has a great impact around the world in seconds. More than 500 million tweets are shared per day, bringing up more information and emotions to be diffused among the large number of users. Negative emotion spreads through the network, raising mental illness,

especially depression. As a result, techniques from different fields have been used to analyze and process social network data to diagnose and suggest the appropriate treatments to mitigate such situations [1, 2]. This work exploits Case bases reasoning approach to detect depression in Twitter users. As a result, this study proposes an automatic adaptive case-based reasoning system for early depression detection on social network and automatic remedy recommendation.

Case-based reasoning (CBR) retrieves, analyzes, and stores new knowledge, making it available for solving problems [3]. CBR were used in medical assistant systems for diagnosing and treatment formulation [4 - 9]. Similarly, machine learning is a significant technique for classifying and predicting illnesses through patterns found in the training dataset [10].

This work extends the previous work in the identification of depression levels (tended to be depressed, depressed, deeply depressed) [11, 12] by adding a remedy recommendation system. It uses a hybrid system of CBR and the Random Forest algorithm (RF-CBR), where Random Forest (RF) outcomes feed the CBR to find treatment for users who are depressed. This study aims to recommend the best remedy to depressed Twitter users, when they are in their early stages of illness, or else recommend the visit to a psychiatrist. The most critical stage in the CBR system is the similarity calculation to find the best fit for the new case from the knowledge base. Researchers have used various techniques to reach for the best similarities the match in the CBR model. For example, [13] and [14] used correlations, while [13] used fuzzy logic to calculate the similarity of users to diagnose depression. On the other hand, [14] showed that a combination of RF with the CBR had the best results and highest accuracy among other techniques. As a result, this study used feature importance measurements for calculating similarity through the integration of RF and CBR (RFCBR).

The main challenge of CBR is the adaptation task, which depends on the domain and application characteristics [7]. Most CBR systems are built as retrieval-only systems, especially CBR in the medical and healthcare domain, as they

leave it to human experts [15, 16]. The adaptation task is often left out due to the complexity of the domain or the difficulty in acquiring the knowledge needed [7, 17]. Some studies have attempted to adopt and explore automatic and semi-automatic adaptation strategies, such as [16, 18-22], but they have all required human interference to complete and/or assess a system's performance. The integration of CBR with other methodologies has been used to overcome the adaptation problem [11]. For this, most CBR systems use either human experts or are rule-based for the adaptation task, while this study is data-driven, which, to our knowledge, is novel and has not been introduced for depression treatment.

The proposed work enhanced the previous research by developing an automatic system called Remedy Recommender Model (RRM) entailing RF and CBR for detecting users' depression levels and in consequent of those outcomes recommends the best remedy. The experiments depend on ground truth data prepared by psychiatrists. They studied the cases in the knowledge base and gave recommendations of remedies accordingly. Three feature importance measurements are used each in a separate experiment to identify the features' weights, namely, overall, permutation, and tree interpreter feature importance measures. The results of this study correspond to [12], showing the best results from retrievals and adaptation using the tree interpreter feature importance measurement.

The rest of the article is organized as follows. Section II illustrates the previous work in both CBR and healthcare. Section III elaborates on the background, while Section IV drafts the proposed approach and solution. Section V discusses the experimental results in addition to comparative analysis with other authors. Finally, Section VI concludes and positions the findings and insights.

II. RELATED WORK

CBR has been used in healthcare for diagnostics and treatment. CASEY is one of the earliest medical expert systems that used CBR for heart failure diagnosis. The system searches for similar cases, finds the differences between a current and a similar case, and transfers the diagnosis of the similar case to the current case. The system uses rule-based domain theory if modification attempts fail or if no similar cases are found [7]. Additionally, Nasiri et al. [9] introduced the DePict CLASS, a cased-based learning assistant system that recommends information retrieved from dementia research based on the ICF framework of the WHO. The system detects and predicts the disease using image classification and text information. Caregivers and domain experts use and update the DePicT CLASS for dementia that is used by caregivers and patients' relatives to find answers in dealing with their problems [9]. Moreover, Mulyana et al. [23] developed a case-based reasoning system for diagnosing types of schizophrenic disorders and mood disorders with their treatment. Medical records of patients with mental disorders were obtained from a mental hospital in Yogyakarta and were used to construct the knowledge base of the system. The system selects the case with the highest similarity to the new problem and recommends its treatment to the new case [23].

To increase the accuracy of the similarity measurements in CBR, the authors in [24 - 26] introduced fuzzy logic with CBR. For example, Ahmed et al. [6] reached an accuracy of 88% in the diagnosis and treatment of stress. Houeland et al. [27] also introduced the random decision tree (RDT), which proved that the hybrid combination outperformed the base algorithms in similarity measures. Similarity was calculated using proximities that were computed using an RF algorithm.

Furthermore, Asim et al. [15] compared the nearest-neighbor and artificial neural network to RF and found that RF contributed to the efficiency of the CBR system they used to identify influential bloggers. They reached an accuracy of 89% when using Gini impurity as weights in the similarity measure of the CBR system. Hseih et al. [28] also compared different classifiers and ensemble classifiers as the best for use with CBR to find Internet-addicted users and were able to reach 89.9% accuracy. In addition, Gu et al. [29] used CBR with a genetic algorithm to diagnose breast cancer, achieving 0.927 accuracy.

In the mental illness sector, several studies used CBR to help diagnose or treat mental illnesses. Mulyana et al. [30] used CBR to diagnose mood disorders, achieving 89.3% accuracy using Modified Tcersky to increase similarity retrieval accuracy. On the other hand, other studies used CBR with different techniques. Kwon et al. [31] used crowd knowledge with CBR to diagnose depression and stress. Furthermore, Rahim et al. [32] used the help of specialists and built an expert system with CBR to diagnose physiological disorders. Table I summarizes some CBR systems in different domains.

TABLE I.   COMPARISON OF CBR SYSTEMS IN DIVERSE DOMAINS

Author/ Domain	Techniques	Results	Similarity	Adaptation
Asim [15] (blogger influence)	RF + ANN + C4.0 + NN	f-measure (85% before adaptation, 91% after adaptation)	Gini entropy	Yes
Mulyana [23] (mental disorder)	CBR	Accuracy: 99%	NA	No
Hsieh [28] (internet addiction)	Ensemble classifier + CBR	Accuracy: 89.9%	Weight of classifier	No
Gu [29] (breast cancer)	CBR	Accuracy: 0.927	Genetic algorithms for weights	No
Mulyana [30] (mood disorder)	CBR	Accuracy: 89.3%	Modifies Tversky Similarity	No
Kwon [31] (depression & stress)	Crowd knowledge + CBR	Accuracy: 92.5%	NA	No
Rahim [32] (phycological disorder)	CBR + expert system	Accuracy: 75%	NA	No

As previously mentioned, most studies have used human experts for adaptation tasks or ignored them. Eknog et al. [14] and Nasisri et al. [9] used experts to perform the adaptation task. On the other hand, Asim et al.'s [15] adaptation was automatic in its use of similarity equations but with no change in the results.

The present study reveals the gap in the treatment of depression and other mental illnesses is that most of the studies did not utilize adaptation or created adaptation with human assistance.

III. BACKGROUND

A. Case-based Reasoning

CBR is found to be a good model because it assimilates problem solving, understand, and learn, and integrate all of them with memory processes. It also amends old solutions to meet new requirements [13]. Thus, in the past few decades, has befitted the medical domain as intelligent computer-aided decision support systems [6, 13]. CBR system depends on reusing previous knowledge to solve new problem through defining the similarities between them. Later, before applying the solution (remedy) CBR revises them then retains applicable solutions (remedies) for further reuses [33 - 36]. These steps of the 4R processes demonstrate the R4 model developed by Aamodt and Plaza [13, 37].

Based on the CBR cycle (Fig. 1), the RETRIEVE process in this article starts with identifying similar cases in the knowledge base using the features' importance and the score of each case. Similarity calculations help find the most similar case with the most similar treatment. A case that has the highest level of similarity will be suggested as the most alike treatment for the case, and this is called the REUSE process. Next, the case will enter the REVISE process, where an adaptation algorithm will be applied to identify if this is the most suitable treatment or if there is another case with better treatment. Finally, after finding the most similar case and recommending the most similar treatment, the RETAIN process can save the case in the knowledge base as a new case.
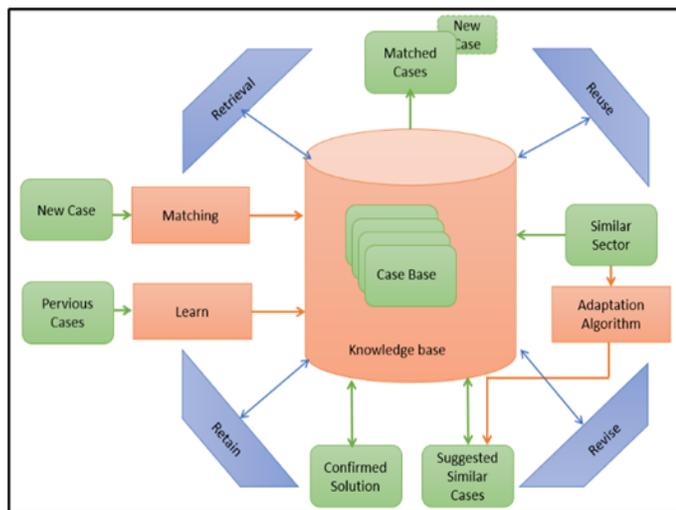
Fig. 1. The CBR Cycle.

B. CBR in Healthcare

CBR's ability to solve problems and its efficiency in the recommendation process has attracted researchers in the healthcare domain [4 - 9]. CBR systems are increasingly used in healthcare due to its utility for the thought process of a physician [13, 38]. Automatic establishment of a facility-adapted knowledge base is highly beneficial to CBR systems in healthcare [39], which is of great significance in medical decision making [13].

CBR has also been used for developing intellectual computer-aided decision support systems in the medical domain in the past few years because of the continuously changing nature of the medical knowledge base and the presence of more than one solution [6, 13, 40].

IV. PROPOSED SYSTEM (RRM)

RRM used for depression detection starts with retrieval of similar cases in the knowledge base which is an important and critical phase to reach an optimal solution for the desired case. The similarity – as first step in the CBR cycle - depends on the features weights to get the most similar cases according features similarities of both cases. The most similar cases will be used to enter the CBR's REUSE stage. Later, adaptation phase is performed in the REVISE stage which is important to recommend the most suitable treatment for depression in its early stages. Finally, the recommended new case will be retained in the knowledge base. This study uses data from [11], data collected consisted of Twitter users who were depressed and non-depressed. A total of 500 users were collected with more than 1M tweets, and 334 users were classified as depressed. The data consisted of user accounts' information and tweets.

The data has been used for running RF to classify users as depressed and non-depressed users and find importance of each feature and the score of each case obtained from the RF probabilities [12]. Importance of features is used as weights for the similarity calculation to retrieve from the training set the most similar case to the case in the test set. Only depressed users' data is used in this study since they are targeted to recommend the best remedy for them. these results based on generating 10 random splits of data to 50% knowledge base and 50% unseen cases. The target is to neutralize the system to any cases temporal order to the system.

Data is trained using 10-fold cross validation to avoid over fitting and then tested on held-out test data. Fig. 2 summarizes the RRM phases. Novelty of this work appears in both retrieval and adaptation phase where the recommended solution for depressed users is proposed.

A. Retrieval Phase

Retrieval of similar cases utilizes Algorithm 1 using local and global similarity equations [33, 34]. The algorithm starts with calculating local similarity of features for the new case and old cases in the knowledge base:

$$sim\ (f_x, f_y) = 1 - \frac{|f_x - f_y|}{f_{max} - f_{min}} \qquad (1)$$
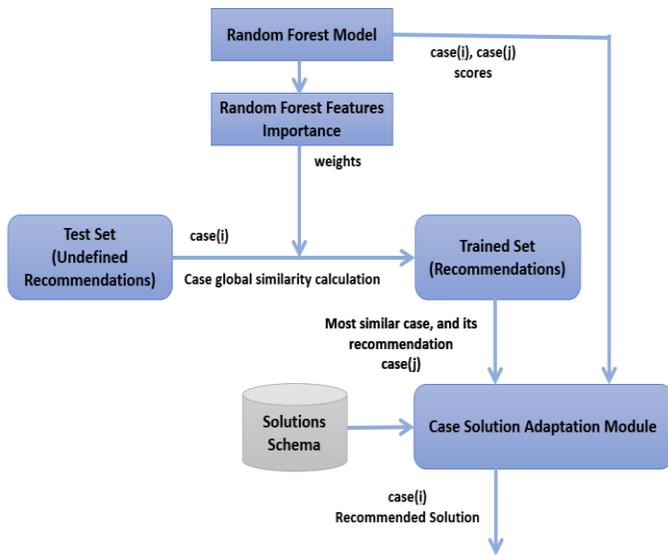
Fig. 2. The RRM System.

where f_x is the feature's value for the new case, f_y is the feature's value for the *y*-th old case, and $f_{max} - f_{min}$ includes the minimum and maximum values for all the old cases included in the database.

Next, the global similarity is calculated, which is the weighted sum of the local similarities of all features used in the analysis. Global similarity is calculated using the following equation:

$$similarity(x, y) = \frac{\sum_{f=1}^{n} w_f \times sim(x_f, y_f)}{\sum_{i=1}^{n} w_f} \qquad (2)$$

Where w_f is the weight of the f-th feature, $similarity(x, y)$ is global similarity between the case x from the training set and the new case y, and $sim(x_f, y_f)$ is the local similarity for the f-th feature between case x and the new case y.

Importance of the features concluded from the RF is used as the weight for each feature. Finding the most similar case is important for the accuracy of the adaptation phase.

To illustrate how algorithm 1 is implemented, this study provides an example from the cases used in this study. This will use a new case and a small case base, containing just two cases and four features as shown in Table II. Importance, minimum, and maximum values of each feature are also shown in Table II. The new case Ci will be compared to the cases stores in the training case repository. The goal is to find the most similar case to the new case to be able to recommend the same remedy. In other words, the remedy used for the case in the repository will be recommended the same remedy. In other words, the remedy used for the case in the repository will be recommended to the new case. Similarity calculations aims to find cases that are analogous in a way their remedy can be reciprocally reused.

Algorithm 1 starts with calculating local similarity for the new case Ci and the two cases using equation (1). Local similarity between C1 and Ci for the (Retweets feature) is:

Algorithm 1: Get_Most_Similar_case (C$_i$, TC)

Input:
C$_i$: input case
TC: training cases in case repository
Output: The most similar case

1:	MaxSim=0
2:	For all training cases *TC*:
3:	For all features in features space:
4:	• Calculate Feature_Local_Sim (case of **C$_i$** and case of TC) //using equation (1)
	End for
5:	• Calculate New_sim = GlobalSim (Feature_Local_Sim of **C$_i$**, Feature_Local_Sim of TC) //using equation (2)
	End for
6:	MaxSim=max (MaxSim, new_sim)
7:	Return the most similar case with MaxSim

TABLE II.   FEATURES' COMPARISON OF NEW CASE AND CASES IN REPOSITORY CORRESPONDING TO THEIR MAXIMUM AND MINIMUM VALUES

Features	C$_1$	C$_2$	New case C$_i$	Importance of features	Maximum Value	Minimum Value
Retweets	26	70	100	0.9	200	0
Hashtags	10	72	20	0.8	123	5
Depress	30	10	56	0.5	350	4
Hate	14	40	20	0.7	120	0

$$\text{Sim}(C_i, C_1) = 1 - \frac{|26-100|}{200-0} = 0.63 \qquad (3)$$

Similarly, local similarity for all features are calculated for the Ci with C1 and C2:

Sim (Ci, C1) = {retweets 0.63, hashtags 0.91, depress 0.92, hate 0.95}

Sim (Ci, C2) = {retweets 0.85, hashtags 0.56, depress 0.87, hate 0.83}

The second step in the retrieval phase is calculating global similarity. Equation 2 is applied to find similarity between cases and the new case. This is done by taking weighted sums of local similarities with the importance of features as weights. Global similarity is calculated as follows in Eq. 4 and 5:

$$\text{Similarity}(C_i, C_1) = \frac{1}{2.9} \times (0.9 \times 0.63 + 0.8 \times 0.91 + 0.5 \times 0.92 + 0.7 \times 0.95) = 0.834 \qquad (4)$$

$$\text{Similarity}(C_i, C_2) = \frac{1}{2.9} \times (0.9 \times 0.85 + 0.8 \times 0.56 + 0.5 \times 0.87 + 0.7 \times 0.83) = 0.769 \qquad (5)$$

Therefore, C1 were chosen to reuse and continue to adaptation phase. C$_i$ is more similar to C$_1$ than C$_2$.

B. Adaptation Phase

Adaptation phase depends on the knowledge base which has three sectors of remedies for different depression levels. Levels of depression are for users who tend to be depressed, depressed users, and users with advanced depression where a doctor consultation is a must. Cases' scores, Sc = {Sc1, Sc2,…Scn} are calculated and retrieved for RF model and saved in the knowledge base with appropriate recommended remedies R = {R1, R2,…Rn}, for each case.

Algorithm 2 illustrates the adaptation phase. For each new case C_i, score Sc_i, and remedy R_i, first, similarity is calculated with the cases in the training set to find the most similar case, C_j. Later, all mean score for remedies in the same sector are calculated. KNN is applied to find the nearest mean score which determines the nearest remedy appropriate for C_i.

In order to illustrate Algorithm 2, this study introduced an example from the study cases. Cases used were introduced in illustrating algorithm 1 above. After applying Algorithm 1, C1 was found to be closer to Ci. As a result, case C1 remedy will be first recommended for new case Ci but adaptation phase will try to reach a better remedy for Ci.

Algorithm 2: RRM Adaptation Pseudo code

Input:

TC: training cases in case repository (knowledge base)

C$_i$: input case

Sc$_i$: Depression classification score of case C_i

CSR: table with columns case, score and remedy

Output: remedy R_i

1: Case C_j = Get_Most_Similar_Case (C_i, TC) // ***using Algorithm 1***
2: RemedySector rs of C_i is the RemedySector of C_j
3: For every remedy r from rs:
4:     -   Obtain Sc(score) from CSR where remedy is r
5:     -   ScoreMean[r,m] where m is the mean of scores of table Sc
6: closeScore = minimum distance between input Sc_i and Scores in ScoreMean
7: R_i is a remedy r obtained from ScoreMean[r, m] where m is closeScore
8: Return remedy R_i

For Algorithm 2, score of new case Ci - derived from the RF result- is used as an input for the algorithm implementation where score Sci is equal to 0.74. Also, CSR containing scores and remedies for repository cases is given in Table III. Since C1 is most similar to new case Ci, remedy of C1 will be used to define the sector. Using classification of sectors in Fig. 3 remedy of C1 is R8 meaning that C1 and R8 is in depressed sector. For each remedy in the depressed case sector, the cases will be retrieved, and the mean score will be calculated. In other words, mean score for cases with remedy in [R6, R7, R8, R9, R10] will be calculated. From Table III, Mean scores of R6= (Sc2 +Sc7)/2= 0.69+0.75/2=0.72. Mean score of R8= (Sc1 +Sc3)/2= 0.785, mean scores of R9 = 0.82, and for R10 = 0.88.

To find best remedy for Ci, Algorithm 2 will apply KNN and find closest mean score to Sci and choose the corresponding remedy for Ci. Since mean score of R6 is the nearest neighbor to score Sci = 0.74 of input case Ci. Sci, Algorithm 2 will return R6 as a better remedy recommended for input case Ci.

TABLE III.   CSR CONTAINS SCORES AND REMEDIES FOR EACH CASE

Cases	Score	Remedy
1	0.85	R8
2	0.69	R6
3	0.72	R8
4	0.74	R1
5	0.32	R2
6	0.88	R10
7	0.75	R6
8	0.41	R3
9	0.82	R9
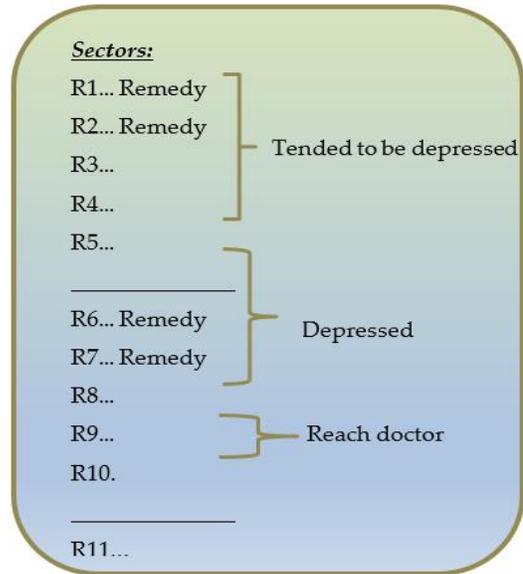
Fig. 3. Classification of Depression Sectors.

V. EXPERIMENTS AND RESULTS

Data used for this experiment has 133 depressed users and 66 not depressed. In an attempt to have ground reality, data for validation of the RRM system results, for each depressed user best recommended solution is assigned with the help of psychologists. This helped to identify how many correct retrievals of similar cases and number of successful adaptations of best solution to each depressed user.

Depending on the score of each case obtained from classification step using RF, the system categorizes the cases to one of the following sectors: tend to be depressed cases, depressed, and deeply depressed as shown in Table IV. Results proved that depression has the exponential distribution of contagious diseases shown in Fig. 4.

TABLE IV. PERCENTAGE OF EACH DEPRESSION LEVEL ACCORDING TO THE SCORES OF CASES

Level	Percentage to data
Tend to be depressed	62%
Depressed	30%
Deeply depressed	8%

A. Retrieval Phase Results

The implementation of CBR starts with applying the similarity equation to find the similarity of the new case to cases in the training set. As explained in the proposed system section retrieval phase is applied to find the most similar cases to the desired case. The results of the implementation showed the accuracy of retrieving similar cases from the knowledge base.

Retrieval experiments employ three different importance feature measures to determine the feature weights used in the similarity equation namely overall, permutation, and tree interpreter feature importance measures. Evaluation measures are applied independently for each experiment. Applying tree interpreter feature importance was able to retrieve successfully 54 users which revealed highest accuracy, 82% as shown in Table V. Precision, recall, and the f-measure revealed tree interpreter is the highest among the other importance criteria reaching 0.766, 0.349, and 0.480, respectively. Overall and permutation feature importance criteria had a smaller number of correct retrievals where the accuracy was 73% and 67%, respectively. Also, precision, recall, and the f-measure indicated that overall and the permutation are less accurate than tree interpreter. Overall results were 0.62, 0.343, and 0.442, respectively while the permutation results were 0.706, 0.343, and 0.472, respectively. Results of retrieval are summarized in Fig. 5.

B. Adaptation Phase Results

Completing the CBR cycle for the experiment, the adaptation phase results are conducted for each cycle and results are recorded according to the type of feature importance measurement used. Tree interpreter feature importance measurement has outperformed other feature importance measurements resulting accuracy of 88% with 58 successful user adaptations as shown in Table VI. Overall and Permutation feature importance criteria had lower result, 76% and 71% were their accuracy results, respectively. Precision, recall, and the f-measure indicated that tree interpreter had higher results than the other measures, reaching 0.762, 0.358, and 0.511, respectively, while overall and permutation reached precision of 0.671 and 0.7401, recall 0.456 and 0.3128, and f-measure 0.543 and 0.439 respectively. Fig. 6 illustrates the adaptation results.

Scores Distribution
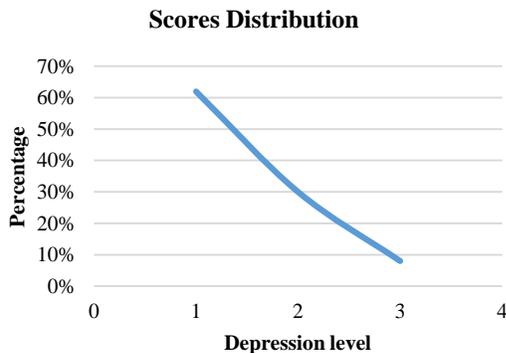
Fig. 4. Exponential Distribution of Depression Levels.

TABLE V. RETRIEVAL PHASE RESULT FOR DIFFERENT IMPORTANCE CRITERIA

Feature Importance Criteria	Retrieval Accuracy	Precision	Recall	F Measure
Overall	67%	0.61189	0.328456	0.427457
Permutation	73%	0.71192	0.30867	0.43063
Tree Interpreter	82%	0.76611	0.34986	0.480356
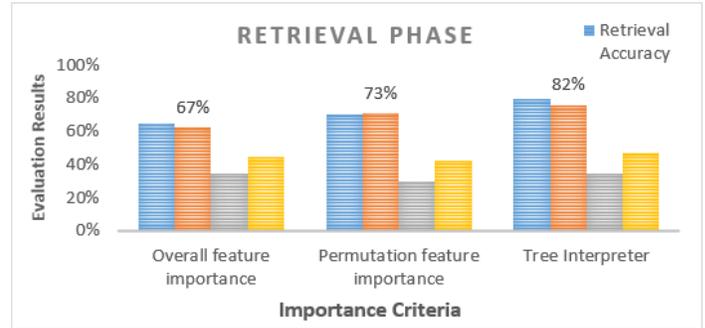
Fig. 5. Statistical Illustration of Retrieval Phase Results.

TABLE VI. ADAPTATION PHASE RESULTS FOR DIFFERENT IMPORTANCE CRITERIA

Feature Importance Criteria	Retrieval Accuracy	Precision	Recall	F Measure
Overall	71%	0.6713	0.4561	0.543161
Permutation	76%	0.7401	0.3128	0.439744
Tree Interpreter	88%	0.7624	0.38519	0.511801
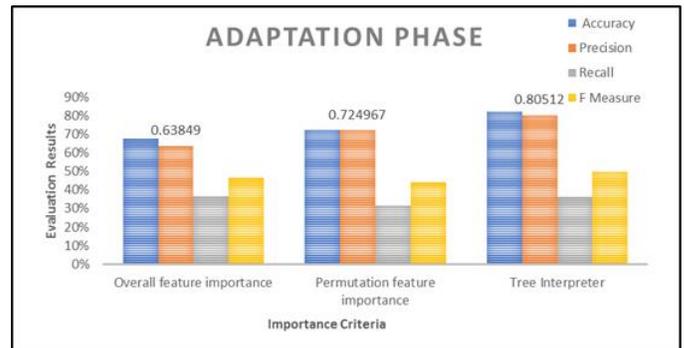
Fig. 6. Statistical Illustration of Adaptation Phase Results.

Results of retrieval and adaptation showed that tree interpreter feature importance results outcomes correspond to the results of [12].

VI. COMPARATIVE STUDY

This comparison has been considered to guarantee the effectiveness of the proposed RRM. Other studies used different data sets and labels and aim to propose objective and non-biased results and analysis, this comparative study will apply the same data to different techniques. As mentioned earlier, different studies have used correlations [14 - 15] and

fuzzy logic [16] to find similarities between cases in the retrieval phase. As a result, those techniques were implemented on this research's data and proved that tree interpreter is best used for feature weights in the similarity equation and the adaptation task.

As shown in Fig. 7 and Fig. 8, this tree interpreter in this study (TI) reached accuracy of 88% while the accuracy of fuzzy logic (Fl) and correlation (Cr) are 71% and 68%, respectively. Also, the precision, recall, and f-measure results in Table VII and Table VIII proved that tree interpreter outperforms fuzzy logic and correlation assessments.

TABLE VII.   COMPARISON OF RETRIEVAL STUDY RESULTS

Importance Criteria	Retrieval Accuracy	Precision	Recall	f-Measure
Tree Interpreter	82%	0.76611	0.34986	0.480356
Correlation	68%	0.5124	0.367341	0.427911
Fuzzy Logic	72%	0.691529	0.30156	0.419977

TABLE VIII.   COMPARISON OF ADAPTATION STUDY RESULTS

Importance Criteria	Adaptation Accuracy	Precision	Recall	f-Measure
Tree Interpreter	88%	0.7624	0.38519	0.511801
Correlation	77%	0.5841	0.38911	0.467071
Fuzzy Logic	78%	0.7518	0.3671	0.493316
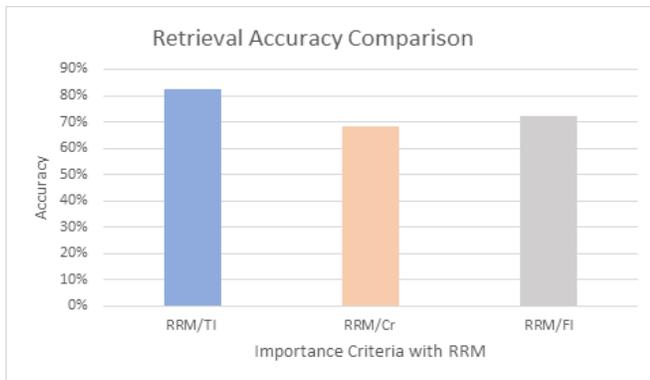
Fig. 7. Comparison of Retrieval Phase Study Accuracy Results.
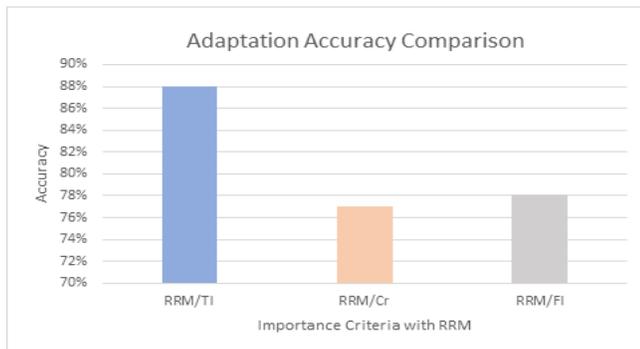
Fig. 8. Comparison of Adaptation Phase Study Accuracy Results.

Reference	Accuracy (%)				
	SVM	Random Forest	KNN	multimodal analysis	CBR
Bohang Chen et al. [42]	74.18	-	-		
Gonzalo A Ruz et al. [43]	81.2	72.5	-		
Priyanka Arora et al. [44]	79.7	84.6	-		
Akshi Kumara et al. [45]	-	81.04	-		
Shakeel Ahmed et al. [46]	-	-	72.0		
Safa, R., et al [47]	-	-	-	83%	
Proposed hybrid RF and CBR system	-	-	-	-	82

The proposed hybrid random forest and Case-based reasoning model showed comparable retrieval accuracy compared to other studies [40 - 47] aimed to depression detection using artificial intelligence and machine learning techniques as shown in Table IX.

VII.   CONCLUSION

This research concludes the best techniques and results for introducing a more accurate retrieved criteria and remedy recommender for depressed user of Twitter. The hybrid system of RF and CBR is a good technique for memory processing, but there were challenges found by previous studies in various fields that CBR system is built as a system for the retrieval of data only. The adaptation task in CBR system is left out, therefore, human interference is needed to complete system performance. To overthrow this dilemma, an automatic system – Remedy Recommendation using (RRM) was developed. This system detects the level of depression for each user and recommends the best solution. As explained in this study, the proposed system is applied to find the most similar cases and recommend remedies accordingly through applying adaptation phase to find best remedy recommended for each user. The system's successful hits exceeded 85% in its optimal hyper parameters configuration. RRM has proven that remedy recommendation done without the interference of human is possible. The main novelty of this work derives from introducing an adaptation task for depression treatment. RRM can be generalized to any type of illness to recommend the best solution using data from different social media platforms. In future work, this study aims to enrich the dataset with raw data collected from other social media platforms, such as Facebook or Snapchat.

REFERENCES

[1] Kuber, A., Kulthe, S., Kosamkar, P., "Detecting Depression in Tweets Using Natural Language Processing and Deep Learning." Lecture Notes in Networks and Systems, 2022, vol 400. Springer, Singapore.

[2] Kursuncu, Ugur, Manas Gaur, Usha Lokala, Krishnaprasad Thirunarayan, Amit Sheth, and I. Budak Arpinar. "Predictive Analysis on Twitter: Techniques and Applications." Lecture Notes in Social Networks, 2018, 67–104.

[3] Perner, Petra. "Mining Sparse and Big Data by Case-Based Reasoning." Procedia Computer Science 35 (2014): 19–33.

[4] Schott, Jonathan M. "Alzheimer's Disease and Other Dementias." Oxford Textbook of Medicine, 2020, 5830–59.

[5] Bach, Kerstin, and Klaus-Dieter Althoff. "Developing Case-Based Reasoning Applications Using Mycbr 3." Case-Based Reasoning Research and Development, 2012, 17–31.

[6] Ahmed, Mobyen Uddin, Shahina Begum, Erik Olsson, Ning Xiong, and Peter Funk. "Case-Based Reasoning for Medical and Industrial Decision Support Systems." Studies in Computational Intelligence, 2010, 7–52.

[7] R. Schmidt and L. Gierl, "Case-based reasoning for medical knowledge-based systems." Studies in Health Technology and Informatics, vol. 77, pp. 720–725, 2000.

[8] S. Nasiri, J. Zenkert, and M. Fathi, "A medical case-based reasoning approach using image classification and text information for recommendation." International Work-Conference on Artificial Neural Networks, 2015, pp. 43–55.

[9] S. Nasiri and M. Fathi, "Case Representation and Similarity Assessment in a Recommender System to Support Dementia Caregivers in Geriatric and Palliative Care." The International Conference on Case-Based Reasoning (ICCBR), 2017, pp. 157–166.

[10] K. Shrestha, "Machine Learning for Depression Diagnosis using Twitter data." International Journal of Computer Engineering in Research Trends, vol. 2, no. 5, pp. 56–61, 2018.

[11] H. S. AlSagri and M. Ykhlef, "Machine Learning-Based Approach for Depression Detection in Twitter Using Content and Activity Features." IEICE Transactions on Information and Systems, vol. 103, no. 8, pp. 1825–1832, 2020.

[12] H. Alsagri and M. Ykhlef, "Quantifying feature importance for detecting depression using random forest." International Journal of Advanced Computer Science and Applications, vol. 11, no. 5, pp. 628–635, 2020.

[13] V. E. Ekong, U. G. Inyang, and E. A. Onibere, "Intelligent decision support system for depression diagnosis based on neuro-fuzzy-CBR hybrid." Modern Applied Science, vol. 6, no. 7, p. 79, 2012.

[14] Y. Asim, B. Raza, A. K. Malik, A. R. Shahaid, and H. Alquhayz, "An adaptive model for identification of influential bloggers based on case-based reasoning using random forest." IEEE Access, vol. 7, pp. 87732–87749, 2019.

[15] N. Choudhury and S. A. Begum, "A survey on case-based reasoning in medicine," International Journal of Advanced Computer Science and Applications, vol. 7, no. 8, pp. 136–144, 2016.

[16] S. Begum, M. U. Ahmed, P. Funk, N. Xiong, and M. Folke, "Case-based reasoning systems in the health sciences: a survey of recent trends and developments," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 41, no. 4, pp. 421–434, 2010.

[17] B. Fuchs and A. Mille, "A knowledge-level task model of adaptation in case-based reasoning," International Conference on Case-Based Reasoning, 1999, pp. 118–131.

[18] I. Bichindaritz, "Semantic interoperability of case bases in biology and medicine," Artificial Intelligence in Medicine, (Special issue on case-based reasoning in the health sciences), vol. 36, no. 2, pp. 177–192, 2006.

[19] D. Brien, J. Glasgow, and D. Munoz, "The application of a case-based reasoning system to attention-deficit hyperactivity disorder," International Conference on Case-Based Reasoning, 2005, pp. 122–136.

[20] F. Díaz, F. Fdez-Riverola, and J. M. Corchado, "gene-CBR: A Case-Based Reasonig Tool for Cancer Diagnosis Using Microarray Data Sets," Computational Intelligence, vol. 22, no. 3-4, pp. 254–268, 2006.

[21] M. d'Aquin, J. Lieber, and A. Napoli, "Adaptation knowledge acquisition: A case study for case-based decision support in oncology," Computational Intelligence, vol. 22, no. 3-4, pp. 161–176, 2006.

[22] D. Glez-Peña, F. Díaz, J. M. Hernández, J. M. Corchado, and F. Fdez-Riverola, "geneCBR: multiple-microarray analysis and Internet gathering information with application for aiding diagnosis in cancer research," BMC Bioinformatics, vol. 10, pp. 187–194, 2009.

[23] S. Mulyana, S. Hartati, and R. Wardoyo, "Case Based Reasoning for Diagnosing Types of Mental Disorders and Their Treatments," International Conference on Soft Computing in Data Science, 2019, pp. 241–251.

[24] P. Bonissone and W. Cheetham, "Fuzzy case-based reasoning for residential property valuation," Handbook of Fuzzy Computation, vol. 15, 1998.

[25] W.-J. Wang, "New similarity measures on fuzzy sets and on elements," Fuzzy Sets and Systems, vol. 85, no. 3, pp. 305–309, 1997.

[26] G. Dvir, G. Langholz, and M. Schneider, "Matching attributes in a fuzzy case-based reasoning," 18th International Conference of the North American Fuzzy Information Processing Society-NAFIPS (Cat. No. 99TH8397), 1999, pp. 33–36.

[27] Houeland, Tor Gunnar. "An efficient random decision tree algorithm for case-based reasoning systems." Twenty-Fourth International FLAIRS Conference. 2011.

[28] W.-H. Hsieh, D.-H. Shih, P.-Y. Shih, and S.-B. Lin, "An ensemble classifier with case-based reasoning system for identifying internet addiction," International Journal of Environmental Research and Public Health, vol. 16, no. 7, p. 1233, 2019.

[29] D. Gu, C. Liang, and H. Zhao, "A case-based reasoning system based on weighted heterogeneous value distance metric for breast cancer diagnosis," Artificial Intelligence in Medicine, vol. 77, pp. 31–47, 2017.

[30] S. Mulyana, S. Hartati, R. Wardoyo, and E. Winarko, "Case-based reasoning with input text processing to diagnose mood [affective] disorders," International Journal of Advanced Research in Artificial Intelligence, vol. 4, no. 9, 2015.

[31] O. Kwon, Y. S. Kim, N. Lee, and Y. Jung, "When collective knowledge meets crowd knowledge in a smart city: A prediction method combining open data keyword analysis and case-based reasoning," Journal of Healthcare Engineering, vol. 2018, 2018.

[32] R. Rahim, W. Purba, M. Khairani, and R. Rosmawati, "Online Expert System for Diagnosis Psychological Disorders Using Case-Based Reasoning Method," Journal of Physics: Conference Series, 2019, vol. 1381, no. 1, p. 12044.

[33] M. Aldayel and M. Ykhlef, "A new sentiment case-based recommender," IEICE Transactions on Information and Systems, vol. 100, no. 7, pp. 1484–1493, 2017.

[34] D. Bridge, M. H. Göker, L. McGinty, and B. Smyth, "Case-based recommender systems," Knowledge Engineering Review, vol. 20, no. 3, p. 315, 2005.

[35] C.-S. Wang and H.-L. Yang, "A recommender mechanism based on case-based reasoning," Expert Systems with Applications, vol. 39, no. 4, pp. 4335–4343, 2012.

[36] F. Lorenzi and F. Ricci, "Case-based recommender systems: A unifying view," IJCAI Workshop on Intelligent Techniques for Web Personalization, 2003, pp. 89–113.

[37] Agnar Aamodt and Enric Plaza. 1994. Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Communications. 7, 1 (March 1994), 39–59.

[38] R. T. Macura and K. Macura, "Case-based reasoning: opportunities and applications in health care." Artificial intelligence in medicine (1997)., 9(1), 1-4.

[39] Rainer Schmidt, Bernhard Pollwein, and Lothar Gierl. "Experiences with Case-Based Reasoning Methods and Prototypes for Medical Knowledge-Based Systems", In Proceedings of the Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (1999): Springer-Verlag, Berlin, Heidelberg, 124–132.

[40] Holt, Alec, Isabelle Bichindaritz, Rainer Schmidt, and Petra Perner. "Medical applications in case-based reasoning." The Knowledge Engineering Review 20, no. 3 (2005): 289-292.

[41] Babu, N.V., Kanaga, E.G.M, "Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review." SN Computer Science. 3, 74 (2022).

[42] Chen B, Cheng L, Chen R, Huang Q, Phoebe Chen Y-P. "Deep neural networks for multiclass sentiment classification.",IEEE 20th International Conference on high performance computing and communications, IEEE 16th International Conference on Smart City, IEEE 4th International Conference on Data Science and Systems 2018; pp. 854–59.

[43] Ruz GA, Henriquez PA, Mascareno A. "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers.", Future Generation Computer Systems. 2020;106:92–104.

[44] Arora P, Arora P. "Mining Twitter data for depression detection.", IEEE International Conference on signal processing and communication (ICSC), 2019; pp. 186–89, https://doi.org/10.1109/ICSC45622.2019.8938353.

[45] Kumar A, Sharma A, Arora A. "Anxious depression prediction in real-time social data." International Conference on advanced engineering, science, management and technology, 2019 (ICAESMT19).

[46] Ahmad S, Asghar MZ, Alotaibi FM, Awan I. "Detection and classification of social media-based extremist affiliations using sentiment analysis techniques.", Human-centric Computing and Information Sciences. 2019;24:1–23.

[47] Safa, R., Bayat, P. and Moghtader, L. "Automatic detection of depression symptoms in twitter using multimodal analysis." The Journal of Supercomputing volume 78, 4709–4744 (2022).

A Novel Hierarchical Shape Analysis based on Sampling Point-Line Distance for Regular and Symmetry Shape Detection

Kehua Xian*

Department of Electro-mechanics and Information Engineering
Sichuan College of Architectural Technology
Sichuan, 618000, China

Abstract—**Regular and symmetry shapes occurred in natural and manufactured objects. Detecting these shapes are essential and still tricky task in computer vision. This paper proposes a novel hierarchical shape detection (HiSD) method, which consists of circularity and roundness detection, and regularity and symmetry detection phases. The first phase recognizes the circular and elliptical shapes using aspect ratio and roundness measurements. The second phase, the main phase in the HiSD, recognizes the regular and symmetry shapes using density distribution measurement (DDM) and the proposed sampling point-line distance distribution (SPLDD) algorithm. The proposed method presets effective with low computation cost shape detection approach which is not sensitive to specific category of objects. It enables to detect different types of objects involving the arbitrary, regular, and symmetry shapes. Experimental results show that the proposed method performs well compared to the existing state-of-the-art algorithms.**

Keywords—Shape recognition; hierarchical shape detection; sampling point-line distance distribution; regular and symmetry shape detection

I. INTRODUCTION

Regular and symmetry shapes occurred in the outdoor urban scenes, indoor built environments, and manufactured objects [1, 2, 24]. Detecting regular and symmetry shapes is essential in computer vision and pattern recognition communities. Regular and symmetry shape detection are broadly used in urban scene recognition [3, 4, 23], vehicle recognition [5, 6], face analysis [7, 8], image reconstruction [9], and driver assistance systems [10]. Moreover, the regular and symmetry shape features are considered salient features that guide eye movements and can thus mainly be used for visual attention detection [11].

Various methods have been developed for regular and symmetry shape detection. However, existing research works suffer from high complexity and it leads to deduct its applicability on real time shape detection applications [22]. As shown in literature, more existing methods focused on regular shapes detection which are limited to detect specific shapes such as circles or polygons. Moreover, these methods cannot detect symmetric shapes and differentiation from regular shapes.

Although many research works investigated regular shapes, the detection of polygon-based shapes has been studied less,

and little progress has been made in recent years [3]. Barnes and Loy [4] presented a posteriori probability approach that defined the continuous log-likelihood of the probability density function for the appearance of regular polygons. Liu and Wang [3] introduced point-line distance distribution (PLDD) to compute each pixel's shape energy for detecting arbitrary triangles, regular polygons, and circles. Among the existing shape detection methods, the PLDD method [3] presented a better performance in terms of accuracy than other existing methods. However, the PLDD method suffers several drawbacks. The main drawback of the PLDD method is the lack of ability to identify symmetry and parallelogram format shapes [3].

Referring to existing drawbacks of current shape detection methods, developing an efficient with low complexity visual shape detection method is required to cover various shapes detection consisting regular and symmetry and other shapes. Therefore, this study presents a novel shape detection method to deal with the challenges, to detect different types of shapes, including the arbitrary, regular, and symmetry shapes with effective performance and low complexity.

The main contributions of this study are as follows, 1) a novel shape detection method developed which is not sensitive to specific category of objects. It enables to detect different types of objects involving the arbitrary, regular, and symmetry shapes, 2) the proposed method is not based on high complexity and computation cost approaches which is applicable in real time applications, 3) the experiments and comparative analyses are performed using different shapes and noise conditions to validate the effectiveness of the proposed method.

The rest of this paper organizes as follows: related methods and the existing shape detection methods are briefly reviewed in Section II. The proposed shape detection method is described in Section III. Section IV presents the experimental results and performance evaluation. Finally, Section V presents the conclusion of the paper. Hierarchical Shape Detection (HiSD) Method.

II. RELATED WORKS

This section discusses existing research works on shape detection methods. The standard and symmetry shape detection methods are categorized into deterministic-based and

*Corresponding Author.

stochastic-based approaches [3]. For the deterministic-based approaches, Hough transform-based methods are commonly applied to infer the center locations, followed by employing edge information obtained by an edge detector [12]. However, the Hough transform-based methods require large storage space and high computational complexity, which results in low processing efficiency. Furthermore, template matching [13] and least square-based methods have been proposed [14] for the discriminative-based approach. For the stochastic-based methods, Dehmeshki and Ye [15] proposed a Genetic Algorithm (GA)-based method for shape recognition that effectively detects regular shapes. However, the GA-based method cannot differentiate the regular and symmetry shapes. Furthermore, other stochastic-based methods have been proposed based on the random sample consensus technique [16] and simulated annealing [17].

III. THE PROPOSED METHOD

The proposed shape detection method in this study consists of two main processes, which are hierarchically structured, as shown in Fig. 1. The first process calculates the roundness and circularity of the shape to identify whether the shape is a circle, ellipse, or not circular. The second and main process aims to detect the regularity and symmetry of the shapes.
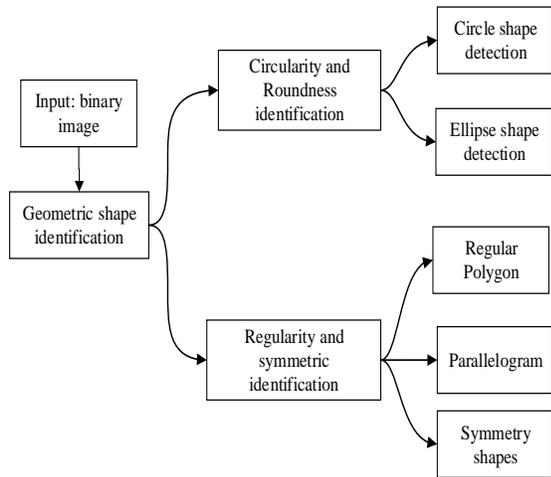
Fig. 1. The Hierarchical Shape Detection Method.

A. Circularity and Roundness Identification

Aspect ratio (*AR*) and roundness factors are used to detect the circularity of the shape. The *AR* is defined as the ratio of length over breadth [18]. *AR* can be calculated through Equation (1) as,

$$AR = \frac{L}{B} \tag{1}$$

where *L* and *B* are the length and breadth of the shape, respectively. Moreover, the roundness feature describes the shape's resemblance to a circle. The roundness factor of a shape will approach 1.0, the closer the shape resembles a circle [18]. The roundness factor (*Ro*) can be obtained by Equation (2),

$$Ro = \frac{1}{\pi}\left(\frac{4*EA}{((D))^2}\right) \tag{2}$$

where *EA* is the enclosed area of the shape and *D* is the diameter.

B. Regularity and Symmetry Identification

This section describes a proposed algorithm to distinguish the shape based on regularity, irregularity, symmetric and asymmetric features. The pseudocode of irregularity and symmetric identification is as follows

Input: *a binary component denotes as S.*
1- *Extract a corner point as P_c from S,*
2- *Compute angles as θ_k for each P_c using Equation (4),*
3- *Determine equal P_c points, initialize N_{eq} and detect shape regularity of S and update the S(T) function,*
4- *For each two connected P_c do:*
 a. *Find mean point as P_m,*
 b. *Generate a perpendicular line as BL that cross the P_m,*
 c. *Measuring the density distribution for two side of BL using calculate integral of are mass based Equation (9) and update symmetry status,*
 d. *Access the symmetry status of the using SPLDD and update symmetry status,*
5- *Distinguish the shape using obtained status*

Output: *shape detection results*

1) Corner points extraction: A corner point is an edge point that occurs when the edge direction changes. In this study, features from the accelerated segment test (FAST) corner detector are used to extract the corner points [19]. FAST corner detector uses a circle of 16 pixels to classify whether a candidate point p is a corner, as presented in [19]. Each pixel in the circle is labeled from integer 1 to 16 clockwise. Suppose a set of N contiguous pixels in the circle are all brighter than the intensity of candidate pixel p (denoted by Ip) plus a threshold value t or all darker than the intensity of candidate pixel p minus threshold value t. In that case, p is classified as a corner. In a mathematic demonstration, for each location on the circle x, the pixel at that position relative to p (denoted by p → x) can have one of three states by Equation:

$$S_{p\to x} = \{d, \quad I_{p\to x} < I_p - t \quad (darker) \; d, \quad I_p - t < I_{p\to x} < I_p + t \quad (similar) \; b, \quad I_p + t \le I_{p\to x} \quad (brighter) \tag{3}$$

Choosing an *x* and computing $Sp{\to}x$ for all $p \in P$ (the set of all pixels in all training images) partitions *P* into three subsets, *Pd, Ps, and Pb*, where each *p* is assigned to [19]. As an example, a pentagon is used to demonstrate the steps. The FAST corner detector can detect the corners of the pentagon, as shown in Fig. 2.
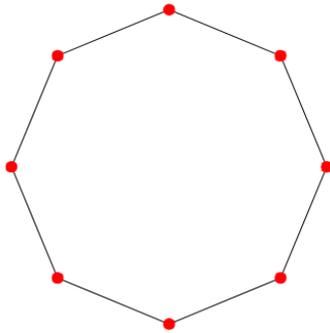
Fig. 2. Detected Corners of a Pentagon Shape.

2) Corners angles computation: As stated in the shape detection algorithm, the angle for each corner is required to identify. A scalar product-based approach is used for corer angle calculation. In this regard, a 5*5 mask is defined to identify the angle for the corner points. The mask is used for intersection point identification, obtained using corner points and sides on the shape. For this identification, the mask is located on the corner point, and the angle is calculated using the extraction of intersection points between the mask and the sides of the shape, as shown in Fig. 3 (P_c is the corner point and P_i, P_j are the intersection points on the shape sides). The mask contains 0 for all elements in M except M_{33} which is 1.

$$M = [0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0]$$

Let us define $P_c = (x_c, y_c)$ as a corner point, $P_i = (x_i, y_i)$ and $P_j = (x_j, y_j)$ as intersection points between mask (M) and polygon sides. The angle for corner point can be obtained using the scalar measurement [20],

$$\theta_k = arccos\left(\frac{\overrightarrow{P_iP_c}.\overrightarrow{P_cP_j}}{\|\overrightarrow{P_iP_c}\|\|\overrightarrow{P_cP_j}\|}\right) \qquad (4)$$

where P_c denotes corner point, P_i and P_j denotes intersection points between mask and polygon sides.
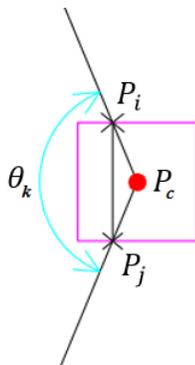
Fig. 3. The Corner Points (P_c) and Intersection Points (P_i, P_j) for Corner Angle Calculation (θ_k).

3) Shape regularity detection: After the corner angle computation, it is required to identify the regularity of the shape. In this regard, a function defines a rule as shown in Equation (5),

$$S(T) = \{(N_{eq} \cong N_T) \text{ and } \left((D(P_{c_i}) \cong D(P_{c_{i+1}})) \forall|u_{i\,N_T}\right) \rightarrow$$
$$S \in Regular \quad (N_{eq} \cong 4) \text{ and } \left(N_{eq} \cong \left[\frac{N_{eq}}{2}\right]\right) \rightarrow S \in$$
$$parallelogram \qquad (5)$$

According the function definition, *S(T)* is defined as the function for determination of shape type. The types of the shape type are regular or parallelogram. This type can be identified using a number of equal angle corners (N_{eq}). As the first statement in *S(T),* If the N_{eq} is approximately equal to total number of corner (N_T) and the distances between corner points are roughly equal, the shape determined as regular shape. Otherwise, if the number of equal angle corners is approximately equal to four and half number of equal angle corners, the shape is determined as a parallelogram.

Euclidean distance is used for the distance measurement of corner points. For instance, the Euclidean distance between *i-th* corner points as P_{c_i} and its neighbor as $P_{c_{i+1}}$ is calculated and denotes as $D(P_{c_i})$. Furthermore, a shape is detected as a parallelogram when N_{eq} is equal to a minimum number of $N_T/2$. For example, Fig. 4 shows a regular polygon shape in which the number of equal corners is equivalent to the number of total corners points, and the distances between all the corner points are the same.
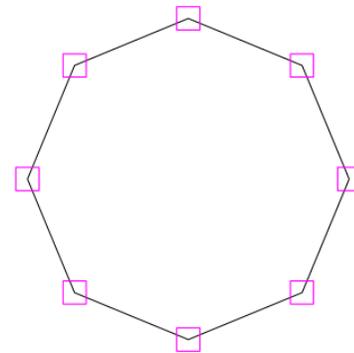
Fig. 4. A Regular Polygon with Same Number of Equal Corners and the Distances between the Corner Points.

4) Symmetry shape detection: Sometimes the shape is not detected as regular and parallelogram. In this case, the symmetric feature of the shape is checked. Fig. 5 illustrates a non-regular symmetric shape.
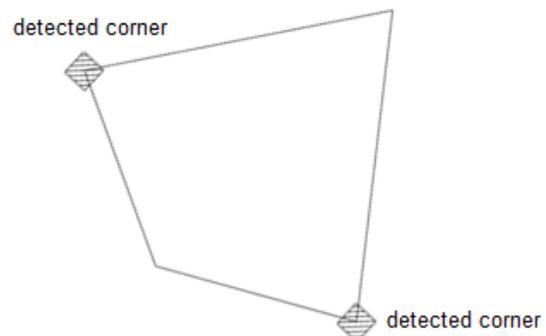
Fig. 5. Illustration of a Symmetric Shape.

In the first step, two corresponding corner points are connected through a line, and the mean point of this line is determined P_m. Based on identified P_m, a perpendicular line is drawn and named baseline *(BL)*. Using the *BL*, two correspondence corner points are identified. The length of *BL* is denoted as L_{bl}, which is measured by measuring the longest line connecting the two farthest points from the *S* perimeter. Furthermore, the details of the algorithm are demonstrated in a synthetic symmetry and asymmetry shape as shown in Fig. 6 and Fig. 7.

a) Density Distribution Measurement (DDM)

The *BL* is devised the *S* into two sections (it is called two sides in a shape which is divided in terms of on *BL*). The individual side is defined as a density function as $f(x, y)$. The *DDM* of both sides of the shape is calculated to measure the area side. The *DDM* is based on calculating integral under each space of *BL* sides. A density distribution function using a two-dimensional Cartesian moment m_{pq}, is the order of $p + q$, as $f(x, y)$, can be defined as,

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^p f(x, y) \, dx \, dy \tag{6}$$

The two-dimensional moment for a $(N \times M)$ discretized image, $g(x, y)$, is:

$$m_{pq} = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} x^p y^p g(x, y) \tag{7}$$

For a given intensity distribution $g(x, y)$ the image moments is defined. A complete moment set of order n consists of all moments m, such that $p + q \leq n$ and contains $1/2(n+1)$ $(n+2)$ elements. The definition of the zeroth order moment of the distribution, f(x, y),

$$m_{00} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy \tag{8}$$

where m_{00} represents the total mass of the given distribution function or section. The zeroth moment represents the total region area. Therefore, the mass of each side can be measured using zeroth moment. After measuring area mass for each side, the obtained values are compared to identify whether they are close or not. A threshold value defines an acceptable range for the obtained values area mass to distinguish the shape as symmetric.

Fig. 6 demonstrates the strategy of regularity and symmetry shape detection shape. As shown in the Fig. 6, the corners of the shape are firstly detected. Secondly, mean metric is calculated using Euclidian distance for each corner pairs as shown in Fig. 6(b). Then a base-line between each corresponding corner points are drawn which intends to divide two are of region for each generated sides of the base-line as shown in Fig. 6(c). Finally, area feature is calculated for each region to check the equality of them. Similar to Fig. 6, the furthest corners of the shape are detected in first step. Then base-line and area region are considered for symmetry measurement and identification.
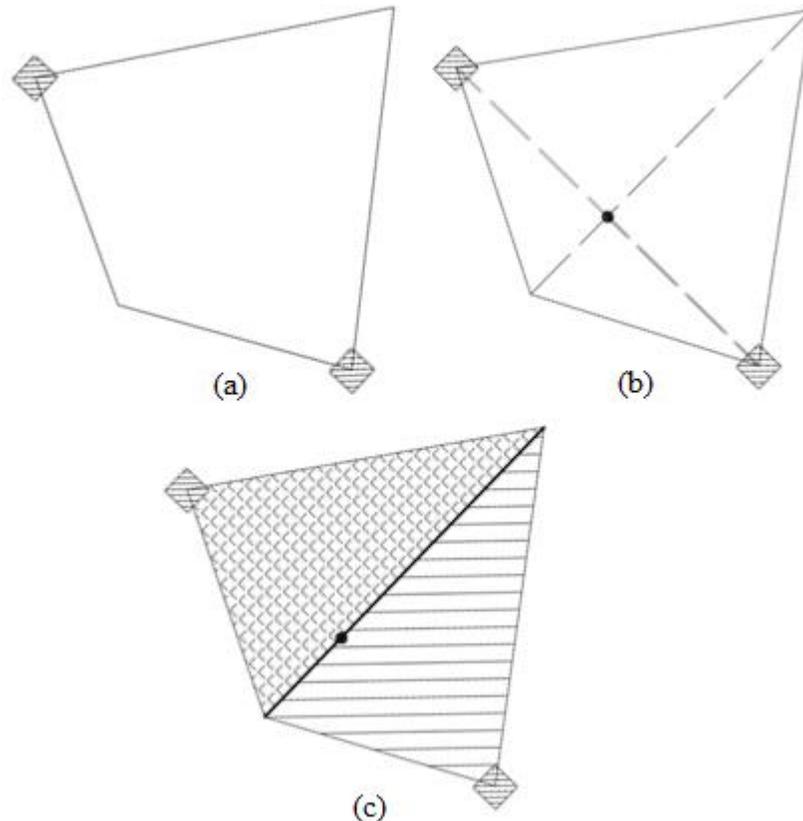
Fig. 6. Demonstration of Shape Regularity and Symmetric Identification for a Symmetric Shape. (a): Detected Corner, (b): Obtain Mean between Two Corner Points (P_m), (c): Calculate the Area for Region of each Sides.
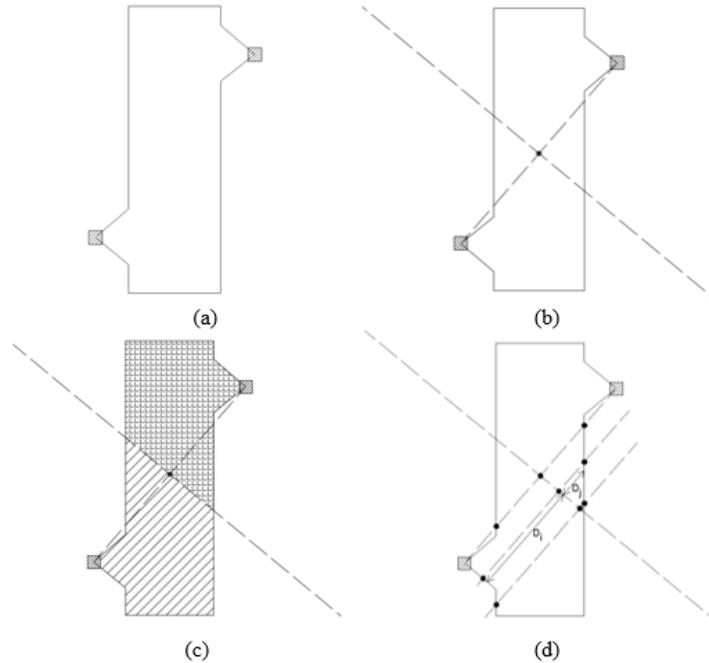
Fig. 7. The Process for Asymmetric Shape Identification.

b) Sampling Point-Line Distance Distribution (SPLDD)

Based on the DDM process, most of the shapes can be detected for symmetry shape identification. However, in some cases like Fig. 7, although the area for two sides of *BL* (the output of the DDM process) are quite similar, they are not symmetric shapes. Meanwhile, using the DDM, they are detected as symmetric shapes though they are not symmetrical.

Our shape detection method considers an algorithm to deal with the *DDM* challenge. The proposed algorithm is adopted from the point line distance distribution (*PLDD*) algorithm [3]. This algorithm is improved with sampling consideration in the *PLDD* algorithm and named Sampling Point-line Distance Distribution (*SPLDD*). The steps of *SPLDD* are as follows: finding the mean point (*Pm*), generating the baseline (*BL*), evaluating the area located on both side of *BL*.

In the *SPLDD* algorithm, as shown in Fig. 8, a set of sample points as P_s are generated on *BL*. In order to generate the P_s points, it is necessary to identify how many P_s are required to allocate on *BL*. For a number of sample points' identification, the length of *BL* (L_{bl}) is considered. Number of P_s denotes as ∂. Different values are tested for ∂ to find the optimum value. In this study, the value of ∂ is identified experimentally, which is $\partial = 10$. Using the value of ∂, the sample points (P_s) are generated. These points are used to draw confluence lines (*CL*). The confluence lines (*CL*) are defined to find the intersection points on *BL*. The *CL* lines touch the sides of the shape *(S)* to find the intersection points around the shape. The intersection points from the *CL* lines are marked as confluences points *(CP)*. Fig. 8 illustrates the details of the proposed *SPLDD* algorithm.
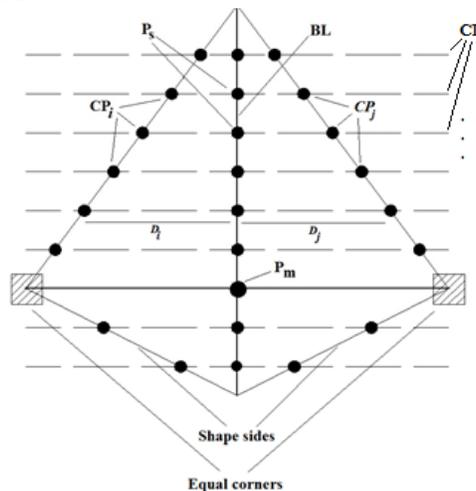
Fig. 8. Graphical Demonstration of *SPLDD* Algorithm.

Since the *CL* lines have an intersection in two opposite sites, two individual points (*CP*) are defined as CP_i (x_i, y_i), CP_j (x_j, y_j). The *CL* lines cross P_s point that defines as P_s (x_s, y_s). The distance between P_s , CP_i and P_s, CP_j (these points are located on a *CL)* are calculated and denoted as D_i and D_j are calculated. This distance measurement uses a Euclidean distance measurement with the following equations,

$$d_i(X) = \sqrt{(x_s - x_i)^2 + (y_s - y_{ij})^2}, \ d_j(Y) =$$

$$\sqrt{(x_s - x_j)^2 + (y_s - y_j)^2} \qquad (9)$$

where the $d(x)$ and $d(y)$ are calculated individually for three points associated with *CL*, however, it is required to perform the distance measurements for all *CP* and *CL*. In this regard, two individual matrices are defined to store all the distance measurements:

$$D(X) = (d_1 (X) \ d_2 (X) \ d_3 (X) \ \vdots \ d_n (X) \,)$$

$$D(Y) = (d_1(Y) \ d_2(Y) \ d_3(Y) \ \vdots \ d_n(Y) \,) \qquad (10)$$

where $d(X)$ and $d(Y)$ contains all the distances between the *CP* and P_s points those are located on the associated *CL*. Using Equation (11), the elements from a matrix are compared to the corresponding distance from another matrix. Consequently, based on the distance comparison, the equal distance of elements is identified, and corresponding points are labeled as True Points (*TP*).

$$\forall \big(d_k (X) \in D(X) \, , d_k (Y) \in D(Y) \big) \qquad (11)$$

$$TP(p_s) = \{ \ p_s \rightarrow true, \qquad d_k (X) \approx d_k (Y), \ p_s \rightarrow false,$$
$$d_k (X) \approx d_k (Y)$$

where the function $TP(p_s)$ contains the true points labeled of p_s points. $d_k (X)$ and $d_k (Y)$ the distance measurements from $d(X)$ and $d(Y)$ matrices. Finally, a number of true points are counted to recognize whether the shape is symmetry or asymmetry. If the number of true points is equal to a defined threshold value, the shape recognizes as symmetry; else, it is asymmetry.

C. Experimental Results and Performance Evaluation

This section presents the experimental results and performance evaluation for the proposed method. For experiments, sample images are firstly generated using addition and applying some image processing operators, including noise and blurring. Secondly, two measurements: the number of correct and false detection variables, are used to measure the performance of the proposed method. Then, the evaluation analysis of the experimental results is presented. The evaluation analysis consists of performance and processing-time measurements.

As mentioned earlier, the sample images are generated based on PLDD experiments [3] for the shape detection experiments. For all the sample images, an image consists of different types of shapes (including regular and symmetry shapes): circles, squares, and regular triangles. Furthermore, for the sample images, Gaussian noise, Pepper & Salt noise, Gaussian blur, and JEPG compression are applied to generate the images for experiments. The generated sample images are organized into different categories-based range of noises scale and compression ratio, as shown in Table I. For example, two generated sample images from Gaussian noise and blur categories are shown in Fig. 9.

TABLE I. GENERATED OF SAMPLE IMAGES UNDER DIFFERENT CATEGORIES

Category	Operator type	Parameter	Range
1	Gaussian noises	Noise variation	0.01–0.04
2	Pepper & Salt noise	Noise density	0.01–0.06
3	Gaussian blur	Scale	0.5 to 3
4	JPEG compressed image	Compression ratio	from 1.33 to 6.02

(a)   (b)
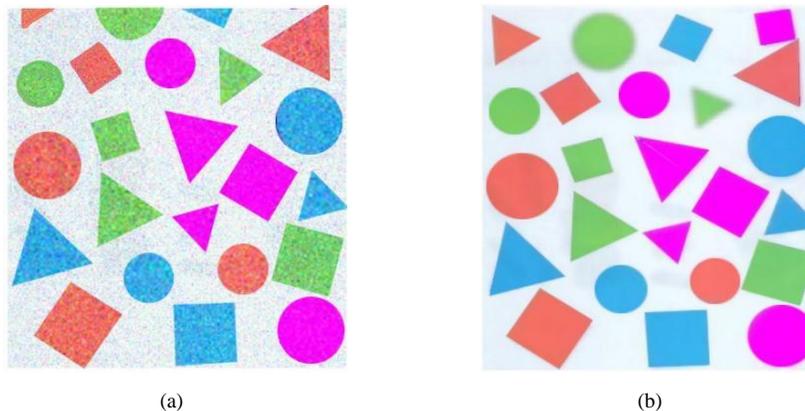
Fig. 9. The Images from Dataset for Shape Detection Experiments and Evaluation (a) Gaussian Noise (Variance = 0.01), (b) Gaussian Blur (Scale = 2.5) [3].

D. Performance Evaluation

This section presents the performance evaluation for the proposed shape detection method. As presented in [3], two measurements are used to evaluate the performance. The measurements are the number of corrected shapes (*CN*) (true detection) and the number of false ones (*FN*) (false detection). The sum of the *CN* and *FN* are calculated using Equations (12) and (12),

$$CN = \sum_{i=1}^{N} N_T$$

$$FN = \sum_{i=1}^{N} N_F \qquad (12)$$

$$Acc = \sum_{i=1}^{N} \frac{CN}{CN+FN}, \qquad (13)$$

where N_T and N_F are represented for true and false measurements. The *Acc* denotes accuracy measurement. The measurements are applied in the dataset containing the generated sample images considering defined noise categories, as shown in Table I. Then, the obtained results from the proposed method are compared to other methods, including the PLDD [3] and radial symmetry nature and direction-discriminated voting (RSNDV) [21] method, as shown in Table II.

As shown in Table II and Table III, in the noises categories, including Gaussian and Pepper & Salt noises, all methods, with increasing noises, the performance would be decreased due to decrease true detection and increased false detection. Based on the comparison, the proposed method achieved higher performance than other methods. The PLDD method also presented better performance when compared to the RSNDV method. The poor performance is because the method uses contour points for shape detection. The RSNDV method may miss some corner points that lead to false detection of shapes.

As shown in Table IV, in contrast to other categories, for both CN and FN measurements, the RSNDV method is the prior method compared to the proposed and PLDD methods. The reason for this can be using the voting approach and contour points features for shape detection in the RSNDV method. Using these features, the shape detector (RSNDV method) can detect shapes more accurately. However, in noise categories, the results are not the same with blur cases.

As shown in Table V, the proposed method roughly achieved better performance than other methods. In high ratio compression, the RSNDV method achieved higher performance in a number of correct shape detection; meanwhile, the PLDD and our proposed method archived promising results in lower compression ratio. However, with a lower compression ratio, our proposed method presented a better performance in false detection. Furthermore, the average processing time for the categories is calculated, and a comparison between our proposed method and other methods is illustrated. To calculate the average processing time, Equation (14) is used,

$$T_a = \frac{1}{N}\sum_{i=1}^{N} T_i \qquad (14)$$

where T_a indicates the average processing time for each category, an individual category's processing time (for example, the processing time for each Gaussian noise), and the number of experiments for each category. Table VI shows the comparison of average processing time for PLDD [3], RSNDV method, and the proposed method.

The processing time presented in Table VI for PLDD and RNSDV is based on Liu and Wang's [3] work. As shown in Table VI, the processing time is significantly reduced to be helpful in real-time applications. It is achieved by using the new SPLDD method, which is based on a sampling of points between the corner points instead of whole contour information.

TABLE II. PERFORMANCE COMPARISON BASED ON GAUSSIAN NOISE (CATEGORY 1)

Methods		Noise Range						Acc
		0.01	0.02	0.025	0.03	0.035	0.04	
PLDD	CN	24	24	23	21	21	18	0.86
	FN	0	2	3	4	4	7	
RSNDV	CN	23	21	23	20	18	19	0.62
	FN	1	10	4	13	27	20	
HiSD	CN	24	23	23	22	20	20	0.92
	FN	0	1	1	2	2	4	

TABLE III. PERFORMANCE COMPARISON BASED ON SALT & PEPPER NOISE (CATEGORY 2)

Methods		Noise Range						Acc
		0.01	0.02	0.03	0.04	0.05	0.06	
PLDD	CN	24	24	23	22	21	19	0.87
	FN	0	0	3	3	8	5	
Gang	CN	24	24	23	23	20	21	0.79
	FN	0	01	3	8	9	14	
HiSD	CN	24	24	23	23	21	20	0.88
	FN	0	0	2	4	6	7	

TABLE IV. PERFORMANCE COMPARISON BASED ON GAUSSIAN BLUR (CATEGORY 3)

Methods		Blur Variation Range						Acc
		0.5	1	1.5	2	2.5	3	
PLDD	CN	24	24	22	22	22	22	0.94
	FN	0	0	2	2	2	2	
RSNDV	CN	24	23	24	24	24	24	0.99
	FN	0	1	0	0	0	0	
HiSD	CN	24	24	23	23	23	23	0.93
	FN	0	0	1	2	3	3	

TABLE V. PERFORMANCE COMPARISON BASED ON JPEG COMPRESSION (CATEGORY 4)

Methods		Ratio Variable					Acc
		1.33	2.14	2.95	4.88	6.02	
PLDD	CN	24	24	24	23	14	0.90
	FN	0	0	0	2	10	
RSNDV	CN	22	23	24	21	21	0.91
	FN	2	1	0	3	6	
HiSD	CN	24	24	24	23	20	0.95
	FN	0	0	0	1	4	

TABLE VI.   COMPARISON OF AVERAGE PROCESSING TIME FOR PLDD, RSNDV METHOD AND THE PROPOSED METHOD (IN TERMS OF SECONDS (S))

8	Gaussian noise	Pepper & Salt	Gaussian blur	JPEG compression
PLDD	24.39	26.73	24.29	25.27
RSNDV	163.68	146.48	70.58	88.77
HiSD	3.21	4.74	3.98	4.43

IV.   CONCLUSION

This paper presented a shape detection method based on hierarchical and sampling point line distance distribution algorithms. The fundamental concepts of different shapes are first defined. The existing shape detection methods are reviewed, and the recent and relevant study is addressed. Based on the literature review, the PLDD method has shown promising results in arbitrary and regular shape detection. However, this method suffers from some challenges, such as a lack of efficiency in shape detection in different types, including rotationally convex symmetry, ellipse, and parallelogram. Therefore, this study proposed a new shape detection method called hierarchical shape detection (HiSD) to detect the shapes in hierarchical processes. In HiSD, a new algorithm with consideration of sampling point line distance distribution (SPLDD) algorithm was proposed to deal with PLDD challenges. The proposed shape detection method is not sensitive to specific category of objects. It enables to detect different types of objects involving the arbitrary, regular, and symmetry shapes. Furthermore, the proposed method is not based on high complexity and computation cost approaches which is applicable in real time applications. The experiments and comparative analyses are performed using different shapes and noise conditions to validate the effectiveness of the proposed method. The experimental results showed that the proposed shape detection method enables detection of different types of shapes, including circle, eclipse, regular, and symmetry shapes in different noise condition scenarios. For future works, this proposed method can be extended for irregular shape detection as they typically exist in real environment. Moreover, the proposed system can be implemented in real-time shape detection to extend the application of current research work.

REFERENCES

[1]   Akbar, H., et al., Bilateral symmetry detection on the basis of Scale Invariant Feature Transform. PloS one, 2014. 9(8): p. e103561.

[2]   Korman, S., et al. Probably approximately symmetric: Fast rigid symmetry detection with global guarantees. in Computer Graphics Forum. 2015. Wiley Online Library.

[3]   Liu, H. and Z. Wang, PLDD: Point-lines distance distribution for detection of arbitrary triangles, regular polygons and circles. Journal of Visual Communication and Image Representation, 2014. 25(2): p. 273-284.

[4]   Barnes, N., G. Loy, and D. Shaw, The regular polygon detector. Pattern Recognition, 2010. 43(3): p. 592-602.

[5]   Teoh, S.S. and T. Bräunl, Symmetry-based monocular vehicle detection system. Machine Vision and Applications, 2012. 23(5): p. 831-842.

[6]   Hsieh, J.-W., L.-C. Chen, and D.-Y. Chen, Symmetrical surf and its applications to vehicle detection and vehicle make and model recognition. Intelligent Transportation Systems, IEEE Transactions on, 2014. 15(1): p. 6-20.

[7]   Xu, Y., et al., Approximately symmetrical face images for image preprocessing in face recognition and sparse representation based classification. Pattern Recognition, 2016. 54: p. 68-82.

[8]   Saha, S. and S. Bandyopadhyay. A symmetry based face detection technique. in Proceedings of the IEEE WIE National Symposium on Emerging Technologies. 2007.

[9]   Chou, C.-Y., et al., Accelerating Image Reconstruction in Dual-Head PET System by GPU and Symmetry Properties. PloS one, 2012. 7(12): p. e50540.

[10]   Loy, G. and N. Barnes. Fast shape-based road sign detection for a driver assistance system. in Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on. 2004. IEEE.

[11]   Kootstra, G., A. Nederveen, and B. De Boer. *Paying attention to symmetry*. in *British Machine Vision Conference (BMVC2008)*. 2008. The British Machine Vision Association and Society for Pattern Recognition.

[12]   Kiryati, N., H. Kälviäinen, and S. Alaoutinen, *Randomized or probabilistic Hough transform: unified performance evaluation.* Pattern Recognition Letters, 2000. 21(13): p. 1157-1164.

[13]   Qi, H., et al., An effective solution for trademark image retrieval by combining shape description and feature matching. Pattern Recognition, 2010. 43(6): p. 2017-2027.

[14]   Chaudhuri, D., A simple least squares method for fitting of ellipses and circles depends on border points of a two-tone image and their 3-D extensions. Pattern Recognition Letters, 2010. 31(9): p. 818-829.

[15]   Dehmeshki, J., et al., *Automated detection of lung nodules in CT images using shape-based genetic algorithm.* Computerized Medical Imaging and Graphics, 2007. 31(6): p. 408-417.

[16]   Fischler, M.A. and R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 1981. 24(6): p. 381-395.

[17]   Birbil, Ş.İ., S.-C. Fang, and R.-L. Sheu, *On the convergence of a population-based global optimization algorithm.* Journal of global optimization, 2004. 30(2-3): p. 301-318.

[18]   Image Metrology, S. *Shape Measurement Parameters*. 2015 15 Feb 2016]; Available from: http://www.imagemet.com/WebHelp6/Default. htm#PnPParameters/Measure_Shape_Parameters.htm.

[19]   Rosten, E. and T. Drummond, Machine learning for high-speed corner detection, in Computer Vision–ECCV 2006. 2006, Springer. p. 430-443.

[20]   Proofwiki. *Angle Between Vectors in Terms of Dot Product*. 2015; Available from: https://proofwiki.org/wiki/Angle_Between_Vectors _in_Terms_of_Dot_Product.

[21]   Gang, W., et al. A Shape Detection Method Based on the Radial Symmetry Nature and Direction-Discriminated Voting. in Image Processing, 2007. ICIP 2007. IEEE International Conference on. 2007.

[22]   Rampone, G., Makin, A.D., Tyson-Carr, J. and Bertamini, M., 2021. Spinning objects and partial occlusion: Smart neural responses to symmetry. Vision research, 188, pp.1-9.

[23]   Xu, J., Cao, W., Liu, B. and Jiang, K., 2021. Object restoration based on extrinsic reflective symmetry plane detection. The Visual Computer, pp.1-16.

[24]   Bartalucci, C., Furferi, R., Governi, L. and Volpe, Y., 2018. A survey of methods for symmetry detection on 3d high point density models in biomedicine. Symmetry, 10(7), p.263.

Development of Automatic Segmentation Techniques using Convolutional Neural Networks to Differentiate Diabetic Foot Ulcers

R V Prakash[1]

Research Scholar
Department of Computer Science & Engineering
S.E.A College of Engineering & Technology
Bengaluru, Karnataka India

Dr. K Sundeep Kumar[2]

Professor
Department of Computer Science & Engineering
S.E.A College of Engineering & Technology
Bengaluru, Karnataka India

Abstract—The quality of computer vision systems to detect abnormalities in various medical imaging processes, such as dual-energy X-ray absorptiometry, magnetic resonance imaging (MRI), ultrasonography, and computed tomography, has significantly improved as a result of recent developments in the field of deep learning. There is discussion of current techniques and algorithms for identifying, categorizing, and detecting DFU. On the small datasets, a variety of techniques based on traditional machine learning and image processing are utilized to find the DFU. These literary works have kept their datasets and algorithms private. Therefore, the need for end-to-end automated systems that can identify DFU of all grades and stages is critical. The study's goals were to create new CNN-based automatic segmentation techniques to separate surrounding skin from DFU on full foot images because surrounding skin serves as a critical visual cue for evaluating the progression of DFU as well as to create reliable and portable deep learning techniques for localizing DFU that can be applied to mobile devices for remote monitoring. The second goal was to examine the various diabetic foot diseases in accordance with well-known medical categorization schemes. According to a computer vision viewpoint, the authors looked at the various DFU circumstances including site, infection, neuropathy, bacterial infection, area, and depth. Machine learning techniques have been utilized in this study to identify key DFU situations as ischemia and bacterial infection.

Keywords—Magnetic resonance imaging (MRI); diabetic foot ulcers (DFU); convolutional neural networks; ischemia& machine learning algorithms & dual-energy x-ray absorptiometry

I. INTRODUCTION

Hyperglycemia (high blood sugar levels) is a chronic illness that causes diabetes. Diabetes can cause serious, life-threatening consequences such renal failure, cardiovascular disease, blindness, and lower limb amputation, which is frequently followed by DFU [6]. According to the World Health Organization's worldwide report on diabetes in 2020, there were 522 million people living with DM in 2018, up from 108 million in 1980. Global incidence among people over the age of 18 increased from 4.7% in 1980 to 8.5% in 2014. According to estimates, there will be 600 million individuals living with DM globally by the end of 2035. According to this estimate, only 20% of these individuals will come from industrialized nations, with the remainder coming from poorer

nations due to low awareness and a lack of healthcare resources [9]. A diabetic patient has a 15%–25% probability of developing DFU at some point, and if necessary precautions are not followed, that might lead to lower limb amputation [8] [10] [15]. However a more recent research [11] [12] suggests higher rates of up to 34%. It has been noted that more than one million patients suffer on average each year [30]. Several significant duties are carried out during the early diagnosis for the evaluation of DFU by monitoring the course of the condition and the number of time-consuming steps made in its treatment and management for each individual case:

1) The examination of the patient's medical history is performed.

2) A specialist in wound or diabetic foot studies the DFU in depth [13] and.

3) Additional tests, such as MRI, CT scans, X-rays, etc., may be helpful in developing the treatment plan.

Leg swelling is a common issue for DFU sufferers, and depending on the specific instance, it may also be itchy and unpleasant. The DFU often has erratic architecture and ambiguous exterior limits. Different phases, such as callus development, redness, blisters, and important tissue types including slough, scaly skin, granulation, and bleeding, all have an impact on the aesthetic appearance of DFU and the skin around it. Clinicians now monitor patients in healthcare settings mostly by visual inspection in order to identify critical problems including infection, area, depth, ischemia, neuropathy, and placement. The likelihood of an infection spreading by DFU is always increased. Therefore, patients must frequently attend healthcare facilities for DFU examination, adding to the cost burden on both patients and healthcare facilities.

The body of existing material on DFU assessment by adaptation of computerized algorithms is still in its infancy. There aren't many computerized approaches for the evaluation of diabetic foot diseases since this way of DFU analysis applying computerized methods is still a relatively new subject. Instead, simple image processing and conventional machine learning are used. The study of computer vision has advanced quickly in recent years, especially with regard to challenging and crucial problems such comprehending pictures from many

domains as spectral, non-medical items, anomalies in medical imaging, and face feature detection [16][19]. Particularly in the area of medical imaging, computer vision algorithms have made significant strides.

The quality of these computer vision systems in detecting abnormalities in different medical imaging, such as "Magnetic Resonance Imaging" (MRI), ultrasonography, dual-energy, X-ray absorptiometry, and computed tomography [26] has been significantly improved by the recent advancement in the field of deep learning. Recent advancements in computer vision and deep learning have made it possible for us to offer complete solutions for DFU recognition. Over a five-year period, a sizable dataset of DFU was gathered from several patients with varying backgrounds at the chosen hospitals in Bengaluru.

End-to-end algorithms, which have the potential to be used to adapt these algorithms to real-world contexts [7] [24] [25] are a major driving force behind the rapidly expanding study fields of medical imaging and computer vision. The capacity of algorithms to detect DFU of various stages and grades has been the main focus of DFU recognition research. Additionally, algorithms must be strong enough to identify DFU in patients with a range of ethical backgrounds. Then, by identifying crucial factors including site, area, depth, infection, and ischemia, a deeper understanding of DFU [18] might be offered. Therefore, creating reliable computer vision algorithms that could analyze the DFU with more accuracy and high precision has the potential to bring about a paradigm change in the treatment of diabetes patients' feet, which would be a cost-efficient, easy, and remote healthcare option.

II.   STATEMENT OF THE PROBLEM

Research in this topic is sparse and frequently exploratory rather than task-focused because it is still in its infancy. The current approaches and techniques for identifying, classifying, and detecting DFU are addressed. On the small datasets, a variety of techniques based on traditional machine learning and image processing are utilized to find the DFU. These literary works never released the datasets and algorithms they used. Therefore, end-to-end automated systems that can identify DFU at all stages and grades are required. It is important to note that no publicly accessible DFU databases are available for study [2] [4]. Modern computer identification methods and medical imaging are mostly dependent on deep learning models at this time. Neural networks are used in deep learning models to somewhat mimic how the human brain operates. Consequently, a sizable collection of DFU photos and professional annotations are required for training the deep learning models. Expert physicians are needed to do these annotations in order to provide ground truth, which makes them expensive. These visual expert annotations in the current DFU dataset are done out by podiatrists who are skilled in DFU. In addition, there may be additional contributing elements, such as the patient's ethnic background and the lighting circumstances.

III.   OBJECTIVES OF THE STUDY

1) To create novel CNN-based automated segmentation techniques to separate surrounding skin from DFU on complete foot pictures since surrounding skin serves as a crucial visual gauge for DFU development.

2) To create reliable and portable deep learning techniques for DFU localization that can be used in mobile devices for remote monitoring.

3) To examine the various diabetic foot diseases in accordance with widely used medical categorization schemes.

IV.   EXPECTED OUTCOME OF THIS STUDY

1) In complete foot photos, experts accurately outlined the DFU and the surrounding skin region. For the first time, the surrounding skin is segmented, which is a crucial sign for physicians evaluating the development of DFU. For the semantic segmentation [22] of DFU and the skin around it, we suggested using two-tier transfer learning segmentation techniques.

2) The huge DFU dataset of 1775 pictures and the Foot Snap dataset are used to evaluate cutting-edge deep learning localization techniques. For remote DFU monitoring, we adapted the durable and lightweight models to portable devices like the Nvidia Jetson TX2 and a Smartphone Android application.

3) According to a computer vision viewpoint, the authors looked at the various DFU circumstances including site, infection, neuropathy, bacterial infection, area, and depth. The identification of crucial DFU circumstances including ischemia and bacterial infection has been done in this research using machine learning methods.

V.   RESEARCH METHODOLOGY

The medical facilities for patients are becoming a major worry, especially for industrialized nations, because to the limited healthcare settings accessible, the growing global population, and the financial strain. The use of computerized telemedicine systems is frequently suggested as a potential remedy for this issue. When it comes to the creation of cutting-edge healthcare systems, the spread of "Information and Communication Technologies" (ICT) brings both opportunities and difficulties. Before establishing an efficient recognition system using DFU pictures, there may be a number of problems for any computerized DFU recognition algorithm that need to be overcome. Several research difficulties are identified based on a study of the current literature that are high levels of similarity within classes between healthy skin and abnormal classes (DFU) in the foot area [1] changes within classes based on the classification of DFU [3] lighting conditions; and ethnicity of the patient.

Another obstacle to accurate diagnosis of DFU is the variations in the visual appearance of the DFU and the skin around it, such as callus development, redness, and blisters, substantial tissue types including slough, granulation, bleeding, and scaly skin depending on the various phases. Because of this, using computer vision algorithms to analyze and recognize DFUs might be exceedingly difficult. It is a difficult annotation work for podiatrists to create ground facts for the segmentation of DFU and the skin around it since these

structures typically have ill-defined outside limits and highly irregular features.

The medical categorization systems say that there are currently no technical ways for identifying the result of DFU. Even for medical professionals, it can be challenging to identify and analyze DFU from a picture. Computer vision algorithms to detect DFU are increasing, but they still have a long way to go before they are as well-established as analysis according to medical categorization systems.

VI. METHODOLOGY

The suggested approach provides feature descriptors, classifiers, and Natural Data-Augmentation utilized in conventional machine learning. This also includes a brief summary of the experimental conditions and deep learning methodologies. To forecast the fate of DFU, the DFU are categorized medically according to many parameters, including size, area, neuropathy, ischemia, and infection. These systems now rely on the clinician's observations and clinical judgments.

These findings will explain why each goal was set and how it was accomplished. With the advancement of computer vision, particularly deep learning techniques, the diagnosis and detection of DFU by a computerized method has become an active study field. In this study, we studied the application of deep learning as well as traditional machine learning for the identification and analysis of DFU. With the help of a traditional machine learning approach, we managed to attain a respectable performance. However, this strategy is highly sluggish for DFU identification jobs because of the several intermediary stages. To identify DFU on the whole foot photos with high accuracy using deep learning, we employed various architectures to train end-to-end models on the DFU dataset with various hyper-parameter values. These techniques have a fast inference rate and can localize and segment numerous DFU. The DFU patch image its region identification of foot ulcer [17] is shown in Fig. 2(a) and Fig. 2(b).

VII. APPLICATION OF CONVOLUTIONAL NEURAL NETWORKS

Deep learning techniques are utilized to perform binary classification to categorize, namely (1) infection and non-infection; (2) ischemia and non-ischemia classes in DFU patches, in order to compare with the traditional characteristics. Modern CNN models like Inception-V3, ResNet50, and InceptionResNetV2 were fine-tuned (transfer learning from pre-trained models) for this purpose.

A revised version of the original Inception architecture, called Inception-V3, was created by the Google team and includes additional capabilities including enhanced normalization and the decomposition of larger convolution kernels[23][27][28] into a number of smaller convolution kernels. In this network, the earliest layers of the design employ depth-wise separable convolutions to speed up the calculations involved in down sampling the input pictures. In order to increase convergence during training, they also devised a batch normalization layer that may reduce internal covariate shift and address the gradient vanishing problem [31].
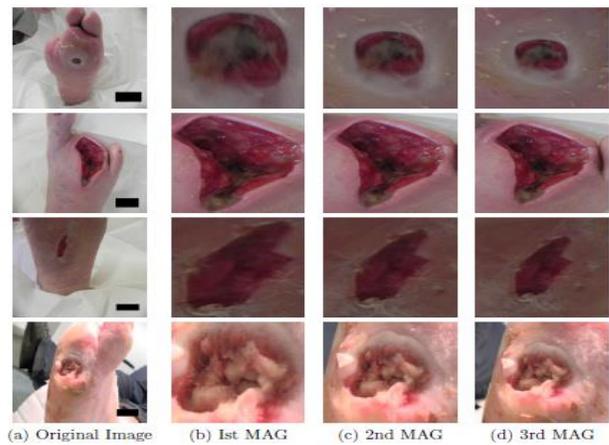
Fig. 1. Representation of Original Image of Various Types of Diabetic Foot Ulcer.
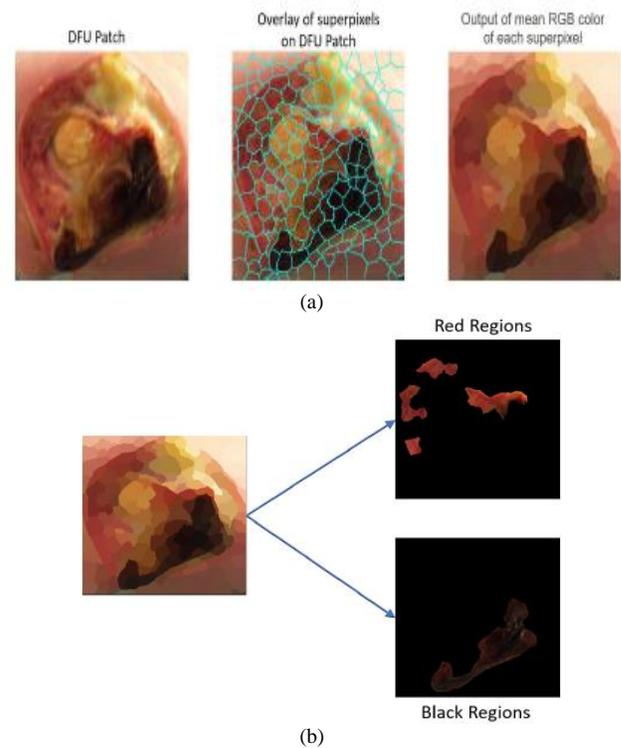
Fig. 2. (a) DFU Patch Image of Diabetic Foot Ulcer, (b) Representation of DFU Patch Image its Region Identification of Diabetic Foot Ulcer.

ResNet50 is a slimmer iteration of ResNet101, which took first place in the ILSVRC classification challenge. The fundamental principle of ResNet is to provide a short-cut link between layers in order to maximize representation from both the initial and subsequent layers during network training. Inception ResNetV2 is an extremely deep network that combines the advantages of residual and inception learning, networks for learning, as their name indicates. It is modeled after the InceptionV3 architecture and uses the remaining connections between the layers to train deeper neural networks [5] which must result in improved performance. On several typical datasets, it produced new, cutting-edge accuracy outcomes.

VIII. RESULTS AND DISCUSSION

We used the five-fold cross-validation approach and divided the datasets for infection and ischemia into three parts: training, validation, and testing [14]. Therefore, we utilized about 6909 patches, 987 patches, and 1974 patches in the training, validation, and testing sets for the ischemia dataset using the suggested approaches, but for the infection dataset, we used 4124 patches, 589 patches, and 1179 patches from the 1459 original foot photos. As previously noted, for the classification job, we employed both TML models and CNNs models, and the input for CML, InceptionV3, Alex Net, and ResNet50 was 256 x 256 RGB photos. The dataset for InceptionResNetV2 was scaled to 299 x 299 pixels.

As our assessment measures, we list Accuracy, Sensitivity, Precision, Specificity, F-Measure, and MCC in Tables I and II. Sensitivity and specificity are regarded as trustworthy assessment measures for classifier completeness in medical imaging [20] [21].

When comparing the results, the TML and CNN approaches outperformed the other methods in the binary classification of ischemia over infection. The average accuracy of all the models in the ischemia dataset is 82.1%, which is much better than the average accuracy in the infection dataset, which is 64.6%. With an average MCC Score for ischemia classification of 64.8% being higher than the infection

classification of 30.1%, MCC score is thought to be a realistic performance metric for the various machine learning algorithms for classification. CNNs fared better than TML models (85.2%) when their performances were compared (compared to 79% for TML models). The accuracy of CNNs (67%) outperformed TML (62.1%) in the infection classification task by a margin of 4.9%. ResNet50 earned the greatest overall score in the ischemia categorization.

1) Discussion and analysis of experiments for the restricted number of medical professionals and healthcare facilities, analysis of DFU situations using automated technologies is crucial. This study performs the preliminary experiment of binary categorization of ischemia and infection of DFU. The primary goal of this experiment is to determine the ischemia and infection situations when computer vision algorithms are most likely to make errors. Few instances of correctly and erroneously identified cases in both binary categories of ischemia and infection are presented in Fig. 1(a), (b), (c), and (d).

Regarding the incorrectly categorized instances, (1) infection and non-infection; (2) ischemia and non-ischemia cases in the DFU have significant intra-class differences and significant inter-class similarities, making it challenging for classifiers to forecast the correct class.

TABLE I. THE PERFORMANCE METRICS FOR THE BINARY CATEGORIZATION OF ISCHEMIA BY CNNs AND CONVENTIONAL MACHINE LEARNING ARE SHOWN IN THE TABLE, WHERE MCC STANDS FOR MATTHEW CORRELATION COEFFICIENT

Modern CNN Models	Accuracy	Sensitivity	Precision	Specify	F-Measure	MCC Score
Bayes Net	0.785 ± 0.022	0.774 ± 0.034	0.809 ± 0.034	0.800 ± 0.027	0.790 ± 0.020	0.572± 0.021
Random Forest	0.780 ± 0.041	0.739 ± 0.049	0.872 ± 0.029	0.842 ± 0.034	0.799 ± 0.033	0.571± 0.034
Multiplayer Perception	0.804 ± 0.022	0.817 ± 0.044	0.787 ± 0.046	0.795 ± 0.031	0.800 ± 0.023	0.610± 0.024
Inception V3 (CNN)	0.841 ± 0.017	0.785 ± 0.045	0.886 ± 0.018	0.898 ± 0.022	0.831 ± 0.021	0.688± 0.022
ResNet50 (CNN)	0.862 ± 0.018	0.797 ± 0.043	0.917 ± 0.015	0.927 ± 0.011	0.852 ± 0.023	0.732± 0.024
Inception ResNetV2 (CNN)	0.853 ± 0.021	0.789 ± 0.054	0.906 ± 0.017	0.917 ± 0.019	0.842 ± 0.027	0.714± 0.026

TABLE II.   THE RESULTS OF THE CLASSIFICATION OF THE INFECTION TASK USING CONVENTIONAL DEEP LEARNING AND CNNS. MATTHEW CORRELATION COEFFICIENT (MCC)

Modern CNN Models	Accuracy	Sensitivity	Precision	Specify	F-Measure	MCC Score
Bayes Net	0.639 ± 0.036	0.619 ± 0.018	0.653 ± 0.039	0.660 ± 0.015	0.622 ± 0.079	0.290± 0.080
Random Forest	0.605 ± 0.025	0.608 ± 0.025	0.607 ± 0.037	0.601 ± 0.069	0.606 ± 0.012	0.211± 0.013
Multiplayer Perception	0.621 ± 0.026	0.680 ± 0.023	0.622 ± 0.057	0.570 ± 0.023	0.627 ± 0.074	0.281± 0.075
Inception V3 (CNN)	0.662 ± 0.014	0.693 ± 0.033	0.653 ± 0.015	0.631 ± 0.034	0.672 ± 0.019	0.325± 0.020
ResNet50 (CNN)	0.673 ± 0.013	0.692 ± 0.051	0.668 ± 0.023	0.654 ± 0.051	0.679 ± 0.019	0.348± 0.021
Inception ResNetV2 (CNN)	0.676 ± 0.014	0.688 ± 0.052	0.672 ± 0.015	0.664 ± 0.033	0.680 ± 0.022	0.352± 0.022

Other variables that affect how these disorders are classified include the lighting, markings, tattoos, and skin tone as a result of the patient's ethnicity. As seen in Fig. 3(a) and Fig. 3(b) of misclassified non-ischemia is hampered by the illumination and the tattoo, respectively.

The ischemia traits in the incorrectly classified ischemia cases (c) and (d) are, in contrast, too subtle for the algorithm to detect. In Fig. 4, it can be shown that situations illustrated in (a) and (b) where blood is present are mistakenly labeled as non-infections; yet, the situation in (b) represents one of the dataset's unusual instances, namely the existence of ischemia and non-infection. The visual signs of illness in cases of misclassified infection were too subdued in these circumstances. The existing realities aren't supported by clinical trials or medical records; they're just based on specialists' eye inspections. Furthermore, prior to the recording of these photos, debridement was mocked with DFU images. As a result, the debridement of DFU eliminated the crucial visual markers of infection, such as colored exudates. Therefore, in the future, the sensitivity and specificity of these algorithms might be enhanced by feeding in ground truth from clinical tests like vascular evaluations (ischemia) and blood tests for detecting the existence of any such bacterial infection.

(a) (b) (c) (d)
Misclassified non-ischemia cases   Misclassified ischemia cases
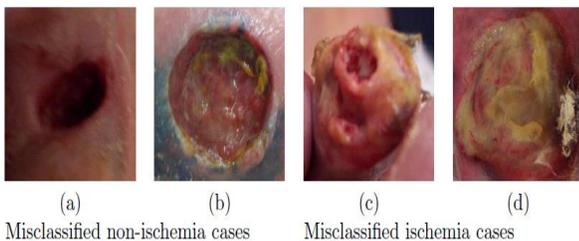
Fig. 3.   Representation of Misclassified Non-Ischemia Cases of Diabetic Foot Ulcer.

(a) (b) (c) (d)
Accurate non-ischemia cases   Accurate ischemia cases
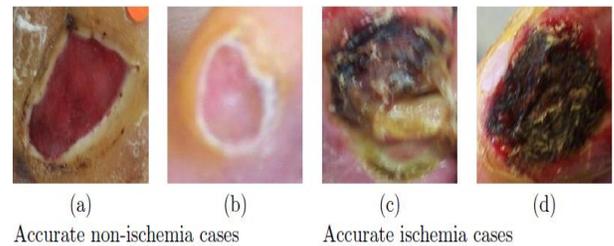
Fig. 4.   Representation of Accurate Non-Ischemia Cases and Accurate Ischemia Cases of Diabetic Foot Ulcer.

IX.   CONCLUSION

These conclusions will explain why each goal was set and how the result was attained. With the advancement of computer vision, particularly deep learning techniques, the diagnosis and detection of DFU by a computerized method has become an active study field. In this study, we studied the application of deep learning as well as traditional machine learning for the identification and analysis of DFU. With the help of a traditional machine learning approach, we managed to attain a respectable performance. Nevertheless, this strategy is highly sluggish for DFU identification tasks because of the several intermediary phases. To identify DFU on the whole foot photos with high accuracy using deep learning, we employed various architectures to train end-to-end models on the DFU dataset with various hyper-parameter values.

These techniques have a fast inference rate and can localize and segment numerous DFU. Then, to demonstrate how the localization techniques may be quickly moved to a portable device, the Nvidia Jetson TX2, to generate inference remotely, a demonstration was conducted. In order to provide real-time DFU localization [29] these deep learning techniques were finally applied in an Android application. In this work, we

developed mobile systems that can help patients and medical professionals in remote settings with DFU diagnosis and follow-up. Although the various algorithms for classification, segmentation, and localization approaches received extremely excellent accuracy, there were some restrictions on the detection of DFU in some specific situations, such as pre-ulcer circumstances and very tiny DFU with subtle traits. The majority of the DFU photos used in the present DFU dataset were taken at Lancashire Teaching Hospital, where DFU had already progressed to a large degree in most cases. Pre-ulcer and modest DFU were only detected in a very small number of patients. Therefore, additional instances of DFU of these grades are required in the DFU dataset in order to strengthen algorithms' ability to recognize these specific DFU.

REFERENCES

[1] Lawrence A Lavery, David G Armstrong, and Lawrence B Harkless, Classification of diabetic foot wounds,The Journal of Foot and Ankle Surgery, 35(6):528-531 , 1996.

[2] YannLeCun, Corinna Cortes, and Christopher JC Burges. The mnist database of handwritten digits, 1998.

[3] Lawrence A Lavery, David G Armstrong, and Lawrence B Harkless. Classification of diabetic foot wounds. The Journal of Foot and Ankle Surgery, 35(6):528{531, 1996.

[4] YannLeCun, Corinna Cortes, and Christopher JC Burges. The mnist database of handwritten digits, 1998.

[5] Alex Krizhevsky, IlyaSutskever, and Geofirey E Hinton. Image net classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097{1105, 2012.

[6] Brett Hewitt, MoiHoon Yap, and Robyn Grant. Manual whisker annotator (mwa): A modular open-source tool. Journal of Open Research Software, 4 (1), 2016.

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, PietroPerona, Deva Ramanan, Piotr Dollar, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In European conference on computer vision, pages 740{755. Springer, 2014.

[8] Sarah Wild, GojkaRoglic, Anders Green, Richard Sicree, and Hilary King. Global prevalence of diabetes estimates for the year 2000 and projections for 2030. Diabetes care, 27(5):1047{1053, 2004.

[9] World Health Organization et al. Global report on diabetes who geneva, 2016.

[10] K Bakker, Jan Apelqvist, BA Lipsky, JJ Van Netten, and NC Schaper. The 2015 iwgdf guidance documents on prevention and management of foot problems in diabetes: development of an evidence-based global consensus. Diabetes/metabolism research and reviews, 32(S1):2{6, 2016.

[11] Andrew JM Boulton, Loretta Vileikyte, Gunnel Ragnarson-Tennvall, and Jan Apelqvist. The global burden of diabetic foot disease. The Lancet, 366 (9498):1719{1724, 2005.

[12] Florencia Aguiree, Alex Brown, Nam Ho Cho, Gisela Dahlquist, Sheree Dodd, Trisha Dunning, Michael Hirst, Christopher Hwang, Dianna Magliano, Chris Patterson, et al. IDF Diabetes Atlas: sixth edition. International Diabetes Federation, 6th edition, 2013.

[13] David G Armstrong, Andrew JM Boulton, and Sicco A Bus. Diabetic foot ulcers and their recurrence. New England Journal of Medicine, 376(24): 2367{2375, 2017.

[14] David G Armstrong, Lawrence A Lavery, and Lawrence B Harkless. Validation of a diabetic wound classification system: the contribution of depth, infection, and ischemia to risk of amputation. Diabetes care, 21(5):855{859, 1998.

[15] Peter Cavanagh, Christopher Attinger, Zulqarali Abbas, ArunBal, Nina Rojas, and Zhang-RongXu. Cost of treating diabetic foot ulcers in very different countries. Diabetes/metabolism research and reviews, 28(S1):107{111, 2012.

[16] Chanjuan Liu, Jaap J van Netten, Je G Van Baal, Sicco A Bus, and Ferdifivan Der Heijden. Automatic detection of diabetic foot complications with infrared thermography by asymmetric analysis. Journal of biomedical optics, 20(2):026003{026003, 2015.

[17] LeiWang, Peder Pedersen, Emmanuel Agu, Diane Strong, and Bengisu Tulu. Area determination of diabetic foot ulcer images using a cascaded two-stage svm based classification. IEEE Transactions on Biomedical Engineering, 2016.

[18] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In European Conference on Computer Vision, pages 818{833. Springer, 2014.

[19] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1{9, 2015.

[20] Marios Anthimopoulos, Stergios Christodoulidis, Lukas Ebner, Andreas Christe, and Stavroula Mougiakakou. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE transactions on medical imaging, 35(5):1207{1216, 2016.

[21] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, ZiyueXu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging, 35(5):1285{1298, 2016.

[22] Ezak Ahmad, Manu Goyal, Jamie S McPhee, Hans Degens, and MoiHoon Yap. Semantic segmentation of human thigh quadriceps muscle in magnetic resonance images. arXiv preprint arXiv:1801.00415, 2018.

[23] Paras Lakhani and Baskaran Sundaram. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology, 284(2):574{582, 2017.

[24] Moi Hoon Yap, Manu Goyal, Fatima Osman, Ezak Ahmad, RobrtMart, Erika Denton, Arne Juette, and Reyer Zwiggelaar. End-to-end breast ultrasound lesions recognition with a deep learning approach. In Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging, volume 10578, page 1057819. International Society for Optics and Photonics, 2018.

[25] MoiHoon Yap, Manu Goyal, Fatima M Osman, Robert Mart, Erika Denton, Arne Juette, and Reyer Zwiggelaar. Breast ultrasound lesions recognition: end-to-end deep learning approaches. Journal of Medical Imaging, 6(1): 011007, 2018.

[26] Simon LF Walsh, Lucio Calandriello, Mario Silva, and Nicola Sverzellati. Deep learning for classifying brotic lung disease on high-resolution computed tomography: a case-cohort study. The Lancet Respiratory Medicine,6(11):837{845, 2018.

[27] M. Goyal, M. H. Yap, N. D. Reeves, S. Rajbhandari, and J. Spragg. Fully convolutional networks for diabetic foot ulcer segmentation. In 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pages 618{623, Oct 2017. doi: 10.1109/SMC.2017.8122675.

[28] Manu Goyal, Neil D Reeves, Adrian K Davison, Satyan Rajbhandari, Jennifer Spragg, and MoiHoon Yap. DFUNet: Convolutional neural networks for diabetic foot ulcer classification. arXiv preprint arXiv:1711.10448, 2017.

[29] Manu Goyal, Neil Reeves, Satyan Rajbhandari, and MoiHoon Yap. Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices. IEEE journal of biomedical and health informatics, 2018.

[30] F William Wagner. The diabetic foot. Orthopedics, 10(1):163{172, 1987.Robert G Frykberg. Diabetic foot ulcers: pathogenesis and management. American family physician, 66(9):1655{1662, 2002.

[31] Zheng Hu, Jiaojiao Zhang, Yun Ge, Handling Vanishing Gradient Problem Using Artificial Derivative, IEEE Access, February 9, 2021 Digital Object Identifier 10.1109/ACCESS.2021.3054915.

Energy Consumption Reduction Strategy and a Load Balancing Mechanism for Cloud Computing in IoT Environment

Tai Zhang, Huigang Li

Hebei Software Institute

Hebei Baoding 071000, China

Abstract—**Modern networks are built to be linked, agile, programmable, and load-efficient in order to overcome the drawbacks of an unbalanced network, such as network congestion, elevated transmission costs, low reliability, and other problems. The many technological devices in our environment have a considerable potential to make the connected world concept a reality. The Internet of Things (IoT) is a research community initiative to bring this idea to life. Cloud computing is crucial to making it happen. The load balancing and scheduling significantly increase the possibility of using resources and provide the grounds for reliability. Even if the intended node is under low or high loading, the load balancing techniques can increase its efficiency. This paper presents a scheduling technique for optimal resource allocation with enhanced particle swarm optimization and virtual machine live migration technique. The proposed technique prevents excessive or low server overloads through optimal allocation and scheduling tasks to physical servers. The proposed strategy was implemented in the cloudsim simulator environment and compared and showed that the proposed method is more effective and is well suited to decreasing execution time and energy consumption. This solution provides grounds to reduce energy consumption in the cloud environment while decreasing execution time. The simulation results showed that the amount of energy consumption compared to particle crowding has decreased by 10% and compared to PSO (Particle Swarm Optimization) scheduling by more than 8%. Also, the execution time has been reduced by 18% compared to particle swarm scheduling and by 8% compared to PSO.**

Keywords—Internet of things; load balancing; cloud computing; virtual machine migration

I. INTRODUCTION

Cloud computing is a computing model based on large computer networks such as the Internet, which provides a new model for the supply, consumption, and delivery of information technology services (including hardware, software, information, and other shared computing resources) using the Internet. Naturally, every change and new concept in the world of technology has its own advantages, problems, and complications [1]. Using cloud computing is not an exception to this rule. Among the advantages of cloud computing, we can mention the lack of time and place restrictions, simple sharing of resources, as well as the reduction of capital and operational costs (the most important advantage), because in fact, cloud computing dynamically provides scalable resources as a service on the Internet and has also put many challenges in

front of experts in this field, among which we can mention things like: resource allocation and load balancing, security, reliability, ownership, data backup and data portability [2, 3, 34]. Meanwhile, resource allocation and load balancing in cloud computing are of great importance. This issue has been discussed in various fields such as operating system, cluster computing and data center management. A resource allocation system in cloud computing can be considered as any mechanism whose purpose is to ensure that the requirements of applications are met [4, 36]. In addition, the resource allocation mechanism must examine the current status of each resource in the cloud environment in order to provide algorithms for better allocation of physical resources or virtual resources and thus reduce operational costs in the cloud environment. It is clear that due to the scale and complexity of these systems, the centralized assignment of tasks to a specific server without considering specific solutions is actually impossible, and also due to the increasing load and volume of requests in advanced data centers and the urgent need to achieve quality. For optimal service, the need for solutions to increase the efficiency of existing servers in the data center is felt. One of the ways to achieve optimal productivity is to use scheduling and load balancing solutions [5, 33]. This technology, with its high potential of cloud computing for storing and processing data remotely, has provided a new computing model. Recently, instead of using domestic resources, many large companies have outsourced them to cloud computing [6]. So users can access their data anywhere in the world, and they do not need high-performance hardware and storage systems because all computing and storage operations are performed by cloud service providers and well-equipped and advanced servers. Meanwhile, the scheduling and resource allocation problem in cloud computing is important because it directly affects the amount of energy consumption and reduction of latency in service provision [7]. A scheduling system in cloud computing can be any mechanism to ensure the provision of application requirements. In addition, the scheduling mechanism should examine the current status of each source in the cloud environment to provide algorithms for better allocation of physical or virtual resources and thus reduce operational costs in the cloud environment. That is because, during the processing, a number of servers might have a high traffic load. During the load distribution among servers with less load, the idle servers can be turned off to reduce energy consumption [8, 35, 26]. In this study, a PSO algorithm is suggested for scheduling and ideal load balancing in the cloud infrastructure

in order to reduce energy usage in cloud computing environments' data centers. The proposed technique prevents server overloads or low load through the optimal assignment of tasks to physical servers. This research is also important from another aspect that with the increase of users and their different requests, the following situations may occur:

- The virtual machine may be performing an operation and not accept another request.

- The request should be made on a new car.

- The request should be applied to a machine that is busy and does not have enough capacity to receive new ones.

- Migration may take place.

Therefore, a solution must be provided to manage these challenges. When the bandwidth of a virtual machine is full, the central cluster sends requests to another machine. At this stage, an algorithm is needed to balance the load and choose the right processing server. Based on this, a solution based on the improved particle community optimization algorithm is presented for efficient scheduling and as a result optimal load balancing in the cloud infrastructure, so that the task execution time can be reduced by broadcasting the requests and in As a result, it helped to manage energy consumption for users. In this research, as an innovation, for the purpose of scheduling, a combined method of particle swarm optimization and virtual machine live migration technique has been used to balance the load using an optimal schedule. The use of combined methods for load balancing has not been much considered in other researches. Also, the goals and contributions of the authors in this research are stated as follows:

- Creating load balance and reducing response time to users' requests.

- Balancing load in cloud network using particle swarm algorithm and task migration.

In the continuation of our paper, it is configured as follows. The next section, which discusses prior works in the area of cloud computing in the IoT context, presents a list of related works. In this section's conclusion, the current works' characteristics are given in table format for different criteria. Section III describes the proposed approach in three parts: problem-solving formulation, H algorithm based on scheduling, and virtual machine migration. The performance evaluation for energy consumption parameters and execution time is given in Section IV. The analysis of the comparison between the suggested approach and the current works is presented as a table in the fifth part. In the last part, the conclusion of the research is given.

II. PREVIOUS WORKS

This section looks at concerns like access control and load balancing for cloud computing networks. Also, the discussion related to the Internet of Things environment is covered. A method based on efficient workload distribution and resource management is suggested in reference [9] employing a cloud computing framework. In this approach, clustering learning techniques are employed to decrease network edge energy consumption as well as processing and communication delays. A similar strategy is presented in reference [10] to build a true edge cloud ecosystem. In the Internet of Things, a capillary computing architecture for orchestrating microservices from edge devices to cloud computing providers is suggested in reference [11]. An Edge-Fog-Cloud environment is described as a distributed cloud for IoT computing in reference [12]. The authors also review a cloud computing offloading strategy for simultaneous localization and mapping of indoor mobile robots in reference [13]. A trust management technique is described in reference [14] to enhance a distinct perspective on the cloud environment and to enable a blockchain-based cloud computing architecture. A safe offloading technique based on machine learning is also suggested in reference [15] for cloud computing, which expands the potential of IoT for applications related to smart cities. A semantic model-based strategy for IoT data description and discovery for IoT-Cloud architecture is presented in reference [16]. They suggested a method for optimizing energy consumption in a set of heterogeneous computing groups to serve various web applications in one of the earliest studies in scheduling and power management [17]. The suggested method decides whether to turn on or off the nodes to reduce total energy consumption while controlling the resources sporadically. In order to reduce power consumption, the request scheduling issue for multi-layer web applications was researched in [18]. It was suggested in [19] to use Power-Aware Tasks Scheduling (EATS) to divide and schedule vast amounts of data in the cloud. This model's primary objective was to improve application efficiency and lower energy usage in subsurface resources. According to [20], the grouped tasks scheduling (GTS) algorithm is used to plan out the tasks in a network of cloud computing services while considering customer needs for service quality. The suggested algorithm creates five groups to split tasks belonging to each group and share characteristics such as user type, task type, job size, and work delay. In order to run task-based efficient programs on distributed operating systems and save energy, a real-time dynamic scheduling system was developed [21]. There is currently no optimal multiprocessor solution for the NP-hard job scheduling problem. Therefore, a polynomial algorithm is suggested in [22] that combines exploratory principles and resource allocation strategies to locate suitable solutions quickly. Table I provides a comparison of the aforementioned techniques. The comparison is based on the purpose, the preferred method, and the result of the work.

TABLE I.   REVIEW OF PREVIOUS WORKS IN THE FIELD OF LOAD BALANCING USING CLOUD COMPUTING

Ref.	Objective	Method Referred	Achievement	Inference
[1]	Utilizing cloud computing for IoT resource allocation and workload distribution	Learning classifier	40% reduction in processing delays	The delay in the transmission of packets is reduced. Reducing energy consumption
[2]	Resource management in cloud computing	Pseudo code-dynamic testing	Increasing throughput and reducing latency	• This solution is affordable, scalable and reliable.
[3]	Developing a new architecture for smart applications that support different IoT workloads in the cloud computing environment.	orchestration based on containers	Many times faster for response time	• A Fog or Cloud resource was successfully offloaded from an Edge node. • Capable of coping with extremely dynamic IoT situations.
[4]	Distributed tasks processing in Cloud computing environment	The lowest processing cost is used in the first method to assign tasks.	Extend time without compromising associated costs	• Display processing and network costs. • Evaluation of Edge, Fog, and cloud computing options in light of device connection congestion.
[5]	Secure offloading for Cloud Computing of things	Machine learning methods, fuzzy neural model	Reduce latency	Cloud selection is based on reinforcement learning, and cloud node availability is estimated using available processing power and remaining node energy.
[6]	Internet of Things data management methods in IoT-Fog-Cloud	Better characteristics of IoT data flow in semantic model	IoT data stream characterization to support semantic data retrieval	• Attention to data storage issues. • The creation of protocols for data discovery for Internet of Things hardware.
[7]	For large-scale Internet of Things systems, enabling efficient access control procedures.	Token management methods with identity-based capabilities	An approach to access management for Internet of Things systems that is scalable	• Achievements of capabilities such as load balancing, decentralized access and lightweight approach
[8]	presenting a framework for predictive analytics using the IoT for mobile devices	utilizing machine learning to analyze data	Data privacy, low cost of data transit to data centers, and quick feedback	• Big data management for IoMT devices

III. PROPOSED APPROACH

This section provides the particle swarm optimization algorithm, and the virtual machine lives migration technique to create optimal load balancing in the cloud infrastructure. Using this solution, the execution time of the tasks decreases, and the energy consumption is also reduced. Accordingly, the formulation of the solution is presented below.

A. Formulation of the Solution

Since mapping the workflow of a program to distributed resources can have several goals, the present study is focused on two goals: energy consumption and traffic consumption. The two targets are formulated as follows.

Studies in [23] have demonstrated that a linear relationship between energy consumption and CPU sufficiency can correctly describe the energy consumption of servers. As a result, the following model describes how much energy a physical machine uses in a cloud environment:

$$P_j = \begin{cases} \left(P_j^{busy} - P_j^{idle}\right) \times U_j^p + P_j^{idle}, & U_j^c > 0 \\ 0 & , \text{ otherwise} \end{cases} \quad (1)$$

Where:

P_j: Energy consumption of the physical machine j

P_j^{idle}: The average energy consumption of the physical machine j when it is idle

P_j^{busy}: The average energy consumption of the physical machine j when it is busy

U_j^p: The normalized amount of processor consumed by the physical machine j

$U_j^c > 0$: The physical machine j is on

The main idea behind this kind of modeling is to convert the less-busy physical machines into idle and then turn them off. In this relation, P_j^{busy} and P_j^{idle} are constant values which are 162 and 215 watts for Dell physical machines.

The following equation is used to calculate the network bandwidth consumption and traffic generated by each physical machine:

$$D_j = \sum_{\forall m \epsilon V_j} \lambda(j,m) \sum_{i=1}^{\rho(j,m)} C_i \quad (2)$$

Where:

D_j: The communication between the physical machine j and other physical machines

$\lambda(j,m)$: Traffic load between the physical machine j and other physical machines

V_j: The set of physical machines communicating with the physical machine j

C_i: The weight of communication link between two physical machines at level i

$\rho(j , m)$: The level of communication between the physical machines m and the physical machine j

Moreover, the approach presented in [24] is used to model resource usage. The model is predicated on the idea that the devices' power consumption and processor productivity truly follow a linear relationship. To put it another way, the knowledge about a task's processing time and CPU efficiency is sufficient to determine its power usage. The definition of efficiency for a resource like ri at any given time is as follows:

$$U_i = \sum_{j=1}^{n} u_{i,j}$$

Where n is the number of currently active tasks and uij is the number of resources task tj uses. As a result, the following formula is used to determine the energy consumption (Ej) of the rj resource:

$$E_j = (P_{max} - P_{min}) \times U_i + P_{min} \tag{3}$$

In this case, Pmax is the maximum power consumption (i.e., at 100% efficiency), and Pmin is the minimum energy consumption per minute when the server is active (or at 1% efficiency).

B. Load Balancing Processes

Based on the above strategies, the main steps for load balancing are as follows:

- Calculating the transfer probability for all its neighbors and selecting the largest one as the next destination.

- Moving to a new node and checking if it is a candidate node; if the answer is yes, a migration should be created and initialized. For advanced migration, you have to go back to the first step.

- Retrograde migration returns to the starting point of the corresponding leading ant and in the same direction as the leading migration. During the route, the information related to the pheromone of each node that the backward migration passed through was updated and if it reached the starting point, it deleted the backward migration.

- Calculating the total resources of the candidate nodes and if the nodes need to perform load balancing operations, these steps will be stopped.

- Finally, in the last step, the load balancing operation should be performed using the live migration of the virtual machine.

- These steps for max-min rules also; it is similar except in the step of calculating the transfer probability.

C. Particle Swarm Optimization based Scheduling

Now, based on the relations presented in the previous section using the PSO algorithm for scheduling is discussed. Other evolutionary algorithms and the PSO algorithm are comparable. The population in PSO is the same as the number of particles in the problem space. Initialization of the particles is random. Every particle will have a compatibility value, which will be evaluated by the compatibility function, which ought to be optimized for every generation. This algorithm attempts to identify the best solution by updating generations after first creating a collection of particles at random [25]. In actuality, each particle chooses the best location (pbest) and position from the gbest particles. Each particle's pbest represents its best outcome to date, while its gbest represents its best compatibility with the entire population. Each generation will update the particle positions and velocities using relations (4) and (5), respectively:

$$v[\ ] = v[\ ] + \tag{4}$$

c1 * rand() * (pbest[ ] - position [ ]) +

c2 * rand() * (gbest[ ] - position[ ])

$$position[\ ] = position[\ ] + v[\ ] \tag{5}$$

The first part of equation (4) is the particle's current velocity. In contrast, the second and third portions are the particle's change in velocity and its rotation toward the best individual experience and collective experience, respectively. The initialization of the particle position and velocity in the PSO algorithm is random. The given tasks in this study are the particles, and the size (dimension) of the particles represents the total number of tasks in the workflow. The processing resource indices are the values assigned to each particle dimension. Each particle, then, represents the mapping of a resource to a job. Each particle is assessed using the compatibility function provided in relations (1) and (2). Equation (3) is used to calculate the velocity of particles, while Equation (3) is used to update their location (4). This assessment continues until it reaches a certain number of repetitions.

The PSO-based scheduling algorithm is presented in Fig. 1. For scheduling by PSO algorithm, the initial step is calculating the traffic cost between resources using equation (2) based on the current network load. The algorithm then distributes completed tasks to resources by PSO's mapping. Then, according to the number of accomplished tasks, the list is ready to be updated, including tasks that their father completes. Following that, the energy wastage cost is calculated. These steps continue until all tasks are scheduled in the workflow.

1: Calculate the cost of traffic between resources using equation (2)
2: Compute PSO ({Task})
3: for all "ready" tasks {Task} ∈ T do
4: Assign tasks to resources sby PSO
5: Dispatch all the mapped tasks
6: Update task list
7: Calculate energy wastage cost using equation (1)
8: Compute PSO ({Task})
9: Perform VM Migration for load balancing.

Fig. 1.   PSO Scheduling Algorithm.

D. Virtual Machine Migration

During the execution process of scheduling tasks on processing servers performed with the help of particle swarm optimization, some servers may fail to execute tasks for reasons such as delays in previous processing tasks. They might have additional traffic load due to the delay in processing tasks, or hardware and software errors may cause additional traffic and overload while some servers are idle. In this case, the processing servers are reviewed periodically, and then the following three conditions are used to determine the status of the processing servers:

- If the server processor achieves its maximum productivity, then the server falls into the category of overloaded machines (Poverload).

- Suppose the server processor productivity is lower than the specified average productivity. In that case, it falls into the category of underload servers, which is usually around 10%. (Punder)

- Other servers that do not meet any of the above requirements are included in the normal server (Pnormal).

To make load balancing, the overload servers should migrate to under-load servers, so loads of servers in overload mode are migrated through the virtual machine migration technique onto another server in Punder or normal mode. Other servers with productivity below 10 percent will be candidates for tasks migration, and therefore they are turned off to reduce power consumption.

IV.   ANALYSIS, EVALUATION AND PERFORMANCE

A model that could simulate and evaluate the cloud infrastructure equipment is required to simulate and evaluate the solution. CloudSim is an extendable simulation tool that enables modeling and simulating cloud computing systems and preparing applications. In reality, this program makes it possible to do integrated modeling, simulation, and assessment of cloud infrastructure and associated services [26]. In order to simulate data centers, 800 hosts have also been defined. A class called a server that expands the Power Model SpecPower class, a subclass of the Power Model class in CloudSim, must be developed in order to establish any virtual server in the software. The CoMon project, a monitoring infrastructure for PlanetLab, has tested real traced data, which has been used to generate the workload. The Round-Robin and ant colony optimization methods are compared to the strategy. The power consumption, execution time, and SLA criteria are taken into account to evaluate the suggested algorithm's effectiveness compared to other methods. Finally, the evaluation results are shown in Fig. 2 to 6.
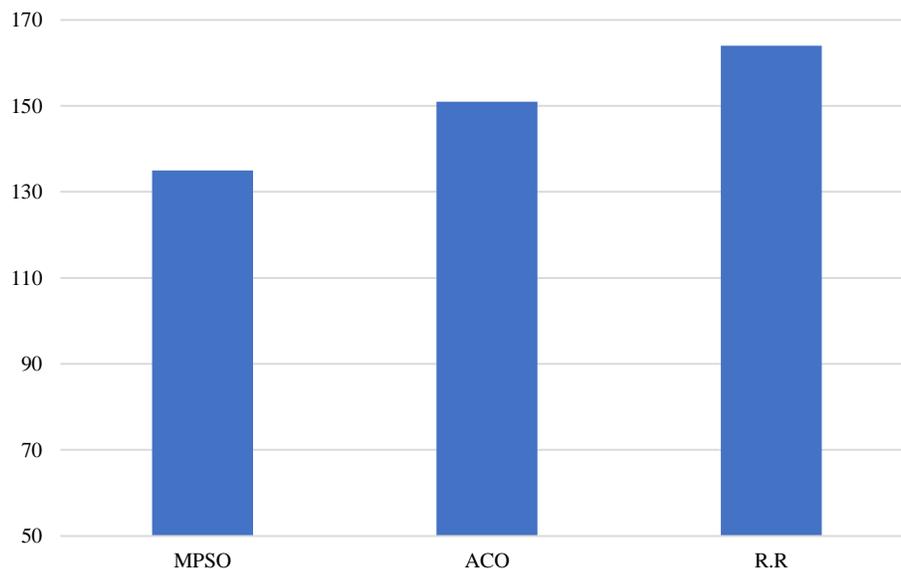
Fig. 2.   Energy Consumption (kw/h).
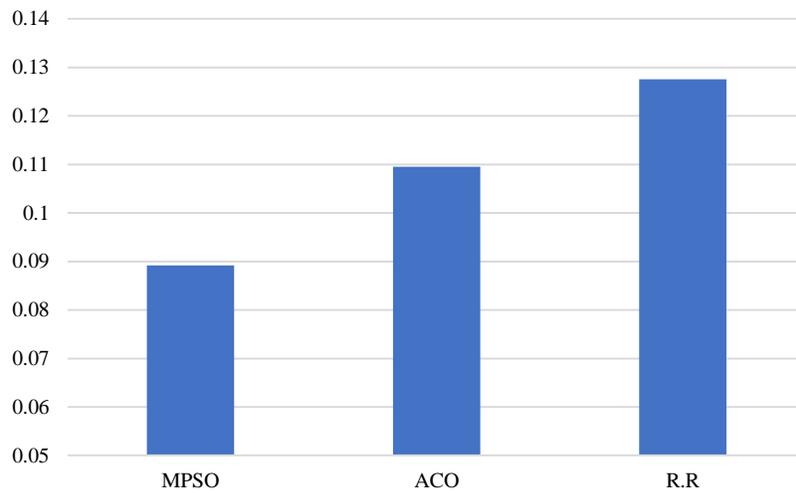
Fig. 3.   Execution Time (Second).
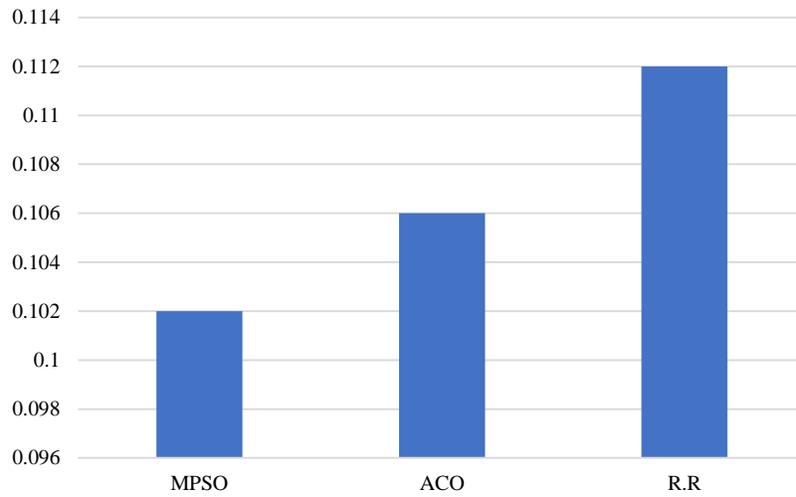
Fig. 4.   Service Level Agreement (%SLA).
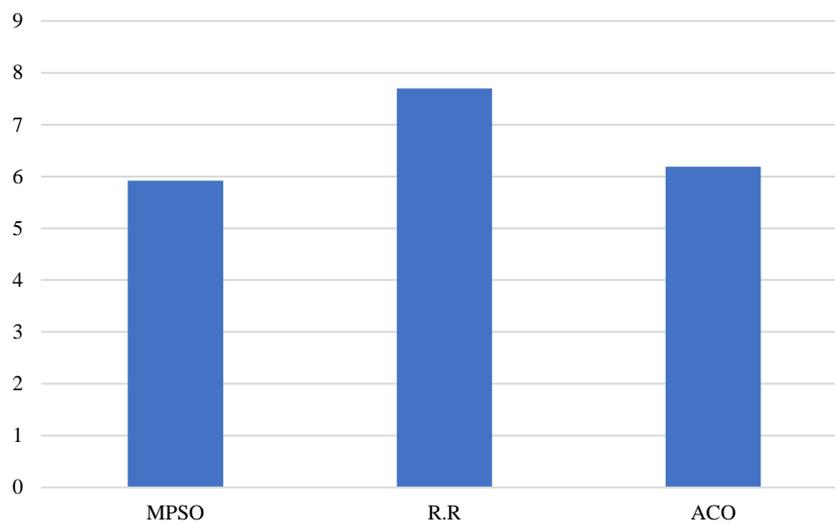
Fig. 5.   Service Level Agreement Degradation due to Migration (%SLA).
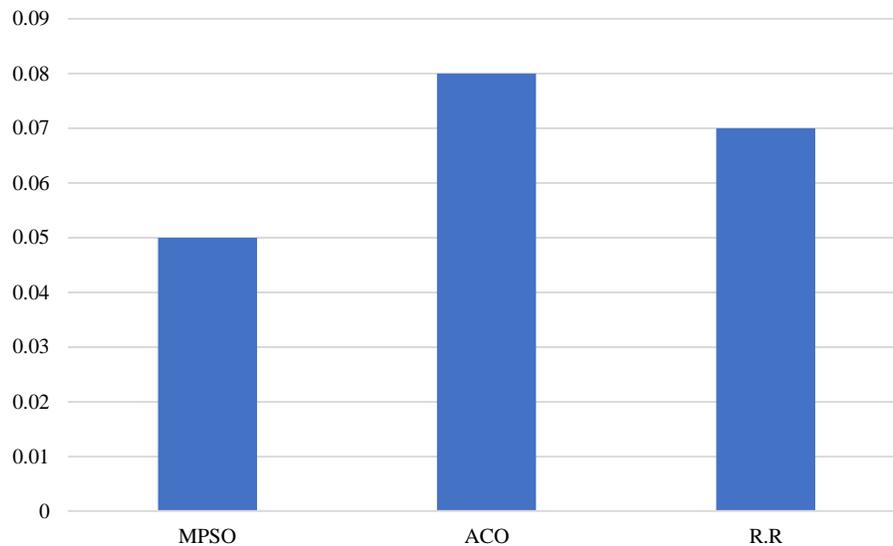
Fig. 6.   Execution Time - VM Reallocation Mean.

According to Fig. 2, the energy consumption in the proposed solution is greatly reduced compared to the other two solutions due to appropriate scheduling and balanced load distribution between processing servers and turning the idle servers off. Accordingly, the scheduling is reduced by 10 and 17% compared to the Ant Colony optimization and Round Robin algorithms. Given the number of resources and energy consumption in each schedule, this reduced energy consumption was predictable.

In addition to reducing energy consumption, reducing demand execution time is also very important in cloud infrastructure. In fact, it is another goal of this research. Accordingly, the execution time in the solution has been reduced to an appropriate level, as shown in Fig. 3. The proposed solution has reduced the execution time by 18 and 28% more than the R.R and ant colony optimization, respectively. This is due to load balancing in the cloud infrastructure by utilizing PSO-based scheduling and the task execution predictor. Moreover, the use of the virtual machine live migration technique allows transferring the processing load to free or idle servers, and the execution time is reduced significantly.

In Fig. 4, the service level agreement violations have been shown. The Service Level Agreement (SLA) is the basis for determining the expected level of service. The service quality parameters in the service level agreement specify the extent to which the provided quality is appropriate. Customers need this contract to ensure the quality level of their services. In fact, the primary purpose of this agreement is to define an official basis for the terms of provided service, such as efficiency or availability. According to the figure, using the load balancing produced by the optimal scheduling, the proposed strategy has provided the servers with higher reliability than the other two strategies, creating better availability in the cloud infrastructure. In other words, the violation percentage of this solution is lower than the service level agreement compared to the other two solutions.

According to Fig. 5, the proposed solution has the least SLA degradation due to migration because all migrations are optimized and adapted to the needs of the infrastructure resulting in the lowest reduction in quality of service. In fact, the service quality parameters in the service level contract determine how appropriate the quality of the provided service is. Customers ensure the quality of the provided service through this contract. As seen in the figure, thanks to the load balancing created by optimizing the timing, the proposed solution has made the servers more reliable than the other two solutions. It thus has higher availability in the cloud infrastructure compared to other solutions. In other words, the percentage of violations of this solution is less than the service level agreement compared to the other two solutions.

Since virtual machine migration can affect runtime, Fig. 6 shows the VM reallocation mean time required for virtual machine reallocation. As it can be seen, the proposed solution in this parameter is also more efficient than other methods. In other words, since this solution examines the status of server resources and allocates them accordingly, it creates the least delay in executing user requests.

V.   COMPARATIVE ANALYSIS

Table II describes the comparative analysis of the suggested method with related research. The comparison is based on features of already published literature, including load balancing, resource management, and energy usage reduction. Almost all publications emphasize the significance of load balancing, except [23, 27] and certain works that allude to resource management, including [19, 20, 22, 23, 25]. Almost all of these studies have not covered the idea of lowering energy usage and scheduling. The references [20, 21, 24, 25] have used edge computing methodologies. They also employ cloud computing in many aspects of their work.

TABLE II. COMPARING THE SUGGESTED APPROACH WITH THE CURRENT APPROACHES

Ref.	Resource management	Load balancing	energy consumption reduction	Scheduling based	Cloud based
[7]	✓	✓	×	×	✓
[8]	✓	✓	×	×	✓
[9]	✓	×	×	✓	✓
[27]	×	✓	×	×	✓
[28]	✓	×	×	✓	✓
[29]	✓	×	×	×	✓
[30]	×	×	×	×	✓
[31]	✓	✓	×	×	✓
[32]	✓	×	×	×	✓
Proposed approach	✓	✓	✓	✓	✓

Also, this paper presents a scheduling technique for optimal resource allocation with particle swarm optimization and virtual machine live migration technique. The proposed technique avoids server overload or underload through optimal allocation and scheduling of tasks to physical servers. The proposed strategy was implemented and compared in the cloudsim simulator environment and showed that the proposed method is more effective and suitable for reducing execution time and energy consumption. This solution provides the basis for reducing energy consumption in the cloud environment and at the same time reducing execution time. Compared to the previous test, due to the increase of ants, it is clearly visible that the virtual machine migration is increasing compared to the previous state which has caused the execution time to decrease. In fact, by applying suitable and optimal migrations, additional load has been transferred from high traffic servers to low traffic servers. As can be seen from Table III, although the amount of energy consumption has increased with the increase in the number of ants, but at the same time, the amount of energy consumption has decreased by approximately 12% compared to the MPSO (Multi_Objective Particle Swarm Optimization) method and 9% compared to Round Robin. Also, by placing Optimum methods and reduction of additional overhead, execution time has also been optimized compared to all three other methods. As compared to the R.R. method, it has reached 27%.

TABLE III. THE DEGREE OF OPTIMALITY OF THE PROPOSED SOLUTION COMPARED TO OTHER SOLUTIONS (100 ANTS)

	PSO	R.R	Random
energy consumption	12%	11%	5%
Virtual machine migration	12%	13%	17%
execution time	5%	29%	16%

VI. CONCLUSIONS

Due to the scale and complexity of the cloud infrastructure, the centralized assignment of tasks to a specific server without considering any specific solutions is impossible. Also, considering the increasing workload and volume of requests in advanced data centers and the urgent need to achieve optimal service quality, solutions should be developed to increase the productivity of existing servers in the data center. One method to achieve optimal productivity is the use of scheduling technics. Load balancing is needed to manage the service providers' resources properly. This study uses a solution based on a swarm particle optimization algorithm and the virtual machine lives migration technique for load balancing. The proposed technique prevents server overloads or low load through the optimal assignment of tasks to physical servers. Finally, it was shown with the help of ClodSim software and in the simulation environment that this method significantly increases the efficiency. For this purpose, the proposed solution was evaluated in two cases with the number of ants 50 and 100. In the first test (number of 50 ants), the solution was compared with the three algorithms ACO, Round Robin and MPSO and it was shown that the proposed solution was able to execute the desired requests in less time than the two algorithms ACO and Round Robin. Efficiency can be provided with the help of the correct migrations that the solution has provided, and the machines that have had a heavy load have been transferred from them to another machine that has a lower load, and on the other hand, the machines that have no load have also been transferred, found correctly and by turning them off, the amount of energy consumption is also reduced; so that compared to the ACO algorithm, the energy consumption has decreased by almost 10% and compared to the round robin algorithm by 13%, but compared to the MPSO, the optimization in terms of energy consumption has not been achieved. In the second experiment, the number of ants increased to 100 ants and in comparison, it was shown that although the reduction of energy consumption is less than the first case, but with optimal virtual machine migrations, the execution time for all three algorithms is between 12 to 27% decreased. This indicates that the solution is optimal compared to other methods.

In the future, the continuation of this research can be extended to the execution time based on demand resource

allocation, because the proposed method does not meet the resource requirements during the execution of requests.

REFERENCES

[1] Kaur, M., & Aron, R. (2020). Energy-aware load balancing in fog cloud computing. *Materials Today: Proceedings*.

[2] Abdulhammed, O. Y. (2022). Load balancing of IoT tasks in the cloud computing by using sparrow search algorithm. *The Journal of Supercomputing*, 78(3), 3266-3287.

[3] Maswood, M. M. S., Rahman, M. R., Alharbi, A. G., & Medhi, D. (2020). A novel strategy to achieve bandwidth cost reduction and load balancing in a cooperative three-layer fog-cloud computing environment. *IEEE Access*, 8, 113737-113750.

[4] Mozaffari, H., & Houmansadr, A. (2020, January). Heterogeneous private information retrieval. In Network and Distributed Systems Security (NDSS) Symposium 2020.

[5] Gowri, V., & Baranidharan, B. (2023). Dynamic Energy Efficient Load Balancing Approach in Fog Computing Environment. In *Intelligent Communication Technologies and Virtual Mobile Networks* (pp. 145-160). Springer, Singapore.

[6] Abdulhammed, O. Y. (2022). Load balancing of IoT tasks in the cloud computing by using sparrow search algorithm. *The Journal of Supercomputing*, 78(3), 3266-3287.

[7] Javadpour, A., Sangaiah, A. K., Pinto, P., Ja'fari, F., Zhang, W., Abadi, A. M. H., & Ahmadi, H. (2022). An Energy-optimized Embedded load balancing using DVFS computing in Cloud Data centers. *Computer Communications*.

[8] Chapnevis, A., Güvenç, I., & Bulut, E. (2020, November). Traffic Shifting based Resource Optimization in Aggregated IoT Communication. In 2020 IEEE 45th Conference on Local Computer Networks (LCN) (pp. 233-243). IEEE.

[9] Lakra, Atul Vikas, and Dharmendra Kumar Yadav. "Multi-objective tasks scheduling algorithm for cloud computing throughput optimization." Procedia Computer Science 48 (2015): 107-113.

[10] Vahidi Farashah, M., Etebarian, A., Azmi, R., & Ebrahimzadeh Dastjerdi, R. (2021). A hybrid recommender system based-on link prediction for movie baskets analysis. *Journal of Big Data*, 8(1), 1-24.

[11] Pourghebleh, B., & Hayyolalam, V. (2020). A comprehensive and systematic review of the load balancing mechanisms in the Internet of Things. Cluster Computing, 23(2), 641-661.

[12] Wei, R. (2022). Load Balancing Optimization of In-Memory Database for Massive Information Processing of Internet of Things (IoTs). Mathematical Problems in Engineering, 2022.

[13] Charandabi, S. E., & Kamyar, K. (2022). Survey of Cryptocurrency Volatility Prediction Literature Using Artificial Neural Networks. Business and Economic Research, 12(1).

[14] Ficco M, Esposito C, Xiang Y, Palmieri F. Pseudo-dynamic testing of realistic edge-fog cloud ecosystems. IEEE Commun Mag. 2017;55(11):98-104.

[15] Taherizadeh S, Stankovski V, Grobelnik M. A capillary computing architecture for dynamic internet of things: orchestration of microservices from edge devices to fog and cloud providers. Sensors. 2018;18(9):29-38.

[16] Mohan N, Kangasharju J. Edge-Fog cloud: a distributed cloud for internet of things computations. Proceedings of the 2016 Cloudification of the Internet of Things (CIoT). Paris, France: IEEE; 2016:1-6.

[17] Sarker VK, Queralta JP, Gia TN, Tenhunen H, Westerlund T. Offloading slam for indoor mobile robots with edge-fog-cloud computing. Paper presented at: Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT); May 2019:1-6; Dhaka, Bangladesh: IEEE.

[18] Kochovski P, Gec S, Stankovski V, Bajec M, Drobintsev PD. Trust management in a blockchain based fog computing platform with trustless smart oracles. Futur Gener Comput Syst. 2019;101:747-759.

[19] Madani, M., Lin, K., & Tarakanova, A. (2021). DSResSol: A sequence-based solubility predictor created with Dilated Squeeze Excitation Residual Networks. International Journal of Molecular Sciences, 22(24), 13555.

[20] Zeng W, Zhang S, Yen IL, Bastani F. Semantic IoT data description and discovery in the IoT-edge-fog-cloud infrastructure. Paper presented at: Proceedings of the 2019 IEEE International Conference on Service-Oriented System Engineering (SOSE); April 2019:106-10609; San Francisco, CA, USA: IEEE.

[21] Bouflous, Z., Ouzzif, M., & Bouragba, K. (2023). Analysis of Load Balancing Algorithms Used in the Cloud Computing Environment: Advantages and Limitations. In *Proceedings of the Future Technologies Conference* (pp. 206-226). Springer, Cham.

[22] Trik, M., Molk, A. M. N. G., Ghasemi, F., & Pouryeganeh, P. (2022). A Hybrid Selection Strategy Based on Traffic Analysis for Improving Performance in Networks on Chip. Journal of Sensors, 2022.

[23] Ismail Leila, Fardoun Abbas," Eats: Energy-aware tasks scheduling in cloud computing systems" Procedia Computer Science, V 83, P 870-877, 2016, Elsevier.

[24] S. Pourjabar and G. S. Choi, "A high-throughput multimode low-density parity-check decoder for 5G New Radio," International Journal of Circuit Theory and Applications, 2021.

[25] Sun, J., Zhang, Y., & Trik, M. (2022). PBPHS: A Profile-Based Predictive Handover Strategy for 5G Networks. *Cybernetics and Systems*, 1-22.

[26] Ali Hend Gamal El Din Hassan,Saroit Imane Aly,Kotb Amira Mohamed, " Grouped tasks scheduling algorithm based on QoS in cloud computing network " Egyptian Informatics Journal, V 18, P 11-19, 2017, Elsevier.

[27] Mozaffari, H., Houmansadr, A., & Venkataramani, A. (2019, December). Blocking-Resilient Communications in Information-Centric Networks using Router Redirection. In 2019 IEEE Globecom Workshops (GC Wkshps) (pp. 1-6). IEEE.

[28] Agrawal, N. (2021). Dynamic load balancing assisted optimized access control mechanism for edge-fog-cloud network in Internet of Things environment. *Concurrency and Computation: Practice and Experience*, 33(21), e6440.

[29] Trick, M., Boukani, B., Emtiyaz, S., Azar, S. R., & Darvandi, F. M. (2014, June). An Overview of through-silicon via–based three dimensional integrated circuits (3D IC) to placement to optimize timing. In International Science Congress Association, Research Journal of Recent Sciences, Manuscript No: ISCARJRS-2013-303.

[30] Pourbemany, J., Mirjalily, G., Abouei, J., & Raouf, A. H. F. (2018, May). Load Balanced Ad-Hoc On-Demand Routing Basedon Weighted Mean Queue Length Metric. In Electrical Engineering (ICEE), Iranian Conference on (pp. 470-475). IEEE.

[31] Rafiee, P., & Mirjalily, G. (2020). Distributed Network Coding-Aware Routing Protocol Incorporating Fuzzy-Logic-BasedForwarders in Wireless Ad hoc Networks. Journal of Network and Systems Management, 28(4), 1279-1315.

[32] Baburao, D., Pavankumar, T., & Prabhu, C. S. R. (2022). A novel application framework for resource optimization, service migration, and load balancing in fog computing environment. Applied Nanoscience, 1-14.

[33] Trik, M., Pour Mozaffari, S., & Bidgoli, A. M. (2021). Providing an adaptive routing along with a hybrid selection strategy to increase efficiency in NoC-based neuromorphic systems. Computational Intelligence and Neuroscience, 2021.

[34] Yamini, R. "Energy aware green task assignment algorithm in clouds." Ianternational Journal for Research in Science and Advance Technology, Issue-1 1 (2018).

[35] Shiri, A., & Khosroshahi, G. K. (2019, April). An FPGA implementation of singular value decomposition. In 2019 27th Iranian Conference on Electrical Engineering (ICEE) (pp. 416-422). IEEE.

[36] Trik, M., Bidgoli, A. M., Vashani, H., & Mozaffari, S. P. (2022). A new adaptive selection strategy for reducing latency in networks on chip. *Integration*.

A Review of Lightweight Object Detection Algorithms for Mobile Augmented Reality

Mohammed Mansoor Nafea[1], Siok Yee Tan[2]

Center for Artificial Intelligence Technology, Faculty of
Information Science and Technology
Universiti Kebangsaan Malaysia, Bangi, 43600 Selangor,
Malaysia

Mohammed Ahmed Jubair[3]

Department of Computer Technical Engineering, College of
Information Technology, Imam Ja'afar Al-Sadiq University
Al-Muthanna 66002, Iraq

Mustafa Tareq Abd[4]

Department of Computer Technology Engineering, Middle East University College
Baghdad, Iraq

Abstract—Augmented Reality (AR) has led to several technologies being at the forefront of innovation and change in every sector and industry. Accelerated advances in Computer Vision (CV), AR, and object detection refined the process of analyzing and comprehending the environment. Object detection has recently drawn a lot of attention as one of the most fundamental and difficult computer vision topics. The traditional object detection techniques are fully computer-based and typically need massive Graphics Processing Unit (GPU) power, while they aren't usually real-time. However, an AR application required real-time superimposed digital data to enable users to improve their field of view. This paper provides a comprehensive review of most of the recent lightweight object detection algorithms that are suitable to be used in AR applications. Four sources including Web of Science, Scopus, IEEE Xplore, and ScienceDirect were included in this review study. A total of ten papers were discussed and analyzed from four perspectives: accuracy, speed, small object detection, and model size. Several interesting challenges are discussed as recommendations for future work in the object detection field.

Keywords—Augmented reality (AR); object detection; computer vision (CV); non-graphics processing unit (Non-GPU); real time

I. INTRODUCTION

As opposed to virtual reality (VR), the term augmented reality (AR) refers to a virtual interface, either in 2D or 3D, that enhances human vision by superimposing additional information (digital content) over the actual surroundings, which results in complete immersion in the imaginary reality[1]. As we can always see and feel the real world around us, it is not possible to get immersed in the virtual world.

AR relies on a device that captures the real world and inverts live animations, virtual objects, texts, data, or sounds observed by the user on a smartphone, tablet, computer, a pair of glasses, or any other on-screen display system [2]. In AR, the virtual information and the real world are synched using geo-localization and embedded sensors (gyroscope, accelerometer) that position the user and adjust the display according to his environment and movements.

Think about driving at night on a mountainous, curved route; also imagine a thick fog covering this area right now. The fog completely hides the trees on either side of the road. Considering how crowded it is, it is challenging to even see the traffic signs from a distance of two to three feet. The fog reflects the headlights' illumination, making driving risky, but stopping could be much riskier. In essence, there is no way to win, Similar cases to this have resulted in numerous multicar accidents. Now imagine being able to view the road's layout on a windshield, displaying the distance between you and the vehicles in front of you, your current speed, exit points, and intersections simply by pushing a button. Even though it's never fully safe to drive in fog, this display can provide the driver access to important information that could help avoid an accident [3].

Again, imagine a new faculty member entering a very large academic library; this new member turns on the camera on his/her cellphone and carefully scans the entire area. Imagine the new member having information about the room he/she is standing on the phone screen. Perhaps the location of the reference books, stacks, and current journals, or where to find help, is recorded by the camera. Imagine a scenario where a new faculty member may have virtual arrows pointing to the book's location., he/she is looking for on display to guide him/her in finding the location of the book. The use of AR makes both scenarios possible [3].

As a method of human-computer interaction, AR merely projects virtual data onto real objects [4]. The accurate superimposition of virtual data onto real objects depends on the detection of real objects and the knowledge of their precise coordinates. Several researchers discussed and elaborated on different challenges faced by AR technology. The following are the main challenges faced by AR technology.

A. Display

One of the main challenges with AR display technology is to create an extensive field of view, high resolution [4], see-through display in a socially suitable shape element. The study by [5] listed some of the challenges in optics and displays that must be addressed, which include offering enough brightness

and sharp display, having a high resolution and extensive area of view, addressing eyestrain, and being in a sunglass-like shape element. Several crucial subjects consist of addressing the AR vergence accommodation challenges, displaying photorealistic content material, and new shape elements which includes contact lenses.

B. Interaction Techniques

This is another challenge which is allowing humans to control AR content material as effortlessly as they do with items within the actual world. One approach that has been explored is utilizing actual gadgets to interact with AR in a method known as Tangible AR [6]. Free-hand gestures are supported by current AR systems; however this will be improved in addition to voice recognition and enabling mixed speech and gesture input in multi-modal interfaces. Future studies might use a variety of different approaches using eye tracking, whole-body input, and various non-verbal indications. Ultimately, [7] recommended more studies on the interaction methods that are not feasible inside the actual world.

C. Social and Ethical Issues

In the long term, the meaningful social and ethical concerns in AR technology are a complicated issue that must be addressed, for instance, identity hacking. However [8] mentioned the privacy implications of seeing individual information in public spaces; the study came up with lists of questionable ethical uses of AR which include deception, surveillance, and behavior modification.

D. Object Detection Techniques

For AR applications like assembly guidance, real-time, scalable object detection is a crucial task. Real-time object detection from RGB images has already been addressed by several Deep Neural Network (DNN) models. Although the majority of current AR and mixed reality (MR) systems can comprehend the 3D geometry of their surroundings, they are unable to recognize and categorize complex items in the real world. Deep Convolutional Neural Networks (CNN) can allow these features; however, it is still challenging to run big networks on mobile devices. It is extremely difficult to offload object detection to the edge or cloud due to the strict requirements for low end-to-end latency and excellent detection accuracy. In [9] proposes an object detector to boost embedded devices' ability to identify objects.

Augmented reality simply displays digital information onto real-world objects as a technique for human-computer interaction. In the era of augmented reality, virtual things created by computers can precisely and instantly overlay physical ones. Predicting bounding boxes and categorizing objects are both steps in the process of object detection which is a key area of computer vision today. Robust detection of objects from natural features of AR is still a complex problem and usually demands high computational time.

DL-based object detection has received significant research interest in current years. For higher picture knowledge techniques, it is vital not only to pay attention to classifying different pictures but to attempt to exactly estimate which objects are present within the pictures and their places (known as object detection) [10]. Many studies have been reported on object recognition using CNN in the field of computer vision.

However, end-to-end DL object identification methods based on regression methods, such as the Single Shot multi-box Detector (SSD) series and the YOLO series, have been successful in detecting objects in real time using GPU-based computers. It is exceedingly difficult to achieve accurate and real-time detection using non-GPU-based PCs and portable devices with low computing capacity because of the high computational needs of many systems. Several researchers discussed that a lot of object detection models have higher computing time which makes their object detection models take a long time to provide the outcomes. The following are the survey's contributions:

1) A comprehensive survey was conducted on the current object detection models for embedded devices published from 2018 to 2022.

2) We summarized numerous object detection methods for mobile and embedded devices.

3) The employment of algorithms to identify their unique and constrained features. We also looked at how these techniques handle problems that arise during the object detection process, as well as the advantages and disadvantages of these techniques for object detection models.

4) Discussion on the current open problems to help direct future studies on improving and enhancing the performance of object detection techniques. The development of the model was done after reviewing many earlier studies from related domains.

There are eight sections in this paper; the introduction and scope of this study are presented in Section I, and the review of the structure of the AR system is presented in Section II. In Section III some applications of object detection were discussed. The most important challenges in the object detection field are presented in Section IV. Previous research on similar topics, the outcome of the analysis, and the limitations and conclusions of such studies are in Sections V, VI, VII, and VIII respectively.

II. AR System

Specific software and hardware are required for the various AR systems, however, the software utilized uses the real-world coordinates via cameras and tracking hardware; the purpose is to convert this location data into an XML file, in software the so-called ARML (Augmented Reality Markup Language) used. The ARML functional blocks establish the relationship between the actual and virtual worlds by identifying the relationship between them; this enables use of virtual items in the real world. These virtual items are controlled by the user's actions that they take [11]. Most devices utilized in AR applications are IoT devices that fall into one of three categories:

- Sensors: Sensors gather information from the real world and send it to an AR app. for example, a mobile device's built-in camera gathers information about a user's environment. Data is processed by the software, which subsequently shows the user predefined content.

With information from cameras or 3D models, the scene's composition is achievable. The tracking device could be (RFID, wireless sensors, accelerometers, GPS, gyroscopes, solid-state compasses, and digital cameras) [11] and have several settings and ranges; they enhance the AR system's tracking accuracy.

- Input devices: These tools let users engage with AR systems. The AR Interface works as a medium between the AR system and these components. The Ikea application's UI could be a good example. The user could move the furniture items in his house by using signs that the program will then interpret as orders [11]. The types of inputs include speaking, blinking, touching, and gesturing, among others. Input devices examples: microphones, touch displays, gesture controllers, styluses, and pointers [12].

- Output devices: Users can interact with the AR system using these tools. Whenever these tools are used for a particular purpose, they are often worn on the user's forehead. HMDs, monitors, and wearable technology are some examples of these devices [13].

Offering a solution in real-time is one of an AR system's main characteristics. The user experience is not the only benefit of AR technology. t offers excellent economic opportunities for service providers and companies [13]. As AR has been more widely used, it allows more e-production of wireless networks, sensors, high-end cameras, cellphones, and other devices, but there are several factors to take into account while designing an AR system and the architecture that supports it. Such as the high quality and how immersive AR is, also the monitoring and rendering slowness [11]. There are some limitations related to the AR systems such as:

A. Hardware Limitations

A variety of smart devices are used to create AR systems. From the least powerful to the most powerful devices that work across multiple environments. Therefore, one of the key goals is to lower these devices' energy consumption to increase their efficiency[14]. In addition to the energy consumption, the cost of the AR devices may easily cost thousands of dollars which makes them unreachable to everyone except the most committed pioneers or early adopters.

B. Software Limitations

The most efficient hardware is useless if it doesn't come with incredible software. Although the architecture of such AR systems has advanced significantly, the software component for such a complex system still has certain challenges to overcome and objectives to achieve [11]. The flexible operating system has to be coded, sized, and powered more efficiently. The following are some examples of operating systems that can be used in the context of AR (FreeRTOS, OpenWSN, and TynyOS). Additionally, there are dedicated browsers for AR, including Mozilla's "Firefox Reality." And several more devices are still under development. Therefore, it is clear that it is needed to design suitable toolkits having the support of various devices and applications across numerous platforms.

C. Lack of Privacy

The perceived risk associated with AR is one of the key privacy concerns. A person's privacy is at risk since AR technology can monitor what they are doing and collects a ton of data about them and their activities.

D. Devices Compatibility

In AR systems, there should be no issues with communication between objects and devices, the compatibility of such devices is one of the main issues with an AR system's design [15]. To address this issue, we should focus on "We must improve semantic exchange between objects and devices and between devices with each other. The semantic web may be used to do this, which can improve the quality of digital content seen via the user interface."

E. User Intervention

The dependency of AR apps on the user and their activity is another issue. Any IoT device must be independent and sensitive [11]. Its role needs to be less obvious to the user so that consumers may have a system that is much more robust, even when there are problems [15]. Fig. 1 presents the components of AR system.
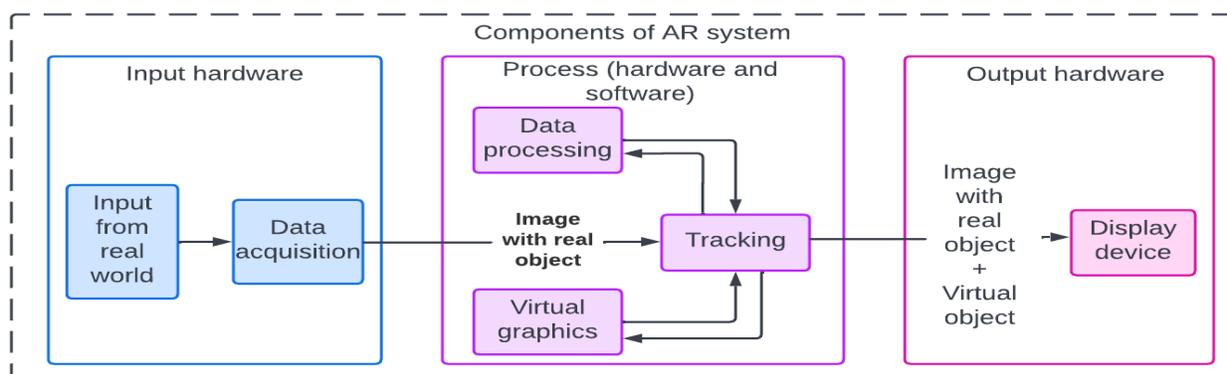
Fig. 1. AR System Components.

III. APPLICATIONS OF OBJECT DETECTION

This section will discuss some recent significant detection applications, such as the detection of pedestrians, faces, traffic signs and light, and texts.

A. Pedestrian Detection

Nowadays, pedestrian detection has received extensive research, which is closely related to pedestrian tracking person re-identification. Before the recent development in DCNN algorithms, some studies combined boosted decision forests with hand-crafted features to develop pedestrian detection methods. also, to address the deformation and occlusion, part-based methods and explicit handling of occlusion are problematic [16].

B. Face Detection

Face detection is vital for numerous face applications and behaves as an essential face recognition pre-processing step. One of the initial computer vision technologies that has successfully supported object detection is face detection, and many of its amazing principles are still having a big impact on object detecting technology today. Face detection is currently used in all parts of life, for digital cameras' "smile" detection, "face swiping" in online retail, face makeup, etc. [16].

C. Text Detection

The issues with text detection in video and images have attracted more notice recently, as shown by the appearance of recent" Robust Reading" competitions in 2003, 2005, 2011, and 2013. There are several main causes for this trend, including a rising number of applications. Text is one of the most artistic methods of communication and can be integrated into documents or scenes as a technique of connecting information. The same problems with computer vision and pattern recognition are connected to text identification in photographs caused by low quality or worse, despite the fact that many studies consider OCR technology (Optical Character Recognition) to be a settled subject. Text detection field still has a big space for search [17].

D. Traffic Sign and Traffic Light Detection

With the advancement of autonomous vehicle technologies, the detection of traffic signs and lights has received a lot of attention nowadays [16]. Even though the computer vision community has mostly highlighted, Traffic Light Recognition has now made a name for itself as a standard basis for pedestrian and general object detection, until 2010, the majority of traffic sign/light detection technologies, except a very small number of works, did not adhere to this paradigm.

IV. CHALLENGES OF OBJECT DETECTION

Finding and identifying various instances of videos and pictures is made easier with the use of a type of computer vision technology called object detection. The most typical object detection problems that data scientists should be aware of are listed below:

A. Small Object Detection

Small objects are typically difficult to detect because of their noisy depiction and low resolution. Currently used object identification pipelines often find small objects by learning the representations of every object at different scales. Typically, such ad hoc systems only provide performance improvements sufficient to offset their computing costs. Numerous real-world applications, such as pedestrian detection and traffic sign identification for improved autonomous driving, frequently involve small objects. However, the detection of small objects is more difficult than that of regular objects, and effective solutions are yet to be established [18]. Even though some of these problems with small object detection have received some attention, most of the proposed methods erroneously increase the feature dimension or enrich the data to improve the effectiveness of small object detection [19].

B. Object Detection Model Size

The present CNN frameworks are deepening and becoming more complex, despite the fact that the accuracy of such network architectures may match or exceed that of human vision, they frequently need extremely high amounts of computer power [20]. Fast object detection techniques have advanced significantly; however, it is still difficult to apply CNN architectures on non-GPU or mobile devices. The usage of CNN-based technologies has brought rise to major research topics in real-time and lightweight network models [21] [22] for object detection in mobile devices. This is driven by the rapid advancement in mobile and embedded intelligent gadgets with low power consumption and limited computing power, such as AR glasses and small intelligent unmanned aerial vehicles (UAVs).

C. Speed

Deep neural network-based methods have frequently performed better than other methods in object detection evaluations. There are two categories of these models - those with Regional Proposal Network (RPN layer) and those without RPN; those with the RPN layer are normally the faster R-CNN & R-FCN. YOLO and SSD are the models without it. YOLO and SSD are also called single-shot detection models. In general, the models with Region Proposal Network are more accurate while the models without Region Proposal Network and SSD are faster. The size of the model could affect the speed of detection so the pruning of model layers will speed up the model [23] [24]. The best models are the models that balance accuracy and speed.

D. Viewpoint Variation

Deep CNNs are only capable of modeling 2D transformation fields using the existing methods for encoding spatial invariance. This fact does not consider that objects in a 2D space are a projection of the ones in 3D and hence, have a limited power to severely shift object viewpoints is not taken into consideration by this [25]. Recently, CNN-based joint object recognition and viewpoint estimate has drawn attention as a potential solution for viewpoint variations. Before predicting the relative stiff transformation between each picture's 2D coordinate and the camera point in 3D space, must first identify the location and kind of objects in an image viewpoint. Estimation, and category classification problems are intrinsically incongruent.

E. Deformation

Sometimes interesting items may be flexible and distorted in unusual ways. Although a person can be detected by an object detector in a different situations, because of his twisted orientations, it will be difficult for the object detector to identify the same person.

The most advanced item detection technology currently available uses Deformable Parts Models (DPM). They do not appear to be the best at representing deformations. The deformations of objects are frequently continuous and not limited to large parts. Variations in popular and efficient object detectors are due to changes in appearance and deformations[26].

F. Occlusion

In real-world images and videos, occlusion is a frequent issue that presents a significant challenge for object detection. For instance, 70% of pedestrians in the Caltech Pedestrian Dataset, are obscured in at least one video sequence frame, and 19% are obscured in every frame. Almost 50% of these pedestrian occlusions were classified as heavy [27]. According to Dollar et al., even with minor occlusion, the performance of typical approaches to detect objects decreases significantly, and with heavy occlusion, it falls even further. Therefore, better performances are provided by object detectors that can learn and infer visible patterns by focus on fix the occlusion issue [28].

G. Shadow and Illumination Conditions

The alteration of illumination is a typical issue with object detection processes. Some techniques are offered to reduce the effect of changes in illumination and shadow caused by moving objects; an example of such techniques is the Moving Object Detection and Shadow Removing under Changing Illumination Condition model. The shadow causes several problems with object localization, segmentation, recognition, and tracking. The shadow may also result in the objects merging, it can cause the shapes of the objects to change, might even result in things going missing and lead to the foreground being confused for the background. It is equally challenging to capture clear moving objects when illumination changes because of the chances of mistaking some background pixels for foreground pixels. The effectiveness of subsequent procedures (such as tracking, recognition, classification, and activity analysis) that require precise moving object detection and accurate acquisition of its exact shape, therefore there is great impact by removing shadows and controlling illumination fluctuations [29, 30].

III. Previous Work

Numerous theoretical and empirical research has covered and elaborated on AR. This study focuses on approaches and procedures that can be utilized to enhance deep learning AR models. The following are previous studies about AR that were conducted by various studies:

In [31], Trident-YOLO, an upgraded version of the YOLOv4-tiny network with better accuracy and real-time speed was suggested. The most significant improvements made by this model are to the set of tools indicated by Alexey and the suggestions of CSP-RFBs and CSP-SPPs, which are appropriate for thin object detection networks. In order to increase the accuracy and recall of lightweight object detection, the network topology was redesigned, and a trident feature pyramid network (Trident-FPN) was suggested by the authors. This Trident-FPN produces a multi-scale feature map of the model while only slightly increasing the computational cost in terms of floating-point operations per second (FLOPs).

The study of [32] introduced the TRC-YOLO, TRC-YOLO proposed the pruning of the YOLOv4-tiny convolution kernel and the addition of an expansive convolution layer to the residual network model to produce an hourglass-shaped Cross-Stage partial Present (CSP) structure. The introduction of TRC-YOLO enhanced the mAP and real-time speed while minimizing the model size which was achieved by minimizing the number of YOLOv4-tiny model parameters. The CSP Res Net module was then enhanced and integrated to boost the model's capacity for feature extraction. In order to obtain higher quality feature images and to enable the model to concentrate on important feature areas and channels, the RFBs module was included to this model.

A mobile inverted bottleneck module is used as the foundation of the feasible and lightweight object detection model presented by [33]; the model was based on deep CNN. Additionally, an improved spatial pyramid pooling was used to concatenate the multi-scale local region characteristics to increase the network's receptive field. The testing results on the aerial picture datasets VEDAI and VisDrone show that the enhanced YOLOv4-tiny model performs significantly better than the original YOLOv4-tiny model. For the VEDAI and VisDrone datasets, the suggested model outperformed the results with mAP of 53.11 percent and 24.73 percent, respectively.

The study by [34] suggests embedded YOLO to enhance the efficiency of low-level features; the study initially suggested a new backbone network topology called the ASU-SPP network but later developed a more straightforward version of the neck network module PANet-Tiny to make computations simpler. Finally, depth-wise separable convolution was employed in the detecting head module to minimize the number of convolution stacks. The embedded YOLO model was compared with the conventional lightweight model after being verified by the COCO test dataset and online tests; it was discovered that the mAP performance was preserved.

In [21], a newly developed lightweight CSL-Module was presented; the new approach showed a comparative performance with previous approaches of a similar nature but due to limited computing resources, two additional components (CSL-Bone and CSL-FPN) were proposed to achieve superior performance with fewer FLOPs. However, achieving low computation cost depends on how the redundant features are generated as the CSL-Module can lower computation costs considerably. Research done at MS-COCO demonstrates that the suggested CSL-Module can approximate the fitting ability of Convolution-3x3.

A real-time object detection approach for non-GPU systems was proposed by [35] in order to help users of low-

configuration computers. Real-time object detection on CPU-based computers is now possible thanks to the optimization of YOLO with OpenCV for CPU-based computing. On some non-GPU machines, the CPU-based YOLO model can detect objects from videos with an accuracy of 80–99 percent and frame rates between 10.12 and 16.29. Comparison with other GPU-based frameworks showed that the proposed model is suitable for CPU-based applications because CPU Based YOLO obtains 31.05 percent mAP.

A study by [35] suggested a brand-new, DL-based lightweight object detection technique. Based on YOLOv4, the study proposed YOLOv4-tiny is proposed to simplify the network topology and minimize parameters, making it appropriate for development on mobile and embedded devices. The proposed approach achieved faster object detection compared to YOLOv4-tiny and YOLOv3-tiny as evidenced by the simulation results; it also achieved an almost similar mean value of average precision as the YOLOv4-tiny. The authors also proposed two identical ResBlock-D modules for the replacement of two CSPBlock modules in the YOLOv4-tiny network to reduce the object detection process. An auxiliary network block that employs two 3x3 convolutions networks, spatial attention, concatenate operation, and channel attention was also proposed for the global feature extraction in order to balance the object detection time and accuracy.

Mixed YOLOv3-LITE was developed by [22] as a mobile and non-GPU compatible lightweight real-time object detection network. The proposed approach supplements the ResBlocks and parallel high-to-low resolution subnetworks that are YOLO-LITE-based. The detector was developed using narrower and shallower convolutional layers compared to those in YOLOv3; this reduces the required level of computation and number of parameters to be trained, hereby improving the network operation speed. These considerations were aimed at solving the problems of limited computing power and excess power consumption in mobile and embedded smart devices.

In [23], a real-time object detection paradigm was developed for mobile devices like laptops or phones without a GPU. The study YOLOLITE to provide a smaller, quicker, and more effective model based on the original YOLOV2 algorithm; the introduction of YOLOLITE increased the accessibility of real-time object identification to a variety of devices. With its success in bringing object detection to non-GPU machines, YOLO-LITE demonstrates the enormous potential of shallow networks for lightweight real-time object detection networks. For such a modest system, running at 21 FPS on a non-GPU computer is quite encouraging, and it demonstrates why batch normalization must be queried for smaller shallow networks. It showcased the capability of shallow networks in fast non-GPU object detection devices. It also proves that shallow networks do not necessarily require batch normalization but reduce the network's overall performance. To sum up, the YOLO-LITE detector is composed of several parts:

1) Input: Image, Patches, Image Pyramid

2) Backbone: Darknet-53

3) Heads: Dense Prediction: one-stage.

A YOLObile framework that relies on compression-compilation co-design was designed in [36] to enable real-time object recognition on mobile devices. The study proposed a brand-new block-punched pruning approach for any kernel size. Furthermore, advanced compiler-assisted optimizations in conjunction with a GPU-CPU collaborative strategy was also suggested to increase computational efficiency on mobile devices. According to experimental findings, the pruning strategy successfully compressed YOLOv4 at a 14 rate with 49.0 mAP. On the Samsung Galaxy S20, a 17 FPS inference performance was reached using the proposed YOLObile framework with the GPU-CPU collaboration strategy. The proposed YOLObile also includes a mobile GPU-CPU collaborative computation strategy to increase the computational effectiveness of DNNs on mobile gadgets. The evaluation shows that the proposed YOLObile framework achieved high hardware parallelism and excellent accuracy. Fig. 2 summarized the percentage of previous work studies that focus on improve each attribute.
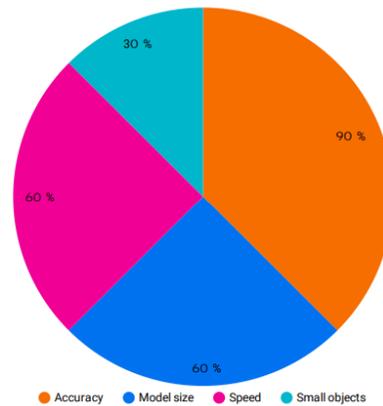
Fig. 2. Applied Approaches.

V. OUTCOME AND FINDINGS

The main finding of this paper is that CNNs object detection models are through an advanced transformation in the field of computer vision. The development of object detection models to be appropriate even for mobile and devices without GPU, which offers customized models and small-scale proof-of-concept studies, is what highlights this transformation. These models' studies provide a foundation for large-scale scientific research issues that employ CNNs. As a result of this tendency, CNNs object detection models will eventually be able to compile a worldwide, digital inventory of everything in order to study object dynamics and their effects on human activity. This study analyzes 10 articles to give a thorough overview of the convolutional neural network (CNN) for object detection in general. Another finding is that always there is a trade-off between speed and accuracy. When focus on increasing accuracy this will make the model more complex which will slow the model down.

TABLE I.   SUMMARY OF PREVIOUS WORK

Reference	Method	Summary	Evaluation Metrics	Dataset Used	Limitations
[31]	Trident-YOLO	Introduce an object detection network that is designed for mobile devices with limited computing power	FPS mAP BFLOPs	PASCAL VOC MS COCO	There is still room for improvement in the detection of complex scenes.
[32]	TRC-YOLO	Proposed a lightweight object detection model based on yolov4-tiny.	FPS mAP	PASCAL VOC MS COCO	Model speed needs to be higher
[33]	Automated object detection on aerial images for limited capacity embedded device	Feasible and lightweight object detection model based on deep CNN	FPS mAP input size F1-score IoU FLOPs	VisDrone VEDAI	The model was based on yolov4 tiny and slower than yolov4 tiny.
[34]	Embedded YOLO	Proposed an ultralightweight target detection network model	FPS mAP parameters Size Latency	MS-COCO	The model's map needs to improve.
[21]	CSLYOLO	Proposed a new lightweight convolution method CSL-Module and proposed two components CSL-Bone and CSL-FPN, both of which achieve better performance with fewer flops.	FPS mAP Params FLOPs	MS-COCO	The model's running speed needs to improve
[42]	CPU Based YOLO	Optimized YOLO with OpenCV to enable real-time object detection on CPU-based devices.	FPS mAP	COCO	Model speed needs to be higher
[24]	YOLOv4-tiny	Proposed Yolov4-tiny based on Yolov4 to reduce the network structure and ensure suitability with mobile and embedded devices.	FPS mAP	MS COCO	Still not implementable for non-GPU devices
[22]	Mixed YOLOv3-LITE	Proposed "an efficient lightweight object detection network that uses a shallow-layer, narrow-channel, and multi-scale feature image parallel fusion" structure	FPS mAP Params GFLOPs Precision Recall F1	PASCAL VOC 2007 & 2012	Need to improve the map
[23]	YOLO-LITE	The proposed detector model runs at "about 21 FPS on a non-GPU computer and 10 FPS after implemented onto a website with only 7 layers and 482 million" FLOPS.	FPS mAP Layers FLOPS Loss	PASCAL VOC 2007 and 2012 COCO 2014	Need to improve the map
[36]	YOLObile	Proposed yolobile framework, a real-time object detection on mobile devices via compression-compilation co-design.	Input Size Backbone Weights FLOPs mAP FPS	COCO dataset	Need to improve detection speed

A. Applied Approaches

The approaches utilized to address the difficulties and resolve the object detection problems are covered in depth in Section V. Tables I and II summarized each of the publications reviewed and how they addressed the difficulties faced by the researchers. As seen in Table II and Fig. 1, most research focus on improving accuracy (90 percent), handling speed (60 percent) have received far more attention than any other problems.

B. Investigated Datasets

Datasets are significant motivators for specific applications and crucial for the development of deep-learning algorithms. The number of case studies and datasets considered has proven to be challenging. It is advantageous and necessary to use a number of datasets, each of which supports a range of parameters and a predetermined composition problem, in order to assess the efficiency of the suggested algorithms. There are many object detection datasets that are readily available in the research domain: COCO [37] [23] [24] [35, 36] http://cocodataset.org/ and PASCAL VOC [23] [38-41] http://host.robots.ox.ac. uk/pascal/VOC/., cifar-10 [42-46]

https://www.cs.toronto.edu/~kriz/cifar.html, bdd100k dataset [47-51] https://www.bdd100k.com/, LISA Traffic Sign Detection Dataset[52-56], KITTI [49, 57-60] http://www.cvlibs.net/datasets/kitti/, SUN-RGB-D [61-65]https://rgbd.cs.princeton.edu/,nuScenes[66-70] https://www.nuscenes.org/, Visual genome [71-75] https://visualgenome.org/ , MPII [76-80] http://human-pose.mpi-inf.mpg.de/ , and imagine [81-85] https://image-net.org/index.php. Some researchers have rarely relied on datasets that are generated synthetically.

VI. LIMITATIONS OF OBJECT DETECTION IN APPLIED APPROACHES

This section discusses a variety of essential object detection techniques as well as object detection issues and constraints. The main problems with object detection systems are their speed, accuracy, difficulty in detecting small objects, and model size. Numerous algorithms have been developed to address these issues, but none of them are completely successful. The enumerated algorithms provide the following solutions to the problems.

- Accuracy

A common metric used to assess the accuracy of object detection models is Mean Average Precision (mAP). The mAP is calculated by comparing the ground-truth bounding box to the detected box; higher mAP scores imply that the associated models have higher detection accuracies. The accuracy metric is the most important factor to assess any object detector. The new models used improved techniques in Neck part like (FPN, PANet, and SPP) and the Residual Blocks in Backbone part also using advanced Head part like head in YOLOv4 to increase the accuracy of bounding boxes position. All these techniques and more are utilized to improve the accuracy of detectors like [23, 24, 36].

- Speed

Beside the accuracy factor, detector speed should be taken in account. The effectiveness of any intelligent system and AR gadgets depends on an effective and quick object detection algorithm. Without real-time detection the model will be useless in field of autonomous vehicles and many other fields. To tackle this issue many researchers suggested new features for example (simple the network structure, reduce parameters, and using simple parts instead of complex ones) in YOLOv4-

tiny the author used FPN without using SPP which used in YOLOv4, additionally YOLO-LIE model minimize YOLOv2 by using less number of backbone layers. However, A constant trade-off always exists between speed and accuracy in larger models [23, 36].

- Small object detection

The detection of small objects is a difficult computer vision task and has been widely used in defense, military, transportation, industry, etc. to improve understanding of small object detection. Object detection has recently made tremendous strides but despite these advancements, there is still significant variation in performance between the detection of large and small objects. To solve the issue of small object detection, some models have been developed, such as those developed in [21, 22, 36] in [21] CSL-YOLO model, because of the huge number of small objects in MS-COCO dataset the author improved new version of feature pyramid network called (CSL-FPN) in this Network Before K-means, a scale limit like Eq. (3) has been included so that the distribution of anchors produced is more in line with the scale of each output layer. In the results, using CSL-FPN has increase the accuracy of detect small objects. However, small objects detection still needs more studies and it's an open issue need more efforts to be improved.

- Model size

Object detection model size has many effects on the speed and accuracy, the trade-off between accuracy in larger models and speed in lightweight versions is continual. The simple network structure and fewer parameters make it ideal for mobile and embedded devices. On other hand the complex structure and more parameters may be caused to improve model accuracy. Modern object detection techniques used in cars now rely significantly on the sensor output from costly radars & depth sensors that make them unsuitable for usage in everyday situations. However, this increase in accuracy may not be useful to address the problem in many real-world applications that demand real-time. performance carried out on a platform with restricted computational resources [22-24, 35]. Simplicity of detector is an open issue due to need of lightweight models in mobile and non-GPU devices. In Table II we can see the summary of improved attributes for 10 studies.

TABLE II. IMPROVED ATTRIBUTES

Author	Year	Method	Speed	Accuracy	Model Size	Small objects
Wang, et al., [31]	2022	Trident-YOLO	✓	✓		
Wang, et al., [32]	2021	TRC-YOLO	✓	✓		
Junos et al., [33]	2022	Automated object detection on aerial images for limited capacity embedded device			✓	✓
Feng et al., [34]	2021	Embedded YOLO	✓	✓		
Zhang et al., [21]	2021	CSLYOLO		✓		✓
Ullah et al., [35]	2020	CPU Based YOLO		✓	✓	
Jiang et al., [24]	2020	YOLOv4-tiny	✓	✓	✓	
Zhao et al., [22]	2020	Mixed YOLOv3-LITE	✓	✓	✓	✓
Huang and Chen [23]	2020	YOLO-LITE	✓	✓	✓	
Cai et al., [36]	2020	YOLObile		✓	✓	

VII. CONCLUSION

The study focused on CNN-based light object detection in the field of AR, and it introduces the CNN's structure, the CNN-based object detection framework, and several techniques for enhancing detection performance. Surveys on the application domains, model components, experience metrics, used datasets, and model performance of lightweight object detection models were also conducted; the challenging problems in the field of AR were also highlighted. Nevertheless, there are still several technical and application-related issues in AR. The performance of CNN in terms of real-time, accuracy, and adaptability was better than that of the traditional approaches, but there is still much potential for improvement. Enhancing the object detection algorithm's structure can minimize the loss of feature information while fully leveraging the relationships between the object and the context. Numerous studies have already been done to address non-GPU and embedded devices. Although the improved methods performed better than the conventional methods, the accuracy still needs to be improved to handle a complex environment. A well-established method specifically for addressing the issue of small object detection is yet to be achieved, however, improvisation in small object methods allows for the achievement of acceptable accuracy values, though it is vulnerable to additional processing time. New methods can be developed in the future by leveraging the strength of the recent trends for improved performance. The current methodologies, for instance, can be improved by hybridizing non-GPU approaches and small object detection with GPU-based approaches.

ACKNOWLEDGMENT

REFERENCES

[1] Chiang, F.-K., X. Shang, and L. Qiao, Augmented reality in vocational training: A systematic review of research and applications. Computers in Human Behavior, 2022. 129: p. 107125.

[2] Tan, S.Y., H. Arshad, and A. Abdullah, An improved colour binary descriptor algorithm for mobile augmented reality. Virtual Reality, 2021. 25(4): p. 1193-1219.

[3] Gao, X., et al., Effects of Augmented-Reality-Based Assisting Interfaces on Drivers' Object-wise Situational Awareness in Highly Autonomous Vehicles. arXiv preprint arXiv:2206.02332, 2022.

[4] Scavarelli, A., A. Arya, and R.J. Teather, Virtual reality and augmented reality in social learning spaces: a literature review. Virtual Reality, 2021. 25(1): p. 257-277.

[5] Lee, Y.-H., T. Zhan, and S.-T. Wu, Prospects and challenges in augmented reality displays. Virtual Real. Intell. Hardw., 2019. 1(1): p. 10-20.

[6] Nizam, S.M., et al., A review of multimodal interaction technique in augmented reality environment. Int. J. Adv. Sci. Eng. Inf. Technol, 2018. 8(4-2): p. 1460.

[7] Goh, E.S., M.S. Sunar, and A.W. Ismail, 3D object manipulation techniques in handheld mobile augmented reality interface: A review. IEEE Access, 2019. 7: p. 40581-40601.

[8] Slater, M., et al., The ethics of realism in virtual and augmented reality. Frontiers in Virtual Reality, 2020. 1: p. 1.

[9] Liu, L., H. Li, and M. Gruteser. Edge assisted real-time object detection for mobile augmented reality. in The 25th annual international conference on mobile computing and networking. 2019.

[10] Li, X., et al. Object detection in the context of mobile augmented reality. in 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). 2020. IEEE.

[11] Arena, F., et al., An Overview of Augmented Reality. Computers, 2022. 11(2): p. 28.

[12] Viglialoro, R.M., et al., Review of the augmented reality systems for shoulder rehabilitation. Information, 2019. 10(5): p. 154.

[13] Ping, J., Y. Liu, and D. Weng. Comparison in depth perception between virtual reality and augmented reality systems. in 2019 ieee conference on virtual reality and 3d user interfaces (vr). 2019. IEEE.

[14] Jeffri, N.F.S. and D.R.A. Rambli, A review of augmented reality systems and their effects on mental workload and task performance. Heliyon, 2021. 7(3): p. e06277.

[15] Solbiati, M., et al., Augmented reality for interventional oncology: proof-of-concept study of a novel high-end guidance system platform. European radiology experimental, 2018. 2(1): p. 1-9.

[16] Zhao, Z.-Q., et al., Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems, 2019. 30(11): p. 3212-3232.

[17] Ye, Q. and D. Doermann, Text detection and recognition in imagery: A survey. IEEE transactions on pattern analysis and machine intelligence, 2014. 37(7): p. 1480-1500.

[18] Kisantal, M., et al., Augmentation for small object detection. arXiv preprint arXiv:1902.07296, 2019.

[19] Liu, Y., et al., A survey of research and application of small object detection based on deep learning. Acta Electonica Sinica, 2020. 48(3): p. 590.

[20] Zaidi, S.S.A., et al., A survey of modern deep learning based object detection models. Digital Signal Processing, 2022: p. 103514.

[21] Zhang, Y.-M., et al., CSL-YOLO: A New Lightweight Object Detection System for Edge Computing. arXiv preprint arXiv:2107.04829, 2021.

[22] Zhao, H., et al., Mixed YOLOv3-LITE: a lightweight real-time object detection method. Sensors, 2020. 20(7): p. 1861.

[23] Huang, R., J. Pedoeem, and C. Chen. YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers. in 2018 IEEE International Conference on Big Data (Big Data). 2018. IEEE.

[24] Jiang, Z., et al., Real-time object detection method based on improved YOLOv4-tiny. arXiv preprint arXiv:2011.04244, 2020.

[25] Wang, X., K. Wang, and S. Lian. Deep consistent illumination in augmented reality. in 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). 2019. IEEE.

[26] Mordan, T., et al., End-to-end learning of latent deformable part-based representations for object detection. International Journal of Computer Vision, 2019. 127(11): p. 1659-1679.

[27] Hebborn, A.K., N. Höhner, and S. Müller. Occlusion matting: realistic occlusion handling for augmented reality applications. in 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). 2017. IEEE.

[28] Wang, A., et al. Robust object detection under occlusion with context-aware compositionalnets. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[29] Shakeri, M. and H. Zhang. Moving object detection under discontinuous change in illumination using tensor low-rank and invariant sparse decomposition. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

[30] Tan, S.Y., H. Arshad, and A. Abdullah, A new illumination invariant feature based on freak descriptor in RGB color space. Journal of Theoretical and Applied Information Technology, 2016. 93(1): p. 207.

[31] Wang, G., et al., Trident‑YOLO: Improving the precision and speed of mobile device object detection. IET Image Processing, 2022. 16(1): p. 145-157.

[32] Wang, G., et al., TRC‑YOLO: A real‑time detection method for lightweight targets based on mobile devices. IET Computer Vision, 2021.

[33] Junos, M.H., A.S.M. Khairuddin, and M. Dahari, Automated object detection on aerial images for limited capacity embedded device using a

lightweight CNN model. Alexandria Engineering Journal, 2022. 61(8): p. 6023-6041.

[34] Feng, W., et al., Embedded YOLO: A Real-Time Object Detector for Small Intelligent Trajectory Cars. Mathematical Problems in Engineering, 2021. 2021.

[35] Ullah, M.B. CPU Based YOLO: A Real Time Object Detection Algorithm. in 2020 IEEE Region 10 Symposium (TENSYMP). 2020. IEEE.

[36] Cai, Y., et al., Yolobile: Real-time object detection on mobile devices via compression-compilation co-design. arXiv preprint arXiv:2009.05697, 2020.

[37] Lin, T.-Y., et al. Microsoft coco: Common objects in context. in European conference on computer vision. 2014. Springer.

[38] Redmon, J. and A. Farhadi. YOLO9000: better, faster, stronger. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[39] Wang, X., A. Shrivastava, and A. Gupta. A-fast-rcnn: Hard positive generation via adversary for object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[40] Redmon, J., et al. You only look once: Unified, real-time object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[41] Ren, S., et al., Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015. 28.

[42] Salimans, T., et al., Improved techniques for training gans. Advances in neural information processing systems, 2016. 29.

[43] He, K., et al. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[44] Zoph, B., et al. Learning transferable architectures for scalable image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[45] Chauhan, K., et al. Robust outlier detection by de-biasing VAE likelihoods. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[46] Huang, G., et al., Multi-scale dense networks for resource efficient image classification. arXiv preprint arXiv:1703.09844, 2017.

[47] Wang, J., et al., Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 2020. 43(10): p. 3349-3364.

[48] Wu, D., et al., Yolop: You only look once for panoptic driving perception. arXiv preprint arXiv:2108.11250, 2021.

[49] Cai, Z. and N. Vasconcelos, Cascade R-CNN: high quality object detection and instance segmentation. IEEE transactions on pattern analysis and machine intelligence, 2019. 43(5): p. 1483-1498.

[50] Hou, Y., et al. Learning lightweight lane detection cnns by self attention distillation. in Proceedings of the IEEE/CVF international conference on computer vision. 2019.

[51] Tran, L.-A., et al., Robustness Enhancement of Object Detection in Advanced Driver Assistance Systems (ADAS). arXiv preprint arXiv:2105.01580, 2021.

[52] Møgelmose, A., D. Liu, and M.M. Trivedi. Traffic sign detection for us roads: Remaining challenges and a case for tracking. in 17th International IEEE Conference on Intelligent Transportation Systems (ITSC). 2014. IEEE.

[53] Møgelmose, A., D. Liu, and M.M. Trivedi, Detection of US traffic signs. IEEE Transactions on Intelligent Transportation Systems, 2015. 16(6): p. 3116-3125.

[54] Lopez-Montiel, M., et al. Evaluation of algorithms for traffic sign detection. in Optics and Photonics for Information Processing XIII. 2019. SPIE.

[55] Zhang, J., et al., A cascaded R-CNN with multiscale attention and imbalanced samples for traffic sign detection. IEEE access, 2020. 8: p. 29742-29754.

[56] Brown, W.S., K. Roy, and X. Yuan. US Traffic Sign Recognition Using CNNs. in Proceedings of SAI Intelligent Systems Conference. 2020. Springer.

[57] Chen, S., et al. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. in Chinese Conference on Biometric Recognition. 2018. Springer.

[58] Shi, S., et al. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[59] Zhu, B., et al., Class-balanced grouping and sampling for point cloud 3d object detection. arXiv preprint arXiv:1908.09492, 2019.

[60] Simon, M., et al. Complexer-yolo: Real-time 3d object detection and tracking on semantic point clouds. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.

[61] Qi, C.R., et al. Deep hough voting for 3d object detection in point clouds. in proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[62] Misra, I., R. Girdhar, and A. Joulin. An end-to-end transformer model for 3d object detection. in Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[63] Xie, Q., et al. Mlcvnet: Multi-level context votenet for 3d object detection. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[64] Zhao, X., et al. Object detection with a unified label space from multiple datasets. in European Conference on Computer Vision. 2020. Springer.

[65] Huang, S., et al., Perspectivenet: 3d object detection from a single rgb image via perspective points. Advances in neural information processing systems, 2019. 32.

[66] Stäcker, L., et al. Deployment of Deep Neural Networks for Object Detection on Edge AI Devices with Runtime Optimization. in Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[67] Yang, Z., et al. 3d-man: 3d multi-frame attention network for object detection. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[68] Han, J., et al., SODA10M: A Large-Scale 2D Self/Semi-Supervised Object Detection Dataset for Autonomous Driving. arXiv preprint arXiv:2106.11118, 2021.

[69] Chen, Q., S. Vora, and O. Beijbom, PolarStream: Streaming Lidar Object Detection and Segmentation with Polar Pillars. arXiv preprint arXiv:2106.07545, 2021.

[70] Zhou, Y., et al. Monocular 3d object detection: An extrinsic parameter free approach. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[71] Kehat, G. and J. Pustejovsky. Neural Metaphor Detection with Visibility Embeddings. in Proceedings of* SEM 2021: The Tenth Joint Conference on Lexical and Computational Semantics. 2021.

[72] Bosch, M., et al., Contextual Sense Making by Fusing Scene Classification, Detections, and Events in Full Motion Video. arXiv preprint arXiv:2001.05979, 2020.

[73] Zhou, L., et al., Object relation detection based on one-shot learning. arXiv preprint arXiv:1807.05857, 2018.

[74] Schwarz, M., et al., RGB-D object detection and semantic segmentation for autonomous manipulation in clutter. The International Journal of Robotics Research, 2018. 37(4-5): p. 437-451.

[75] Yu, R., et al. Visual relationship detection with internal and external linguistic knowledge distillation. in Proceedings of the IEEE international conference on computer vision. 2017.

[76] Simon, T., et al. Hand keypoint detection in single images using multiview bootstrapping. in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2017.

[77] Rogez, G., P. Weinzaepfel, and C. Schmid, Lcr-net++: Multi-person 2d and 3d pose detection in natural images. IEEE transactions on pattern analysis and machine intelligence, 2019. 42(5): p. 1146-1161.

[78] Luvizon, D.C., H. Tabia, and D. Picard, Human pose regression by combining indirect part detection and contextual information. Computers & Graphics, 2019. 85: p. 15-22.

[79] Sumer, O., T. Dencker, and B. Ommer. Self-supervised learning of pose embeddings from spatiotemporal relations in videos. in Proceedings of the IEEE International Conference on Computer Vision. 2017.

[80] Sekii, T. Pose proposal networks. in Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[81] Simonyan, K. and A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[82] Chen, B.-C., et al. Efficient object embedding for spliced image retrieval. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[83] Gupta, V., et al., Deep learning-based automatic detection of poorly positioned mammograms to minimize patient return visits for repeat imaging: A real-world application. arXiv preprint arXiv:2009.13580, 2020.

[84] Gao, Y., et al., Utilizing the instability in weakly supervised object detection. arXiv preprint arXiv:1906.06023, 2019.

[85] Joung, S., et al. Cylindrical convolutional networks for joint object detection and viewpoint estimation. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

The Influence of Virtual Secure Mode (VSM) on Memory Acquisition

Niken Dwi Wahyu Cahyani[1], Erwid M Jadied[2], Endro Ariyanto[4]

Informatics Faculty
Telkom University
Bandung, Indonesia

Nurul Hidayah Ab Rahman[3]

Centre for Information Security Research
Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia
Parit Raja, Malaysia

Abstract—Recently, acquiring the Random Access Memory (RAM) full memory and access data is gaining significant interest in digital forensics. However, a security feature on the Windows operating system - Virtual Secure Mode (VSM) - presents challenges to the acquisition process by causing a system crash known as a Blue Screen of Death (BSoD). The crash is likely to occur when memory acquisition tools are being used. Subsequently, it disrupts the goal of memory acquisition since the system must be restarted, and the RAM content is no longer available. This study analyzes the implications of VSM on memory acquisition tools as well as examines to what extent its impact on the acquisition process. Two memory acquisition tools, namely FTK Imager and Belkasoft RAM Capturer, were used to conduct the acquisition process. Static and dynamic code analyses were performed by using reverse engineering techniques that are disassembler and debugger. The results were compared based on the percentage of unreadable memory between active and inactive VSM. Static analysis showed that there is no difference between all applications' functions for both active and inactive VSM. Further Bugcheck analysis of the MEMORY.DMP is pointed to the ad_driver.sys module in FTK Imager that causes the system to crash. The percentage of unreadable memory while running on active VSM and inactive VSM for Belkasoft is about 0.6% and 0.0021%, respectively. These results are significant as a reference to digital investigators as consistent with the importance of RAM dump in live forensics.

Keywords—Live forensics; memory acquisition; virtualization; virtual secure mode

I. INTRODUCTION

As defined by the Digital Forensics Research Workshop (DFRWS), digital forensics is the use of scientifically derived and proven methods to preserve, collect, validate, analyze, and present admissible digital data that meet the court requirements [1]. Digital data originated from electronic devices that have data storage capability, including smartphones, digital cameras, and even printers. There are two types of digital data, namely: (i) volatile data – data that will be lost when there is no electrical power on the devices, and (ii) non-volatile data – data that is still stored in the device's storage media even though the power is turned off. RAM forensics or memory forensics involves collecting and examining volatile data. It becomes a priority to undertake the live acquisition if an electronic device is on, considering the data will be lost when the device is turned off. Furthermore, some cyber security incidents require RAM forensics such as malware attacks, due to its behavior

that could leave no trail on non-volatile memory [2]. As an example, a study [3] was able to identify Advanced Encryption Standard (AES) keys in the memory of a ransomware process by examining memory dumps using live forensics tools. It further indicates that artifacts from memory forensics are not limited to evidence collection, yet they could be utilized to minimize the impact of cyber incidents.

While there has been significant development in advanced computing architecture, it poses challenges to memory forensics practices. For example, the use of a recent security feature known as Virtual Secure Mode (VSM), which was started from Windows 10 and Windows Server 2016 operating system, complicates the acquisition of volatile data in memory. It has been highlighted in [4] that the use of some acquisition tools (e.g., Magnet RAM Capturer, FTK Imager) to undertake live forensics causes the system to crash. Subsequently, the volatile data (i.e., the initial object of the acquisition) is no longer available since the operating system will restart the system [4], [5]. However, much work remains to be done in the technical analysis such as what happens to the system when the VSM feature is active that affects the tools during the memory acquisition process. This motivates the direction of this study to carry out further investigations on the VSM environment.

It has been noted that not all memory acquisition tools can complete the acquisition process during an active VSM environment. Therefore, this study aims to conduct a technical analysis of the VSM effects on the live memory acquisition process using two cases. The first case is a successful memory acquisition by using the Belkasoft RAM Capturer tool, and the second case is an unsuccessful memory acquisition by using the FTK Imager tool. Reverse engineering techniques are applied to analyze the behavior of the system. The main methods used are static and dynamic code analysis using IDA disassembler and Windbg debugger. Additionally, event analysis is conducted by examining event logs collected by the operating system to facilitate our understanding of the impact.

While previous works have been studying VSM and identifying the BSoD for live forensic tools, our study may become the initial research investigating the impact of VSM on live forensics. The results are discussed with technical data produced from reverse engineering techniques. Static analysis results can be used to understand the tools' program code that could directly lead to crash events. The dynamic analysis could demonstrate the tools' behavior in their running state that

may (or may not) cause the impact, and how they interact with the operating system as the manager of the computer system including memory. We contribute to the underlying methodology that applies static, dynamic, and event analysis in examining the behavior of VSM and how it impacts the running memory acquisition tools.

The remaining of this paper is organized as follows: Section II discusses the related work, Section III describes the materials and methods employed in this study, Section IV discusses the results, Section V presents the conclusion, and Section VI highlights the limitations and future work.

II. RELATED WORK

Collecting, and preserving the data of Random Access Memory (RAM) for forensic analysis is considered critical in live forensics. It contains many valuable forensic interest artifacts, including processes running on the computer. Examples of the content's use are to examine security incidents and get data from encrypted containers when it is being opened. The importance of memory forensic acquisition has attracted significant interest in recent years. Arfeen et al. [6] developed a framework for memory acquisition periodically to analyze process behavior while it is running and reside in memory to help ransomware detection. Prakoso et al. [7] examined how Metasploit attacks on Windows 10 can be analyzed using live forensics techniques on the volatile memory. The study used three well-known RAM acquisition tools, namely: FTK Imager, Dumpit, and Magnet RAM Capture. Volatility was used as the analysis tool. The results showed that RAM's live forensics can obtain key artifacts including the attacker's IP address and evidence of malware. Kazim et al. [8] identified chat artifacts of an instant messaging tool including master encryption keys that are encrypted by Bitlocker and Truecrypt, from memory dumps of Windows 7 computers. The memory dump was been analyzed using analysis tools such as Volatility and Rekall [9], [10]. The results confirm the necessity of deploying mechanisms to collect RAM from local and remote systems to support the RAM acquisition, for incident responder teams.

Choosing the appropriate tools for the acquisition and analysis of memory forensics depends upon the compatibility between digital devices and operating systems, which may pose challenges to investigators [11]. Therefore, many existing studies have attempted to understand the strengths and weaknesses of memory acquisition tools. A study in [12] compared four tools, namely Windows Memory Reader, Belkasoft's Live Ram Capturer, ProDiscover, and FTK Imager, to examine their performance in capturing memory including their ease of use. Another study in [13] showed the differences in processing time, memory usage, registry key, and DLL for FTK Imager, Belkasoft RAM Capturer, Memoryze, DumpIt, and Magnet RAM Capturer. Similarly, [14] also examined how the combination of Belkasoft RAM Capturer, FTK Imager, and Winhex can be utilized to obtain data for the Line app in Windows 8.1. Prakoso et al. [7] identified that FTK Imager, Dumpit, and Magnet RAM Capture, have the same performance in acquiring the targeted artifact of a Metasploit attack in Windows 10 based on their acquisition results comparison.

With the important role of memory acquisition and analysis in digital forensics, it indicates that any issues that may hinder these processes shall be examined, including VSM. VSM is a Windows 10 technology for creating and managing a secure operating system environment [15]. The secure environment is designed to be a place for the execution of critical security functions, protecting it from attacks directed against the operating system. VSM uses virtualization as its base [16].Virtualization on a machine run by an emulator, commonly known as a hypervisor. Microsoft gives a particular name to its hypervisor system which is Hyper-V, while the virtual machine is known as a partition (e.g., Partition A and Partition B). Hyper-V virtualizes hardware resources for each partition and manages these virtual resources, including virtual memory and CPU.

Details architecture of a Windows environment that supports VSM shows that Hyper-V occupies the root partition [16]. The partition houses two environment modes, namely: (i) kernel, and (ii) user. Each environment operates on a separate domain, called the Virtual Trust Level (VTL). VTL enforces isolation in three aspects. First is memory access in which each VTL has a set of memory access protections that prevent an allocated VTL's memory from being accessed by entities in another VTL. Second is the virtual processor state, where each VTL has a set of private virtual processor registers associated with it. Third, interrupt in which each VTL has a separate interrupt system to prevent interference from entities operating in other VTLs during sending and processing of interrupts.

A study on Windows 10 reported that Hyper-V implements two VTLs: VTL 0 and VTL 1[16]. VTL 0 hosts a traditional Windows environment. Users running in VTL 0 are referred to as normal environment users, while the running kernels are known as normal kernels. VTL 1, on the other hand, is the place for the Windows environment to perform security-critical functionality. The environment is referred to as a safe environment. VTL-based memory access protection enforced by Hyper-V can be further referred to in [16]. The study shows the memory region's contents that are part of the memory dump of a VSM-enabled Windows environment mapped to the lsaIso.exe trustlet. The question mark character ('?') indicates unreadable memory because it cannot be accessed beyond the isolation limits implemented by VTL 1, where lsaIso.exe operates. A report in [4] discussed the effect of VSM that causes BSoD on several tools including Magnet RAM Capture v1.1.1 and FTK Imager Lite v3.1.1, however other tools such as Belkasoft RAM Capturer and Passmark osforensics v 5.1.1001 were not included.

There are existing studies that have demonstrated static code analysis and dynamic code analysis. For example, a study by Hirst [17] showed an acquisition test on no-quiescent virtual machines that utilized dynamic code analysis. Another study in [18] identified memory acquisition challenges that misuse two architectural features, which are physical address layout and secure container. The authors acknowledged them as a new class of anti-memory forensic techniques. Significantly, these studies provided key guidance on the methods that can be referred to conduct testing and observation.

Yehuda et al. [19] proposed a hypervisor-based memory acquisition tool by extending the Volatility framework and implementing it in ARM64-bit kernels. The authors showed how their proposed tool can reduce the processor's consumption, maintain the coherent state of the memory dump, and generate fewer tradeoffs for network and disk acquisition. The tool successfully conducts memory acquisition without facing any difficulties caused by security and privilege levels in Linux OS and ARM processors, called Trust Zone which divides accesses into secure and non-secure ones.

Nevertheless, the study on technical analysis of the VSM effect on the live memory acquisition process is still limited. There have been significant studies of VSM architecture on Windows 10, including the details of VSM initialization activity performed by the Windows loader during the boot process, and the communication interface on VSM [16], [20]. However, the explanation still lacks the technical impacts of VSM on the memory acquisition process. In this study, therefore, we attempt to examine the memory acquisition in Windows-based OS, especially in Windows 10 that enabled the VSM feature in Intel machines to manage virtual trust levels for kernel and user processes.

III. RESEARCH METHOD

This study applied reverse engineering as it is a widely recognized technique in digital forensics to process and interpret data [21], [22]. The methods used to analyze the impact of VSM on the memory acquisition tool are static and dynamic code analysis using the IDA disassembler and windbg debugger tools.

Event analysis is conducted using the operating system's event logs for further correlation with the findings from the static and dynamic code analysis. The complete research stages are presented in Fig. 1. The hardware and software specifications used in this research are presented in Tables I and II, respectively.

Experiments in this study are conducted in two environments, (1) VSM-enabled, and (2) non-VSM-enabled. VSM feature is enabled through the BIOS by setting the "Intel Virtualization Technology" option to "Enabled." The BIOS used in this study is from the American Megatrends vendor, version 309, with VBIOS Version 1054.I021x441UAR.002.
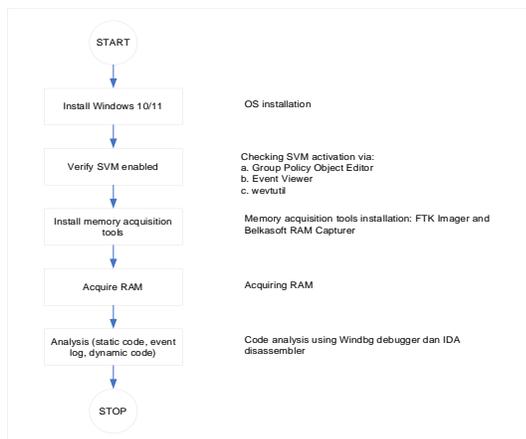
Fig. 1.   Research Stages.

TABLE I.   HARDWARE SPECIFICATION

No	Name	Specification	Function
1	Processor	Intel(R) Core(TM) i3-7020U CPU @ 2.30GHz 2.30 GHz	To execute a program
2	RAM	20,0 GB (19,9 GB usable)	To store data and instructions for a process

TABLE II.   SOFTWARE SPECIFICATION

No	Name	Specification	Function
1	Microsoft Windows 10	Edition: Windows 10 Enterprise Version: 21H1 OS build: 19043.1526 Experience: Windows Feature Experience Pack 120.2212.4170.0	Operating System
2	AccessData FTK Imager	Version 4.5.0.3	Acquisition Tools
3	Belkasoft RAM Capturer	Modified date 22/10/2018	Acquisition Tools
4	IDA PRO	7.5 SP3 x64	Disassembler
5	Diaphora	Version 2	Program diffing tool
6	Windbg Preview	Version 1.2202.7001.0	Debugger
7	Event Viewer	Version 1.0	Event Viewer

The steps taken to obtain data to be analyzed are presented in Table III.

TABLE III.   EXPERIMENTS AND ANALYSIS STEPS

Static Code Analysis		
No		*Steps*
1		VSM feature setting [Active/Non-Active]
	a	Disassembling executable files of memory acquisition apps (.exe) [FTK Imager/Belkasoft RAM Capturer] by using IDA PRO
	b	Running the diffting plugin to compare the apps' functions in VSM vs. non-VSM environments by using Diaphora on IDA PRO
2		Save the results
Note: 4 (four) files were produced: Assembly codes for: • FTK Imager • Belkasoft RAM Capturer SQLite files for: • FTK Imager • Belkasoft RAM Capturer		
Event Log Analysis		
No		Steps
1		Acquiring memory by using FTK Imager and Belkasoft RAM Capturer in active VSM and non-active VSM environments
2		Opening Event Viewer
3		Copying event log:
	a	Event Application
	b	Event Security
	c	Event System
Note: 3 (three) files were produced: Application, Security dan System event logs		
Dynamic Code Analysis		
No		Steps
1		Choosing the target executable files
2		Preparing the required symbols
3		Starting the debugging

IV. RESULTS AND DISCUSSION

This section presents the results of experiments and discusses the memory acquisition, static code, event log, and dynamic code analyses that have been carried out.

A. Memory Acquisition Analysis

It has been observed that the FTK Imager has successfully acquired a memory dump in a non-VSM-enabled. On the other hand, no memory dump was generated when VSM was active because the system experienced a Blue Screen of Death (BSoD). Meanwhile, Belkasoft RAM Capturer managed to acquire memory dump in both VSM environments. Therefore, this section will analyze the differences in the results of memory acquisition from the FTK Imager application in a non-enabled- VSM environment (non-VSM), and Belkasoft RAM Capturer in both VSM environments.

All three memory dumps generated by the memory acquisition applications have the same size according to the measured memory capacity of 21.4 GB (23,068,672,000 bytes). Here, we will focus on the contents of the memory dump, which has the value "?????????????????...??????" as a mark of memory locations that are not readable by applications (see Fig. 2).

The data shows that the unreadable memory space of active VSM is larger than non-VSM (the sign … refers to the other 65 rows that are not displayed). It indicates that VSM enforces more limitations on physical memory access than non-VSM. The limitation can be correlated with the implementation of memory access protection for each VTL, especially for VTL1 which runs in a safe environment [16].

By calculating the portion of memory with the value "?????????????????...??????" (see Fig. 2), there are 122 MB of memory size for Belkasoft in VSM-enabled mode. It is about 0.6% of the memory size. Meanwhile, Belkasoft and FTK Imager in non-VSM mode are 0.4 MB and 0.5 MB, respectively. It is only about 0.0021% and 0.0025% of the memory size. The comparison of the unreadable memory percentage between VSM-enabled and non-VSM-enabled is significant, as consistent with the importance of memory data for a live forensic investigation.

B. Static Code Analysis

The analysis compares the differences between an active VSM state and when VSM is not active. In this section, the experiment results are grouped based on two types of memory acquisition applications tested, namely FTK Imager and Belkasoft RAM Capturer. The assembly codes are derived from the machine code from the disassembler process using IDA PRO.

A python IDA plugin called Diaphora is used to generate an SQLite file that lists the functions identified from the assembly code generated by IDA PRO. The main purpose of using Diaphora is to examine differences in the functions of the FTK Imager application for active and inactive VSM conditions and the Belkasoft RAM Capturer application.

(a) FTK Imager	(b) Belkasoft RAM Capturer	
NON-VSM	VSM	NON-VSM
0003BC410-0003C5FFF	0002C2000-0002C5FFF	0003BC410-0003C5FFF
140B88220-140B8831F	0003BD000-0003CFFFF	0009D0000-0009D0FFF
1463D8220-1463D8342	1065AC000-1065ACFFF	002CDB000-002CDBFFF
148A02000-148A03FFF	1173D0000-1173D0FFF	002DD6000-002DD6FFF
148AFB410-148AFBFFF	1187D9000-1187D9FFF	0050CA000-0050CAFFF
14A5FC000-14A5FCFFF	1189C6000-1189C7FFF	0051C9000-0051CCFFF
14A5FE000-14A5FFFFF	118BC8000-118BC9FFF	0052CD000-0052CDFFF
14A601000-14A601FFF	118BD1000-118BD1FFF	0056C6000-0056C6FFF
18A319000-18A319FFF	118FCA000-11BFCAFFF	0057C8000-0057C8FFF
...	...	...
540DD5000-540DD9FFF	44D82B000-44D82BFFF	545DD3000-545DD3FFF
540DDB000-540DDC649	44DC28000-44DC28FFF	54CC41EA0-54CC41F9F
541DCB000-541DCBFFF	44DC2C000-44DC2CFFF	
5459C6000-5459CC120	450D4B410-450D4BFFF	
5459CC181-5459CFFFF	451267000-451271649	
5479D3000-5479D3FFF	463EB4000-463EBBFFF	
5479DA000-5479DAFFF	4AF92D2E0-4AF92DFFF	
	4AFF55000-4AFF55FFF	
	4B0C50000-4B0C50FFF	
	4B1C763D0-4B1C76FFF	
	4B5198000-4B5198FFF	
	4B7087610-4B7087FFF	
	4B718C4D0-4B718DFFF	
	4B72993A0-4B7299FFF	
	4BB415000-4BB415FFF	
	4FA01D000-4FA01DFFF	
	50CFA8610-50CFA8FFF	
	537CBC410-537CCDFFF	
	53C4AA320-53C5C4FFF	
	559E16EA0-559E16F9F	

Fig. 2.   Memory Sectors that are Unreadable by FTK Imager and Belkasoft RAM Capturer.

Diaphora succeeded in recognizing 32483 functions from the assembly code of the FTK Imager application. The results were compared in terms of names, order of contents, hash values, and relative virtual address (RVA) values of all these functions. It is identified that all functions of these assembly codes are equal when run in both enabled VSM and non-enabled VSM. The number of files with the status of "100% equal", "Perfect match, same name," "Same order and hash," and "Same RVA and hash" are 21760, 3, 6507, and 4213 files, respectively. Likewise, for the 38 functions that have been recognized from the Belkasoft RAM Capturer application, all of them are also identified to be the same. It has been observed that 15 files with the status of "100% equal", 3 files with the status of "Perfect match, same name," six files with the status of "Same order and hash," and 14 files with the status of "Same RVA and hash".

Based on these findings, it can be deduced that FTK Imager and Belkasoft RAM Capturer applications do not have different functions in their static code for both environments (i.e., VSM is enabled and non-enabled). This further indicates that the VSM environment does not affect the overall running characteristics of the application.

C. Event Log Analysis

Windows operating system generates three event logs which are Application, System, and Security logs. In this study, we focused on observing events from Application and System

logs because they contain key information when the system crashes and restarts.

We identified the records of events associated with FTK Imager application crashes during the memory acquisition process, and when VSM is enabled from the event Application log. Detailed information is presented in Fig. 3. It describes the error name called BlueScreen and informs that this crash event has the data stored in the MEMORY.DMP file.

We observed more information from the System log events. Fig. 4 reports an event with an "error" status. This status is captured from the second experiment scenario; when the VSM is not activated. Detailed information can be found in the General field, stating that the VSM feature is not activated and the Hypervisor as a virtualization emulator fails to run. This information confirms the environment in which we did not activate the VSM. While this setting can be checked from the BIOS configuration, this "error" status notified us that this virtualization-based enablement policy should be mandatory in Windows 10. This situation may lead to anti-forensics, where the implementation of security control prevents digital forensic tools to operate.

The captured information about the error when the FTK Imager is running on the active VSM is presented in Fig. 5.

It is likely indicating the cause of the blue screen and the record of the crash event that forced the system to reboot. The operating system provides the information in their Bugcheck error in Event Viewer. Bugcheck error will record the BSoD event, and its basic error code to identify what caused the BSoD. The fourth row in this Bugcheck provides information that the system is rebooted and IsolatedUserMode is active.

Fig. 3.   Selected Significant Events of FTK Imager in Non-Active VSM.

Fig. 4.   Selected Significant Events of FTK Imager in Non-Active VSM.

Fig. 5.   Selected Significant Events of FTK Imager in Active VSM.

We have the same observation about the active IsolatedUserMode when Belkasoft RAM Capturer runs in active VSM (see Fig. 6). Other key points in Fig. 5 are shown in rows 5 and 6. These two rows indicate a Hypervisor failure to handle CVE-2018-3646. Further examination of Common Vulnerabilities and Exposure (CVE) suggests that the vulnerability is related to the possibility of unauthorized disclosure of information [23]. A possible explanation for this failure could be associated with the existence of a memory space isolation system that caused the memory acquisition tools unable to access the information.

An additional analysis of the MEMORY.DMP file was undertaken to obtain further information on the "Bugcheck" event. We used the Windbg application and ran the command !analyze -v (see Fig. 7). The Bugcheck analysis was carried out on the MEMORY.DMP file supports that the crash is related to the FTK Imager application. The associated module is ad_driver, and the image name is ad_driver.sys. The file directory is located at C:\Users\[UserName]\AppData\Local\Temp. This is consistent with the information on the BSoD screen, which indicates an error has occurred in the driver.sys. Furthermore, Windbg provides more information about this error by indicating that the driver.sys in question is related to ad_driver.

D. Dynamic Code Analysis

Dynamic code analysis examines the application's behavior while the operating system executes it. Interaction from the user will affect the direction of execution. The dynamic code analysis is performed on the FTK Imager application with an active VSM environment. The aim is to observe the application's behavior related to the BSoD error.

Fig. 6.   Selected Significant Events of Belkasoft RAM Capturer in Active VSM.

```
***************************************************************************
*                         Bugcheck Analysis                               *
***************************************************************************
FILE_IN_CAB:  MEMORY.DMP
BUGCHECK_CODE:  3b
BUGCHECK_P1: c0000005
BUGCHECK_P2: fffff803362939d0
BUGCHECK_P3: ffffab8f63b96c10
BUGCHECK_P4: 0
CONTEXT:  ffffab8f63b96c10 -- (.cxr 0xffffab8f63b96c10)
rax=0000000000000000 rbx=0000000000000000 rcx=ffffd302a40fe000
rdx=00002cfd5c603000 rsi=0000000002600000 rdi=0000000000700000
rip=fffff803362939d0 rsp=ffffab8f63b97618 rbp=ffffd302a40fd000
 r8=0000000000001000  r9=0000000000000080 r10=7ffffffffffffffc
r11=ffffd302a40fd000 r12=ffffffffffffffff r13=0000000000001000
r14=ffffd302b2dbc160 r15=ffffd302b4030a10
iopl=0         nv up ei ng nz na pe cy
cs=0010  ss=0018  ds=002b  es=002b  fs=0053  gs=002b             efl=00050283
ad_driver+0x39d0:
fffff803`362939d0 488b440af8      mov     rax,qword ptr [rdx+rcx-8] ds:002b:00000000`00700ff8=????????????????
Resetting default scope

PROCESS_NAME:  FTK Imager.exe
STACK_TEXT:
ffffab8f`63b97618 fffff803`3629256a     : 00000000`c00000bb ffffff803`3a8e681c ffffab8f`63b976a0 01d8134d`cdf32b29 : ad_driver+0x39d0
ffffab8f`63b97620 fffff803`36291110     : 00000000`02600000 fffff803`34ff5f01 00000000`00700000 00000000`000008c4 : ad_driver+0x256a
ffffab8f`63b976b0 fffff803`34c8f825     : ffffd302`b31fe0a0 ffffd302`00000000 00000000`00000000 00000000`00000001 : ad_driver+0x1110
ffffab8f`63b97700 fffff803`35075b58     : ffffab8f`63b97a80 ffffd302`b31fe0a0 00000000`00000001 ffffd302`b40020c0 : nt!IofCallDriver+0x55
ffffab8f`63b97740 fffff803`35075957     : 00000000`00000000 ffffab8f`63b97a80 00000000`00000000 ffffab8f`63b97a80 : nt!IopSynchronousServiceTail+0x1a8
ffffab8f`63b977e0 fffff803`35074cd6     : 00000000`00000000 00000000`00000000 00000000`00000000 00000000`00000000 : nt!IopXxxControlFile+0xc67
ffffab8f`63b97920 fffff803`34e08cb5     : 00000000`00000000 00000000`00000000 00000000`00000000 00000000`00000000 : nt!NtDeviceIoControlFile+0x56
ffffab8f`63b97990 00007ffa`9ef0ce54     : 00000000`00000000 00000000`00000000 00000000`00000000 00000000`00000000 : nt!KiSystemServiceCopyEnd+0x25
00000000`006f6918 00000000`00000000     : 00000000`00000000 00000000`00000000 00000000`00000000 00000000`00000000 : 0x00007ffa`9ef0ce54
SYMBOL_NAME:  ad_driver+39d0
MODULE_NAME: ad_driver
IMAGE_NAME:  ad_driver.sys
STACK_COMMAND:  .cxr 0xffffab8f63b96c10 ; kb
BUCKET_ID_FUNC_OFFSET:  39d0
FAILURE_BUCKET_ID:  AV_ad_driver!unknown_function
OS_VERSION:  10.0.19041.1
BUILDLAB_STR:  vb_release
OSPLATFORM_TYPE:  x64
OSNAME:  Windows 10
FAILURE_ID_HASH:  {8f6b899e-895f-35a5-567c-c877346fcd6e}
Followup:     MachineOwner
---------
```

Fig. 7.   Summary of Bugcheck Analysis Results.

The analysis commenced by selecting the "Start debugging" menu in the Windbg Preview application and selecting the executable file from the FTK Imager application. The debugger downloaded the symbol file "ProfUISad64.pdb" to perform the debugging process. The following commands are typed on the "Command" page to control the process:

- To load symbols:

    - .symfix
    - .reload

- To run the FTK Imager application:

    - g

As a result of executing those commands, we identified that the last module before the system crash was C:\Windows\system32\mssprxy.dll. The module is recorded from the debugger as a module that is loaded before the user clicks the "Capture Memory" button. This is an unexpected finding because the information from the event log analysis suggests the module that caused the crash is ad_driver.sys. Therefore, other scenarios in dynamic code analysis shall be considered to find the very last module loaded by the operating system before the crash happens.

V.   CONCLUSION

This study aims to conduct a technical analysis of the effects of VSM on the memory acquisition process. Two cases were observed that are: (1) a successful acquisition process by using the Belkasoft RAM Capturer, and (2) an unsuccessful acquisition process by using the FTK Imager. The static analysis results of the two applications did not show any differences in the program code when the application machine code disassembler was carried out, both when VSM was enabled and non-enabled. It is concluded that the VSM environment does not affect the program modules of the application.

Meanwhile, Application event analysis comprises logs of system crashes and is stored in the MEMORY.DMP file. Bugcheck analysis of the dump file shows the cause of the system experiencing BSoD when it executes the ad_driver.sys module. Furthermore, results from dynamic analysis explained the behavior of the FTK Imager application just before the BSoD occurs, and it is identified that the application accesses the C:\Windows\system32\mssprxy.dll module.

VI.   FUTURE WORK

This study highlights the impact of VSM on the memory acquisition process that causes the loss of memory artifacts when the process is halted and the system restarts. However, this study is limited to two memory acquisition tools running on the Windows operating system, which respond differently to the activation of the VSM feature. More importantly, the difference opens more directions for future work. Investigating the impact on other tools and operating systems would present more significant results to be compared. Testing environments should involve different scenarios in dynamic code analysis and conduct an in-depth analysis of the ad_driver.sys module content. This is to seek further understanding of how the module causes the system crashes.

REFERENCES

[1] Collective work of all attendees, "Digital Forensic Research Workshop," in Proceedings of The Digital Forensic Research Conference (DFRWS) (2001), Aug. 2001.

[2] A. Case and G. G. Richard, "Memory forensics: The path forward," Digit Investig, vol. 20, pp. 23–33, 2017, doi: https://doi.org/10.1016/j.diin.2016.12.004.

[3] S. R. Davies, R. Macfarlane, and W. J. Buchanan, "Evaluation of live forensic techniques in ransomware attack mitigation," Forensic Science International: Digital Investigation, vol. 33, p. 300979, 2020, doi: https://doi.org/10.1016/j.fsidi.2020.300979.

[4] J. Hale, "Memory Acquisition and Virtual Secure Mode," https://df-stream.com/2017/08/memory-acquisition-and-virtual-secure/, 2017.

[5] H. K. Brendmo, "Live forensics on the Windows 10 secure kernel," Norwegian University of Science and Technology, 2017.

[6] A. Arfeen, M. Asim Khan, O. Zafar, and U. Ahsan, "Process based volatile memory forensics for ransomware detection," Concurr Comput, vol. 34, no. 4, Feb. 2022, doi: 10.1002/cpe.6672.

[7] D. C. Prakoso, I. Riadi, and Y. Prayudi, "Detection of Metasploit Attacks Using RAM Forensic on Proprietary Operating Systems," Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, pp. 155–160, May 2020, doi: 10.22219/kinetik.v5i2.1037.

[8] A. Kazim, F. Almaeeni, S. al Ali, F. Iqbal, and K. Al-Hussaeni, "Memory Forensics: Recovering Chat Messages and Encryption Master Key," in 2019 10th International Conference on Information and Communication Systems (ICICS), Jun. 2019, pp. 58–64. doi: 10.1109/IACS.2019.8809179.

[9] S. Anson, Ed., "Acquiring Memory," in Applied Incident Response, Wiley, 2019, pp. 103–131. doi: 10.1002/9781119560302.ch5.

[10] S. Anson, Ed., "Memory Analysis," in Applied Incident Response, Wiley, 2019, pp. 235–275. doi: 10.1002/9781119560302.ch9.

[11] G. M. Jones and S. G. Winster, "An Insight into Digital Forensics: History, Frameworks, Types and Tools," in Cyber Security and Digital Forensics, Wiley, 2022, pp. 105–125. doi: 10.1002/9781119795667.ch6.

[12] R. J. McDown, C. Varol, L. Carvajal, and L. Chen, "In-Depth Analysis of Computer Memory Acquisition Software for Forensic Purposes," J Forensic Sci, vol. 61, pp. S110–S116, Jan. 2016, doi: 10.1111/1556-4029.12979.

[13] M. N. Faiz and W. A. Prabowo, "Comparison of Acquisition Software for Digital Forensics Purposes," Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, pp. 37–44, Nov. 2018, doi: 10.22219/kinetik.v4i1.687.

[14] I. Riadi, S. Sunardi, and M. E. Rauli, "Live Forensics Analysis of Line App on Proprietary Operating System," Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, pp. 305–314, Oct. 2019, doi: 10.22219/kinetik.v4i4.850.

[15] Microsoft, "Virtual Secure Mode," Microsoft Learn, Jul. 07, 2022. https://learn.microsoft.com/en-us/virtualization/hyper-v-on-windows/tlfs/vsm (accessed Oct. 05, 2022).

[16] A. Milenkoski and D. Phillips, "Virtual Secure Mode: Architecture Overview," 2019.

[17] N. W. Hirst, "The implications of virtual machine introspection for digital forensics on nonquiescent virtual machines," NAVAL POSTGRADUATE SCHOOL MONTEREY CA, 2011.

[18] N. Zhang, R. Zhang, K. Sun, W. Lou, Y. T. Hou, and S. Jajodia, "Memory Forensic Challenges under Misused Architectural Features," IEEE Transactions on Information Forensics and Security, vol. 13, no. 9, pp. 2345–2358, Sep. 2018, doi: 10.1109/TIFS.2018.2819119.

[19] R. ben Yehuda, E. Shlingbaum, Y. Gershfeld, S. Tayouri, and N. J. Zaidenberg, "Hypervisor memory acquisition for ARM," Forensic Science International: Digital Investigation, vol. 37, Jun. 2021, doi: 10.1016/j.fsidi.2020.301106.

[20] A. Milenkoski, "Virtual Secure Mode: Communication Interfaces," 2019.

[21] A. M. Marshall and R. Paige, "Requirements in digital forensics method definition: Observations from a UK study," Digit Investig, vol. 27, pp. 23–29, 2018, doi: https://doi.org/10.1016/j.diin.2018.09.004.

[22] R. Nordvik, H. Georges, F. Toolan, and S. Axelsson, "Reverse engineering of ReFS," Digit Investig, vol. 30, pp. 127–147, 2019, doi: https://doi.org/10.1016/j.diin.2019.07.004.

[23] National Vulnerability Database, "CVE-2018-3646," The MITRE Corporation, Dec. 28, 2017. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2018-3646 (accessed Oct. 05, 2022).

Optimizing Faculty Workloads and Room Utilization using Heuristically Enhanced WOA

Lea D. Austero[1], Ruji P. Medina[4]
Graduate Programs
Technological Institute of the
Philippines
Quezon City, Philippines

Ariel M. Sison[2]
School of Engineering and
Technology
Emilio Aguinaldo College
Manila, Philippines

Junrie B. Matias[3]
College of Computing and
Information Sciences
Caraga State University
Butuan City, Philippines

Abstract—The creation and generation of schedules that are free of conflicts manually every academic semester present higher education institutions with a duty that is laborious and demanding of their resources. The course timetabling optimization, as an education timetabling problem, is a popular example of an NP-hard combinatorial problem. Numerous attempts have been made over the course of the past few decades to find a solution to this problem, but no one has yet developed a foolproof approach that can examine all alternatives to find the best method. The promising swarm-based optimization algorithm called Whale Optimization Algorithm was heuristically enhanced in the present study and is called HEWOA. It was designed as a solution to the course timetabling problem discussed in the current study. HEWOA was able to generate an efficient timetable for the large dataset of 1700 events for an average time of 14.92 seconds only, with an average generation of 7.2 and a best time of 8.38 seconds. These results reveal that the performance of HEWOA was better than that of various hybrids of the Genetic Algorithm that was compared in the present study.

Keywords—*Heuristics; mutation; optimization; swarm; timetabling; whale optimization algorithm*

I. INTRODUCTION

The application of automated procedures to a time-consuming and resource-intensive task often leads to increased efficiency and productivity, as well as time and cost savings. Among these processes are the preparation and creation of academic schedules. Timetabling problem was solved manually through trial and error, but this was not the greatest option. At present, scientific methods are used to address the problem [1], [2]. Timetabling problems, better known as the university course timetabling problem (UCTTP), are known to be NP-hard, meaning the problem cannot be solved exactly in polynomial time as its size and complexity increase exponentially [3]. It involves allocating non-overlapping classes to given resources such as classrooms and teachers in space-time [4]. The number of courses, the average number of lectures per day, the desired free timeslots each day, and the targeted off-days in a week are a few of the constraints that influence the design of the educational timetable [1], [5]. In the scheduling problem, there are two types of constraints: hard constraints and soft constraints. Hard constraints are rules or restrictions that cannot be broken. Soft constraints are requirements that, if not violated, can improve the effectiveness of the timetable. A timetable is considered efficient if it is able to solve the problem while adhering to all of the hard constraints specified [6].

The process of scheduling classes is often carried out with the assistance of specialized models that are adapted to meet the requirements of the particular educational establishment in question. A significant amount of work devoted to scheduling makes use of streamlined models to investigate and evaluate the performance of various scheduling strategies. The vast majority of research on course scheduling focuses on modeling and computational results, with very little attention paid to actual implementation in the real world [7].

Several strategies have been used to solve course timetabling using benchmark and real-world datasets. This problem was solved over decades using optimization approaches. Heuristic approaches helped resolve timetabling's complex behavior and model [8]. Evolutionary methods are frequently used in solving course timetabling; however, these existing methods were not able to quickly tests all alternatives to find the best solution [8]–[10]. Recently, various research has employed the Whale Optimization Algorithm (WOA), which is appreciated as a simple, flexible, and competitive swarm-based metaheuristic algorithm [11][12][13]. Despite its potential, it has inherent flaws that must be addressed before it can effectively address optimization issues such as course timetabling. WOA, like most metaheuristic algorithms, struggles to have a balanced local and global search.

The present study is an application and enhancement of the algorithm used in the previous work of WOA [14] in solving course timetabling. In this work, WOA was integrated with heuristic mutation to improve further the performance of WOA in solving optimization problems such as timetabling. The aim of the present work is to introduce HEWOA as a heuristically enhanced variant of the WOA, which is used in solving course timetabling problems. A literature review is presented in the next section of this paper, which contains a discussion on timetabling, solutions for solving UCTTP, and WOA. The third section presents the particulars of the methodology, which includes the problem definition, the architecture, and the HEWOA. Section IV presented the observations, results, and discussions on the experiments conducted in solving the timetabling problem. Finally, the conclusion of this study is presented in Section V.

II. LITERATURE REVIEW

A. Timetabling

The process of allocating resources to discrete objects in space-time to achieve desired goals within a given set of constraints is called timetabling. Timetabling encompasses many research-intensive fields. In education timetabling, the timetabling for the course and examination is the most studied. It is challenging to execute a course scheduling solution with the same approach to a problem because each institution has unique characteristics and constraints or limitations [15]. Universities and colleges in the Philippines use a manual procedure to schedule classes, increasing work for program heads and making it difficult to analyze every timetable combination [16]. In each country, accrediting government and private agencies require state universities and colleges to adhere to specific policies and criteria for scheduling classes. One of the recommendations suggests limiting the number of preparations for each faculty member to no more than four distinct subjects so that they have just enough work to do during the semester. The quality of instruction may degrade if faculty members teach more subject courses than the university deems optimal [17], which also leads to student's poor academic performance[18]. This case is often violated in actual practice. There are also cases of an unbalanced allocation of workloads among faculty members, in which some have more than four preparations while others have fewer than four.

More research is encouraged in solving UCTTP as it is unique across institutions due to policies and regulations. A general solution that could solve all the concerns in UCTTP does not exist [9]. Apart from being effective, optimization algorithms should consider simplicity and adaptability to a range of varying real-world UCTTP [10]. Thus, adapting the implementation of state-of-the-art methods on real-world UCTTP is still open to be explored by researchers.

B. Solutions in Solving UCTTP

Over the past decade, several works have shown substantial advancements in timetabling techniques and algorithms. These methods were created to address either benchmark datasets or real-world datasets [9]. In terms of the quantity and hard and soft constraints, benchmark and real-world UCTTPs differ. Creation solutions for benchmark datasets are often generalized and intended for comparing algorithms. The benchmark datasets utilized in international timetabling competitions, such as Socha[19], ITC-02[20], and ITC-07 [21], are the most popular testbeds among researchers in comparing algorithms. Real-world UCTTP, on the other hand, emphasizes the applicability of solutions in academic institutions. Due to varied legislation, educational systems, and cultures, even real-world UCTTP vary in terms of their criteria [10].

Metaheuristic methods promise precise and optimal timetable scheduling solutions and are popular for timetabling and other optimization challenges. They are simple to implement, faster than the standard mathematical-based optimization process, and achieve optimal results [11]. Evolutionary methods like Genetic algorithms, ant colony, local search, simulated annealing, and tabu search are frequently used in course scheduling; however, none of these was considered the best [8], [9]. Hybrid techniques or combining two or more algorithms are also prevalent and have produced more high-quality outcomes than other techniques, as proven in prior studies [22]–[24]. Hybrid methods are appropriate for maximizing the benefits of separate techniques. Single solution-based meta-heuristics and population-based meta-heuristics are the most popular approaches for the benchmark UCTTP, while in the case of real-world or actual datasets such as the one used in this work, the most popular methods used include single solution-based meta-heuristic, Operations Research, population-based meta-heuristic, hyper-heuristic and hybrid approaches [9].

C. Whale Optimization Algorithm

The Whale Optimization Algorithm (WOA) is a swarm-based optimization algorithm that is inspired by the hunting behavior of humpback whales[13]. WOA is among the most promising and competitive optimization techniques [11], [25]. The whales, while encircling the prey, create specific bubbles along a circular path. The bubble-net attacking technique assists in exploitation. The prey for the search state of WOA represents the exploration phase. For exploitation, the whale position is updated using either spiral movement or shrinking encirclement. For exploration, the humpback whale finds the best solution and updates its position according to other whales. Having this inspiration, WOA is composed of three operators: encircling prey, bubble-net attacking method, and search prey.

The following are the relevant equations implemented by WOA on its operations. Eq. 1-4 captures the procedures for the encircling prey.

$$\vec{D} = |\vec{C}.\vec{X}^*(t) - \vec{X}(t)| \tag{1}$$

$$\vec{X}(t + 1) = \vec{X}^*(t) - \vec{A}.\vec{D} \tag{2}$$

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \tag{3}$$

$$\vec{C} = 2.\vec{r} \tag{4}$$

where t indicates the current iteration, $\vec{A}$ and $\vec{C}$ are coefficient vectors, and $X*$ is the position vector of the best solution obtained so far. $\vec{X}$ is the position vector, $|\ |$ is the absolute value, and is an element-by-element multiplication, and $\vec{r}$ is a random vector in [0, 1]. It should be noted that $X*$ should be updated in each iteration if there is a better solution. The vectors $\vec{A}$ and $\vec{C}$ are calculated as shown in Eq. 3 and 4, respectively.

Bubble-net attacking method:

$$\vec{X}(t + 1) = \vec{D}'.e^{bl}.\cos(2\pi l) + \vec{X}^*(t) \tag{5}$$

$$f(x) = \begin{cases} \vec{X}^*(t) - \vec{A}.\vec{D}, & \text{if } p < 0.5 \\ \vec{D}'.e^{bl}.\cos(2\pi l) + \vec{X}^*(t), & \text{if } p \geq 0.5 \end{cases} \tag{6}$$

Where p represents a constant for explaining the shape of the logarithmic spiral and l is a random number uniformly distributed in the range of [-1, 1].

and Search for prey:

$$\vec{D} = |\vec{C}.\vec{X}_{rand} - \vec{X}| \tag{7}$$

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A}.\vec{D} \qquad (8)$$

where $\vec{X}_{rand}$ is nominated arbitrarily from whales in the current iteration.

Many recent research has employed the Whale Optimization Algorithm (WOA), which is appreciated as a simple, flexible, and competitive swarm-based metaheuristic algorithm [11][12][13]. Whale bubble-net hunting inspired WOA's algorithm. Its effectiveness and adaptability attract researchers from many fields. It is used in electrical, computer, aeronautical, and construction engineering [26]. Despite WOA's promising features, it has some unavoidable flaws, including being designed for continuous search space [13], requiring too many parameters tuning [27], having no theoretical convergent property [28], and having a probability distribution that changes with iterations [12]. It may also prematurely converge, trapping it in local optima [29], [30]. WOA, like most metaheuristic algorithms, struggles to balance local and global searches. The present study is an application and enhancement of the algorithm used in the previous work of WOA [14] in solving course timetabling. In this work, WOA was integrated with heuristic mutation.

III. METHODOLOGY

A. Problem Definition

In the allocation of schedules to resources, the course timetabling problem must fulfill both hard and soft restrictions. This section presents the problem description and objective functions to be used in the implementation of the timetable. This study's constraints and objective functions are similar to the previous work [14] but were implemented with a different algorithm to enhance WOA.

- Timetable Guidelines

The number of courses, the average number of lectures per day, the desired free timeslots each day, and the desired off-days in a week are a few of the variables that influence the design of an educational timetable. Each course in a curriculum is a class which could be a lecture, a lab, or both. Each of these classes is allocated a teacher and a classroom at a time that should not conflict with other classes scheduled for that day. Classrooms are utilized either for lecture classes or laboratory work. The mathematical formulation must fulfill all the relevant variables in order to produce a timetable that is both efficient and feasible. The following variables have been taken into consideration throughout this study. Let:

- E be the set of scheduled classes for a teacher and students with specified courses and classrooms,

- S be the set of students grouped through blocks in a program,

- T be the set of teachers wherein each teacher can handle many courses with a maximum of 4 unique subjects to handle,

- C is the set of courses wherein each lecture unit in a course is equivalent to one hour, whereas each laboratory unit is equivalent to three hours of class;

- R be the set of rooms, either lecture or laboratory, that will be assigned to classes, and;

- K is the set of period slots in a day from 7:30 in the morning to 7:30 in the evening, wherein the days of the week are paired as Monday-Thursday, Tuesday-Friday, and Wednesday-Saturday.

- Constraints

The constraints in course timetabling problems are classified as soft and hard. Soft constraints are optional, while hard constraints must be satisfied completely [9], [13]. For this work, the minimum requirements set as policies by the Accrediting Agency of Chartered Colleges and Universities in the Philippines and the Commission on Higher Education for state universities such as Bicol University are also taken into account in these constraints. Table I presents the constraints considered in this study which were identified as the most common constraints being used in solving course timetabling [9].

TABLE I. THE CONSTRAINTS

Code	Type	Description
H1	Hard	No teacher may be assigned to the same group of students in two separate classes.
H2	Hard	A teacher should handle only one course in one classroom at each time slot.
H3	Hard	Exactly only one class is assigned per timeslot per day in a classroom.
H4	Hard	The size of the classroom should be considered.
H5	Hard	For all required courses for a group of students must be given a scheduled
H6	Hard	All the teaching periods required in the curriculum must be given a schedule
H7	Hard	An uninterrupted period of time required for a class should be assigned precisely on a given day.
S1	Soft	At least one timeslot in a day should be vacant for a group of students
S2	Soft	The maximum number of straight classroom teaching hours is three in a schedule
S3	Soft	The total number of teaching hours in a day should not exceed 6 hours.

- Objective Functions

The degree to which the timetable can satisfy the constraints effectively will determine how much each solution will cost. In each generation of the candidate solution, a time slot is allotted to a pair of student and faculty groups associated with a classroom and a course. This is done with the goal of satisfying all hard requirements while keeping the expense of satisfying soft constraints to a minimum. The problem may also be expressed using the formulation that follows. Let X be the set of all possible solutions, $HC = \{h_1 \dots h_6\}$ the set of hard constraints, $SC = \{s_1 \dots s_5\}$ the set of soft constraints and $x \subseteq X$ the set of all candidate solutions.

The goal of this procedure is to find the most efficient timetable with the least cost, as presented in Eq. 1 by the penalty charged per violation of a hard constraint:

$$hp(x) = \sum_{t \in T} f_1(x,t) + \sum_{t \in T} f_2(x,t) + \sum_{r \in R} f_3(x,r) + \sum_{r \in R} f_4(x,r) + \sum_{s \in S} f_5(x,s) + \sum_{s \in S} f_6(x,s) + \sum_{s \in S} f_7(x,s) \quad (9)$$

TABLE II. THE OBJECTIVE FUNCTIONS FOR THE HARD CONSTRAINTS

Function	Penalty instance	Purpose
$f_1(x,t)$	Two or more classes are assigned to one teacher to handle	Ensures no conflict in faculty loading
$f_2(x,t)$	More than one teacher is assigned to a class	Ensures that only one teacher is assigned to a class
$f_3(x,r)$	One classroom is assigned with multiple classes.	Ensures no conflict in the classroom assignment
$f_4(x,r)$	If the classroom size is not considered.	Ensures that the classroom can accommodate the class size
$f_5(x,s)$	a course for a block of students has no schedule assigned	Ensures that all courses enrolled by the students are assigned a schedule
$f_6(x,s)$	The number of required hours on the curriculum is not satisfied.	Ensures that the required hours offered in the curriculum are equal to the scheduled classes
$f_7(x,s)$	More than a total of six hours of teaching load for a teacher	Ensures that there is a balance distribution of workload within a week

The penalty function for hard constraints is represented by Equation 9, and the value of the objective function for each solution can be determined as follows when soft constraints are also considered:

$$f(x) = \left(\sum_{s \in S} f_8(x,s) + \sum_{t \in T} f_9(x,t) + \sum_{t \in T} f_{10}(x,t) \right) + (hp(x) \times W) \quad (10)$$

The function $f(x)$ shown in Eq. 10 represents the sum of all penalties from hard and soft constraints. The function examines each solution for possible violations on the constraints wherein a value of one (1) will be given had there been violations; otherwise, zero (0). It can be observed that there is a variable W that is applied to the hard constraints $hp(x)$, where a value of three (3) will be multiplied by the counted violations on hard constraints. This was done so that the objective function would give more weight to violations on hard constraints.

TABLE III. THE OBJECTIVE FUNCTIONS FOR THE SOFT CONSTRAINTS

Function	Penalty instance	Purpose
$f_8(x,s)$	There is no vacant period for a day to a group of students	Ensures that students have at least one vacant period daily
$f_9(x,s)$	The total teaching hours for a course totaled to more than 3 hours	Ensures that the schedules are not straight and that having such would be tiresome both for the students and teachers.
$f_{10}(x,s)$	the total teaching hours to be handled by a teacher is more than 6 hours	Ensures that the schedules are distributed throughout the week.

B. Architecture

The courses and classes to be included in the scheduling were collected from various colleges at Bicol University. The lecture and laboratory rooms available for utilization and faculty members teaching these courses were also identified.

These raw data were used in the generation of the timetable. Various sizes of datasets were considered for the experimentations to test the algorithms: 400, 800, 1200, and 1700 as the largest dataset.

Shown in Fig. 1 is the framework of the method for generating an optimized and fairly distributed timetable. The classes to be scheduled, faculty schedule, room availability, and specified constraints serve as the inputs. These data will be processed using HEWOA as objective functions. A penalty will be applied to violations of constraints, and the process will be repeated through generations until the cost approaches zero to produce an efficient and optimized timetable.
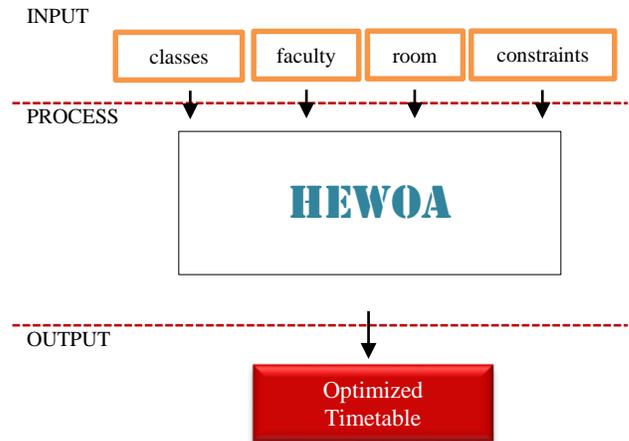
Fig. 1. Framework of the Study.

C. The Heuristically Enhanced Whale Optimization Algorithm

In this work, preliminary results implemented in solving course timetabling using WOA indicate that updating the entire solution using these traditional procedures of WOA has a greater likelihood of destroying rather than improving it. This is because the whale operator is designed to update all values using the same parameter values, and the event number is large. Additionally, the more constraints applied to a problem, the more complicated the search process becomes. A solution emerged during the experimentation: when the first procedure (encircling prey) was not used, the performance of WOA was improved. This is due to the nature of the equation, which has a high probability of destroying the solution. Moreover, the heuristic mutation operator is integrated into the process to enhance further WOA's exploration and exploitation capability.

Fig. 2 now depicts the pseudocode of HEWOA, taking into account the aforementioned modifications to the method as well as the section where the heuristic mutation seen in Fig. 3 will be implemented. It is shown that the encircling procedure was removed from the process.

Instead of updating the position of the current search whale using the encircling prey process, remove it, retaining the Bubble-net attacking method for the exploitation phase and searching for prey for the exploration phase.

```
Initialize the whales population Xᵢ(i = 1, 2,...,n)
Calculate the fitness of each search agent
X*=the best search agent
  while (t < maximum number of iterations)
    for each search agent
    Update a, A, C, l, and p
      if1 (p<0.5)
        if2(|A| < 1)
          • Select a random search agent (Xrand)
          • Update the position of the current search
            agent by the Eq. 8
      else if1 (p ≥ 0.5)
        Update the position of the current search by
        the Eq. 5
      end if1
      Update the position of the current search
      agent using Heuristic Mutation (Fig. 4)
    end for
    Check if any search agent goes beyond the search
    space and amend it Calculate the fitness of each
    search agent Update X* if there is a better
    solution
    t=t+1
  end while
return X*
```

Fig. 2.   The HEWOA Pseudocode.

Accordingly, in WOAs encircling prey section, search agents update to the best agent. It is easy to trap the algorithm in a local solution, causing premature convergence [31]. Consequently, updating the position of the current search agent using heuristic mutation is integrated to further improve the algorithm's exploitation and exploration capability. In this work, the heuristic mutation focuses on invalid classes and repairs them using random pairs of room and period.

- The Encoding Method

The genetic operator avoids illegal offspring by encoding all events in each candidate solution in the same index, as shown in Fig. 3.
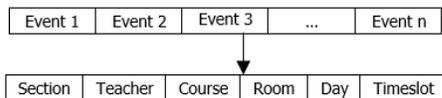
Fig. 3.   The Representation of the Solution.

The subset of solutions in the current generation is a 3-dimension array containing scheduled events. Each of these events contains the codes of other constraints such as student section, teacher, course, room, day, and timeslot. This data structure is similar to the one used in the work of [32]. In addition, prioritizing constraints are applied, and the events of the teacher who has the highest workload and the events of the section of students who have the most classes that need to be scheduled are attached to the solution first.

- Fitness Function

The whale's fitness is the weighted sum of penalty cost based on Equations 9 and 10. Every constraint violation incurs a penalty of 1. However, the cost for hard constraints is multiplied by the weight value W such that the algorithm prioritizes these constraints while finding and ranking all candidate solutions. In addition, the impact of soft constraints will highlight the superior solution since it meets preferences. In other words, a solution with the same cost in hard constraints will be differentiated by the number of violations on soft constraints.

- Heuristic Mutation

Mutation modifies genes to create new individuals. The heuristic mutation is designed to produce better offspring wherein a set of chromosomes is transformed from a parent by exchanging some genes (neighborhood) [33]. In Fig. 4, the operator gets all invalid gene indexes and then mutates them.

The operator does not implement mutation probability parameters in mutating invalid genes since it generates ten percent random genes or classes. The operator selects one of these randomly generated genes that fulfill stringent criteria. If no valid gene is found, the room or periods are altered at random (day and timeslots). It guarantees that randomly selected times are distinct from nearby occurrences in order to ensure diversity. Additionally, the rate of ten percent random genes or classes can be decreased or increased during the configuration or before running the algorithm.
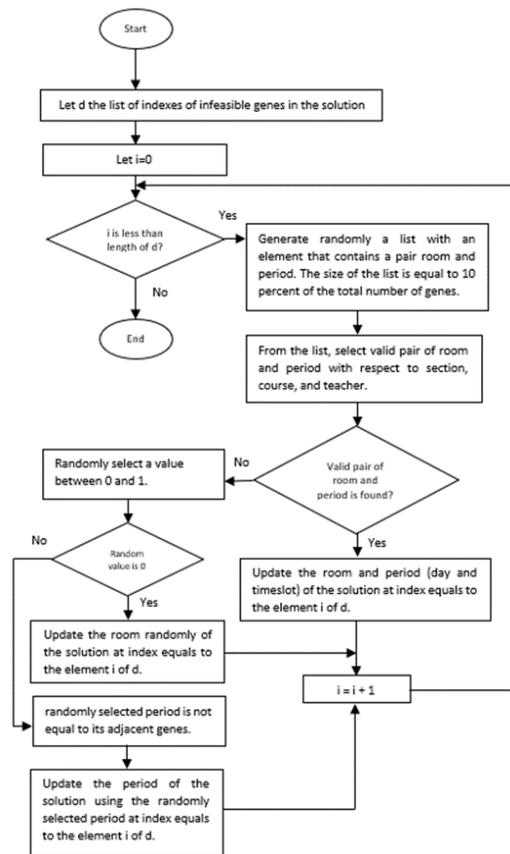
Fig. 4.   Heuristic Mutation Applied to WOA.

IV.   RESULTS AND DISCUSSION

Using real-world data, the efficiency of HEWOA was evaluated. These are actual datasets that include the courses to be scheduled for the various programs offered by Bicol University, Legazpi City, Albay, Philippines, as well as the rooms being utilized for the conduct of classes and the assigned faculty that will handle the class. The HEWOA uses a population of 10 whales and will stop generating solutions if all hard constraints are satisfied by any whale or when it reaches an iteration of 1000.

TABLE IV.   PERFORMANCE OF THE HEWOA COMPARED TO OTHER
METHODS

Methods	Events	Average Generation	Best Time	Average Time	SD
GA Using Heuristic Mutation	400	3.4	0.590s	1.061s	0.353
	800	5.2	2.370s	**3.455s**	0.650
	1200	8.0	6.430s	8.554s	1.402
	1700	15.5	16.670s	21.670s	4.247
GA Using Invalid Genes Focused Random Resetting Mutation	400	4.1	1.280s	1.595s	0.198
	800	7.0	5.145s	7.608s	1.475
	1200	14.0	16.670s	23.015s	3.979
	1700	33.9	55.700s	81.355s	16.402
HEWOA	400	**2.4**	**0.458s**	**0.74s**	0.160
	800	**4.1**	**2.009s**	3.514s	0.812
	1200	**5.8**	**5.299s**	**8.362s**	1.964
	1700	**7.2**	**8.384s**	**14.942s**	4.671
EWOA [14]	117	-	-	3s	-
	195	-	-	15s	-
	304	-	-	270s	-
Guided GA [34]	878	136.5	-	-	-
	1140	409.5	-	-	-
Parallel GA and Local Search [35]	166	900	-	-	-
Greedy and Genetic Fusion Algorithm [36]	300	900	-	-	-

Table IV shows the various sizes of events or the number of classes used to test the various techniques. As part of the execution of the generating schedules in the timetable, the objective functions (Eq. 9) are executed, which assess penalties (Tables II and III) for schedules that violate the constraints stated in Table I. These costs determine the fitness function. Initial generations or runs would incur corresponding penalties, which would gradually decrease until they approached zero, at which point there would be no violations.

The performance of the HEWOA was compared to other competitive methods that solved course timetabling: GA using heuristic mutation; GA using invalid genes focused on random resetting mutation; guided GA [34], parallel GA and local search [35]; and greedy and genetic fusion algorithm [36]. The result in Table IV of their performances in terms of total execution time is based on ten (10) runs per method. In terms of average generation, HEWOA was able to perform better compared to other work which uses Hybrid GA as the base method. In terms of execution time, HEWOA was also able to perform better except on a dataset with 800 classes.

Fig. 5, 6, and 7 illustrate the pace at which the indicated techniques are approaching closer to zero, which is defined as the state in which no more penalties are incurred on the constraints. The research indicates that HEWOA is the approach with the fastest pace among these solutions, followed by GA with heuristic mutation. The least performing method is the GA when random resetting mutation is applied to infeasible genes.

An example of manually generated faculty workloads that shows unoptimized class schedules and loading is shown in Fig. 8. It can be observed in the workload that on Tuesday and Friday, the faculty had teaching hours of more than 6 hours. Thus, violating soft constraint 3, the total number of teaching hours per day should not exceed 6 hours. The same scenario can be repeated with other faculty since that method of plotting schedules is manual, which is prone to errors, especially when the schedulers are adjusting or transferring loads from one faculty to another.

Table V shows an example of faculty workloads generated by HEWOA. It can be seen that the workloads are equally distributed in the week, although the schedules do not show the consultation hours. It is also noted that that days are paired, like Monday-Thursday and Tuesday-Friday. In each timeslot, the total hour is one and a half hours, and the other one and a half hours are to be lectured on a corresponding pair day.
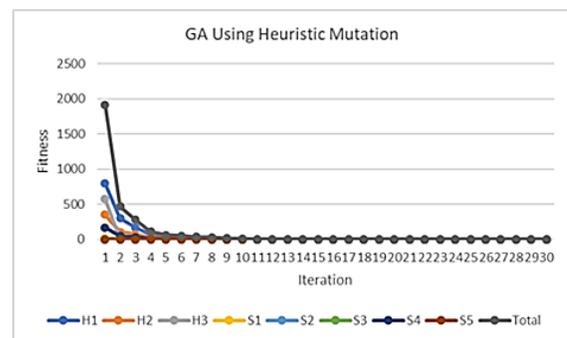
Fig. 5.   Performance of Genetic Algorithm when a Heuristic Mutation is Applied.
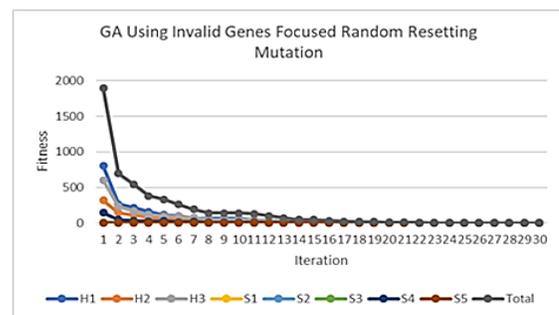
Fig. 6.   Performance of Genetic Algorithm when Random Resetting Mutation is Applied on Infeasible Genes.
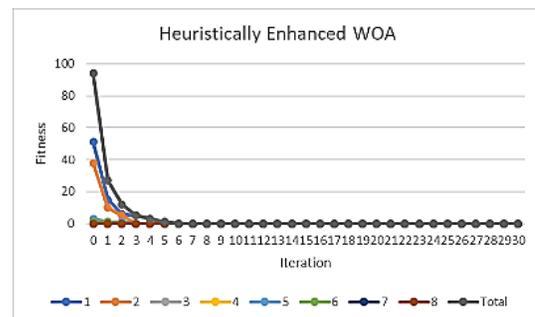
Fig. 7.   Performance of HEWOA when the Heuristic Mutation is Integrated.

Time	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
07:00AM-07:30AM						
07:30AM-08:00AM	ITE 10-AC1 NSB-CCIS MSIT LAB LAB			ITE 10-AC1 NSB-CCIS MSIT LAB LAB		
08:00AM-08:30AM		ITE 10-IJ2 NSB-CCIS ST 4 LEC			ITE 10-IJ2 NSB-CCIS ST 4 LEC	IT 115-Y1Y21 NSB-CCIS NET LAB LEC
08:30AM-09:00AM						
09:00AM-09:30AM	OFFICIAL WORKING HOURS	ITE 10-IJ2 HIRAYA-CCIS CL2 LAB		OFFICIAL WORKING HOURS	ITE 10-IJ2 HIRAYA-CCIS CL2 LAB	
09:30AM-10:00AM						
10:00AM-10:30AM			CONSULTATION- -			IT 115-Y1Y21 NSB-CCIS NET LAB LAB
10:30AM-11:00AM	ITE 10-AC1 NSB-CCIS ST 4 LEC	ITE 10-IK2 HIRAYA-CCIS CL2 LAB		ITE 10-AC1 NSB-CCIS ST 4 LEC	ITE 10-IK2 HIRAYA-CCIS CL2 LAB	
11:00AM-11:30AM						
11:30AM-12:00PM						
12:00PM-12:30PM	ITE 10-BC1 HIRAYA-CCIS CL4 LAB			ITE 10-BC1 HIRAYA-CCIS CL4 LAB		
12:30PM-01:00PM						
01:00PM-01:30PM		OFFICIAL WORKING HOURS			OFFICIAL WORKING HOURS	
01:30PM-02:00PM	CONSULTATION- -	ITE 10-MO1 HIRAYA-CCIS CL2 LAB		CONSULTATION- -	ITE 10-MO1 HIRAYA-CCIS CL2 LAB	
02:00PM-02:30PM						
02:30PM-03:00PM						
03:00PM-03:30PM	OFFICIAL WORKING HOURS	ITE 10-NO1 HIRAYA-CCIS CL2 LAB		OFFICIAL WORKING HOURS	ITE 10-NO1 HIRAYA-CCIS CL2 LAB	
03:30PM-04:00PM						
04:00PM-04:30PM						
04:30PM-05:00PM		ITE 10-MO1 NSB-CCIS ST 3 LEC			ITE 10-MO1 NSB-CCIS ST 3 LEC	
05:00PM-05:30PM						
05:30PM-06:00PM						
06:00PM-06:30PM						
06:30PM-07:00PM						
07:00PM-07:30PM						
07:30PM-08:00PM						
08:00PM-08:30PM						
08:30PM-09:00PM						

Fig. 8. Example of Faculty Workloads Generated Manually Captured from the University Database.

TABLE V. EXAMPLE OF FACULTY WORKLOADS GENERATED BY HEWOA CAPTURED FROM THE UNIVERSITY DATABASE

Faculty	Section	Course Code	Room	Day	Day Description	Timeslot	Timeslot Description
114	BSIS 1-105	ITE 10	18.03	1	Mon-Thu	3	10:30-12:00
114	BSIT 2-113	IS 105	18.04	1	Mon-Thu	4	12:00-13:30
114	BSIS 1-105	ITE 10	7.02	1	Mon-Thu	8	18:00-19:30
114	BSIS 1-107	ITE 10	18.01	2	T-TF	3	10:30-12:00
114	BSIT 2-113	IS 105	7.02	2	T-TF	4	12:00-13:30
114	BSIS 3-111	IS 116	7.01	2	T-TF	6	15:00-16:30

On room utilization, plotting classes manually can result in an inefficient room allocation. These would result in more classroom usage; it could increase the cost of maintenance and energy. In this work, the utilization of classrooms and laboratories is optimized since the HEWOA can produce timetables with fewer rooms compared to classrooms. For example, when we retrieved the schedules in one semester with 1700 classes, the classrooms and laboratories utilized were more than 120 compared to schedules generated by HEWOA and GA, which only used 91.

A. Implications

Population-based metaheuristics such as Genetic Algorithms, Particle Swarm Optimization, and Ant Colony Optimization are superior to other methods in solution space exploration. Still, these approaches require a higher processing time necessary to generate solutions of high quality [37]. A good quality solution means no violations of the specified constraints whose purpose is explained and presented in Tables II and III. It was observed in the presented results in Table IV and Fig. 5, 6, and 7 that HEWOA was able to generate good-quality solutions that require lesser computational time. HEWOA is a multi-objective implementation of WOA and can be used for other similar solutions.

The typical technique of timetabling is inefficient, manual, and not very robust against changes [10]. As a result, the number of course conflicts is considerable, which reduces the effectiveness of the instruction [38], [39]. The utilization of an intelligent approach for UCTTP has been the subject of much research and widespread debate worldwide. Using the method in the present work would aid in developing efficient software for course scheduling in academic institutions.

V. CONCLUSION

This paper introduces HEWOA, a heuristically enhanced Whale Optimization Algorithm designed to solve UCTTP. The results of the experiments on various sizes of real-world data indicate that both GA and HEWOA could generate a feasible and efficient timetable that could satisfy all the identified constraints set by educational institutions. Generating class schedules using GA and HEWOA can optimize classroom and laboratory utilization which could help decrease the cost of maintenance and energy. Optimized classroom and laboratory utilization could also help increase the number of students since there would be more available resources for classes. Faculty workloads also can be improved using automated scheduling; thus, it helps satisfy soft constraints and enhance the quality of schedules.

Moreover, it is observed that HEWOA outperformed Hybrid GAs [34] and other methods [35], [36] in terms of execution time and average solution generation for the majority of utilized event sizes.

Lastly, future work would include testing the performance of this method using benchmark UCTTP, which provides different constraints than the real-world data used in this study.

REFERENCES

[1] R. Ganguli and S. Roy, "A Study on Course Timetable Scheduling using Graph Coloring Approach," 2017. [Online]. Available: http://www.ripublication.com

[2] R. R. Iwańkowicz and M. Taraska, "Self-classification of assembly database using evolutionary method," Assembly Automation, vol. 38, no. 3, pp. 268–281, May 2018, doi: 10.1108/AA-06-2017-071.

[3] M. Nouiri, A. Bekrar, A. Jemai, S. Niar, and A. C. Ammari, "An effective and distributed particle swarm optimization algorithm for flexible job-shop scheduling problem," J Intell Manuf, vol. 29, no. 3, pp. 603–615, Mar. 2018, doi: 10.1007/s10845-015-1039-3.

[4] Ş. Gür and T. Eren, "Scheduling and planning in service systems with goal programming: Literature review," Mathematics, vol. 6, no. 11, Nov. 2018, doi: 10.3390/math6110265.

[5] M. Andrey, V. Voronkin, P. Svetlana, and S. Alexey, "Software and hardware infrastructure for timetables scheduling in university," Oct. 2018, pp. 15–20. [Online]. Available: http://ceur-ws.org

[6] J. S. Tan, S. L. Goh, G. Kendall, and N. R. Sabar, "A survey of the state-of-the-art of optimisation methodologies in school timetabling problems," Expert Syst Appl, vol. 165, Mar. 2021, doi: 10.1016/j.eswa.2020.113943.

[7] M. Mühlenthaler, "Real-world academic course timetabling," Lecture Notes in Economics and Mathematical Systems, vol. 678, pp. 107–128, 2015, doi: 10.1007/978-3-319-12799-6_4.

[8] H. Alghamdi, T. Alsubait, H. Alhakami, and A. Baz, "A Review of Optimization Algorithms for University Timetable Scheduling," 2020. [Online]. Available: www.etasr.com

[9] M. C. Chen, S. N. Sze, S. L. Goh, N. R. Sabar, and G. Kendall, "A Survey of University Course Timetabling Problem: Perspectives, Trends and Opportunities," IEEE Access, vol. 9, pp. 106515–106529, 2021, doi: 10.1109/ACCESS.2021.3100613.

[10] R. A. Oude Vrielink, E. A. Jansen, E. W. Hans, and J. van Hillegersberg, "Practices in timetabling in higher education institutions: a systematic review," Ann Oper Res, vol. 275, no. 1, pp. 145–160, Apr. 2019, doi: 10.1007/s10479-017-2688-8.

[11] U. Can and B. Alatas, "Performance comparisons of current metaheuristic algorithms on unconstrained optimization problems," Periodicals of Engineering and Natural Sciences, vol. 5, no. 3, pp. 328–340, 2017, doi: 10.21533/pen.v5i3.120.

[12] Y. Ling, Y. Zhou, and Q. Luo, "Lévy Flight Trajectory-Based Whale Optimization Algorithm for Global Optimization," IEEE Access, vol. 5, pp. 6168–6186, 2017, doi: 10.1109/ACCESS.2017.2695498.

[13] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," Advances in Engineering Software, vol. 95, pp. 51–67, May 2016, doi: 10.1016/j.advengsoft.2016.01.008.

[14] L. D. Austero, A. M. Sison, J. B. Matias, and R. P. Medina, "Solving course timetabling problem using Whale Optimization Algorithm," in 2022 8th International Conference on Information Technology Trends (ITT), May 2022, pp. 160–166. doi: 10.1109/ITT56123.2022.9863951.

[15] J. B. Matias, A. C. Fajardo, and R. P. Medina, "A fair course timetabling using genetic algorithm with guided search technique," in 2018 5th

[16] International Conference on Business and Industrial Research (ICBIR), 2018, pp. 77–82.

[16] H. Bellardo, "Preference-driven university course scheduling system," 2010.

[17] The New Times, "Heavy workload affects quality of teaching," https://www.newtimes.co.rw/author/36/times-reporter, Sep. 19, 2017.

[18] I. Gwambombo, "The Effect of teacher's workload on students' academic performance in community secondary schools," 2013.

[19] R. Lewis and B. Paechter, "Finding feasible timetables using group-based operators," IEEE Transactions on Evolutionary Computation, vol. 11, no. 3, pp. 397–413, Jun. 2007, doi: 10.1109/TEVC.2006.885162.

[20] Y. Bykov, "The description of the algorithm for international timetabling competition," 2014. [Online]. Available: https://www.researchgate.net/publication/228388131

[21] L. di Gaspero, B. Mccollum, and A. Schaerf, "The Second International Timetabling Competition (ITC-2007): Curriculum-based Course Timetabling (Track 3) Multiobjective Timetabling with Fairness View project PHD works View project The Second International Timetabling Competition (ITC-2007): Curriculum-based Course Timetabling (Track 3)-preliminary presentation," 2007. [Online]. Available: http://www.idsia.ch/

[22] S. L. Goh, G. Kendall, and N. R. Sabar, "Improved local search approaches to solve the post enrolment course timetabling problem," Eur J Oper Res, vol. 261, no. 1, pp. 17–29, Aug. 2017, doi: 10.1016/j.ejor.2017.01.040.

[23] P. Daru Kusuma and A. Sayid Albana, "University Course Timetabling Model in Joint Courses Program to Minimize the Number of Unserved Requests," IJACSA) International Journal of Advanced Computer Science and Applications, vol. 12, no. 10, pp. 121–127, 2021, [Online]. Available: www.ijacsa.thesai.org

[24] S. Al-Negheimish, F. Alnuhait, H. Albrahim, S. Al-Mogherah, M. Alrajhi, and M. Hosny, "An Intelligent Bio-Inspired Algorithm for the Faculty Scheduling Problem," IJACSA) International Journal of Advanced Computer Science and Applications, vol. 9, no. 5, 2018, [Online]. Available: www.ijacsa.thesai.org

[25] H. M. Mohammed, S. U. Umar, and T. A. Rashid, "A Systematic and Meta-Analysis Survey of Whale Optimization Algorithm," Computational Intelligence and Neuroscience, vol. 2019. Hindawi Limited, 2019. doi: 10.1155/2019/8718571.

[26] N. Rana, M. S. A. Latiff, S. M. Abdulhamid, and H. Chiroma, "Whale optimization algorithm: a systematic review of contemporary applications, modifications and developments," Neural Computing and Applications, vol. 32, no. 20. Springer Science and Business Media Deutschland GmbH, pp. 16245–16277, Oct. 01, 2020. doi: 10.1007/s00521-020-04849-z.

[27] R. Sivalingam, S. Chinnamuthu, and S. S. Dash, "A modified whale optimization algorithm-based adaptive fuzzy logic PID controller for load frequency control of autonomous power generation systems," Automatika, vol. 58, no. 4, pp. 410–421, 2017, doi: 10.1080/00051144.2018.1465688.

[28] R. Salgotra, U. Singh, and S. Saha, "On Some Improved Versions of Whale Optimization Algorithm," Arab J Sci Eng, vol. 44, no. 11, pp. 9653–9691, Nov. 2019, doi: 10.1007/s13369-019-04016-0.

[29] H. S. Alamri, Y. A. Alsariera, and K. Z. Zamli, "Opposition-based Whale Optimization Algorithm," J Comput Theor Nanosci, vol. 24, no. 10, pp. 7461–7464, 2018, doi: 10.1166/asl.2018.12959.

[30] N. Singh and H. Hachimi, "A New Hybrid Whale Optimizer Algorithm with Mean Strategy of Grey Wolf Optimizer for Global Optimization," Mathematical and Computational Applications, vol. 23, no. 1, p. 14, Mar. 2018, doi: 10.3390/mca23010014.

[31] K. Lu and Z. Ma, "A modified whale optimization algorithm for parameter estimation of software reliability growth models," Journal of Algorithms & Computational Technology, vol. 15, p. 17483026211034442, 2021.

[32] J. B. Matias, A. C. Fajardo, and R. P. Medina, "Examining genetic algorithm with guided search and self-adaptive neighborhood strategies for curriculum-based course timetable problem," in 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), 2018, pp. 1–6.

[33] W. Ho and P. Ji, "A hybrid genetic algorithm for component sequencing and feeder arrangement," J Intell Manuf, vol. 15, no. 3, pp. 307–315, Jun. 2004, doi: 10.1023/B:JIMS.0000026569.88191.46.

[34] M. Fachrie and A. F. Waluyo, "Guided Genetic Algorithm to Solve University Course Timetabling with Dynamic Time Slot," in 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020, Dec. 2020, pp. 583–587. doi: 10.1109/ISRITI51436.2020.9315448.

[35] D. Kristiadi and R. Hartanto, "Genetic Algorithm for lecturing schedule optimization," IJCCS (Indonesian Journal of Computing and Cybernetics Systems), vol. 13, no. 1, p. 83, Jan. 2019, doi: 10.22146/ijccs.43038.

[36] K. Wang, W. Shang, M. Liu, and W. Lin, "A Greedy and Genetic Fusion Algorithm for Solving Course Timetabling Problem," in 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), 2018, pp. 344–349.

[37] C. K. Teoh, A. Wibowo, and M. S. Ngadiman, "Review of state of the art for metaheuristic techniques in Academic Scheduling Problems," Artif Intell Rev, vol. 44, no. 1, pp. 1–21, Jun. 2015, doi: 10.1007/s10462-013-9399-6.

[38] Ç. H. Erdoğan and R. Topuz, "Investigation of Stress Perceptions of Physical Education Teachers," Asian Journal of Education and Training, vol. 6, no. 2, pp. 144–148, 2020, doi: 10.20448/journal.522.2020.62.144.148.

[39] J. A. Bowden and P. J. Green, "Completion mindsets and contexts in doctoral education: Pursuing efficiency and quality with integrity," 2019, pp. 77–99. doi: 10.1007/978-981-13-6990-2_5.

Transformation Model of Smallholder Oil Palm Supply Chain Ecosystem using Blockchain-Smart Contract

Irawan Afrianto[1], Taufik Djatna[2], Yandra Arkeman[3], Irman Hermadi[4]
Computer Science Department, IPB University, Bogor, Indonesia[1, 4]
Agro-industrial Technology Department, IPB University, Bogor, Indonesia[2, 3]

Abstract—The development of new technology has the potential to disrupt and transform an existing system as well as information technology. This study aims to build a proposed model in order to transform an old system into a Blockchain-based system. The smallholder oil palm supply chain currently uses a traditional information system and technology, hence its integrity, transparency, and security are vulnerable. To solve this problem, frontier technology is needed, such as a blockchain, which has trust, transparency, and traceable characteristics to improve performance and quality. The method used in this study was digital transformation in the context of operational processes and technology. In addition, the As-Is To-Be model was used as a mechanism to develop a transformation model for the smallholder oil palm supply chain system. Specifically, the As-Is model was implemented in identifying and analyzing the information system and technology in an already existing system, while the To-Be was used to determine the blockchain potential and characteristics. This becomes the proposed model for the transformation of the old system into a Blockchain technology-based. Also, a prospective and mechanism for system transformation were produced in the aspect of transactions, data, and architecture, as well as the flow of change strategies needed in the transformation of the blockchain-based smallholder oil palm supply chain system.

Keywords—*Transformation model; smallholder oil palm supply chain; blockchain; smart contract*

I. INTRODUCTION

The oil palm industry is one of the leading sectors in Indonesia as indicated by its plantation area and productivity. Also, the oil palm supply chain, specifically the FFB management, has been extensively studied from different perspectives.

It has been observed that the smallholder oil palm agroindustry supply chain has a variety of chain models, with many actors and regulations [1]. The independent actors in the oil palm agroindustry supply chain, such as traders are more concerned with the profit element. Therefore, formal supply chain institutions, such as cooperatives need to improve their roles and functions in order to play an active role, which brings benefits and improves the welfare of an independent oil palm smallholder [2][3].

Currently, an information system is used for processing transactions in the smallholder oil palm supply chain environment. This developed system is still centralized, thereby allowing data vulnerability occurrence. Also, the role of users who still interact a lot with the system leads to input errors and other results that are not desired. This simply means data security is not guaranteed, and it is vulnerable to being exploited by other parties. Another problem with a centralized system is that it has a mechanism controlled only by one entity. This implies that the entity can fully control the activities and data contained in the system. Consequently, data are easily manipulated by system owners or hackers, thereby reducing the trust and data transparency in the operation of a product, as well as tracing an asset's provenance [4]. It has been discovered that the digital contract in the existing system was only an agreement between the oil palm mill and the cooperative but the contractual clauses contained therein are often not observed. This leads to manipulation and non-transparency in the running system. In addition, the centralized system poses a high operational challenge in securing the servers and data from hackers. It involves the cost of maintaining data authenticity due to a large amount of manual work and paperwork involved in the business processes [5].

These problems encourage the development of blockchain-based smart contract technology, as an alternative solution to eliminate and reduce challenges in data transactions and a centralized system. The smart contract's implementation reduces human intervention, supports transactions, and increases trust between entities [6]. Meanwhile, blockchain technology adoption targets several domains, such as supply chain, healthcare system, electronic voting, proof of location, distributed cloud storage, and even human resources management and recruitment [7]. According to [8]–[10] blockchain is managed by a distributed peer-to-peer system in a decentralized infrastructure secured by a cryptographic hashing mechanism and with the support of a smart contract and consensus system, which is effective in collaboration between organizations and individuals.

The transformation of a system becomes important when a new technology emerges and has the potential to improve the existing system's performance [11]. This study aims to develop a technological transformation model needed by an existing system to be blockchain technology-based. Subsequently, the following questions were answered

1) What model and technology is used in the current supply chain system for smallholder oil palm?

2) What is the desired model when blockchain is used for the proposed system?

3) Which system and technology transformation mechanisms are required?

Since the existing smallholder oil palm supply chain system is digital-based, the approach selected is a partial blockchain technology adoption mechanism model [12]. In this scenario, the transformation is performed by considering the presence of an information technology/information system existing previously and focusing on the transformation of technology and system architecture. This current study provides an overview of the system transformation mechanism from "traditional" digital solutions (As-Is) to model a system by applying "frontier" digital solutions (To-Be) technology - blockchain-based system.

This article is structured as follows, Section I discusses the problem background in the current system, potential solutions with blockchain technology, as well as the objectives and contributions. Section II entails the literature reviews, namely smallholder oil palm supply chain, blockchain technology, smart contract, and types of blockchain platforms. Furthermore, Section III describes the proposed methodology used in conducting the study phases. Section IV contains the results and discussion, starting with the existing system analysis, the proposed transformation model, and how it can be achieved, as well as discussions related to similar studies and future developments. Finally, the conclusions of the study results are presented in Section V.

II. LITERATURE REVIEW

A. Smallholder Oil Palm Supply Chain

The smallholder oil palm supply chain in Indonesia varies in different regions. However, one similarity between the many flows is that they always involve intermediaries, called middlemen. An example of an existing supply chain is shown in Fig. 1, in which the FFB is harvested from farmers' plantations and is then collected by cooperatives/farmer groups. Furthermore, the FFB is picked up by small agents, who distribute them to large agents, then to Delivery Order (DO) holders, before arriving at PKS. This supply chain is both long and also detrimental to farmers because every party takes profit along the process. This reduces the impact on prices and the profits that independent smallholders are able to earn [13].
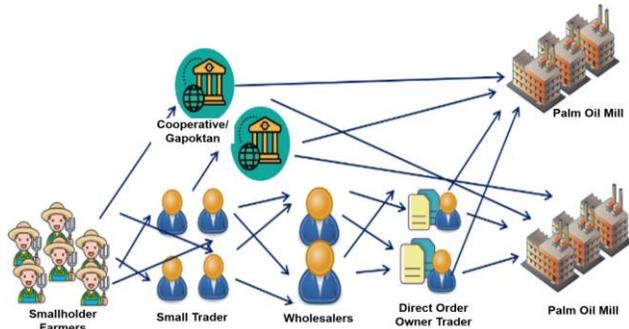
Smallholder oil palm farmers do not have a great ability to obtain capital and assistance due to a lack of knowledge as well as access to information and capital that can ensure their sustainability [14]. To increase their prosperity and capacity, standardization and a good plantation management program need to be implemented [15]. Also, institutions capable of uniting smallholder oil palm farmers need to be developed. This can be encouraged through the existence of the farmers' groups (Gapoktan), as well as cooperatives that become facilitators in regulating and supporting smallholders [16]. As reported in [17][18], the ease of obtaining certification for smallholder oil palm farmers needs to be supported by the existing supply chain ecosystem in order to provide quality assurance and products produced.

B. Blockchain Technology

Blockchain is a ledger system, such as a master ledger, in which every transaction that has ever existed is recorded in the form of a decentralized database network. All existing transaction data are recorded in a block entity with each connected to a pre-existing block like a chain [19].

The use of blockchain technology is capable of transforming the conventional supply chain ecosystem with its benefits. In a Blockchain-based supply chain, there are several participants in the network working and interacting together. Each participant sends transactions specifically according to their task in the network [20][21]. In China, a traceability model for agricultural products was developed with blockchain technology and IoT using RFID [22]. As reported, this blockchain technology is capable of developing supply chain business models by reengineering the traditional supply chain to become blockchain-based [23]. Meanwhile, the implementation of blockchain-based logistics operations produces a system that provides a faster and more accurate traceability model [24].

Blockchain has an indelible data structure formed by a series of data blocks that are connected linearly in a time sequence [25]. Information is stored in each block and encrypted with asymmetric cryptographic algorithms to ensure the security of data access and transmission. In the system, each data block records and updates the node data as well as the transaction information according to the defined consensus algorithm of the distributed nodes. The validity of the data is verifiable for all blocks using a hash algorithm [26]. Fig. 2 shows a simple blockchain architecture consisting of a set of blocks containing transactions with a header, transaction, and hash component [27].

Fig. 1. Smallholder Oil Palm Supply Chain Flow.
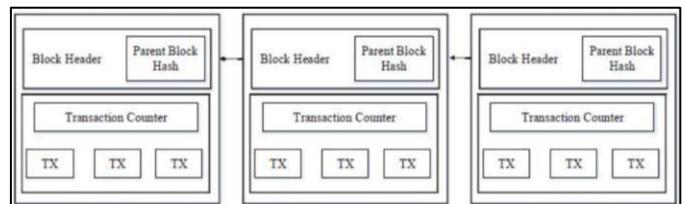
Fig. 2. Blockchain Architecture [27].

C. Smart Contract

It is believed that blockchain combined with smart contract technology has great potential to facilitate a business process [28]. A smart contract is considered a computerized transaction protocol that executes a series of contract clauses introduced in the nineties, while blockchain was the first technology to officially support its implementation [29]. In the literature, a smart contract and blockchain are inseparable terms that together form the second generation of blockchain technology [30][31]. The smart contract execution on the blockchain can be implemented in both public and private blockchain environments with conditioned functionality [32][33].

As described in [34], a smart contract is built with the specific purpose of executing a complete set of instructions on the blockchain. Furthermore, it is a computer program containing contractual agreements between entities generated by the user and extracted from the environment (blockchain) [35]. The purpose of a smart contract is to make the agreement efficient, secure, and increase trust between entities [36].

A smart contract is a working mechanism that involves digital assets and two or more parties. Some or all parties input assets, which are automatically redistributed among users according to a formula and on certain unknown data at the time of contract is initiated [37]. Based on the mechanism, a smart contract has five developmental stages, namely 1) negotiation, 2) development, 3) deployment, 4) maintenance, and 5) learning and self-destruction [38]. Also, a smart contract is capable of reducing human intervention in business process flows in the system environment [39], and can audit automatically in order to quickly and efficiently complete the work expected of users [40].

D. Types of Blockchain Platforms

In the development of a system, two types of blockchain platforms can be used. The first is a public blockchain, which emphasizes the importance of anonymity and decentralization, such as the Ethereum blockchain. The second is a private / consortium blockchain which has the characteristics of being more focused on efficiency and centralized to some extent, like the Hyperledger Fabric blockchain [41].

It is important to note that Ethereum was launched in 2015 with a programmable Blockchain. Due to this capability, Blockchain can be used to build decentralized applications according to user needs. As stated in [42], the characteristics of the Ethereum Blockchain are stateful and Turing Completeness, hence it is possible to store data other than transactions and create complex programs. In addition, the blockchain on Ethereum is a public ledger used to record transactions, which means all data from the initial launch of Ethereum are recorded in 23 Blockchains [43]. Since the ledger is publicly accessible, Ethereum connects each ledger block using a hash algorithm. Based on this mechanism, changing one transaction data requires replacing the hash of the next blocks. Ethereum is generally viewed as a transaction-based state machine [44][45] and a stateful Blockchain, as assets are stored in blocks, not just transactions. It was observed that there are three Merkle Patricia Trees, being a modification of the Merkle Tree on

Bitcoin for transactions, accounts, and receipts, thereby leading to three roots stored in the block header. Another study found that there is more data on an Ethereum block than on Bitcoin, which supports the process of creating decentralized applications using the Ethereum Blockchain [46].

Meanwhile, Hyperledger Fabric is an open-source technology of a distributed ledger platform, which is blockchain-enabled [47]. Furthermore, it is a technology developed by the Linux Foundation which is modular for industrial use. Hyperledger Fabric has a smart contract called chain code [48], written with general-purpose programming languages such as Java, Go, and Node.js. It also uses a different approach to a smart contract in general, namely execute-order-validate [49]. Transactions are first executed on a smart contract, then it is ordered via a certain consensus protocol and finally validated with an endorsement policy before being written to the ledger [50]. Hyperledger Fabric has the following main components, such as assets, shared ledger, chain code, peer nodes, channels, organizations, membership service provider (MSP), and ordering services [51][52].

III. PROPOSED METHODOLOGY

This study aims to develop an existing system and transform it into a blockchain technology-based model. The approach used was the As-Is To-Be Model [53], in which the current system is the reference (As-*Is model*) for the development of a new blockchain-based model (*To-Be model*). This is achieved by developing a transformation model that bridges changes from the existing system to the new form. A limitation of digital transformation within the framework of technological transformation was provided, which includes network and communication, hardware and software, and data technologies [54]. Furthermore, the transformation vision approach is at the operational process level and relates to the study conducted by MIT and Capgemini Consulting [55] used to support technological change in the smallholder oil palm supply chain system. Fig. 3 shows the stages of the study performed.

In the first stage, analysis and evaluation of the existing system were conducted. It includes the activities for analyzing the transaction process contained in the existing system. It is important to mention that transaction includes activities or business processes performed by each actor in the current system. The existing system architecture as well as the data contained in the running system was then analyzed based on the model and database used. This architecture is a program structure or computer system consisting of software components, its externally visible characteristics, and its relationship.

In the second stage, a literature review discussing blockchain technology in relation to the transaction mechanism occurring in the blockchain environment, data processed and stored, as well as the architecture supporting the development of a Blockchain-based system was conducted. The final stage involves the development of a transformation model from the old system to a Blockchain technology-based model. This transformation model includes the transaction, the data, and the architecture.

Fig. 3.   Study Flow.

Fig. 4.   Scope of Study Results.

Fig. 5.   Existing System Smallholder Oil Palm Supply Chain.

IV. RESULT AND DISCUSSION

This section entails the results of the study conducted, which include the discussion of the analysis and current system evaluation, the proposed transformation model, and its change strategy. Fig. 4 shows the discussion flow of the results, with the current system used as an As-Is model, which provides an overview of the operating conditions of transactions, as well as data and system architecture. Meanwhile, the To-Be model was proposed for a new system transformation using Blockchain technology. To produce the To-Be model, a transformation strategy is needed, which is described using a flowchart. Specifically, the transformation employed includes changes to the transaction mechanism, data, and technology architecture as part of the operational processes in the smallholder oil palm supply chain system to achieve the objectives.

A. Existing System of Smallholders Oil Palm Supply Chain

The current oil palm supply chain uses a compact mechanism, specifically for the smallholder, which includes only three actors, namely small farmers, cooperatives/farmer groups, and finally the oil palm mill. This aims to reduce the length of the existing supply chain, as well as to utilize institutions, such as cooperation and better welfare for smallholder oil palm farmers.

Fig. 5 shows the current smallholder supply chain work system, in which the flow of information in the system occurs in all supply chain actors, while the FFB products flow from upstream to downstream. In this supply chain, information system and technology was used to support operational transactions conducted by the actors, from procurement requests to FFB product fulfilment.

The current information system starts with a digital contract mechanism. Conceptually, a digital contract is a form of an agreement made electronically through the interaction between the offered party and the electronic system. In the current system, two digital contracts were used to initiate operations as well as the demand and supply transactions for fresh fruit bunches (FFB) from an oil palm mill to cooperatives, and also from cooperatives to smallholders. Fig. 6 shows the interaction of the digital contract in the current smallholder oil palm supply chain.
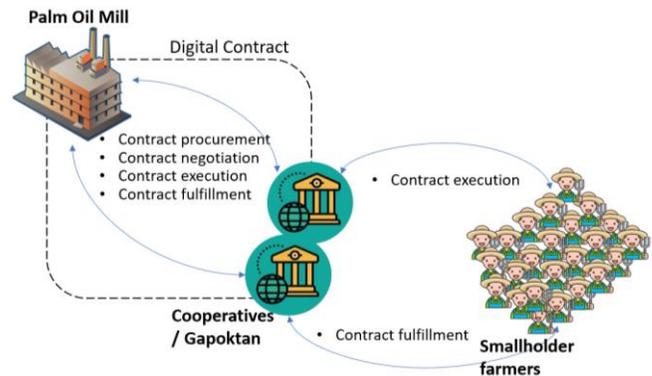
Fig. 6.   Interaction of Digital Contract of Existing System Smallholder Oil Palm Supply Chain.

Meanwhile, Fig. 7 shows the current FFB business process. The activities occurring indicated that the contract initiation process was performed by an oil palm mill, offering the purchase of FFB to cooperatives. Furthermore, the cooperative party that gets a contract offer from the oil palm mill evaluates the approval. When there is a mutual agreement from the negotiations between the oil palm mill and the cooperatives, then a digital contract is created and stored in the system. The next stage is to provide a Direct Order (DO)

as a work letter and implement the digital contract. The DO was given by oil palm mill to cooperatives, and cooperatives pass the DO on to smallholder oil palm farmers who are members. It is important to note that the smallholder oil palm farmers act as suppliers who are obliged to fulfill the DO by entering FFB production data simultaneously. Meanwhile, the cooperatives serve as 1) collectors, 2) distributors, and 3) sellers. This implies that the task of cooperatives is at the center of the current information system because the transactions they conducted are the most compared to other actors in the supply chain system. As a collector, the Cooperative monitors the FFB sent by independent smallholders and collects it into a large batch of FFB ready to be sent to the oil palm mill. Also, as a distributor, the Cooperative has a role to provide transportation tools to pick up and carry the FFB products, but as a seller, they are to ensure that the price obtained still meets the standard price and the mutually agreed contract. The Oil palm mill as a consumer is expected to check the suitability of the FFB products provided by the Cooperative and pay in accordance with the agreement.

Fig. 7. Process Business of the Current Smallholder Oil Palm Supply Chain System.

Table I shows operational transactions performed by farmers which includes farmers' biodata, data on how their plantations and oil palm plantations are managed, and other supporting data. The next transaction is regarding the sale and purchase of FFB, including its fulfilment transaction based on the DO provided by the Cooperative, such as the sales process to the Cooperative as well as information on the farmers' performance level from profiles and the performed transaction records.

Meanwhile, Table II shows the transactions made by Cooperatives, being the intermediary that connects the oil palm farmers with coconut mills. The transaction starts with 1) the recording of the cooperative profile and supporting data in the system. 2) Contract creation and Direct Order (DO) for FFB request transactions. 3) Purchasing the FFB from farmers and selling it to an oil palm mill. The next operation is as a

distributor that transports the FFB from a smallholder to oil palm mill. The process of receiving payments both from an oil palm mill and a farmer is performed by the cooperative. Based on the Cooperative's transactions and profile results, their work performance is explained as follows.

TABLE I. SMALLHOLDER FARMER TRANSACTIONS

No	Actors	Identity Process	Transactions
1	Farmer	Profile	Member Data
2	Farmer	Profile	Farm
3	Farmer	Profile	Land Process
4	Farmer	Profile	Supporting data
5	Farmer	Operational	Transaction fulfillment
6	Farmer	Operational	Sales
7	Farmer	Operational	History DO
8	Farmer	Operational	Performance

TABLE II. COOPERATIVE TRANSACTIONS

No	Actors	Identity Process	Transactions
1	Cooperative	Profile	Member Data
2	Cooperative	Profile	Supporting data
3	Cooperative	Operational	Transactions History
4	Cooperative	Operational	procurement
5	Cooperative	Operational	fulfillment
6	Cooperative	Operational	Delivery
7	Cooperative	Operational	Transactions (procurement and fulfillment)
8	Cooperative	Operational	Payment
9	Cooperative	Operational	Contract History
10	Cooperative	Operational	History DO
11	Cooperative	Operational	Performance

The next operational transaction takes place at the oil palm mill as shown in Table III. It starts with the initiation of contracts between the oil palm mill and cooperatives, then continues with the making of DO and registration of FFB suppliers. The sale and purchase transaction was performed with the cooperative, a partner based on the agreed contract and DO. The FFB payment process was conducted by the oil palm mill at a mutually agreed time.

TABLE III. OIL PALM MILL TRANSACTIONS

No	Actors	Identity Process	Transactions
1	Oil Palm Mill	Profile	Member Data
2	Oil Palm Mill	Profile	Supporting data
3	Oil Palm Mill	Operational	procurement
4	Oil Palm Mill	Operational	Purchase
5	Oil Palm Mill	Operational	Payment
6	Oil Palm Mill	Operational	Contract History
7	Oil Palm Mill	Operational	DO History
8	Oil Palm Mill	Operational	Supplier
9	Oil Palm Mill	Operational	Performance

The government, such as the provincial or district plantation agency acts as an external actor that regulates, configures the system, processes the FFB reference price data, and monitors smallholder oil palm supply chain transactions in their area as shown in Table IV.

TABLE IV. AGENCY TRANSACTIONS

No	Actors	Identity Process	Transactions
1	Agency	Operational	FFB-Province Reference Price Data
2	Agency	Operational	System Configuration

According to Fig. 8, a database was developed to store profile data, manage FFB transaction data, and calculate the performance of each actor in the smallholder oil palm supply chain. The model in the current system includes six interconnected table entities for storing the data used in the digital contract creation and the transaction executed by oil palm mill, cooperatives, and smallholders. Subsequently, the table entails oil palm mill, cooperatives, contracts, contract details, smallholders, and supporting data in the form of plantation profiles.
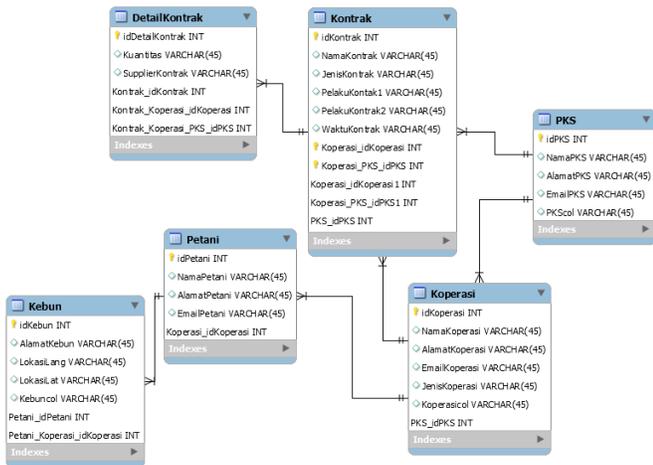
Fig. 8. Existing Database Scheme.

The web-based technology was employed in the current system architecture for it to be accessible by all supply chain actors. The system was developed using platform-based web technology with a centralized database model on a server. In accessing the system, supply chain actors were required to register in order to obtain access rights to perform transactions according to their privileges. Fig. 9 shows the current system architecture.
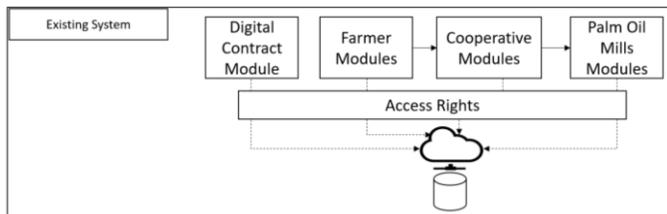
Fig. 9. Digital Contract Table Relation Schema.

B. Proposed Blockchain-Based Smallholders Oil Palm Supply Chain System Transformation Model

A blockchain technology-based system was the system model developed in the smallholder oil palm supply chain. Fig. 10 shows the logical changes made to the proposed system by adding smart contract technology and a blockchain network. A smart contract is generally used to automate manual work performed by users, as well as increase the accuracy and compliance with data entered in the system. Meanwhile, the use of the blockchain network serves as a data guarantor in a more secure, difficult-to-manipulate, and immutable environment.
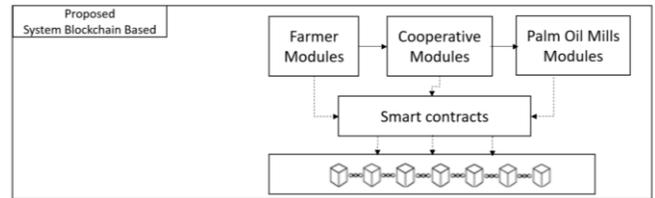
Fig. 10. Logic Model of Proposed System Smallholder Oil Palm Supply Chain.

Fig. 11 shows the differences between traditional application development and those that are blockchain-based. This difference affected the application development on the user side, the database model, and the network architecture used. Consequently, the changes caused a system transformation from a traditional model to one that is blockchain-based. The first transformation was the transformation of the designed application to manage transactions (transaction transformation) by changing the traditional concept into Decentralized applications (DApps) used to transact in the blockchain system. The second was the transformation of a traditional database system into a blockchain-based data model, both on-chain and off-chain data (data transformation). Meanwhile, the third was the blockchain network transformation used to store and disseminate transaction records to all nodes in a blockchain-based system (network/architecture transformation).
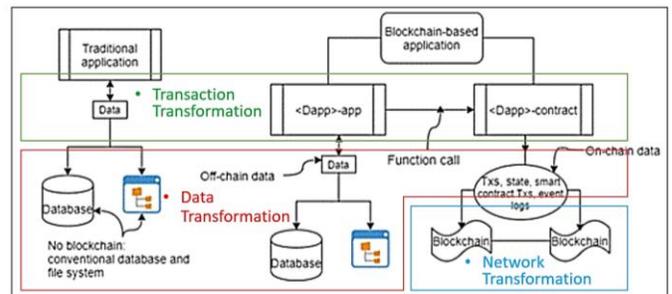
Fig. 11. Proposed Transformation Model of Smallholder Oil Palm Supply Chain Blockchain-based.

1) Transaction transformation model: It is important to reiterate that the first transformation conducted was to change the current system transactions into those executable in the blockchain environment. This mechanism is shown in Fig. 12 and it begins with identifying the transaction process for each supply chain actor, followed by determining whether the

transaction needs to be stored on the blockchain environment or not. When the transaction requires a secure, transparent data environment, and traceability requirements, it was added/modified by adding a smart contract in the system development environment.
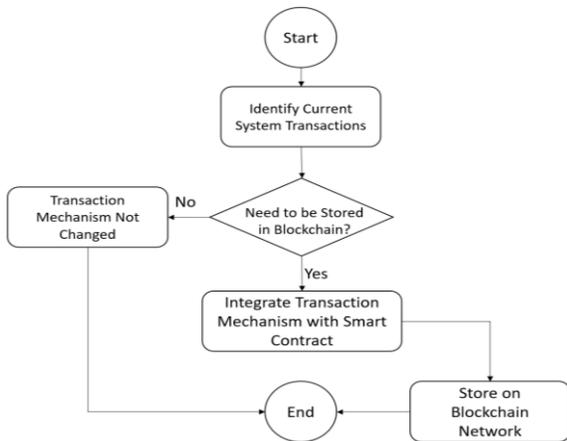
Fig. 12. Blockchain-based Smallholder Oil Palm Supply Chain Transaction Transformation Strategy.

The identification and analysis of transactions existing in the current system consist of actors in the supply chain, namely farmers, cooperatives, oil palm mill, and related agencies, such as admins. There are two process identities, namely the identity for storing profile data and that of operational data. Meanwhile, the business processes are operational data that can be executed by any user based on their access rights. Table V shows the results of the identification and analysis of actors, transactions, and recommendations for the proposed smart contract.

The proposed smart contract architecture model stores data related to the profile of each smallholder oil palm supply chain actor, such as farmer profiles, cooperatives, and oil palm mills. The profile data of supply chain actors are in the form of identity data of actors and other supporting data according to their roles. Meanwhile, smart contract transactions are designed to store transactional data occurring in smallholder oil palm supply chain activities, ranging from the main contract, DO manufacturing, FFB sourcing transactions, FFB fulfilment transactions, performance reviews, and payment mechanisms. The proposed smart contract architecture is shown in Fig. 13.

Fig. 14 shows a sequence diagram for the transaction of profiles of actors in the FFB supply chain, namely farmers, cooperatives, and oil palm mills. This helps in describing a scenario or a series of steps taken in response to an event to produce a certain output, including the changes that occurred.

It is important to reiterate that a smart contract is an executable code running on the blockchain to facilitate, execute, and enforce agreement terms. A smart contract's principal goal was to automatically execute the terms of the agreement once the specified conditions are met [56]. An illustration of the smart contract code is shown in Fig. 15, which described the transactions and actions performed.

TABLE V. PROPOSED SMART CONTRACT ON BLOCKCHAIN-BASED SMALLHOLDER OIL PALM SUPPLY CHAIN SYSTEM

No	Actors	Identity Process	Business Process	Proposed Smart Contract
1	Farmer	Profile	Member Data	Profile Smart Contract
2	Farmer	Profile	Farm	
3	Farmer	Profile	Land Process	
4	Farmer	Profile	Supporting data	
5	Farmer	Operational	Transaction (fulfillment)	Transaction Smart Contract
6	Farmer	Operational	Sales system	
7	Farmer	Operational	History DO	
8	Farmer	Operational	Performance	
9	Cooperative	Profile	Member Data	Profile Smart Contract
10	Cooperative	Profile	Supporting data	
11	Cooperative	Operational	Transaction History	Transaction Smart Contract
12	Cooperative	Operational	Transactions (procurement and fulfillment)	
13	Cooperative	Operational	Payment system	
14	Cooperative	Operational	Contract History	
	Cooperative	Operational	History DO	
15	Cooperative	Operational	Performance	
16	Oil Palm Mill	Profile	Member Data	Profile Smart Contract
17	Oil Palm Mill	Profile	Supporting data	
18	Oil Palm Mill	Operational	Transaction (procurement)	Transaction Smart Contract
19	Oil Palm Mill	Operational	Purchase	
20	Oil Palm Mill	Operational	Payment system	
21	Oil Palm Mill	Operational	Contract History	
	Oil Palm Mill	Operational	History DO	
22	Oil Palm Mill	Operational	Supplier	
23	Oil Palm Mill	Operational	Performance	
24	Admin (Agencies)	Operational	FFB/Province Reference Price Data	Transaction Smart Contract
25	Admin (Agencies)	Operational	System Configuration	
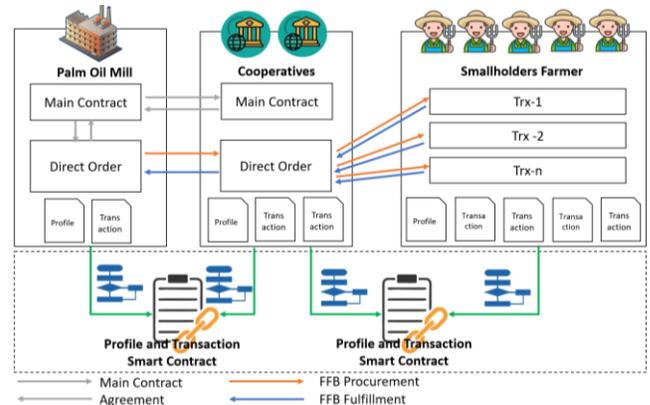
Fig. 13. Proposed Smart Contract Architecture Smallholder Oil Palm Supply Chain.
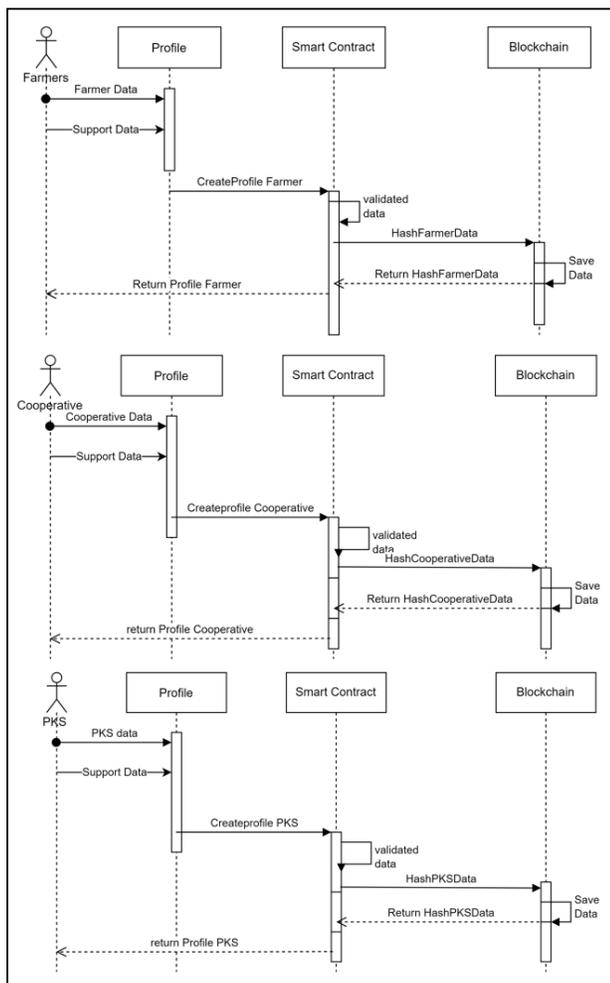
Fig. 14. Sequence Diagram Proposed Smart Contract on Smallholder Oil Palm Supply Chain.

Fig. 15. Structure of Smart Contract on Smallholder Oil Palm Supply Chain.

2) Data transformation model: Data transformation is part of the development of the current system into a blockchain-based model. Also, data changes and adjustments are needed in determining the data to be stored in the internal database environment and those to be stored on the blockchain network. Fig. 16 shows the mechanism for adapting the existing data in the current system into a blockchain-based model.

The databases managed by centralized authorities typically store multiple copies of data for secure server management. A copy of the data is very important because those stored are prone to corruption. The decentralized nature of blockchain

has a major influence on the changes in how each industry implement digital transaction system. Furthermore, blockchain allows anyone who accesses it to work independently and eliminates any need for centralized control. For example, a server-based database acts as a centralized processing unit, thereby allowing the database to work dynamically from small to large scale. The client and server communicate with each other over a secure connection. It was observed that blockchain is based on distributed ledger technology and the network is enabled with a secure, peer-to-peer cryptographic protocol. Meanwhile, it shows another way of working, in which the accessor can only read and write. It is important to note that the data stored on a blockchain cannot be deleted or replaced. This is because its nature is fixed and it supports immutability, hence the inputted data is not disturbed or manipulated.

Fig. 17 shows the transformation of the proposed data model from a centralized to a decentralized database. The decentralized data model on the blockchain network is facilitated by the existence of an on-chain transaction mechanism stored in the blockchain network environment. It also involves an off-chain transaction mechanism, which is a data model stored outside the blockchain network. The proposed model uses a NoSQL database type to facilitate the development of a blockchain-based system, as shown in Fig. 18.
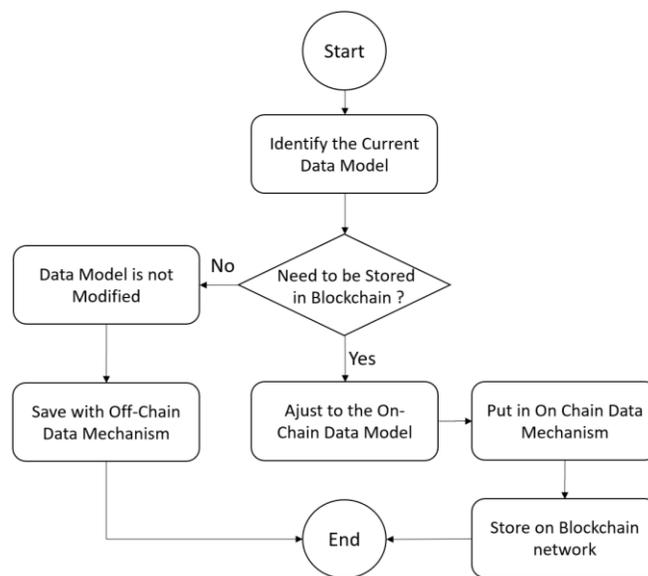
Fig. 16. Blockchain-based Smallholder Oil Palm Supply Chain Data Transformation Strategy.
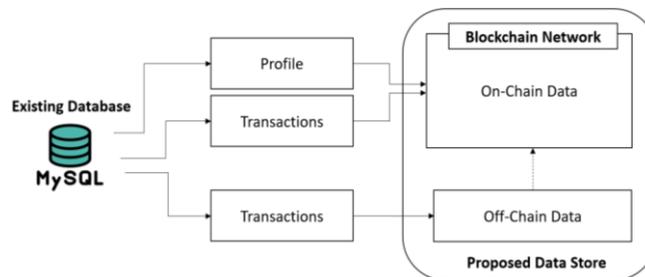
Fig. 17. Proposed Decentralized Data Model Transformation.

Fig. 18. Example of a NoSQL-based Data Model.

The data format used in database development, which is stored in a No-SQL and the process of sending it to the blockchain network with json formatted APIs is shown in Fig. 19.

Fig. 19. On-chain and off-chain Data Models in Smallholder Oil Palm Supply Chain.

3) Network architecture transformation model: Network transformation is the evolution of the current model to a blockchain-based system. Specifically, the current system works with a centralized network, while the blockchain network is decentralized and distributed.

Fig. 20 shows the mechanism for developing a blockchain network architecture on a smallholder oil palm supply chain system. In addition, the stages of change include 1) an analysis of the web platform technology used in the current system, 2) the development towards a web platform that supports blockchain, namely by building a blockchain network, 3) the deployment of a smart contract on a blockchain network, and 4) end-user application development.

The network architecture in the smallholder oil palm supply chain is shown in Fig. 21. Starting from an end-user, applications interact with a smart contract to perform operations on the blockchain network. Consequently, the blockchain stores these records immutably and return the changes in response to the client application.

A smart contract works by updating the ledger in the blockchain network, which represents the immutable data on all transactions and manages a world state denoting the current data state. Furthermore, it is capable of executing the state stored in the state DB and requesting transaction records contained in the blockchain. Client applications interact with a smart contract to perform operations on the blockchain network. This causes the blockchain to store these records immutably and return the changes in response to the client application.
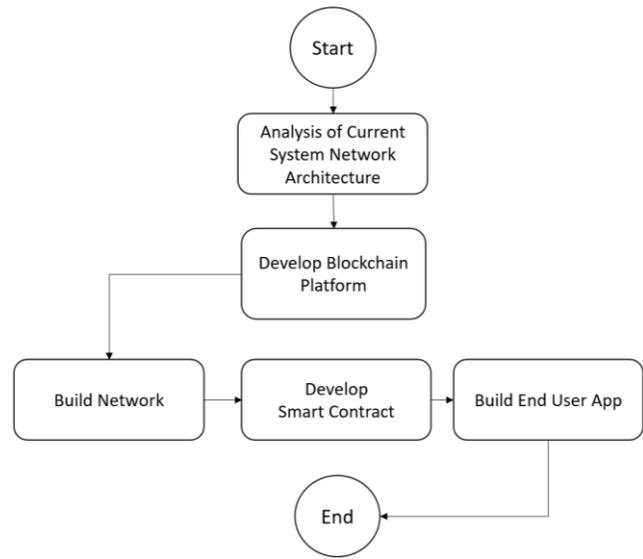
Fig. 20. Blockchain-based Smallholder Oil Palm Supply Chain Network Architecture Transformation Strategy.
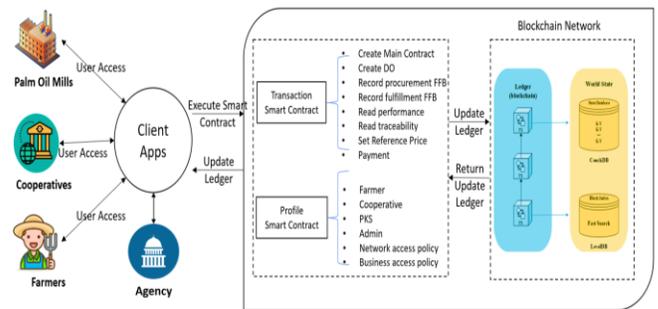
Fig. 21. Proposed Network Architecture in Smallholder Oil Palm Supply Chain.

This is consistent with the study in [57], that the development of a smart contract started from the re-engineering of existing business processes. According to [58], the smart contract's function is applied as an automation mechanism and data integrity guard. This means the smart contract has the ability as a secure data storage mechanism, particularly for access to a smallholder oil palm supply chain [59]. However, this study was only completed at the transformation model stage, hence it is necessary to conduct a maturity assessment of the digital transformation model [60]. It also needs to be developed towards a physical model in the form of developing a prototype system, in which the data, transactions, APIs used, and the blockchain platform was tested [61].

V. CONCLUSION

This study has led to a transformation model from the current smallholder oil palm supply chain to a blockchain-smart contract-based system. The system transformation produced consisted of elements of transaction, data, and

network architecture transformations. The results were obtained by taking a digital transformation approach, particularly at the technological level, and focusing on the operational processes. The proposed system model was derived from the partial integration mechanism for the adoption of blockchain technology. This is because the current system uses an information technology/information system in its working process. The results represented the As-Is To-Be model in developing the system transformation. Specifically, the current system was used as an As-Is model, which was analyzed and evaluated, while the To-Be was the result of analysis and collaboration with blockchain technology, serving as the solution. Furthermore, a flow of transformation strategies in transactions, data, and network architecture was produced. The transaction transformation was based on business processes and those occurring within the smallholder oil palm supply chain. This transaction is analyzed in order to get a picture of the parts that are still being conducted manually by the user, have low accuracy, and lack integrity. Based on the investigation results, a form of change and adjustment of transactions was obtained using the smart contract mechanism. Regarding the data transformation, the results were achieved through the analysis performed on the current system model, which was developed with data in accordance with blockchain technology. The evaluation results showed that some data in the smallholder oil palm supply chain need to be included in the blockchain in order to maintain its security, integrity, and transparency. Consequently, the direction in this section was to apply on-chain and off-chain data storage mechanisms. The last transformation was the network architecture, in which the current system employs an architectural model and the network used was centralized, while blockchain technology was distributed and decentralized. This indicates the need to develop a network that is able to distribute records/ledgers to all nodes and user applications. It is important to mention that these applications are developed based on Dapps capable of working in a blockchain environment with smart contract by utilizing APIs as a medium for connecting data to be stored in the network. This study's results were practically useful as a guide, direction, and strategy for developing the existing model for it to be transformed into a new system based on blockchain-smart contract technology. In the future, studies have to be directed toward the development of physical models and prototypes of the blockchain-smart contract-based smallholder oil palm supply chain system. This can be implemented using a suitable platform with a smart contract model and executed on a blockchain-based prototype environment.

REFERENCES

[1] T. Djatna, M. Asrol, and T. Baidawi, "SCOR-Based Information Modeling for Managing Supply Chain Performance of Palm Oil Industry at Riau and Jambi Provinces, Indonesia," Int. J. Supply Chain Manag., vol. 9, no. 5, pp. 75–89, 2020.

[2] S. Raharja et al., "Institutional strengthening model of oil palm independent smallholder in Riau and Jambi Provinces, Indonesia," Heliyon, vol. 6, no. 5, p. e03875, 2020, doi: 10.1016/j.heliyon.2020.e03875.

[3] H. Heryani et al., "Institutional Development in the Supply Chain System of Oil Palm Agroindustry in South Kalimantan," Int. J. Technol., vol. 13, no. 3, pp. 643–654, 2022, doi: 10.14716/ijtech.v13i3.4754.

[4] H. Halaburda, "Blockchain revolution without the blockchain?," Commun. ACM, vol. 61, no. 7, pp. 27–29, 2018.

[5] I. Önder and H. Treiblmaier, "Blockchain and tourism: Three research propositions," Ann. Tour. Res., vol. 72, pp. 180–182, 2018, doi: https://doi.org/10.1016/j.annals.2018.03.005.

[6] F. Idelberger, G. Governatori, R. Riveret, and G. Sartor, "Evaluation of logic-based smart contracts for blockchain systems," in International symposium on rules and rule markup languages for the semantic web, 2016, pp. 167–183.

[7] M. M. H. Onik, M. H. Miraz, and C.-S. Kim, "A recruitment and human resource management technique using blockchain technology for industry 4.0," in Smart Cities Symposium 2018, 2018, pp. 1–6.

[8] M. Conoscenti, A. Vetro, and J. C. De Martin, "Blockchain for the Internet of Things: A systematic literature review," in 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), 2016, pp. 1–6.

[9] O. Bermeo-Almeida, M. Cardenas-Rodriguez, T. Samaniego-Cobo, E. Ferruzola-Gómez, R. Cabezas-Cabezas, and W. Bazán-Vera, "Blockchain in agriculture: A systematic literature review," in International Conference on Technologies and Innovation, 2018, pp. 44–56.

[10] A. Gorkhali, L. Li, and A. Shrestha, "Blockchain: A literature review," J. Manag. Anal., vol. 7, no. 3, pp. 321–343, 2020.

[11] F. Zaoui and N. Souissi, "Roadmap for digital transformation: A literature review," Procedia Comput. Sci., vol. 175, pp. 621–628, 2020, doi: 10.1016/j.procs.2020.07.090.

[12] I. Afrianto, T. Djatna, Y. Arkeman, I. S. Sitanggang, and I. Hermadi, "Disrupting Agro-industry Supply Chain in Indonesia With Blockchain Technology: Current and Future Challenges," in 2020 8th International Conference on Cyber and IT Service Management (CITSM), 2020, pp. 1–6.

[13] Marimin et al., "Supply chain performance measurement and improvement of palm oil agroindustry: A case study at Riau and Jambi Province," IOP Conf. Ser. Earth Environ. Sci., vol. 443, no. 1, 2020, doi: 10.1088/1755-1315/443/1/012056.

[14] N. S. Rahayu, A. A. Nugroho, and R. R. Yusuf, "Exclusion of Smallholders in the Indonesia Palm Oil Industry," KnE Soc. Sci., vol. 2022, pp. 1158–1182, 2022, doi: 10.18502/kss.v7i9.11010.

[15] H. Bakhtary, F. Haupt, C. Luttrell, D. Landholm, and I. Jelsma, "Promoting sustainable oil palm production by independent smallholders in Indonesia," Clim. Focus Meridian Inst., no. February, pp. 1–12, 2021, doi: 10.13140/RG.2.2.35216.74242.

[16] K. Anwar, D. Tampubolon, and T. Handoko, "Institutional Strategy of Palm Oil Independent Smallholders: A Case Study in Indonesia," J. Asian Financ. Econ. Bus., vol. 8, no. 4, pp. 229–238, 2021, doi: 10.13106/jafeb.2021.vol8.no4.0529.

[17] J. D. Watts et al., "Challenges faced by smallholders in achieving sustainable palm oil certification in Indonesia," World Dev., vol. 146, p. 105565, 2021, doi: 10.1016/j.worlddev.2021.105565.

[18] E. P. Pramudya et al., "Incentives for Palm Oil Smallholders in Mandatory Certification in Indonesia," Land, vol. 11, no. 4, pp. 1–28, 2022, doi: 10.3390/land11040576.

[19] A. Pieroni, N. Scarpato, L. Di Nunzio, F. Fallucchi, and M. Raso, "Smarter City: Smart energy grid based on Blockchain technology," Int. J. Adv. Sci. Eng. Inf. Technol., vol. 8, no. 1, pp. 298–306, 2018, doi: 10.18517/ijaseit.8.1.4954.

[20] A. Litke, D. Anagnostopoulos, and T. Varvarigou, "Blockchains for Supply Chain Management: Architectural Elements and Challenges

Towards a Global Scale Deployment," Logistics, vol. 3, no. 1, p. 5, 2019, doi: 10.3390/logistics3010005.

[21] I. Afrianto, T. Djatna, Y. Arkeman, I. Sukaesih Sitanggang, and I. Hermadi, "Disrupting Agro-industry Supply Chain in Indonesia with Blockchain Technology: Current and Future Challenges," 2020, doi: 10.1109/CITSM50537.2020.9268872.

[22] F. Tian, "An agri-food supply chain traceability system for China based on RFID & blockchain technology," 2016 13th Int. Conf. Serv. Syst. Serv. Manag. ICSSSM 2016, 2016, doi: 10.1109/ICSSSM.2016.7538424.

[23] L. Cheng, J. Liu, C. Su, K. Liang, G. Xu, and W. Wang, "Polynomial-based modifiable blockchain structure for removing fraud transactions," Futur. Gener. Comput. Syst., vol. 99, pp. 154–163, 2019, doi: https://doi.org/10.1016/j.future.2019.04.028.

[24] P. Helo and Y. Hao, "Blockchains in operations and supply chains: A model and reference implementation," Comput. Ind. Eng., vol. 136, no. July, pp. 242–251, 2019, doi: 10.1016/j.cie.2019.07.023.

[25] I. Afrianto and Y. Heryanto, "Design and Implementation of Work Training Certificate Verification Based On Public Blockchain Platform," in 2020 Fifth International Conference on Informatics and Computing (ICIC), 2020, pp. 1–8.

[26] Y. Xinyi, Z. Yi, and Y. He, "Technical Characteristics and Model of Blockchain," pp. 562–566, 2018, doi: 10.1109/controlo.2018.8439793.

[27] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," Proc. - 2017 IEEE 6th Int. Congr. Big Data, BigData Congr. 2017, no. October, pp. 557–564, 2017, doi: 10.1109/BigDataCongress.2017.85.

[28] J.-C. Cheng, N.-Y. Lee, C. Chi, and Y.-H. Chen, "Blockchain and smart contract for digital certificate," in 2018 IEEE international conference on applied system invention (ICASI), 2018, pp. 1046–1051.

[29] B. K. Mohanta, S. S. Panda, and D. Jena, "An overview of smart contract and use cases in blockchain technology," in 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2018, pp. 1–4.

[30] M. Swan, "Chapter Five - Blockchain for Business: Next-Generation Enterprise Artificial Intelligence Systems," in Blockchain Technology: Platforms, Tools and Use Cases, vol. 111, P. Raj and G. C. B. T.-A. in C. Deka, Eds. Elsevier, 2018, pp. 121–162.

[31] I. Afrianto, C. R. Moa, S. Atin, and I. Rosyidin, "Prototype Blockchain Based Smart Contract For Freelance Marketplace System," in 2021 Sixth International Conference on Informatics and Computing (ICIC), 2021, pp. 1–8.

[32] D. Macrinici, C. Cartofeanu, and S. Gao, "Smart contract applications within blockchain technology: A systematic mapping study," Telemat. Informatics, vol. 35, no. 8, pp. 2337–2354, 2018, doi: https://doi.org/10.1016/j.tele.2018.10.004.

[33] R. Yuan, Y.-B. Xia, H.-B. Chen, B.-Y. Zang, and J. Xie, "Shadoweth: Private smart contract on public blockchain," J. Comput. Sci. Technol., vol. 33, no. 3, pp. 542–556, 2018.

[34] D. Tapscott and A. Tapscott, Blockchain revolution: how the technology behind bitcoin is changing money, business, and the world. Penguin, 2016.

[35] F. Idelberger and G. Governatori, "Rule Technologies. Research, Tools, and Applications," vol. 9718, no. November 2017, pp. 167–183, 2016, doi: 10.1007/978-3-319-42019-6.

[36] I. Afrianto, T. Djatna, Y. Arkeman, I. Hermadi, and I. S. Sitanggang, "Block Chain Technology Architecture For Supply Chain Traceability Of Fisheries Products In Indonesia: Future Challenge," J. Eng. Sci. Technol. Spec. Issue INCITEST2020, pp. 41–49, 2020.

[37] W. Reijers, F. O'Brolcháin, and P. Haynes, "Governance in Blockchain Technologies & Social Contract Theories," Ledger, vol. 1, pp. 134–151, 2016, doi: 10.5195/ledger.2016.62.

[38] L. Wang, X. Shen, J. Li, J. Shao, and Y. Yang, "Cryptographic primitives in blockchains," J. Netw. Comput. Appl., vol. 127, pp. 43–58, 2019, doi: https://doi.org/10.1016/j.jnca.2018.11.003.

[39] Y. Cao, "3 - Energy Internet blockchain technology," W. Su and A. Q. B. T.-T. E. I. Huang, Eds. Woodhead Publishing, 2019, pp. 45–64.

[40] Y. Zhao, X. Meng, S. Wang, and T. C. Edwin Cheng, Contract Analysis and Design for Supply Chains with Stochastic Demand, vol. 234. 2015.

[41] M. N. M. Bhutta et al., "A Survey on Blockchain Technology: Evolution, Architecture and Security," IEEE Access, vol. 9, pp. 61048–61073, 2021, doi: 10.1109/ACCESS.2021.3072849.

[42] R. A. Canessane, N. Srinivasan, A. Beuria, A. Singh, and B. M. Kumar, "Decentralised applications using ethereum blockchain," in 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), 2019, vol. 1, pp. 75–79.

[43] C. Rupa, D. Midhunchakkaravarthy, M. K. Hasan, H. Alhumyani, and R. A. Saeed, "Industry 5.0: Ethereum blockchain technology based DApp smart contract," Math. Biosci. Eng., vol. 18, no. 5, pp. 7010–7027, 2021.

[44] M. S. Abhijith, A. T. M. Achuthan, M. Alan Babu, and K. Shyam Krishna, "Enhanced Pharmaceutical Supply Chain Management Using Ethereum Blockchain," Int. J. Innov. Sci. Res. Technol. ISSN, no. 2456–2165, 2021.

[45] S. K. Panda and S. C. Satapathy, "An investigation into smart contract deployment on ethereum platform using Web3. js and solidity using blockchain," in Data Engineering and Intelligent Computing, Springer, 2021, pp. 549–561.

[46] P. Kamboj, S. Khare, and S. Pal, "User authentication using Blockchain based smart contract in role-based access control," Peer-to-Peer Netw. Appl., vol. 14, no. 5, pp. 2961–2976, 2021.

[47] D. Li, W. E. Wong, and J. Guo, "A survey on blockchain for enterprise using hyperledger fabric and composer," in 2019 6th International Conference on Dependable Systems and Their Applications (DSA), 2020, pp. 71–80.

[48] H. Mukne, P. Pai, S. Raut, and D. Ambawade, "Land record management using hyperledger fabric and IPFS," in 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1–8.

[49] Q. Nasir, I. A. Qasse, M. Abu Talib, and A. B. Nassif, "Performance analysis of hyperledger fabric platforms," Secur. Commun. Networks, vol. 2018, 2018.

[50] D. Ravi, S. Ramachandran, R. Vignesh, V. R. Falmari, and M. Brindha, "Privacy preserving transparent supply chain management through Hyperledger Fabric," Blockchain Res. Appl., vol. 3, no. 2, p. 100072, 2022.

[51] H. Sukhwani, N. Wang, K. S. Trivedi, and A. Rindos, "Performance modeling of hyperledger fabric (permissioned blockchain network)," in 2018 IEEE 17th International Symposium on Network Computing and Applications (NCA), 2018, pp. 1–8.

[52] "Peers — Tài liệu hyperledger-fabricdocs main." https://hyperledger-fabric.readthedocs.io/vi/latest/peers/peers.html (accessed Sep. 05, 2022).

[53] E. Gonçalves and I. Monteiro, "Reporting the Usage of iStar in a Model-Based Industrial Project to Evolve an e-Commerce Application," CEUR Workshop Proc., vol. 2983, pp. 8–14, 2021.

[54] K. Pousttchi, A. Gleiss, B. Buzzi, and M. Kohlhagen, "Technology impact types for digital transformation," Proc. - 21st IEEE Conf. Bus. Informatics, CBI 2019, vol. 1, pp. 487–494, 2019, doi: 10.1109/CBI.2019.00063.

[55] Capgemini Consulting, "Accelerating Digital Transformation," Digit. Transform. Rev., no. May, p. 62, 2013.

[56] M. Alharby and A. van Moorsel, "Blockchain Based Smart Contracts : A Systematic Mapping Study," pp. 125–140, 2017, doi: 10.5121/csit.2017.71011.

[57] I. Sarker and B. Datta, "Re-designing the pension business processes for achieving technology-driven reforms through blockchain adoption: A proposed architecture," Technol. Forecast. Soc. Change, vol. 174, no. May 2020, p. 121059, 2022, doi: 10.1016/j.techfore.2021.121059.

[58] I. A. Omar, R. Jayaraman, M. S. Debe, K. Salah, I. Yaqoob, and M. Omar, "Automating Procurement Contracts in the Healthcare Supply Chain Using Blockchain Smart Contracts," IEEE Access, vol. 9, pp. 37397–37409, 2021, doi: 10.1109/ACCESS.2021.3062471.

[59] Z. Sun, D. Han, D. Li, X. Wang, C.-C. Chang, and Z. Wu, "A blockchain-based secure storage scheme for medical information," EURASIP J. Wirel. Commun. Netw., vol. 2022, no. 1, pp. 1–25, 2022.

[60] E. Gökalp and V. Martinez, "Digital transformation maturity assessment: development of the digital transformation capability maturity model," Int. J. Prod. Res., 2021, doi: 10.1080/00207543.2021.1991020.

[61] L. Hang, I. Ullah, and D.-H. Kim, "A secure fish farm platform based on blockchain for agriculture data integrity," Comput. Electron. Agric., vol. 170, p. 105251, 2020, doi: https://doi.org/10.1016/j.compag.2020.105251.

Speckle Reduction in Medical Ultrasound Imaging based on Visual Perception Model

Yasser M. Kadah[1]*, Ahmed F. Elnokrashy[2], Ubaid M. Alsaggaf[3], Abou-Bakr M. Youssef[4]

Electrical and Computer Engineering Department, King Abdulaziz University, Jeddah, Saudi Arabia[1, 3]
Electrical Engineering Department, Benha University, Benha, Egypt[2]
Biomedical Engineering Department, Cairo University, Giza, Egypt[1, 4]

Abstract—**Ultrasound imaging technology is one of the most important clinical imaging modalities due to its safety, low cost, in addition to its versatile applications. The main technical problem in this technology is its noisy appearance due to the presence of speckle, which makes reading imaging more difficult. In this study, a new method of speckle reduction in medical ultrasound images is proposed based on adaptive shifting of the contrast sensitivity function of human vision using a bias field map estimated from the original image. The aim of this work is to have an effective image enhancement strategy that reduces speckle while preserving diagnostically useful image features and allowing practical implementation in real-time for medical ultrasound imaging applications. The new method is used to improve the visual perception of image quality of ultrasound images by adding a local brightness bias to the areas with speckle noise. This allows the variations in image pixels due to speckle noise to be better perceived by the human observer because of the visual perception model. The performance of the proposed method is objectively assessed using quantitative image quality metrics and compared to previous methods. Furthermore, given that image quality perception is subjective, the level of added bias is controlled by a single parameter that accommodates the different needs for different users and applications. This method has potential to offer better viewing conditions of ultrasound images, which translates to higher diagnostic accuracy.**

Keywords—Contrast sensitivity function; image quality metrics; speckle reduction; ultrasound imaging

I. INTRODUCTION

Ultrasound imaging technology is one of the most important clinical imaging modalities due to its safety, low cost, in addition to its versatile soft tissue imaging applications that include abdominal imaging, echocardiography, and obstetrics and gynaecology [1]. Given that this technology relies on sending low intensity acoustic waves into the body and receiving the reflected and scattered echoes to reconstruct a cross sectional image of the body, it is inherently safe and can be considered as the safest imaging modality available today. This is evident by being the only imaging modality allowed for use on pregnant women to follow up the growth of the fetus and assess its biophysical profile in addition to detecting any anomalies present. Another advantage of ultrasound imaging technology is its availability in a wide range of forms including basic hand-held or portable units that cost a few thousand dollars to sophisticated dedicated echocardiography systems that cost hundreds of thousands of dollars. This allows ultrasound imaging to be popular for use in rural and low-income communities as well as in large,

specialized hospitals. Future direction in this technology indicates that it will likely become as essential to medical practice as the stethoscope and it is predicted that general practitioners will soon have small ultrasound units in their clinical practice.

Despite the many advantages of ultrasound imaging technology, it has a clear shortcoming in the quality of its image as compared to other modalities such as x-rays, magnetic resonance imaging or computed tomography. Ultrasound images generally look very noisy to the viewer and requires some training to correlate the anatomy to the acquired ultrasound images. This is mainly due to the presence of speckle that sometimes obscure pathological changes in the body and hence may cause errors in the diagnosis. Therefore, the problem of speckle reduction in ultrasound images has been the focus of research of many academic and industrial research groups since the early beginnings of ultrasound imaging and is expected to remain so given its impact on this technology [2].

Speckle noise is an unavoidable direct result of the physics of ultrasound imaging. Ultrasound imaging is done by sending an ultrasonic pulse through the body from an ultrasonic transducer arranged in 1D or 2D array form. This ultrasonic pulse propagates through the tissues and interact with its different components yielding reflected waves from specular reflectors and scattering from point reflectors [1]. The difference between specular and point reflectors is mainly in size whereas specular reflectors are larger than the wavelength of ultrasound waves while scatterers are significantly smaller than this wavelength. This wavelength is a fraction of a millimeter in the usual range of ultrasound imaging frequencies in the 2-15 MHz. Therefore, tissue interfaces and major blood vessels behave like specular reflectors while blood capillaries and cells within the extracellular space act as scatterers [3][4]. So, in a tissue like the liver, hepatocytes scatter ultrasound waves independently and the backscattered part is what is received by the ultrasound transducer. Given that such scattering depends on the ultrasound transducer frequency and orientation as well as the complex 3D structure of the tissue, the received scattered waves from the many cells involved interfere and produce a pattern of partial constructive and destructive interference points that show as random noise in the image. The difference between speckle noise and true random noise is that random noise change with time and hence can be reduced by simple averaging. On the other hand, speckle noise pattern remains the same as long as the imaging

*Corresponding Author

conditions remain the same (i.e., probe orientation, frequency, and tissue being imaged). Therefore, it does not improve with averaging and other more sophisticated methods be utilized to reduce its effect on image quality [5][6].

II. LITERATURE REVIEW

Several approaches have been proposed to address the problem of speckle reduction. Such approaches can be broadly classified as acquisition and post-processing methods. The acquisition methods try to reduce speckle through the acquisition of multiple versions of the same slice with each acquired with different beamforming (e.g., using steering, different focal point, different frequency, etc.). Then, since the speckle pattern is different when beamforming changes, using simple averaging of these acquisitions will result in an image with reduced speckle [1][7][8]. Despite the intuitiveness of the approach and apparent simplicity, it practically requires reprogramming of the acquisition protocols of different applications, which adds significant cost in addition to being a cumbersome constraint in some applications that require high frame rate and/or special ordering of acquisitions such as in 3D and 4D imaging.

On the other hand, the second approach based on preprocessing was the focus of virtually all research groups working on this problem since the 1980s until today. Its basic premise is to start from the reconstructed image and apply different filters to improve the speckle pattern. Hence, it requires only a computer and access to the frame buffer of the ultrasound imaging system to perform its job. Moreover, with the advent of modern computing platforms and massively parallel processing hardware such graphics processing units (GPUs) that all became well within budget for ultrasound imaging systems, this approach seems like the logical starting point for practical purposes. From the technical point of view, post-processing methods can be categorized into linear, nonlinear, diffusion-based, and wavelet-based filtering methods [2]. The linear filters include such techniques as First-Order Statistics Filtering, Local Statistics Filtering with Higher Moments, and Homogeneous Mask Area Filtering [9][10][11]. The nonlinear filtering methods include Median Filtering, Linear Scaling Filter, Geometric Filtering, and Homomorphic Filtering [12][13][14][15]. The diffusion-based methods include several variants of Anisotropic Diffusion Filtering [16][17][18][19][20]. The wavelet-based methods mainly work using wavelet shrinkage using different wavelet families and levels of composition [21][22][23][24]. Hybrids between the above methods were also introduced [25][26][27].

Despite the relative success achieved by the present methods in reducing speckle in ultrasound images, there remains several problems that hinder the practicality of their use in clinical ultrasound. First, there is a gap between what the quality of the processed image means to researchers and clinicians. That is, the outcome of these methods may look smoother and hence better to the researchers but to the contrary such appearance may hide some features in the image that are important to the clinical sonographers. So, there is a need to develop better quality metrics that more accurately reflect the clinical view of the image quality [28]. Second, the focus only on reconstructed images may be the easiest way to

address the speckle reduction problem but it certainly is not the optimal one. If a better solution to this problem is to be achieved, the different aspects of image formation must all be taken into consideration and optimized together to get the best results. This includes optimizing acquisition, beamforming, dynamic range compression, detection, sampling, quantization, scan conversion, and image reconstruction algorithms. Third, even though parallel processing has been around for a while now, very few reconstruction algorithms have been designed to work on parallel processing platforms. This is a clear disadvantage because they will not run in real-time and/or pose a scheduling problem in the processing chain in ultrasound imaging systems especially in complex, time-critical applications such as 4D imaging.

One of the important aspects of ultrasound image quality is how it is perceived by a human sonographer. In particular, the psychophysiological aspects of human vision have a profound effect on such things as contrast sensitivity and details resolution that influence diagnostic quality [28]. In [29], Campbell and Robson presented the concept of contrast sensitivity function (CSF) that define the effects of contrast and spatial frequency content on the visual detectability by a human observer. In Fig. 1, this effect is illustrated using a simulated grating with different spatial frequency and contrast simulated based on algorithms in [30]. As can be observed, the variations in the grating are not detectible above the CSF superimposed on the grating as a dashed line. This suggests that reducing the contrast for high spatial frequency image details can reduce their detectability by a human user. In the context of speckle reduction, the goal is to make their less detectible by the sonographer. Given that the speckle noise has high spatial frequency content, this suggests that reducing their contrast would reduce their visibility by a human sonographer.
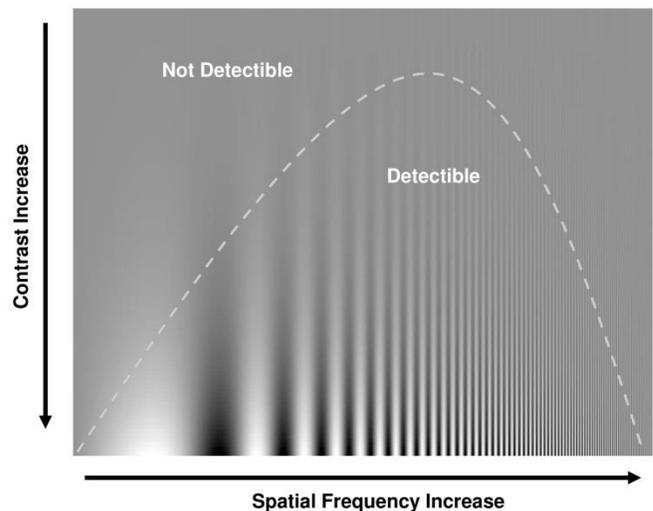
Fig. 1. The Campbell-Robson Contrast Sensitivity Function Chart where Visual Detection of Variation Depends on Both Contrast and Spatial Frequency as Outlined by the Separation with the Dashed Line.

In this work, a new method of speckle reduction in medical ultrasound imaging is developed whereby a user-controlled image display method adds a bias field map estimated from the original image that reduces speckle visibility. The new method relies on the contrast sensitivity function of the human visual perception to push the high spatial frequency content in speckle out of the visual detectability zone. This strategy does not affect the speckle pattern itself or change its statistical independence but rather allows the user to control how it is displayed to match subjective needs. The new method is experimentally verified using real ultrasound imaging data collected from a research system and compared to previous speckle reduction methods using quantitative image quality metrics.

III. Methodology

The speckle pattern in ultrasound images is characterized by a pseudo-random pattern with high contrast and high spatial frequency content due to its origin from the interference of many back-scattered waves. As a result, unlike random noise, it contains information about underlying tissue and may contribute diagnostic information in some cases. On the other hand, in many cases, the speckle pattern reduces the perceived ultrasound image quality and hence control over its visibility in the image is needed. In this work, it is proposed to take advantage of the characteristics of the contrast sensitivity function of the human visual system to develop a method that allows the visibility of speckle to be controlled by the user through appropriate contrast modification. This approach recognizes the importance of speckle as containing diagnostic information and addresses the subjective nature of image quality perception. According to this approach, reducing the contrast of speckle can improve image quality by reducing their visual detectability. The contrast is defined as the ratio of the maximum intensity to the average intensity. In order to verify this concept, a preliminary experiment was conducted where a speckle pattern from a real ultrasound imaging experiment was displayed along with the same pattern with different uniform intensity bias added to it. As this bias intensity increases, the contrast becomes lower according to its definition. In Fig. 2, the image to the left is the original

speckle pattern while the other images show the same pattern with an increasing constant bias intensity added to the original as one moves to the right. As can be observed, the coarse visual appearance of the original speckle pattern gets finer as the bias intensity increases. Even though this confirms the validity of the concept, it is still not practical to just add a constant bias to the whole image because it may affect the contrast sensitivity of other areas of the image not containing speckle. A more suitable approach is to design a spatially-variant bias map that estimates a smooth local brightness level in different parts of the image. In this work, using a simple image denoising technique such as 2D median filtering with a large kernel as the bias field is proposed. This allows the differentiation of speckle from specular reflectors in the image. The weighted averaging of this bias map and the original image would result in a variable degree of speckle visibility depending on how the weight is selected and allows the user to select the level of speckle suppression subjectively. Given that, the weighted average may have dynamic range compared to the original, a simple adjustment of the dynamic range is done to maintain the appropriate display quality. A block diagram of the new method is shown in Fig. 3. The method accepts an original image in the form of a reconstructed image or as raw data collected by the imaging system. The original image is fed into two processing blocks. The first uses one of the present image-denoising filters for bias map estimation. In this work, a 2D median filter was used with a kernel size of nine to result in a fairly constant bias within the speckle pattern areas while varying with major interfaces to maintain contrast. It is conceptually possible to use other techniques to estimate that local bias field that serves as a local estimate of average intensity in the neighborhood. Then, the output from the image denoising filter and original image are used to obtain the final image using a weighted average. The weight used is selected by the user depending on the subjective desired outcome and application and applied to the original image. In order to ensure that the resultant image is displayed properly on the imaging system monitor, the dynamic range of the final image is adjusted to fit the dynamic range of the monitors (for example, 8-bit gray scale).
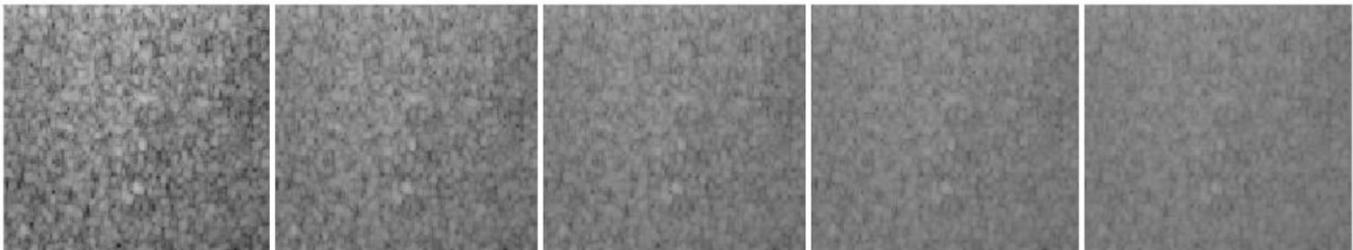
Fig. 2. Diagram Showing the Results of Adding a Constant Intensity Bias to the Original Image Shown to the Left. The Speckle Pattern Detectibility Decreases As Intensity Bias Increases From Left to Right Images.
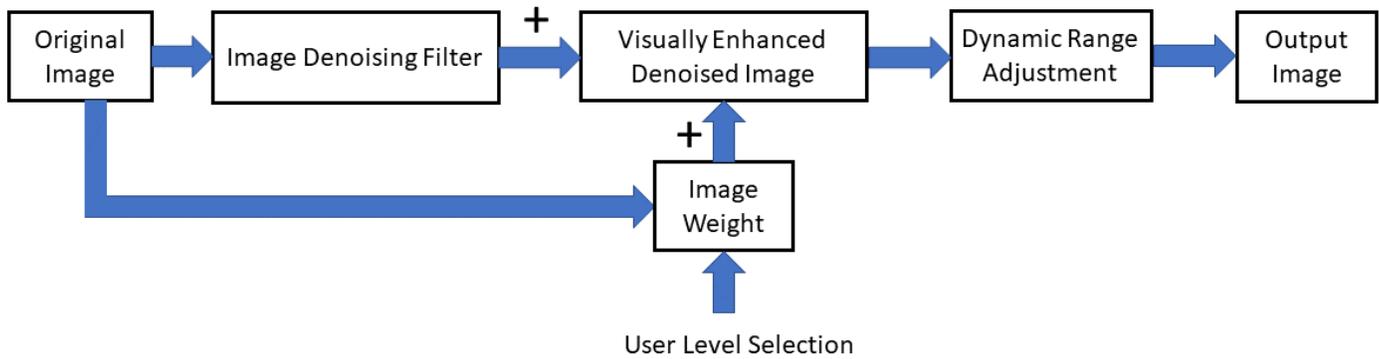
Fig. 3. Diagram Showing the Steps of the Porposed Technique Whereby Intensity Bias Map is Estimated using a Simple Image Denoising Filter and then used to Reduce Speckle Visibility in the Original Image.

IV. EXPERIMENTAL VERIFICATION

The ultrasound imaging data were collected using a Digison Digital Ultrasound Research system (Mashreq Company, Egypt). The system was equipped with a custom research interface to control image acquisition with ability to access and save raw radiofrequency sampled data for each collected image line. In order to make sure that the images are representative of different applications, the images were collected using several ultrasound array probes including convex array abdominal probe, small parts linear probe and a tight convex array endo-cavity probe. The imaging experiments were done for different clinical applications on human volunteers, as well as on a quality control tissue-mimicking phantom (Multi-Tissue Ultrasound Phantom CIRS Model 040GSE, CIRS Inc., U.S.A.). In each imaging experiment, a specific region was imaged using a specific imaging probe and a total of 10 images were collected for each application. The total number of imaging experiments done was 26 with a total number of images of 260. The research interface allowed the collection of raw image data at a sampling rate of 50 M Samples/s at 16 bits of quantization. Signal processing using filter-based Hilbert transformation for peak detection then resampling to obtain a total of 512 data samples per line (or stick). The number of image lines was 128 and this 512×128 array represents the stick data. The application of the new technique was done on the stick data using a bias field estimation from a 2D median filter with kernel size of nine and the weighting factor used to combine the bias field and original image was takes as 0.5. The dynamic range modification was done using simple window/level operation to ensure that the histogram of the combined image extends over the available display dynamic range of 8-bits. The image quality of output images is estimated using two quantitative image quality metrics of structural similarity index (SSIN) [31], and universal quality metric [32]. Given that each application may have different image characteristics, the image quality metrics from the 10 images collected for each of the 26 different imaging experiments were averaged to provide more reliable comparisons.

The final image reconstruction was subsequently performed using scan conversion and/or interpolation according to the array geometry and dimensions to display the image in the correct spatial format. All processing was done on Matlab 2022b (MathWorks, Inc.) using an educational license available through King Abdulaziz University. The computing platform consists of a personal computer with 11th generation Intel® Core™ i7-11700F running at 2.50 GHz clock and using a 64-bit Windows 11 Home Edition, with 32 GB of RAM.

V. RESULTS AND DISCUSSION

The output image results from the new technique as compared to the original images and four representative techniques covering the current approaches in speckle reduction as applied to sample applications are shown in Fig. 4. The previous techniques considered are wavelet denoising [21][23], relaxed median (RMedian) denoising [14][15], speckle reducing anisotropic diffusion (SRAD) [17][16][18], and local statistics based filtering (Lee) [9][10]. In each of these techniques, the original technique is implemented with the implementation details suggested in the most recent variant. As can be observed, the new technique shows finer speckle pattern and less blurring as compared to present techniques. This is particularly evident in the linear array example at the bottom of the page where the texture is significantly smoother compared to other techniques without having a blurring problem as found with the techniques based on wavelet denoising and lee filter.

The performance of different techniques was quantitatively assessed using the structural similarity index and universal quality image quality metrics for each imaging experiment. The results are shown in Fig. 5. Each point on these plots represents the average of the respective image quality metric computed over the 10 images collected in each experiment. As can be observed, the proposed method had better metrics in the majority of experiments followed by SRAD and RMedian techniques.
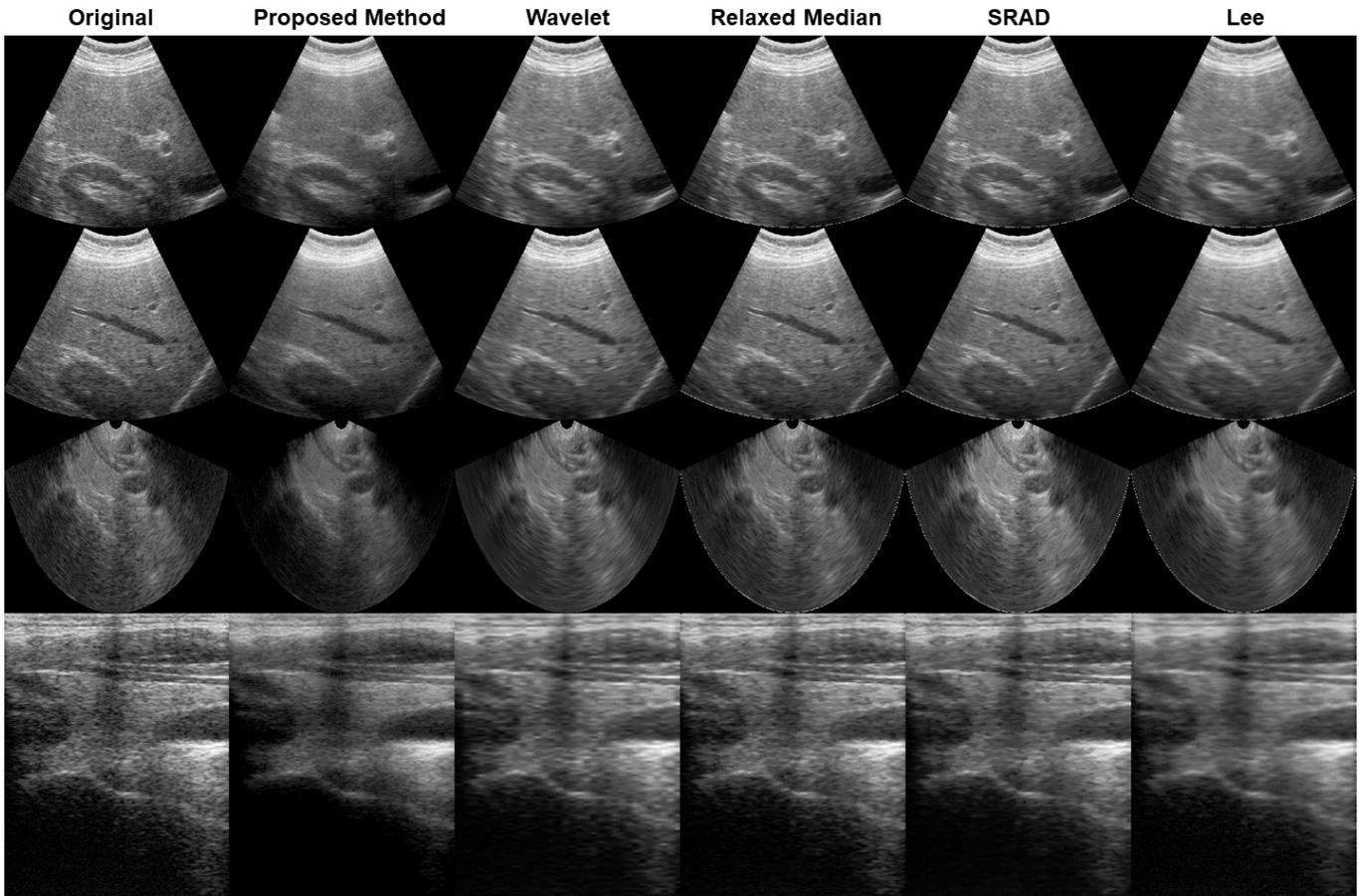
Fig. 4.   Fig. 4. Diagram Showing the Output Image Results from the Porposed Technique as Compared to the Original Images and Four Previous Techniques. As Can Be Observed, the Proposed Method Offers Finer, Less Detectible Speckle Pattern Without Image Blurring.
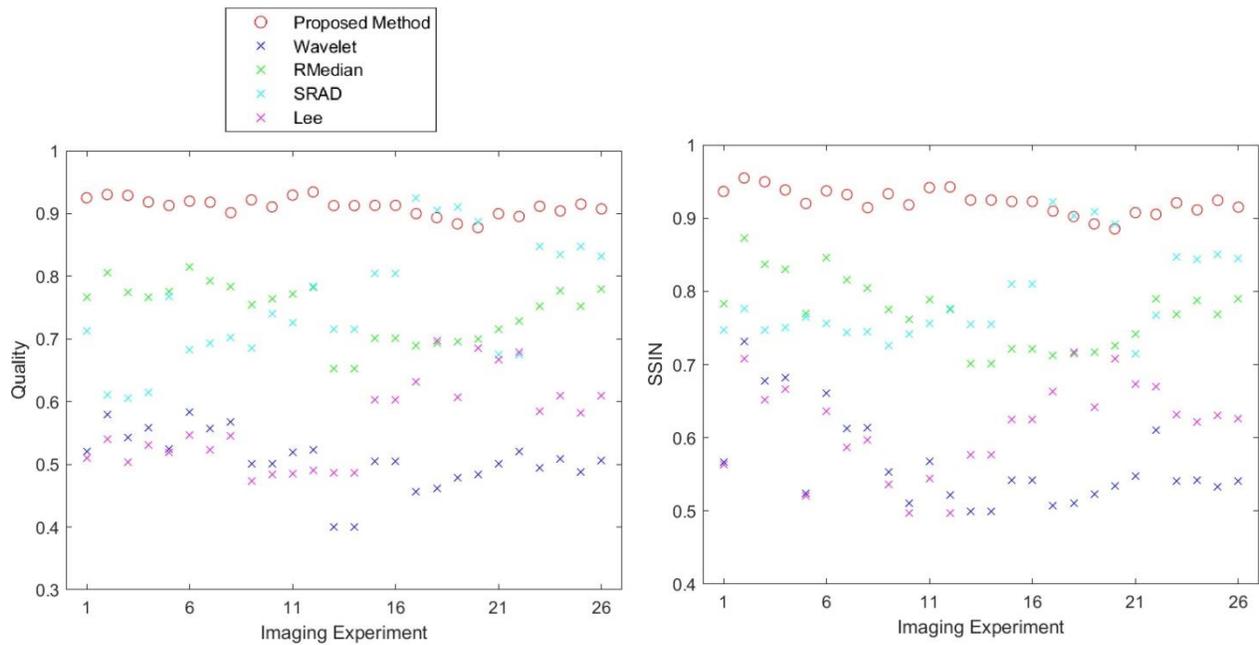
Fig. 5.   The Performance Comparison of Different Techniques as Evaluated by Two Quantitative Image Quality Metrics for Each Imaging Experiment.
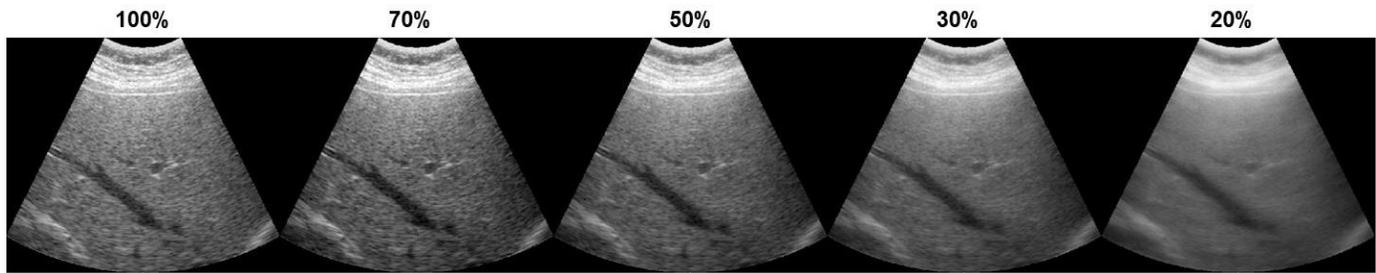
Fig. 6.   Diagram Showing the Output Image for Different user Level Selections Starting from the Original Image (No Intensity Bias) on the Left With Higher Contribution of Bias Intensity From the Left to the Right Images.

To demonstrate the value of user level selection over the speckle visibility in the output image, the output image from the same experiment using levels of 100% (original image without adding intensity bias field), 70%, 50% (used in generating previous results), 30% and 20% are provided in Fig. 6. As can be seen, the level of speckle visibility decreases as one moves to a lower level of original image contribution to the weighted average. This comes at the expense of possible blurring depending on how the bias field is generated. Therefore, a selection of a level around 50% seems to strike a balance between the two requirements of reduced speckle and blurring. This allows the sonographer to adjust the level to what he peceives as the optimal image quality for improved diagnosis.

The challenge facing the broad adoption of the new methodology are mainly due to the way the implementation in real clinical settings. Usually, the sonographer looks at the real-time images on the monitor of the ultrasound imaging system to make the diagnosis while doing the scan. If the implementation is not done on the ultrasound imaging system itself, an external computer must be connected to the system to acquire the images, process them in real-time and then display them on a medical grade monitor. The computational complexity of the new method is essentially $O(N^2)$, which is of the same order as the image reconstruction process itself and does not pose any problems in real-time processing on modern computing platforms. The main problem in the external processing option lies in how the extensive measurements and calculations packages usually offered on ultrasound imaging systems can be accessed and used during the scan. Since asking the sonographer to use an external computer to control and display ultrasound images on an external monitor and use the built-in ultrasound imaging for measurements and calculations might be inconvenient to the doctor, the implementation as an integrated original equipment manufacturer module might be the best alternative.

VI. CONCLUSIONS

A new method of speckle reduction in medical ultrasound images based on adaptive shifting on the contrast sensitivity function of human vision is proposed. The new method offers an effective image enhancement strategy that reduces speckle while preserving diagnostically useful image features and allowing practical implementation in real-time for medical ultrasound imaging applications. The new method is used to improve the visual perception of image quality of ultrasound images by adding a local contrast bias to the areas with speckle noise. This allows the variations in image pixels due to speckle noise to be better perceived by the human observer as a result of the visual perception model. Furthermore, given that image quality perception is subjective, the level of added bias is controlled by a single parameter that accommodate the different needs for different users and applications. The new method is experimentally verified using real ultrasound imaging data from 26 imaging experiments with 10 images in each. The results are evaluated qualitatively by comparing appearance of images and quantitatively using two image quality metrics. The results demonstrate the performance of the proposed method and indicate its potential to offer better viewing conditions of ultrasound images, which translates to higher diagnostic accuracy.

REFERENCES

[1]   P. R. Hoskins, K. Martin, A. Thrush, Diagnostic Ultrasound: Physics and Equipment, 2nd ed., Cambridge University Press, 2010.

[2]   C. P. Loizou, C.S. Pattichis, Despeckle Filtering for Ultrasound Imaging and Video, Volume I: Algorithms and Software, 2nd ed., Morgan & Claypool, 2015.

[3]   C. B. Burckhardt, "Speckle in ultrasound B-mode scans," IEEE Trans. Sonics Ultrasonics, vol. SU-25, no. 1, pp. 1–6, 1978.

[4]   R. F. Wagner, S.W. Smith, J.M. Sandrik, H. Lopez, "Statistics of speckle in ultrasound B-scans," IEEE Trans. Sonics Ultrasonics, vol. 30, pp. 156–163, 1983.

[5]   E. Krupinski, H. Kundel, P. Judy, C. Nodine, "The medical image perception society, key issues for image perception research," Radiology, vol. 209, pp. 611–612, 1998.

[6]   P. G. Gobbi, Modeling the Optical and Visual Performance of the Human Eye, SPIE Press, 2013.

[7]   A. Perperidis, D. Cusack, A. White, N. McDicken, T. MacGillivray, T. Anderson, "Temporal Compounding: A Novel Implementation and Its Impact on Quality and Diagnostic Value in Echocardiography," Ultrasound in Medicine & Biology, vol. 41, no. 6, pp. 1749-1765, 2015.

[8]   C. P. Loizou, C. S. Pattichis, Despeckle Filtering for Ultrasound Imaging and Video, Volume II: Selected Applications, 2nd ed., Morgan & Claypool, 2015.

[9]   J. S. Lee, "Digital image enhancement and noise filtering by using local statistics," IEEE Trans. Pattern Anal. Mach. Intell., PAMI-2, no. 2, pp. 165–168, 1980.

[10] O. Rubel, V. Lukin, A. Rubel, K. Egiazarian, "Selection of lee filter window size based on despeckling efficiency prediction for sentinel SAR images," Remote Sensing, vol. 13, no. 10, p.1887, 2021.

[11] A. F. de Araujo, C. E. Constantinou, J. Tavares, "Smoothing of ultrasound images using a new selective average filter," Expert Systems with Applications, vol. 60, pp. 96-106, 2016.

[12] J. Saniie, T. Wang, N. Bilgutay, "Analysis of homomorphic processing for ultrasonic grain signal characterization," IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 3, pp. 365–375, 1989.

[13] M. A. Gungor, I. Karagoz, "The homogeneity map method for speckle reduction in diagnostic ultrasound images," Measurement, vol. 68, pp. 100-110, 2015.

[14] A. B. Hamza, P. L. Luque-Escamilla, J. Martínez-Aroza, R. Román-Roldán, "Removing noise and preserving details with relaxed median filters," Journal of mathematical imaging and vision, vol. 11, no. 2, pp.161-177, 1999.

[15] K. Chauhan, R. K. Chauhan, A. Saini, "Enhancement and Despeckling of Echocardiographic Images," In Soft Computing Based Medical Image Analysis, Academic Press, pp. 61-79, 2018.

[16] P. Perona, J. Malik, "Scale-space and edge detection using anisotropic diffusion," IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 7, pp. 629–639, July 1990.

[17] Y. Yongjian, S. T. Acton, "Speckle reducing anisotropic diffusion," IEEE Trans. Image Process., vol. 11, no. 11, pp. 1260–1270, November 2002.

[18] H. Choi, J. Jeong, "Speckle noise reduction for ultrasound images by using speckle reducing anisotropic diffusion and Bayes threshold," Journal of X-ray Science and Technology, vol. 27, no. 5, pp.885-898, 2019.

[19] R. G. Dantas, E. T. Costa, "Ultrasound speckle reduction using modified gabor filters," IEEE Trans Ultrason Ferroelec Freq Cont, vol. 54, no. 3, pp. 530-538, 2007.

[20] K. Z. Abdel-Monem, A. M. Youssef, Y. M. Kadah, "Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion," IEEE Trans. Biomed Eng, vol. 49, no. 9, pp. 997-1014, Sept. 2002.

[21] D. L. Donoho, "Denoising by soft thresholding," IEEE Trans. Inform. Theory, vol. 41, pp. 613–627, 1995.

[22] S. Gupta, R. C. Chauhan, S. C. Sexana, "Wavelet-based statistical approach for speckle reduction in medical ultrasound images," Med Biol Eng Comput, vol. 42, pp. 189–192, 2004.

[23] A. K. Bedi, R. K. Sunkaria, "Ultrasound speckle reduction using adaptive wavelet thresholding," Multidimensional Systems and Signal Processing, vol. 33, no. 2, pp.275-300, 2022.

[24] J. Kang, J. Y. Lee, Y. Yoo, "A new feature-enhanced speckle reduction method based on multiscale analysis for ultrasound B-mode imaging," IEEE Trans Biomed Eng, vol. 63, no. 6, pp. 1178 – 1191, 2016.

[25] J. Zhang, G. Lin, L. Wu, C. Wang, Y. Cheng, "Wavelet and fast bilateral filter based de-speckling method for medical ultrasound images," Biomed Sig Proc Cont, vol. 18, pp. 1-10, 2015.

[26] B. A. Abrahim, Z. A. Mustafa, I. A. Yassine, N. Zayed, Y. M. Kadah, "Hybrid Total Variation and Wavelet Thresholding Speckle Reduction for Medical Ultrasound Imaging," J Med Imag Health Inform, vol. 2, pp. 114-124, 2012.

[27] Z. A. Mustafa, B. A. Abrahim, I. A. Yassine, N. Zayed, Y. M. Kadah, "Wavelet Domain Bilateral Filtering with Subband Mixing for Magnetic Resonance Image Enhancement," J Med Imag Health Inform, vol. 2, pp. 230-237, 2012.

[28] S. H. Schwartz, Visual Perception: A Clinical Orientation, McGraw-Hill Medical Pub. Division, 2004.

[29] F. W. Campbell, J. G. Robson. "Application of Fourier analysis to the visibility of gratings," The Journal of physiology, vol. 197, no. 3, p.551, 1968.

[30] Images for Campbell-Robson CSF chart, Available at: ttps://visiome.neuroinf.jp/database/item/3181 (Accessed on October 29, 2022).

[31] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, "Image quality assessment: From error measurement to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, April 2004.

[32] Z. Wang and A. Bovik, "A universal quality index," IEEE Signal Process. Lett., vol. 9, no. 3, pp. 81–84, March 2002.

Multi-level Video Captioning based on Label Classification using Machine Learning Techniques

J. Vaishnavi, Dr. V. Narmatha

Department of Computer and Information Science
Faculty of Science, Annamalai University, Annamalai Nagar, India

Abstract—Video captioning is the heuristic and most essential task in the current world to save time by converting long and highly content-rich videos into simple and readable reports in text form. It is narrating the events happening in videos in natural language sentences. It makes the way to many more interesting tasks by the use of labels, tags, and terms such as video content retrieval, video search, video tagging, etc. Video captioning is currently being attempted by many researchers using some exciting Deep learning techniques. But this approach is to find the best of machine learning for the process of captioning videos in a different way. The novel part of the proposed approach is classifying videos by using the labels existing in video frames that belong to the various categories and producing consecutive Multi-Level captions that describe the entire video in a round-robin way. Informative features are extracted from the video frames such as Gray Level Co-occurrence Matrix (GLCM) features, Hu moments, and Statistical features to provide optimal results. This model is designed with two superior and optimal classifiers such as Support Vector Machine (SVM) and Naive Bayes separately. The models are demonstrated with the prevailing standard dataset Microsoft Research Video Description corpus (MSVD) and evaluated by the benchmark classification metrics such as Accuracy, Precision, Recall, and F1-Score.

Keywords—*Video captioning; label classification; Hu moments; GLCM; statistical features; SVM; Naive Bayes*

I. INTRODUCTION

Video is the most used and needed multimedia which is preferred over images and text. Social media influences and increased the usage of videos. It plays an essential role in the everyday lives of people. Video is the combination of audio and scenes which narrates with a lively touch. Hearing-impaired people suffer to extract the complete information from the video. Video captioning is the current attention-seeking computer vision [1] task which is focused on by many budding researchers. Video captioning solves the issue by narrating the video in human-understandable language. It conquers almost every field, especially from education to entertainment. The vast application of video captioning is increasing every day with prime applications such as content summarization, human-robot interaction, reports, tagging, classification, video indexing, and video surveillance [2].

Video captioning is the process of describing the frames of the input video with natural sentences. It is also a task with some complexity such as, it must capture every frame of the input video and extracting the essential features by which the frames are classified with labels to provide accurate captions.

The video captioning task has more different views and possibilities to generate captions. Machine learning techniques contribute to various methods in this field. Here is the attempt to attain the best of machine learning classifiers to classify the frames along with captions based on the labels which are categorized into different categories. Firstly, videos are converted into frames and the input frames are resized by 200×200 as constant. Some essential edge, shape and texture features are extracted from processed frames. Two standard multi-class machine learning classifiers are utilized to classify the frames with appropriate captions.

The proposed model is designed to generate the captions for the videos with less time complexity and high accuracy and also to create captions for each input video frame with particular timestamps. It will be utilized in the crime branch and hearing-impaired people will learn about the happenings of the video fruitfully. Video captioning is approached by various researchers in various methods which include different techniques and cost-effective machines. But this approach is to provide a different dimension of view towards video captioning. The proposed approach is introduced to overcome the issue of utilizing cost-effective resources like GPU (Graphical Processing Unit). The proposed method is constructed with superior machine learning techniques for both feature extraction and classification and the comparative results are analyzed to declare the best model.

The article is constructed based on the following sections. Section II discusses the related works of video captioning. The proposed method is clearly explained with the needed equation and structure in Section III. The preprocessing step is discussed in sub-section IIIC. The elaborate information about the utilized benchmark dataset is given in Section IV. Section V shows the comparative study of two standard classifiers with their computed results. The article is concluded in Section VI.

II. RELATED STUDY

The field of video captioning is evolved with different techniques which various researchers develop. Still, it is considered a challenging task due to its complexities. It includes some sub-tasks such as event detection, localization, object classification, etc. The model discussed the event detection from the high-content sports video. The events are detected by using audio-visual features and classified from the multiple genres of sports video with the employment of the standard Machine learning classifier such as the Support vector machine. Video captioning is performed by utilizing

the template-based techniques [3, 4, 5] for the generation of captions for the videos in the earlier stages.

Dynamic captioning model is introduced to show the variation of the speech signal based on the volume of the audio. It positions the caption which indicates the speaker for a better understanding of hard-of-hearing people. It explores various techniques for the betterment of the model to provide satisfactory captioning, especially for hearing impaired people such as face detection and recognition, speech–text correspondences and visual alignment, etc. The model is tested in real-time with 20 video clips of 60 hearing-impaired people. Video captioning is the extension of an image captioning task. Image captioning is the task of illustrating the content of the image. Graph-based automatic captioning of the image is proposed in which is superior to other methods such as LDA, HAM, and EM with the advantages of not defining the parameters but initializing values for only two constant parameters. This model is demonstrated with the dataset Corel image database. Video is the most preferred media by people when compared to other media. It has various choices. More types of videos are available nowadays effortlessly. The need for an automatic choice selection of videos is essential to avoid wasting energy and conserve time. To overcome the issue, the preferred choice of video for the viewers is automated by the model proposed by the researchers. It is modeled by extracting the fused features of visuals and closed captions with the Hidden Markov model.

In the era of video captioning, traditional machine learning techniques are utilized in the foremost phase of video captioning evolution such as the Hidden Markov model and post-action grammar techniques [6] with various features such as hand-crafted features and object-centric features [7], etc. Videos are captioned easily by initially classifying them where it belongs to and categorized based on the content of the video [8, 9] and activities [10] that exist in the video. Visual features [11, 12, 13] are considered one of the essential features extracted to generate captions for the video. The embedding spaces are constructed between the input videos and the natural language sentences [14, 15].

Visual grounding [16, 22] is also one of the major video-related tasks which utilize visual reasoning. Hidden features [17] are extracted by using pos sequence features. A gated fusion network [18] is utilized for captioning videos. Image captioning techniques are highly functional in advanced techniques in the current time which motivate many tasks such as video captioning [19]. The superior models are designed with an attention mechanism [20] to caption the videos for boosting the performance and also by utilizing convolutions in both encoding and decoding phases. One of the major differences between image captioning and video captioning is extracting the temporal information [21] which is essential to generating captions for videos. Various features are extracted for captioning videos such as visual, object, spatial, etc. Audio features [23] are also extracted for captioning as multi-model features [24] along with speech features. Additional memory modules [25] are exploited to evade coherent captions for the video. Videos are pre-processed to generate appropriate captions in which temporal segmentation [26] in a video between various events is essential in video captioning tasks.

The memory attended recurrent network (MARN) model [27] aims to match the visuals and term that describes the visual. Visual reasoning is adopted in [28] Reasoning module network (RMN) for location and time. Transformers are utilized for the caption generation process. An accelerated masked transformer [29] is utilized in the decoding phase which generates captions especially with localizing tasks. The videos are categorized based on domains designed as domain-specific decoders [30]. Image captioning is also the one of the cause of video captioning. It makes the video captioning processes easier. Image captioning process is taken as the subsidiary content for video captioning process to enlarge the diversity [31]. Activity net captions and Microsoft coco image datasets are utilized for the captioning process which actually enlarges the diversity.

III. PROPOSED APPROACH

In the Proposed Model, the videos are categorized into various categories which are based on labels and generate Multi-Level natural language sentences based on the extracted features. The Multi-Level captions are generated as it describes the content of the video with 7 to 10 sentences successively. The videos are illustrated in a nutshell that truly meets the need of the model as saves time than spending more watching and understanding videos. The process of providing captions in this model consists of two vital parts Feature extraction and classification. The essential features are extracted such as Gray Level Co-occurrence Matrix (GLCM) features, Hu moments, and statistical features to detect and classify the frames. Two foremost and appropriate classifiers which work better in classifying images and videos are employed in this model such as Support Vector Machine (SVM) and Naive Bayes. Preprocessing of video frames is performed for the enrichment of informative frames to generate appropriate captions for the videos. Fig. 1 shows some of the video frames from the MSVD dataset.

Fig. 1. Frames of different Categories of MSVD Dataset. (A). A Man Sits with Two Dogs, (B). A Girl Walks on the Parade, (C). A Car on the Road, (D). A Man is Talking with Children and an Old Lady, (E). A Man is Driving a Vehicle, (F) A Dog is Sitting Near a Bag.

A. Feature Extraction

1) GLCM Features: Gray Level Co-occurrence Matrix (GLCM) shows the various combinations of pixels based on the brightness values as gray level values in an image. GLCM also has another name "Gray Tone Spatial Dependency Matrix". GLCM matrix is performed based on the orders of the texture calculations. Normalize GLCM is needed to get the value one as the sum of its elements. An element in GLCM

after normalization defines the probability of pair of pixels which explores the gray values of spatial relationship in the image. The following steps are involved to create a Normalized GLCM Matrix. Fig. 2 shows the extracted features for the further classification.

a) Arrange and quantize the specific parameters (make the intensities of the pixels arranged in a needed number of gray levels) in an image data.

b) Create the square matrix of GLCM with specific order N×N, where N denotes the Number of Levels.

c) A symmetric matrix is introduced here with GLCM Matrix.

d) Dividing each element to normalize GLCM Matrix by the sum of all elements.
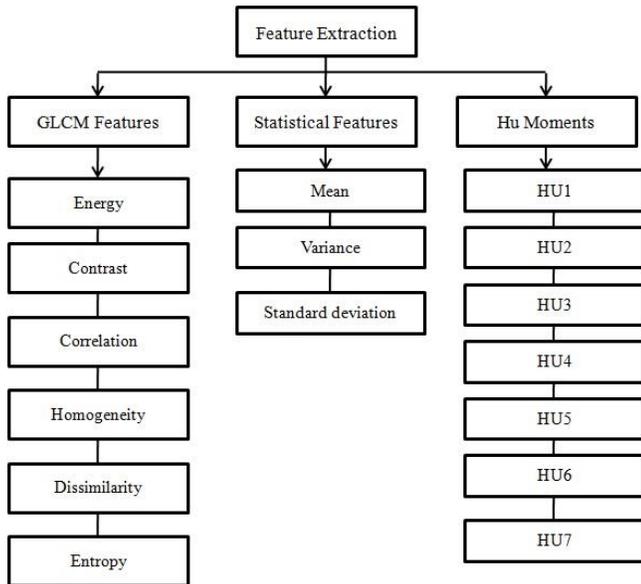
Fig. 2.   Features Extracted for Classification.

Here, six different GLCM features are extracted for the betterment of predicting exact captions for the input videos such as Energy (Angular second moment), Contrast, Correlation, Dissimilarity, Entropy, and Homogeneity.

2) Energy: It is the sum value of the squared elements in the matrix of the occurrence of gray. It is also named Angular Second Moment (ASM) and Uniformity. It ranges the values from zero to one. The constant image has one as the energy value. Eq. (1) explains the calculation of the energy.

$$Energy = \sum_{x,y=0}^{N-1} R(x,y)^2 \qquad (1)$$

3) Contrast: It mainly considers the intensity between the pixels that existed in the entire image. Here frames are considered images. So, the intensity between one pixel and the nearby pixel is calculated to get the contrast value. It is also named variance and inertia. The intensity contrast is measured for the entire pixels with the neighboring pixel of an image. The contrast value ranges from zero as per the constant image has the value zero as its contrast, the range of contrast value is

from zero to (size (GLCM, 1) -1) two. Contrast is calculated using Eq. (2).

$$Contrast = \sum_x \sum_y |x-y|^2 R(x,y) \qquad (2)$$

4) Correlation: Correlation is the estimation of how each pixel is related to the other pixel. It calculates the relation between the neighborhood pixels in the entire image. It values one for the positivity-related pixel and -1 for the negatively related pixel of an image and the value zero is assigned for a constant image. Correlation ranges the values between -1 and +1. By using Eq. (3), correlation is calculated.

$$Correlation = \sum_x \sum_y \frac{((x-\mu_x)(y-\mu_y)R(x,y))}{\sigma_x \sigma_y} \qquad (3)$$

5) Homogeneity: Homogeneity is the evaluation of similarity and nearness of the distribution of various elements in the Gray Level Co-occurrence matrix to its diagonals. It calculates the similarity between pixels of an image. It compares a pixel with the neighbor one. It ranges the values between zero and one. The diagonal of GLCM has one as the homogeneity value. Homogeneity is calculated between two pixels are calculated by using the Eq. (4).

$$Homogeneity = \sum_x \sum_y \frac{1}{1+|x-y|^2} R(x,y) \qquad (4)$$

6) Entropy: It is the average image information that calculates the amount of randomness in an image. It is measured based on the position of the pixel in a region (x, y). The Entropy of an image is computed by measuring the Entropy of the pixel values inside the two-dimensional region centered at each pixel position (x, y). By using equation (6), entropy is calculated. Table I shows the values of various GLCM features for sample frames. By using Eq. (5), entropy is calculated.

$$Entropy = \sum_x \sum_y R(x,y) \log R(x,y) \qquad (5)$$

7) Dissimilarity: It concentrated on the interesting region of the frame. It requires calculating the mean absolute difference and distance between two pixels of an image. It is the calculation of local intensity variation between two pixels which is nearby one another. The larger value means the higher differences; meanwhile smaller value resembles less difference between two pixels of the region. Dissimilarity is calculated by using the Eq. (6).

$$Dissimilarity = \sum_x \sum_y |x-y| R(x,y) \qquad (6)$$

TABLE I.   THE GLCM VALUES FOR SAMPLE FRAMES

Frames	Contrast	Correlation	Energy	Homogeneity	Entropy	Dissimilarity
1	3.59	9.74	4.21	7.89	1.56	1.10
2	5.79	8.05	4.18	8.24	2.87	2.81
3	2.35	9.21	3.49	9.17	1.17	1.35
4	8.14	8.03	7.59	9.13	6.62	8.14
5	4.73	8.57	3.16	8.28	1.59	3.22

8) Hu Moments: Hu moments are specially for shape matching. Hu moments are the essential features to detect the object from an image or frame which is existed in any direction or orientation. It makes the model identify the object or anything that existed despite any direction. Image moments are the weighted sum of the pixel of intensities in an image. These moments take into account the fundamental measures of objects such as area, centroid, orientation, and other needed properties. The central moment of an image gives the information that is constant to translation in an image. It needs to be invariant to translation, rotation, and scale. Hu moments are able to get seven values despite any transformation of an image based on the central moments. Central moments are calculated by the equation (7).

$$\mu_{xy} = \sum_i \sum_y (i - \bar{i})^x (j - \bar{j})^y I(i,j) \tag{7}$$

Here comes, Hu moments that characterize the existence in a frame. It detects the object through its shape. The strong point of Hu moments is stated that it traces the outline or edges of an object by using the shape feature vector to detect its shape. The difference between two different shapes of the object is measured by using the similarity metric. In seven moments, six moments are computed depending on the constant measures of scale, translation, reflection, and rotation. The seventh moment measures the changes in the image reflection. The seven Hu moments are calculated by the equations (8 - 14), Table II shows the values of Hu moments features for sample frames.

$$Hu_1 = H_{02} + H_{20} \tag{8}$$

$$Hu_2 = (H_{20} - H_{02})^2 + 4(H_{11})^2 \tag{9}$$

$$Hu_3 = (H_{30} - 3H_{12})^2 + 3(H_{03} - 3H_{21})^2 \tag{10}$$

$$Hu_4 = (H_{03} + H_{21})^2 + (H_{30} + H_{12})^2 \tag{11}$$

$$Hu_5 = (H_{30} - 3H_{12})(H_{30} + H_{12})[(H_{30} + H_{12})^2 - 3(H_{03} + H_{21})^2] + (3H_{21} - H_{03})(H_{03} + H_{21})[3(H_{30} + H_{12})^2 - (H_{03} + H_{21})^2] \tag{12}$$

$$Hu_6 = (H_{20} - H_{02})[(H_{30} + H_{12})^2 - 7(H_{03} + H_{21})^2] + 4H_{11}(H_{30} + H_{12})(H_{03} + H_{21}) \tag{13}$$

$$Hu_7 = (3H_{21} - H_{03})(H_{30} + H_{12})[(H_{30} + H_{12})^2 - 3(H_{03} - H_{21})^2] + (H_{30} - 3H_{12})(H_{03} + H_{21})[3(H_{30} + H_{12})^2 - (H_{03} + H_{21})^2] \tag{14}$$

9) Statistical features: Statistical features are one of the vital features which are needed to extract the content of an image for classification. It classifies based on the orders. Some of the most utilized statistical features by existing researchers are mean, median, standard deviation and variance, etc. A statistical feature depends on the texture of an image or frame. Intensity distribution information of an image is provided as the textual feature. The probability of intensity level distribution in histograms is utilized for the statistical feature calculation. Mean, Variance and standard deviation are the statistical features that are extracted from the input video frames for further purposes. Mean is the measure of the

average intensity value in an image that represents the brightness of an image based on the results of the mean calculation. The image is bright if it has a high mean value otherwise the image will have a low mean value. Variance is the measure of the difference between every pixel point and the mean value, and it is estimated by determining the difference between each pixel point and the mean value and squaring the differences and then dividing the square sums by the data points. All disparities from the mean in all directions are calculated by variance. Standard deviation is the measure of the contrast of intensity values in gray. The high contrast has resulted in a high value and the low contrast has resulted in a low standard deviation value. Table III shows the values of statistical features for sample frames.

TABLE II.   THE HU MOMENT VALUES FOR SAMPLE FRAMES

Frames	HU1	HU2	HU3	HU4	HU5	HU6	HU7
1	1.25	3.54	2.02	5.00	-6.13	-6.78	2.13
2	2.75	2.03	1.85	6.97	1.69	3.52	-2.63
3	2.14	1.42	1.72	2.03	-2.14	6.55	3.04
4	6.97	1.96	5.98	2.14	-6.12	1.23	1.17
5	3.72	5.89	1.53	3.56	1.96	4.56	-1.06

TABLE III.   THE STATISTICAL VALUES FOR SAMPLE FRAMES

Frames	Mean	Variance	Standard deviation
1	1.00	2.15	4.85
2	6.15	1.56	4.02
3	1.11	3.58	5.99
4	2.44	0.83	0.83
5	9.47	1.74	4.01
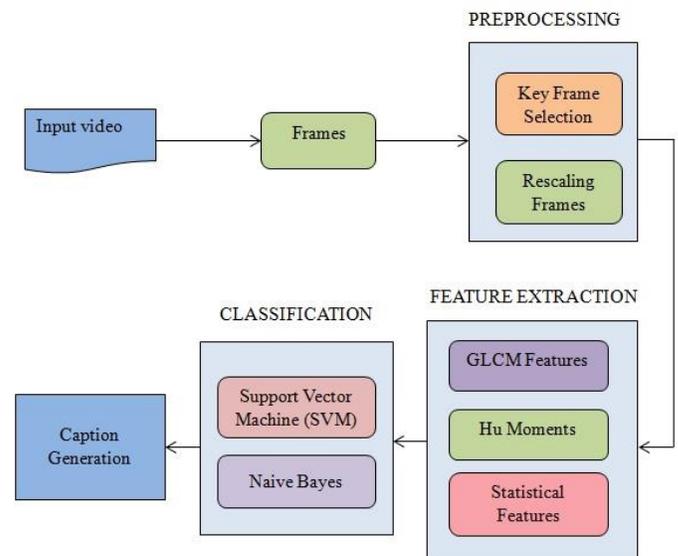
Fig. 3.   Overall Architecture of the Model.

B. Classification

1) Support vector machine: Support Vector Machine (SVM) is the standard Machine learning classifier that is abundantly used for many image-related tasks which also gives high accuracy for the most classification tasks. It is a type of supervised model that splits data as training and testing. The training part is taken to learn the logic of the frames and the testing part is utilized to check whether the learned phase matches the actual frame. It works better in high-dimensional spaces. Fig. 3 explains the overall architecture of the model proposed.

2) Naïve Bayes: Naive Bayes is the most utilized multi-class classifier that can decide in less time compared to other classifiers. Probability of the existing object influences more in classifying video frames. It is the probabilistic classifier that works better in the high dimensional space. It classifies the input video frames into various categories which are categorized based on the actions. Finally, the input video is classified with the particular appropriate captions.

C. Preprocessing

Preprocessing is the foremost step that simplifies the process by reducing the computational cost of processing the entire frames of a single video. It makes the model simpler and more comfortable for the further process of generating captions. Videos are converted into various numbers of frames. Fig. 4 shows the conversion of video into frames.
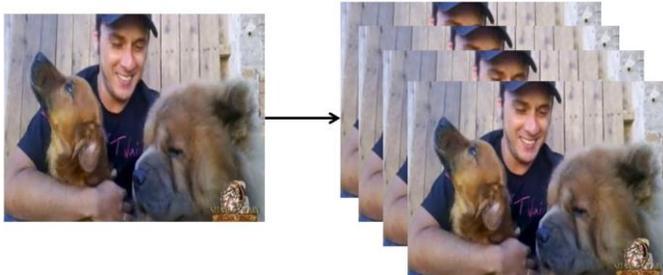
Fig. 4. Conversion of Video Into Frames.

Videos are collected from the MSVD dataset with an average duration of 10 to 25 seconds. The numbers of frames are varied for every video based on the events happening and the duration of the video. The sizes of the frames are different for extracting features. It can reduce the accuracy of predicting sentences for a particular video. To overcome the issue, the Rescaling process is acquired by changing the size of every frame of input videos to a constant size of 200×200.
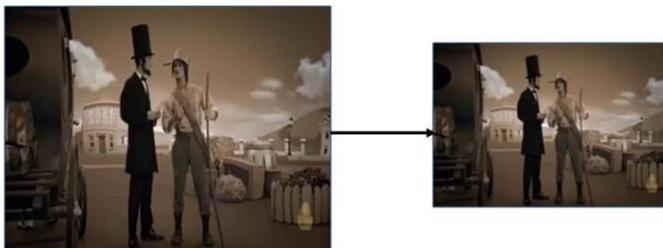
Fig. 5. Rescaling of Video Frames Into Standard Size.

The size and also the number of frames are made constant. The 50 constant key frames are extracted from every video frame. This step boosted the model by speeding up the learning process. Rescaling of video frames into standard size of 200x200 is shown in Fig. 5.

IV. DATASET

The Microsoft Research Video Corpus (MSVD) dataset is the benchmark video dataset with multi-lingual captions which attracts the researchers to demonstrate their attempts at video-related tasks. MSVD Dataset is collected by AMT (Amazon Mechanical Turk) workers during the summer of 2010. It is a collection of 1970 short random YouTube video clips of a duration of 10 to 25 seconds. Each video consists of seven to ten ground truth sentences that describe the contents of the videos simultaneously. In the sum of 20000 parallel descriptions are collected for video clips with 16000 exclusive vocabularies. The data has the split as per the ratio of 80:20 for training and testing.

V. PERFORMANCE COMPARISON

The performance of two standard techniques such as Naive Bayes and Support vector machine are evaluated using the metrics Accuracy, Precision, Recall, and F1 – score. Table IV shows the evaluation of Naive Bayes with performance metrics. Accuracy for Naive Bayes is 76.99 %. Prediction, Recall, and F1–score are scored as 72.52%, 69.89%, and 68.13% respectively. Fig. 6 shows the performances of the Naive Bayes classifier.

Table V shows the performance evaluation of the Support vector machine classifier with 80.12% accuracy, 73.85% for precision, 71.56% for recall, and 70.95 F1–score. Fig. 7 is the chart that represents the performance evaluation of the Support vector machine.

TABLE IV. PERFORMANCE EVALUATION OF NAIVE BAYES

Performance Metrics	Values (%)
Accuracy	76.99
Precision	72.52
Recall	69.89
F1-Score	68.13
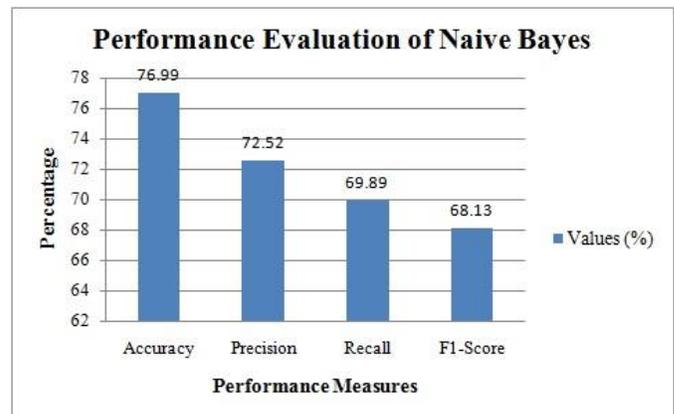
Fig. 6. Performance Evaluation of Naive Bayes.

TABLE V.   PERFORMANCE EVALUATION OF SUPPORT VECTOR MACHINE

Performance Metrics	Values (%)
Accuracy	80.12
Precision	73.85
Recall	71.56
F1-Score	70.95
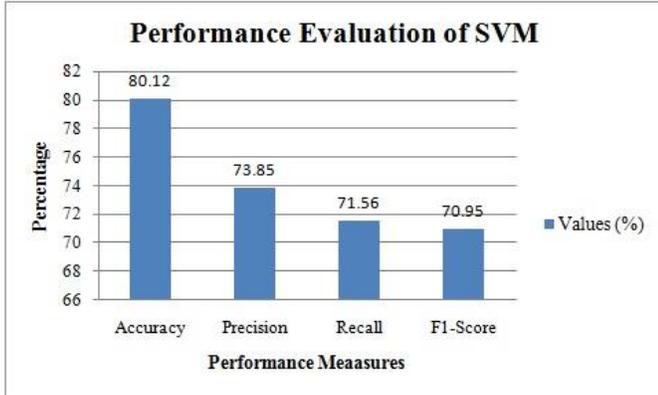
Fig. 7.   Performance Evaluation of Support Vector Machine.
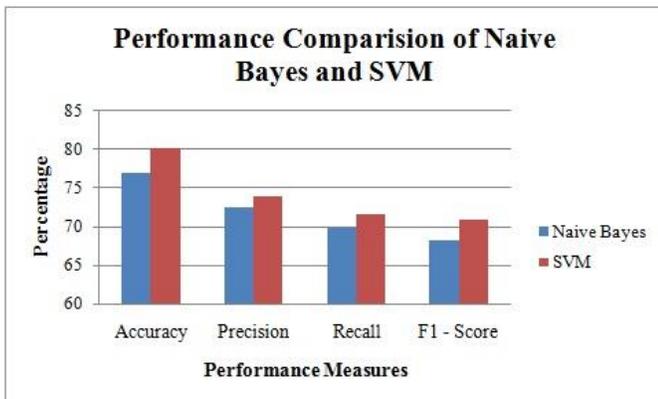
Fig. 8.   Comparative Study of the Performances of Naive Bayes and Support Vector Machine.

a girl applied eye makeup to a guy   a woman is applying makeup

Fig. 9.   The Result of Two Models. Fig (a). The Result of Naive Bayes, Fig (b). The Result of the Support Vector Machine.

The comparison of the performances of two classifiers such as the Support vector machine and Naive Bayes is clearly shown in Fig. 8. The chart depicts the performances of two classifiers which shows that the support vector machine outperforms Naive Bayes classifier. The results of the two models are shown in Fig. 9. The result of SVM (i.e. (A)) describes the video more precisely than Naive Bayes (i.e. (B)).

VI. CONCLUSION

Video Captioning is a fascinating task that made many researchers make an impact in the field in a unique way. This attempt is to show the different approaches to caption videos with machine learning techniques. The proposed architecture is designed based on classifying video frames with labels that are categorized into different categories and generating Multi-Level captions that explain the happenings of the videos consecutively. The model is structured with two major parts, Feature Extraction and Classification. The feature extraction part involves extracting the essential features to provide exact results such as some Gray Level Co-occurrence Matrix features, Hu moments, and statistical features. The classification process has experimented with two standard classifiers separately which show the comparative performance for classifying videos. The comparative results show that the Support Vector Machine (SVM) overperforms the Naive Bayes classifier. It also shows that the SVM is the appropriate classifier for classifying frames by using labels for other video-related tasks. It also provides the route to some other interesting video-related tasks by labels and tags. The proposed model is examined with the MSVD benchmark dataset and the classification performance is evaluated using Accuracy, Precision, Recall, and F1-Score.

REFERENCES

[1]   He, X. Deng, "Deep Learning for Image-to-Text Generation: A Technical Overview," IEEE Signal Processing Magazine, 34(6), 109–116, 2017.

[2]   N. Aafaq, S. Main, W. Liu, S. Z. Gilani, M.Shah, "Video description: A survey of Methods, Datasets, and Evaluation Metrics," ACM Computing Surveys, Vol. 52, no. 6, 2019.

[3]   S. Guadarrama, N. Krishnamoorthy, G. Malkarnenkar, S. Venugopalan, R. Mooney, T. Darrell, K. Saenko, "YouTube2Text: Recognizing and Describing Arbitrary Activities Using Semantic Hierarchies and Zero-Shot Recognition," IEEE International Conference on Computer Vision (ICCV), pp. 2712-2719, 2013.

[4]   N. Krishnamoorthy, G. Malkarnenkar, R. Mooney, K. Saenko, S. Guadarrama, "Generating Natural-Language Video Descriptions Using Text-Mined Knowledge," AAAI Conference on Artificial Intelligence, 2013.

[5]   M. Rohrbach, W. Qiu, I. Titov, S. Thater, M. Pinkal, B. Schiele, "Translating Video Content to Natural Language Descriptions," IEEE International Conference on Computer Vision (ICCV), pp. 433-440, 2013.

[6]   H. Pirsiavash, D. Ramanan, "Parsing videos of actions with segmental grammars," IEEE Conference on Computer Vision and Pattern Recognition, pp. 612– 619, 2014.

[7]   L. Zhou, C. Xu, J. J. Corso, "Towards automatic learning of procedures from web instructional videos," AAAI, 2018.

[8]   A. Gaidon, Z. Harchaoui, C. Schmid, "Temporal localization of actions with actoms," IEEE transactions on pattern analysis and machine intelligence, 35(11), 2782–2795, 2013.

[9]   G. Gkioxari, J. Malik, "Finding action tubes," IEEE Conference on Computer Vision and Pattern Recognition, pp. 759–768, 2015.

[10] L. Wang, Y. Qiao, X. Tang, "Video action detection with relational dynamic-poselets," European Conference on Computer Vision, pp. 565–580, Springer, 2014.

[11] M. Gygli, H. Grabner, L. Van Gool, "Video summarization by learning sub modular mixtures of objectives," IEEE Conference on Computer Vision and Pattern Recognition, pp. 3090–3098, 2015.

[12] H. Yang, B. Wang, S. Lin, D. Wipf, M. Guo, B. Guo, "Unsupervised extraction of video highlights via robust recurrent auto-encoders," IEEE International Conference on Computer Vision, pp. 4633–4641, 2015.

[13] T. Yao, T. Mei, Y. Rui, "Highlight detection with pairwise deep ranking for first- person video summarization," IEEE Conference on Computer Vision and Pattern Recognition, pp. 982–990, 2016.

[14] M. Otani, Y. Nakashima, E. Rahtu, J. Heikkila, N. Yokoya, "Learning joint representations of videos and sentences with web image search," European Conference on Computer Vision, pp. 651–667, Springer, 2016.

[15] R. Xu, C. Xiong, W. Chen, JJ. Corso, "Jointly modeling deep video and compositional text to bridge vision and language in a unified framework," AAAI, vol. 5, page 6, 2015.

[16] R. Hong, D. Liu, X. Mo, X. He, H. Zhang, "Learning to compose and reason with language tree structures for visual grounding," T-PAMI, 2019.

[17] J. Hou, X. Wu, W. Zhao, J. Luo, Y. Jia, "Joint syntax representation learning and visual cue translation for video captioning," IEEE International Conference on Computer Vision (ICCV), 2019.

[18] B. Wang, L. Ma, W. Zhang, W. Jiang, J. Wang, W. Liu, "Controllable video captioning with pos sequence guidance based on gated fusion network," IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[19] X. Yang, H. Zhang, J. Cai, "Learning to collocate neural modules for image captioning," IEEE International Conference on Computer Vision (ICCV), 2019.

[20] J. Chen, Y. Pan, Y. Li, T. Yao, H. Chao, T. Mei, "Temporal deformable convolutional encoder- decoder networks for video captioning," AAAI Conference on Artificial Intelligence, pp. 8167–8174, 2019.

[21] J. Yu, J. Li, Z. Yu, Q. Huang, "Multimodal transformer with multi-view visual representation for image captioning," IEEE Transaction Circuits Systems Video Technology, 4467– 4680, 30, 2020.

[22] Z. Yu, Y. Song, J. Yu, M. Wang, Q. Huang, "Intra and inter modal multilinear pooling with multitask learning for video grounding," Neural Processing Letter, 1–17, 2020.

[23] V. Iashin, E. Rahtu, "A better use of audio-visual cues: dense video captioning with bi- modal transformer," British Machine Vision Conference,2020.

[24] V. Iashin, E. Rahtu, "Multi-modal dense video captioning," Computer Vision and Pattern Recognition Workshops, pp. 958–959, 2020.

[25] J. Lei, L. Wang, Y. Shen, D. Yu, TL. Berg, M. Bansal, "Mart: memory-augmented recurrent transformer for coherent video paragraph captioning," ACL, 2020.

[26] A. Sasithradevi, S. Mohamed, R. Mansoor, "A new pyramidal opponent color-shape model based video shot boundary detection," Journal of Computer Visual Communication & Image Represention, vol. 67, 102754, 2020.

[27] W. Pei, J. Zhang, X. Wang, L. Ke, X. Shen, YW. Tai, "Memory-Attended Recurrent Neural Network for Video Captioning," 2020.

[28] G. Tan, D. Liu, M. Wang, ZJ. Zha, "Learning to Discretely Compose Reasoning Module Networks for Video Captioning," 2020.

[29] Y. Zhou, H. nanjia, "Accelerated masked transformer for dense video captioning," Elsevier, Neuro computing, 2021.

[30] M. Hemalatha, C. Chandra Sekhar, "Domain-specific semantics guided approach to video captioning," IEEE, 978-1-7281-6553-0, 2020.

[31] J. Vaishnavi, Dr. V. Narmatha, "Video Captioning based on Image Captioning as Subsidiary Content", International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), IEEE, 2022.

Students' Perspective on Sustaining Education and Promoting Humanising Education through e-Learning

A Qualitative Focus Group Study

Aidrina Sofiadin
Department of Information Systems
International Islamic University Malaysia, Kuala Lumpur, Malaysia

Abstract—The COVID-19 pandemic has shifted the education sector towards an e-learning approach to sustain education. Sustaining student education through e-learning significantly impacts student learning experiences and outcomes, which can be influenced by e-learning infrastructure, e-learning materials, and learning engagement. The concept of humanising education refers to the learning process that reflects students' moralities and values. Focus group discussion is a practical approach to assess students' perspectives on sustaining and humanising their education through e-learning. This qualitative focus group study aimed to discuss the e-learning factors that will sustain education and promote humanising education in the virtual learning environment. Thirty students from Information Technology (IT) and business fields have participated and provided a different view on sustaining education through e-learning. Thematic analysis was used to analyse the focus group data. Five themes were identified: (1) e-learning technologies and infrastructure; (2) e-learning principles: pedagogy and materials; (3) health and wellness; (4) equality; and (5) engagement: communication and collaboration. The analysis enlightened the e-learning design for sustainable e-learning. In addition, this paper outlines how the findings support sustainable development goals.

Keywords—Education; sustainability; humanizing education; sustainable e-learning

I. Introduction

The COVID-19 pandemic has shifted students learning into the online environment to sustain education. Despite of this, the pandemic has exposed education inequalities among students due to lack of access to devices and Internet connection. Consequently, the pandemic has affected the world 2030 agenda for sustainable development in 2015 that intend to improve public health, education, and economic growth while reducing inequality and poverty [1].

In 2021, the Sustainable Development Goals (SDG) report indicated that most of the progression of the goals had been affected due to the pandemic [2]. This is not limited to education only, but also caused lingering threat to food systems, public health, global workforce, and economy [2,3]. Continuous threats were later reported in the SDG report 2022 [4]. Moreover, the pandemic has prolonged health effect among everyone around the globe due to lung and heart scarring. Since many have lost their loved ones due to the COVID-19 pandemic, the health and well-being among the society are worrisome. E-learning has become the main drive to support student's education as it is believed to reduce the

transmission of COVID-19. To sustain education through e-learning, SDG [1] on (1) SDG#3: Good Health and Well-being; (2) SDG#4: Quality education; (3) SDG#9: Industry, Innovation, and Infrastructure; and (4) SDG#10: Reduced inequalities need to be considered.

Even though e-learning helps students to continue their studies throughout the pandemic, there is an alarming increase in mental illness among students. Since the COVID-19 outbreak, a high percentage of depression and anxiety among the Malaysian population aged above 18 years was reported in 2020 [5]. This finding was supported by [6], which indicates the high percentage of depression, anxiety, and stress among higher education students in Malaysia. The research aims to discover students' perspectives on sustaining their education and promoting humanising education through online learning environment. The research findings indicated that the five themes provide insights on designing online education that cater the student's learning ability as a human being and a learning environment that nurture good health and well-being. The outcomes from this research will support the development on SDG#3: Good Health and Well-being and SDG#4: Quality Education through e-learning. This study is organized as follows: literature review, methods, results and discussions, and conclusion.

II. Literature Review

E-learning has been introduced for decades before the COVID-19 pandemic outbreak. Many educational institutions have adopted e-learning; however, e-learning was mainly used as an education resource repository where students can obtain learning materials. Indeed, the COVID-19 pandemic has pressured education institutions to fully migrate to online learning to sustain learners' education. Many studies were conducted on e-learning during the pandemic. However, there are challenges among students, instructors, and institutions to migrate to e-learning. Furthermore, the new norm of online education has also led to health issues such as mental health, depression, and physical health.

Since the pandemic, a huge number of studies have been conducted on students' perspectives on e-learning during the COVID-19 outbreak, see Table I. Many methods such as online survey, literature review, structural equation modelling, and qualitative phenomenological approach were used to identify the impact of e-learning on its stakeholders during the pandemic. Based on these findings, most are concerned with

how e-learning impacts the students' health, education structure, e-learning policies, and teaching and learning styles.

TABLE I.   THE LIST OF PREVIOUS STUDIES ON STUDENTS' PERSPECTIVES ON E-LEARNING DURING THE PANDEMIC

Methods	Findings
Online survey	The status of mental health among university students was greatly affected during the pandemic. Most students have a good perception of e-learning [6].
Literature review	The COVID-19 global lockdown has changed the accessibility and structure of education [7].
Questionnaire survey	Faculty members have a positive attitude toward e-learning for disabled students as e-learning promotes education interaction during the pandemic [8].
Structural equation modelling	Factors of E-learning quality during the COVID-19 pandemic are assurance, reliability, responsiveness, and content [9].
Structural equation modelling	Teachers' online instructional innovation and creativity improve students' engagement [10].
Qualitative phenomenological approach	Inconsistency and awareness in e-learning policy directions on e-learning during the COVID-19 pandemic needs to be acknowledged [11].
Survey	Usability, interaction, and quality would motivate students to use the e-learning system during the COVID-19 pandemic [12].
Systematic Literature Review (SLR)	Network infrastructure, hardware requirements, implementation complexity, and training requirements are the factors for online exam endorsement [13].
Systematic Literature Review (SLR)	Attention is needed on Information and Communication Technology (ICT)-based teaching methods for learning Islamic education. Policymakers should develop better policies for Information Technology (IT) awareness and acceptance [14].
Online survey	The impact of the COVID-19 lockdown on students' mental health due to self-isolation using e-learning [15].

The findings on Table I indicate the need for support systems that enrich student wellness. Furthermore, wellness among students in aspects of physical, social, spiritual, emotional, and intellectual health need to be neglected. A report indicates that 76% of students and 73% of staff highlighted that sustaining their well-being is the main challenge [16]. Some universities have implemented health education programs to support students' academics by promoting and maintaining a healthy and positive lifestyle [17, 18, 19]. At the same time, wellness among university staff was considered by many universities such as Curtin University [19], International Islamic University Malaysia [20], Monash University Malaysia [21], and The University of Texas of Austin [22], to name a few, which provide services and support for their staff well-being.

Table II shows the findings on students' wellness and education development during the pandemic [23, 24, 16, 25, 26]. The statistics of negative impact on students are alarming. Consequently, these issues are detaining the development of the SDG especially on promoting quality education and good health and wellbeing. Based on SDG report in 2022 [4], the global learning crisis has worsened due to severe education system interruptions caused by the COVID-19 pandemic. Many countries have invested huge amount of money to recover the education services and facilities and focus on improving the education as many education institutions are re-

open. However, the report indicates that there is a lack of psychosocial support for students in which only 20 per cent of countries provide support for student's mental health and psychosocial. Not only among students, but also among young people around the world due to limited access to mental health care service. The report suggested that it is crucial to include mental health support as part of the post pandemic plans. Therefore, this significant measure is essential for the education institutions to consider due to rapid increase of depression and anxiety issues among students.

TABLE II.   REPORTS ON WELLNESS IN HIGHER EDUCATION IN 2020

Author	Year	Report	Wellness issues
Healthy Minds Network and the American College Health Association	2020	The Impact of COVID-19 on College Student Well-being.	66% report on financial stress and mental stress.
Eisenberg, Lipson, Heinze, Zhou, Talaski and Patterson	2020	The Healthy Minds Study: Fall 2020 Data Report	79% depression, 68% anxiety, 11% eating disorder, 42% feel loneliness in some of the time, and 13% on suicidality ideation.
Salesforce.org	2020	Connected Student Report: Insights into Global Higher Education Trends from over 2,000 Students and Staff.	Financial anxiety and mental health among students
Baldwin, Towler, Oliver II and Datta	2017	An examination of college student wellness	Emotional wellness (stress issue) is the top challenge, followed by Physical wellness (health behavior)
Curtin University	2016	Curtin Wellbeing Surveys	28.8% depression or anxiety 27.5% suicidal thoughts

Thus, to reduce the gaps between mental performance and mental health, there is a need of a support system that tackles both education and wellness to achieve sustainable education. In addition, the introduction of humanizing education concept can become the inspiration to improve students learning and health. Since the use of e-learning continues along with traditional learning, e-learning may have the potential to promote humanize education while improving students' well-being.

III. METHODS

This study aims to investigate how e-learning can sustain education and promote humanized education through focus groups discussion to assess students' perspectives on sustaining their education and promoting humanising education through e-learning. This study used a qualitative research approach to acquire student perspectives. Qualitative methods [27] have been widely used to identify students' needs and perspectives. Thirty students were selected and participated in the focus group discussion. The participants are from various areas of study such as Law, Business, Education, Health Science, and Engineering. The participants from different disciplinary were selected to obtain a different view on sustaining education through e-learning. The participants

were divided into six groups, where each group consisted of five members. The allocation of the group members was done randomly. The participants' demographic data were collected, and descriptive statistics were used to describe the participants' demographic data. However, this paper will focus on the qualitative data obtained from the focus group. Qualitative method was adopted due advised by Hasselblad and Olsson [28] on their publication that qualitative data is necessary to access user opinions on using technology. Thus, to analyse the qualitative data, qualitative coding was applied to increase the analysis validity, fair contribution of opinions among the participants, and transparency on reviewing data analysis.

Semi-structured interviews were implemented since it enables flexibility in achieving the research questions and explore the participants' perspectives [29] on e-learning potential in sustaining education and promoting humanizing education. Table III presents a set of open-ended questions for a semi-structured interview that was developed to facilitate the focus group discussions. Discussions were conducted online via Microsoft Teams and Nearpod for an online collaboration board for 45 minutes. The focus group started with a general question on the use of e-learning during the pandemic and later to a more focused one, covering e-learning accessibility, materials, activities, and impact. During the focus group discussions, participants' opinions were based on their own words and thoughts. The participants were informed that the data from the focus group discussions would be used for research and publication.

Thematic analysis was used to analyse the qualitative data to report themes. This analysis was used due to its practical data analysis approach for the qualitative method [4,30]. Recorded online focus group discussions were transcribed, and data from the Nearpod was generated. Along the focus group discussion, mind map was produced for each interview session to track the topics and sub-topics that were highlighted. These mind maps which represent the themes and highlighted issues, later assisted the data analysis. The use of this mind maps practice in qualitative data analysis was supported by Fearnley [31] that indicates the use of mind map in qualitative data analysis, will provide transparent and broader view of contents discussed in interviews.

TABLE III.   THE OPEN-ENDED QUESTIONS WERE PREPARED PRIOR TO THE FOCUS GROUP DISCUSSIONS

No.	Main questions
1.	What are your perspectives on learning in an online learning environment?
2.	What were your expectations of e-learning and the use of mixed reality?
3.	Did you aware of e-learning limitations?
4.	How can e-learning help you to sustain your education?
5.	What are your perspectives on humanising education?
6.	How can e-learning help you to experience humanistic education?
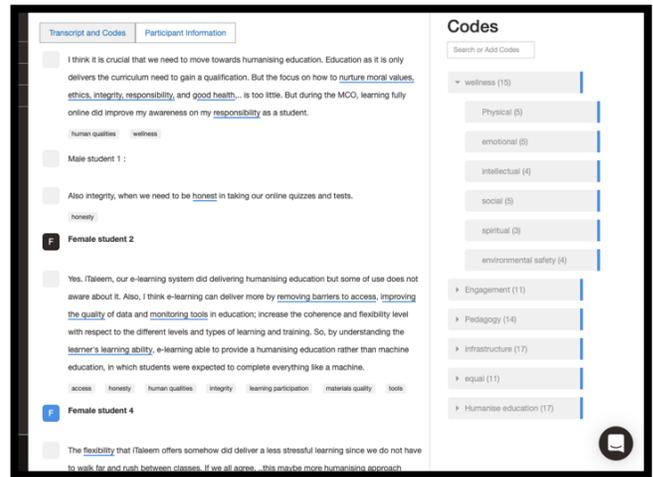
Fig. 1.   Screenshot of Qualitative Coding Process using Delve Qualitative Analysis Tool.

The transcripts were uploaded to Delve qualitative analysis tool and codes were identified. Fig. 1 shown above is an example of one of the transcripts that was analyzed using Delve. A combination of deductive and inductive coding approaches was used where the data analysis started with the inductive coding with codes were developed from the themes categories and the data. To increase the data validity and reduce potential biases, the second round of analysis was conducted using deductive coding in which the analysis start using the codes generated in the first round of analysis. The transcriptions were read again to assign excerpts to codes. Then, inductive coding applied with new codes. These new codes were later categorized into the themes.

IV. RESULTS AND DISCUSSION

Results were taken from the six focus groups. There were 30 participants between the ages of 20 to 25 years. Among the 30 participants, 13 were males, and 17 were female students. They are from five areas of study: Law, Business, Education, Health Science, and Engineering. Based on the focus group discussions, this research found that many students prefer online education as it offers more flexibility, time-consumption, interactive activities, and e-learning materials.

The participants' satisfaction level is high on the flexibility offered by e-learning since students can access their learning activities anytime, anywhere. Participants believe that e-learning reduces their transportation time to a physical class. E-learning allows students instantly access their online class without spending time driving or walking to class. In comparison with learning activities in class, e-learning activities are more attractive. The use of an online collaboration board, forum, and quiz, allow students to re-access the exercises again. Participants believe that they participated more in online class activities than in physical class activities. Some participants have more confidence in learning in an online environment since they cannot see people they believe will look down on them. Since e-learning provides a repository for learning materials, students have more access to attractive materials such as videos, audio on case studies, and reading materials. However, they are not

satisfied with the social environment because of their inability to feel the student life on campus.

The qualitative data analysis from the focus groups identified five main themes pertaining to sustaining education through e-learning: 1) e-learning technologies and infrastructure, (2) e-learning principles, (3) health and wellness, (4) equality, and (5) engagement. Table IV outlines the main themes and sub-themes.

Main Themes	Sub-themes
E-learning technologies and infrastructure	Network and connectivity, tools, hardware, software, service, immersive technologies, and emerging technologies.
E-learning principles	Pedagogy and materials quality
Health and wellness	Physical, emotional, intellectual, social, spiritual, and environmental safety.
Equality	An inclusive learning environment, equal learning participation, accommodating of different learning styles and disabilities, and fair use of technology.
Engagement	Communication, collaboration, and social interaction

A. E-learning Technologies and Infrastructure

E-learning infrastructure is essential to ensure online learning takes place without any interruptions. The potential loss of internet connection or poor internet connection may create barriers for students to perform online learning. The second concern raised by students was the support service.

"..removing barriers to access, improving the quality of data and monitoring tools in education; increase the coherence and flexibility level with respect to the different levels and types of learning and training. So, by understanding the learner's learning ability, e-learning able to provide a humanising education rather than machine education, in which students were expected to complete everything like a machine." (Male student).

Most students from the focus groups expressed their online learning environment as an interactive way to learn. Some of the students noted that the variety type of learning materials and activities have made e-learning more engaging compared to face-to-face learning. Even though some of them confessed that e-learning have lack of communication and social engagement, a few of the students disagreed and remarked that e-learning provide wider social connectivity.

"I communicate with more people via e-learning and I am more engaged with my group assignment discussion by using e-learning chat…. I think my connectivity will be limited if I only meet people in my classes." (Female student).

A few students pointed out that the use of immersive technology could extend students' learning experience and discover a real-world situation beyond the classroom. This could enrich human sense and quality. Furthermore, VR is believed to provide higher learning opportunity on international exposure due to VR ability to simulate real-life learning environment that are beyond the classroom. This finding tally with Ignatius [32] perspective on how digitalisation should be integrated together with basic human skills and values to re-humanising education.

"Mixed reality can provide more learning exposure on international business practice and technology development and implementation." (Female student).

"The use of virtual reality or augmented reality can increase learning outcome since it allows students to do some tasks beyond the classroom." (Male student).

Interestingly, a few students proposed AI as a coach in e-learning that will provide advice on students' learning progress and recommend healthy learning styles.

"I wish e-learning can help me to advise me to become a better person. It can teach or guide me to become ethical person by give me some information what good or bad practice are. Maybe the use of AI technology in e-learning will help me to increase my human values." (Female student).

Students wanted e-learning as a system that would take care of their education and health at the same time. They believe e-learning should be a system where they can continuously learn even after graduation. E-learning should be a technology that allows them to learn how to improve their knowledge, skills, moral values, and health. A student suggested that the use of AI may assist on learning to be an ethical person rather than just academic matters. Students also pointed out that the e-learning tools should be easy to use and do not require a lot of internet data, which may lead to crashes during online learning activities. Further advantages of VR and AI will be discussed specifically in the following themes.

B. E-learning Principles

Overall, the course syllabus, assessments, and learning outcomes are the same when conducting an online or physical approach. However, the evaluation methods, such as instructions and rules, need to be communicated to the students. A proper lesson plan needs to be shared with the students. Interestingly, a student pointed out that students should prepare their own study plan to organise their studies between online classes. The e-learning materials should be in good quality, attractive and easy to access. Nimehchisalem [33] also discovered that education could be humanised by ensuring the classroom materials were developed with both a human approach and a discovery approach, in which the curriculum can cater to a learner's needs.

"Make sure the content in e-learning is more attractive and easier to access to all students. It is also important to consist of ethical values to ensure students are aware of ethical practice in business and developing technology." (Male student).

"… the project or assignment should not be based on the textbooks or any other reading material. Our learning activities or exercises should promote creative thinking towards a better world. I think if we aim for better world that create harmony relationship between people and environment, then maybe that what it means by humanizing education……. so, e-learning can deliver this by provide some online learning interactivity that promote creative thinking towards making better future and provide some case study on climate change

crisis so that we are aware of the climate change impact in our future." (Female student).

Some participants were worried about the massive number of materials available in e-learning as it consumed more time to study than the indicated student learning time (SLT) in the course outline. Thus, this leads to health issues that lead to stress, anxiety, fatigue, headache, and eye strain. Students believe that awareness of ethics and values is important for them to learn in humanising education. In direction of humanizing education through e-learning, a few students suggested that e-learning content should correspond to promoting moral values and sustainable practices. Awareness and knowledge on best practices, ethics, morality, and values need to be delivered through e-learning activities and materials as a guide toward humanized future leaders.

"…to promote humanizing education, we should not look at education is a way to get more money in the future. It should be look as a way to make our planet and environment better. For example, most people think that business and entrepreneurship are about making a lot of money and be rich. It should create thinking on how business will help our economy become more stable and provide better service and product to our community to improve life and our environment. So, I think e-learning can provide some awareness on this since it can deliver a variety of digital information to the student.".

This may increase the quality of education for all that supports the fourth SDG. The finding also indicates that moral values need to be added as part of education quality criteria, which is supporting Balakrishanan [34], Juhary [35], and Khilji [36] statements.

C. Health and Wellness

Due to online learning during the COVID-19 lockdown, mental health has become a significant concern. Students had substantial concerns about their health in terms of physical and mental. The tremendous amount of time spent online learning has led to physical health problems such as eye strain, back pain, and headaches. The vast amount of learning tasks has caused mental health issues such as stress, depression, and anxiety.

"Be more concern with the student's mental and physical health and help them to keep motivated to sustain the quality of the education. Lectures can play a role as an online motivator." (Female student).

"e-learning can integrate the use of AR or VR to allow active learning that involve movements rather than just sitting on a chair that are causing back pains and bad body posture."

A participant pointed out that the lecturer could play a role in motivating the students through the learning process. It should be noted that the SLT should be revised for the online learning approach as it involves high contact with the computer screen and less body movement. In addition, a student suggested that e-learning should not only consist academic achievements but also the summary of achievements of soft skills and wellness criteria, that will reflect humanise

education. One student added that students need to be treated and educated as human to humanizing education.

"We are not a robot or a business product… to create humanizing education we need to be treated as human who have limitation in performing tasks and remembering things. We do not have the energy to learn the whole day. We go to class during daylight and do our homework at night."

Due to the increasing of mental illness among students [16], the findings from the focus group discussions recognized a few factors that cause a decline in the student's health, namely feeling isolated, lack of social activity, screen time, and learning materials. Thus, to ensure the students to perform better learning achievement and to support the recommendation by the United Nations in the Sustainable Development Goals 2022 Report [4], there is a need to redesign the e-learning materials, activities, and infrastructure as an alternative to provide health and psychosocial support to overcome the health and wellness issues among students.

D. Equitable

Since e-learning was used before the pandemic, some students have accessed e-learning using their personal computers or university computers. Equal access to e-learning is important for students to be able to access all the e-learning materials without any problem. For instance, as pointed out by a participant:

"Looking at the pandemic impact on our education, e-learning has saved us in sustaining our education…..I also believe e-learning provide equality education to both female and male, or even young and old learners." (Male student).

"Basic education like learn to use software and hardware need to be free and accessible for everyone. Like Finland, their education is free and there is no exam or grades that make them compete."

Due to the COVID-19 lockdown and strict movement control order, some students do not have access to e-learning as they do not afford to buy their own personal computer. This is an issue for those students to sustain their education during the pandemic. Some participants suggested government funds and donation programmes provide a personal computer to those unfortunate ones to sustain their education. The post-pandemic has led to an unstable economy, affecting many educational budgets. Thus, to sustain education on international exposure, the use of virtual reality could provide an opportunity for students to learn abroad virtually.

"The use of virtual reality and augmented reality can reduce education difference among students since this mixed reality will allow all students to have the opportunity to discover about other countries' development by virtually travelling there. This will reduce unfairness for some students who were unable to travel abroad to discover new things." (Female student).

E-learning provides a platform that can allow everyone to have opportunity to learn and have an enduring education throughout a lifetime. However, e-learning need to be accessible and provide free basic education to public. This will ensure that no one gets left behind regardless their location

and lifestyle status. The emerging technology such as VR could provide balance the number of students who obtain international exposure during their learning experience.

E. Engagement

An online learning environment may provide flexibility in communication and collaboration. Participants agreed that online activities such as collaboration boards, forums, and quizzes are more attractive and easier to re-access for revision. Some students highlighted that the online activities engaged their social and emotional learning as they are more confident learning online.

"The relationship between students and teachers should be established to promote learning engagement. For example, teachers should give personal feedback on online assessment rather than just computer-generated feedback." (Female student).

"Like Pokemon Go, the use of augmented reality has changed the game experience and increased gamers engagement since it is more attractive. So, I think mixed reality will promote better learning engagement since the learning experience will be beyond the classroom." (Male student).

The use of virtual and augmented reality in education was believed could sustain student learning and promote humanising education. To support this finding, the use of virtual experience needs to be reviewed together with human processes, experiences, and ethics, as suggested by Trkman and Cerne [37]. Furthermore, the use of chat in e-learning improves interactions between the students and lecturers as it provides fast feedback and response. However, this interaction may increase students' engagement if the lecturer responds promptly to the chat or instant messaging. Students were worried about their interaction more with their devices than people, and they are concerned if their learning process does not lead them towards a humanising education. Losing social interaction in campus life was the primary concern expressed by students during the focus group discussions since this may lead to an educational environment without soul. This relates to Razak [38] statement on the importance of education with soul in humanistic education. Instead, online learning during the pandemic feels like a factory production of learners. Students were given tasks without knowing their challenges learning online during the pandemic. Some students highlighted the need to be in a campus environment to enhance their social networks and improve their mental and physical health, which supports their learning engagement.

V. Conclusion

This study identified five themes, namely e-learning technologies and infrastructure, e-learning principles, health and wellness, equality, and engagement that could sustain students' education and promote humanistic education using e-learning. Furthermore, the concept of humanizing education can relate to the student's ability to sustain their education. For students to sustain their education through e-learning, the design of the online learning course should consider students' health, learning time, learning styles, engagement, and emerging technologies that can assist in delivering education

with soul. The e-learning system needs to be publicly accessible to deliver essential education to everyone.

Furthermore, a participant indicates that a sustainable education can be achieved if equal access to quality education is provided. Thus, an e-learning infrastructure should provide a good network, sufficient learning tools and a support system to ensure students can access their learning without interruption. Also, the university needs to ensure that students have a compatible device to access e-learning. The e-learning courses should be designed to meet different student's backgrounds.

The use of mixed reality is able to promote the concept of humanising education. The e-learning materials need to be attractive and interactive to sustain the education quality. In order to make education humanised, e-learning materials and content need to be designed based on students' personal learning styles. As a result, these five themes supported the SDGs, hence, sustaining education and promoting a humanising education. Overall, students in this study appear to have high acceptance and satisfaction levels in using e-learning. Also, they believe that e-learning has fulfilled their learning needs to sustain their education during the pandemic. This study suggests future studies on improvements in quality assurance on online learning to ensure balance between intellectual, emotional, spiritual, and physical aspects. Future research could focus on developing a support system that look after students' mental health and motivate their lifelong learning to ensure sustainable and humanistic education. The findings in this paper could lead to new education support systems that considers both mental and mind performance and health. Consequently, this will lead to the development of sustainable, healthy, and prosperous communities.

References

[1] United Nations, "Transforming our world: the 2030 Agenda for Sustainable Development," 2015. [Online]. Available: https://documents-dds-ny.un.org/doc/UNDOC/GEN/N15/291/89/PDF/N1529189.pdf?OpenEle ment.

[2] United Nations, "The Sustainable Development Goals Report 2021," 2021. [Online]. Available: https://unstats.un.org/sdgs/report/2021/ [Accessed July 27, 2022].

[3] A. Haleem and M. Javaid, "Effects of COVID-19 pandemic in daily life," *Elsevier Public Health Emergency Collection,* vol. 10, no.2, pp. 79-7, 2020.

[4] United Nations, "The Sustainable Development Goals Report 2022," 2022. [Online]. Available: https://unstats.un.org/sdgs/report/2022/The-Sustainable-Development-Goals-Report-2022.pdf [Accessed August 2, 2022].

[5] L.P. Wong, H. Alias, A.A. Md Fuzi, I.S. Omar, A.Mohamad Nor, M.P.Tan, et al. "Escalating progression of mental health disorders during the COVID-19 pandemic: Evidence from a nationwide survey." 2020, *PLoS One,* vol.16, no.3, 2020, [Online]. Available: .https://doi.org/10.1371/journal.pone.0248916 [Accessed July 22, 2022].

[6] F.M. Moy, and Y.H. Yit , "Perception towards E-Learning and COVID-19 on the Mental Health Status of University Students in Malaysia." *Science Progress*, vol. 104, no.3, 2021. [Online]. Available:

https://journals.sagepub.com/doi/10.1177/00368504211029812 [Accessed July 22, 2022].

[7] H. Onyeaka, C. K. Anumudu, Z. T. Al-Sharify, E. Egele-Godswill, and P. Mbaegbu. "COVID-19 Pandemic: A Review of the Global Lockdown and Its Far-Reaching Effects." *Science Progress*, 2021. [Online]. Available: https://doi.org/10.1177/00368504211019854. [Accessed July 22, 2022].

[8] A. Mohammed Ali, "E-Learning for Students With Disabilities During COVID-19: Faculty Attitude and Perception." *SAGE Open*, vol.11, no.4, 2021.[Online]. Available: https://doi.org/10.1177/21582440211054494. [Accessed July 20, 2022].

[9] C. Saxena, H. Baber, and P. Kumar. "Examining the Moderating Effect of Perceived Benefits of Maintaining Social Distance on E-Learning Quality During COVID-19 Pandemic." *Journal of Educational Technology Systems*, vol. 49, no. 4, pp. 532–54. 2021. .[Online]. Available: https://doi.org/10.1177/0047239520977798. [Accessed July 23, 2022].

[10] U. Ahmed, A. I. Ismail, M. Fati, and M.A. Akour. "E-Learning during COVID-19: Understanding the Nexus between Instructional Innovation, E-Psychological Capital, and Online Behavioural Engagement." *Management in Education*, 2021. [Online]. Available: https://doi.org/10.1177/08920206211053101. [Accessed July 22, 2022].

[11] M. A. Islam, S. Nur, and M.S. Talukder. "E-Learning in the Time of COVID-19: Lived Experiences of Three University Teachers from Two Countries." *E-Learning and Digital Media,* vol. 18, no. 6, pp. 557–80. 2021.[Online].Available:https://doi.org/10.1177/20427530211022924. [Accessed July 24, 2022].

[12] A. Altalbe, "Antecedents of Actual Usage of e-Learning System in High Education During COVID-19 Pandemic: Moderation Effect of Instructor Support," *IEEE Access*, vol. 9, pp. 93119-93136, 2021, [Online]. Available:https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=944807 1 [Accessed July 28, 2022].

[13] A. W. Muzaffar, M. Tahir, M. W. Anwar, Q. Chaudry, S. R. Mir and Y. Rasheed, "A Systematic Review of Online Exams Solutions in E-Learning: Techniques, Tools, and Global Adoption," *IEEE Xplore*, vol. 9, pp. 32689-32712, 2021, [Online]. Available: https://ieeexplore.ieee.org/document/9357335 [Accessed June 12, 2022].

[14] A. Qazi, G. Hardaker, I. S. Ahmad, M. Darwich, J. Z. Maitama and A. Dayani, "The Role of Information & Communication Technology in Elearning Environments: A Systematic Review," *IEEE Access*, vol. 9, pp. 45539-45551, 2021, [Online]. Available: https://ieeexplore.ieee.org/document/9381199. [Accessed July 28, 2022].

[15] A. Agarwal, S. Sharma, V. Kumar and M. Kaur, "Effect of E-learning on public health and environment during COVID-19 lockdown," *Big Data Mining and Analytics*, vol. 4, no. 2, pp. 104-115, 2021, [Online]. Available: https://ieeexplore.ieee.org/document/9343920. [Accessed July 28, 2022].

[16] Salesforce, "Connected Student Report: Insights into Global Higher Education Trends from over 2,000 Students and Staff." Second Edition. [Online]. Available: https://www.salesforce.org/wp-content/uploads /2021/06/connected-student-report-second-edition-06-23-21.pdf [Accessed September 5, 2022].

[17] Roger Williams University, " Health and Counseling: Health Education Program." [Online]. Available: https://www.rwu.edu/undergraduate/ student-life/health-and-counseling/health-education-program [Accessed September 8, 2022].

[18] The University of Arizona, "Health & Wellness for Students." [Online]. Available: https://www.arizona.edu/health-wellness-students [Accessed September 8, 2022].

[19] Curtin University (2022). Health and well-being. [Online]. Available: https://www.curtin.edu.au/study/campus-life/health-wellbeing/ [Accessed September 8, 2022].

[20] International Islamic University Malaysia, "Counselling and Career Service Centre." [Online]. Available: https://www.iium.edu.my/centre /ccsc [Accessed September 8, 2022].

[21] Monash University Malaysia, "Support Service :Counselling Health & Wellness." [Online]. Available: https://www.monash.edu.my/student-services/support-services/counselling-and-wellness[Accessed September 8, 2022].

[22] The University of Texas, "Campus & Community: Health & Wellness." [Online]. Available: https://www.utexas.edu/campus-life/health-and-wellness [Accessed September 8, 2022].

[23] Healthy Minds Network and the American College Health Association, "The Impact of COVID-19 on College Student Well-being." [Online]. Available:https://healthymindsnetwork.org/wp-content/uploads/2020/07/Healthy_Minds_NCHA_COVID_Survey_Report _FINAL.pdf [Accessed September 3, 2022].

[24] D. Eisenberg, S. K. Lipson, J. Heinze, S. Zhou, A. Talaski and A. Patterson, "The Healthy Minds Study: Fall 2020 Data Report." [Online]. Available: https://healthymindsnetwork.org/wp-content/uploads/2021/02/HMS-Fall-2020-National-Data-Report.pdf [Accessed September 3, 2022].

[25] D. R. Baldwin, K. Towler, M. D. Oliver II and S. Datta, "An examination of college student wellness: A research and liberal arts perspective." *Health Psychology Open.* pp. 1-9. 2017. [Online]. Available:https://journals.sagepub.com/doi/pdf/10.1177/2055102917719 563[Accessed September 4, 2022].

[26] Curtin University, "Curtin Wellbeing Surveys" 2016. [Online]. Available:https://www.facebook.com/CurtinWBS/photos/a.7943142940 79382/794314247412720/ [Accessed September 4, 2022].

[27] M.E. Kiger and L. Varpio, "Thematic analysis of qualitative data: AMEE Guide No. 131", *Medical Teacher*, vol. 42, no.8, pp. 846-854, 2020.[Online].Available:https://doi.org/10.1080/0142159X.2020.175503 0 [Accessed September 7, 2022].

[28] A. Hasselblad and L. Olsson, "The Importance of Including Qualitative Data in Technology Evaluation - Investigating the Technology Implementation Evaluation Score (TIES)," *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2020, pp. 489-493.

[29] G. McCracken, *The long interview.* Newbury Park, CA: SAGE, 1988.

[30] L.S. Nowell, J.M. Norris, D. E. White, and N. J. Moules. "Thematic Analysis: Striving to Meet the Trustworthiness Criteria." *International Journal of Qualitative Methods*, 2017. [Online]. Available:https://doi.org/10.1177/1609406917733847. [Accessed September 8, 2022].

[31] C. Fearnley, *Mind mapping in qualitative data analysis: Managing interview data in interdisciplinary and multi-sited research projects.* Geo: Geography and Environment. John Wiley & Sons Ltd, 2022. [E-book] Available: wileyonlinelibrary.com/journal/geo2

[32] C. Ignatius, " Re-humanising Education. Business Today." August 8, 2022.[Online].Available:https://www.businesstoday.com.my/2022/08/08 /re-humanising-education/ [Accessed September 16, 2022].

[33] V. Nimehchisalem, "Interview with Brian Tomlinson on Humanising Education*." International Journal of Education and Literacy Studies.* vol. 4, no.2, pp.101-106, 2016.

[34] V. Balakrishanan, " Humanising education" The Star. January 30, 2022. [Online].Available:https://www.thestar.com.my/news/education/2022/01 /30/humanising-education [Accessed September 16, 2022].

[35] J. Juhary, "Humanising Online Activities: Lessons Learned from Digitizing Work-the-Walk (WTW) Approach." *International Journal of Information and Education Technology,* vol. 12, no. 10, 2022.

[36] S. E. Khilji, "An Approach for Humanising Leadership Education: Building Learning Community & Stakeholder Engagement," *Journal of Management education,* vol.46, no. 3, pp. 439 – 471, 2022.

[37] P. Trkman, and M. Cerne, "Humanising digital life: Reducing emissions while enhancing value-adding human processes." *International Journal of Information Management.* vol. 63. P. 102443. 2022. [Online]. Available:https://www.sciencedirect.com/science/article/pii/S026840122 1001365. [Accessed September 16, 2022].

[38] D.A Razak, " Humanising education" The Sun Daily. January 21, 2019. [Online]. Available: https://www.thesundaily.my/opinion/humanising-education-NH393483 [Accessed September 16, 2022].

Diagnosis of Carcinoma from Histopathology Images using DA-Deep Convnets Model

K. Abinaya, B. Sivakumar*

Department of Computing Technologies
SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamilnadu

Abstract—Cancer is a major origin of mortality around the globe, responsible for roughly high morbidity and mortality in 2020, or almost one per six deaths. Cervical, lung, and breast are the most common types of cancers. Cervical is the fourth highest common in women worldwide. Cervical would then kill approximately 4,280 women. Infections that cause, such as human papillomavirus (HPV) and hepatitis, account for approximately 30% of cases in low- and lower-middle-income countries. Many cancers are curable if detected as early as possible. In this proposed work, developed the DA-Deep convnets model (Data augmentation with a deep, Convolutional Neural Network) for the detection of cervical cancer from biopsy images. Deep Convolutional Neural Network presents one of the most applied DL approaches in medical imaging. Today, enhancements in image analysis and processing, particularly medical imaging, have become a major factor in the improvement of various systems in areas such as medical prognosis, treatment, and diagnosis. Based on our proposed model we achieved 99.2% accuracy in detecting the input image has cancer or not.

Keywords—*Cancer; cervical cancer; convolutional neural network; deep learning*

I. INTRODUCTION

The American Society's estimates for cervical in the US in 2022[16]: "There will be approximately 14,100 new cases of invasive cervical given a diagnosis." Oncology encompasses a wide range of diseases that can affect different parts of the body, collectively known as cancer. There are also terms like cancer and neoplasms in use. The term "metastasis" refers to the spread of cancer cells beyond the normal boundaries of the body. Diagnosing for a person looks to the doctor when they found some abnormal manifestation in the body. After that, the doctor checks that person's manifestation and medical history. Based on the symptoms the doctor will suggest a health check test. Some people may not have the manifestation too. They are able to detect only through medical checks like a biopsy, or CT scan. Sometimes doctors will find after the screening test only such as Pap test or mammography.

A. Types of Cervical Cancer

A lab microscope is used to classify cervical s and pre-cancers. Squamous cell carcinomas and adenocarcinomas are the most common types of cervical [18].

1) In most cases of cervical (up to 9 out of 10), squamous cells are present. They arise from ectocervical cells. The most common site of squamous cell carcinoma is the transformation zone.

2) Adenocarcinomas is one of the majority of cervical cancers, it is generated from the glandular cells Adens. Cervical adenocarcinoma evolves from endocervical mucus-producing gland cells.

3) Adenosquamous or mixed carcinomas are identified when both features are detected in the cervical.

B. Standard Screening Tests for Cervical

1) The HPV Test: By taking the sample piece of the DNA (Deoxyribonucleic Acid) from the cervix the doctor is able to identify, this procedure is referred to as the HPV (Human Papilloma Virus) test. HPV will not detect cancer. It will determine only the presence of HPV that cause cervical. When the person has HPV type 16 and 18 there is a possible cervical risk. In the test, doctor will collect the cell from the cervix and look at the changes caused by the HPV. Visual Inspection with Acetic Acid (VIA) is another method for detecting precancerous cervical lesions. A pathologist applies acetic acid (diluted vinegar) to the cervix to see if there are any changes in the cells. Women are given the results of the VIA test right away. In a Cone biopsy, the doctor will detach a cone-shaped piece of tissue from the cervix, called conization. The tissue includes the transformation zone where cervical pre-s and s are most likely to start also sometimes it will completely remove the early cancerous too. Few general concepts are explained in less technical terms to help you better understand how doctors decide if cancer is presently using the Shape and size of the cells, Nucleus shape, and size.

Based on the above-mentioned, the pathologist identifies cancer from the biopsy images using microscopy manually through the more number of slices in tissue after the progress. So, when they are diagnosed with the biopsy images, they may have an error as a manual error. So, the result may come from those having cancer who don't have it and vice versa. So, in the proposed work, the automatic identification of cancer through the deep convolutional neural network is suggested. For getting more accuracy in the proposed model, generate the input with data augmentation.

II. RELATED WORK

A survey was made about cervical cancer by the American society as more new cases were found as well as more number of death due to cervical cancer. In most cases, cervical cancer

---

*Corresponding Author.

is prevented using the HPV vaccination based on the screening test. When diagnose earlier as possible it will be managed effectively. If cancer is found at later stages with suitable/relevant treatment and palliative therapy then it is possible to control it.

S. N. Eldin et al. [1] proposed the diagnosis of the breast using the biopsy images using the best three models of deep learning techniques such as Resnet50 & 101, Densenet 169 model, without doing the preprocessing of the images and with preprocessing the input images. Also in this work, they tell about the various preprocessing techniques based on the performance of the models. D. Bardou et al. [2] work regarding the classification of breasts automatically using machine learning and deep learning approaches. Classification is based on the benign or malignant with subclasses. They used two approaches for the classification extracting the feature using two coding models using SVM in the machine learning model and the second approach using the Convolutional Neural Network model in deep learning techniques. In this work, they generated the images using data augmentation techniques to increase the accuracy of the result. E. CENGİL et al. [3] detect lung cancer using the 3D Convolutional Neural Network model with the implementation of TensorFlow. Here the identification of the cancer is based on the CT images; also here they classified lung cancer as benign stage and malignant stage. Here they used only the less number of a dataset for detection and classification. Y. Li et al. [4] proposed work for the identification of gastric cancer as well as its classification. Here they did the pre-processing technique as cropping because of the large size of images in the given dataset so that the deep learning model can learn and train properly using the given dataset. Here they classify lung cancer based on the patch-based classification and the slice-based classification. Van Eekelen L et al. [5] here worked for the segmentation of the major cell and the tissue types in the bone marrow trephine biopsy images. Along with the input they have biopsy images for the different wide range of persons. Based on the given input they detect and segment using VGG 16 model. They also get the result from the pathologist and compare that result with the automatic detection of cancer using the deep learning model and give the accuracy compared with the pathologist data.

Koh J et al. [6] here worked on the detection of breasts using CT–images using the deep learning model. In this proposed work they also analyze their result with the internal and the external dataset. X. Dong et al. [7] proposed lung cancer detection and segmentation using the CT images with the help of Hybridized Fully Convolutional Neural Network deep learning model with the help of mathematical calculation. Here they performed the data augmentation during the training process. Identify the size using the pyramid structure and the texture classifier used to detect the normal and abnormal cells from the given set of images. Here they use the method to reduce the false-positive results. 2D feature maps are included for better feature extraction. Singla C, et al. [8], Mammogram screening plays an important screening for

the detection of breast cancer. It has a very poor-quality image so it is very difficult for them to identify cancer from the given input. So, to enhance the image for the training they did the image enhancement with the help of the filters. They use the filters such as High PSNR and low MSE values. R. Roslidar et al., [9] in this work, detect breast cancer using Thermography images. Based on the thermograms screening model, doctors are able to detect cancer by observing the temperature distribution in the breast. Here they detect and classify cancer using a deep learning model. Actually to identify breast cancer with large tissue and early detection, can go for the thermography screening technique. So based on the temperature distribution only it is able to find out. It is very difficult for the pathologist to find out manually. So, to overcome these problems they used automatic detection using the deep learning model. To achieve a better result, they use the image enhancement model as a pre-processing method such as mean subtraction, and local contrast normalization. They also did the resizing such as width and height modified images. For the segmentation part, they use the ROI segmentation. L. Wei et al., [10] in this proposed work, classify using the dermoscopy images with the help of a deep learning model. Detection of lightweight skin cancer based on the image pre-processing, training, and fusion of an image. To overcome the overfitting problem in training progress for given input they did the data augmentation method. In pre-processing steps, they construct the positive pair and also construct the negative pair. Here for the segmentation, they use the Mobile net-V1 architecture. They loaded the input images with preprocessing, and in the training process they took 50 epochs. This paper also showed, the with and without discriminator model process accuracy.

1) S. Sharanyaa et al., [11, 12] based on the deep learning model in this paper, they identify early gastric. It is identified using endoscopy and CT scans with lightweight techniques. Before training, the noise part images are removed as pre-processing and the feature extraction is done by the color threshold algorithm. Here they also said about the part of the images having a greater number of the red color band in the given images. They proposed the Deep color net model for the identification of cancer. Find the maximum correlation between the pixels from the images using a test vector. They also highlighted the more spread area using the higher match score. Some pre-processing is also done such as resizing and the RGB image to grayscale conversion. Feature extraction and feature mapping are also done here for pre-processing the data.

Based on the survey of the paper the early detection of cervical cancer is much needed in the world nowadays. Because if it is found early, it can avoid the most last stage of cancer and also reduce the death rate. Due to manual detection sometimes it may get errors. Also it takes more time to detect cancer. So to overcome these problems it's better to get the automatic detection in the medical field with the use of a deep learning model as the convolutional neural network.

III. MATERIALS AND METHODS

In this section, here discussed the data set pre-processing of the images for accurate results. Also, the method used for automatic detection of cervical cancer is Convolutional Neural Network algorithm.

A. Data Set

In the proposed work, the data was collected from the Kaggle dataset. The dataset contains both normal biopsy images as well as cancer biopsy images. In the deep learning model, it needs to train both the normal and the images then it can ably learn by the Convolutional process during the training process. This dataset is again divided for training, testing, and validation processes. It contains the dataset in png format as an image format. So Convolutional Neural Network can learn the feature from the images by feature extraction in the convolution process. The data was in the zipped file format after that it will unzip the file for the data pre-processing for further steps.

B. Data Pre-processing

1) Data augmentation: Data augmentation was used to create additional images in the proposed work, and the Google Colab platform's Image data generator was used to put this idea into action. It is possible to bring the image data generator into Collab by way of the Keras preprocessing. The image has been rotated by 40 degrees, out of a possible 360 degrees. Every flipped picture will find its permanent home in the 3D model. The image has become distorted on all sides. Width shift = 0.2, Horizontal shift = true, Fill mode = nearest allows you to make an object in another image appear narrower or larger than it did in the original. The image has been flipped both horizontally and vertically. When an image is flipped, the pixels are rearranged, but the image's original qualities remain unchanged. Some photographs' vertical inversion is meaningless, but it has uses in cosmology and photography studies of microscopic objects. This is shown in Fig.1 along with the corresponding collab python code. Given algorithm 1 explains about the augmentation process.
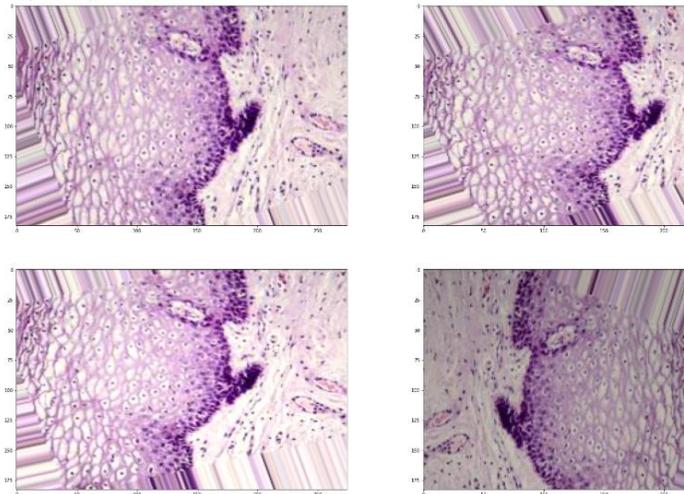
Fig. 1. Input Images with Data Augmentation.

Algorithm 1 Data Augmentation
Input: Normal and Cancer Images
Output: Augmented Images
for i to n images in dataset do
R[i]=>Rotate range from 0 to 30
R[i]=>Zoom range from 0 to 0.2
R[i]=>width shift range as 0.2
R[i]=>Height shift range as 0.2
End for

2) Image training and validation: The dataset is split 80:20 so that the model can be trained and validated on separate data sets, avoiding potentially biased data. All of the photos in the folders that have been designated to correspond with the designated class names will have their titles applied to them automatically. DataLoader also brings in the train's annotated pictures and its data tracks. This separates our dataset into two classes: "normal," representing a biopsy from a healthy person, and ",", representing a random sample of data.

C. Proposed Model Architecture

1) Neural network: The goal of deep learning algorithms is to produce results that are competitive with human judgments by repeatedly analyzing data by a predetermined logical framework. This is achieved by deep learning, which uses a network of interconnected neural processors to create a complex computational model.
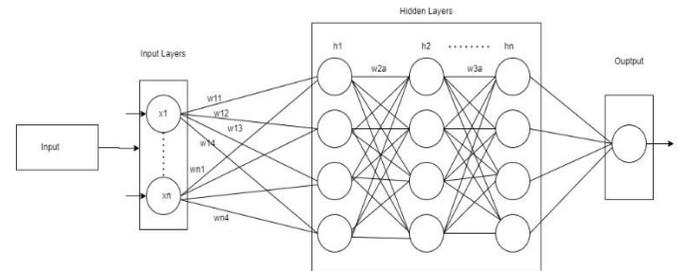
Fig. 2. Neural Network Architecture.

The Input layer in Fig. 2 above receives the data from the Output layer and does nothing with it other than pass it on. Weight, activation function, a deep network, and dense layers all work together in the hidden layer to process the input. Also it shows the connections between neurons and their associated weights (w) and hidden layers (h). What do weights mean? The weight assigned to a neuron indicates how strongly it is connected to other neurons in the network. Every neuron's output is computed using this method. Only the first input (with weights) is included in the Eq. (1) below.

$$Output = F\{[(x1 * w11) + (x1 * w12) + (x1 * w13)] + [(xn * wn1) + (xn * wn2) + (xn * wn3) + (xn * wn4)]\} \quad (1)$$

The last hidden layer is then used to produce the output in the output layer. The four neurons in Fig. 2 will be used to generate the desired result. Each neuron in the network will have four weights used in the calculation.

2) Convolutional neural network

a) Convolutional layer: The convolutional layer is the first layer of a Convolutional Neural Network. Filters occupy a small portion of the input image's height and width but extend throughout its depth. It's been programmed to identify a specific class of features in the images, it's being shown. The filter or kernel is iterated through every possible location on the input matrix as part of the convolution layer. If the input image is I by I pixels and the filter is f by f, then the dimensions of the convolved output can be determined using below Eq. (2),

$$Dimension\ of\ the\ convolved\ ouput =$$
$$[size\ of\ the\ input\ image - filter\ size + 1] *$$
$$of\ [\ size\ of\ the\ input\ image - filter\ size + 1\ ] \quad (2)$$

The Eq. (3) for determining padding is as follows:

$$Output\ size = [i + 2p - f + 1] * [i + 2p - f + 1] \quad (3)$$

The input image's padded size is determined by the Eq. (4) provided,

$$Size\ of\ padded\ input\ image = \ (i + 2p) * (\ i + 2p) \quad (4)$$

Convolutional layer output can be obtained by performing the below Eq. (5),

$$Output\ for\ convolutional\ Layer = \frac{i - f + 2p}{s} + 1 \quad (5)$$

i represents the input image dimension i x i, f represents the filter size f, p represents the padding value, s represents stride value. After the convolutional layer, the output size is calculated to be 3x3, which is the result of the operations of padding, stride, and convolution.

The Eq.(6) used to determine how many pooling layers should be used.

$$Dimension\ of\ pooling\ Layer = \frac{input - filter}{stride} + 1 \quad (6)$$

Consider input as 4 x 4, Filter value as 2, Stride value as 2. Finally it calculates the input, weight, and bias value to predict the output using Eq. (7),

$$Ouput = ReLu[(i1 * w1) + (i2 * w2) + (in * wn) + Bias \quad (7)$$

i-represent the input layer, w-represent the weight, b-represent the bias value

ReLu is the activation function.

b) Dropout: As a workaround, a dropout layer is implemented. In order to reduce the training time and overall size of the model, this layer prunes the neural network of some of its neurons. The neural network has a dropout threshold of 0.3, at which point 30% of the nodes are removed at random.

c) Assertion of Activation: The model's final and most crucial parameter, the activation function is essential to any convolutional neural network. It merely chooses which bits of data should be sent forward and which ones should be left floating around indefinitely. Based on this, let's use Adam as an optimizer with a learning rate of 0.001. Consequently, our model's process is depicted in Fig. 3. It is based on all of the aforementioned layers, and it processes the operations and gives the output by training the images through all of the layers.
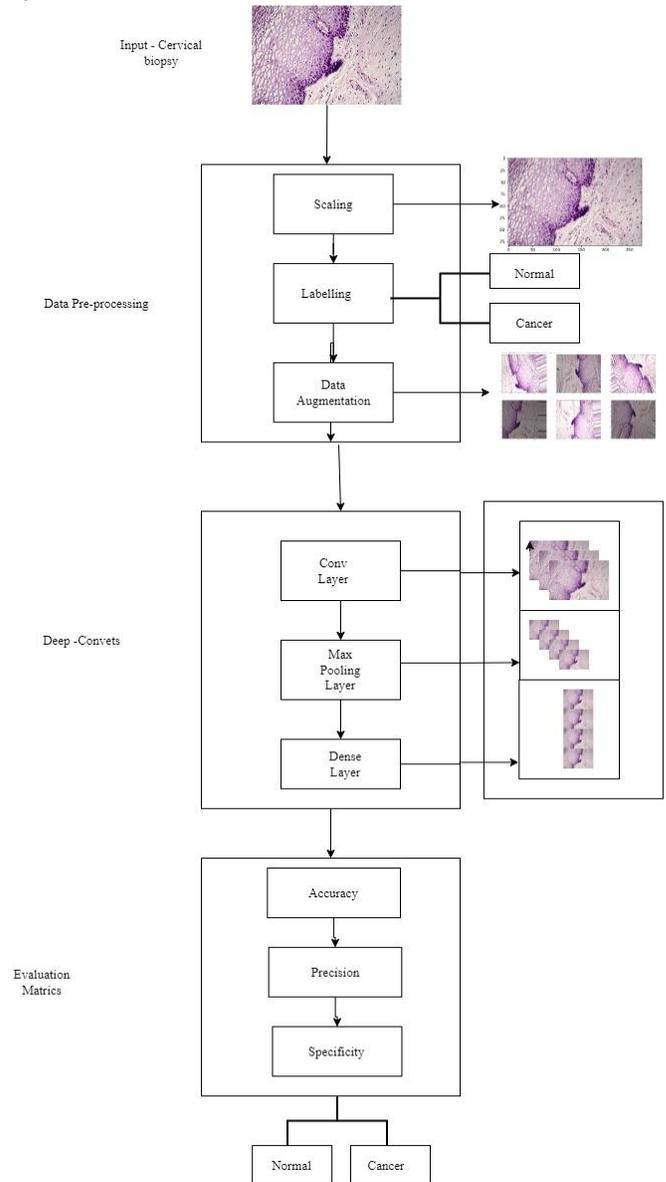
Fig. 3. Block Diagram of the Proposed DA-Deep Convents Architecture.

IV. RESULT AND DISCUSSION

Using the Kaggle dataset, the proposed DA-Convent model was implemented in Google Colab. We also compared unaltered images to those that had data superimposed on them. Data augmentation for the input images, while it takes more time, will result in higher precision than the raw input data.

Fig. 4(a) and Fig. 4(b) show the training and validation of accuracy and loss value. Here for data augmentation, consider a number of the epochs as four to validate the accuracy and the loss function for the given set of images.

(a)

(b)
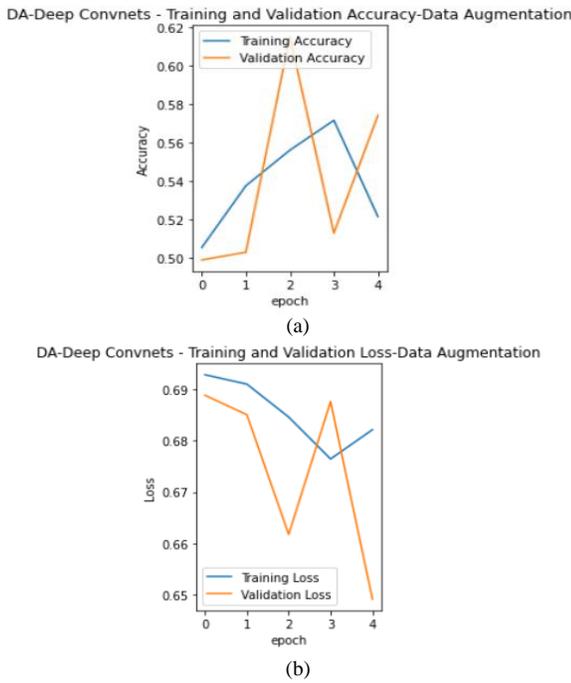
Fig. 4. (a) Traning and Validation Accuracy for Data Augmentation (b). Traning and Validation loss for Data Augmentation

A comparison of the results obtained with and without data augmentation, as well as an explanation of how the data is changing, will be provided as an output. Therapeutic identification methods include measures like precision, recall, f-score, and accuracy. A confusion matrix can also be used to assess the quality of the model. The Confusion Matrix displays how the classification component becomes confused while making predictions. The confusion matrix displays four categories: the expected classes from the original problem as columns, and the actual classes as rows. The following Eq. (8), (9) and (10) is used to determine the accuracy and loss value of the DA-Deep Convnets model.

True Positive: Normal cell, detect as Normal cell.

True Negative: Really cancer cell, detect as cancer cell.

False Positive: Normal cell, detect as cancer cell.

False Negative: Really cancer cell, detect as normal cell.

$$Accuracy = \frac{Number\ of\ classes\ correctly\ classified}{total\ number\ of\ classes\ taken} = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

$$Precision\ for\ Positive\ class = \frac{True\ Positive}{Number\ of\ cases\ predicted\ as\ positive} = \frac{TP}{TP+FP} \tag{9}$$

$$Precision\ for\ Negative\ class = \frac{True\ Negative}{Number\ of\ cases\ predicted\ as\ Negative} = \frac{TN}{TN+FN} \tag{10}$$

Based on the above equation, here calculated the performance for the given set of images and achieved more accuracy compared to the previous one. During the training process, here achieve more accuracy as well as reduce the loss function, as shown in Fig. 5(a) and Fig. 5(b).

(a)

(b)
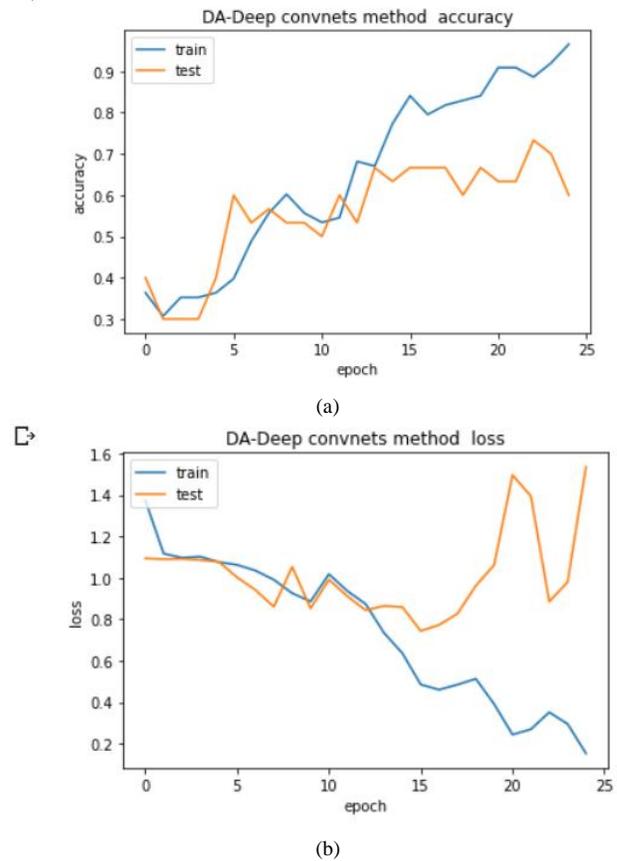
Fig. 5. (a) Training and Testing Accuracy for DA-Deep Convnets Net Model with 25 Epoch, (b). Training and Testing Loss for DA-Deep Convnets Net Model with 25 Epochs.
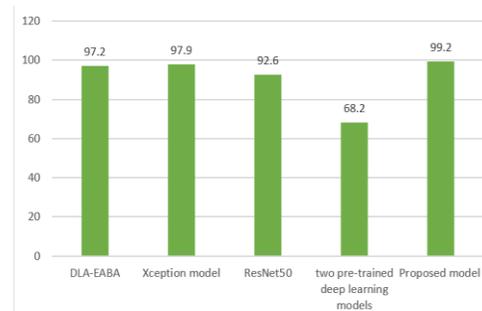
Fig. 6. Comparative Experiment Results of Proposed Architecture with different Models for Accuracy.

Fig. 6 shows the various comparative analysis with different models with the proposed model with reference [13, 14, 15, 17].

V. CONCLUSION AND FUTURE WORK

In the proposed work, an intelligent health care design is developed for the detection of cervical cancer using the biopsy image. Currently, the incidence of cervical cancer is at an all-time high. One of the most important aspects of a patient's recovery is a timely diagnosis. One of the most popular visual recognition tasks is using Convolutional Neural Networks. Every day, CNN's position in the medical section grows stronger. In proposed DA-Deep Convnets model to create a

more reliable model to assist dermatologists in accurately identifying cervical cancer. Accuracy in categorization and the top two was achieved through the use of data augmentation. With the help of data augmentation and the convolutional neural network, accuracy was achieved by 99.2%. In future work, we plan to do the classification and segmentation of cervical cancer for a more accurate process with more number of classifications.

REFERENCES

[1] S. N. Eldin, J. K. Hamdy, G. T. Adnan, M. Hossam, N. Elmasry and A. Mohammed, "Deep Learning Approach for Breast Cancer Diagnosis from Microscopy Biopsy Images," 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), 2021, pp. 216-222, doi: 10.1109/MIUCC 52538.2021.9447653.

[2] D. Bardou, K. Zhang and S. M. Ahmad. 2018.Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks, in IEEE Access, vol. 6, pp. 24680-24693, , doi: 10.1109/ACCESS.2018.2831280.

[3] E. CENGİL and A. ÇINAR. 2018.A Deep Learning Based Approach to Lung Cancer Identification, International Conference on Artificial Intelligence and Data Processing (IDAP), pp. 1-5, doi: 10.1109/IDAP.2018.8620723.

[4] Y. Li, X. Li, X. Xie and L. Shen. 2018 Deep learning based gastric cancer identification.IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 182-185, doi: 10.1109/ISBI.2018.8363550.

[5] Van Eekelen L, Pinckaers H, van den Brand M, Hebeda KM, Litjens G. 2022.Using deep learning for quantification of cellularity and cell lineages in bone marrow biopsies and comparison to normal age-related variation. Pathology;54(3):318-327. doi:10.1016/j.pathol.2021.07.011.

[6] Koh J, Yoon Y, Kim S, Han K, Kim EK. 2022.Deep Learning for the Detection of Breast Cancers on Chest Computed Tomography. Clinical Breast Cancer.;22(1):26-31. doi:10.1016/j.clbc.2021.04.015.

[7] X. Dong, Y. Zhou, L. Wang, J. Peng, Y. Lou and Y. Fan. 2020.Liver Cancer Detection Using Hybridized Fully Convolutional Neural Network Based on Deep Learning Framework,in IEEE Access, vol. 8, pp. 129889-129898,doi: 10.1109/ACCESS.2020.3006362.

[8] Singla C, Sarangi PK, Sahoo AK, Singh PK. 2022.Deep learning enhancement on mammogram images for breast cancer detection. Materials Today: Proceedings.;49:3098-3104. doi:10.1016/j.matpr. 2020.10.951.

[9] R. Roslidar et al. 2020 A Review on Recent Progress in Thermal Imaging and Deep Learning Approaches for Breast Cancer Detection, in IEEE Access, vol. 8, pp. 116176-116194, doi: 10.1109/ACCESS. 2020.3004056.

[10] L. Wei, K. Ding and H. Hu, Automatic Skin Cancer Detection in Dermoscopy Images Based on Ensemble Lightweight Deep Learning Network, in IEEE Access, vol. 8, pp. 99633-99647, 2020, doi: 10.1109/ACCESS.2020.2997710.

[11] S. Sharanyaa, S. Vijayalakshmi, M. Therasa, U. Kumaran and R. Deepika. 2022 DCNET: A Novel Implementation of Gastric Cancer Detection System through Deep Learning Convolution Networks, 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), pp. 1-5, doi: 10.1109/ICACTA54488. 2022.9752960.

[12] S. Iqbal et al., 2021.Prostate Cancer Detection Using Deep Learning and Traditional Techniques,in IEEE Access, vol. 9, pp. 27085-27100, doi: 10.1109/ACCESS.2021.3057654.

[13] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He and J. Fan.2020.Deep Learning Assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis,in IEEE Access, vol. 8, pp. 96946-96954, doi: 10.1109/ACCESS.2020.2993536.

[14] Jain R, Gupta M, Taneja S, Hemanth DJ. 2020.Deep learning based detection and analysis of COVID-19 on chest X-ray images,Applied Intelligence. 2020;51(3):1690-1700. doi:10.1007/s10489-020-01902-1.

[15] Ismael AM, Şengür A. 2020.Deep learning approaches for COVID-19 detection based on chest X-ray images.Expert Systems with Applications. 2021;164:114054. doi:10.1016/j.eswa.2020.114054.

[16] https://www.cancer.org/cancer/cervical-cancer/about/key-statistics.html.

[17] Alyafeai Z, Ghouti L. A fully-automated deep learning pipeline for cervical cancer classification. Expert Systems with Applications. 2020;141:112951. doi:10.1016/j.eswa.2019. 112951.

[18] https://www.cancer.org/cancer/cervical-cancer/about/what-is-cervical-cancer.html.

KMIT-Pathology: Digital Pathology AI Platform for Cancer Biomarkers Identification on Whole Slide Images

Digital Pathology AI Platform for Cancer Biomarker Identification

Dr. Rajasekaran Subramanian[1], Dr. R. Devika Rubi[2], Dr. Rohit Tapadia[3], Rochan Singh[4]

Associate Professor, Dept of Computer Science and Engineering, Keshav Memorial Institute of Technology, Hyderabad, India[1, 2]
Consultant – Pathology, Tapadia Diagnostics Centre, Hyderabad, India[3]
Research Intern, Dept of Computer Science and Engineering, Keshav Memorial Institute of Technology, Hyderabad, India[4]

Abstract—**Analysis and identification of cancer imaging bio markers on biopsy tissues are done through optical microscope. Digital tissue scanners and Deep learning models automate this task and produce unbiased diagnostics. The digital tissue scanner is called as virtual microscopy which digitize the glass slide tissues and the digitized images are called as Whole Slide Images (WSI). They are multi-layered (level) images having high resolution, huge in size and stored as a pyramidal tiff file. As normal web browsers are unable to handle WSI, a special web imaging platform is needed to obtain, store, visualize and process WSI. This platform must provide basic facilities for uploading, viewing and annotating WSI which are the inputs to the deep learning models. The integration of deep learning models with the platform and the WSI database provides a complete solution to cancer diagnostics and detection. This paper proposes two AI deep learning models for the diagnostics and the detection of cancer imaging bio markers on breast cancer and prostate cancer WSI. Efficientnet deep learning model is used to detect ISUP (International Society of Urologic Pathologists) grading for prostate cancer which is trained and tested by 5000 prostate WSI and produces 80% accuracy with 0.6898 quadratic weighted kappa (QWK) score. R2Unet model is used to identify tubule structures for breast cancer which is a morphological component to grade breast cancer. The model is trained and tested by 17432 WSI tiles and generates f1 metric accuracy as 0.9961 with mean_io_u 0.8612. The paper also shows the complete execution of these two Deep learning models (from uploading WSI to visualize the AI detected results) on the newly developed WSI imaging web platform.**

Keywords—*AI for cancer prediction and diagnostics; deep learning for WSI analysis; tubule prediction on breast cancer; ISUP grading for prostate cancer; WSI imaging platform*

I. INTRODUCTION

Optical microscope is the fundamental device for analysing, diagnosing and identifying disease bio markers on the tissue slides. It helps to identify morphological and cell artifacts of the disease. The heavy workload and continuous diagnosis through naked eye on microscope by pathologist may lead to a biased analysis for the disease. Digital scanners remove all these limitations and introduce a new research field called Digital pathology. Digital scanners capture high resolution images of the glass slide tissues and produce Whole Slide Images (WSI) by scanning the conventional glass tissue slides which are the most usable imaging modality by the pathologists worldwide. WSI system, also called as virtual microscopy, consists of two parts namely a scanner (hardware) and a virtual slide viewer (software). Scanner scans the conventional glass slide into digital image (file) and virtual slide viewer helps to view and analyse these digital images.

The digital WSI files are larger than any other healthcare digital images such as radiology images. These files are multi gigabyte in size (about 1600 megapixels) and requires a storage space about 4.6 GB [1]. They are stored in various types of image formats used by whole-slide imaging equipment sellers, out of which the common ones are TIFF, OME-TIFF, JPEG, NDPI or SVS. All these vendor specific formats are handled in the WSI platform by Open Slide [2] and Bio-Formats which are C and java-based software tools developed by Carnegie Mellon University and OME Consortium [3] respectively. The main processes related to WSI are obtaining, storing, visualizing and processing the digitized WSI. As these files are constructed as a multi-layered pyramid structure which provides a microscopy visual effect to the users to view and navigate (pan and zoom) slides on the computer screen.

Deep learning is used to improve diagnostic accuracy in digital pathology. Various pathologists from different locations can collaborate on the web based WSI platform. The whole slide images can be visualized in the web browser and the pathologists can annotate (mark) the affected (cancer) region on the WSI with the help of annotation tools. These annotations can then be used as inputs for training deep learning models to detect and analyse the cancer regions.

Currently, there are a lot of WSI viewers available in the market. A few examples are Digital slide Archive (DSA) which has an inbuilt WSI viewer [4], Orbit image analysis which is a WSI viewer/analysis tool [5], Orthanc WSI is a web viewer [6], OpenSeadragon AJAX image viewer [7] is a web image viewer for reading and viewing WSI, SlideRunner is an open-source cell annotations and visualization tool [8] and QUPATH is a desktop application for WSI data analysis [9].

This paper proposes two AI deep learning models for Tubule structure detection in Breast Cancer and ISUP Grading detection in Prostate Cancer using a newly developed web imaging platform called as KMIT-Pathology. It also describes the important features of the WSI imaging platform such as obtaining/storing WSI, WSI/Tiles retrieval, annotations, visualizations, images analysis and AI predictions.

This paper is organized as follows; Section II discusses the related works. Section III describes the proposed web platform and deep learning models for cancer diagnostics and detection. Section IV presents the Results. Section V gives the Conclusion, Limitations and Further enhancement of cancer detection, deep learning models and the web-based imaging platform for WSI.

II. RELATED WORKS

The authors Tiago Marques Godinho [10] proposed an architecture to integrate digital pathology workflows with a universal Picture Archive and Communications Systems (PACS) using solely DICOM standard data formats and communications. They suggested that storing a WSI using tiled organization shown in Fig. 1 (Middle) is an efficient way than storing it in a single frame organization shown in Fig. 1 (Left). In single frame organization the entire image pixels are stored sequential array where as in tiled organization the entire image is stored as a series of rectangular regions which makes easy to directly access any 2D sub regions of the image.

Image pyramid organization shown in Fig. 1 (Right) helps to perform panning and zooming vision of WSI. Pyramid organization contains various zoom level tiled images of same WSI at various layers of the pyramid. The top layer of the pyramid contains the lowest resolution of the WSI called thumb nail whereas the bottom layer contains highest resolution tiled images of WSI. This structure felicitates the zooming/panning visualization of WSI. Pyramid organization is the accepted visualization standard for WSI by DICOM [11].

David A Gutman et al. [4] developed an open-source web-based platform called Digital Slide Archive (DSA) for digital pathology. DSA allows users to manage large collections of pathology images. DSA has an image viewer which can be used to view the pathology images. DSA represents the WSI (whole slide image) data in a pyramid structure that contains several images at different resolutions. Partitioning of WSI into tiles is done by the DSA for efficient image analysis and visualization which makes DSA scalable. They also discussed several image analysis algorithms and pipelines for WSI images.

OpenSeadragon (OSD) a JavaScript based open-source web viewer uses deep zoom images (DZI) developed by Microsoft to save pyramidal file for WSI. This helps OSD to extract/display querying image tiles rather than downloading the entire WSI. OSD provides efficient zoom and navigation on WSI through DZI file format and the different file formats of the various vendors are handled by open slide, a C based library developed by Carnegie Mellon University.
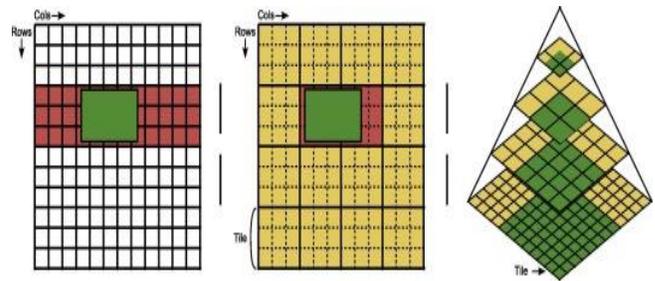
Fig. 1. Storage Structure of WSI (Single Frame Structure, Tiled Structure and Pyramidal Structure of WSI).

Slide Runner [8] is a python based open source WSI tool for massive cell annotation and visualization developed at Pattern Recognition Lab, Fried rich-Alexander-Universität Endanger-Nürnberg and the Institute of Veterinary Pathology, Freie University Berlin. It can perform various cell annotations such as centre of the cell, boundary of the cell, single cell annotation and multi cell annotations by a single click. It uses open slide to manage vendor formats of WSI and provide plugins for various image processing operations such as threshold, segmentation and normalization.

QuPath is a java-based desktop software for WSI imaging analysis. It provides annotation, segmentation and classification for individual nuclei or cells. It performs cellular map and morphological feature extraction of the entire tissue section. It uses ImageJ and open CV libraries for image processing operations. Users can create their custom workflow by using scripting functionalities and add their extensions along with the existing QuPath tools. Neofytos Dimitriou et al. [12] discuss the various deep neural networks training models for WSI and emphasize the different methodologies involved in these network models. They also discuss various image formats of WSI and the lack of a universal image format. They give the insights for better performance of deep learning training by using patches of WSI than the entire WSI.

Famke Aeffner et al. [13] gives a brief introduction for modern digital pathology concepts which includes digitization of glass slides, software tools and the complement between computational image analysis and GUI for displaying the WSI. They also discuss in detail about the importance of a pathologist's role in image analysis and emerging role of AI in digital pathology. The authors, Laura Barisoni et al. [14] describe the evolution of digital pathology and computational image analysis which brings a lot of new changes in the pathology ecosystem. They also talk about the use of computer vision and AI techniques in the conversion of digital pathology images into mine able data, extraction of image information, developing techniques to diagnostic diseases and help to accurately identify patients at risk.

Breast cancer is the most leading cancer in the world. Histological grading is used to evaluate the behaviour and prognosis of breast cancer on H&E images. It quantifies the tumour architecture and the cytology deviation of the breast cancer tissue against the normal. Nottingham Grading system is a globally accepted breast cancer grading system in which Tubule identification is one of the morphological components to grade breast cancer classification [15]. The authors, Xiao

Jian Tan et al. [16] used independent data set and support vector machine to classify tubule and non-tubule. They used three parameters spatial angle, distribution and harmonic mean to measure the amount of tubule and non-tubule formation and they got the classification accuracy as 98%.

Ajay Basavanhally et al. [17] used O'Callaghan neighbourhood method to identify tubule on H&E breast tissue. They involved not only low-level attributes for training but also include the spatial distances of lumen, cytoplasm and nuclei to detect tubule. They used colour convolution method to detect nuclei. Colour gradient based active contour model and hierarchical neighbourhood algorithm are used to detect lumen areas in tubule which produced 80% classification accuracy.

Prostate cancer is the deadliest cancer affects and destroys the stroma and normal arrangements of prostate glands. International Society of Urologic Pathologists (ISUP) grades the morphological /cytological patterns of the cells in prostate cancer tissues. Gleason Scoring System (GSS) is a part of ISUP grading which helps ISUP to find major and minor patterns of the cells. These major and minor patterns decide the value of ISUP which ultimately decides the grade of the prostate cancer. Yuchen Li et al. [18] combined different neural networks such as multi scale parallel branch convolution neural network (MP B-CNN) and DeepLabv architecture to identify Gleason grading and pattern segmentation for prostate cancer. The system used 1211 WSI for training and testing and produce mean IOU as 77.29 with quadratic weighted kappa (QWK) as 0.77.

III. PROCEDURE AND APPROACH

The WSI imaging platforms QuPath, Slide Runner and Digital Slide Archive (DSA) provide facilities for Data Annotations, Images uploading, retrieval and visualizations. QuPath is a desktop-oriented platform. All these platforms provide tools to perform WSI image processing techniques. There exists a gap in digital pathology which handles WSI images and require a web-based platform. This platform should perform an end-to-end operation such as images uploading, retrieval, annotations, visualizations with AI predictions.

The proposed KMIT-Pathology platform is designed to fill-in this identified gap by performing above listed end-to-end operations. This section describes the various features of KMIT-Pathology platform such as Description of pyramidal data structure of WSI, Annotations and visualization of WSI. This section also discusses about the deep learning models to identify tubule on breast cancer and to find ISUP grading for prostate cancer.

A. WSI Visualization and Annotations

KMIT-Pathology uses pyramidal structure files of WSI which are stored as .tiff and .SVS formats. Pyramidal images are stored with different levels of resolutions. The base level is having full resolution and as it goes up the resolution of levels decreases [19]. The thumbnail (the top most level) is a very low-resolution version of the image. Fig. 2 represents pyramidal image having various levels with different resolutions. SVS files are the digital slide images produced by

Aperio digital pathology slide scanner [20]. SVS image format is essentially based upon the Tiled TIFF format, it also stores the tiles in pyramidal form. It has all the essential features of TIFF format with few additional features such as storing the thumbnail image and new non-standard compression tags (33003 and 33005) [21].
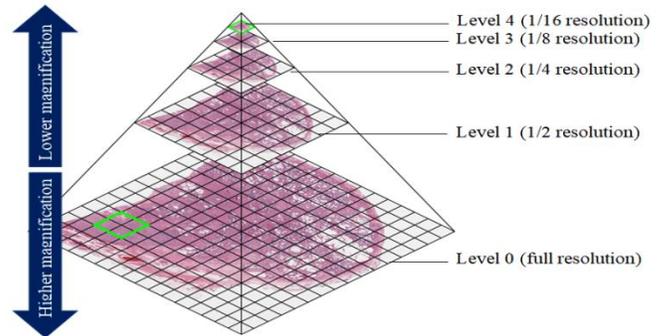
Fig. 2.   Pyramidal Image with different Levels of Resolution.
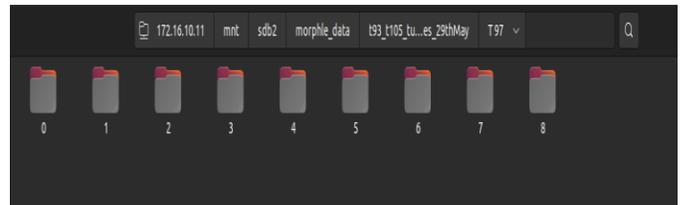
Fig. 3.   Pyramidal Tiff (Level Wise Folders) of WSI.

Morphle continum 240 scanner generates pyramidal TIFF files as a list of sequential folders (level wise resolution) with each folder containing a number of tiles belonging to the resolution of the respective levels as shown in Fig. 3.

(a)

(b)

(c)

(d)

(e)

(f)
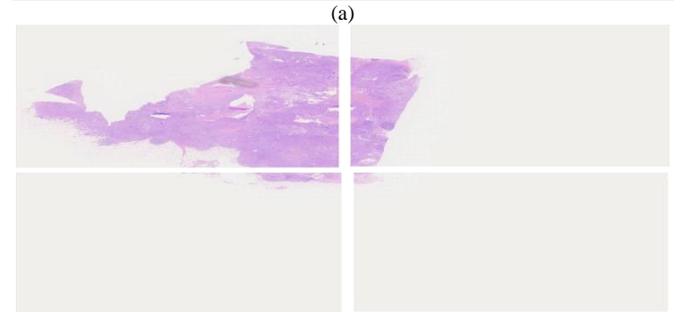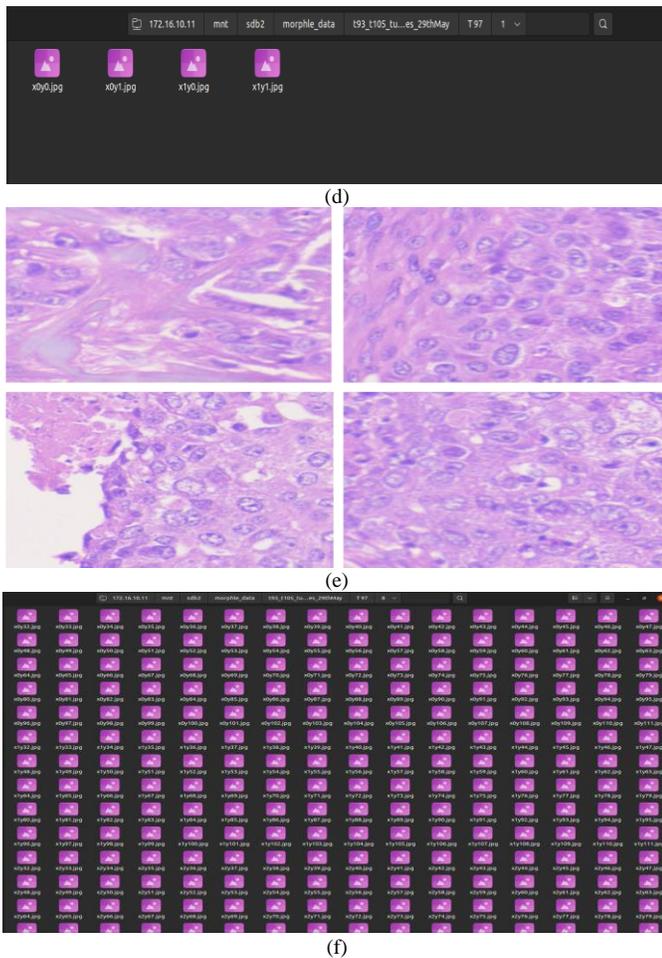
Fig. 4. (a) Level 0 Resolution Image, Level 0 Low Resolution Image is shown in Fig. 4 (a), Fig 4(b). Level 1 Resolution Image, Level 1 Low Resolution Images are Shown in Fig. 4 (b), Fig 4 (c). Level 0 Images Folder, Level 0 Low Resolution Images Folder is Shown in Fig. 4(c), Fig 4 (d). Level 1 Images Folder, Level 1 Low Resolution Images Folders are Shown in Fig. 4(d), Fig 4 (e). Level 8 High Resolution Images, Level 8 High Resolution Images are Shown in Fig. 4(e), Fig 4 (f). Level 8 Images Folders, Level 8 High Resolution Images' Folders are Shown in Fig. 4 (f).

KMIT-Pathology uses zero to eight folders to view various zoom levels from 0x to 80x. Level zero (Folder 0) is the thumbnail of WSI and level eight contains high resolution tiles of WSI. Each level contains twice the number of tiles than the previous level. Level zero consists of one image, level one contains four images and level eight consists of thousands of tiled images which are shown in Fig. 4(a) to Fig. 4(f). OpenSlide is used for zooming and panning operations (for visualization) on tiles which are present in various levels of WSI as shown in Fig. 4(a) to Fig. 4(f). Visualization of a WSI (at Zoom level 1) on KMIT-Pathology platform is shown in Fig. 5.

Annotations on WSI done by pathologist shows the diseased region (ground truth) and deep learning models learns the features of the diseased regions from ground truth and identifying (or predicting) the diseased region on any new image, if it exists. Even though WSI can be visualized at various zoom levels (resolution), the annotations performed at any level will be reflected only at the base level (high

resolution level). So, KMIT-Pathology provides facilities to annotate on base level tiles as well as on whole WSI. For example, as mitosis annotations (Fig 6(a)) requires base level annotations whereas tubule annotations (Fig. 6(b)) depend on structures which can be performed on whole WSI at any level.

Fig. 5. Visualization of a WSI.
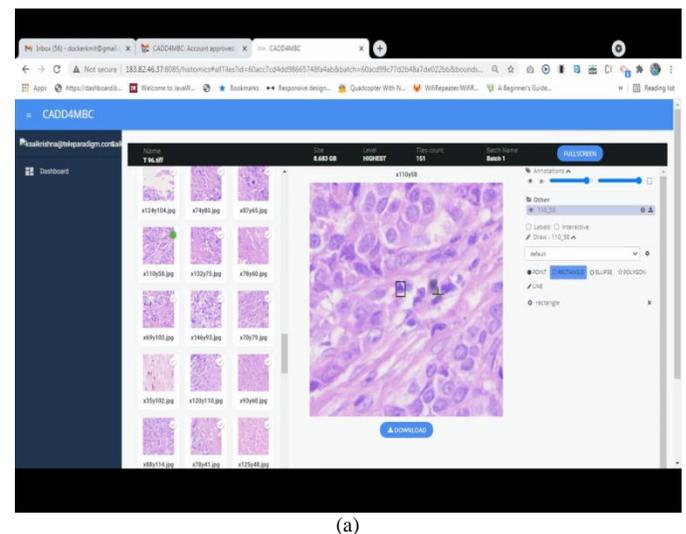
(a)
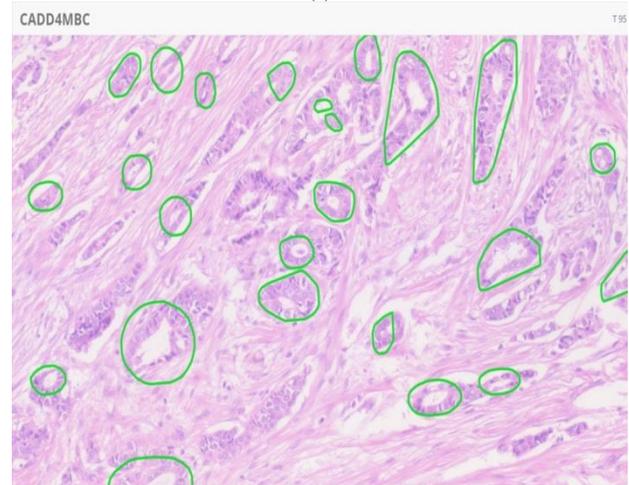
(b)

Fig. 6. (a). Annotations on WSI-Tiles by KMIT-Patholog, (b). Annotations on whole WSI by KMIT-Pathology.

605 | P a g e

KMIT-Pathology provides drawing tools needed for annotations and having a provision to download the .json files of the annotations which are later used by deep learning models for prediction.

B. Deep Learning Models

KMIT-Pathology performs WSI image analysis, primarily for Cancer detection and diagnostics. This section discusses about two deep learning models to identify tubule on Breast Cancer WSI images and ISUP grading for Prostate cancer using WSI images.

1) Tubules identification on breast cancer WSI using KMIT-Pathology platform: Histological grading is used to evaluate the behaviour and prognosis of breast cancer on H&E images. Histological grading provides the degree of differentiation in the morphological assessment of breast cancer. Nottingham grading system (NGS) was globally accepted as a histological guide for breast cancer grading classification. Identification of tubule is one of the major components in NGS system [22]. Tubules are a special type of Breast Carcinoma which are composed of tubular structures. They have an open lumina, lined by a single layer of epithelial cells. 2 - 6 % of malignant tumours are Tubular [23]. They define the aggressiveness of cancer and become disorganized as they progress. In Benign cases, the tubules that are almost round and mostly identical, but in malignant cases they are highly random and identifying boundaries is difficult for experts as well.

Deep learning model R2Unet is used for training and testing to identify tubule on breast WSI. There are around 50 breast cancer WSI of size varies from 8 GB to 12 GB are collected and annotated by three pathologists on WSI using KMIT-Pathology web platform. The annotations are cropped from WSI and the mask are created to run deep learning models.
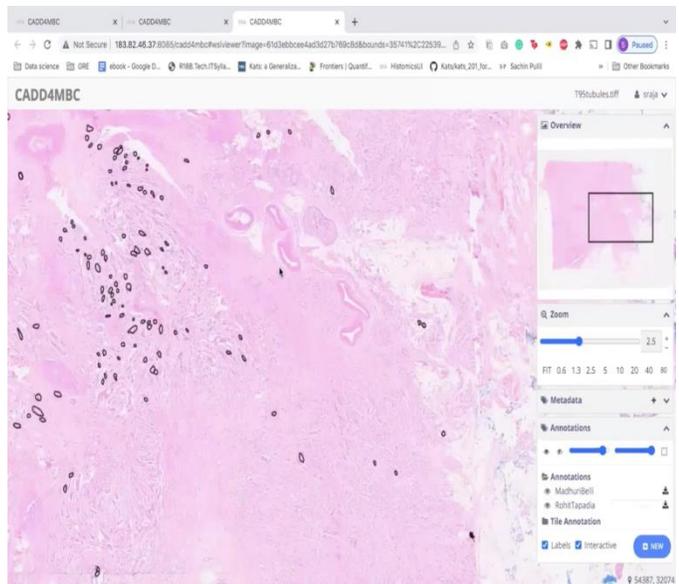
Fig. 7. Tubule Identification on Breast Cancer WSI using R2Unet.

The model uses 140000 tiles of size 256*256 of magnification 40x for training and testing and the model got accuracy of F1 score 0.9295. The tubule identification on breast cancer WSI is shown in Fig. 7.

2) ISUP grade prediction for prostate cancer using KMIT-Pathology: Prostate cancer is the second largest cancer which causes more death in men worldwide. Lack of expert pathologist delays the diagnosis and creates human biased decisions in analysis and the diagnosis of the cancer. AI driven computer aided diagnosis (CAD) system automatically predict ISUP grading for prostate cancer and also helps the pathologist for early prediction and diagnosis.AI driven KMIT-ISUP grade prediction model for prostate cancer [24] uses EfficientNet to predict Gleason scores. ISUP grade depends on the Gleason Scoring (GS) system. GS system defines the structural growth pattern of cancer cells in the tissue. ISUP grading is calculated with respect to the combined score of GS majority and minority scores.

KMIT-ISUP grade identification deep learning model predicts GS score and the respective ISUP grade. KMIT-ISUP model uses Prostate Cancer Grade Assessment (PANDA) challenge dataset provided by Radboud University Medical Center and Karolinska Institute [PANDA Challenge, 21]. The model uses 5000 WSI and the respective semantic masks for various Gleason scores.
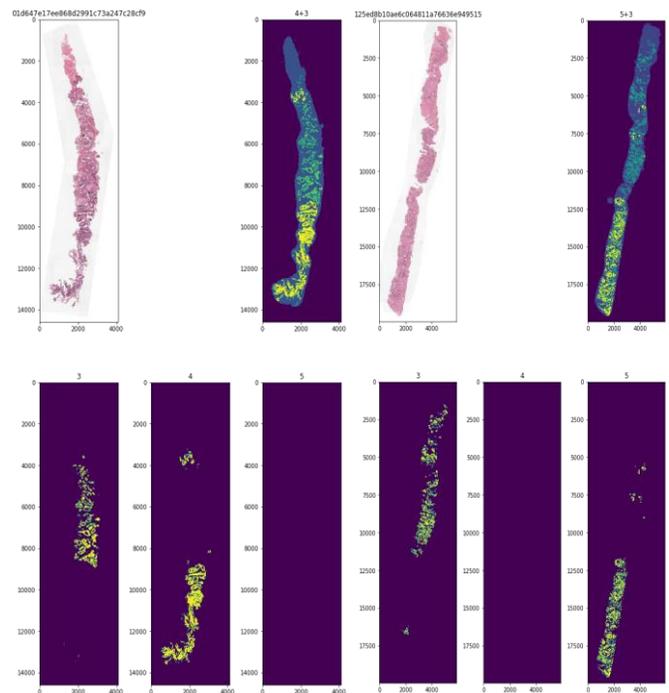
Fig. 8. Sample Prostate WSIs and the Masks with Gleason Score Labels.
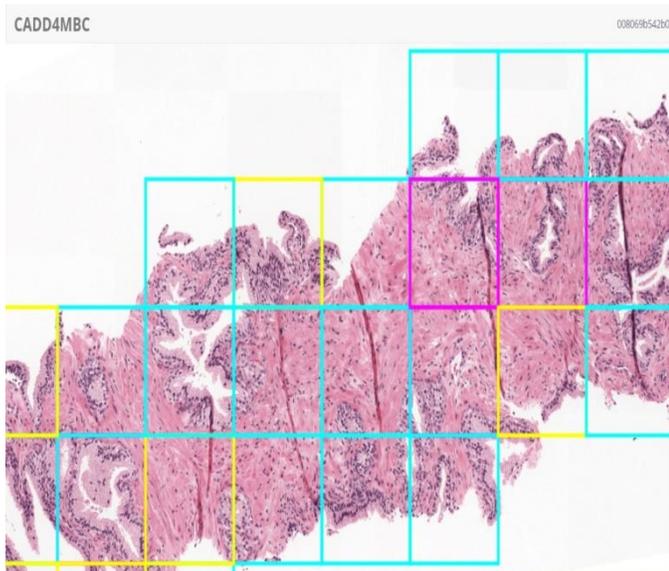
Fig. 9. Prostate WSI with Gleason Score Identification.

The WSI images and the masks are divided into 92783 patches of size 224x224 for training and testing and the model produces 80% of accuracy for ISUP grading prediction. Sample prostate WSIs and the semantic masks for ISUP 3 (4+3) and ISUP 4 (5+3) are shown in Fig. 8. The majority and the minority Gleason score (GS) for ISUP grade 3 is GS 4 and GS 3 whereas for ISUP grade 4 the majority and the minority Gleason scores (GS) are GS5 and GS3. The mask for GS3, GS4 and GS5 are generated using the mask of whole WSI which are also shown in Fig 8. Deep learning predictions for Gleason scores of prostate WSI patches and the color representation for the various GS are shown in Fig. 9. Gleason scores of the patches of WSI are sorted and the two topmost scores (majority and minority patterns) are considered to calculate ISUP grade of the WSI.

IV. RESULTS

KMIT-Pathology imaging platform is an End-to-End digitized pathology imaging platform having various necessary features like WSI data storage in Pyramidal structure with respective folders structures, Pathology Annotations tools for both complete WSI image and for Image-tiles, visualization of AI-Predictions. The deep learning models for Breast Cancer and Prostate Cancer, which are detailed in this publication are developed using KMIT-Pathology platform. Thus, proposed KMIT-Pathology imaging platform has been effectively showcased by using its various features like WSI Images storage and retrieval, Images Annotations, AI-Predictions visualizations etc.

This paper also discusses about the implementation of two AI deep learning models viz. 1) to identify Tubule on Breast Cancer WSI 2) ISUP grading on Prostate cancer by using KMIT-Pathology platform. Tubule deep learning model R2Unet uses 140000 tiles of size 256*256 of magnification 40x for training and testing to identify tubule and the model got accuracy of F1 score 0.9295. The deep learning model EfficientNet is used to predict Gleason scores of the prostate cancer which uses 92783 tiles of size 224x224 for training and testing and the model produces 80% of accuracy for ISUP grading prediction.

Both the deep learning model used KMIT-Pathology for Image/WSI uploading, retrieval, storage, image annotations, AI based cancer bio-markers identification and visualizations of the AI results.

The other imaging platforms like QuPath, OpenSeaDragon Slide Runner and Digital Slide Archive (DSA) provide facilities for Data Storage, Annotations and Visualization. Whereas, KMIT-Pathology includes all the above-mentioned features, along with AI-Deep Learning Models to identify Cancer Biomarkers on WSI images.

V. CONCLUSION AND THE FUTURE ENHANCEMENT

The invention of digital scanners gives a new era to pathology labs and opens a new research area called Digital pathology. Digital scanners capture high resolution images of the glass slide tissues. These virtual microscopy images need to be stored, visualized, analyzed and produce AI driven predictions. Various people located in different places like lab technicians, pathologists and AI engineers are needed to collaborate to perform these operations. So, they need a common web platform to interact with each other. KMIT-Pathology is a web-based platform for WSI image analysis. This platform provides facilities to perform upload, annotate and visualize WSI images. The integration of AI models with the web-based platform helps to identify cancer imaging biomarkers.

This paper also demonstrates the usage of KMIT-Pathology platform for End-to-End Cancer Bio-markers identification for Breast Cancer and Prostate Cancer WSI images.

REFERENCES

[1] DICOM Whole Slide Imaging (WSI) [Internet]. Rosslyn: National Electrical Manufacturers Association (NEMA), c2020 [updated 2020 May 8; cited 2021 Dec 21], Available from: https://dicom.nema.org/Dicom/DICOMWSI/.

[2] Goode A, Gilbert B, Harkes J, Jukic D, Satyanarayanan M. "OpenSlide: A vendor-neutral software foundation for digital pathology", J Pathol Inform, 2013; 4:27.

[3] Linkert M, et al. "Metadata matters: access to image data in the real world", J Cell Biol. 2010;189:777–782.

[4] Gutman DA, Khalilia M, Lee S, Nalisnik M, Mullen Z, Beezley J, Chittajallu DR, Manthey D, Cooper LAD. "The Digital Slide Archive: A Software Platform for Management, Integration, and Analysis of Histology for Cancer Research", Cancer research vol. 77, 21 (2017), e75-e78, doi:10.1158/0008-5472.CAN-17-0629.

[5] Stritt M, Stalder AK, Vezzali E. Orbit Image Analysis. "An open-source whole slide image analysis tool", PLoS computational biology vol. 16,2 e1007313. 5 Feb. 2020, doi:10.1371/journal.pcbi.1007313.

[6] Orthanc [Internet]. Dept of Medical Physics, University Hospital of Liege, Osimis S.A., c2017-2021 [cited 2021 Dec 21], Available from: https://www.orthanc-server.com/static.php?page=wsi.

[7] OpenSeadragon project, OpenSeadragon. 2013. [Last accessed on 2013 Aug 26]. Available from: http://openseadragon.github.io.

[8] Aubreville, M., Bertram, C., Klopfleisch, R., Maier, A. (2018). SlideRunner. In: Maier, A., Deserno, T., Handels, H., Maier-Hein, K., Palm, C., Tolxdorff, T. (eds) Bildverarbeitung für die Medizin 2018. Informatik aktuell. Springer Vieweg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-56537-7_81.

[9] Bankhead, P., Loughrey, M.B., Fernández, J.A. et al. "QuPath: Open source software for digital pa-thology image analysis", Sci Rep 7, 16878 (2017), https://doi.org/10.1038/s41598-017-17204-5.

[10] Marques Godinho T, Lebre R, Silva LB, Costa C. "An efficient architecture to support digital pathology in standard medical imaging repositories", Journal of biomedical informatics vol. 71 (2017): 190-197, doi:10.1016/j.jbi.2017.06.009.

[11] A.C.R. Nema. "Digital imaging and communications in medicine (DICOM)", Supplement 145: Whole Slide Imaging in Pathology, ed.: Part, 2009.

[12] Dimitriou Neofytos, Arandjelović Ognjen, Caie Peter D. "Deep Learning for Whole Slide Image Analysis: An Overview", Frontiers in Medicine, 6, 2019, 2296-858X.

[13] Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, Lujan GM, Molani MA, Parwani AV, Lillard K, Turner OC, Vemuri VNP, Yuil-Valdes AG, Bowman D. "Introduction to Digital Image Analysis in Whole-slide Imaging: A White Paper from the Digital Pathology Association", Journal of pathology, 2019, Apr 24;10:15. PMID: 30984469; PMCID: PMC6437786.

[14] Barisoni, L., Lafata, K.J., Hewitt, S.M, Anant Madabhushi, Ulysses G.J. Balis. "Digital pathology and computational image analysis in nephropathology", Nat Rev Nephrol 16, 669–685 (2020), https://doi.org/10.1038/s41581-020-0321-6.

[15] M. Veta, J.P.W. Pluim, P.J. van Diest, M.A. Viergever. "Breast Cancer Histopathology Image Analysis: A Review", IEEE Trans. Biomed. Eng., 61 (5) (2014), pp. 1400-1411, 10.1109/TBME.2014.2303852.

[16] Xiao Jian Tan, Nazahah Mustafa, Mohd Yusoff Mashor, Khairul Shakir Ab Rahman. "A novel quantitative measurement method for irregular tubules in breast carcinoma", Engineering Science and Technology, an International Journal, Volume 31, 2022, 101051, ISSN 2215-0986, https://doi.org/10.1016/j.jestch.2021.08.008.

[17] Ajay Basavanhally, Elaine Yu, Jun Xu, Shridar Ganesan, Michael Feldman, John Tomaszewski, and Anant Madabhushi. "Incorporating domain knowledge for tubule detection in breast histopathology using O'Callaghan neighborhoods", Proc. SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis, 796310 (4 March 2011); https://doi.org/10.1117/12.878092.

[18] Y. Li et al. "Automated Gleason Grading and Gleason Pattern Region Segmentation Based on Deep Learning for Pathological Images of Prostate Cancer", IEEE Access, vol. 8, pp. 117714-117725, 2020, doi: 10.1109/ACCESS.2020.3005180.

[19] Aperio format [internet].: Openslide. c2021 [cited 2021 Dec 21], Available from: https://openslide.org/formats/aperio/.

[20] Aperio Digital Pathology Slide Scanners [internet].: Leica Biosystems Nussloch GmbH 2021, c2021 [cited 2021 Dec 21], Available from: https://www.leicabiosystems.com/digital-pathology/scan/.

[21] Geojs [internet].: Kitware. c2021 [cited 2021 Dec 21], Available from: https://opengeoscience.github.io/geojs/.

[22] Elston CW, Ellis IO. "Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer:experience from a large study with long-term follow-up", Histo-pathology 19(5): 403–410, https://doi.org/10.1111/j.1365-2559. 1991. tb00229.x.

[23] Roychowdhury M. Tubular [Internet]. PathologyOutlines.com website. [Last Accessed June 25, 2022], https://www.pathologyoutlines.com/topic/breastmalignanttubular.html.

[24] Rajasekaran S, Rohit Tapadia, Devika Rubi R, Aadithya Pabbisetty. "ISUP grading for prostate cancer pathology images using deep learning", International Journal of Medical Science and Current Research (IJMSCR), Vol.4, (5) (2021), 01-09.

Visually Impaired Person Assistance Based on Tensor FlowLite Technology

Nethravathi B[1], Srinivasa H P[2], Hithesh Kumar P[3], Amulya S[4], Bhoomika S[5], Banashree S Dalawai[6], Chakshu Manjunath[7]

Department of Information Science and Engineering, JSS Academy of Technical Education,
Bangalore-60, Karnataka, India[1, 4, 5, 6, 7]
Department of Computer Science and Engineering, T. John Institute of Technology, Bangalore-83, Karnataka, India[2]
Department of Computer Science and Engineering, JSS Academy of Technical Education, Bangalore-60, Karnataka, India[3]

Abstract—The most exciting thing about computer visualization is to detect a Real time object application system. This is abundantly used in many areas. With the more increase of development of deep learning such as self-driving cars, robots, safety tracking, and guiding visually impaired people, many algorithms have improved to find the relationship between video analysis and images analysis. Entire algorithms behave uniquely in the network architecture, and they have the same goal of detecting numerous objects in a composite image. It is very important to use our technology to train visually impaired people whenever they need them, as they are visually impaired and limit the movement of people in unknown places. This paper offers an application system that will identify all the possible day-to-day objects of our surroundings, and on the other side, it promotes speech feedback to the person about the sudden as well as far objects around them. This project was advanced using two different algorithms: Yolo and Yolo-v3, tested to the same criteria to measure its accuracy and performance. The SSD_MobileNet model is used in Yolo Tensor Flow and the Darknet model is used in Yolo_v3. Speech feedback: A Python library incorporated to convert statements to speech-to-speech. Both algorithms are analyzed using a web camera in a variety of circumstances to measure the correctness of the algorithm in every aspect.

Keywords—Tensor flow; SSD; Yolo; Yolo_v3; gifts; deep learning

I. INTRODUCTION

Human optical coordination is extremely precise and can handle multitasking even with unconscious notice. If you have a large amount of data, you need a more perfect system to appropriately detect and identify multiple objects at once. You can use better algorithms to train your computer to recognize multiple objects in an image with high precision. Object recognition is the utmost difficult application of computer visualization because this requires a comprehensive understanding of the image. This system was developed using a laptop, webcam, and Bluetooth headphones. TensorFlow is used to develop intelligent object detection algorithms. Models of IoT, embedded, and mobile devices are run by using tensor flow lite. It consists of low latency and low binary size, making it easy to design devices at the edge of your network. This advances potential, connectivity, and confidentiality. An unlimited length of speeches can be read by using google translate text to speech API which is a python library interface of google text to speech (gtts).

1) Deep-Learning: Deep learning comes under a branch of machine learning, a multi-tiered semantic network. These try to mimic the behavior of the human brain, permitting it to learn from the vast amount of data, even distant beyond its capabilities [1]. Single-layer semantic networks can provide limited estimates, but additional hidden layers help improve placement and accuracy.

2) Tensor flow lite: A version of TensorFlow is tensor flow lite. Tensor flow lite is used because it can run TensorFlow models on IoT, embedded, and mobile devices. TensorFlow Lite has low latency and a small binary size, which makes it easy to design devices on the network. It advances expectancy, confidentiality, and connectivity. Google Text-to-Speech (gtts), a Python library interface that uses the Google-Translate Text to Speech API. This can read language objects of infinite length. The system accurately detects objects up to 12-13 feet away.

3) Tensor flow framework: It is an open-source Machine learning framework used by many applications developed by Google using Google datasets. TensorFlow develops datasets for many applications, has a flexible architecture that is written in Python or C++, and has various available frameworks such as Keras, Theano,Caffe, and Torch. Among these, Tensor Flow provides robust machine learning and easy model building. TensorFlow Lite is used for embedded mobile IoT models. Tensor flow lite has two important components as listed below:
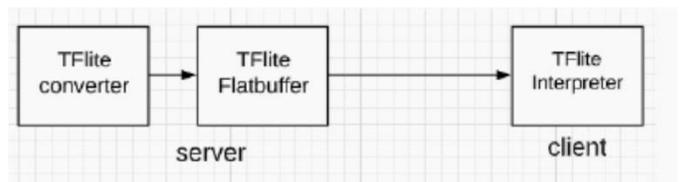
Fig. 1. Tensor Flow Components.

- Interpreter
- Converter

The interpreter is one which makes calculations based on the input data which have been given, and the Converter creates a flat buffer file (see Fig. 1). This file is given by the client's device. This uses the TensorFlow Lite Interpreter file internally. TensorFlow lite supports APIs in numerous languages such as Objective-C, C ++ and Python Java, Swift. Model optimization tools of tensor flow lite reduce the size of the model and improve performance by not losing its accuracy. TensorFlow Lite features an effective flat buffer model format optimized for small size and portability. Flat Buffer is an effective cross-platform serialization library for Rust, C #, Java, JavaScript C ++, and Typescript. Storage-efficient Flat Buffer trains algorithms to accelerate datasets, classify the data, and predict results precisely. When the input data is entered into the model, the weights are adjusted until the model fits properly. To avoid over fitting in the model cross-validation process is done.

4) Yolo V3: Yolo v3 You Look Only Once is the algorithm proposed by Redmond compared to the method used in the detection algorithms before YOLO V3. This is also useful for discovery bind boxes and class dividers that instantly capture the potential of a class (see Fig. 2). YOLO V3 takes a very different approach to object detection, providing up-to-date results that are superior to other real-time, large margin detection algorithms. Each of these N grids is responsible for content ingestion and retrieval. In parallel, these grids predict binding box bindings in terms of cell bindings [11]. This process handles both cell detection and detection from a single image. This saves a lot of computational effort because multiple cells predict the same by predicting different merge boxes. Duplicate forecasts are generated.
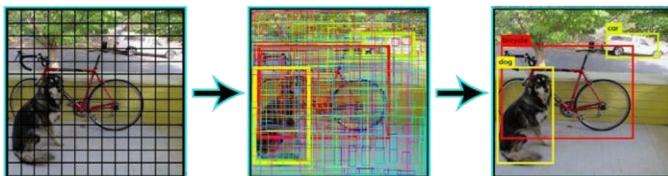
Fig. 2. Working of Yolo V3.

5) CNN: Convolution Neural Network is a concept of deep-learning algorithm that can capture an image and assign a value (readable weight and bias) to the various elements /elements in the image and be able to distinguish one from the other. The pre-processing required for a convolution network is very low when compared to other partition Algorithms. Earlier times, filters were used to be made by hand with adequate training convolution networks that can read these filters and symbols.

The organization of the paper starts with the literature survey, where total 12 papers were analyzed with merits and demits. Methodology is explained with system architecture and use case diagram, followed by results and discussion.

II. LITERATURE SURVEY

Literature review aimed to look at various learning procedures and see if they could be applied to our use case. In [1], it has been proposed that humans use their eyes and mind to see and perceive the playing field and surrounding objects. In blind people, these talents are misplaced or impaired to varying degrees. Their eyes cannot perform the duty of sight. This paper is about designing and implementing a portable utility that helps people accurately determine their distance from them without seeing their environment. The proposed device is equipped with a CNN-based real-time data acquisition technology known as YOLO (You Look Only Once) and a separate digital camera on a Raspberry Pi board. The system additionally measures inference space and logically promises these statistics to the visually impaired. In [2], a geometry-based approach evaluates the entire 3D model. This way there is no need to isolate (or split) the object of interest. This system does not need to isolate the object of interest behind the surrounding. Compared to other methods, the proposed system is tested with real data collected from places such as kitchens or restaurants. The work proposed in this review meets the requirements of high accuracy, execution time, and compliance with the VIP criteria. An analytical data set has been published. In [3], this program does not need to isolate the object of interest behind the surrounding compared to other methods, the proposed system is tested with real data collected from places such as kitchens or restaurants. The work proposed in this review meets the requirements of high accuracy, execution time, and compliance with the VIP criteria. An analytical data set has been published. This system predicts his or her distance for approximate accuracy level. In [4], humans use their eyes and mind to see and perceive the playing field and surrounding objects. In blind people, these talents are misplaced or impaired to varying degrees. Their eyes cannot perform the duty of sight. This article is about designing and implementing a portable utility that helps people accurately determine their distance from them without seeing their environment. The proposed device is equipped with a CNN-based real-time data acquisition technology known as YOLO and a separate digital camera on a Raspberry Pi board. The gadget additionally measures inference space and logically promises these statistics to the visually impaired. In [5], in most cases, ongoing support is needed in almost all situations, especially in daily work. Other issues include difficulty recognizing people and detecting obstacles. To calculate this avoidance, this combines many of the available technologies and is integrated into one device with many features that you can use, for blind people. It takes less time to train the model. In [6], its functionality is easily compromised by creating complex training that incorporates many small image elements with more quality content object finders and scene separators, standard acquisition properties, and other modifications, improving acquisition function. Briefly review specific activities, including key discovery, facial recognition, and pedestrian detection. In [7], here the system gathers all information on motionless cameras to detect moving objects in digital videos. The system is developed based on the visual flow rate and usage as well as the combination. Object limitations are set using blob analysis; to obtain the audios, a

central filter is used and the objects which are needed are removed using thresholding algorithms for natural operation. The accuracy is less than 65%. For this purpose, they suggest that 3D counterparts that are not specific to the most widely used 2D point detectors are insufficient, and suggest another option to support these points of interest, developing a data-based awareness algorithm with a spatial window temporarily. It supports only the static data set and performance is less than 65% K-Means. An algorithm with auto encoders is used. In [9], the experimental data, they achieved higher error rates of 1 and 5 of 37.5% and 17.0%, which is much better than the prior art. Some of these are surveyed by a completely linked layer with a maximum level of integration and a final 1000 Softmax. This used unsaturated neurons to speed up training and more efficient GPUs for convolutional performance. In the experimental data, they achieved higher error rates of 1 and 5 of 37.5% and 17.0%, which is much better than the prior art. A top-five test score of 15.3%, compared to 26.2% obtained by the top runners-up. In [10] this work, they introduced the Regional Proposal Network (RPN) which shares the features of complete image conversion and acquisition network that allows most cost-effective regional proposals. Regional Proposal Network is an entirely synchronized network that predicts object parameters and resistance scores in each area. RPN trained completely to produce high-quality regional proposals, which is used from

Fast R-CNN for recognition. The RPN shows the integrated network where it is headed. Accuracy is high but large number of datasets need to be trained, so it consumes more time and space and computation power. The code has been made public. In [11], test results show the 7-bounds of conversion, the system can provide excellent real-time performance. This shows a 7 layer YOLO with an 11x11 grid cell that can accurately detect small people and vehicles. This provides accuracy and real-time performance and can only identify the person and car. In [12-15], use background-for-back classification methods to extract motion content and produce embedded videos. Then obtained directional vectors based on silhouette computations. They also used the flexible aspect of human movement to make smooth decisions over time and minimize errors in job perception. Our monocular method tolerates moderate viewing changes and can be used for both advanced and side views for most functions. This method does not do any small sample but instead is effective for all small windows. This improves the overall accuracy of the entire detector, it is not supported for other compound scoring functions. By using kernels, and by using this approach they demonstrated the great benefits of working on the three datasets which are publicly available. Surprisingly, they showed that a single solid HOG filter can exceed the modern partially crippled component. The detailed analysis of existing work is depicted in Table I.

TABLE I. THE DETAILED ANALYSIS OF EXISTING WORK

Reference paper No.	Special Features	Merits	Demerits	Conclusion and Future Work
[1]	The proposed MedGlasses system includes a couple of wearable glasses, (AI)- primarily based intelligence drug tablet box, a cellular tool app, and a cloud-primarily based facts control platform. The results shown that recognition accuracy of 95.1% can be achieved.	The improves the safety of medication for the visually impaired people	1.It is limited to only medicines. 2.It is not suitable for other objects.	This system efficiently moderates the problem of drug connections that are caused by taking wrong drugs, hence it decreases the cost of medical management and provides visually impaired patients with a safe medicinal facility.
[2]	This proposes a different framework for the detection of the objects in our daily activities and this system provides its related information which existing such as size, and security path for grasping on a surface. It contains pipelines which are the combinations of a series of point cloud representation and table plane detection, objects detection, and the full model estimation through a robust system.	It has met the requirements Of high precision, performing time, and correctness.	It requires an enormous amount of data to train. Accuracy is not up to the mark	In this framework, they considered the advantages of deep learning (e.g. RCNN, YOLO) that it can be an efficient method for exploration tasks, and the geometry-based approach Model This approach does not require separation or segmentation of the object of interest from the background.
[3]	The proposed system used YOLO methodology and a system with a single camera which is attached to a Raspberry Pi board also estimates the distance of the object and detects it and later sends this data to the blind person the method of audio feedback.	It gives accuracy above 98.8%.	We need large pieces of data to train the model.	This system concluded that it can detect the objects with its distance the accuracy of 98.8%
[4]	This application was developed using the features of the Python and OpenCV libraries, and this eventually ported to the Raspberry Pi3 Model and B+ platform. It has high Processing as well as includes the Extended identification of far distance objects accurately	It consumes less time to train the model.	1.It is expensive. 2.Its processing speed is very less.	This proposed system addresses various optimum distances and objects with less value of image source measurement
[5]	This is a voice assistance system to guide visually blind people in their daily activities. The device used is the various combination of technologies and consolidates all of these into a solitary multipurpose device that is used by the blind people	It consumes less time to train the model.	It is expensive. Its processing speed is very less.	This system discusses the different systems and algorithms for training models to achieve the accuracy of the detection.
[6]	This model behaves differently in network construction, learning strategies, optimization features, etc. it also	Improves the detection	specifies the object. It is not suitable to	It concludes several detection tasks including salient object detection, face

	provides an overview of the deep learning-based object detection environment.	performance further.	detect all the objects.	recognization, traffic signal recognization, etc.
[7]	Here the system gathers all information on motionless cameras to detect moving objects in digital videos. The system is developed based on the visual flow rate and usage as well as the combination. To obtain the audios, a central filter is used and the objects which are needed are removed using thresholding algorithms for natural operation.	Applied many algorithms to detect the object and trained using many algorithms	This is used to detect and track moving objects but accuracy is less than 65%.	It is concluded that the proposed system will accurately detect and track the correct moving objects in moving video.
[8]	In this work it has been developed an allowance of ideas of spatiotemporal case Due to this reason, it showed the direct 3D counterparts to frequently used 2D interest point detection	K-Means Algorithm with autoencoders	It supports only the static data set and performance is less than 65%.	Devised the recognition algorithm related to spatiotemporal data It resulted in recognization based on the variety of the dataset with a large amount of data
[9]	In the experimental data, they achieved higher error rates of 1 and 5 of 37.5% and 17.0%, which is much better than the prior art. Some of these are surveyed by a completely linked layer with a maximum level of integration and a final 1000 Softmax.	achieved in between the top 5 error rates of 37.5% and 17.0%	It needs more amount of data and it consumes more time to train the model.	Used the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the other system error rate.
[10]	used Region Proposal Network which shares full-image convolutional features with the detection network, so it enables nearly cost-free region proposals. The RPN is trained end-tend to create high-quality region proposals, which are used by Fast R-CNN for the detection of an object	Precisely identifies the object that is needed	Accuracy is a high but large number of datasets needed.	Simple, useful, and the code can be freely available Code is open-source with public availability.
[11]	This system used the methodology of YOLO and seven layers of convulsion network(CNN) The various size of grid cells are used for the processing of the object detection.	This system can provide excellent detection accuracy and real-time object detection	Trained only for less dataset.	The visual quality assessment using real-world images that shows the 7layer of YOLO with 11x11 grid cells correctly detects people and small vehicles
[12]	This system does not perform any of the sub-functions but it minimizes all of the MMOD which are used to progress any object identification method which is straightforward in the parameters, such as HOG or bag of visual word models.	MMOD optimizes the accuracy of the entire object detectors.	It does not support any complex functions, that uses the kernel	This approach shows significant performance improvement on three public datasets. Surprisingly it shows, that a single hard HOG filter can perform a deformable part model.

III. METHODOLOGY

A. Modules

1) Image Capturing: Proposed System consists of a camera, placed in the classroom to capture all the students. From these captured image frames, using the opencv and system will detect the student's face in the captured image using Haar cascade face detection technique. 2. Training phase: In the training phase we are applying CNN algorithm.3. Object recognition: i) Find object location: Object localization from the input video stream can be identified. Image classification from object detection, which goes through a ConvNet that results in a vector of features fed to a softmax to classify the object. Neural network have a few more output units that encompass a bounding box. In particular, we add four more numbers, which identify the x and y coordinates of the upper left corner and the height and width of the box (bx, by, bh, bw). ii) Find object movement: The Python's built in deque datatype to efficiently store the past N points the object has been detected and tracked at. Libraries imutils also used by collection of OpenCV and Python convenience functions. By checking the X, Y co-ordinate values and tracking them,

the object movement can be detected. This process is further explained in Fig. 3, a use-case diagram.
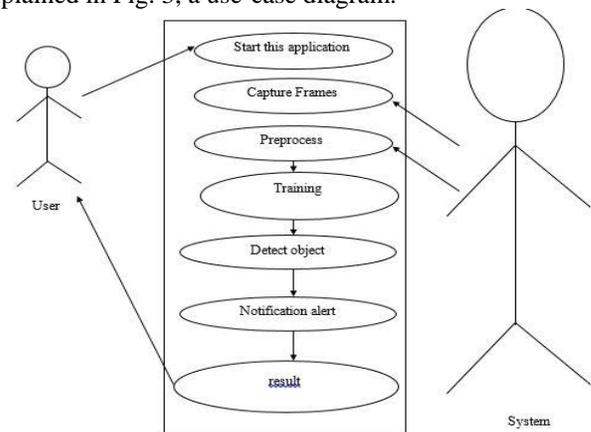
Fig. 3. Use-case Diagram.

2) System architecture: The architecture of Object detection system for blind people is described in Fig. 4. In this system input is real-time object through a camera, the captured image is stored for pre-processing, then the object is identified

through the captured image, Then the calculation of distance of the objects from the person and generates the audio signal for the identified objects, transfers the audio signal in the form of audio feedback. When user starts the application, the system captures the frame, the object capture within the frame undergoes for pre-processing in the system. After preprocessing, the system is trained to detect object, the object within the frame is detected, if multiple object is found within the frame, then the distance between those objects are also detected. The notification alert of those detected object is sent as a result to the user. So, the user can be aware of any obstacles in their path. By using this application even the blind people live their lives independently.

Fig. 6.   Multiple Object Detection.

Fig. 7.   The Graph of Model Accuracy.

The above Fig. 7 shows the graph of model accuracy in which x-axis is represented as epoch involved in training and y-axis is represented as accuracy. The graph shows the accuracy for 200 epoch in this dataset that has been divided as train and validation data. This graph shows the overall model accuracy of our implementation, that as the epoch increases accuracy increases.

Fig. 4.   System Architecture.

IV.   RESULTS AND DISCUSSIONS

In the above Fig. 5(a) shows that the person is captured in a frame the system detects the object within the frame then it sends audio feedback as person and also even in the form of text in the figure the person is moved to right so the system sends the audio and text feedback has person is move to right. The object is captured within the frame of 0.8769 pixel. Similarly in the Fig. 5(b) the object known as cat is captured within the frame of 0.6921 pixel and in the Fig. 5(c) the object dog is captured within the frame of 0.8645 pixel.

When multiple objects are detected within the frame the frame is clustered for each object then the system calculates the distance between the each object. In the above Fig. 6, objects such as person, cat and cell phone are captured within the frame. The system calculates the distance between each object and sends the result in the form of audio and text in console feedback.

Fig. 8.   The Graph of Model Loss.

The above Fig. 8 shows the graph of model loss in which x-axis is represented as epoch involved in training and y-axis is represented as loss. The graph shows the loss for 200 epoch in this dataset has been divided as train and validation data this graph shows the overall model loss of the implementation that as the epoch increases, loss reduced.

(a)   (b)   (c)

Fig. 5.   (a) Person (b) Cat (c) Dog.

V. CONCLUSION

Developed a desktop application using python, the system automatically detects the objects and makes the voice alert to blind people. Object detection is used to search objects in the real world in images of the world, often found in blind scenes. It depends on the location and cameras that are used to detect the object. This architecture is proven and is reliable for blind people. Since the use of tensor flow lite, its response time is quick and even it requires small power, making it suitable for transferable applications. There are many things in the data set, and all of these are basic things that anyone needed in their daily activities. The biggest attraction of this work is that the output signal is in the voice listening format and blind people make use of headphones which makes it easy to guide visually impaired persons, so the overall accuracy rate is 98% which is better than previous implementations. Thus it allows users to become self-reliant without having to seek help.

REFERENCES

[1] W. Chang, L. Chen, C. Hsu, J. Chen, T. Yang, and C. Lin, "MedGlasses: A Wearable Smart-Glasses-Based Drug Pill Recognition System Using Deep Learning for Visually Impaired Chronic Patients," in IEEE Access, vol. 8, pp. 17013-17024, 2020.

[2] V. Le, H. Vu and T. T. Nguyen, "A Frame-work assisting the Visually Impaired People: Common Object Detection and Pose Estimation in Surrounding Environment," 2018 5th NAFOSTED Conference on Information and Computer Science (NICS), Ho Chi Minh City, 2018, pp. 216-221s.

[3] S. Duman, A. Elewi and Z. Yetgin, "Design and Implementation of an Embedded Real-Time System for Guiding Visually Impaired Individuals," 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey, 2019, pp. 1-5.

[4] L. Tepelea, I. Bucio, C. Grava, I. Gavrilut and A. Gacsádi, "A Vision Module for Visually Impaired People by Using Raspberry PI Platform," 2019 15th International Conference on Engineering of Modern Electric Systems (EMES), Oradea, Romania, 2019, pp. 209-212.

[5] I. Joe Louis Paul, S. Sasirekha, S. Mohanavalli, C. Jayashree, P. MoohanaPriya and K. Monika, "Smart Eye for Visually Impaired-An aid to help the blind people," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-5.

[6] P. Vyavahare and S. Habeeb, "Assistant for Visually Impaired using Computer Vision," 2018 1st International Conference on Advanced Research in Engineering Sciences (ARES), Dubai, United Arab Emirates, 2018, pp. 1-7.

[7] O. Stephen, D. Mishra, and M. Sain, "Real-Time object detection and multilingual speech synthesis," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-3.

[8] M. A. Khan Shishir, S. Rashid Fahim, F. M. Habib, and T. Farah, "Eye Assistant: Using a mobile application to help the visually impaired," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-4.

[9] C. Jayawardena, B. K. Balasuriya, N. P. Lokuhettiarachchi, and A. R. M. D. N. Ranasinghe, "Intelligent Platform for Visually Impaired Children for Learning Indoor and Outdoor Objects," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 2572-2577.

[10] S. Pehlivan, M. Unay, and A. Akan, "Designing an Obstacle Detection and Alerting System for Visually Impaired People on Side". Medical Technologies Congress (TIPTEKNO), 2019.

[11] M. Putra, Z. Yussof, K. Lim, and S. Salim "Convolutional neural network for the person and car detection using Yolo framework", Journal of Telecommunication, Electronic and Computer Engineering, Vol.10, PP 67-71.

[12] M. Singh, A. Basu and M. K. Mandal, "Human Activity Recognition Based on Silhouette Directionality," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 9, pp. 1280-1292, Sept. 2008, doi: 10.1109/TCSVT.2008.928888.

[13] Ghaith Al-refai and Mohammed Al-refai, "Road Object Detection using Yolov3 and Kitti Dataset" International Journal of Advanced Computer Science and Applications(IJACSA), 11(8), 2020.

[14] Mohana and HV Ravish Aradhya, "Object Detection and Tracking using Deep Learning and Artificial Intelligence for Video Surveillance Applications" International Journal of Advanced Computer Science and Applications(IJACSA), 10(12), 2019.

[15] Hafsa Ouchra and Abdessamad Belangour, "Object Detection Approaches in Images: A Weighted Scoring Model based Comparative Study" International Journal of Advanced Computer Science and Applications(IJACSA), 12(8), 2021.

Design of Robust Quasi Decentralized Type-2 Fuzzy Load Frequency Controller for Multi Area Power System

Jesraj Tataji Dundi[1], Rama Sudha
Kasibhatla[3]
Department of Electrical Engineering
AUCOE (A), Andhra University
Visakhapatnam, India[1, 3]

Anand Gondesi [2]
Department of Electrical and
Electronics Engineering
Dr.L.B College of Engineering
Visakhapatnam, India

A. Chandrasekhar [4]
Department of Engineering
Mathematics
GITAM University
Visakhapatnam, India

Abstract—**Interconnected power systems receive power through tie lines. Sudden perturbation in load causes uneven power distribution issues resulting in sudden changes in the voltage and frequency in the given system (tie-line power exchange error). The Load Frequency Controller (LFC) has an ability to stabilize the system for the above mentioned disturbances. In this paper, a novel load frequency controller based on Type-2 Fuzzy Quasi-Decentralized Functional Observer (T2FQFO) is proposed. In the proposed methodology the observer gains are obtained mathematically which guarantees, the stability of the system. The efficacy of the proposed technique has been tested on an IEEE standard testing systems. The results shows the proposed T2FQFO has higher performance when compared with Fuzzy Quasi-Decentralized Functional observer, Quasi- Decentralized Functional observer and classical state observer. And the results say that the peak over shoot and settling time have been improved by 25% (Approx.) by Type-2 Quasi Decentralized Functional Observer ((T2FQFO) than other observers.**

Keywords—Load frequency control; type-2 fuzzy quasi-decentralized functional observer; fuzzy quasi-decentralized functional observers; state observer; type-2 fuzzy controller

I. Introduction

State Estimation and its analysis constitute a fundamental aspect of science and engineering. Information about the internal state of a given system can be retrieved simply by measuring the values of inputs and outputs to the system, under this process. The term system state stands for the internal condition of a system at any given instant of time. Dynamic State Estimation (DSE) [7] refers to analysis of internal characteristics that trigger changes in the given system. These are majorly applicable to most engineering and science disciplines (viz. Electrical, Electronics, Civil, Mechanical, Aerospace, Chemical, etc. [13].).

In Control System, an Observer refers to a sub-system that is employed for estimating the internal state/condition of another system (of which it is a part) solely by analyzing the input and output to the system. This idea was first introduced by D. Luenberger in 1966 [21]. Some of the commonly used observers are state observer, functional observer, sliding observer, bounding observer [4]. The terms Observer and State Observer are almost synonymous and are implemented using computers in industrial applications.

Functional observers on the other hand, work using probabilistic/statistical approaches. They are usually available in reduced forms. For LTI (linear time invariant) systems, [3] such observers can still be designed even if the system becomes sun-observable. The paper primarily discusses the application of the aforesaid technique for power systems where exposure to disturbances like sudden changes in the load, modification in system configuration, loss of transmission lines, generator failures, etc.[16][17] are common phenomena.

In this system, there is every possibility of random values creeping into the measured data set. Nevertheless, the state estimation technique being probabilistic in nature can be carried out with random constituents. Hence, probabilistic or set-membership approaches are preferred for data analysis [9]. The common schemes include Method of Moments Estimation, Minimum Variance Unbiased Estimator (MVUE), Maximum a Posterior (MAP), Maximum Likelihood Estimation (MLE), Bayes Least Squared Error (BLSE), Bayes Estimation, Nonlinear System Identification, Wiener Filter, Kalman Filter, Particle Filter [2][6].

Collection of various sub-systems that facilitate power supply or transfer is referred to as a power system. Multiple units/areas in a power system connected via transmission lines are known as interconnected power systems. State observers may be used to measure the magnitude of the transmitted power, the transmission lines being used are commonly known as tie-lines [5]. All units/areas must have identical line frequencies. Load frequency control (LFC) [18] method prevents deviations of frequency and tie-lines for all areas/units under steady state. The main objective of this paper is a comprehensive study of linear and non-linear dynamic state estimation techniques for fault detection and LFC using PMU's [19]. A modern power system is constituted by various types of generators, transmission networks, measuring devices [8], etc. with properties similar to a general power system.

The paper describes the work (methodology) in Section II, the results obtained and the description in Section III and the conclusion in Section IV below.

A. Literature Review

M. Darouach et al. [1] introduced a simple method to design a full-order observer for linear systems with unknown inputs provided with the necessary and sufficient conditions for the existence of the observer. This method reduces the design procedure of full-order observers with unknown inputs to a standard one where the inputs are known. The existence conditions are given, and it was shown that these conditions are generally adopted for unknown inputs observer problem.

Chuang Liu, Hak-Keung Lam et al. [9] in their paper the authors investigated the stability of a Takagi-Sugano fuzzy model-based (FMB) functional observer control system. If the state-feedback control cannot measure the system state, the fuzzy function observer is designed to estimate the control input instead of directly estimating the system state. A fuzzy function observer can reduce the observer order, which determines a large number of observer gains.

Therefore, we propose a new form of fuzzy function observer that facilitates stability analysis so that the gain of the observer can be obtained numerically while guaranteeing stability. The proposed form also shows using the separation principle to design the fuzzy controller and the fuzzy function observer separately. To design fuzzy controllers with system stability in mind, we use higher derivatives of the Lyapunov function (HODLF) to reduce the conservativeness of the stability condition. HODLF generalizes the commonly used first derivative.

By exploiting the properties of the membership functions and the dynamics of the FMB control system, convex and relaxed stability conditions can be derived. A simulation example is provided to demonstrate the feasibility of the proposed relaxation of the stability constraint and the designed fuzzy function observer controller. Based on the proposed fuzzy function observer, operators can easily achieve stable observer gains. To confirm efficacy and mitigation, LIU et al. designed a fuzzy function observer controller for the nonlinear system 1639 via Holdf.

Further a more advanced techniques can be applied to meet the boundary conditions of the derivation of membership functions. And by extending the technique in discrete-time linear functional observer, the discrete-time fuzzy functional observer can also be investigated.

M. Darouach et al. [13], presented a simple and straightforward method to design full and reduced order observers for linear time-invariant descriptor systems and these are presented in this paper. The approach for the reduced order observer design is based on the generalized Sylvester equation. Sufficient conditions for the existence of the observers are given. An illustrative example is included. The reduced-order observer design method is based on the new resolution method of the constrained generalized Sylvester equation. It was shown that the existence conditions of the observer generalize those adopted in D. N. Shields et al. for the observer design of square descriptors systems problem. An extension to a less restrictive conditions is under study.

K. Rama Sudha et al. [16] have described a method based on the type 2 fuzzy system 'T2FS'. For LFC (Load Frequency Control) in power systems including Superconducting Magnetic Energy Storage Units (SMES) of two-area connected reheat heating systems. Therefore the author proposed a Type 2 (T2) fuzzy approach for load frequency control in a two-section connected reheat power plant, considering the Generation Rate Constraint (GRC), Boiler Dynamics (BD), and SMES. A distinct advantage of this controller is its lack of sensitivity to large load changes and system parameter variations, even in the presence of nonlinearities. The proposed method is tested in a dual-range power system to demonstrate its robust performance under different ranges of load changes. We compare the performance of a type 2 (T2) fuzzy controller with an optimal PID controller (Khamsum's optimal PID) and a fuzzy PI controller (type 1 fuzzy) controller in the presence of GRC, BD, and SMES. Simulation results confirm the high robustness of the proposed small-power-capacity SMES controller against various disturbances and system uncertainties compared with SMES in previous studies.

II. Power System DSE and LFC using Functional Observers

Application of Quasi decentralized functional observer (QDFO) to two area inter-connected linear power system with a single tie-line model has been depicted [11]. Development of Functional Observer (FO) [3] model for highly interconnected power systems have been under taken and is presented. The LFC signal has been generated here using a control signal. The signal needed is acquired directly from the functional observer (FO) instead of forming a control signal out of a linear combination of individual state signals. Tie-line power measurements, voltage magnitudes along with phase angles and current measurements combined with current voltage measurements of the PMU are employed for the Quasi decentralized functional observer. Power system state is primarily analyzed using an estimation algorithm, 'observability' analysis, bad data recognition etc. [7]. The basic idea and its analysis were formulated in this paper, under five corollaries.

A. Corollary -I

It might not be possible to measure all state variables under practical circumstances either due to complications related to data acquisition or cost issues. As a result, for all states of the state vector, a state estimation technique is required to obtain state feedback [19].

Let $\hat{x}(t)$ = estimate of the state vector x(t)

State Observer/Estimator

$$\dot{x} = \alpha x + \beta u \tag{1}$$

$$y = \gamma x \tag{2}$$

Here x serves as the state vector, y is the output (which includes an estimation of the state vector x).

The control signal is represented by 'u'.

Open-loop Observer (L= [0], $\alpha obs=\alpha$)

$$\dot{\hat{x}}(t) = \alpha \hat{x}(t) + \beta u \tag{3}$$

Failure of the observer takes place due to disturbances and modeling error.

Open-Loop Estimation Error

$$\tilde{x}(t) = x(t) - \hat{x}(t) \tag{4}$$

$$\dot{\tilde{x}}(t) = \dot{x}(t) + \hat{\dot{x}}(t) = \alpha\tilde{x}(t) \tag{5}$$

Hence,

$$\tilde{x}(t) = e^{\alpha t}\tilde{x}(0) \tag{6}$$

Error Dynamics:

There is no rectification for modeling imperfections, and state matrix α is unstable and has unbounded error.

$$\dot{x} = \alpha\hat{x} + \beta u + \mathcal{L}(y - \gamma\hat{x}) = \alpha\hat{x} + \beta u + \mathcal{L}y \tag{7}$$

$$\alpha_{obs} = \alpha - \mathcal{L}\gamma \tag{8}$$

Where $\mathcal{L}$ is the observer gain

Eigen values are assigned to the matrix $\mathcal{L}$. Appropriate estimation requires the equation to be reduced in terms of an open loop observer as shown in the Fig. 1.
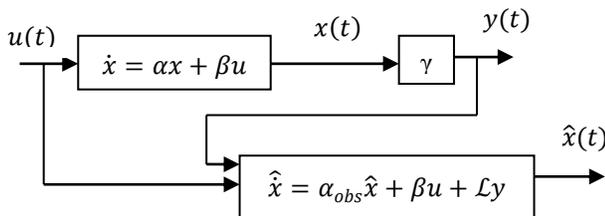
Fig. 1. Block Diagram of Observer.

Theorem1:

The system (γ, α) is observable if and only if the dual system (α^T, γ^T) is controllable [19].

Eigen values could be arbitrarily assigned (stable) by using state feedback provided (α^T, γ^T) is controllable.

$$\dot{x} = \alpha^T x + \gamma^T u \tag{9}$$

$$u = -\mathcal{L}^T x \tag{10}$$

$$\dot{x} = (\alpha^T - \gamma^T\mathcal{L}^T)x \tag{11}$$

$$(\alpha^T - \gamma^T\mathcal{L}^T)^T = \alpha - \mathcal{L}\gamma \tag{12}$$

Similar Eigen values are obtained.

B. *Corollary -II*

In application to the load frequency control, Functional Observability is better than State Observability [20].

Functional observer can be stated as follows.

$$\dot{x}(t) = \alpha x(t) + \beta u(t) \tag{13}$$

$$y(t) = \gamma x(t) \tag{14}$$

$$z(t) = \mathcal{L}x(t) \tag{15}$$

Here $x(t) \in R^n, u(t) \in R^m$ are the state input vectors, $y(t) \in R^p$ is the state output vector and $z(t) \in R^r$ is the vector

to be estimated. $\alpha \in R^{nxn}$, $\beta \in R^{nxm}$, $\gamma \in R^{pxn}$ and $\mathcal{L} \in R^{rxn}$ are matrices of known constants. A functional observer should be a dynamic system that should be able to track z(t) asymptotically. It is theorized to have the following structure:

$$\dot{w}(t) = \eta w(t) + \lambda y(t) + \mu u(t) \tag{16}$$

$$\hat{z}(t) = gw(t) + \text{€}y(t) \tag{17}$$

α, β and γ are the system matrices and η, λ, μ, D and € are the observer matrices and which are defined as:

$\alpha \in R^{nxn}$ $\qquad$ $N \in R^{qxq}$

$\beta \in R^{nxm}$ $\qquad$ $\lambda \in R^{qxp}$

$\gamma \in R^{pxn}$ $\mu \in R^{qxm}$

$\mathcal{L} \in R^{rxn}$ $D \in R^{rxp}$ $\quad$ $\text{€} \in R^{rxp}$

Theorem 2:

If and only if the following conditions are met, the completely observable qth order functional observer of (16) and (17) will estimate Lx(t).

- A stability matrix,η

$$\lambda\gamma = P\alpha - \eta P \tag{18}$$

$$\mu = P\beta \tag{19}$$

$$\mathcal{L} = DP + \text{€}\gamma \tag{20}$$

Observer error in state estimation can be defined as

$$e(t) \triangleq w(t) - Px(t) \tag{21}$$

Derivation and upon substitution of the system and observer equations, we get

$$\dot{e}(t) =$$
$$w(t) + P\dot{x}(t) = \eta w(t) + \lambda\gamma\dot{x}(t) + \mu u(t) - P\alpha x(t) - P\beta u(t) \tag{22}$$

Applying conditions (13) and (14) yields

$$\dot{e}(t) = \eta e(t) \tag{23}$$

The differential solution above is in the form of an exponential function

$$e(t) = e^{\eta t} \tag{24}$$

in which the observer dynamics are controlled by the variable N.

Applying these conditions, we get:

$$\lim_{t\to\infty} e(t) = w(t) - Px(t) \tag{25}$$

Further simplification, leads to the following we get

$$e_z(t) = \hat{z}(t) - \mathcal{L}x(t) = D(w(t) - Px(t)) \tag{26}$$

The above equation should reach zero asymptotically as expected α and η do not share common Eigen values but confirms that P will have a unique solution. Now X(t) and e(t) are easily derived by using above defined conditions.

$$\dot{x}(t) = \alpha x(t) + \beta u(t) = \alpha x(t) - \beta(Dw(t) + \text{€}y(t))$$
$$= (\alpha + \beta\mathcal{L})x(t) + (\beta D)e(t) \tag{27}$$

$$\dot{e}(t) = \eta e(t) \tag{28}$$

This results in a composite system similar to the full-state observer as

$$\begin{bmatrix} \dot{x}(t) \\ \dot{e}(t) \end{bmatrix} = \begin{bmatrix} \alpha + \beta\mathcal{L} & \beta D \\ 0 & \eta \end{bmatrix} \begin{bmatrix} x(t) \\ e(t) \end{bmatrix} \tag{29}$$

Aside from the fact that the control law, $\mathcal{L}x(t)$ is used instead of $-Kx(t)$. Though there is variation in notation, they are similar to each other. Judging the constraints involved, an rth order observer is required to achieve functional state of the system, where the order of the system r is as small as possible. The observer matrices should therefore guarantee the ease of Eigen value assignment and simplicity of the control algorithm so that it can be readily applied.

For order estimation, the matrix ranks are taken into consideration.

$$\text{rank} \begin{bmatrix} L\alpha \\ \gamma\alpha \\ \gamma \\ \mathcal{L} \end{bmatrix} = \text{rank} \begin{bmatrix} \gamma\alpha \\ \gamma \\ \mathcal{L} \end{bmatrix} \tag{30}$$

$$\text{rank} \begin{bmatrix} s\mathcal{L} - L\alpha \\ \gamma \\ \mathcal{L} \end{bmatrix} = \text{rank} \begin{bmatrix} \gamma\alpha \\ \gamma \\ \mathcal{L} \end{bmatrix} \quad s \in \gamma, R(s) \geq 0 \tag{31}$$

The condition is satisfied when the ranks on the LHS and RHS are equal. The author in [15] shows that this condition is equal to the delectability of the pair (F, g), where

$$F = \mathcal{L}\alpha\mathcal{L}^+ - \mathcal{L}\alpha(I - \mathcal{L}^+\mathcal{L}) \begin{bmatrix} \gamma\alpha(I - \mathcal{L}^+\mathcal{L}) \\ \gamma(I - \mathcal{L}^+\mathcal{L}) \end{bmatrix}^+ \begin{bmatrix} \gamma\alpha\mathcal{L}^+ \\ \gamma\mathcal{L}^+ \end{bmatrix} \tag{32}$$

$$g = \left(I - \begin{bmatrix} \gamma\alpha(I - \mathcal{L}^+\mathcal{L}) \\ \gamma(I - \mathcal{L}^+\mathcal{L}) \end{bmatrix} \begin{bmatrix} \gamma\alpha(I - \mathcal{L}^+\mathcal{L}) \\ \gamma(I - \mathcal{L}^+\mathcal{L}) \end{bmatrix}^+ \right) \begin{bmatrix} \gamma\alpha\mathcal{L}^+ \\ \gamma\mathcal{L}^+ \end{bmatrix} \tag{33}$$

Where, $\mathcal{L}+$ denotes the Moore-Penrose generalized inverse of matrix $\mathcal{L}$. Moreover, if matrices λ, μ and E satisfy Theorem 1, a Hurwitz matrix η is given by

$$\eta = F - Zg \tag{34}$$

For F-Zg to be stable, the Matrix Z is obtained by any pole placement method. Matrices E and K are obtained according to the equation,

$$[E \quad K] = \mathcal{L}\overline{\alpha}\Sigma^+ + Z(I - \Sigma\Sigma^+) \tag{35}$$

Where, $\overline{\alpha} = \alpha(I - \mathcal{L}^+\mathcal{L})$, $\overline{\gamma} = \gamma(I - \mathcal{L}^+\mathcal{L})$ and $\Sigma = \begin{bmatrix} C\overline{A} \\ \overline{C} \end{bmatrix}$

Matrix λ, μ are obtained according to

$$\lambda = K + \eta E \tag{36}$$

$$\mu = (\mathcal{L} - E\gamma)\beta \tag{37}$$

All of the required observer parameters can be easily computed using this algorithm, which results in a functional observer of form

$$\dot{w}(t) = \eta w(t) + \lambda y(t) + \mu u(t) \tag{38}$$

$$\hat{z}(t) = w(t) + Ey(t) \tag{39}$$
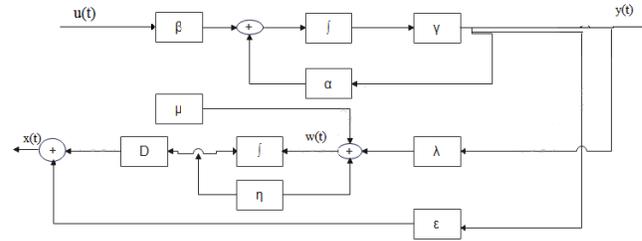
Fig. 2. Schematic of a Functional Observer.

C. Corollary –III

Functional Observer based Conventional Controller:

Due to increasing complexity in power distribution network, a simplified assumption that considering all the generators in a given area into one single generation unit, transmission lines and various bus bars are lumped into one single entity might not be appropriate enough for any complex power system network. This paper presents an analysis of a Quasi-decentralized Functional Observer (QDFO) project to control the tie-line power and frequency of multi-area interconnected power system with real time considerations [10]. Also, linear system (two-area) connected with a single tie-line model was considered for the generation of control signals and Quasi-Decentralized Functional Observer(QDFO) is applied to it for the control. Further it is formulated with FO [3] approach for LFC (Load Frequency Control) of highly interconnected power networks. In the process of generation of the LFC signal, it is required to estimate the control signal. It is more rational to estimate the desired signal directly using the functional observer (FO) than estimating all the individual states and then linearly combining those individual state estimates to construct the control signal [4]. And the control signal is generated by a Quasi-Decentralized Functional Observer (QDFO) in this case. PMU measurements of voltage and current magnitudes and phase angles, as well as tie-line power measurements, are used in the Functional Observer (FO) estimate process [2]. The suggested FO-based controllers have less complicated architecture and comparable performance to Full Order observers. Further, the Functional Observability (FO) criterion is less rigorous than the state observability requirement, and the design and analysis of the observer of overall network topology [9] is take into account.

Design Algorithm

1) *L is* Partitioned according to $\mathcal{L} = \begin{bmatrix} \mathcal{L}_1 \\ \mathcal{L}_2 \end{bmatrix} = \begin{bmatrix} g\gamma \\ \mathcal{L}_2 \end{bmatrix}$

2) Checking if condition (26) is met, if yes continue, otherwise a sliding mode functional observer doesn't exist.

3) Calculating *F* using (32) and *G* using (33).

4) Using (34) and any pole placement method obtain *Z* to make *η* Hurwitz.

5) Calculate *E* and *K* using (35)

6) Use (36) to calculate *λ*.

7) Calculate *μ* according to (37).

Design of FO based Controller:

Let ê be the estimate of e.

$$\dot{\hat{e}} = \alpha\hat{e} - \beta\Gamma_c^T\hat{e} + \Gamma_o(e_1 - \hat{e}_1) \qquad (40)$$

Where, $\Gamma_o \equiv [\gamma_1^o, \gamma_2^o, \dots, \gamma_n^o]^T \epsilon R^n$ is the observer gain vector.

Observer error is given as,

$$\tilde{e} = \alpha_o\tilde{e} + \beta(g(x)(u^* - u_{PID} - u_d - u_s) - d) \qquad (41)$$

As (γ, α) is observable, the observer gain vector, Γ_o can be stringently Hurwitz with a symmetric positive definite matrix, P and a positive definite matrix, Q_o.

If $\mathcal{L}(.)$ is taken to be the Laplace transform, the transform function can be written as $\mathcal{L}(\tilde{e}_1)$ by choosing $\mathcal{M}(s)$ such that $\mathcal{M}(s) = s^m + b_1 s^{m-1} + \dots + b_m$, $m < n$ and $\mathcal{M}^{-1}(s)$ is a proper stable transfer function and $\eta(s)\mathcal{M}(s)$ is a proper SPR transfer function. Hence $\mathcal{L}(\tilde{e}_1)$ can be written as:

$$\mathcal{L}(\tilde{e}_1) = \eta(s)\mathcal{M}(s)\mathcal{M}^{-1}(s)\mathcal{L}(g(x)(u^* - u_{PID}) - d) - \eta(s)\mathcal{M}(s)\mathcal{M}^{-1}(s)\mathcal{L}(g(x)u_d) - \eta(s)\mathcal{M}(s)\mathcal{M}^{-1}(s)\mathcal{L}(g(x)u_s) + \eta(s)\mathcal{M}(s)\mathcal{L}(u^* - u_{PID}) - d) - \mathcal{L}(u^* - u_{PID}) \qquad (42)$$

Denote the function φ such that

$$\mathcal{L}(\varphi) = \mathcal{M}(s)^{-1}\mathcal{L}(g(x)(u^* - u_{PID}) - d) - \mathcal{L}(u^* - u_{PID}),$$

Hence the dynamce equation can be written as

$$\tilde{e}_1 = \gamma_m^T\tilde{e} \qquad (43)$$

Where, $B_m = [0,0,\dots,0,0,\dots,b_1,b_2,\dots b_m]^T \epsilon \mathfrak{R}^n$,

$C_m = [1,0,\dots 0]^T \epsilon \mathfrak{R}^n$

For further analysis, there is necessity to assume the following assumptions

Assumption 1: The uncertain non-linear function f(x) for the states is bounded by an upper bound function $f^u(x)$, i.e. $f(x) \leq f^u(x)$. The uncertain non-linear function g(x) related with the input is bounded by $g_l \leq \| g(x) \| \leq g^u$ where both upper and lower boundaries g^u and g_l are positive constants.

Assumption 2: the function φ is bounded by $\| \varphi \| \leq \varepsilon$ where ε is a positive constant.

Defining and differentiating V with respect to t, we get

$$\dot{V} = \frac{1}{2}\Big((\tilde{e}^T\alpha_0^T P\tilde{e} + \tilde{e}^T P\alpha_0\tilde{e}) + (u^* - u_{PID} + \varphi - \tilde{u}_d - \tilde{u}_s)^T\beta_m^T P\tilde{e} + \tilde{e}^T P\beta_m(u^* - u_{PID} + \varphi - \tilde{u}_d - \tilde{u}_s)\Big) \quad (44)$$

But it is given that,

$$\alpha_0^T P + P\alpha_0 = -Q \qquad (45)$$

And $P\beta_m = \gamma_m$ $\qquad (46)$

Where, $Q = Q^T > 0$. Substituting (45), (46) and (43) in (44) we get,

$$V \leq -\frac{1}{2}|\tilde{e}^T Q\tilde{e}| + |e_1||u^* - u_{PID}| + \tilde{e}_1(\varphi - \bar{u}_d) - \tilde{e}_1\bar{u}_s \quad (47)$$

Based on the above assumptions, u_d can be designed such that $\tilde{e}_1(\varphi - \bar{u}_d) \leq 0$.

$$u_d = \begin{cases} \varepsilon + k, & \text{if } \tilde{e}\_1 \geq 0 \text{ and } \varphi > 0 \\ 0 & \text{if } \tilde{e}_1 \geq 0 \text{ and } \varphi < 0 \\ 0, & \text{if } \tilde{e}_1 \geq 0 \text{ and } \varphi > 0 \\ -(\varepsilon + k), & \text{if } \tilde{e}_1 \geq 0 \text{ and } \varphi < 0 \end{cases} \qquad (48)$$

Where, k is a positive constant. Substituting u_d in (47),

$$\dot{V} \leq -\frac{1}{2}|\tilde{e}^T Q\tilde{e}| + |\tilde{e}_1||u^* - u_{PID}| - \tilde{e}_1\tilde{u}_s$$

$$\leq -\frac{1}{2}\lambda_{min}(Q)|\tilde{e}_1|^2 + |\tilde{e}_1||u^* - u_{PID}| - \tilde{e}_1\tilde{u}_s \qquad (49)$$

$$\dot{V} \leq -1/2\,\lambda\_min\,(Q)\,|\tilde{e}\_1|{}^2 + |\tilde{e}\_1|(1/|g\_1|\,(|f^u\,(\acute{x})| + |y\_m^{((n))}| + |\Gamma\_e^T\hat{e}|) + |u\_PID|) - \tilde{e}\_1\,u\_s) \quad (50)$$

$$V_d(K_{PID}) \leq -\frac{1}{2}\lambda_{min}(Q)|\tilde{e}_1|^2 + |\tilde{e}_1|\Big(\frac{1}{|g_1|}\big(|f^u(\hat{x})| + |y_m^{(n)}| + |\Gamma_e^T\hat{e}|\big) + |u_{PID}|\Big) \qquad (51)$$

If K_{PID} estimated by MGA results in $V_d(K_{PID}) < 0$, $\dot{V} < 0$ is satisfied: given that u_s is not applied to the input in (10). Conversely, if $V_d(K_{PID}) > 0$, u_s should be applied leading to the condition $\dot{V} < 0$. For stabilizing the effect caused by u_s being either excluded or included in PID controller, define a gate function.

$$\theta(V_d) = \begin{cases} 0, & \text{if } V_d < 0; \\ \frac{V_d}{V_m}, & \text{if } 0 \leq V_d < V_m \\ 1, & \text{if } V_d \geq V_m \end{cases} \qquad (52)$$

Where, V_m is a positive constant. Now the supervisory control with the gate function becomes

$$u_s \leq \theta(V_d)\text{sgn}(\tilde{e}_1)\Big(\frac{1}{|g_1|}\big(|f^u(\hat{x})| + |y_m^{(n)}| + |\Gamma_e^T\hat{e}|\big) + |u_{PID}|\Big) \qquad (53)$$

D. Corollary IV

In recent years, elements like heuristics, reasoning gained importance due to their ease in more flexible control theories. Fuzzy logic (FL), optimization algorithms like Genetic Algorithm (GA) and Artificial Neural Network (ANN) may be applied for further refinement of this model to facilitate the development of advanced/intelligent controllers. Fuzzy logic (FL) is advantageous in terms of its logical designing and decision making but its drawback lies in the fact that the solution takes a complex form if more number of variables are involved. To estimate the control input specifically when the output y is measurable instead of system state x [15], the author proposed Fuzzy Functional Observer [14]. The T–S fuzzy model (1) is assumed to be as [9]:

$$\dot{x} = \sum_{i=1}^p w_i(y)(\alpha_i x + \beta_i \breve{u}) \qquad (54)$$

$$y = \gamma x \qquad (55)$$

Where, $y \in R^l$ is the output of the system, $\gamma \in R^{l \times n}$ is the output matrix. Now the Fuzzy controller becomes:

$$u = \sum_{j=1}^{p} w_i(y)u_j = \sum_{j=1}^{p} w_i(y)g_j x \qquad (56)$$

Where, $u_j = gG_j x \in R^m$ is the control input in j^{th} rule.

Assuming rank $(\gamma) = l$, rank $(g_j) = m$, i.e., γ and g_j are of full row rank, the following fuzzy functional observer is proposed to estimate the control input u:

$$\dot{z}_j = \sum_{i=1}^{p} w_i(y)\big(\eta_{ij}z_j + \lambda_{ij}y + \mu_{ij}\breve{u}\big) \; \forall j \qquad (57)$$

$$\breve{u}_j = z_j + E_j y \forall j \qquad (58)$$

$$\breve{u} = \sum_{j=1}^{p} w_j(y)\breve{u}_j \qquad (59)$$

Where, $z_j \in R^m, \breve{u}_j \in R^m$, $\breve{u} \in R^m$ are the observer state, estimated control input in j^{th} rule, estimated control input respectively; $\eta_{ij} \in R^{m\times m}$, $\lambda_{ij} \in R^{m\times l}$, $\mu_{ij} \in R^{m\times m}$, $E_j \in R^{m\times l}$ are the observer gains to be designed.

Estimation error

$$e_j = Q_j x - z_j \text{ , where } Q_j = g_j - E_j\gamma \qquad (60)$$

$$\dot{x} = \sum_{i=1}^{p}\sum_{l=1}^{p} h_{il}(\alpha_i x + \beta_i G_l x - \beta_i \sum_{k=1}^{p} w_k e_k) \qquad (61)$$

$$\dot{e}_j = \sum_{i=1}^{p}\sum_{l=1}^{p} h_{il}\Big((\phi_{ij} + \alpha_{ij}g_l)x + \eta_{ij}e_j - \alpha_{ij}\sum_{k=1}^{p} w_k e_k\Big)\forall j \quad (62)$$

Where $h_{il} = w_i w_l$, $\phi_{ij} = Q_j \alpha_i - \eta_{ij}Q_j - J_{ij}\gamma$, $\alpha_{ij} = Q_j \beta_i - \mu_{ij}$

To make the system asymptotically stable, g_j , η_{ij} , λ_{ij}, μ_{ij} and E_j is to be determine.

Now on imposing the constraints, we get

$$\dot{x}_a = \sum_{i=1}^{p}\sum_{l=1}^{p} h_{il}\Gamma_{il}x_a \qquad (63)$$

$$\text{Where,} \Gamma_{il} = \begin{bmatrix} \alpha_i + \beta_i g_l & -\beta_i w_1 & -\beta_i w_2 \\ 0 & \eta_{i1} & 0 \\ 0 & 0 & \eta_{i2} \end{bmatrix}$$

Theorem 3:

The error systems are to be asymptotically stable if there are matrices, $X^T \in R^{m\times m}$, $Y_{ij} \in R^{m\times 2l}$, i,j=1, 2, . . . , p such that it satisfies the following:

$X > 0$

$XF_{ij} - Y_{ij}M_{ij} + F_{ij}^T X - M_{ij}^T Y_{ij}^T < 0 \quad \forall i, j$

$\widetilde{E}_{i_1 j} = \widetilde{E}_{i_2 j} \forall i_1 < i_2, j$

Where,

$$F_{ij} = g_j\alpha_i g_j^+ - g_j\bar{\alpha}_{ij}\Sigma_{ij}^+ \begin{bmatrix} \gamma\alpha_i g_j^+ \\ \gamma g_j^+ \end{bmatrix} \qquad (64)$$

$$\mathcal{M}_{ij} = \big(I - \Sigma_{ij}\Sigma_{ij}^+\big)\begin{bmatrix} \gamma\alpha_i g_j^+ \\ \gamma g_j^+ \end{bmatrix} \qquad (65)$$

$$\bar{\alpha}_{ij} = \alpha_i\big(I - g_j^+ g_j\big) \qquad (66)$$

$$\Sigma_{ij} = \begin{bmatrix} \gamma\bar{\alpha}_{ij} \\ \bar{\gamma}_j \end{bmatrix} \qquad (67)$$

$$\bar{\gamma}_j = \gamma\big(I - g_j^+ g_j\big) \qquad (68)$$

Now the controller gain G_j can determineby

$$\begin{bmatrix} \widetilde{E}_{ij} & \widetilde{K}_{ij} \end{bmatrix} = Xg_j\bar{\alpha}_{ij}\Sigma_{ij}^+ + Y_{ij}\big(I - \Sigma_{ij}\Sigma_{ij}^+\big); \; Z_{ij} = X^{-1}Y_{ij} \qquad (69)$$

Using the equation of constraint, we can have

$$N_{ij} = g_j\alpha_i g_j^+ - E_j\gamma\alpha_i g_j^+ - \big(\lambda_{ij} - \eta_{ij}E_j\big)\gamma g_j^+ \forall i, j$$

$$K_{ij} = \lambda_{ij} - \eta_{ij}E_j \qquad (70)$$

The error system now becomes

$$\dot{e}_j = \sum_{i=1}^{p} w_i\big(F_{ij} - Z_{ij}\mathcal{M}_{ij}\big)e_j \; \forall j \qquad (71)$$

Applying the Lyapunov function, time derivative of $\dot{V}(e_j)$ is obtain as:

$$\dot{V}(e_j) = \sum_{i=1}^{p} w_i e_j^T \big(XF_{ij} - Y_{ij}\mathcal{M}_{ij} + F_{ij}^T X - \mathcal{M}_{ij}^T Y_{ij}^T\big)e_j \qquad (72)$$

Where $Y_{ij} = XZ_{ij}$.

In order to determine the observer gains satisfying the constraints, it is necessary to calculate the following terms from the above equations:

1) η_{ij}
2) E_j and the intermediate variable K_{ij}
3) λ_{ij}
4) μ_{ij}.

Hence observer gains can be obtained.

E. Corollary V

Fuzzy Functional Observer Vs Type-2 Fuzzy Functional Observer:

Type-2 fuzzy systems can model complex uncertainties better than type-1 and it is important to develop type-2 fuzzy systems [12] for enhanced performance. Literature shows that type-2 fuzzy systems gained much importance in recent times. In this paper, T2FFO based on Lyapunov theorem have been modeled to investigate closed-loop stability. Numerous techniques have been presented in this paper for tuning free parameters in the design of optimal type-2 fuzzy systems. Few techniques like instance recursive orthogonal least-squares algorithm, multi-objective genetic optimization [3], steepest descent method, etc. can be employed. Taking the computational cost into account, as the type-reduction part is more, simplified type-2 fuzzy systems have been proposed. Also the direct defuzzification for type-2 fuzzy systems has been proposed and compared with type-reduction. The design of adaptive fuzzy controllers have some challenges like approximation error effects, computational cost, external disturbances and state estimation errors unswervingly on the stability of closed-loop system. Hence a novel robust observer for a class of uncertain nonlinear systems using a new adaptive compensator based indirect adaptive type-2 fuzzy controller [12] is presented to eliminate the disturbances. As it is being employed 3-dimensional type-2 membership functions, the "course of dimensionality" problem can be solved and this controller can be applied to the higher order systems.

Fuzzy systems with normal, orthogonal, consistent and complete triangular membership function in the antecedent. The systems are dynamic and their consequents are in the form of state equation and output equation. The system rules are in the form

$$R^{i_1 \dots i_m} \text{ IF } u(k) \text{ is } \widetilde{U}^{i_1 \dots i_m} \text{ THEN}$$
$$\begin{cases} x^{i_1 \dots i_m}(k+1) = F^{i_1 \dots i_m}.x(k) + G^{i_1 \dots i_m}.u(k) \\ y(k) = C^{i_1 \dots i_m}.x(k) + (x(k), u(k)) \end{cases} \quad (73)$$

With $\widetilde{U}^{i_1 \dots i_m}$ orthogonal triangular type-2 membership functions such that, $\widetilde{U}^{i_1 \dots i_m} = [\widetilde{U}_1^{i_1} \dots \dots \widetilde{U}_m^{i_m}]^T$. The state vector is $x(k) = [x_1(k) \dots \dots x_n(k)]^T$, the fuzzy input vector is $u(k) = [u_1(k) \dots \dots u_m(k)]^T$, and the output vector is $y(k) = [y_1(k) \dots \dots y_p(k)]^T$, $m \le n$, $i_j = 1, \dots, N_j$, $j = 1, \dots, n$. The constant matrices in the consequents are $F^{i_1 \dots i_m} \in \mathbb{R}^{n \times n}$, $G^{i_1 \dots i_m} \in \mathbb{R}^{n \times m}$, $C^{i_1 \dots i_m} \in \mathbb{R}^{p \times n}$ and the uncertainty in output is given by the state vector function and control input

$$D(x(k), u(k)) = [D_1(x(k), u(k)), \dots \dots, D_p(x(k), u(k))]^T \quad (74)$$

Without loss of generality, $D_j, j = 1, \dots, p$, are fuzzy sets with membership functions is assume.

$$\mu_{D_j}(d_j) = \begin{cases} \text{nonzero,} & \underline{S_j}\left(x_j(k)\right) \le d_j \le \overline{S_j}\left(x_j(k)\right), \quad j = 1, \dots \dots, n \\ 0, & \text{elsewhere} \end{cases} \quad (75)$$

Type-2 Fuzzy PID controller with 3 input Type-2 Fuzzy Logic Controller (T2FLC) structure with coupled rules is shown in Fig. 3.
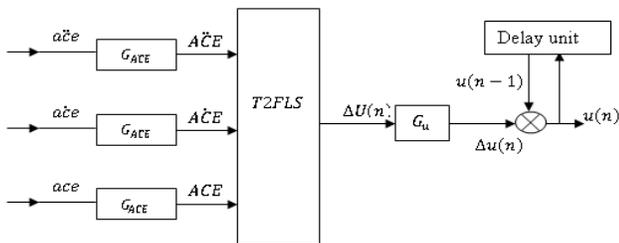
Fig. 3. Structure of the Interval Type-2 Fuzzy PID Controller.

Here ACE, (ACE)˙ and (ACE)¨ are the three input variables which are fuzzified by two interval type-2 fuzzy as positive and negative as shown below in Fig. 2 and represented by and Membership functions for ACE are given as:
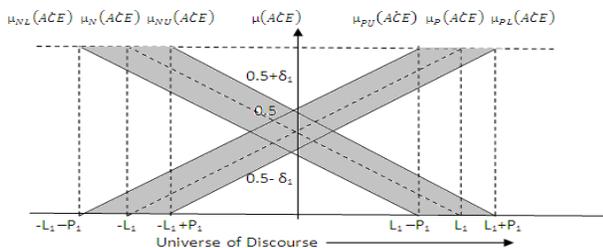
Fig. 4. Membership Functions of the Interval Type-2 Fuzzy Sets for the Scaled Input Variables.

TABLE I. CONTROL RULES FOR T2 FUZZY CONTROLLER

	ΔACE			
	N	P	N	N
ΔAĊE	Z	N	P	P
	P	N	N	N

In this paper it is assumed that membership functions of the antecedents of all rules are consistent, triangular, complete and orthogonal (Fig. 4).

The block diagram of the plant and observer is as shown in Fig. 5 where $u(k)$, $y(k)$ and $\hat{x}(k)$ denotes the plant input, output and state estimate respectively. The system input is a fuzzy variable and the system output is crisp and fuzzy for systems.
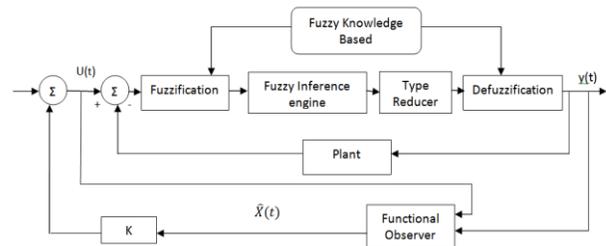
Fig. 5. Block Diagram of Plant and Observer.

III. RESULTS AND DISCUSSIONS

Simulation of the proposed technique has been carried out using Matlab / Simulink with an increase in demand of the first area ΔPD_1 and second area ΔPD_2. Testing with wide perturbations in input have been also carried out over both the areas i.e., demand of the first area ΔPD_1 and the second area ΔPD_2. It is apparent from the Fig. 6 – Fig. 12 that the system response is faster in terms of control and frequency deviations using proposed methodology are also nullified as a result. Thus the theorized model provides better control and frequency damping when performance comparison is made with fuzzy functional observer, functional observer, functional observer without controller and Luenberger/state observer, under all operating conditions.

Table I lists down the robustness of performance under various operating conditions numerically. Here the settling time, overshoot and undershoot have been calculated for different operating points. The Simulation results are shown for 10% band of step load change for operating point of Appendix A. According to Table I, the proposed T2FQFO based controller is better compared to the FQFO based controller, Functional Observer based controller, Functional Observer and the Luenberger observer based controller.

A. Result Analysis:

The Peak over Shoot, Undershoot and settling time at different operating points with different observers have been tabulated in the Table II shown below. From the table we can say that, Peak over shoot had been improved by 25.37% (approximately) and the settling time have improved by 8% using Type-2 Quasi Decentralized Functional Observer over Fuzzy Decentralized Functional Observer. Peak over shoot

and the settling time had been improved by 46.77% (approximately) and 28.52% respectively by using Type-2 Quasi Decentralized Functional Observer over Fuzzy Decentralized Functional Observer. Peak over shoot and the settling time had been improved by 58.91% (approximately) and 47.7% respectively by using Type-2 Quasi Decentralized Functional Observer over Quasi Decentralized Functional Observer. Peak over shoot and the settling time had been improved by 53.28% (approximately) and 62.98% respectively by using Type-2 Quasi Decentralized Functional Observer over Full order Luenberger Observer. These parameters are taken at different operating points and these parameters are improving at different operating points. The parameters are more improved at operating point "5".

TABLE II. ΔF_1 (T) RESPONSE PERFORMANCE IN VARIOUS CONTROL STRATEGIES

Operating Point	Controller	Over Shoot (P.U)	Under Shoot (P.U)	Settling time(sec)
1	Type-2 Quasi Decentralized Functional Observer	0.05013	-0.06942	3 . 4 0 1
	Fuzzy Decentralized Functional Observer	0.06718	-0.0868	3 . 6 9 4
	Quasi Decentralized Functional Observer	0.09419	-0.0990	4 . 7 5 8
	Full order Luenberger Observer	0.1220	-0.0990	6 . 5 0 5
	Functional Observer without Controller	0.1073	-0.0990	9 . 1 8 7
2	Type-2 Quasi Decentralized Functional Observer	0.0478	-0.0704	3 . 0 1
	Fuzzy Decentralized Functional Observer	0.0702	-0.0968	3 . 2 9
	Quasi Decentralized Functional Observer	0.1039	-0.1147	4 . 6 1 6
	Full order Luenberger Observer	0.1163	-0.1147	6 . 1 2 5
	Functional Observer without Controller	0.1214	-0.1147	6 . 9 6 7
3	Type-2 Quasi Decentralized Functional Observer	0.0487	-0.07038	2 . 9 5 0
	Fuzzy Decentralized Functional Observer	0.0693	-0.0943	3 . 2 4
	Quasi Decentralized Functional Observer	0.1025	-0.1108	4 . 8 0 4
	Full order Luenberger Observer	0.1143	-0.1108	6 . 3 2 6
	Functional Observer without Controller	0.1188	-0.1108	7 . 2 4 1
4	Type-2 Quasi Decentralized Functional Observer	0.04818	-0.07052	3 . 0 1 9
	Fuzzy Decentralized Functional Observer	0.06941	-0.09437	3 . 2 2
	Quasi Decentralized Functional Observer	0.1026	-0.1107	4 . 7 3 5
	Full order Luenberger Observer	0.1144	-0.1107	6 . 7 7 3
	Functional Observer without Controller	0.1189	-0.1107	7 . 3 1 2
5	Type-2 Quasi Decentralized Functional Observer	0.0509	-0.0676	4 . 1 4

	Fuzzy Decentralized Functional Observer	0.0636	-0.08113	4 . 4 0
	Quasi Decentralized Functional Observer	0.0851	-0.0909	6 . 8 0 8
	Full order Luenberger Observer	0.0917	-0.0909	6 . 9 8
	Functional Observer without Controller	0.0956	-0.0909	> 1 0
6	Type-2 Quasi Decentralized Functional Observer	0.0487	-0.07032	3 . 3 9 6
	Fuzzy Decentralized Functional Observer	0.07032	-0.09322	3 . 7 2 6
	Quasi Decentralized Functional Observer	0.1008	-0.1073	4 . 8 5 7
	Full order Luenberger Observer	0.1119	-0.1073	6 . 4 0 7
	Functional Observer without Controller	0.1162	-0.1073	7 . 2 1 4
7	Type-2 Quasi Decentralized Functional Observer	0.05057	-0.06799	3 . 8 9 6
	Fuzzy Decentralized Functional Observer	0.06447	-0.08239	4 . 0 1 2
	Quasi Decentralized Functional Observer	0.08722	-0.09273	6 . 5 4 2
	Full order Luenberger Observer	0.09461	-0.09273	7 . 1 7 4
	Functional Observer without Controller	0.098938	-0.09273	> 1
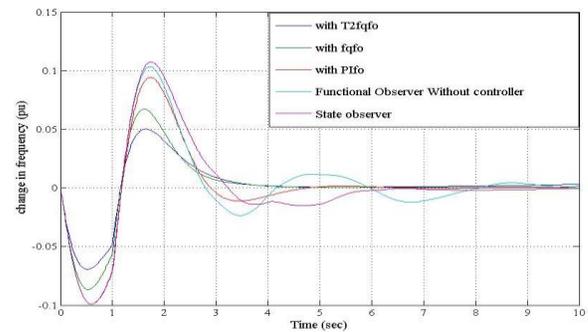
Fig. 6.   Change in Frequency with Step Increase in Demand at Operating Point 1.
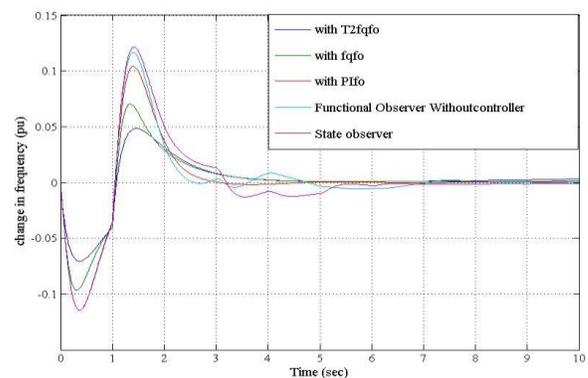
Fig. 7.   Change in Frequency with Step Increase in Demand at Operating Point 2.
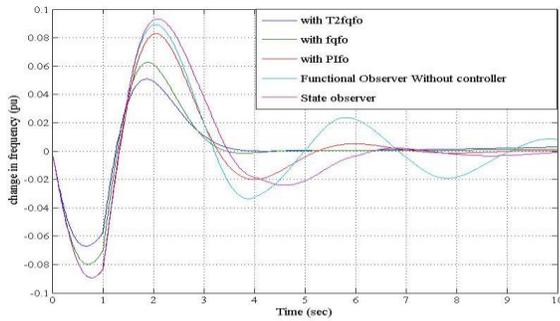
Fig. 8. Change in Frequency with Step Increase in Demand at Operating Point 3.
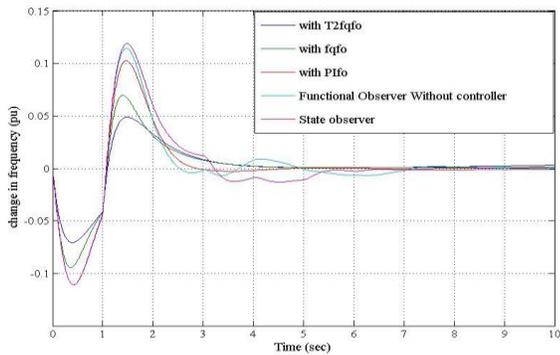
Fig. 9. Change in Frequency with Step Increase in Demand at Operating Point 4.
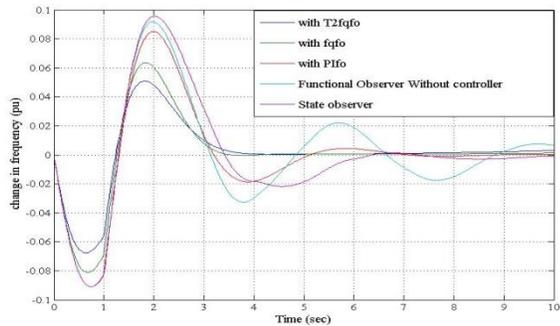
Fig. 10. Change in Frequency with Step Increase in Demand at Operating Point 5.
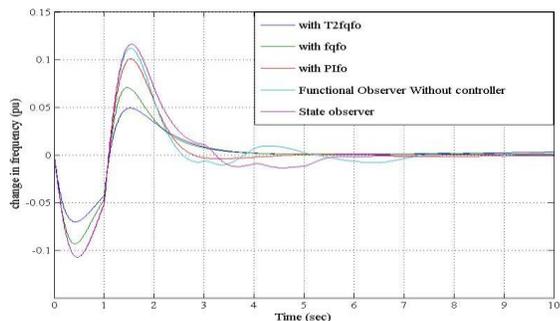
Fig. 11. Change in Frequency with Step Increase in Demand at Operating Point 6.
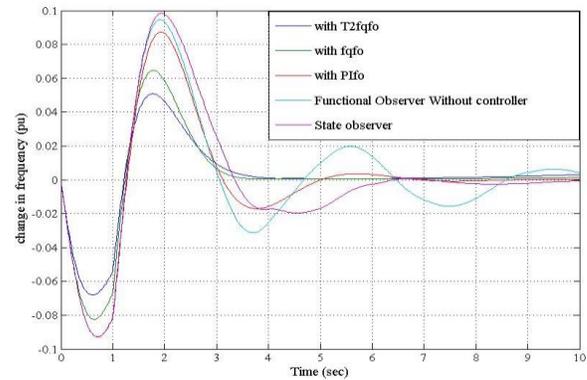
Fig. 12. Change in Frequency with Step Increase in Demand at Operating Point 7.

IV. CONCLUSIONS

In this present paper a novel Type-2 Fuzzy Functional Observer (T2FQFO) as a solution to the as a solution to the problem of load frequency control is proposed and applied to multi-area power system. Testing for disturbance attenuation and precise reference frequency tracking under different load (operating) conditions has been carried out over a typical two area interconnected reheat thermal power system with parameter uncertainties of a wide range. Competence of the proposed observer model is tested by performance comparison with FQFO based control, Functional Observer (With and without conventional controller) and Luenberger observer pertaining to settling time, maximum overshoots/undershoots under a variety of operating situations. Simulation results obtained show robustness pertaining to stability and consistency of performance of the suggested observer modeled herewith. The work may be extended by designing an optimization based design methodology using Genetic Algorithm-Fuzzy controllers or Honey Bee Mating optimization algorithm for tuning Type-2 Fuzzy PI/PID controllers. The LFC design can also be done using certain techniques like Active Disturbance rejection control for Type-2 Fuzzy system. The work may be extended with drawing extra degree of freedom framing as Type-3 Fuzzy. The relationship between the appropriate FOUs for a Type-2 FLC and the uncertainties in the plant parameters is still unsolved. If it is solved, the applications of Type-2 FLCs will be greatly promoted.

REFERENCES

[1] M. Darouach, "Full Order Observers for Linear Systems with Unknown Inputs", IEEE Trans. Autom. Control, vol. 39, no. 3, pp. 606–609, Mar. 1994.

[2] H.Beides and G.Heydt, "Dynamic state estimation of power system harmonics using Kalman filter methodology," IEEE Transaction on Power Delivery,vol.6, no.4, pp.1663-1670,1991.

[3] Hieu Trinh, Tyrone Fernand, "QuasiDecentralized Functional Observers for the LFC of Interconnected Power Systems", IEEE Transactions on Power Systems, 28, no. 3, Aug. 2013.

[4] H. Trinh, T. Fernand, "Functional Observers for Dynamical Systems", Berlin, Germany: Springer Verlag, 2012.

[5] K. R. Sudha, R. Vijaya Santhi, "Robust Decentralized Load Frequency Control of Interconnected Power System with Generation Rate Constraint using Type-2 Fuzzy Approach", Science Direct Electrical Power and Energy Systems, 33, pp 699–707, 2011.

[6] L.Zhao and A.Abur, "Multi area state estimation using synchronized phasor measurement," IEEE Transaction on power system, vol.20, no.2, pp. 611-617,2005.

[7] K.Shih and S.Huang, "Application of a robust algorithm for dynamic state estimation of a power system," vol.17, no.1, pp.141-147, 2002.

[8] E.Farantatos, G.K.Stefopoulos, G.J.Cokkinides and A.Meliopoulous, "PMU Based dynamic state estimation for electric power system," in IEEE power and energy society general meeting,2009,pp.1-8.

[9] Chuang Liu, Hak-Keung Lam, Tyrone Fernando, and Herbert Ho-Ching Iu, "Design of Fuzzy Functional Observer-Controller via Higher Order Derivatives of Lyapunov Function for Nonlinear Systems," IEEE Transactions on Cybernetics, vol.pp, Issue: 99, pp.1-11,2016.

[10] M. Aldeen, J. Marsh, "DecentralisedObserver based Control Scheme for Interconnected Dynamical Systems with Unknown Inputs", IEEE Proceedings, Part D, Control Theory and Applications 146,pp.349–358, 1999.

[11] M. Aldeen, H. Trinh, "Load Frequency Control of Interconnected Power Systems via Constrained Feedback Control Schemes", Computers and Electrical Engineering 20, pp.71–88, 1994.

[12] Mendel JM, "Advances in Type-2 Fuzzy Sets and Systems", Inform Science 177, pp.84–110, 2007.

[13] M. Darouach and M. Boutayeb, Design of observers for descriptor systems,IEEE Transactions on Automatic Control, vol. 40, pp. 1323-1327, 1995.

[14] Michal Polanský, and Cemal Ardil" Robust Fuzzy Observer Design for Nonlinear Systems" International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:5, No:5, 2011.

[15] X. J. Ma, Z. Q. Sun, "Analysis and design of fuzzy controller and Fuzzy observer" IEEE Trans. Fuzzy Systems. 1998, vol. 6, no. 1, pp. 41–51..

[16] KR Sudha, RV Santhi "Load frequency control of an interconnected reheat thermal system using type-2 fuzzy system including SMES units", International Journal of Electrical Power and Energy Systems 43 (1), 1383-1392, 2012.

[17] KR Sudha, YB Raju, AC Sekhar "Fuzzy C-Means clustering for robust decentralized load frequency control of interconnected power system with Generation Rate Constraint" International Journal of Electrical Power and Energy Systems 37 (1), 58-66, 2012.

[18] R Vijaya Santhi, KR Sudha, S Prameela Devi ,"Robust Load Frequency Control of multi-area interconnected system including SMES units using Type-2 Fuzzy controller, Fuzzy Systems (FUZZY), 2013 IEEE International Conference on, pp 1-7,2013.

[19] Fang Chen, Xueshan Han,Zhiyuan Pan,Li Han "State Estimation Model and AlgorithmIncluding PMU" DRPT2008 6-9 April 2008 Nanjing China.

[20] Tyrone Fernando, Kianoush Emami, Shenglong Yu, Herbert Ho-Ching Iu, and Kit Po Wong, Fel, "A Novel Quasi-Decentralized Functional Observer Approach to LFC of Interconnected Power Systems",IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 31, NO. 4, JULY 2016.

[21] D G Luenberger, "Observer for multivariable system," IEEE Transactions on Automatic control, Vol.11, no.2, pp.190-197, 1966.

EEG-Based Silent Speech Interface and its Challenges: A Survey

Nilam Fitriah, Hasballah Zakaria, Tati Latifah Erawati Rajab*
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia

Abstract—**People with speech disorders could have social and welfare difficulties. Therefore, the silent speech interface (SSI) is needed to help them communicate. This interface decodes the speech from a human's biosignal. The brain signals contain information from speech production to cover people with numerous speech disorders. Brain signals can be acquired non-invasively by electroencephalograph (EEG) and later transformed into the features for the input of speech pattern recognition. This review discusses the advancement of EEG-based SSI research and its current challenges. It mainly discussed the acquisition protocol, spectral-spatial-temporal characterization of EEG-based imagined speech, classification techniques with leave-one-subject or session-out cross-validation, and related real-world environmental issues. It aims to aid future imagined speech decoding research in exploring the proper methods to overcome the problems.**

Keywords—Imagined speech; silent speech interface; electroencephalograph (EEG); speech recognition

I. INTRODUCTION

Communication is essential in daily human life. People would hardly communicate in noisy circumstances, in quiet environments where no sounds are allowed, in secret conversation, or when they have speech disorders.

Speech disorders could negatively affect a person's social life and welfare. WHO reported that in 2011 there was 3,6% of the world's population experienced extreme difficulty living in their community due to speech disorders [1]. Moreover, they also were hindered from getting a job, as stated by ILO in 2017 for 4,1% of Indonesian citizens [2]. Additionally, research conducted in the United States found that one in 13 adults experience speech disorders annually [3]. In this circumstance, an interface to assist communication becomes more necessary than ever.

One of the interfaces intended to help people with speech impairments to communicate is the silent speech interface (SSI), which converts the biosignal into a speech. Human speech can be categorized into overt speech (with sound), silent articulation (articulator moves but no sound), and covert speech (no sound and movement) [4]. The latter is also called silent speech or imagined speech as our focus of discussion.

The causes of speech disorders can be the absence of

knowledge to speak experienced by deaf people, articulation problems, neurologic dysfunction (e.g., stroke), and paralysis (i.e., tetraplegia, muscular degenerative diseases, locked-in syndrome, or coma patients) [4]. Most causes come from brain disorders [5]. The applicable sensors mainly record brain activity, such as electroencephalograph (EEG) or electro-corticograph (ECoG). While ECoG has a higher Signal-to-Noise Ratio (SNR) than EEG, its invasive electrode placement can have a clinical risk. Moreover, ECoG only covers a specific area, while EEG has a broader coverage than ECoG. Hence, EEG is considered safer than ECoG.

This review aims to present the development of EEG-based speech imagery studies and assist researchers in finding a solution to achieve better accuracy and solve the real problem. It focused on EEG application to decode imagined speech as a pipeline consisting of signal acquisition, signal preprocessing, feature extraction, classification techniques, and the real-world application challenges that were still unnoticed. The process flow of the review is in Fig. 1, and the summary of the discussed references is in Table I.
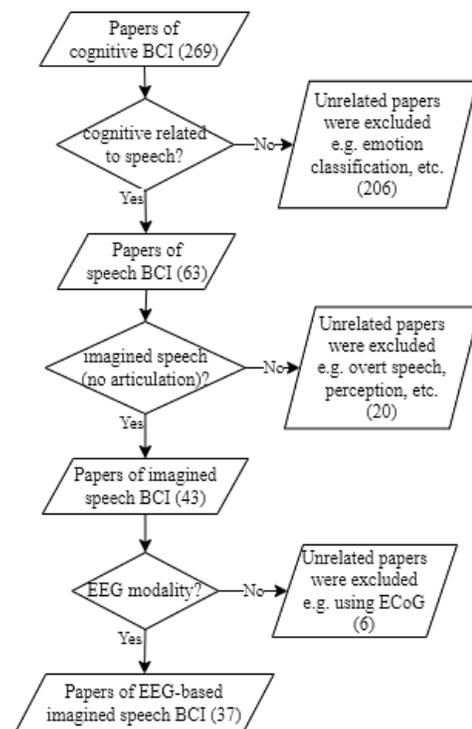
Fig. 1. Review Processes.

TABLE I.   EEG-BASED SSI STUDIES

Ref.	Cue	Speech	Freq. (Hz)	Feature extraction	Classification	Validation
[6]	Audiovisual	"to", "two", "too", "here", "hear"	1-20	FFT	Min. least-squares	CV
[7]	Visual	0-9; NATO phonetic alphabet (a-e); "yes", "no", etc.; sentence	0.9-60	STFT+LDA	HMM	CV
[8]	Visual	"alpha", "bravo", …, "echo"	1-300	DTCWT, LDA	HMM	CV
[9]			3-18	Hilbert transform, spectral feature	matched-filter	CV
[10]	Audiovisual	"/ba/", "/ku/"	3-20	spot of interest (SOI)	LDA	CV
[11]			4-25	AR	kNN	CV
[12]			1-45	CSP	SVM	CV
[13]	Visual	"/a/", "/u/"		Statistics, geometric mean, energy sum, entropy, wavelength	LDA	CV
[14]			8-40	MFCC	kNN/SVM	CV
[15]	Audio	"/aa/", "/ae/", "/l/", "/r/", "/m/", "/n/", "/uu/", "/ow/", "/s/", "/z/"	4-28	Spectrogram	LDA	CV
[16]	Visual	"two", "to", "four", "for"	N/A	Voltage	DT	CV
[17]	Visual	(Korean) 3 ("sam"), 5("oo"), 9 ("gu"), 10 ("sib"); cheek ("ppyam"), nose ("ko"), eye ("nun"), mouth ("ib")	1-100	spectrogram, STFT	SVM	CV
[18]	Visual	(Chinese) "左" ("left"), "壹"(1)	6-30	CSP, DWT, AR	SVM	CV
[19]	Visual	/um/	4-20	AR	LDA	CV
[20]	Audio (question)	(Arabic) "yes", "no"	0-48	DWT	SVM, SOM, LDA	CV
[21]	Audio	"/a/","/i/","/u/","/e/","/o/"	1-100	Statistics	EL	CV
[22]	Audio (question)	(Indian & English): "yes", "no"	0-40	Spectral power of FFT	ANN	CV
[23]			1-50	Statistics	SVM	CV
[24]	Audiovisual	"/iy/", "/uw/", "/piy/", "/tiy/", "/diy/", "/m/", "/n/", "pat", "pot", "knew", "gnaw"		Statistics, MFCC, nonlinear features	SVM	CV
[25]				DWT	DNN	CV
[26]				Riemannian manifold	RVM	CV
[27]				channel cross-covariance (CCV)	CNN+LSTM+DAE	CV
[28]	Visual	"/a/", "/i/", "/u/", "in", "out", "up", "cooperate", "independent"	8-70	channel cross-correlation matrix	LSTM	CV
[29]				DWT	DNN	CV
[30]				Bag of Features (BoF)	RNN	LOSO-CV
[30]			4-25			
[31]	Visual	(Spain): "arriba" ("up"), "abajo" ("down"), "izquierda" ("left"), "derecha" ("right"), "seleccionar" ("select")	4-25	DWT	RF	CV
[32]			0-64	Statistics, RWE		
[33]			40-50	Bag of Features (BoF)	NB + TL	CV
[34]			2-40	RWE	RF	CV
[35]		(Spain) "/a/", "/e/", "/i/", "/o/", "/u/", "arriba" ("up"), "abajo" ("down"), "izquierda" ("left"), "adelante" ("forward"), "atrás" ("backward")		RWE	CNN	CV
[36]	Audiovisual			CNN layer, FBCSP	CNN	CV
[37]				CNN layer, FBCSP	CNN + TL	LOSO-CV
[38]				word embedding + Siamese encoder	kNN	CV
[39]	Visual	"/a/", "/e/", "/i/", "/o/", "/t/"	0.5-220	phase per band (Hilbert transform)	SVM	CV
[40]	Audio	"/a/", "/e/", "/i/", "/o/", "/u/", "yes", "no", "left", "right".	0.1-70	RMS, zero-crossing rate, moving window average, kurtosis, and power spectral entropy	RNN	CV
[41]	Audio	"Hi Bixby", "Call Mom", "Open Camera", "What's the weather".	0.5-70		RNN	CV
[42]	Audio	"ambulance", "clock", "hello", "yes", "light", "help me", "pain", "stop", "thank you", "toilet", "TV", "water".	0.5-40	CSP	LDA	CV
[43]	Audio	"go", "back", "left", "right", and "stop"	0.5-60	covariance and MaxLCor	ELM	CV
[44]	Audio	"hello", "help me", "stop", "yes", "thank you"	0.5-128	DWT, MaxLCor	SVM	CV
[45]	Visual	ten words for every vowel: "a" ("can", …, "tap"), "e" ("bed", …, "vex"), etc.	0.5-50	coherence, PDC, DTF, transfer entropy	DBN	LOSO-CV

ANN = artificial neural network (NN), AR = autoregression, CNN = convolutional NN, CSP = common spatial pattern, DAE = Deep Autoencoder, DNN = deep NN, DT = decision tree, DTCWT = Double-Tree Complex Wavelet Transform (WT), DTF = direct transfer function, DWT = Discrete WT, ELM = extreme learning, FBCSP = filter bank CSP, FFT = Fast Fourier Transform (FT), HMM = hidden Markov model, kNN = k-nearest neighbour, LDA = linear discriminant analysis, LSTM = long short-term memory, MaxLCor = Maximum Linear Cross-correlation Coefficient, MFCC = Mel Frequency Cepstral Coefficients, NB = naïve Bayes, PDC = partial directed coherence, RF = random forest, RMS = root mean square, RNN = recurrent NN, RVM = relevance vector machine (VM), RWE = relative wavelet energy, SOM = self-organizing map (clustering), STFT = Short Time FT, SVM = support VM, TL = transfer learning.

This paper did not compare the accuracies quantitatively between studies due to the different techniques, datasets, or computation environments. Hence, the comparison would not be apple-to-apple. Discussing techniques and challenges is more worthwhile than the accuracy comparison to find the right solutions. The remainder of this review is organized as follows. Section II discusses the data acquisition process. Section III describes the signal processing used in the reviewed studies. Section IV deals with Spectral, Spatial and Temporal analysis. Section V explains the classification and feature extraction techniques that were categorized further into non-deep learning and deep learning, including the modified validation method. This paper describes the challenges of the application of EEG-based SSI in Section VI. Finally, Section VII draws some conclusions.

II. DATA ACQUISITION

The research subjects were given the cue of speech (vowels or words) shown from the monitor, heard from the earphones, or both (audiovisual cue). If the cue presentation was before the speech imagery, they had to memorize the cue, and this would separate the imagined speech task from the reading/listening task. While in the simultaneous cues, subjects performed the imagined speech task with the reading/listening task at the same time. Besides, the most active brain parts for listening and reading are different, i.e., the listening process involves the temporal lobes, and the reading process involves the occipital lobes. Thus, cue format and presentation regarding the time of imagined speech can affect different active brain areas.

The acquisition protocol of the Arizona State University dataset [26] used only visual cues, as illustrated in Fig. 2. The cue presentation was simultaneous with the imagined speech recording. The subject performed speech imagery at each "beep" sound and continued the same pattern until the visual cue disappeared (7 x T second). They used three short words ("in", "out", and "up") and two long words ("cooperate" and "independent"). For the longer words, T is more than one second. This protocol was applied to 15 subjects (11 males, one left-handed, age 22-32) and the EEG signals recording used 60-channel EEG at 1000 Hz and two-channel electrooculography (EOG) to capture ocular artefacts. They applied Common Spatial Pattern (CSP) to get the most active brain areas. The results showed that brain activity almost entirely focused on the left frontal, middle and parietal sides of the brain, as the location of the motor cortex and Broca and Wernicke's area. Even in the rest state, the brain remains highly active.

The other previous study that used visual cues presented the cues separately from the imagined speech state [23], as shown in Fig. 3, named the KARAONE dataset. The goal was to differentiate between the pronounced speech performed in the cue state and the imagined speech after the cue state. They used seven phonemes; "/iy/", "/uw/", "/piy/", "/tiy/", "/diy/", "/m/", "/n/", and four words from Kent's list [46], i.e. "pat", "pot", "knew", and "gnaw". This protocol was conducted on 12 subjects (8 males, all right-handed, age 27.4±5) and recorded with 62-channel EEG at 1024 Hz. They found that central brain areas in temporal locations had discriminative features.

A different protocol was employed by Coretto et al. [34], as illustrated in Fig. 4. They used both audio and visual cues before subjects performed imagined speech. No specific reason why they decided to use both cue types. They employed five vowels; "/a/", "/e/", "/i/", "/o/", "/u/", and five Spanish words; "arriba" ("up"), "abajo" ("down"), "izquierda" ("left"), "derecha" ("right"), "atrás" ("backward"), and "adelante" ("forward"). This protocol was conducted on 15 subjects (8 males, one left-handed, averaged age 25) and recorded with six-channel EEG at 1024 Hz. Moreover, there was still no further study with spatial analysis to observe which brain parts are more involved in both cues.

On the other side, a Brain-Computer Interface (BCI) open dataset [47] for the speech imagery study only employed audio cues. They used five common words; "hello", "help me", "stop", "thank you", and "yes" and conducted this protocol on 15 subjects using 64-channel EEG at 256 Hz. The cue presentations were given independently from the imagined speech state, as shown in Fig. 5.
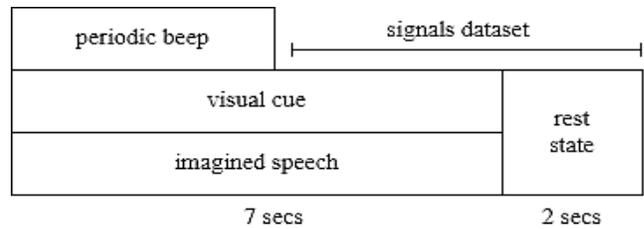
Fig. 2. Arizona State University Dataset's Acquisition Protocol.
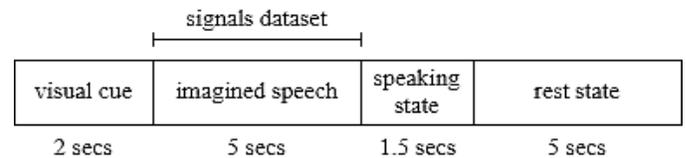
Fig. 3. KARAONE Dataset's Acquisition Protocol.
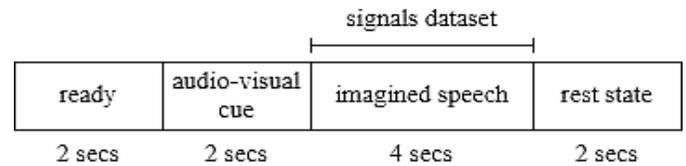
Fig. 4. Coretto Dataset's Acquisition Protocol.
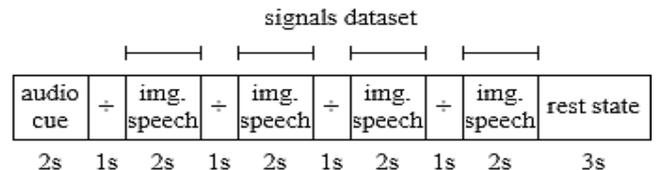
Fig. 5. BCI 2020 Track 3 Dataset's Acquisition Protocol.

Besides the cue format and its presentation timing, the duration of each state is an important issue. In the KARAONE dataset, the length of the imagined speech state was five seconds with no repetition. Meanwhile, Arizona State University's dataset [26] and Coretto's [34] applied repetition

during one imagined speech state using the beep sound. The Arizona State University dataset had seven repetitions in one imagined speech state. Coretto et al. applied repetition only for words, but not vowels, i.e., the subject must complete the task within four seconds of the vowel imagined speech state. BCI dataset [47] also used repetitions, but rather than using the beep sound, it applied a fixation cross shown from the monitor. The purpose of repetition block use was to maintain focus consistency [8]. Moreover, the longer the duration of the imagined speech, the easier for the subject to get sleepy because they performed the imagined speech state in silence.

While performing imagined speech, subjects were asked to refrain from moving articulation, swallowing, moving their eyes, or blinking. These restrictions aim to reduce the muscular and ocular artefacts as the significant artefacts in the EEG signal. These artefacts are generated by the signal from muscle activity surrounding the head, particularly the region near the articulator and eye activity. The range of muscular and ocular signals has intersected with the range of EEG signals [48].

The previous studies used different types of speech, such as vowels [26], phonemes [23], syllables [10], words [6]–[8], [16]–[18], [23], [26], [31], [32], sentences [7], and binary questions [20], [22]. The use of speech parts (vowels, phonemes, and syllables) mainly was to observe the brain when planning to produce the sound of words. Meanwhile, using words or binary questions to observe the brain when planning to respond/send a message earlier than sound production. Thus, the former is more syntactic, whereas the latter is more semantic. The distribution of the speech used in this related research is shown in Fig. 6.

Using syllables or vowels as the cue was for their discriminative sounds or articulator movements. To maintain the number of nasals, plosives, and vowels, Zhao and Rudzicz [23] used phonemes and short words with similar sounds. DaSalla et al. [12] only used the vowels "/a/" and "/u/" because of their articulation differences; "/a/" with an opened mouth regulated by digastricus muscles and "/u/" with rounded lips controlled by orbicularis oris muscles. On the other side, Nguyen et al. [26] utilized vowels to discover the relationships between vowel voices and model accuracies. They reported that different voices had an impact on the performances. Then, the use of the syllables "/ba/" and "/ku/" [9], [10] was mainly to classify the rhythm.

Most SSI studies employed the word cues for several reasons. It represents natural communication, such as asking, responding, or making a statement [18], [32] and semantically differentiates homonyms [6], [16], [17], e.g., "two" and "to". Additionally, employing words with different lengths could help the model to differentiate words (such as "in" and "cooperate") [26] or repetition [8] affected the model and use words as commands for moving the cursor on the monitor ("up"/"down") [31].
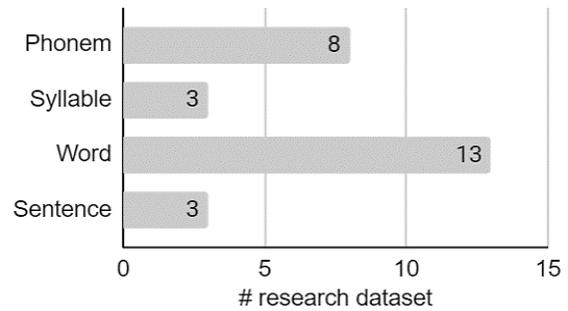
Fig. 6. Used Speech Cues.

Wester [7] also applied several words grouped based on their usability; digits (0-9), alpha corpus, phone-related words, MP3 player commands, Graduate Record Examination (GRE) Corpus, and Lecture Corpus. The alpha corpus consists of "alpha", "beta", "theta", "delta", and "echo". The phone-related words covered some phone communication words, such as "yes", "no", "accept", "deny", and "wait". The phrases for MP3 player commands were "start", "back", "next", "louder", and "turn down". GRE Corpus represent the rarely used words like "brittle" or "profundity". Lecture Corpus to examine the sequence of words as a sentence such as "good afternoon, ladies and gentlemen, welcome to the interact centre. My name is …, thank you for your attention, any questions."

The cognitive aspect is vital in the question cue because subjects must first comprehend the question. Then, different questions with the same answer could raise other cognition. For example, the answer to "Are you a scholar?" and "Are you a human?" are assumed to be "yes", but the cognition is different. Furthermore, the brain activity to respond with a "yes" is different from saying a "yes" with no intention. It will need the adaptability of the model to overcome the difference.

The multilanguage issue represents the other chance for SSI study. Many languages have been applied, e.g. English [6], [7], [16], [23], [26], Chinese [18], Hindi [22], Arabic [20], Korean [17], and Spanish [31]. Suppose the model can classify words from one language and their equivalent in another. In that case, this sparks the chance to build a complete set of EEG-based imagined speech datasets, regardless of the source languages of speech.

III. SIGNAL PREPROCESSING

Even though some restrictions, e.g. to move, were applied in the recording protocol, artefacts and noises are still unavoidable in EEG signals acquisition because of the low frequency and voltage, which are easily interfered with muscular/ocular artefacts and other noises. It still becomes a challenging issue for EEG studies in data cleaning without losing significant information/features for later analysis or pattern recognition. The noise removal must be performed before the downsampling step to prevent the downsampled

values from being falsely interpreted as noise. Previous studies used Independent Component Analysis (ICA) [49], [50] or artefact detectors based on the joint use of spatial-temporal features (ADJUST) [51].

If EEG acquisition used a high sampling rate, e.g., 1000Hz, most studies applied the downsampling process to lower the computational complexity. According to the Nyquist theorem and the brain signal frequency range of 0.5-100 Hz [48], 256 Hz covers more than twice the commonly observed maximum frequency. From the 37 reviewed studies, only seven studies used it ([11], [12], [17], [21], [26], [34], [42]). There will be thread off for the downsampling; the smaller the sample size, the lower the needed computation resources, but the more important features lost. Thus, using the original sample size could help observe the discriminative speech recognition features while also considering the available resources.

IV. SPECTRAL, SPATIAL, AND TEMPORAL ANALYSIS

Choosing only certain frequency bands, such as the alpha and beta bands, could decrease the number of features [20] because alpha and beta bands contain discriminative information. Statistic calculation (e.g. maximum value, average, standard deviation, kurtosis, and others) in beta, delta, and theta gave higher accuracy than the other bands in imagined vowel classification [13] with 81.25-98.75% accuracy for classifying the combined task, e.g., features from imagined speech state of "/a/" was combined with rest state, "/a/" and "/u/", and so on. However, the higher gamma band is not discriminative for speech imagery [11], [18], [31], [39] since muscular artefacts produce high gamma activity [52]. Additionally, from the reviewed papers, only a few studies (8 of 37) used it, as shown inFig. 7. Two of them reported that the gamma did not provide discriminative characteristics to decode speech [39] except for a speech with articulation [21].

When observing which frequency bands give the highest accuracy, there are some considerations about physiological activities associated with the specific bands. The high-frequency bands, e.g., beta or gamma, are dominated by muscular artefacts. Even though these waves correspond to the concentration or active attention [53], thus, they could have information for imagined speech recognition. Meanwhile, the low-frequency bands do not correlate with concentration [54]. The alpha waves correspond to relaxed awareness, the theta waves appear when the consciousness moves to drowsiness, and the delta waves are further away from concentration since they relate to deep sleep. Furthermore, low-frequency bands often get interferences with ocular artefacts or lead movements [55]. Since cognitive task, e.g. imagined speech, requires concentration, the role of low-frequency bands that yield higher accuracy than the high ones needs more examination.

EEG spatial analysis can give a better insight into which brain area has essential information for imagined speech recognition; the electrodes can be selected further. When a conversation happens, the most active regions are the auditory cortex, motor cortex, Broca's area, and Wernicke area [56]. Broca's area is in the inferior frontal gyrus, while Wernicke's area is in the superior temporal gyrus, with the arcuate fasciculus connecting both of them to build auditory-motor interaction [57], as illustrated inFig. 8. Some SSI research has

validated this brain region [7], [18], [22], [26]. Furthermore, in early speech production, the auditory potential existed in the superior temporal gyrus (STG) [58], located in the temporal area in both hemispheres.

Most of the reviewed studies reported high accuracy achievement by using the features extracted from frontal lobes (location of the Broca's area) [9], [10], [17], [26], [42] and temporal lobes (location of the Wernicke's area) [17], [21], [39], [42], followed by parietal lobes [9], [17], [26]. The others found that occipital lobes contributed to achieving the highest accuracy, which is obvious since they employed visual cues in their protocols [10], [39].

Suppose the EEG signals were treated as an event-related potential (ERP) during imagined speech production. The signals can be aligned to the onset of imagined speech production and then averaged to focus between the preceding and following onset time [59]. Previous studies have examined it and found that the potential in the left hemisphere arises one-two seconds before speech production [36] and appears a moment before cued speech is produced [60]. One study reported that speech-related ERP reached the peak at 350ms after the cue [12], and the highest significant level (from the paired t-test) existed between 400-600ms after the cue in the frontal area [17]. Moreover, from the five alphabets ("/a/", "/e/", "/i/", "/o/", and "/t/") classification result, data in 100-600ms after the cue gave the highest accuracy of 46.61% (chance level of 20%) [39].
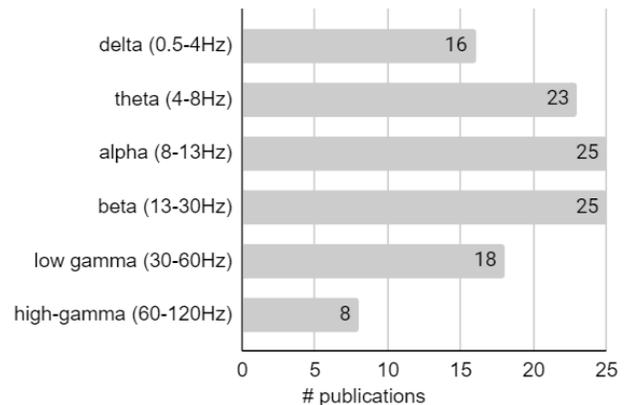
Fig. 7.   Observed Frequency Bands in EEG-Based SSI.
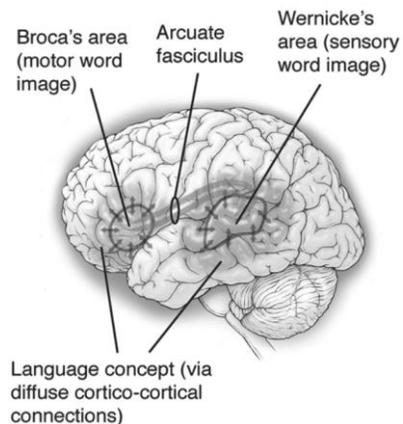
Fig. 8.   Language Organization in the Left Brain Hemisphere [57].

V. CLASSIFICATION AND FEATURE EXTRACTION TECHNIQUES

A. Non-Deep Learning

Most features were the result of the transformation to the frequency domain, as shown inFig. 9, e.g., Fourier, Wavelet, or Hilbert-Huang Transform. Common Spatial Patterns (CSP) and Principal Component Analysis (PCA) are other extraction methods. Filter bank CSP (FBCSP) [61] was the state-of-the-art for feature selection [35] or classification [36], [37], [62].

Since EEG signals are also time-series data, some studies applied auto-regression (AR) [11], [18], [19] or Mel Frequency Cepstral Coefficient (MFCC) [24], [41]. MFCC was more discriminative than the AR coefficient for vowel recognition in DaSalla's dataset [12] by yielding an accuracy of 75% [14] and performed better than statistics and nonlinear features in the KARAONE dataset [23]. MFCC gained an accuracy of 19.69% (chance level of 9.09% for 11 classes), while the accuracy of statistics features was 15.91%, and the accuracy of nonlinear features was 14.67% [24]. Several studies also treated EEG signals as a sequence of words by applying the Bag of Features (BoF) [30], [33]. In text pattern recognition, BoF was often used to represent a word existence using the feature values calculated from its previous words.

The other feature types are connectivity features, which relate to the brain's neural pathways when subjects perform a specific task, such as imagined speech production. They are structural [63], functional [64], and effective connectivity [65]. Structural connectivity refers to the tracts of white matter that physically interconnect brain regions. Functional connectivity refers to the statistical dependence (i.e. correlation) of time-series data between a pair of brain regions influenced by structural connectivity. Meanwhile, effective connectivity refers to a causal model representing the interactions between connected neurons.

Few studies of imagined speech recognition have considered applying functional or effective connectivity features. Qureshi et al. [43] used functional connectivity features fed into an extreme learning machine (ELM) to classify imagined speech of five words. These features were covariance and maximum linear cross-correlation coefficient (MaxLCor), with the same calculation for phase-only time-series data. MaxLCor is one of the spatial connectivity features to measure functional connectivity and extract EEG characteristics by calculating the normalized product of two time-series signals and then measuring their similarities [66]. They reported that covariance features yielded the highest accuracy of 87.90% on binary classification. Pawar et al. [44] also used MaxLCor combined with DWT features to classify imagined speech words [47] and achieved an accuracy of 40.64±2.45% (chance level 20%). Furthermore, Chengaiyan et al. [45] identified vowels and consonants by applying brain connectivity features on each frequency band; coherence [67] as functional connectivity and partial directed coherence (PDC) [68], direct transfer function (DTF) [69], and transfer entropy [70] as effective connectivity. They fed the features into deep learning methods, a Recurrent Neural Network (RNN) and a Deep Belief Network (DBN), where RNN gave lower accuracy of 72% than DBN with an accuracy of 80%.

Many different features were classified with Support Vector Machine (SVM) ([17], [18], [20], [39]) or Linear Discriminant Analysis (LDA) ([7], [13], [20]) as shown in Fig. 1, since both are good at separating discriminative values. Meanwhile, the Decision Tree (DT) [16] and its ensemble variant called Random Forest (RF) [31], [32], [34] were used due to their capability to distinguish between the classes and process a large number of features. The RF feed with Wavelet features outperformed SVM in the same dataset [31], [32].
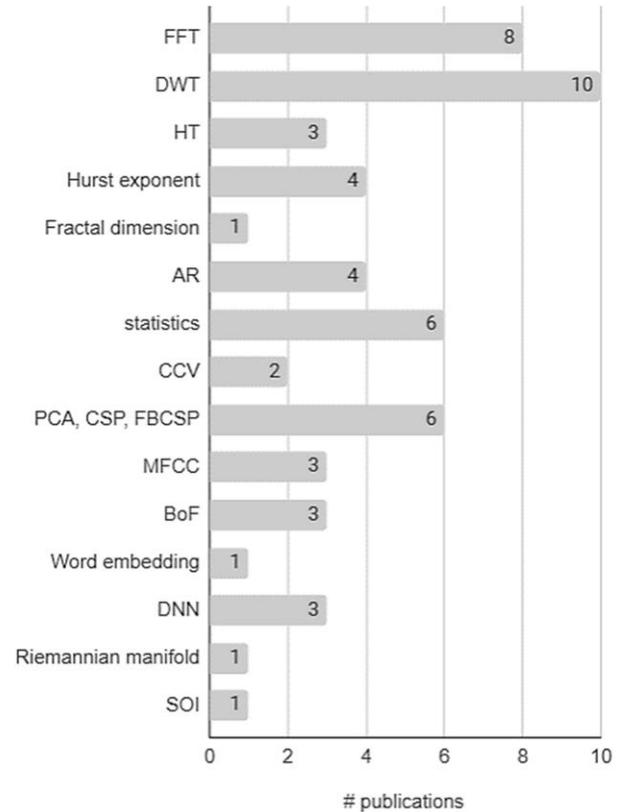
Fig. 9.   Features in EEG-Based SSI.
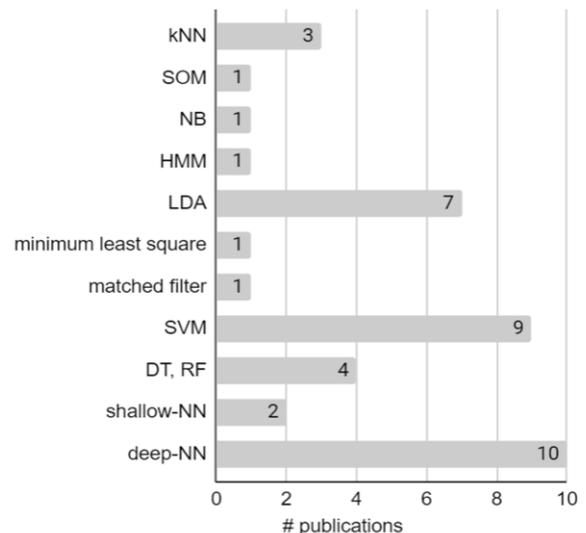
Fig. 10.   Classification Methods in EEG-Based SSI.

Another classification technique formerly used in imagined speech classification was k-Nearest Neighbor (kNN). When combined with MFCC features, it gained higher accuracy than SVM and the Hidden Markov Model (HMM) to identify the vowel of "/a/" and "/u/" from the DaSalla's dataset [12] with kNN's averaged accuracy of 86.89% compared with 75.83% and 70.56% respectively [14]. It also still outperformed SVM in classifying "/a/", "/e/", "/i/", "/o/", and "/u/" in the additional datasets of the same study [14]. Since kNN depends on the centroids for the k classes derived during training and the test dataset was classified based on the majority class of its neighbours, new samples may require k-NN retraining.

The shallow Artificial Neural Network (ANN), in the form of ELM, was also applied due to its architecture of layers and nodes that can achieve good generalization. With only using statistical values of the signals to classify vowels, ELM's accuracy was higher than SVM or LDA, in which ELM's accuracy was 87.07% compared with 51.07% for SVM and 81.98% for LDA [21]. Furthermore, with average power features to classify "yes" and "no", ANN also performed better than SVM and RF, with 92.18% compared with 83.07% and 79.95%, respectively [22].

The ANN's capacity for generalization motivated further research using deep learning (DL). While non-DL techniques depend on the features input, DL uses its layers to learn the data characteristics directly.

B. Deep Learning

Several attempts employed deep learning (DL) to extract features, e.g., Convolutional Neural Network (CNN) to extract spatial features and Recurrent Neural Network (RNN) for temporal characteristics [27]. These feature vectors were concatenated in the form of a channel covariance matrix as the input for Deep Autoencoder (DAE) classifier. Siamese Network increased the distance of different labelled samples and vice versa, which gave higher accuracy ($31.40 \pm 2.73\%$) [38] than the baselines [33], [34], [37], [62] by using the same Coretto's dataset [34].

Although DL has become the common feature and representation learning method, it is also well-known for being data-hungry. A small-size dataset might generate a final model that overfits. Therefore, only a few previous studies have used DL compared with non-DL, e.g., SVM or LDA. Furthermore, it is impossible to cover the whole language corpus, as subjects would feel uncomfortable for more than 30 minutes of EEG recording.

C. Model Evaluation

Although many reported models and their features claimed their highest accuracy, different datasets and experiment environments caused the accuracies to be incomparable since those factors can affect the chosen discriminative features and the model training process. Furthermore, the cognitive variance across subjects [71] needs more consideration since it caused the recognition model's accuracy to be more accurate when it was trained and tested in one specific subject's data but lower accuracy when it was tested to recognize the other's. This problem is called the inter-subject case. Hence, the model's accuracy in most brain signal studies was evaluated by each

person, i.e., the accuracy was calculated for each subject before the whole-averaged accuracy was reported. From the reviewed collected references, the validation was categorized into three versions.

The first type of validation is the usual cross-validation (CV) in machine learning treated subject-wise. The subjects' datasets were gathered into one massive dataset. Then it was split into training, validation, and testing parts with the configurable percentages for each part, e.g., training and validation took 60% and 20%, respectively, for k-fold cross-validation. The remaining 20% for each subject was kept unlabeled for the model to later predict in the testing session. In this type of CV, the training of the model uses the features from all subjects that could make the model achieve very high accuracy because it also learned a part of data whose the same variance as the testing dataset. Still, it became weak when facing the subjects' cognitive variance as an inter-subject problem exists. This case is essential to be solved but still unnoticed by many previous studies.

The second type is leave-one-subject-out cross-validation (LOSO-CV). It is similar to leave-one-out cross-validation (LOOCV) by using one subject's dataset as a validation dataset. This method aimed to measure the robustness of the trained model related to the inter-subject issue. It can prevent the model from peeking at the test dataset variance and overfitting. This method aimed to measure the robustness of the trained model related to the inter-subject issue. It also can be extended to be the leave-N-subject-out CV. Only a few gathered studies, listed in Table I, employed LOSO-CV.

The third type is leave-one-session-out cross-validation (LOSeO-CV) to face the intra-subject problem, i.e., the model's accuracy degrades when recognizing a new recording dataset of the same subjects whose datasets were used for training the model before. This validation type is the modification of LOSO-CV with a different perspective. Even though some studies were aware of the intra-subject problem [7], [15], they did not apply LOSeO-CV since their goal was to get higher accuracy than the baseline with the current data distribution only.

It is essential to note that the comparison of the higher accuracy achieved from the general CV with the lower accuracy gained from LOSO-CV was irrelevant. It is because LOSO-CV aimed to prepare the model to become adaptive to different subjects' data distribution due to the cognitive variance. Besides, the general CV only considers the current data distribution and potentially peeks distribution information from the same subject in the testing dataset. Thus, the model tends to be overfitting.

On the recognition of Arizona State University's dataset [26] in Table II, the researchers could train their model using the general CV on the deep learning model, and they achieved higher accuracy than the baseline (49%), with the highest accuracy being $96,79 \pm 4.19\%$ [28] for vowel recognition only. Another study also used deep learning with the general CV that boosted the accuracy of long-word recognition to 81.65%, higher than the baseline of 66.20%. Although, for short-long word discrimination, a deep learning implementation [29] still did not achieve higher accuracy than the baseline (77.60% of

80.10%). Meanwhile, another research [30] used LOSO-CV while recognizing only the long words. It reported a lower accuracy of 62,99 ± 4.78% than the reported accuracy in Table II. There was still no further observation for short-word classification with higher accuracy than the baseline. Moreover, deep learning also successfully yielded higher accuracy on the KARAONE dataset [23] for multi-class classification (i.e. not a binary classification), as shown in Table III, with an accuracy of 57.15% [25] higher than the baseline (33.3%) [24]. These KARAONE dataset classification studies used the general CV for validation.

Before the development of Coretto's dataset, Torres-Garcia et al. constructed an EEG-based imagined speech dataset [31] with five similar words to Coretto's dataset; Torres-Garcia used "seleccionar" ("select") rather than "adelante" ("forward") and "atrás" ("backward"). The non-deep-learning model achieved the highest accuracy (70.33%) for this dataset, with a general CV for its validation, as shown in Table IV. One deep learning implementation [30] with a transfer learning approach still gained slightly lower accuracy (65.65%) validated by LOSO-CV.

Similar to Arizona State University's dataset, the studies on Coretto's dataset also achieved the highest accuracy by implementing a deep learning model (30% for vowel classification [35] and 62.37% for word classification [36]). It was validated by a general CV, as shown in Table V. Further research validated the deep learning model with LOSO-CV to classify the vowels. It successfully yielded higher accuracy (32.75%) [37] than the baseline (30%) [35] in recognizing vowels. Thus, the model became quite robust since it could still accurately recognize the unseen subject's dataset.

The other reviewed studies used a general CV for validation; DaSalla's EEG-based imagined speech dataset consists of "a" and "u" speech [12], and Dzmura's consists of "ba" and "ku" [9]. There was still no deep learning exploration for DaSalla's dataset. Although, the researchers can achieve higher accuracy than the baseline for binary classification, as shown in Table VI, by using different feature extraction algorithms. Meanwhile, the baseline study of Dzmura's dataset still had the highest accuracy (74.25%) with the spectral feature and matched-filter classification.

From observing several EEG-based imagined speech datasets, some deep-learning studies yielded higher accuracy than the baseline studies validated by general CV, e.g. in Arizona State University's, Torres-Garcia's, Coretto's, and KARAONE datasets. Although, the deep learning models gained lower accuracy with LOSO-CV, e.g. in Arizona State University's dataset, Torres-Garcia's dataset, and Coretto's dataset. Some studies applied the transfer learning approach to build a more robust model with LOSO-CV validation; one successfully gained higher accuracy [37], but the other still got slightly lower accuracy [30]. Nevertheless, the transfer learning approach could have the capability to train the robust model. Additionally, the non-deep-learning models could gain higher accuracy when the informative features fed into them, as in DaSalla's and Dzmura's datasets. The accuracy of the trained models from the same datasets validated with the general CV in Table II-VII can become the benchmark for further studies.

TABLE II. ACCURACY WITH GENERAL CROSS-VALIDATION FOR ARIZONA STATE UNIVERSITY'S DATASET

Ref.	Accuracy (%)			
	Vowel	Short word	Long word	Short vs long word
[26]	49.00	**50.10**	66.20	**80.10**
[27]	-	-	**81.65**	-
[28]	**96.79**	-	-	-
[29]	-	-	-	77.60

TABLE III. ACCURACY WITH GENERAL CROSS-VALIDATION FOR KARAONE DATASET

Ref.	Binary-class Accuracy (%)					Multi-class Accuracy (%)
	Bila-bial	Nasal	C/V	/uw/	/iy/	
[23]	56.64	63.50	18.08	79.16	59.6	-
[24]	-	-	-	-	-	33.3%
[25]	-	-	-	-	-	**57.15%**

TABLE IV. ACCURACY WITH GENERAL CROSS-VALIDATION FOR TORRES-GARCIA'S DATASET

Ref.	Accuracy (%)
[31]	41.21
[32]	**70.33**
[33]	61.02

TABLE V. ACCURACY WITH GENERAL CROSS-VALIDATION FOR CORETTO'S DATASET

Ref.	Accuracy (%)	
	Vowel	Word
[34]	22.72	19.60
[35]	**30.00**	24.97
[36]	-	**62.37**
[38]	-	31.40

TABLE VI. ACCURACY WITH GENERAL CROSS-VALIDATION FOR DASALLA'S DATASET

Ref.	Binary-class Accuracy (%)		
	/a/-rest	/u/-rest	/a/-/u/
[12]	72.33	78	62.67
[14]	75.00	**93.83**	**91.83**
[13]	**75.83**	77.5	72.5

TABLE VII. ACCURACY WITH GENERAL CROSS-VALIDATION FOR D'ZMURA'S DATASET

Ref.	Accuracy (%)
[9]	**74.25**
[10]	58.05
[11]	68.83

VI. CURRENT CHALLENGES

A. Laboratory Environment

In the imagined speech decoding research, EEG signals were recorded in a conducive laboratory environment with a proper procedure to minimize the noise and artefacts. When the interface is intended to be practical, e.g., for patients at the hospital or as an in-house assistive tool, it will face a noisier environment and inevitable artefacts.

The previously discussed artefact removal approaches still require validation since they only focused on increasing the accuracy without reexamining the effectiveness of artefact removal. The acquisition should have relaxed restrictions, such

as allowing subjects to swallow or blink during imagined speech production, to validate the result of artefact removal.

B. Related Channels

Much spatial information could be observed using many electrodes, but it will be less convenient, less impractical, and have bigger feature dimensionality. This issue also happens in motor imagery BCI. Several approaches to overcome it are channel selection, spatial filter, and feature selection [32]. Feature selection aimed to select the most discriminative features, spatial filter to extract characteristics from channels employed, and channel selection to choose several channels with similar/better accuracy as the whole channels.

C. Time-Lock

In EEG signals, there is temporal information related to the onset time and spatial features related to the brain area. Spatial and temporal information of overt speech and imagined speech correlated [42]; the spatial pattern is not significantly different, but the temporal one is. It is due to the difficulty in determining the time-lock of the imagined speech, compared with the overt speech whose time-lock is easily detected from the voice. Besides, the time-lock can be different in the different sessions.

D. Intra-Subject and Inter-Subject Problem

One main problem of EEG-based imagined speech studies is the limited speech data. Should the recording cover the whole vocabulary in a language, the subjects will be exhausted, and it will need a very long time. Besides, there are different patterns produced even if a person imagines the same speech, i.e., the inter-subject issue [15], [41], or if he imagines the same speech at a different time, i.e., the intra-subject problem [42]. Thus, the model must be adaptive to recognize new data from new sessions/subjects without training from scratch.

One approach to solving the adaptation issue is transfer learning (TL). Traditional machine learning assumes that the distribution of the present learned data and the future data are the same. In contrast, TL assumes their domain is different, or the future labelling task may differ. Some studies reported that TL did not decrease the accuracy [36], [37]. However, as the accuracy was still poor (35-60%), it needs further observation.

E. Connectivity

The brain works as a neural pathway, and its existing connectivity can contain informative features for cognitive tasks, including speech imagery. Functional connectivity (correlation between brain areas) and effective connectivity (the causal model of brain areas' interaction) become potential characteristics to help identify imagined speech.

F. Subject Limitation

Current SSI studies were limited to healthy subjects. The subjects' brains must be in good condition to record signals. When the study includes brain-impaired participants, the problem-related brain area might affect the data acquired and its recognition accuracy, which needs more profound observation. Additionally, observing subjects with health issues, such as the locked-in syndrome (LIS) patient, was also challenging. The moment the LIS subject began producing the speech could not be identified precisely, although the subject has been instructed to confirm the speech production attempt

[72]. Other health disorders, e.g., Aphasia, Apraxia, Dysarthria, laryngectomy (i.e. removal of the larynx by operation procedure), and tracheostomy, also have specific conditions. For example, the brain activity in speech production for a new laryngectomy patient may differ from a one-year patient.

G. Online Learning

There would be a requirement for online learning where the model training is simultaneous with EEG recording. It could exploit the users' feedback to retrain the model and recognize the pattern more accurately. Moreover, the subjects could also be trained to modify their brain waves to adapt to BCI [73]. Currently, most studies used the offline training (, i.e., outside the recording session). Although few studies performed online learning [7], [20], the performance was low and inconsistent.

VII. CONCLUSIONS

This review discussed the pipeline of EEG-based SSI, which consists of signal acquisition, signal preprocessing, feature extraction, and classification, to see problems that often arise in each step. The acquisition process needs a proper design of subject inclusion, cue format, and speech types according to the purpose of the study, including the challenges that need answers to apply the decoding in the real world while maintaining the high accuracy achieved in a lab environment. These challenges deal with handling noises and artefacts, the trade-off between the number of channels and spatial features and onset time determination to gain discriminative temporal characteristics. Besides, variance shifts due to different recording sessions or users that demand an adaptive model and its validation need consideration. The inclusion of brain-impaired subjects and the potential of online learning could make the interface more applicable. To conclude, this review suggests that it is crucial to start by building the proper pipeline and taking problems in each step into consideration to overcome the challenges. High accuracy is insufficient to make the model applicable in the real world.

REFERENCES

[1] WHO, "World Report On Disability," 2011. [Online]. Available:https://www.who.int/disabilities/world_report/2011/report.pdf

[2] ILO and LPEM FEB UI, "Memetakan Penyandang Disabilitas (PD) di Pasar Tenaga Kerja Indonesia," 2017. [Online]. Available: https://www.ilo.org/jakarta/whatwedo/publications/WCMS_587668/lang--en/index.htm

[3] N. Bhattacharyya, "The prevalence of voice problems among adults in the United States," Laryngoscope, vol. 124, no. 10, pp. 2359–2362, Oct. 2014, doi: 10.1002/lary.24740.

[4] F. Bocquelet, T. Hueber, L. Girin, S. Chabardès, and B. Yvert, "Key considerations in designing a speech brain-computer interface," J. Physiol., vol. 110, no. 4, pp. 392–401, Nov. 2016, doi: 10.1016/j.jphysparis.2017.07.002.

[5] J. A. Gonzalez-Lopez, A. Gomez-Alanis, J. M. Martin Donas, J. L. Perez-Cordoba, and A. M. Gomez, "Silent Speech Interfaces for Speech Restoration: A Review," IEEE Access, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3026579.

[6] P. Suppes, Z.-L. Lu, and B. Han, "Brain wave recognition of words," Proc. Natl. Acad. Sci., vol. 94, no. 26, pp. 14965–14969, Dec. 1997, doi: 10.1073/pnas.94.26.14965.

[7] M. Wester, "Unspoken Speech: Speech Recognition Based On Electroencephalography," Universität Karlsruhe, 2006. [Online]. Available: https://www.researchgate.net/publication/36453500_Unspoken_Speech_-_Speech_Recognition_based_on_Electroencephalography

[8] A. Porbadnigk, M. Wester, J. P. Calliess, and T. Schultz, "EEG-Based Speech Recognition - Impact of Temporal Effects," in Proceedings of the International Conference on Bio-inspired Systems and Signal Processing, 2009, no. January, pp. 376–381. doi: 10.5220/0001554303760381.

[9] M. D'Zmura, S. Deng, T. Lappas, S. Thorpe, and R. Srinivasan, "Toward EEG Sensing of Imagined Speech," in Proceedings of the 13th International Conference on Human-Computer Interaction, 2009, pp. 40–48. doi: 10.1007/978-3-642-02574-7_5.

[10] S. Deng, R. Srinivasan, T. Lappas, and M. D'Zmura, "EEG classification of imagined syllable rhythm using Hilbert spectrum methods," J. Neural Eng., vol. 7, no. 4, p. 046006, Aug. 2010, doi: 10.1088/1741-2560/7/4/046006.

[11] K. Brigham and B. V. K. V. Kumar, "Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy," in 2010 4th International Conference on Bioinformatics and Biomedical Engineering, Jun. 2010, pp. 1–4. doi: 10.1109/ICBBE.2010.5515807.

[12] C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," Neural Networks, vol. 22, no. 9, pp. 1334–1339, Nov. 2009, doi: 10.1016/j.neunet.2009.05.008.

[13] B. M. Idrees and O. Farooq, "Vowel classification using wavelet decomposition during speech imagery," in 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), Feb. 2016, pp. 636–640. doi: 10.1109/SPIN.2016.7566774.

[14] A. Riaz, S. Akhtar, S. Iftikhar, A. A. Khan, and A. Salman, "Inter comparison of classification techniques for vowel speech imagery using EEG sensors," in The 2014 2nd International Conference on Systems and Informatics (ICSAI 2014), Nov. 2014, no. Icsai, pp. 712–717. doi: 10.1109/ICSAI.2014.7009378.

[15] X. Chi, J. B. Hagedorn, D. Schoonover, and M. D. Zmura, "EEG-Based Discrimination of Imagined Speech Phonemes," Int. J. Bioelectromagn., vol. 13, no. 4, pp. 201–206, 2011, [Online]. Available: https://pdfs.semanticscholar.org/b74f/c325556d1a7b5eb05fe90cde1f0c891357a3.pdf

[16] C. M. Spooner, E. Viirre, and B. Chase, "From Explicit to Implicit Speech Recognition," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8027 LNAI, 2013, pp. 502–511. doi: 10.1007/978-3-642-39454-6_54.

[17] T. Kim, J. Lee, H. Choi, H. Lee, I.-Y. Kim, and D. P. Jang, "Meaning based covert speech classification for brain-computer interface based on electroencephalography," in 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), Nov. 2013, pp. 53–56. doi: 10.1109/NER.2013.6695869.

[18] L. Wang, X. Zhang, X. Zhong, and Y. Zhang, "Analysis and classification of speech imagery EEG for BCI," Biomed. Signal Process. Control, vol. 8, no. 6, pp. 901–908, Nov. 2013, doi: 10.1016/j.bspc.2013.07.011.

[19] Y. Song and F. Sepulveda, "Classifying speech related vs. idle state towards onset detection in brain-computer interfaces overt, inhibited overt, and covert speech sound production vs. idle state," in 2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings, Oct. 2014, pp. 568–571. doi: 10.1109/BioCAS.2014.6981789.

[20] M. Salama, L. Elsherif, H. Lashin, and T. Gamal, "Recognition of Unspoken Words Using Electrode Electroencephalograhic Signals," in COGNITIVE 2014 : The Sixth International Conference on Advanced Cognitive Technologies and Applications, 2014, pp. 51–55. [Online]. Available: https://bu.edu.eg/portal/uploads/Engineering, Shoubra/Electrical Engineering/3513/publications/May ahmed salama mohamed_Recog of unspoken words.pdf

[21] B. Min, J. Kim, H. Park, and B. Lee, "Vowel Imagery Decoding toward Silent Speech BCI Using Extreme Learning Machine with Electroencephalogram," Biomed Res. Int., pp. 1–11, 2016, doi: 10.1155/2016/2618265.

[22] A. Balaji et al., "EEG-based classification of bilingual unspoken speech using ANN," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jul. 2017, pp. 1022–1025. doi: 10.1109/EMBC.2017.8037000.

[23] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Apr. 2015, pp. 992–996. doi: 10.1109/ICASSP.2015.7178118.

[24] C. Cooney, R. Folli, and D. Coyle, "Mel Frequency Cepstral Coefficients Enhance Imagined Speech Decoding Accuracy from EEG," in 29th Irish Signals and Systems Conference (ISSC), Jun. 2018, pp. 1–7. doi: 10.1109/ISSC.2018.8585291.

[25] J. T. Panachakel, A. G. Ramakrishnan, and T. V. Ananthapadmanabha, "Decoding Imagined Speech using Wavelet Features and Deep Neural Networks," in 2019 IEEE 16th India Council International Conference (INDICON), Dec. 2019, pp. 1–4. doi: 10.1109/INDICON47234.2019.9028925.

[26] C. H. Nguyen, G. K. Karavas, and P. Artemiadis, "Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features," J. Neural Eng., vol. 15, no. 1, p. 016002, Feb. 2018, doi: 10.1088/1741-2552/aa8235.

[27] P. Saha and S. Fels, "Hierarchical Deep Feature Learning for Decoding Imagined Speech from EEG," in Proceedings of the AAAI Conference on Artificial Intelligence, Jul. 2019, vol. 33, pp. 10019–10020. doi: 10.1609/aaai.v33i01.330110019.

[28] M. Parhi and A. H. Tewfik, "Classifying Imaginary Vowels from Frontal Lobe EEG via Deep Learning," in 2020 28th European Signal Processing Conference (EUSIPCO), Jan. 2021, pp. 1195–1199. doi: 10.23919/Eusipco47968.2020.9287599.

[29] J. T. Panachakel, A. G. Ramakrishnan, and T. V. Ananthapadmanabha, "A Novel Deep Learning Architecture for Decoding Imagined Speech from EEG," IEEE Austria Int. Biomed. Eng. Conf. (AIBEC 2019), Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.09374

[30] M. Jiménez-Guarneros and P. Gómez-Gil, "Standardization-refinement domain adaptation method for cross-subject EEG-based classification in imagined speech recognition," Pattern Recognit. Lett., vol. 141, pp. 54–60, Jan. 2021, doi: 10.1016/j.patrec.2020.11.013.

[31] A. A. Torres-García, C. A. Reyes-García, and L. Villaseñor-Pineda, "Toward a Silent Speech Interface Based on Unspoken Speech," in Proceedings of the International Conference on Bio-inspired Systems and Signal Processing, 2012, pp. 370–373. doi: 10.5220/0003769603700373.

[32] A. A. Torres-García, C. A. Reyes-García, L. Villaseñor-Pineda, and G. García-Aguilar, "Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification," Expert Syst. Appl., vol. 59, pp. 1–12, Oct. 2016, doi: 10.1016/j.eswa.2016.04.011.

[33] J. S. García-Salinas, L. Villaseñor-Pineda, C. A. Reyes-García, and A. A. Torres-García, "Transfer learning in imagined speech EEG-based BCIs," Biomed. Signal Process. Control, vol. 50, pp. 151–157, Apr. 2019, doi: 10.1016/j.bspc.2019.01.006.

[34] G. A. Pressel Coretto, I. E. Gareis, and H. L. Rufiner, "Open access database of EEG signals recorded during imagined speech," in 12th International Symposium on Medical Information Processing and Analysis, Jan. 2017, p. 1016002. doi: 10.1117/12.2255697.

[35] C. Cooney, A. Korik, R. Folli, and D. Coyle, "Evaluation of Hyperparameter Optimization in Machine and Deep Learning Methods for Decoding Imagined Speech EEG," Sensors, vol. 20, no. 16, p. 4629, Aug. 2020, doi: 10.3390/s20164629.

[36] C. Cooney, A. Korik, R. Folli, and D. H. Coyle, "Classification of Imagined Spoken Word-pairs using Convolutional Neural Networks," in Proceedings of the 8th Graz Brain Computer Interface Conference 2019, 2019, pp. 338–343. doi: 10.3217/978-3-85125-682-6-62.

[37] C. Cooney, R. Folli, and D. Coyle, "Optimizing Layers Improves CNN Generalization and Transfer Learning for Imagined Speech Decoding from EEG," in 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Oct. 2019, pp. 1311–1316. doi: 10.1109/SMC.2019.8914246.

[38] D. Y. Lee, M. Lee, and S. W. Lee, "Classification of Imagined Speech Using Siamese Neural Network," in 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Oct. 2020, pp. 2979–2984. doi: 10.1109/SMC42975.2020.9282982.

[39] Y. Wang, P. Wang, and Y. Yu, "Decoding English Alphabet Letters Using EEG Phase Information," Front. Neurosci., vol. 12, p. 62, Feb. 2018, doi: 10.3389/fnins.2018.00062.

[40] G. Krishna, C. Tran, J. Yu, and A. H. Tewfik, "Speech Recognition with No Speech or with Noisy Speech," in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2019, pp. 1090–1094. doi: 10.1109/ICASSP.2019.8683453.

[41] G. Krishna, C. Tran, Y. Han, M. Carnahan, and A. H. Tewfik, "Speech Synthesis Using EEG," in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2020, pp. 1235–1238. doi: 10.1109/ICASSP40776.2020.9053340.

[42] S.-H. Lee, M. Lee, and S.-W. Lee, "EEG Representations of Spatial and Temporal Features in Imagined Speech and Overt Speech," in Pattern Recognition, 2020, pp. 387–400. doi: 10.1007/978-3-030-41299-9_30.

[43] M. N. I. Qureshi, B. Min, H.-J. Park, D. Cho, W. Choi, and B. Lee, "Multiclass Classification of Word Imagination Speech With Hybrid Connectivity Features," IEEE Trans. Biomed. Eng., vol. 65, no. 10, pp. 2168–2177, Oct. 2018, doi: 10.1109/TBME.2017.2786251.

[44] D. Pawar and S. Dhage, "Imagined Speech Classification using EEG based Brain-Computer Interface," in 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Apr. 2022, pp. 662–666. doi: 10.1109/CSNT54456.2022.9787644.

[45] S. Chengaiyan, A. S. Retnapandian, and K. Anandan, "Identification of vowels in consonant–vowel–consonant words from speech imagery based EEG signals," Cogn. Neurodyn., vol. 14, no. 1, pp. 1–19, Feb. 2020, doi: 10.1007/s11571-019-09558-5.

[46] R. D. Kent, G. Weismer, J. F. Kent, and J. C. Rosenbek, "Toward Phonetic Intelligibility Testing in Dysarthria," J. Speech Hear. Disord., vol. 54, no. 4, pp. 482–499, Nov. 1989, doi: 10.1044/jshd.5404.482.

[47] D. Pal, S. Palit, and A. Dey, "Brain Computer Interface: A Review," in Lecture Notes in Electrical Engineering, vol. 786, 2022, pp. 25–35. doi: 10.1007/978-981-16-4035-3_3.

[48] B. Onaral and A. Cohen, "Biomedical Signals," in Medical Devices and Systems, 3rd ed., J. D. Bronzino, Ed. CRC Press, 2006, pp. 1-1-1–22. doi: 10.1201/9781420003864.sec1.

[49] S. Vorobyov and A. Cichocki, "Blind noise reduction for multisensory signals using ICA and subspace filtering, with application to EEG analysis," Biol. Cybern., vol. 86, no. 4, pp. 293–303, Apr. 2002, doi: 10.1007/s00422-001-0298-6.

[50] S. Cruces, L. Castedo, and A. Cichocki, "Novel blind source separation algorithms using cumulants," in 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100), Dec. 2002, vol. 5, no. 1–4, pp. 3152–3155. doi: 10.1109/ICASSP.2000.861206.

[51] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," Psychophysiology, vol. 48, no. 2, pp. 229–240, Feb. 2011, doi: 10.1111/j.1469-8986.2010.01061.x.

[52] S. D. Muthukumaraswamy, "High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations," Front. Hum. Neurosci., vol. 7, 2013, doi: 10.3389/fnhum.2013.00138.

[53] P. Georgieva, F. Silva, M. Milanova, and N. Kasabov, "EEG Signal Processing for Brain–Computer Interfaces," in Springer Handbook of Bio-/Neuroinformatics, no. June 2016, Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 797–812. doi: 10.1007/978-3-642-30574-0_46.

[54] S. Sanei and J. A. Chambers, "Introduction to EEG," in EEG Signal Processing, West Sussex, England: John Wiley & Sons Ltd, 2013, pp. 1–34. doi: 10.1002/9780470511923.ch1.

[55] A. van Boxtel, "Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles," Psychophysiology, vol. 38, no. 1, p. S004857720199016X, Jan. 2001, doi: 10.1017/S004857720199016X.

[56] G. Hesslow, "Conscious thought as simulation of behaviour and perception," Trends Cogn. Sci., vol. 6, no. 6, pp. 242–247, Jun. 2002, doi: 10.1016/S1364-6613(02)01913-7.

[57] E. F. Chang, K. P. Raygor, and M. S. Berger, "Contemporary model of language organization: an overview for neurosurgeons," J. Neurosurg., vol. 122, no. 2, pp. 250–261, Feb. 2015, doi: 10.3171/2014.10.JNS132647.

[58] G. Hickok and D. Poeppel, "Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language," Cognition, vol. 92, no. 1–2, pp. 67–99, May 2004, doi: 10.1016/j.cognition.2003.10.011.

[59] T. Schultz, M. Wand, T. Hueber, D. J. Krusienski, C. Herff, and J. S. Brumberg, "Biosignal-Based Spoken Communication: A Survey," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 25, no. 12, pp. 2257–2271, Dec. 2017, doi: 10.1109/TASLP.2017.2752365.

[60] J. Prescott and G. Andrews, "Early and late components of the contingent negative variation prior to manual and speech responses in stutterers and non-stutterers," Int. J. Psychophysiol., vol. 2, no. 2, pp. 121–130, Nov. 1984, doi: 10.1016/0167-8760(84)90005-9.

[61] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface," in 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Jun. 2008, pp. 2390–2397. doi: 10.1109/IJCNN.2008.4634130.

[62] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," Hum. Brain Mapp., vol. 38, no. 11, pp. 5391–5420, Nov. 2017, doi: 10.1002/hbm.23730.

[63] K. J. Friston, "Functional and Effective Connectivity: A Review," Brain Connect., vol. 1, no. 1, pp. 13–36, 2011, doi: 10.1089/brain.2011.0008.

[64] P. Babaeeghazvini, L. M. Rueda-Delgado, J. Gooijers, S. P. Swinnen, and A. Daffertshofer, "Brain Structural and Functional Connectivity: A Review of Combined Works of Diffusion Magnetic Resonance Imaging and Electro-Encephalography," Front. Hum. Neurosci., vol. 15, no. October, 2021, doi: 10.3389/fnhum.2021.721206.

[65] K. E. Stephan and K. J. Friston, "Analyzing effective connectivity with functional magnetic resonance imaging," WIREs Cogn. Sci., vol. 1, no. 3, pp. 446–459, May 2010, doi: 10.1002/wcs.58.

[66] C. Meisel and C. Kuehn, "Scaling Effects and Spatio-Temporal Multilevel Dynamics in Epileptic Seizures," PLoS One, vol. 7, no. 2, p. e30371, Feb. 2012, doi: 10.1371/journal.pone.0030371.

[67] R. W. Thatcher, D. North, and C. Biver, "EEG and intelligence: Relations between EEG coherence, EEG phase delay and power," Clin. Neurophysiol., vol. 116, no. 9, pp. 2129–2141, Sep. 2005, doi: 10.1016/j.clinph.2005.04.026.

[68] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," Biol. Cybern., vol. 84, no. 6, pp. 463–474, May 2001, doi: 10.1007/PL00007990.

[69] M. J. Kaminski and K. J. Blinowska, "A new method of the description of the information flow in the brain structures," Biol. Cybern., vol. 65, no. 3, pp. 203–210, 1991, doi: 10.1007/BF00198091.

[70] T. Schreiber, "Measuring Information Transfer," Phys. Rev. Lett., vol. 85, no. 2, pp. 461–464, Jul. 2000, doi: 10.1103/PhysRevLett.85.461.

[71] D. Dash, P. Ferrari, and J. Wang, "Spatial and Spectral Fingerprint in the Brain: Speaker Identification from Single Trial MEG Signals," in Interspeech 2019, Sep. 2019, vol. 2019-Septe, pp. 1203–1207. doi: 10.21437/Interspeech.2019-3105.

[72] J. S. Brumberg, E. J. Wright, D. S. Andreasen, F. H. Guenther, and P. R. Kennedy, "Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex," Front. Neurosci., vol. 5, no. 65, May 2011, doi: 10.3389/fnins.2011.00065.

[73] N. Birbaumer et al., "The thought translation device (TTD) for completely paralyzed patients," IEEE Trans. Rehabil. Eng., vol. 8, no. 2, pp. 190–193, Jun. 2000, doi: 10.1109/86.847812.

Deeply Learned Invariant Features for Component-based Facial Recognition

Adam Hassan[1], Serestina Viriri[2]

Sudan University of Science and Technology, College of Computer Science & Information Technology, Khartoum, Sudan[1]
University of KwaZulu-Natal, School of Mathematics, Statistics & Computer Science, Durban, South Africa[2]

Abstract—**Face recognition underage variation is a challenging problem. It is a difficult task because ageing is an intrinsic variation, not like pose and illumination, which can be controlled. We propose an approach to extract invariant features to improve facial recognition using facial components. Can facial recognition over age progression be improved by resizing independently each individual facial component? The individual facial components: eyes, mouth, and nose were extracted using the Viola-Jones algorithm. Then we utilize the eyes region rectangle with upper coordinates to detect the forehead and lower coordinates with the nose rectangle to detect the cheeks. The proposed work uses Convolutional Neural Network with an ideal input image size for each facial component according to many experiments. We sum up component scores by applying weighted fusion for a final decision. The experiments prove that the nose component provides the highest score contribution among other ones, and the cheeks are the lowest. The experiments were conducted on two different facial databases-MORPH, and FG-NET databases. The proposed work achieves a state-of-the-art accuracy that reaches 100% on the FG-NET dataset and the results obtained on the MORPH dataset outperform the accuracy results of the related works in the literature.**

Keywords—Invariant features; facial components; facial recognition; convolutional neural network; weighted fusion

I. INTRODUCTION

Facial recognition basically is the activity of verifying or identifying a person's identity through facial characteristics. It captures facial features, analyses, and performs patterns comparison to know and determine the identity, or to decide whether the person is the correct one. However, the difficulties that emerge in designing an invariant face recognition system comprise variations in illumination, pose, and age.

One of the most distinguished biometric attributes is facial features, which are the most compatible with Machine Readable Travel Documents [1]. The advantage of facial features is that they can be captured at a distance without permission.

Facial ageing is not a controllable process throughout human life and cannot be avoidable, not like other variations which are flexible to handle during the image acquisition period [2]. Besides, facial ageing images are not free from extrinsic variations such as illumination, pose, and expression which add up to the challenge of finding an appropriate method for age-invariant facial recognition [3]. These

approaches succeeded in finding distinctive features. The approaches are divided mainly into two groups. One of the groups is local appearance-based techniques which are used to extract local features. The whole face is portioned into small patches [4]. The other group is Key-points-based techniques specified to detect points of interest in the full-face image.

A local binary pattern (LBP) is a texture extractor used to extract distinctive features from objects [5]. It is applied in many applications like face recognition [6], texture segmentation, texture classification, and facial expression recognition. Khoi et al. [7] present a fast face recognition model using LBP and its variants based on the content approach. Karaaba et al. [8] proposed multi-HOG that combined different histograms of oriented gradients for robust face recognition.

Prince et al. in [9] used LDA based on probabilistic linear discriminant analysis (PLDA) to maximize discriminability, which is applicable for face recognition with pose variation. Perlibakas et al. [10] used Gabor features to eliminate redundancy to obtain the best descriptors for face recognition, then, used cosine similarity for evaluation.

This paper illustrates subject recognition through the facial components during ageing and relevant variations due to illumination, pose, expressions, resolutions and facial image occultation; thus, face recognition with both intrinsic and extrinsic variation is a challenging process. Moreover, extracting facial components such as the forehead and cheeks is additional work.

Several researchers performed double feature extractors for better representation as a single concatenated vector [2] or separated to vote the robust to increase the accuracy and performance simultaneously [11]. However, features extracted using hand-crafted descriptors do not have sufficient capacity to adequately represent the appearance of the face [12]. The effectiveness of deep learning has been proven to extract highly discriminative features for promising results in the field of facial ageing research [13]. A suitable alternative is to train a convolutional neural network model to identify and extract discriminative features.

The remainder of the paper is organized as follows. Section II gives a brief literature review and related work. Section III describes our methodology and techniques. Section IV discusses the experiments and analysis. Section V draws paper conclusions and future work.

II. Literature Review

Research in recognition based on facial components has seen less attention compared to the attention being given to developing approaches that use global face descriptors. For instance, holistic representations like densely sampled feature extractors such as Local Binary Patterns (LBP) [14] and Scale Invariant Feature Transforms (SIFT) descriptors [15] are heavily used in automated face recognition [16]. In contrast, research concerning facial component-based representations such as features that are extracted from specific facial components is rarely addressed in the literature.

Component-based facial recognition is an alternative to using the full-face approach. A distinguished comparison between the two approaches is found in [17]. The authors proposed a component-based and two others full-face methods. Experiments showed the superiority of the components approach on the whole face with the consideration of the robustness against pose variations.

Wang Lijia et al. [18] used Procrustes analysis, which is insensitive to rotation, translation and scale to align facial components. They used a random measurement matrix to extract components' features and applied the Gradient projection Algorithm for classification. The method proved that component-based approaches matched better than holistic approaches in terms of accuracy rate.

Early research [19] illustrated that the upper face components are more discriminant than other components. Authors in [18] solved the one training sample problem by using a component-based linear discriminant analysis (LDA) method through five facial parts to construct component bunches.

Multi-feature extraction is also used to extract textural and shape from facial components such as eyes, mouth, and nose. The algorithm is flexible enough to be satisfied with the non-occluded facial parts to perform the recognition [20].

Authors in [21] proposed decision-level fusion containing 34 region classifiers decision-level fusion is performed with majority voting.

Boussaad et al. [22] used a pre-trained CNN model on resized components to cope with Alex Net input layer size, then Discriminant Correlation Analysis for fusion and Support Vector Machine for classification.

Component-based methods have proven effective when used to handle age-invariant features. For instance, [22] applied a Discriminant Correlation Analysis (DCA) as a feature-level fusion on separated components features and performed facial classification using a Support Vector Machine (SVM) and obtained a 97.87% as recognition accuracy rate.

This work proposes a component-based approach for age-invariant face recognition using deeply learned features extracted from separated components (eyes, nose, mouth, forehead, and cheeks), then matching score level fusion is performed, and cosine similarity is used for classification.

III. Methodology and Techniques

We propose a methodology to study invariant features for component-based facial recognition during age progression. Fig. 1 illustrates a flow diagram of our proposed methodology that consists of sequential steps. Firstly, facial images are pre-processed through facial component detection, cropping, resizing, and data augmentation. Secondly, the convolutional neural Network CNN extracts discriminative features through the augmented facial components for training the CNN model for validation and gaining the accuracy of testing sets. Thirdly, we fuse the scores obtained by every component for a final overall decision.

A. Facial Datasets

In this work, we use the two freely available datasets MORPH (Album II) [23] and FGNET [24] to perform age-invariant facial recognition. The MORPH dataset consists of 55,134 facial images belonging to 13,617 classes with an age range that extends from 16 to 77 years. The FGNET dataset includes 1002 images of a total number of 82 classes, and its age range begins from 0 to 69 years. The two sets of data that include age variation are also exposed to lighting, expressions, and head position. The Morph dataset contains 9260 images that belong to 754 classes that differ in age. The images are categorized into classes; each contains images of different ages, not exceeding the 5 years age gap. Two separate datasets were created and randomly selected. 70% of images are considered to train the CNN network, while the rest 30% remain for testing.

B. Facial Components Detection

Face alignment is a technique for identifying the geometric structure of the human face. In previous work [11] the angle of inclination of the straight line between the outer corners of the two eyes is used to align the face horizontally. The facial component algorithm using Viola and Jones performs well with optimum execution complexity [25]. However, we use the locations of the eyes, nose, and mouth to detect additional components such as the forehead and cheeks. Fig. 2(a) shows whole face and component detection, and Fig. 2(b) the cropped components analyzed in this work. The three other components are supposed to be discriminating for the facial recognition model of our work.
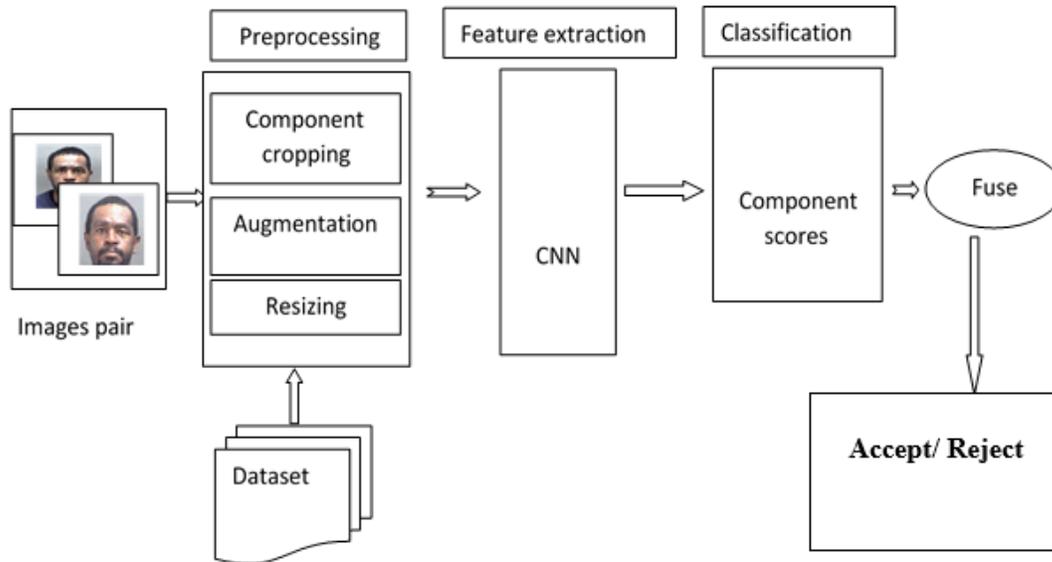
Fig. 1. Proposed Methodology Framework.

Algorithm 1 describes the other three components depending on Viola and Jones. Many images from the databases used in this work have different pose variations. Therefore, this work employs face alignment as in [11] for correcting horizontal facial images before applying even Viola and Jones detection. We get the benefit from the detected facial components and rectangles boundaries. Specifically, the coordinates of the top left, the width, and the height of the pre-detected components, are crucial to detecting the other components.

Algorithm 1: Components Extraction (forehead, left cheek, and right cheek) – from the Face dataset

---

Input**:** Training set *M*, with *m* classes

nj = number of images in a given class

Bbox: boundary box

for *i* = 1 *to m* **do**

  for *j* = 1 *to nj* **do**

  Bbox (forehead) → Bbox (eyes pair)

  Bbox (left check) → Bbox (eyes pair) + Bbox (nose)

  Bbox (left check) → Bbox (eyes pair) + Bbox (nose)

  end for

 end for

C. Image Pre-processing

Initially, images in the datasets were pre-processed to enhance performance and gain improved accuracy results. Facial components, specifically the eyes, nose, mouth, forehead, and cheeks are detected and cropped using the Viola-Jones algorithm for the first three components and the proposed algorithm illustrated above for the last two facial components. Then the images are translated in different directions, rotated and resided to increase the number of images in the training datasets. Finally, the images are capable enough to be fed to the convolutional neural network with RGB color channels.

D. Feature Extraction and Classification

The convolutional neural network is fed with images after completely being preprocessed. In this work, we propose age-invariant facial recognition using its components and utilizing a Convolutional neural network. CNN is the preferable and most powerful employed algorithm in the area of deep learning [26]. The clear advantage of CNN is that relevant features are automatically identified away from human supervision [27].

1) Convolutional Neural Network: CNN architecture is still an open problem. That, the size of training data determines the best number of layers and filters to avoid over fitting [28].

In this work, extract parts or components of the face like eyes, nose, mouth, forehead, and cheeks instead of a full face to recognize a person. We conduct the work using three convolution layers, two max pooling, and one fully connected layer. The multi-layer neural networks built to recognize discriminant features from the origin pixels of images preceded by suitable pre-processing for the intended purpose. In CNN architectures, the main layers for successful models are convolution layers followed by down-sampling in pooling layers and concluding with fully connected layers.

2) Cosine similarity: Cosine similarity is a suitable choice for metric learning due to its special property of providing similarity between the intervals -1 to +1 [29]. Cosine similarity (CS) between two vectors x and y is defined as the following:

$$\text{CS}(x, y) = \frac{X^T Y}{||x|| \, ||y||} \tag{1}$$
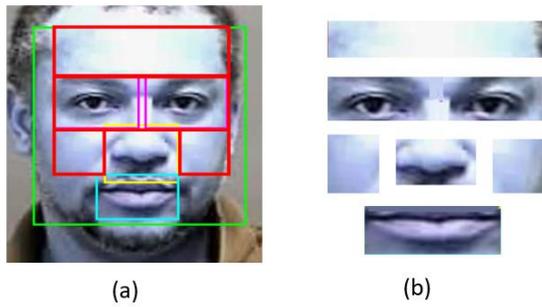
Fig. 2. (a) Facial Components Detection and (b) Facial Components Cropping.

3) Euclidean distance: Euclidean distance [30] is used to evaluate the performance between two feature vectors of pairs. Euclidian distance for two image pairs featuring vectors x and y can be calculated as follows:

$$d(x, y) = \|x - y\| \tag{2}$$

Given two image feature sets, $x = \{x_1, x_2 \dots x_n\}$,

And $y = \{y_1, y_2 \dots y_n\}$, the function defines similarity distance between the two sets adopting the minimum distance as the following.

$$h_{min}(X, Y) = \min d_{(x \epsilon X, y \epsilon Y)}(x,y) \tag{3}$$

Cosine similarity and Euclidean distance measures receive the feature vectors extracted by CNN, then calculate the similarity and distance between two image pair feature vectors as mentioned above. When the result yields less than the threshold, the two faces are regarded as the same identity, otherwise, regarded as different identities.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

MORPH (Album II) [23] is one of the data sets on which we conduct our experiments. The dataset is divided into two main experimental groups: 1) age gap 0-1 which is composed of 2590 images belonging to 494 classes. 2) Age gap 1-5, which includes 5335 images from 942 classes. Then we select randomly seventy per cent of the images in the dataset for the training process and the rest for testing. From both training and testing datasets, six different sub-datasets of eyes, nose, mouth, forehead, left cheek, and right cheek was created. Different CNN networks with the same number of sub-datasets are built. The matched image pairs with the highest similarity are considered as the same identity. To enhance the performance of the networks, we conducted various experiments for each component with different input sizes.

For experimental analysis, accuracy and error rates are calculated. As a result, the most appropriate learning rate and appropriate input image size for each component are adopted using the following equations:

$$Accuracy = \frac{\text{No. of testing samples classified correctly}}{\text{Total no.of testing samples}} \tag{4}$$

$$Error\ rate = \frac{\text{No. of misclassification}}{\text{Total no.of testing samples}} \tag{5}$$

To recognize the image ideally, the input image size plays a crucial role based on the extracted complex features. We notice that image expansion or compression degrades the accuracy. So, many input sizes are chosen to analyze and depending upon performance parameters, an applicable size for all components is recommended.

Learning rate is a hyper-parameter that specifies the adjustment in the weights of the network depending on the loss gradient descent function. It specifies how fast or slow the network will reach the optimal weights. If the learning rate is too high it will skip the optimal solution and if it is too low then, too much iteration will be spent to converge to reach the best values. Thus, utilizing a good learning rate is important.

The experiments involved various input sizes and learning rates and all results were recorded. Generally, six different networks are conducted, one for each component. Each network is trained using only one component and tested using the same component but from other classes not seen during the training phase. All relevant combinations of size and learning rate are carried out and the loss is computed.

For the eyes region component, throughout all various combinations, the size of 64x64 and 0.001 learning rate show the least error of 1.075 and accuracy of 98.925 at the training phase. Thus, it is recommended to use the mentioned size and learning rate for the eye region component for promising results. Fig. 3 illustrates the findings.

Image size learning rate	20x80	32x32	64x64	50x120
0.001	1.6553	1.8372	1.0750	1.6634
0.003	1.8532	1.8104	1.6755	1.7860
0.01	1.9821	1.8206	1.7632	1.6173
0.03	1.9857	2.2741	2.2163	2.1927
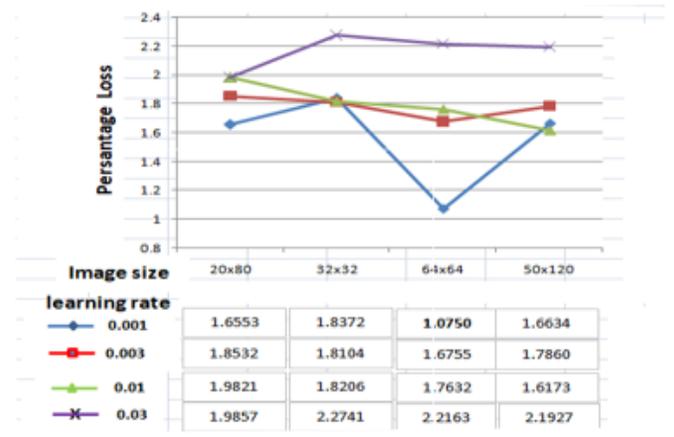
Fig. 3. Percentage Loss Graph for Various Input Image Sizes for Eyes Component.

For the nose component, as depicted in Fig. 4, the size of 32x32 and 0.003 learning rate show the least error of 0.5871% and accuracy of 99.0315. Therefore, the preferable size and learning rate for the component are 32x32 and 0.003 for the trained dataset. Similarly, Fig. 5 shows that the size of 60x80 and 0.001 learning rate for the mouth component produced the least error of 5.7469 and gives the accuracy of 93.9512%.

Two additional components are extracted- forehead and cheeks, using algorithm 1. Following a similar way to find the best image size and learning rate, the goal is to utilize various input sizes and combinations of learning rates. It is noticeable, as shown in Fig. 6 that the size of 40x64 and 0.001 learning rate are the best combination examined that gives the least error rate of 7.1472 and accuracy of 92.8528 for the forehead component.

Image size	32x32	64x64	100x70	
learning rate				
0.001	0.7514	0.7012	0.6934	0.6934
0.003	0.5871	0.9104	0.9107	0.9107
0.01	1.2536	0.9104	1.2542	1.4105
0.03	0.8249	0.8251	1.6139	2.9013
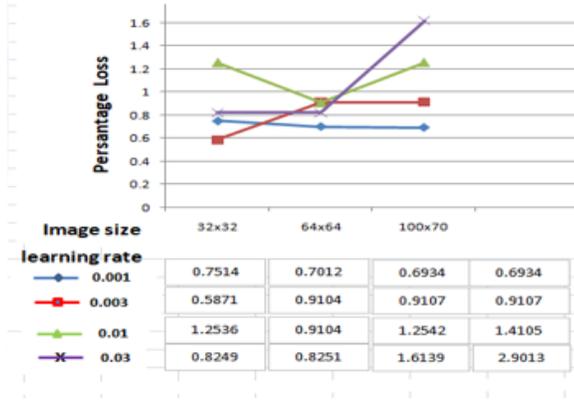
Fig. 4.   Percentage Loss Graph for Various Input Image Sizes for Nose Component.

Image size	32x32	64x64	100x70	128x128
learning rate				
0.001	9.4833	10.2517	11.0274	12.7167
0.003	12.6872	11.7981	13.7292	14.2893
0.01	12.9256	13.3281	14.5186	14.8642
0.03	11.8925	12.2398	14.9735	15.3782
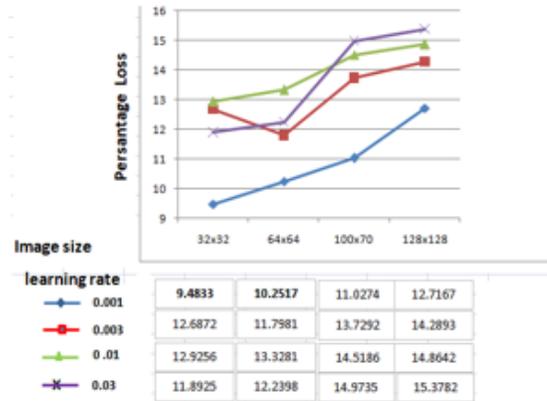
Fig. 7.   Percentage Loss Graph for Various Input Image Sizes for Left Cheek Component.

Image size	32x32	30x40	60x80	100x120
learning rate				
0.001	0.7514	0.7012	0.6934	0.6934
0.003	0.5871	0.9104	0.9107	0.9107
0.01	1.2536	0.9104	1.2542	1.4105
0.03	0.8249	0.8251	1.6139	2.9013
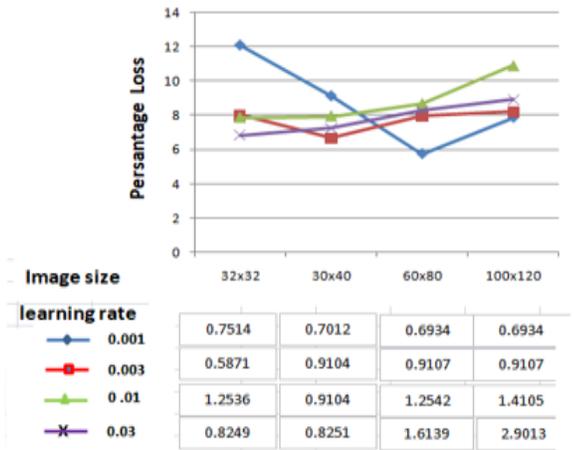
Fig. 5.   Percentage Loss Graph for Various Input Image Sizes for Mouth Component.

Image size	32x32	64x64	100x70	128x128
learning rate				
0.001	9.5488	10.2586	11.7835	13.0012
0.003	12.8965	11.9967	14.0021	14.8975
0.01	13.0067	13.4351	14.7698	14.9876
0.03	12.0132	12.4532	15.1745	15.7533
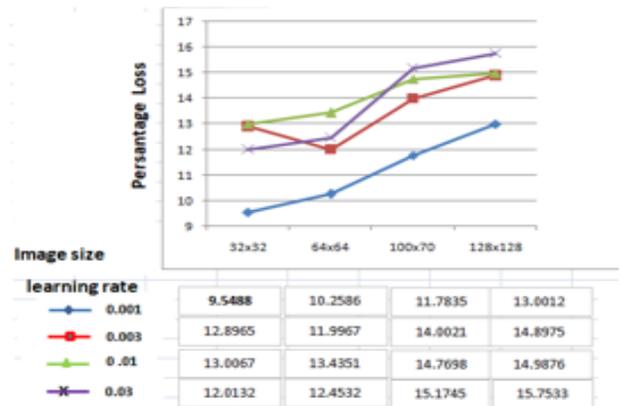
Fig. 8.   Percentage Loss Graph for Various Input Image Sizes for Right Cheek Component.

Image size	32x32	40x64	60x80	100x120
learning rate				
0.001	14.0267	7.1472	8.3056	9.1858
0.003	10.1034	8.6715	9.7962	10.2871
0.01	9.6872	9.1582	10.0217	12.3512
0.03	8.6739	9.4892	10.3623	10.5973
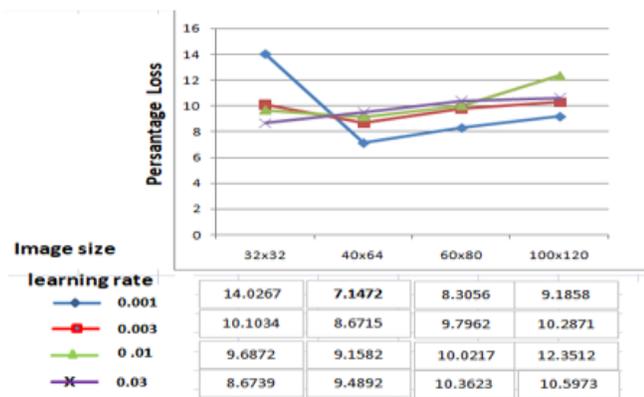
Fig. 6.   Percentage Loss Graph for Various Input Image Sizes for Forehead Component.

The other two components are the left and right cheeks which give close results. Fig. 7 and Fig. 8 depicted the results and it's clear to deduce that the best size for both is 32x32 and the combined learning rate is 0.001. The best training accuracy is 90.5167% and 90.4512 for left and right cheek respectively.

A. Training Parameters

This work selects three convolutional layers, two max pooling, one fully connected layer, and a classification layer. All convolution layers are immediately followed by a ReLU (Rectified Linear Unit) activation function. The max-pooling size is 2×2 and stride 2 for down sampling. Images in datasets are RGB color and each component has its applicable resized input image and the best learning rate.

Images in each CNN of the mentioned components are divided into mini-batches. The mini-batch is set to size 100. Each batch is trained and convoluted with a fixed number of generated filters combined with a determined bias and a computed constant value.

In the phase forward pass for each hidden layer neuron, the activation is calculated as:

$$net_h = \sum_{i=1}^{n} w_i * x_i + b \qquad (6)$$

Where net_h represents the total net input for the determinant hidden layer, n is the number of filter weights, w_i is i_{th} filter weight, x is the input neuron, b: bias. Perform the same process for successive layers. Repeat this process for the output layer neurons, using the output from the prior hidden layer neurons as input for the subsequent hidden layer. To get

the total error (E_{total}), squared error function is applied for each output neuron as follows.

$$E \text{ total} = \frac{1}{2}\sum_{i=1}^{n} (T_i - Y_i)^2 \tag{7}$$

Where n is the number of output neurons, T_i is the target output for neuron *i* and, Y_i is the output of the neuron *i* calculated by a forward convolution pass.

The Backwards Pass: the purpose is to update all weights in the network to provide the calculated output to be closer the target output, by minimizing the error for each output neuron and update the whole network**.** Now the crucial role is to know how much the change in each weight will affect the overall error. In other words, the role is to calculate the partial derivative of the total error with respect to each weight. The step describes the partial derivative of E_{total} with respect to w_i . That means to obtain the gradient with respect to w_i. Properly, the following chain rule is applied:

$$\frac{\partial E \text{ total}}{\partial w_i} = \frac{\partial E \text{ total}}{\partial out_{01}} * \frac{\partial out_{01}}{\partial net_{01}} * \frac{\partial net_{01}}{\partial w_i} \tag{8}$$

The final step is minimizing error by subtracting the value multiplied by some learning rate from the current weight. The following is actually what happens to update each weight:

$$w_i^+ = w_{i-} \eta * \frac{\partial e}{\partial w_i} \tag{9}$$

When all weights are updated after a repeated process, we roll into a forward pass using the updated weights. Final activations fed to the fully connected layer transform the learned neurons into a new embedding vector. The distance between the impostor pair is enlarged while the distance between the genuine pair is minimized.

B. Data Augmentation

Data augmentation balances the samples size of training set when some classes have abundant samples while the rest lack appropriate number of samples [26] . Data augmentation is a technique that prevents the network from over fitting through discriminant features of the training images [31]. The input facial component images are translated horizontally and vertically in the range [-30, 30]. Then, we rotate the images and resize them to the ideal input size for each component.

C. Integrating Facial Components

In this work, to verify components of the same type, metric learning to discriminate the similarity is used. Many research works applied Mahalanobis distance learning [32] and Euclidean distance [33] but the drawback of the Mahalanobis distance is the equal adding up of the variance normalized squared distances of the features. However, the key issue of Euclidian norm that it gives the same importance to any direction. So, metric learning that capable with angular distribution to calculate similarity is needed. The best choice is cosine distance to deal with metric learning that provide reliable classification [34].

D. Combining Components Scores

All scores of the facial components are fused at the matching score level. The obtained output matching scores of each of these components are weighted and combined. Fusion at the matching score level is usually preferred, as it is relatively easy to access and combine the scores presented by the different modalities [36]. The adopted matching score in this work is the same as the algorithm specified in [35] but, with a cosine similarity output of each facial component.

E. Score Normalization

Each component is treated independently, and its own score is computed. Then, each component's scores are normalized given a set of n raw component matching scores $\{S_k\}$, k = 1, 2... n. For the sake of this work, n should be 1 up to 6 the number of analysed facial components. The corresponding normalized scores S_k' is given by Min-max normalization as the following:

$$s_k' = \frac{s_k - \min(\{s_k\})}{\max(\{s_k\}) - \min(\{s_k\})} \tag{10}$$

Where min and max are the minimum and maximum seen at learning phase, respectively, of the given set $\{S_k\}$ of component matching scores.

F. Score Weighting

Each component score is given a weighing to achieve an overall fusion. Let s_e' s_n' , s_m', s_f', and s_c' be the normalized scores for a specific class - *i* and t_1, t_2, t_3, t_4, and t_5 are the thresholds of the eyes, nose, mouth, forehead, and cheeks respectively. Then the initial weights of components scores are computed as follows:

$$w_e'^i = \frac{s_e'^i}{t_1 + s_e'^i} \tag{11}$$

$$w_n'^i = \frac{s_n'^i}{t_2 + s_n'^i} \tag{12}$$

$$w_m'^i = \frac{s_m'^i}{t_3 + s_m'^i} \tag{13}$$

$$w_f'^i = \frac{s_f'^i}{t_4 + s_f'^i} \tag{14}$$

$$w_c'^i = \frac{s_c'^i}{t_5 + s_c'^i} \tag{15}$$

The preliminary weights $w_e'^i$, $w_n'^i$, $w_m'^i$, $w_f'^i$, and $w_c'^i$ related to eyes, nose, mouth, forehead, and cheeks respectively. Next, the fusion weights for the ith class are computed respectively, as follows:

$$w_e^i = \frac{w_e'^i}{w_e'^i + w_n'^i + w_m'^i + w_f'^i + w_c'^i} \tag{16}$$

$$w_n^i = \frac{w_n'^i}{w_e'^i + w_n'^i + w_m'^i + w_f'^i + w_c'^i} \tag{17}$$

$$w_m^i = \frac{w_m'^i}{w_e'^i + w_n'^i + w_m'^i + w_f'^i + w_c'^i} \tag{18}$$

$$w_f^i = \frac{w_f'^i}{w_e'^i + w_n'^i + w_m'^i + w_f'^i + w_c'^i} \tag{19}$$

$$w_c^i = \frac{w_c'^i}{w_e'^i + w_n'^i + w_m'^i + w_f'^i + w_c'^i} \tag{20}$$

Then the fusion score is computed as follows:

$$S_{fuse} = w_e^i s_e' + w_n^i s_n' + w_m^i s_m' + w_f^i s_f' + w_c^i s_c' \qquad (21)$$

Cosine similarity between the two facial image pairs is the similarity score between the two faces. Lastly, this score is compared with a threshold to decide whether two faces belong to the same class or not. The threshold *is* selected from the training set so that False Acceptance Rate is equivalent to False Rejection Rate. The decision function defined in Equation (22) verifies the class.

$$Decision\ (S_{fuse}) = \begin{cases} Accept, if > threshold \\ Reiect. otherwise \end{cases} \qquad (22)$$

The ROC curve of the experiments on MORPH dataset is depicted in Fig. 9. The curve shows that the cheek component has the lowest accuracy while the nose has the highest contribution that reflects its stability during aging.
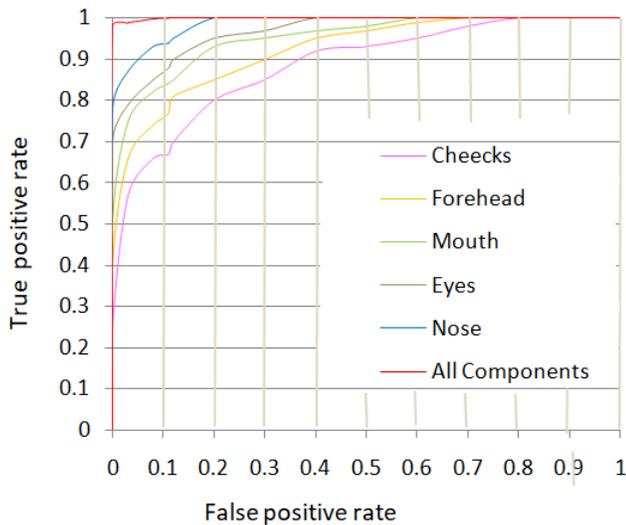
Fig. 9.   MORPH Results on 1-5 Year Age Gap Data Set.

The Tables I to IV display per component accuracy performance and detail how the proposed method improves the performance using two-facial component score fusion. Moreover, it is noticeable that the nose component has the highest accuracy, and the cheeks are the lowest. Performance is enhanced when scores of the two components are fused. Eyes and nose combination increases the accuracy for both data subsets. However, the accuracy decreased dramatically when forehead and cheek scores fused. But, when scores of three components or more are fused the accuracy is highly improved. We compared our work with the previous most related works. The comparison results are shown in Table V.

TABLE II.   PER COMPONENT TAR (%) AT 1% FAR, MORPH 1-5 YEAR AGE GAP SET

Components	Accuracy %	
Eyes	86.92	
Nose	95.10	
Mouth	84.28	
Forehead	77.62	
Cheeks	70.33	
fusion	**Enhancement%**	**Accuracy%**
Forehead + Cheeks	0.10	77.72
Mouth + forehead	1.20	85.48
Eyes+ forehead	2.12	89.04
Eyes+ mouth	2.43	89.35
Nose + cheeks	1.05	96.15
Nose+ forehead	2.24	97.34
Nose + mouth	2.80	97.9
Nose +eyes	3.20	98.3
All components	4.80	99.9

TABLE I.   PER COMPONENT TAR (%) AT 1% FAR, MORPH 0-1 YEAR AGE GAP SET

Components	Accuracy %	
Eyes	90.31	
Nose	95.12	
Mouth	87.07	
Forehead	82.52	
Cheeks	77.45	
fusion	**Enhancement%**	**Accuracy%**
Forehead + Cheeks	0.58	83.1
Mouth + forehead	1.53	88.6
Eyes+ forehead	2.55	92.86
Eyes+ mouth	2.95	93.26
Nose + cheeks	1.46	96.58
Nose+ forehead	2.82	97.94
Nose + mouth	3.25	98.37
Nose +eyes	3.56	98.68
All components	4.83	99.95

TABLE III.   PER COMPONENT TAR (%) AT 1% FAR FG-NET 0-1 YEAR AGE GAP SET

Components	Accuracy %	
Eyes	90.81	
Nose	95.8	
Mouth	88.12	
Forehead	86.72	
Cheeks	80.04	
fusion	**Enhancement%**	**Accuracy%**
Forehead + Cheeks	0.51	87.23
Mouth + forehead	1.5	89.62
Eyes+ forehead	2.87	93.68
Eyes+ mouth	3.23	94.04
Nose + cheeks	1.87	97.67
Nose+ forehead	2.91	98.71
Nose + mouth	3.25	99.05
Nose +eyes	3.67	99.47
All components	4.2	100

TABLE IV.   PER COMPONENT TAR (%) AT 1% FAR, FG-NET 1-5 YEAR AGE GAP SET

Components	Accuracy %	
Eyes	89.24	
Nose	94.92	
Mouth	86.05	
Forehead	82.29	
Cheeks	73.68	
fusion	**Enhancement%**	**Accuracy%**
Forehead + Cheeks	0.44	82.73
Mouth + forehead	1.1	87.15
Eyes+ forehead	2.23	91.47
Eyes+ mouth	2.75	91.99
Nose + cheeks	1.73	96.65
Nose+ forehead	2.52	97.44
Nose + mouth	3.1	98.02
Nose +eyes	3.54	98.46
All components	5.08	100

TABLE V.   COMPARISON OF OUR RESULT WITH COMPONENT BASED ALGORITHMS

Description	Component used	Component representation	Face database	Accuracy	
Component-based LDA method with component Bunches [18]	L. eye, R. eye, nose, R. mouth, L. mouth	Pixel representation	FERET	93.57%	
Component-based with procruetes analysis [19]	Eyes, nose, and mouth	Random measurement matrix	Extended Yalu face database	96%	
Resized components to pre-trained model [20]	Periocular region, nose, mouth	Pre-trained AlexNet CNN	FG-NET database	98.31%	
Components with landmarks [2]	Eyes ,Nose, Mouth and Eyebrows	Multi-scale local binary pattern and scale-invariant feature transform	PCSO and MORPH database	97.60%	
The Proposed work	Eyes, nose, mouth, cheeks and forehead	Convolutional Neural Network	MORPH database	99.90 %-99.95 %	
			FG-NET data-base	100 %	

V. CONCLUSIONS AND FUTURE WORK

This paper addressed the challenge of facial recognition on aging subjects using Convolutional Neural Network. Facial components such as eyes, nose, mouse, forehead and cheeks are resized independently, each as ideal input image size. Each subset of the same type classified indecently using cosine similarity. Weighted fusion is utilized to sum up all facial components scores for a final decision. We tested our work using FG-NET and MORPH publicly available datasets. The proposed work achieved a state-of-the art accuracy of 100% on FG-NET dataset and the results obtained on MORPH dataset outperform accuracy results obtained in the literature. Our future work will focus on facial components that give high score contributions and updating the weights of the candidate components to improve overall performance.

REFERENCES

[1]   Harakannanavar, Sunil S., C. R. Prashanth, Vidyashree Kanabur, Veena I. Puranikmath, and K. B. Raja, "Technical Challenges, Performance Metrics and Advancements in Face Recognition System," International Journal of Computer Sciences and Engineering 7, no., pp. 836-847, 3 2019.

[2]   Otto, Charles, Hu Han, and Anil Jain, "How does aging affect facial components?.," in 7584, Berlin, Heidelberg, 2012.

[3]   Liao, Shengcai, Anil K. Jain, and Stan Z. Li, "Partial face recognition: Alignment-free approach," IEEE Transactions on pattern analysis and machine intelligence, vol. 35, no. 5, pp. 1193-1205, 6 Sep 2012.

[4]   Kortli, Yassin, Maher Jridi, Ayman Al Falou, and Mohamed Atri, "A novel face detection approach using local binary pattern histogram and support vector machine," in International Conference on Advanced Systems and Electric Technologies (IC_ASET), 2018.

[5]   Ojala, Timo, Matti Pietikäinen, and David Harwood, "A comparative study of texture measures with classification based on featured distributions," Pattern recognition, vol. 29, no. 1, pp. 51-59., 1996.

[6]   Napoléon, Thibault, and Ayman Alfalou, "Pose invariant face recognition: 3D model from single photo," Optics and Lasers in Engineering, vol. 89, no. 1, pp. 150-161, 1 Feb 2017.

[7]   Khoi, Phan, Lam Huu Thien, and Hoai Viet Vo, "Face retrieval based on local binary pattern and its variants: a comprehensive study," International Journal of Advanced Computer Science and Applications, vol. 7, no. 6, 2016.

[8]   Karaaba, Mahir, Olarik Surinta, Lambert Schomaker, and Marco A. Wiering, "Robust face recognition by computing distances from multiple histograms of oriented gradients," IEEE Symposium Series on Computational Intelligence, pp. 203-209), 7 Dec 2015.

[9]   Prince, Simon, Peng Li, Yun Fu, Umar Mohammed, and James Elder., "Probabilistic models for inference about identity," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 1, pp. 144-157., May 2011.

[10] "Face recognition using principal component analysis and log-gabor filters," arXiv preprint cs/0605025., 7 May 2006.

[11] Hassan, Adam, and Serestina Viriri, "Invariant feature extraction for component-based facial recognition," International Journal of Advanced Computer Science and Applications, vol. 11, no. 11, 2020.

[12] Zhai, Huanhuan, Chunping Liu, Husheng Dong, Yi Ji, Yun Guo, and Shengrong Gong., "Face verification across aging based on deep convolutional networks and local binary patterns," International conference on intelligent science and big data engineering, pp. 341-350, 14 Jun 2015.

[13] Yousaf, Adeel, Muhammad Junaid Khan, Muhammad Jaleed Khan, Adil M. Siddiqui, and Khurram Khurshid., "A robust and efficient convolutional deep learning framework for age-invariant face recognition," Expert Systems, vol. 37, no. 3, p. e12503, 2020.

[14] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 7, pp. 971-987., 7 Aug 2002.

[15] Klare, Brendan, Zhifeng Li, and Anil K. Jain, "Matching forensic sketches to mug shot photos," IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 3, pp. 639-646, 14 Oct 2010.

[16] Klare, Brendan, and Anil K. Jain, "On a taxonomy of facial features," in IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2010.

[17] Heisele, Bernd, Purdy Ho, Jane Wu, and Tomaso Poggio., "Face recognition: component-based versus global approaches," Computer vision and image understanding,, vol. 91, no. 1-2, pp. 6-21, Jul 2003.

[18] Wang, Lijia, Hua Zhang, and Zhenjie Wang., "Component based representation for face recognition.," 2015.

[19] Sadr, Javid, Izzat Jarudi, and Pawan Sinha, "The role of eyebrows in face recognition," vol. 32, no. 3, pp. 285-293, 2003.

[20] Gumede, A., Serestina Viriri, and M. Gwetu., "[Hybrid Component-based Face Recognition," in Conference on Information Communication Technology and Society (ICTAS),, 2017.

[21] Seal, Ayan, Debotosh Bhattacharjee, Mita Nasipuri, Consuelo Gonzalo-Martin, and Ernestina Menasalvas, "À-trous wavelet transform-based hybrid image fusion for face recognition using region classifiers," Expert Systems, vol. 35, no. 6, p. e12307, 2018.

[22] Boussaad, Leila, and Aldjia Boucetta, "An effective component-based age-invariant face recognition using Discriminant Correlation Analysis," Journal of King Saud University-Computer and Information Sciences, 25 Aug 2020.

[23] Sahoo, Tapan Kumar, and Haider Banka, "Multi-feature-based facial age estimation using an incomplete facial aging database," Arabian Journal for Science and Engineering, vol. 43, no. 12, pp. 8057-8078, 2018.

[24] Panis, Gabriel, Andreas Lanitis, Nicholas Tsapatsoulis, and Timothy F. Cootes, ". Overview of research on facial ageing using the FG-NET ageing database," Iet Biometrics, vol. 5, no. 2, pp. 37-46, 2016.

[25] Viola, Paul, and Michael Jones, "Rapid object detection using a boosted cascade of simple features," in In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR, 2001.

[26] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017.

[27] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. and Chen, "Recent advances in convolutional neural networks," Pattern recognition, vol. 77, no. 1, pp. 354-377, 2018.

[28] Hu, Guosheng, Yongxin Yang, Dong Yi, Josef Kittler, William Christmas, Stan Z. Li, and Timothy Hospedales, "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition," in n Proceedings of the IEEE international conference on computer vision workshops, 2015.

[29] Nguyen, Hieu V., and Li Bai., "Cosine similarity metric learning for face verification," in In Asian conference on computer vision, Berlin, Heidelberg., 2010.

[30] El Khiyari, Hachim, and Harry Wechsler, "Age invariant face recognition using convolutional neural networks and set distances," Journal of Information Security, vol. 8, no. 3, p. 174, 2017.

[31] Van Dyk, David A., and Xiao-Li Meng, "The art of data augmentation," Journal of Computational and Graphical Statistics, vol. 10, no. 1, pp. 1-50, 2001.

[32] Hu, Junlin, Jiwen Lu, and Yap-Peng Tan., "Discriminative Deep Metric Learning for Face Verification in The Wild," in In Proceedings of the IEEE conference on computer vision and pattern recognition, 2014.

[33] Hu, Yiqun, Ajmal S. Mian, and Robyn Owens, "Sparse Approximated Nearest Points for Image Set Classification," In CVPR, pp. 121-128, 20 Jun 2011.

[34] Huang, Xixian, Xiongjun Zeng, Qingxiang Wu, Yu Lu, Xi Huang, and Hua Zheng., ".[Face Verification Based on Deep Learning for Person Trackingin Hazardous Goods Factories," Processes, vol. 10, no. 2, p. 380, 2022.

[35] Viriri, Serestina, and Jules R. Tapamo, "Integrating iris and signature traits for personal authentication using user-specific weighting. Sensors," Sensors, vol. 12, no. 4, pp. 4324-4338, 2012.

[36] Jain, Anil K., and Arun Ross, "Multi biometric systems," Communications of the ACM, vol. 47, no. 1, pp. 34-40., 1 Jan 2004.

Research on the Design of Online Teaching Platform of College Dance Course based on IGA Algorithm

Yunyun Xu

Hunan First Normal University

Cheng Nan Academy

Changsha, 410002, China

Abstract—As a comprehensive art form, dance plays an integral role in developing the overall quality of students. However, with the increasing progress of IT technology, the traditional classroom-based teaching mode of dance course can no longer adapt to the current educational environment. This research designs an e-teaching platform system for college dance lessons based on the IGA algorithm (Improved Genetic Algorithm, IGA). First, it establishes a mathematical model of artificial intelligence questions and then proposes the functional design of an online teaching platform for college dance course based on the IGA algorithm. The feasibility of the proposed online teaching platform system of college dance course based on the IGA algorithm is validated via a number of testes. When the iteration count is 500, the success rate of the IGA algorithm reaches 99%; when the iteration count is 100, the average fitness is 0.929; when the iteration count is 100 times, the moderate calculation value is 0.936. While for the traditional Genetic algorithm (GA), the results are 88.6%, 0.73, and 0.752, respectively. By comparing with the traditional teaching mode based on GA algorithm, the proposed method based on IGA algorithm is obviously superior in many aspects.

Keywords—Online teaching; improved genetic algorithm; college dance course; intelligent test set; simulated annealing

I. Introduction

With the progress of distance internet teaching, scholars in China and foreign countries have been thinking about the reform of traditional classroom-based teaching mode of art courses [1]. Therefore, the establishment of an online teaching platform system of dance course is particularly important for the future development of college dance course teaching [2]. In general, online teaching in colleges and universities in China is still in the early stage. Therefore, it is of great practical significance to propose an online teaching platform for art courses [3]. The genetic algorithm is a computational model that finds the optimal solution or approximate solution by simulating the evolution process. It has the advantages of being able to solve complex, multi-constrained problems and seek global parallel solutions [4]. Genetic algorithm has attracted the focus of many scholars because of its ability to cope with the questions in the platform, initialize the population, and perform genetic evolution to accurately group the questions [5]. However, genetic algorithms can easily fall into a state of premature maturity. In this study, the genetic algorithm is improved to obtain the IGA algorithm and the improved algorithm is used to design a mathematical model of the test problem, so as to enrich and improve the online dance teaching

system in universities. Meanwhile, the current research on online dance teaching in universities in China mainly focuses on the level of online dance teaching promotion, while the innovation of this study lies in detailing the strategies and methods to carry out, which are highly operable and close to the teachers of dance in universities, which will provide new ideas for the construction of online platforms for teaching art courses.

II. Related Work

Genetic Algorithm (GA) is a novel algorithm developed rapidly in recent years. Aiming at the low accuracy of traditional methods in evaluating the quality of university mathematics classes, Yang Yun proposed a genetic algorithm-based quality evaluation method for the reform of university mathematics teaching model. The results revealed that the goodness of fit of the modified genetic algorithm exceeded 90%, and the error was significantly reduced to 0.01-0.04 [6]. In order to improve the search speed and achieve the effect of short-term optimization, Y Zhang et al. summarized various intelligent algorithms used in the current research of related research groups, and proposed a computing method of group parallel evolution. The results showed that the algorithm can prevent the results from falling into the prematurity and has strong convergence in the group experiment [7]. Dong-Jie Li's team proposed a calculation method CGA that combines chaos algorithm and genetic algorithm, in order to overcome the defects that BP neural network is likely to drop into local minimum and converge slowly in the process of gesture action recognition. Simulation results showed that the proposed algorithm can improve real-timeliness and accuracy of gesture recognition. [8]. Hua Wang and Lingwei Wang proposed an improved classical genetic algorithm to address the instability of the traditional computer network system caused by many factors such as the huge system. The simulation experiment results indicated that the algorithm can improve the fluency of the algorithm, thereby improving the availability of the network and reducing the occurrence of error in the network [9]. Aiming at current problem of college English course arrangement in China, J Xu proposed an improved genetic algorithm. Experimental results indicated that the proposed algorithm improved the convergence speed and individual diversity, and prevented the results from easily falling into local optimum. [10].

Online teaching has become more and more important, and there are many ways to improve the existing problems of

online teaching. Wang designed a control system for a web-based distance learning platform. According to the experimental results, the average computing time of resource scheduling consumed by the system is 32.34 ms, which is 12.6% and 23.7% faster than the other two traditional methods, respectively [11]. For the purpose of improving the prediction of professional education evaluation results, Gao et al. proposed a simulated annealing algorithm on the basis of simulated annealing fish swarm algorithm. According to the results, the forecasting system of professional education lesson assessment performance constructed based on simulated annealing algorithm can effectively predict the examination results [12]. To improve the detection of word sequences within the English classroom, Deng Bowen designed an analytical model on the basis of group encoding and modified genetic algorithm. It was demonstrated that the model well achieved subsequent sub-learning target tasks, indicating strong reliability [13]. In order to play the role of virtual reality technology in college dance course teaching, Shi discussed the application of VR technology in dance course education [14]. The MC Tsai's team proposed a computer curriculum design method based on the actual teaching methods of AL and ML, and the results showed that this metheology usefully improved learning outcomes and learning participation of online computer courses [15].

The above research fully shows that genetic algorithm has been widely used in many fields, and there are also many researches on online teaching platforms. However, there are still a variety of problems that need to be solved when combining genetic algorithm with online teaching platform system for university dance course. Therefore, the research will improve GA algorithm to obtain the IGA algorithm, and embed the algorithm into the universities e-dancing teaching platform system.

III. DESIGN METHOD OF ONLINE TEACHING PLATFORM OF COLLEGE DANCE COURSE BASED ON IGA ALGORITHM

A. Research on Artificial Intelligence Test Questions and Establishment of Mathematical Models of Test Questions

The problem research of artificial intelligence group volume refers to the problem of seeking the best answer or near-optimal answer under the restriction of multiple constraints [16]. Combined with previous studies on the reform of online courses in universities, this research assumes that the target test paper must meet the following attributes: Question type (QT), Number of questions (NQ), Difficulty of exam paper (Difficulty of exam paper), DP), knowledge point (KP) and discrimination (Distinction of test papers, DI). Therefore, the mathematical model of the quiz problem involved in this paper can be defined as M.

$$M = \{QT, NQ, DP, KP, DI, EP\} \tag{1}$$

From the above equation, it can be seen that the question of the examination paper is a six-tuple question, and the mathematical model matrix of the test paper is established according to this question as follows.

$$m = \begin{bmatrix} a_{1,1} & \cdots & a_{1,5} \\ \vdots & \ddots & \vdots \\ a_{t,1} & \cdots & a_{t,5} \end{bmatrix} \tag{2}$$

Where t represents the count of questions in the examination paper, $a_{t,1}$, to $a_{t,5}$ represent the t attributes of the first question, respectively.

On the basis of summarizing actual teaching experiences of college dance courses and combined with the mathematical model of the test problem, the constraints required for the test set are set as follows.

Constraints QT, which stipulates that the question types of the test questions must be composed of fixed test question types.

Constraint NQ, which specifies an overall number of questions on the test paper.

Constraints DP, set according to the subject requirements of the candidates, specifies the difficulty level of the examination paper. The difficulty level can be calculated according to the ratio of the difficulty of all test questions to the total score. The formula is as follows.

$$DP = \Sigma_{i=1}^{r} P_i \times C_i / \Sigma_{i=1}^{r} P_i \tag{3}$$

Where r is the total number of item IDs in the test paper, C_i is the difficulty coefficient of the i-th item ID, and P_i represents the actual score of the i-th item ID .

Constraints KP stipulates the knowledge points and scope of the examination paper for the corresponding course, in order to help the teachers to have an accurate understanding of the students' mastery of the knowledge points. In this study, the knowledge point coverage is calculated based on the percentage of knowledge point scores. Assuming that the total number of knowledge points examined in the test paper is d, the formula for calculating the coverage rate of knowledge points is as follows KP_i.

$$KP_i = KS_i / \Sigma_{i=1}^{r} KS_i \tag{4}$$

$$\sum_{i=1}^{d} KP_i = 1 \tag{5}$$

Where KP_i represents the percentage of the score of the i th knowledge point in the mark of the examination paper, and KS_i represents the mark of the i th knowledge point.

Constraints DI, based on which the teacher can master the specific learning status of students. The formula of the discrimination degree DI is calculated as follows.

$$DI = \Sigma_{i=1}^{r} Q_i \times P_i / \Sigma_{i=1}^{r} Q_i \tag{6}$$

Where Q_i represents the discrimination coefficient of the first i question, P_i represents the actual score of the first question, and DI is the overall discrimination.

The choice of the fitness function (FF) directly affects the convergence speed of the IGA algorithm and even determines whether it can find the optimal solution. In this study, the IGA algorithm has satisfied the constraints of the question type QT and the number of test questions NQ at the beginning of the initial population design. Therefore, the remaining constraints affecting FF include the difficulty of the DP test paper, the knowledge points of KP and the discrimination of the DI test paper.

The weight of the above four influences is $W_i\left(i \leq 4\right)$. The difference between the anticipated and realized value of the constraints corresponding to different test papers is expressed as $dv_i\left(i \leq 4\right)$: In conclusion, the f equation for the objective function of present study is as follows.

$$f = \sum_{i=1}^{4} dv_i \times W_i \tag{7}$$

$$\sum_{i=1}^{4} dv_i = 1 \tag{8}$$

It can be concluded that the tinier the value of f, a closer the outcome of individual evolution is to the expectation. Hence, in order to obtain the individual best solution in a group volume, the required objective function f value must be as small as possible. Meanwhile, assuming that the fitness scores reflect the evolution of the individual positively and ensure that the time complexity of calculating the fitness function is small, the calculation formula f of the relationship between the objective function and FF is proposed as follows.

$$g = 1/1 + f \tag{9}$$

Where g represents the FF of the grouping problem.

B. Design Scheme based on IGA Algorithm

The core component of the online teaching platform of college dance course is artificial intelligence. Through previous researches, it is found that the traditional genetic algorithm (GA) has excellent performance in solving multi-constraint combinatorial optimization problems [17-18]. Therefore, the GA algorithm is adopted to address the artificial intelligence quiz problem in this paper. When the GA algorithm realizes the process of intelligent paper grouping, it is often prone to fall into precociousness and has weak local search ability. Therefore, this study improves the GA algorithm to obtain IGA algorithm, and carries out the mat lab simulation experiment to demonstrate the feasibility of the IGA algorithm. The results indicate that the IGA algorithm proposed in this study performs outperforms in terms of test composition.

Chromosome coding is designed by means of piecewise real number coding. Taking the pre-class thinking test paper of the course "Latin Dance in College" as an example, there are five MCQs, three fill-in-the-blank questions, and two short-answer questions in this test paper. The chromosome code of the test paper is shown in Table I.

TABLE I. THE CHROMOSOME CODES OF EXAM PAPERS FOR THE COURSE OF "COLLEGE LATIN DANCE"

Type of question	Topic ID no.
MCQs	3; 16; 72; 26; 42
Fill -in-the-blank questions	86; 28; 34
Short answer questions	2; 48

In the above table, the question type is the chromosome of the test paper, and the question ID number corresponding to each question type is the chromosome code of the test paper.

After the chromosome is encoded, the population will be initialized. To avoid individuals with high level of similarity between topics and the poor quality of the papers, this study will delete the initialized population and the genetic evolution of the offspring, until only one individual remains. Before initialization, it is necessary to calculate individual similarity S_{PQ}, and preset the threshold T of individual similarity. According to previous research, the T value of this study is set to 35%. If the calculation value S_{PQ} is S_{PQ} greater than the set T value, it needs to delete the individual; if the S_{PQ} value is less than the set T value, keep the individual. In this way, it can be ensured that the value between the next generation populations S_{PQ} is less than the set T value, and this method can prevent the algorithm from falling into local convergence in the early stage. In the calculation of the value S_{PQ}, the method of information entropy calculation is adopted, which is expressed as below.

$$H\left(k\right) = -\frac{\sum_{i=1}^{l}\sum_{j=1}^{n} p_{ij} \log p_{ij}}{l} \tag{10}$$

Where k represents the number of individuals; l represents the number of individual genes; n represents the number of alleles available for selection $\{g_1, g_2, \cdots, g_n\}$; p_{ij} represents the j-th allele in each individual P. Therefore, the similarity between S_{PQ} individuals is Q, which can be expressed as follows.

$$S_{PQ} = 1/1 + H\left(2\right) \tag{11}$$

Where $H(2)$ represents the information entropy between P & Q individuals.

After the design of the initialized population is completed, the selection operator needs to be designed, and the direction of the population genetics mainly depends on the selection operator algorithm. At present, the most commonly used method of selection operator algorithm is the roulette selection method. The approach can determine if the individual is transmittable for the following generation based on the individual fitness value applied in the odds ratio. The benefit of roulette selection method is that the individual with better fitness has a high percentage of being selected for evolution, and those with poor individual fitness have a small percentage of being selected for evolution. The process is illustrated in the chart below.
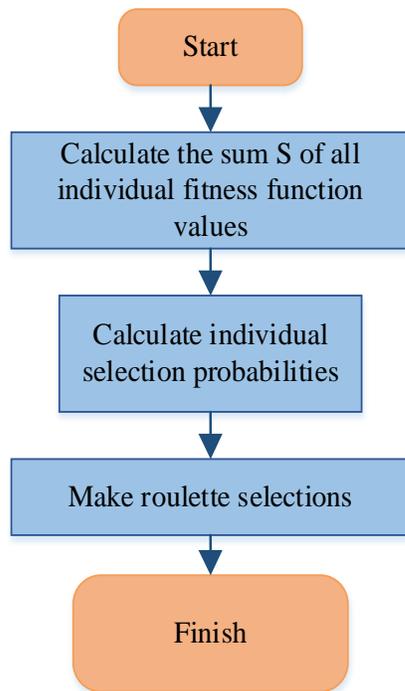
Fig. 1. Roulette Selection Flow Chart.

The calculation formula of the sum S of the individual fitness function values in Fig. 1 is as follows.

$$F = \sum_{i=1}^{n} f_i \qquad (12)$$

Here, f_i stands for the fitness function value of the i th individual.

The formula for calculating the probability of individual selection is as follows.

$$P_{si} = f_i / F \qquad (13)$$

Although the roulette selection method can make excellent individuals obtain high genetic probability, but there is also the possibility that excellent individuals are not inherited. At present, the selection operator algorithms that are widely used include the optimal individual preservation method and the sorting selection method. This research will combine the advantages of the above methods, and choose the offspring based on the IGA algorithm and combining the roulette selection method and the ranked selection method. The specific process is to use the ranked selection method to set the top 15% of individuals for inheritance, and then use the roulette wheel to select the remaining individuals.

Traditional GA algorithm is prone to fall into precociousness and has weak local search ability [19-20]. To address these drawbacks of the traditional GA algorithm, acceptance probability will be added to the IGA algorithm to perform simulated annealing of the crossover variant operation individuals in the study. The searching capaability of the GA algorithm is improved. The detailed formula is as follows.

$$p = 1, f(old) < f(new) \qquad (14)$$

$$p = \exp\left[-f(old) - f(new)/T_t\right], f(old) \geq f(new) \qquad (15)$$

Where $f(old)$ is the old individual, $f(new)$ is the new individual, and T_t is the temperature at time t. If the new test paper from the group shows that the individual is better, then this individual will be selected for iteration; if the new test paper produced by the group shows the individual is poor, a certain probability P will be selected for subsequent iterations.

The temperature T decreased with time t. The initial temperature is set as T_0, and the specific calculation formula of the temperature drop is as follows.

$$T_{t+1} = \beta T_t \qquad (16)$$

Where β is a constant, which will be set to 0.95 in this study.

Since the IGA algorithm is an evolutionary algorithm, this study sets two conditions for the algorithm. If the result satisfies either one of the conditions, it can be judged to be terminated, and then the result is output, that is, the algorithm is completed. The detailed process is illustrated in Fig. 2.

C. Functional Design of Online Teaching Platform of College Dance Course Based on IGA Algorithm

To solve the problems existing in the dance course online teaching platform in colleges and universities, intensive research was conducted, and the results showed that the students' uneven foundation in the subject directly affects their enthusiasm for learning the course. Since this research marks and categorizes the online teaching materials, the students are able to choose course materials at different difficulty levels, and implement hierarchical and classified learning. Based on the requirements of the online teaching platform of college dance course, a functional framework of college dance course online teaching platform based on IGA algorithm is designed, as shown in Fig. 3.
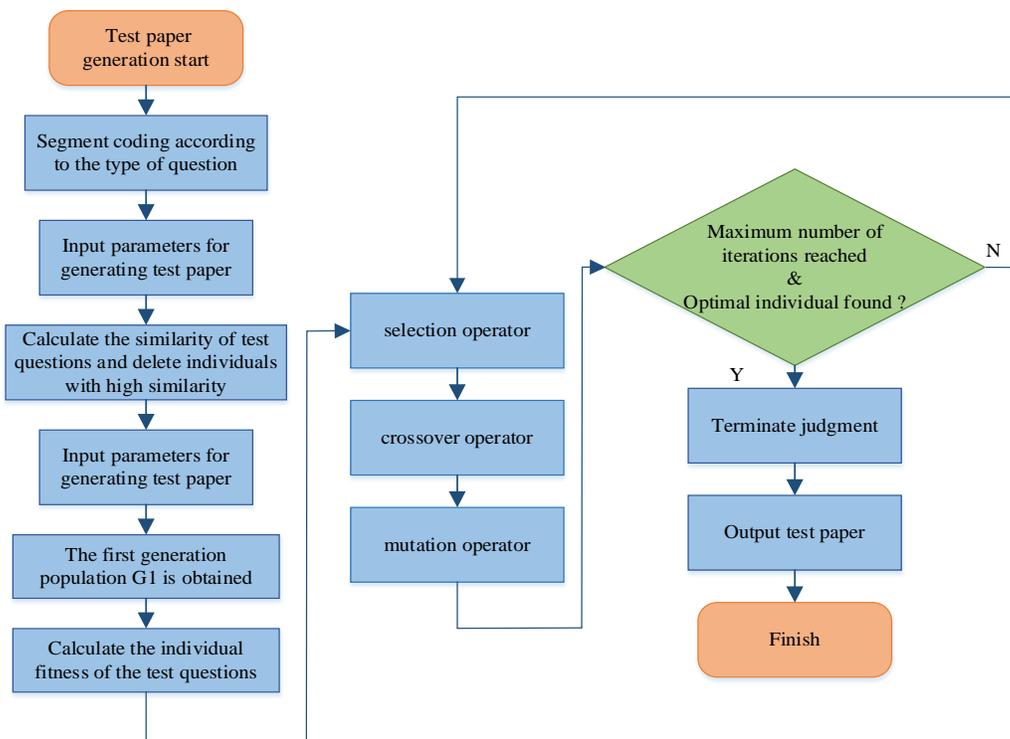
Fig. 2.   Group Volume Flow Chart based on Improved Genetic Algorithm.
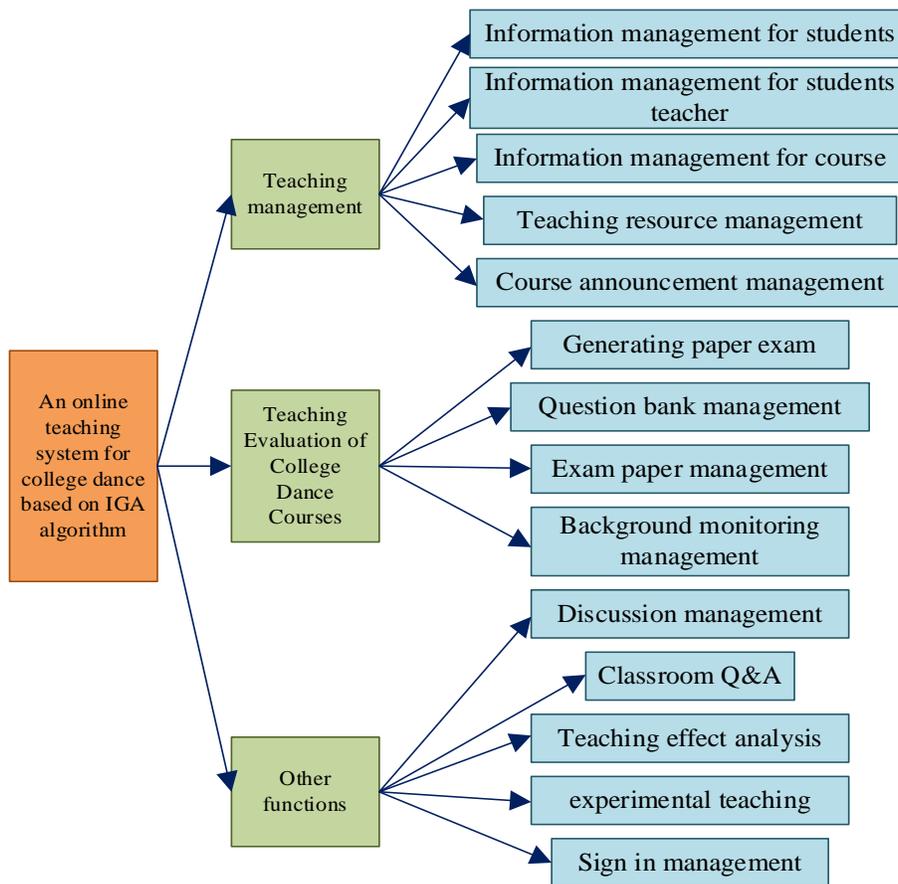
Fig. 3.   Functional Framework of College Dance Course Online Teaching System.

A block diagram of the design of the online dance teaching platform can be seen in Fig. 3 and summarizes the specific needs of the system into three parts.

In the first part, the functional requirements of the management module of the online teaching system include four blocks, namely classroom instructor management, dance teaching resource management, elective student management and background administrator. The specific content is to label and classify the teaching videos and teaching courseware of dance course. Among them, the label attributes are divided by grading settings. For example, teachers can set the first-level labels of the course by themselves, such as course theme, difficulty level of knowledge points, etc. Then, it is subdivided into second-level labels under the first-level label. For instance, the difficulty level of knowledge points is divided into three second-level labels: easy, medium, and difficult. Through the label classification of different teaching resources, the classification results of each label are obtained, so that students who choose this course can accurately select teaching resources from the classified teaching resources according to their own needs.

In the second part, the functional requirements of the evaluation module of the online teaching system of dance course in universities refer to using the IGA algorithm to realize the system artificial intelligence group test and provide the background monitoring. First, the IGA algorithm can effectively solve the optimization problem of artificial intelligence, so as to help teachers conduct targeted pre-class thinking tests for students, unit quizzes in the course, and organize online exams. Second, teachers can monitor students' learning, exams, and sign-in situations through the background.

In the third part, the functional requirements include answering questions in college dance classes, sign-in management, etc. Teachers are able to combine major functions of the dance course online teaching system with the functions of the third part so as to realize the role of supporting teaching.

The IGA algorithm proposed in this study has the following features, its incorporation of acceptance probabilities for simulated annealing of individuals after cross-variance operations, which can effectively reduce the problem of results falling into a local optimum at an early stage, is a major feature of this study to optimise the search capability of traditional GA algorithms.

IGA algorithm-based online teaching platform of college dance course runs on a computer with 2.3GHz, Intel Core i5 processor. The framework design adopts the B/S architecture design mode, the front-end framework adopts Bootstrap, and the back-end framework adopts Springboot. The software is implemented in Java language, combined with Intellij idea 2019, maven3.6.1 and other development tools.

IV. IMPLEMENTATION AND TESTING OF THE ONLINE TEACHING PLATFORM OF COLLEGE DANCE COURSE BASED ON THE IGA ALGORITHM

A. Experiment on the Performance of the Teaching Platform based on IGA Algorithm

The data set used in this experiment is randomly generated by the loop of the randperm function in Matlab. There are 570 test questions in the generated dataset, including 300 theoretical questions, 120 basic movement test questions, 90 work performance questions, and 60 choreography questions. The specific data set is shown in Table II.

According to the test question design and Table II, it can be concluded that the questions ID from 1-300 are theoretical questions, each question is three points; 301-420 questions are basic movement test questions, each question is 10 points; 421-510 questions are work performance questions, each question is 20 points; 511-570 questions are choreography questions, each question is 20 points. The difficulty coefficient of the test questions is randomly generated, including four kinds of difficulty, among which 0.0-0.25 represents easy, 0.26-0.5 represents medium difficulty, 0.51-0.75 represents difficult, and 0.76-1 represents very difficult. The knowledge points of the test questions are set to 10 types. The discrimination coefficient ranges from zero to one. If the discrimination degree of the test question is less than 0.2, it means that the gap between the test questions is small, which cannot reflect the real level of the students.

This study will compare four aspects, including running time, success rate, average fitness, and optimal fitness. Through the comparative analysis, the performance of the IGA algorithm in the paper grouping is studied. After the experiment, the comparison results of the traditional GA algorithm and the IGA algorithm obtained are shown in Fig. 4.

The above picture shows the difference between the genetic algorithm and the IGA algorithm. It can be seen from Fig. 4(a) that in the process of creating papers, the running time of the traditional genetic algorithm increases with the number of iterations. Although the running time of the IGA algorithm also increases with the number of iterations, its growth rate is significantly slower and the overall running time is shorter compared with the traditional genetic algorithm. From Fig. 4(b), it can be seen that the number of iterations is 100, the success rate of the traditional genetic algorithm is about 86.7%, while the success rate of the IGA algorithm is as high as 95.3%. There is no significant change in the success rate between the two algorithms as the number of iterations increases. When the iteration count reaches 500, the success ratio of traditional genetic algorithm is stable at 88.6%, while the success rate of the IGA algorithm is as high as 99%. In short, the success rate of the IGA algorithm is higher.

From Fig. 4(c), it can be seen that the average individual fitness of the IGA algorithm stabilizes at 0.929 when the number of iterations reaches 100, while the average individual fitness of traditional genetic algorithm is only 0.73. The average individual fitness of the traditional genetic algorithm increases significantly with the number of iterations. Comparing the two algorithms, it is clearly seen that the IGA algorithm has higher average individual fitness and requires less iterations.

TABLE II.   EXPERIMENTAL TEST DATA SET TABLE FOR THE IGA ALGORITHM-BASED TEACHING PLATFORM

Question ID	Type	Score	Difficulty _	Knowledge point	Distinction
1	1	3	0.34	6	0.42
2	1	3	0.68	1	0.31
3	1	3	0.91	10	0.67
4	1	3	0.22	5	0.41
5	1	3	0.47	2	0.38
...	...	...	...	...	...
570	4	20	0.86	9	0.49

(a)Test Paper Auto-production runtime comparison chart

(b)Test Paper Auto-production success rate comparison chart

(c)Fitness mean comparison chart

(d)Optimal fitness value comparison chart
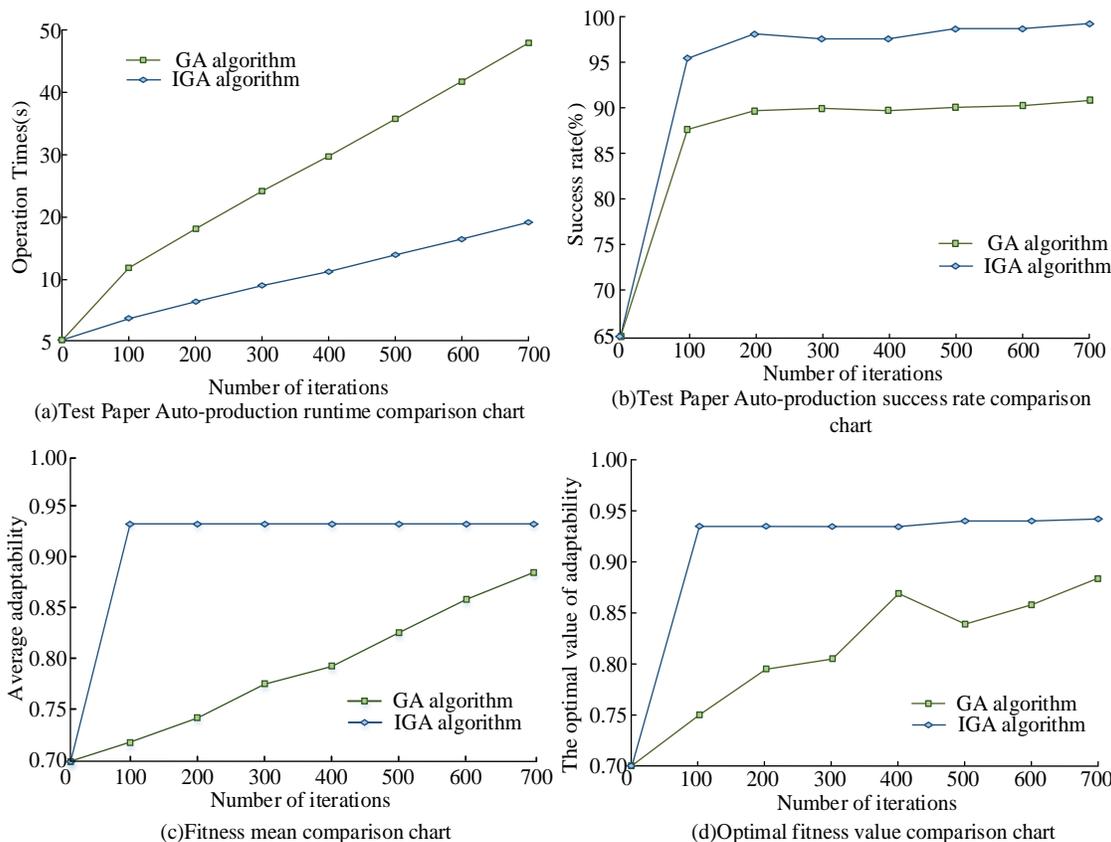
Fig. 4.   Comparison between GA Algorithm and IGA Algorithm.

According to the setting of formula (12), when the fitness calculation value is higher, the individual test papers obtained by the grouped papers are better. From Fig. 4(d), we can observe that when the iteration number reaches 100. Optimal value of the individual fitness of the traditional genetic algorithm is about 0.752, while that of the IGA algorithm is about 0.936. The results show that the optimal values of individual fitness for the two algorithms both increase with the iteration number. Comparing the optimal value of individual fitness of the traditional genetic algorithm and the IGA algorithm, the individual value of the individual fitness of the test papers grouped by the IGA algorithm is obviously greater as compared with the GA algorithm, which stabilizes at about 0.94 when the number of iterations is only 100. In short, the IGA algorithm performs better in terms of the best value of individual fitness of test papers.

To sum up, compared with the traditional genetic algorithm, the IGA algorithm proposed in this study has significantly improved the functions in all the above four aspects, indicating that the IGA algorithm can effectively improve the success rate, speed, memory consumption and the ability to produce quality papers in a short period of time in the online teaching platform for university dance.

B. Comparative Experiments on the Performance of the Teaching Platform System under the Conditions of Different Parameter

This study conducted a comparative analysis of the performance of online teaching platform system under the condition of different parameters. Considering the common characteristics of the GA algorithm, the size of the initial population will affect the results of the test. The initial population size was set to five different parameters, including 10, 50, 100, 150 and 200. Then, the comparison of fitness average (favg), the optimal value of the fitness function (Maximum Fitness value, fmax) and group volume uptime of the traditional genetic algorithm and the IGA algorithm are shown in Fig. 5.

(a)Comparison table of average fitness values of
different population sizes

(b)Comparison table of optimal fitness function
values for different population sizes

(b)Comparison table of running time of different
population sizes
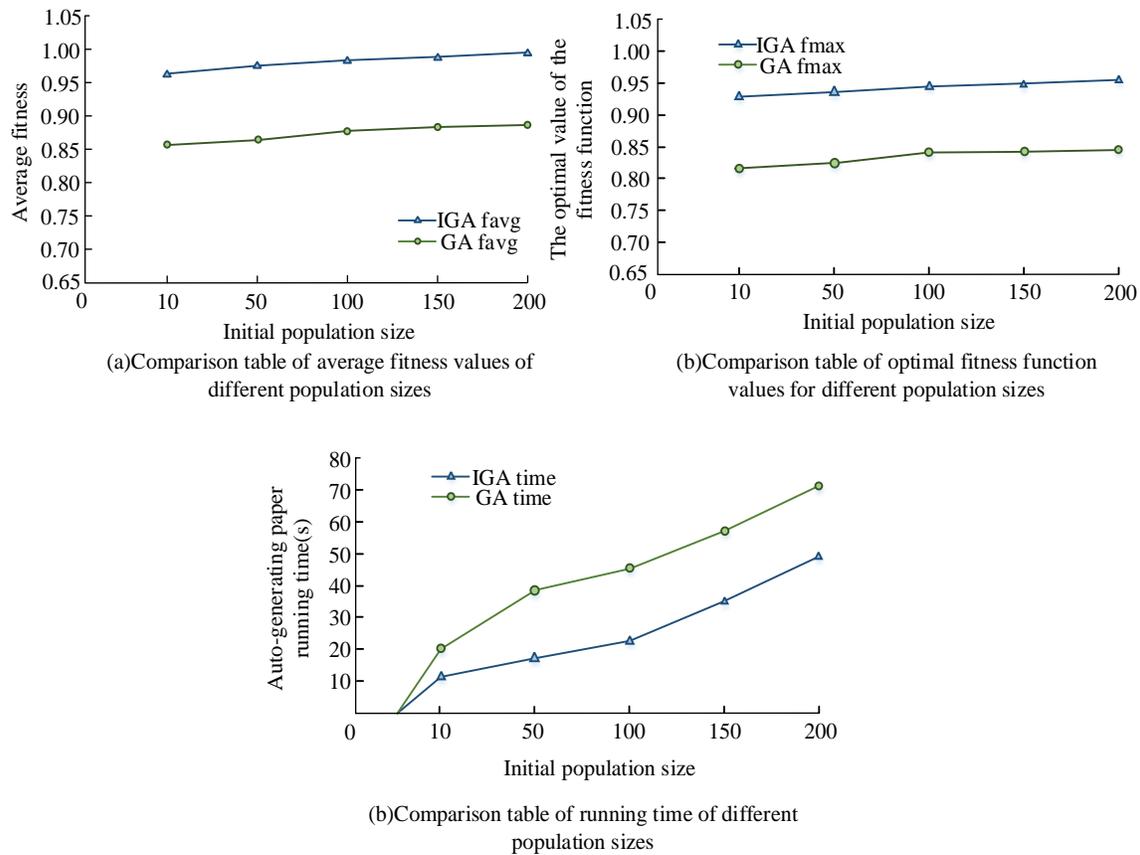
Fig. 5.   Performance Comparison of Teaching Platform System under Different Parameters.

As can be seen from the above figure, the size of the initial population has various degrees of influence on the three aspects of the algorithm. From Fig. 5(a) and 5(b), it can be seen that while the size of the initial population increases, both favg and fmax increase, indicating that the quality of the groups is better. However, as seen in Fig. 5(c), the time required to create a volume increases significantly as the size of the initial population increases. In summary, when the size of the initial population is at a value of 150, the quality of the rolls is excellent and the time required for the rolls is short, indicating the best value for the initial population.

Fig. 6 shows the relationship between the crossover probability (P_c) and the individual fitness of the IGA algorithm. As P_c increases, the value of A shows a trend of rising first and then falling. When the P_c value is 0.8, both IGA f_{avg} and IGA f_{max} reach the maximum value, indicating that when the P_c value is set to 0.8, the quality of the test papers produced by the IGA algorithm is the best.
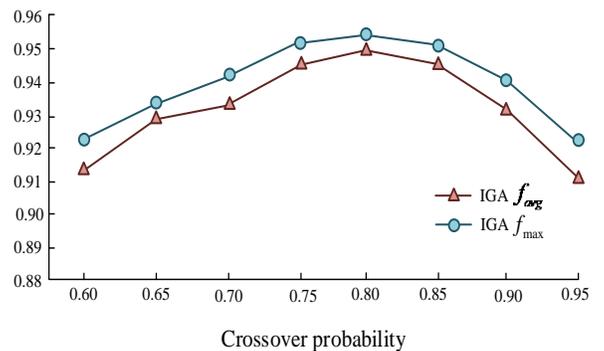
Fig. 6.   Comparison of Different Cross-Probability Experiments.
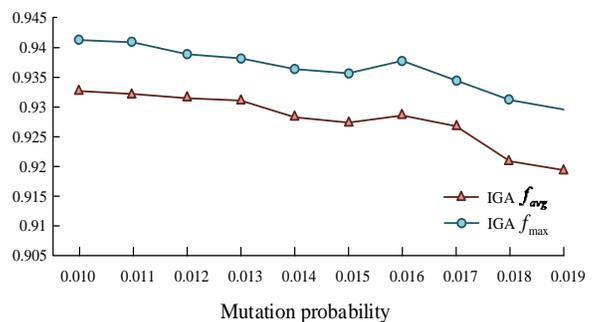
Fig. 7.   Comparison of Mutation Probability.

Fig. 7 shows the relationship between the mutation probability (P_m) and the individual fitness of the IGA algorithm. The results show that the best mark of the individual fitness and the average fitness of the group test P_m show a downward trend as the value increases. When the P_m value is 0.01, both IGA f_{avg} and IGA f_{max} reach the maximum value, indicating that when the P_c value isset to 0.1, the quality of the test paper produced by the IGA algorithm is the best.

In the IGA algorithm proposed in this study, the inclusion of a simulated annealing process allows it to effectively reduce the problem of results falling into a local optimum at an early stage and improve the success rate of the grouped volumes, which is the aim of this study to optimise the search capability of the traditional GA algorithm.

C. Performance Testing of the Online Teaching Platform System

To validate the feasibility of the online teaching system of college dance course based on IGA algorithm in actual operation, the teaching platform system will be functionally tested in this study. The functional test environment of the teaching platform system will run on a computer with 2.3GHz, Intel Core i5 processor. Functional tests will be conducted on the student function block and the teacher function block.

As can be seen from Table III, when the number of system visitors is 50, the response time of the traditional GA-based online teaching platform is 0.812 seconds, with a central processor occupancy of 6.06%. When the number of visitors to the system was 400, the response time of the online teaching platform based on the IGA algorithm was 6.158 seconds, with a central processor occupancy rate of 89.32%.

The system property of university dance online teaching platform based on IGA algorithm can directly reflect whether it meets the design request or not. In the performance testing environment of this experiment, JMeter 5.4.1 software will be used to automate the test in Win10 environment and Fiddle software will be used for script recording. The number of visits to the system is tested, and the specific operation is to simulate the number of concurrent users in the actual running process of the teaching platform. The test results are shown in Table IV.

As can be seen from Table IV, when the number of system visitors is 50, the response time of the online teaching platform based on the IGA algorithm is 0.532 seconds and the CPU occupation rate is 8.12%. When the number of system visitors is 400, the response time of the online teaching platform based on IGA algorithm is 5.52 seconds and the CPU occupancy rate is 68.12%, showing qualified performance in the number of visitors. The comparison with the conventional algorithm in Table III shows a significant improvement in response time, percentage of central processor occupation and balanced response time.

The online teaching platform based on IGA algorithm proposed in this study can meet the requirements of 400 users at the same time, indicating that the system meets the design request.

TABLE III. PERFORMANCE TEST TABLE OF ONLINE TEACHING PLATFORM OF COLLEGE DANCE COURSE BASED ON GA ALGORITHM

Number of concurrent users	response time/s	CPU usage/%	Balanced response time /s
50	0.532	4.06	0.01064
100	1.201	8.21	0.01201
150	1.989	16.42	0.01326
200	2.654	24.52	0.01327
300	3.992	47.31	0.01330
400	5.52	68.12	0.01380

TABLE IV. PERFORMANCE TEST TABLE OF ONLINE TEACHING PLATFORM OF COLLEGE DANCE COURSE BASED ON IGA ALGORITHM

Number of concurrent users	response time/s	CPU usage/%	Balanced response time /s
50	0.532	4.06	0.01064
100	1.201	8.21	0.01201
150	1.989	16.42	0.01326
200	2.654	24.52	0.01327
300	3.992	47.31	0.01330
400	5.52	68.12	0.01380

V. CONCLUSION

This research designs an online teaching platform system for college dance course based on IGA algorithm. In the research, the mathematical model of the intelligent question-making problem is applied, and the constructed model is verified. The results show that the success rate of the online teaching platform system for college dance courses based on IGA algorithm is 99% when the number of iterations is 500 times; when the number of iterations is 100, the average fitness is 0.929; when the number of iterations is 100 times, the moderate calculation value is 0.936. However, for the traditional GA algorithm, the results are 88.6%, 0.73 and 0.752, respectively. By comparing the GA algorithm and the IGA algorithm, the mathematic model of the test-setting problem based on the IGA algorithm is obviously superior to the GA algorithm in all aspects. At the same time, the research shows that the crossover probability and mutation probability are respectively $P_c = 0.8$. When the number of visitors to the

system is 50, the response time of the online teaching platform based on the IGA algorithm is 0.532 seconds, and the CPU occupancy rate is 8.12%. Compared to traditional genetic algorithms, the IGA algorithm proposed in this study shows significant improvements in all four of these areas, indicating that the IGA algorithm can effectively improve the success rate, speed, memory consumption and the ability to produce high quality papers in a short period of time on an online university dance teaching platform. Although the method proposed in the study can provide teachers and students with functions such as personalized teaching, intelligent test detection and teaching evaluation, the system lacks real-time monitoring of students' learning behavior and corresponding data analysis, so the follow-up research will conduct further research on this problem.

RREFERENCES

[1] H.-W. Chao et al. "Exploring the Effects of Blended Learning, Flipped Learning, and Online Remedial Teaching on Improving Students' Learning Performance and Motivation". International Journal of Technology and Human Interaction, vol. 17, pp. 98-114, 2021.

[2] Łukasz Tomczyk et al. "Evaluation of the Functionality of a New E-Learning Platform vs. Previous Experiences in E-Learning and the Self-Assessment of Own Digital Literacy", Sustainability, vol. 12, pp. 1-22, 2020.

[3] C. Jiang & R. Wang. "Design and Implementation of the Dance Teaching Live Broadcasting System Based on the Distance Education Platform", International English Education Research, vol. 2, pp. 27-29, 2019.

[4] Liang Z. Y. et al. "Hybrid Algorithm Based on Genetic Simulated Annealing Algorithm for Complex Multiproduct Scheduling Problem with Zero-Wait Constraint", Mathematical Problems in Engineering, vol. 4, pp. 1-21, 2021.

[5] Chen N. N. & Bai Z. Y. "The Application of Improved Genetic Algorithm and Least Square Method in Computer Mathematical Modeling", Application of Intelligent Systems in Multi-modal Information Analytics, vol. 1234, pp. 415-418, 2021.

[6] Y. Yang. "Quality Evaluation Method of a Mathematics Teaching Model Reform Based on an Improved Genetic Algorithm", Scientific Programming, vol. 18, pp. 1-10, 2021.

[7] Y. Zhang et al. "Reasearch on Shared Intelligent Test Paper Generating Algorithm Based on Multi Branches Tree", Procedia Environmental Sciences, vol. 14, pp. 69-83, 2019.

[8] ZH Xia et al. "Gesture Recognition Algorithm of Human Motion Target Based on Deep Neural Network", Mobile Information Systems, vol. 3, pp. 1-12, 2021.

[9] H. Wang & L. W. Wang. "Computer Network Design Based on Improved Genetic Algorithm", Application of Intelligent Systems in Multi-modal Information Analytics, vol. 1385, pp. 865-870, 2021.

[10] J. Xu. "Improved Genetic Algorithm to Solve the Scheduling Problem of College English Courses", Complexity, vol. 3, pp. 1:11, 2021.

[11] J. F. Wang. "Network distance education platform control system based on big data", International journal of internet protocol technology, vol. 12, pp. 173-180, 2019.

[12] Y. Gao et al. "Prediction of Vocational Education Coursework Assessment Results Based on the Simulated Annealing Neural Network", Wireless Communications and Mobile Computing, vol. 2, pp. 1-11, 2022.

[13] B. W. Deng, "Word order detection in English classroom teaching based on improved genetic algorithm of block coding", Journal of Intelligent and Fuzzy Systems, vol. 40, pp. 1-12, 2020.

[14] J. X. Shi. "Application Research and Analysis of Computer Virtual Reality Technology in College Dance Teaching", Cyber Security Intelligence and Analytics, vol. 1343, pp. 63-68, 2021.

[15] M. C. Tsai et al. "Exploring the effects of web-mediated activity-based learning and meaningful learning on improving students' learning effects, learning engagement, and academic motivation", Universal Access in the Information Society, vol. 19, pp. 783-798, 2021.

[16] Y. F. Wang & S. R. Wang. "Intelligent Test Paper Generation Based on Dynamic Programming Algorithm". Journal of Physics Conference Series, vol. 1682, pp. 1-5, 2020.

[17] Y. M. Chen. "Dance Teaching Activities in the Context of Informatization", Journal of Physics Conference Series, vol. 1852, pp. 022049, 2021.

[18] Shang M. L. "Thoughts on the Reform of Higher Vocational Dance Teaching Based on Big Data", vol. 1852, pp. 022084, 2021.

[19] Stevens Kym "Building an Authentic Cultural Curriculum through Tertiary Cultural Dance", Arts and Humanities in Higher Education: An International Journal of Theory, vol. 19, pp. 264-284, 2020.

[20] R. Huang. "The Scientific Training Programs for College Dance Students", Open Journal of Social Sciences, vol. 7, pp. 233-239, 2019.

Early-Warning Dropout Visualization Tool for Secondary Schools: Using Machine Learning, QR Code, GIS and Mobile Application Techniques

Judith Leo

School of Computational and Communication Science and Engineering (CoCSE)
The Nelson Mandela African Institution of Science and Technology
Arusha, Tanzania

Abstract—Investment in education through the provision of secondary school to the community is geared to develop human capital in Tanzania. However, these investments have been hampered by unacceptable higher rates of school dropouts, which seriously affect female students, since most schools do not have effective mechanisms for quality data management for immediate and effective decision making. Therefore, this study aims to solve the problem of data management from the school level in order to assist higher levels to receive appropriate and effective data on time through the use of emerging technologies such as machine learning, QR codes, and mobile application. To implement this solution, the study has explored the predictors of school dropout using a mixed approach with questionnaires and interview discussion. 600 participants participated in problem identification in the Arusha region. Through the use of design science research methodology, Unified Modeling Language, MYSQL, QR codes and mobile application techniques were integrated with Support Vector Machine to develop the proposed solution. Finally, the evaluation process considered 100 participants, and the results showed that an average of 89% of participants provided positive feedback on the functionalities of the developed tool to prevent dropouts in secondary schools in Africa at large.

Keywords—*Dropout; education; girls; machine learning; students; QR code; mobile application*

I. INTRODUCTION

Yearly in Tanzania, 30% of girls drop out before reaching form four compared to 15% of boys. Overall, more than 70,000 girls were expelled from school due to pregnancy between 2003 and 2015 [1]–[3]. As a result, girls aged between 15-24 years old are 2.5 times more likely to contract HIV/AIDS than their male counterparts [4]–[8]. Additionally, in Tanzania, especially at the village, ward, and district levels, female counterparts are not given their rights because of traditional beliefs, cultural practices, child labor, girls' menstruation periods, a long distance from home to school, poverty, etc.; hence, gender-based violence is still significantly high [9], [10], [19]–[24], [11]–[18]. Furthermore, the variation in completion and dropout rates in Tanzania differs from one region to another; for example, in Arusha, the percentage of dropouts of total students enrolled in 2015 was 8.9% and 2.3%, respectively, making Arusha the leading region in school dropout [25]–[28](BEST, 2016). Moreover, it has been noted that school dropout is more dominant in lower

secondary levels than in Ato primary levels. The major factor contributing to this high dropout rate in secondary school is early pregnancy, which in turn makes the victims young women, with a poor economy and a high probability of dropout from HIV/AIDS patients [29][30], as briefly described in Fig. 1. Fig. 1 shows a young adolescent secondary school girl who is faced with many challenges, such as child labor, home chores, cultural practices such as early marriage and girls not supposed to be educated, long distances between the home premises and school, and menstruation issues to mention a few [26], [31]–[37].
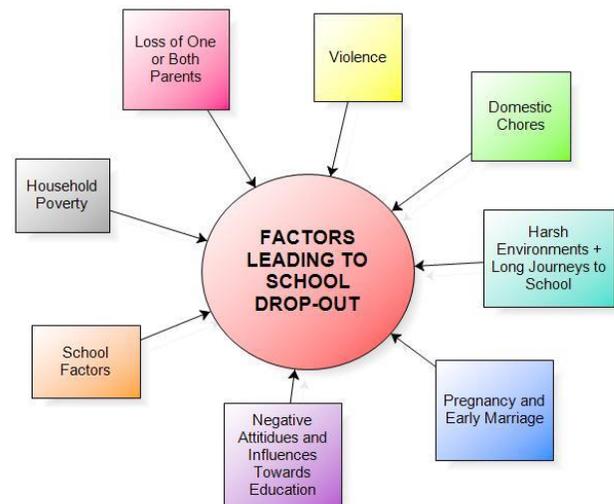
Fig. 1. Description of the Challenges that Lead to School Dropout [38].

The government of Tanzania has made efforts to ensure equal opportunities for boys and girls in schools. However, many initiatives have been implemented using the top-down approach, which focus on ensuring that the government uses an electronic-based information system to collect and store data from lower levels, such as schools, wards, and districts [3][39]. The top-down approach has proven to produce poor quality data and delays in data submission and decision making. This is because the lower levels (schools, wards, and district education offices) do not have proper mechanisms for data collection and storage. Additionally, most schools in Tanzania still use manual systems to collect data; for instance,

the class teacher has to document the attendance of more than thirty (30) students twice a day [40]–[42]. The activity of attendance and collection is tedious to the extent that it is not done every day, and sometimes wrong information is captured in the attendance book. Furthermore, there are no effective mechanisms to track and ensure that students enrolled in schools complete their studies, according to [38], [43]–[46]. Moreover, there are poor-quality datasets that can be used to develop demand and data-driven solutions for immediate action, effective planning, campaigns, and policies on matters related to the elimination and reduction of school dropouts [47]–[50]. Therefore, this research study aims to solve the problem of data collection, on-time data submission, effective data visualization, and availability of quality datasets from the lower levels (schools) in order to assist higher levels of receiving appropriate and effective data on time through the use of emerging technologies such as machine learning (ML), QR codes, statistical visualization, and mobile applications, which are cheap to operate, understandable, and sustainable. In connection with the proposed technical solution, the study, with the collaboration of the government of Tanzania through the Ministry of Education, Science and Technology and other related stakeholders such as HakiElimu, also aims to conduct campaigns, workshops, seminars for secondary school students, teachers, parents, and related domains of education sectors on matters related to the elimination and reduction of school dropout.

II. MATERIALS AND METHODS

This section presents the methodologies of the research study, including the study area, participants, sampling procedures, tools, and methods used throughout the research activities. The study followed the design science research methodology (DSRM), which consists of three stages: problem identification, solution development, and validation [51]. The DSRM was chosen because it is an outcome-based method. It also offers iterations through the research stages and provides guidelines for evaluation, with a focus on improving the functional performance of each interface. The study took a total period of three years from 2017 to 2020 to observe the impact (outcome) of the research study based on the initial focus group of participants.

A. Study Area

This study was conducted in six districts of the Arusha region in Tanzania. This is because, according to the literature, the Arusha region has a high school dropout rate, and the region can well represent the situation that is happening in both rural and urban areas based on its existing geographical position in the country and the nature of its people [52]. The region is positioned in such a way that it attracts many tourists; therefore, natives, including their children (local people in the area), tend to depend on tourism activities instead of focusing on acquiring education. In addition, the majority of people in Arusha are Maasai who still practice and seriously follow their culture and customs; hence, most young adolescent girls are not allowed to go to school, practice early marriage, and also have babies at a very young age, to mention a few. This study considered six districts in the Arusha region, namely Meru, Longido, Ngorongoro, Karatu,

Moduli, and Arusha urban and rural areas, as shown in Fig. 2. The urban and rural areas of Arusha were considered as one district.
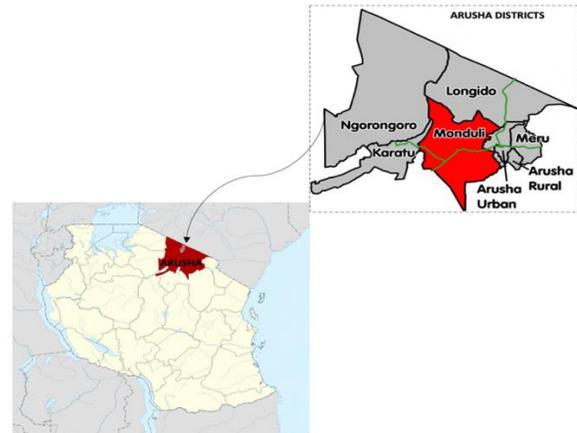
Fig. 2. Districts of Arusha Region in Tanzania.

B. Problem Identification Stage

In the problem identification stage, an intensive literature review was conducted on various strategies and mechanisms that have been implemented so far in the area, especially the information and communication technology (ICT), -based solutions, to understand the existing challenges that led to continued secondary school dropout. In the process, a mixed design approach, qualitative and quantitative methods with focus group discussions, and administered questionnaires were conducted, as shown in Fig. 3, in order to receive feedback and constructive comments from the participants. This feedback and comments were used to identify the requirements for the design, development, and validation of the proposed solution Early-warning Dropout Visualization Tool (EDV-Tool).

The study population comprised 600 participants, including teachers, students, parents, education officers, health officers, and policymakers, in the Arusha region. In each district, 100 participants were systematically selected to form the sample. In general, 100 participants in each district were distributed as follows: 20% teachers, 20% students, 20% parents, 20% education officers, 10% health officers, and 10% policymakers, as shown in Table I. The collected data were then coded and analyzed using R programming software to determine the most appearing or mentioned features and functionalities of the proposed solution.

C. Design and Development of EDV-Tool

The rapid application development (RAD) method was followed in the design and development processes of the EDV-Tool. RAD was chosen because it allows short and iterative development cycles with effective results. Five iterations were conducted between development processes. The tool was developed using MySQL, JSON, JavaScript, Java Server Pages, Python (ML), QR codes, mobile apps, and ML techniques.

D. Evaluation of EDV-Tool

The evaluation of the developed EDV-Tool was conducted in the last year of the research study (2020). The evaluation considered only 100 participants from the urban and rural districts of Arusha. This is because the urban and rural districts of Arusha represented both urban and rural features and challenges well based on the nature and responses received during the data collection stage. During the evaluation process, a mixed-design approach of quantitative and qualitative methods was deployed as the study design. Focus group discussions were used in the process to obtain a large amount of feedback and participants. Focus group discussions and interviewer-administered questionnaires focused on understanding the performance of the tool in terms of both functional and nonfunctional features. Focus group discussions were limited to a maximum of 15 participants to ensure equal participation. Audio recorders, mobile phones, and laptops were used to record, transcribe, and code the discussions and interviews. The interfaces and functionalities (requirements) of the tool were then evaluated in four stages: unit, integration, system, and validation tests. During unit testing, each module was tested independently before integration with other modules of the tool. Integration testing was then performed after integrating one module with another. These tests were performed to eliminate any syntax or logical errors and maintain the compatibility of all modules of the

tool. After successfully integrating all modules, a black-box testing methodology was used to verify the functionalities of the tool that were identified during the requirement identification and collection processes. The black-box testing methodology is a type of testing approach in which the tester tests the system without having knowledge of the internal codes. Finally, a validation test was performed by deploying the developed tool to 100 selected participants from the Arusha urban and rural districts. Data from the questionnaires were analyzed statistically using the R programming software.

III. RESULTS

This section briefly presents the results of each stage taken towards the development of the EDV-Tool. The content in this section is presented in four subsections, which include the results of problem identification, design, development (description of EDV-Tool), and evaluation.

A. Results of Problem Identification

The data collection exercise conducted in the Arusha district using focus group discussions and administered questionnaires. Tables I and II briefly describe the selected participants and collected requirements, and Fig. 3 presents the most identified existing predictors of dropout in secondary school.

TABLE I. DESCRIPTION OF THE PARTICIPANTS DEMOGRAPHIC INFORMATION

SN	Characteristics of Participants		Total Sample (N=600)	Percentage Distribution(%)	Description
1.	Age in years				
		≤ 20	260	43.33%	-The average age group for the study was 28 years old.
		21-35	240	40%	
		≥ 36	100	16.67%	
2.	Level of Education				
		Primary	60	10%	-Secondary comprised of participants with junior and senior high school levels at a ratio of 2:3 respectively.
		Secondary	200	33.33%	
		Certificate Level	140	23.33%	
		Diploma Level	100	16.67%	-Non-formal education comprised of participants who do not have any kind of education whereby majority were parents.
		University	100	16.67%	
3.	Disctrict				
		Arusha urban & rural	100	16.67%	-In each district, 100 participants were selected for problem identification. Whereas, for evaluation only 100 participants from Arusha urban & rural district were involved.
		Meru	100	16.67%	
		Longido	100	16.67%	
		Karatu	100	16.67%	
		Monduli	100	16.67%	
		Ngorongoro	100	16.67%	
4.	Gender				
		Female	360	60%	-Majority of participants were Female at Female to Male ratio of 3:2.
		Male	240	40%	
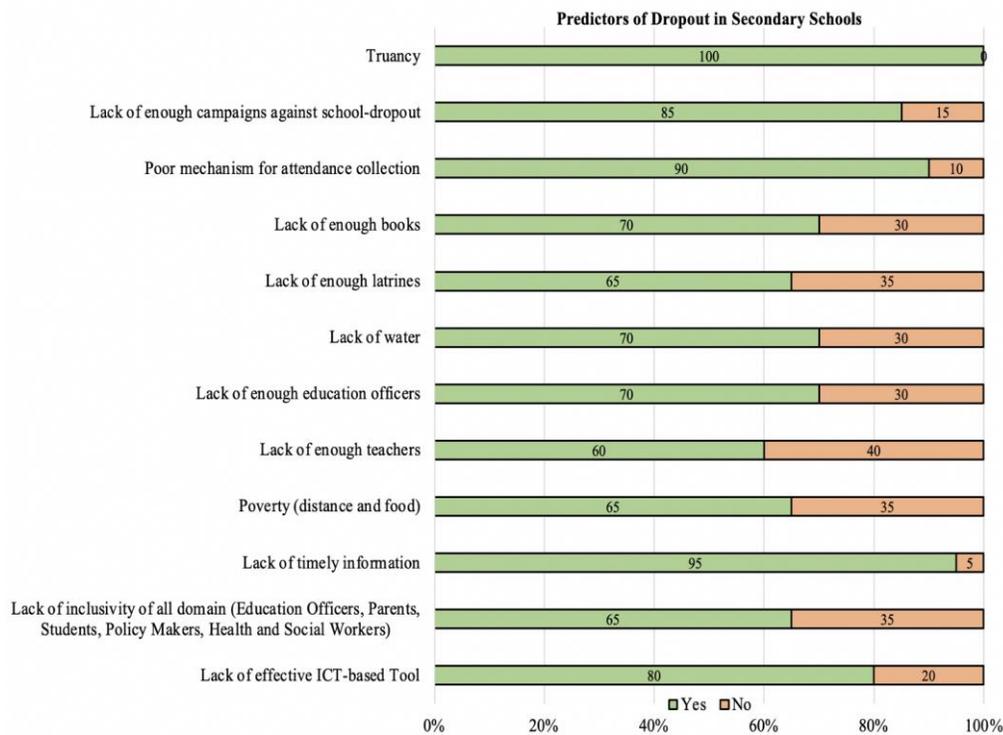
Fig. 3.   Predictors of Dropout in Secondary School in Tanzania.

TABLE II.   DESCRIPTION OF FUNCTIONAL AND NON-FUNCTIONAL REQUIREMENTS

Functional Requirements		
SN	**Requirement and its Description**	
1.	Generate visual, excel-format, and map-based reports	
2.	Search records per keywords, location, and image.	
3.	Submit all essential data on time.	
Non-functional Requirements		
SN	**Requirement**	**Description**
1.	Security	The system should be secured at both data access and transmission levels. All users should access data based on their usability levels and uniques credentials (user-names and passwords).
2.	Accessibility	The system should be available at all times and easy to use. The attendance collection module should capture data both when on-line and offline.
3.	Efficiency	The system should be reliable, affordable, and scalable.
4.	Interoperability	The system should be compatible with all browsers and android OS.
5.	Maintainability	The system should be easy to maintaining to the extent that when one service is under maintenance should not hinder other services from operation.

B. Results of EDV-Tool Design

Fig. 4 shows the conceptual framework of the EDV-Tool, which describes the flow of data and main users. In connection with Fig. 4, Table III describes the main users of the EDV-Tool and a brief explanation of their activities in terms of the data that they submitted to and received from the tool. Fig. 5 shows the ML model development approach that was deployed, starting from the users sending the required data to the online database (DB). The dataset was then cleaned and transformed into the appropriate format. Because the data are not well balanced at a ratio of 0.43 (Yes Dropout (30%) and No Dropout (70)), the SMOTE oversampling techniques were used to balance and then, the best features for prediction of the dropout status truancy rate, lack of sufficient teachers, lack of digital system for data collection, lack of sufficient latrines, and water were selected using random forest (RF) techniques. The model was then developed considering three supervised ML algorithms: support vector (SV), extra tree (ET), and XGBoost classifiers. Using sensitivity [0.999, 0.941, 0.892], accuracy [0.732, 0.587, 0.522], and F1-score [0.821, 0.756, 0.744], respectively, the models were validated, and then the Wilcoxon signal ranking was used to select the best model based on the three validation techniques. Finally, the user acceptance procedure was performed using 100 stakeholders from Arusha's urban and rural districts. Based on this approach, the SV classifier was selected as the best model for developing the development of EDV-Tool. Fig. 6 describes the flow of information to enable attendance data from the school level to be captured using either online or offline modes based on the status of Internet connectivity in the area.
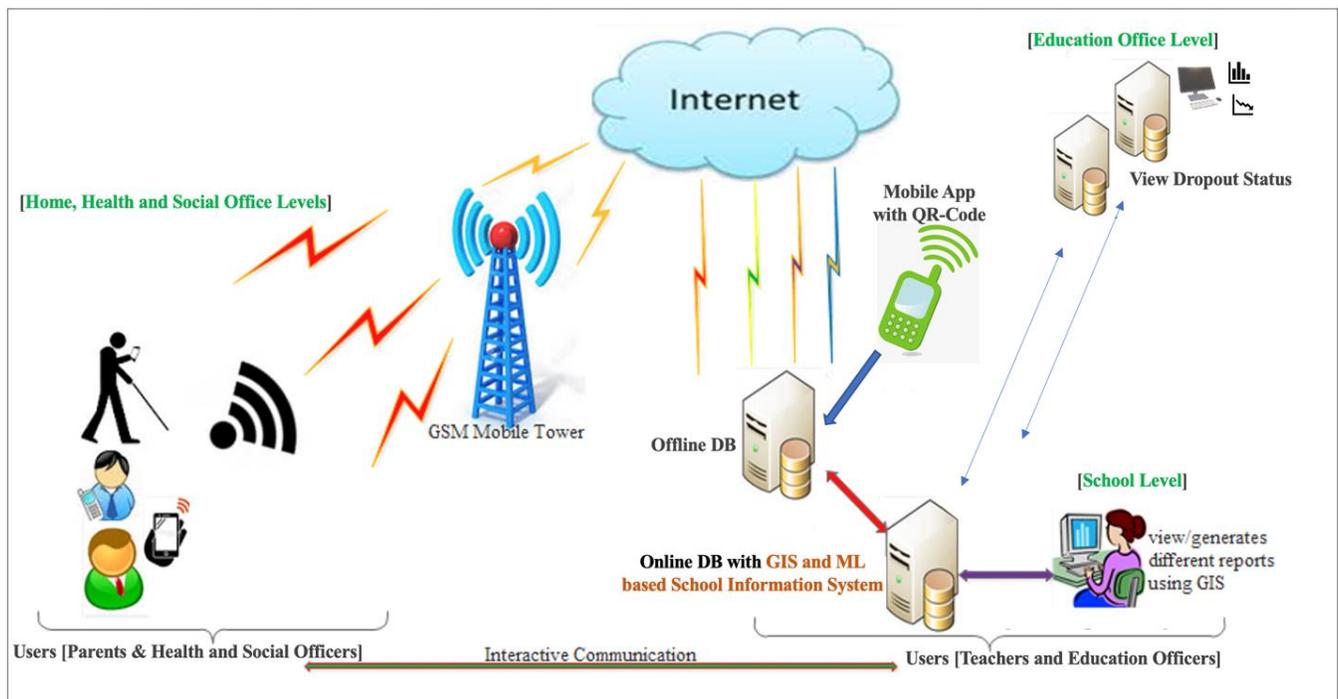
Fig. 4.   Conceptual Framework of EDV-Tool.

TABLE III.   DESCRIPTION OF USERS AND THEIR ACTIVITIES IN THE EDV-TOOL

S/N	Name of User	Submitted Data in EDV-Tool	Received Data from EDV-Tool
1.	Teachers	-Feed data into the EDV system -Update pupil status into the system	- System user
2.	Education Officer	-Visualization of system output/information for decision and planning purposes.	- System user - Decision making
3.	Students	-Visualize system information -Help in identifying schools with high-dropout. -Locating pupils affected by dropout.	- System user
4.	Parents	-Key decision maker and planner all matters pertaining education in the district. -Visualize system information	- System user
5.	Hospitals and Social Officers	-Victims of dropout will provide data on the causes of dropout.	- Beneficiaries of project outcomes
6.	Policy Makers and Community	-Receive data for policy making -Get knowledge and information	- Beneficiaries of project outcomes
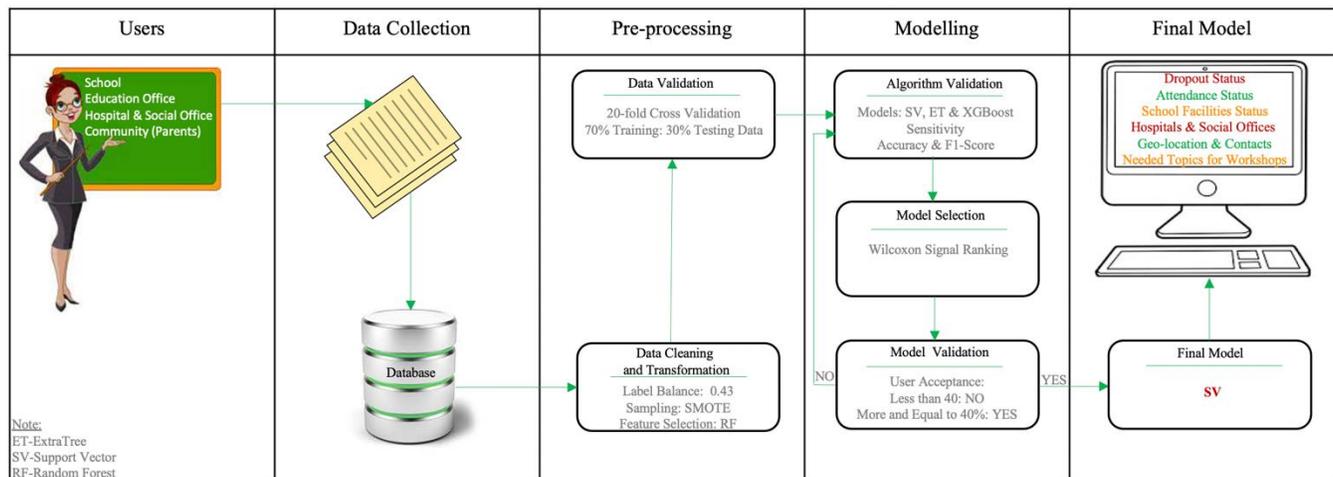
Fig. 5.   Machine Learning Model Development Approach.
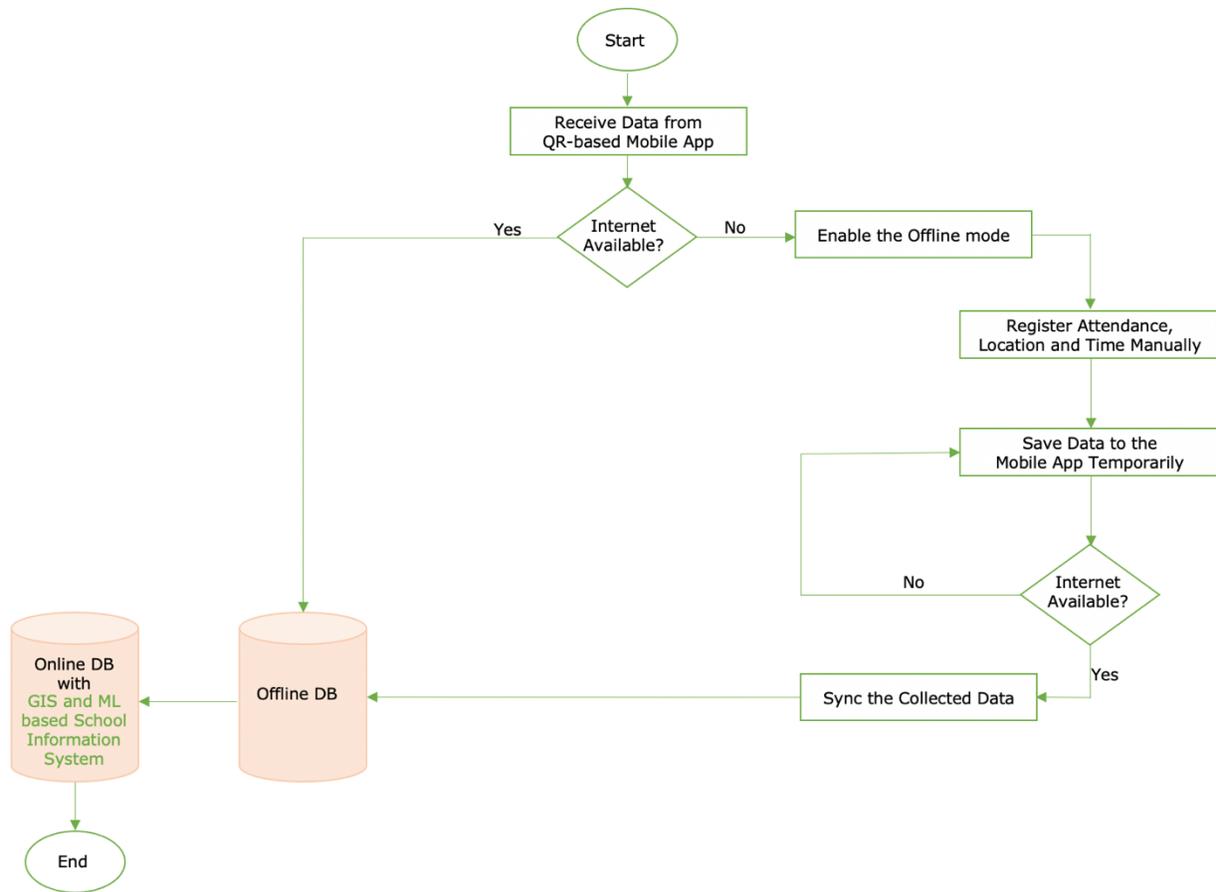
Fig. 6. Flowchart of EDV-Tool Describing Online and Offline Modes.

C. Results of EDV-Tool Development

The EDV-Tool consists of (i) a mobile application that is integrated with the QR code and visualization techniques to facilitate attendance collection from both teachers and students, easy to understand graphical representation of school dropout status, and timely sharing of information to the users; (ii) Geographic Information System (GIS) and ML-based school information system, which is responsible for processing the integrated data from school (teachers and students), and education, health, and social offices and then transforms the data into required formats such as Excel, graphs, and map-based formats; and (iii) a communication subsystem that integrates data between offline and online databases, and sends data to be displayed by the EDV-Tool.

The main features of the developed system are as follows: secured login and report sample interfaces, as shown in Fig. 7. Fig. 8 shows the training on the use of EDV-Tool which was done at Murieth secondary school and the exhibition of the EDV-Tool to "SIKU YA VIJANA KIMATAIFA which was held on 2018 and also, describe the campaigns and seminars which were done to different stakeholders on elimination of the school dropout.
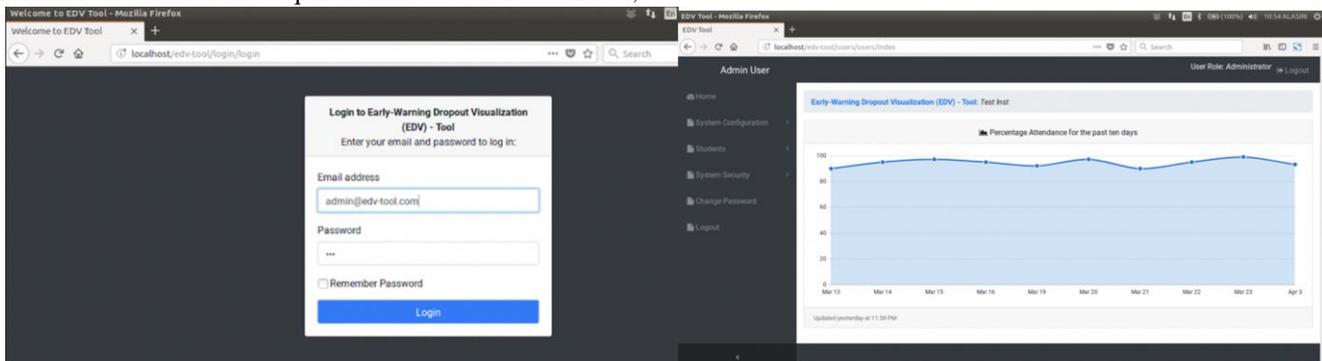
Fig. 7. Interfaces of EDV-Tool Describing the Login Section and Report Sample.

D. Results of EDV-Tool Evaluation

Overall, at an average of 89% of the participants agreed that the system performs well in terms of visual reports, essential data for all domains, easy to search data, security level, accessibility level, effectiveness, interoperability, and maintainability as described in Fig. 9. The remaining 11%.

of the average evaluation results represent the majority of participants who strictly practice cultural activities and do not believe in educating girls in their community. In addition, the analysis results of the 100 participants, as described in Fig. 10, showed that there was a ratio of 45:55 for male to female ratio and 60:40 for participants with education higher than secondary school (O-level) to those with education lower than secondary school.

Fig. 8.   Results of Evaluation of EDV-Tool.
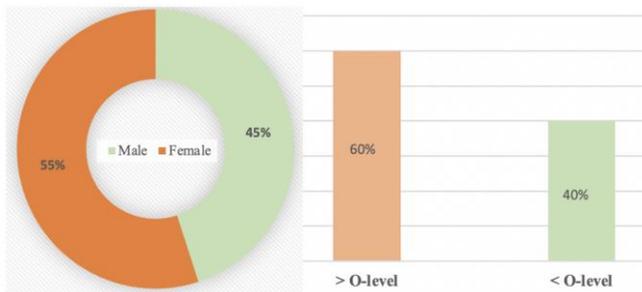
Fig. 9.   Description of the Participants in Terms of Gender and Education Levels.

IV.   DISCUSSION

The developed EDV tool is a combination of not only an ICT based system but also integrated measures and strategies such as involvement of all key players of education under one system, planned campaigns, seminars, and workshops to educate, motivate, and encourage students on the importance of education, elimination of bad-cultural practices such as early marriages and pregnancies, and HIV-AIDS education. Therefore, the tool captured all essential requirements collected during the problem identification stage (data collection and literature review), as described in Fig. 3. In addition, the tool enables quality data collection, on-time data submission, and effective data visualization from the low level (e.g., school – village – ward–district) to higher levels (government-Ministry) through the use of emerging technologies such as ML, QR-code, mobile apps, GIS, and visualization techniques that are inexpensive, easy to use, and

sustainable. Furthermore, the tool can be accessed both via mobile devices and the web to monitor the attendance of students and staff, as well as other statistical information regarding school resources or facilities such as number of toilets, availability of tap water or water, number of classrooms, teachers, and students. Hence, through the generated data, the ML techniques enable the tool to predict the students who are at risk of school dropout weekly, and then, the status of dropout and non-dropout is shared with the teachers, parents, and education officers. The tool has proven to provide understandable and early representations of school dropout status, and as a result, the whole scenario assists in the reduction of HIV/AIDS risk behaviors since the students, especially adolescent girls, who are at the risk of dropping out, will be controlled and monitored by sending daily attendance to their parents via SMS and special alerts to the school and education offices. Based on the school dropout status, special meetings with parents can be easily arranged in order to discuss how the students can be assisted, and sometimes special interventions such as campaigns, workshops, and seminars are arranged with the help of other domains such as education officers, social workers, medical experts, policy makers, and the community. Lastly, the tool contains geo-location and contact information of nearby hospitals; therefore, any required action for incidences, such as rape and education on early pregnancies and HIV, can be easily accessed.

Apart from these major contributions towards the prevention of school dropout, other issues were learned during the different processes of tool development and

implementation. These contributions include: (i) It is important to understand the timetable (school examination, marking of the examination, sports day, work deadlines, etc.) of the school before visiting in order to select the best time for data collection, campaigns, evaluation, and other activities related to the study. (ii) Some schools do not even have windows and doors in their classrooms and walls that surround them for security purposes. Such schools find it unusual to be given ICT-based tools and solutions such as the EDV-Tool instead of being supported to build walls and fix-in classrooms with the required doors and windows. (iii) Some schools have the issue of poverty as the major problem that leads to most students not being able to attend classes and courses until the end of the day without going back home for lunch. (iv) Some schools in the Maasai tribe do not cooperate well with female researchers; therefore, planning to visit the schools should involve a male figure who will initiate the communication. Therefore, based on these findings, it is very important to prepare as early as possible measures and strategies to overcome these dynamics and shortcomings, as well as to justify that the intention of the study is still valid and flexible enough to adopt other measures that will come out during the data collection meetings and discussions on the required solutions with stakeholders.

V. CONCLUSION

Efforts to prevent school dropouts in Tanzania entail reducing and eliminating risk factors across all essential domains (i.e., individual, peers, family, school, community, culture, government level, and its associated subdomains) in the education sector. Therefore, the developed and implemented EDV-Tool has taken into consideration these factors by being a combination of school-community-based mechanisms and strategies that have involved key players from schools, communities (parents), education offices (ward, district, and government levels), health, and social offices. In addition, the tool has facilitated the use of open data from TAMISEMI as well as encouraging the flow of information from low levels (schools and communities) to government levels, and on-time access of school dropout status to all essential domains in the education sector and to the highest organ, which is the Ministry of Education, Science and Technology through the linkage with the education officers and schools. In addition, based on the data collected via the EDV-Tool, the schools were able to identify the required number of seminars and select appropriate topics for discussion in terms of seminars, campaigns, and workshops on matters related to the prevention of school dropout and its risk factors. Therefore, through the use of the EDV-Tool mechanism and strategies, all associated dropout risk behaviors such as truancy, HIV-risk behaviors, and bad-cultural practices such as early pregnancies and marriages will be timely identified and special measures will be timely implemented with the involvement of all essential domains. The study recommends the implementation of the EDV-Tool in all schools in Tanzania, especially those in rural areas, to reduce the rate of school dropout and facilitate timely collection and sharing of quality data, which is essential for effective and timely decision making on school dropout issues

at large. Future research should focus on conducting the same study in other regions of Tanzania.

VII. DATA AVAILABILITY STATEMENT

The data used to support the results of this research study are available from the corresponding author upon request.

CONFICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] MOHCDGEC, "National Adolescent Health and Development Strategy 2018 - 2022," Minist. Heal. Community Dev. Gender, Elder. Child., pp. 1–41, 2018.

[2] URT, "National Five Year Development Plan 2021/22-2025/26: Realising Competitiveness and Industrialisation for Human Development," United Repub. Tanzania, no. June, p. 321, 2021.

[3] Tanzania, "Education Sector Performance Report 2018/2019," Development, vol. 7, no. 5,371,780,231.09, pp. 2,274,923,575.00-29.08, 2019.

[4] J. Damas, "Examining the factors for girls' dropout in secondary schools in Meatu district council," 2016.

[5] G. Subrahmanyam, "Gender perspectives on causes and effects of school dropouts," Sida, p. 111, 2016, [Online]. Available: http://www.sida.se/publications.

[6] The Secretary of State, "Girls' education: towards a better future for all," Dep. Int. Dev., vol. 1, no. 1, pp. 1–44, 2016, [Online]. Available: http://dictionary.cambridge.org/dictionary/english/challenge%5Cnhttp:// www.prb.org/Publications/Reports/2003/EmpoweringWomenDevelopin gSocietyFemaleEducationintheMiddleEastandNorthAfrica.aspx%5Cntt p://www.aup.edu.pk/sj_pdf/SOCIO-ECONOMIC AND CULTURAL CONS.

[7] "THE SUSTAINABLE DEVELOPMENT GOALS (SDGS) AND AFRICA ' S AGENDA 2063."

[8] R. B. Johnston, "Arsenic and the 2030 Agenda for sustainable development," Arsen. Res. Glob. Sustain. - Proc. 6th Int. Congr. Arsen. Environ. AS 2016, pp. 12–14, 2016, doi: 10.1201/b20466-7.

[9] T. Smith, "Influence of cultural practices on soybean nectar production," 2016, doi: 10.1603/ice.2016.109936.

[10] G. Campbell, K. A. Roberts, and N. Sarkaria, "Other Harmful Traditional Practices," Harmful Tradit. Pract., pp. 119–127, 2020, doi: 10.1057/978-1-137-53312-8_8.

[11] T. L. Michubu, "a Study of the Causes of Dropout in Public Primary Schools in Ndoleli Division, Igembe North District, Kenya," pp. 1–41, 2012.

[12] C. Wolman, R. Bruininks, and M. L. Thurlow, "Dropouts and Dropout Programs: Implications for Special Education," Remedial Spec. Educ., vol. 10, no. 5, pp. 6–20, 1989, doi: 10.1177/074193258901000504.

[13] G. R. Ole Kinisa, "Effectiveness of Educational Policy in Curbing School Dropout in Secondary Schools in Tanzania: A Case of Dodoma City," Int. J. Sci. Res. Publ., vol. 9, no. 5, p. p8916, 2019, doi: 10.29322/ijsrp.9.05.2019.p8916.

[14] T. Baasher, R. H. O. Bannerman, H. Rushwan, and I. Sharaf, "Traditional Practices Affecting the Health of Women and Children," Stud. Fam. Plann., vol. 13, no. 6/7, p. 222, 1982, doi: 10.2307/1965452.

[15] G. Offorma, "GIRL-CHILD EDUCATION IN AFRICA," no. July 2009, 2015.

[16] A. M. Mohamed, D. H. K. Mberia, and D. W. Muturi, "Influence of Socio-Cultural Practices on Girl Child Participation in Secondary Schools in Garowe, Puntland," IOSR J. Humanit. Soc. Sci., vol. 22, no. 05, pp. 78–85, 2017, doi: 10.9790/0837-2205077885.

[17] D. Sekiwu, "Investigating the relationship between school attendance and academic performance in universal primary education: The case of Uganda," African Educ. Res. J., vol. 8, no. 2, pp. 152–160, 2020, doi: 10.30918/aerj.82.20.017.

[18] M. a Rihani, "5 Benefits of Girls ' Secondary Education," no. 202, 1825.

[19] C. F. Hjorth et al., "Mental health and school dropout across educational levels and genders: A 4.8-year follow-up study," BMC Public Health, vol. 16, no. 1, pp. 1–12, 2016, doi: 10.1186/s12889-016-3622-8.

[20] N. C. Andiema, "Influence of Culture on Girl Child Education in Central Pokot Sub County, Kenya," East African J. Educ. Stud., vol. 3, no. 1, pp. 26–38, 2021, doi: 10.37284/eajes.3.1.279.

[21] O. D. Akinyi and C. E. Musani, "Economic Factors Affecting Girls Academic Performance (Kcse) In Mixed Secondary Schools: A Case Of Nakuru Municipality," Eur. J. Econ. Law Polit., vol. 03, no. 01, pp. 18–52, 2016, doi: 10.19044/elp.v3no1a4.

[22] D. J. Dockery, "School Dropout Indicators, Trends, and Interventions for School Counselors Donna J. Dockery Virginia Commonwealth University."

[23] H. A. Sackey, The Determinants of School Attendance and Attainment in Ghana : A Gender Perspective, no. 12. 2007.

[24] J. P. Casey, "Understanding High Dropout Rates in Primary School Education in Mozambique," no. June, p. 73, 2014, [Online]. Available: http://www.diva-portal.org/smash/get/diva2:732524/FULLTEXT01.pdf.

[25] dkk Xian, Gou Li, "A study on Demonetization and its Impact on Corruption and Black Money," Saudi J. Humanit. Soc. Sci., vol. 2, no. 5, pp. 597–610, 2011, doi: 10.21276/sjhss.

[26] "Out Rates in Community Secondary Schools in Tanzania : a," 2014.

[27] S. I. Segumba et al., "Factors Leading To Problems of Drop Out in Primary School Pupils in Temeke District," 2015, [Online]. Available: http://repository.out.ac.tz/1464/1/SALEHE_IDDI_SEGUMBA.pdf.

[28] T. H. E. Barriers, U. I. N. Action, and K. E. Y. Achievements, "Girls ' Education in Tanzania Girls ' Education : Tanzania At a Glance," pp. 2002–2003, 2006.

[29] United Nations, The fourth domain for gender equality, decision-making and power. 2010.

[30] H. George, "An Investigation on the factors influencing primary school pupils' dropout in Ruangwa district, Lindi region," Open Univ. Tanzania, vol. 1, no. 2, pp. 274–282, 2017.

[31] C. Awinia, "Free Basic Education and Gender Disparities in Tanzania: An Empirical Assessment of Challenges and Policy Options," Huria J., vol. 26, no. 2, pp. 1–22, 2019.

[32] National Dropout Prevention Center/ Network, "Effective Strategies for Dropout Prevention," 15 Eff. Strateg. Dropout Prev., 2014, [Online]. Available: http://www.dropoutprevention.org/effective-strategies.

[33] J. Abdulkadir and R. B. Zainuddin, "Cultural Practices as a Hindrance to Girl Child Education Attainment in Northern Nigeria," Asian J. Multidiscip. Stud., vol. 6, no. 3, pp. 27–31, 2018.

[34] S. Saryal et al., "Forced Marriage in the UK A scoping study on the experience of women from," Natl. Online Resour. Cent. Violence Against Women, vol. 3, no. 11, pp. 531–551, 2004, [Online]. Available: www.nber.org/papers/w15355%5Cnhttp://www.vawnet.org/Assoc_File VAWnet/AR_PsychConsequences.pdf%5Cnhttp://www.ijsrp.org/research-paper-0913/ijsrp-p21116.pdf%5Cnhttp://www.soas.ac.uk/honourcrimes/resources/file55689.pdf.

[35] C, "Exploring Factors Affecting Girls ' Education at Secondary Level : A Case of Karak District , Pakistan," J. Educ. Pract., vol. 6, no. 19, pp. 95–110, 2015, [Online]. Available: www.iiste.org.

[36] Berhane R., "The impact of harmful traditional practices on the girl child," Sierra, no. September, pp. 25–28, 2006, [Online]. Available: http://www.un.org/womenwatch/daw/egm/elim-disc-viol-girlchild/ExpertPapers/EP.4 Raswork.pdf.

[37] T. Prodromou and D. Pratt, "Making sense of stochastic variation and causality in a virtual environment," Technol. Knowl. Learn., vol. 18, no. 3, pp. 121–147, 2013, doi: 10.1007/s10758-013-9210-4.

[38] D. C. Mendoza Cazarez, "Factors Affecting School Dropout and Completion in Mexico: Does Agency Matter?," J. Hum. Dev. Capab., vol. 20, no. 3, pp. 311–328, 2019, doi: 10.1080/19452829.2019.1609917.

[39] Camfed, "Project Reference Number 5101 Tanzania and Zimbabwe Baseline A New Equilibrium for Girls," no. 5101, pp. 1–262, 2014.

[40] J. Muir, "Student Attendance: Is It Important, and What Do Students Think?," Transactions, vol. 6, no. 2, pp. 50–69, 2009, doi: 10.11120/tran.2009.06020050.

[41] T. J. Lee-St. John et al., "The Long-Term Impact of Systemic Student Support in Elementary School: Reducing High School Dropout," AERA Open, vol. 4, no. 4, p. 233285841879908, 2018, doi: 10.1177/2332858418799085.

[42] E. Can, F. Oya Aktas, and I. Tuzun Arpacioglu, "The Reasons of School Dropouts in Higher Education: Babaeski Vocational College Case," Univers. J. Educ. Res., vol. 5, no. 12A, pp. 84–88, 2017, doi: 10.13189/ujer.2017.051313.

[43] K. De Witte, S. Cabus, G. Thyssen, W. Groot, and H. M. Van Den Brink, "A critical review of the literature on school dropout," Educ. Res. Rev., vol. 10, no. October 2017, pp. 13–28, 2013, doi: 10.1016/j.edurev.2013.05.002.

[44] H. Malcolm, V. Wilson, J. Davidson, and S. Kirk, "Absence from School : A study of its causes and effects in seven LEAs," Absence from Sch. A study its causes Eff. seven LEAs, p. 88, 2003.

[45] FCDO, "Every girl goes to school, stays safe, and learns: Five years of global action," pp. 1–966, 2021, [Online]. Available: https://www.brookings.edu/wp-content/uploads/2016/07/todays-challenges-girls-educationv6.pdf.

[46] K. H. Okemwa, "a Research Project Submitted in Partial Fulfillment of the Requirements for the Award of the Degree of Master of Education of the University of Nairobi," 2014.

[47] M. J. Khan and J. Ahmed, "Child education in the time of pandemic: Learning loss and dropout," Child. Youth Serv. Rev., vol. 127, no. October 2020, p. 106065, 2021, doi: 10.1016/j.childyouth.2021.106065.

[48] N. Mduma, K. Kalegele, and D. Machuve, "Machine learning approach for reducing students dropout rates," Int. J. Adv. Comput. Res., vol. 9, no. 42, pp. 156–169, 2019, doi: 10.19101/ijacr.2018.839045.

[49] B. Prenkaj, P. Velardi, G. Stilo, D. Distante, and S. Faralli, "A Survey of Machine Learning Approaches for Student Dropout Prediction in Online Courses," ACM Comput. Surv., vol. 53, no. 3, 2020, doi: 10.1145/3388792.

[50] D. A. Blair, "Facts and Figures," Br. Med. J., vol. 1, no. 5165, p. 62, 1960, doi: 10.1136/bmj.1.5165.62.

[51] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," J. Manag. Inf. Syst., vol. 24, no. 3, pp. 45–77, 2007, doi: 10.2753/MIS0742-1222240302.

[52] A. Mwamatandala and M. S. Muneja, "Effect of School Management on Teachers' Commitment: A Case of Government Secondary Schools in Arusha City, Tanzania," East African J. Educ. Soc. Sci., vol. 1, no. 3, pp. 87–96, 2020, doi: 10.46606/eajess2020v01i03.0046.

The Best Techniques to Deal with Unbalanced Sequential Text Data in Deep Learning

Sumarni Adi[1]*, Awaliyatul Hikmah[2], Bety Wulan Sari[3], Andi Sunyoto[4], Ainul Yaqin[5], Mardhiya Hayaty[6]

Department of Information Systems, Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia[1, 3]
Department of Informatics, Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia[2, 5, 6]
Magister of Informatics Engineering, Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia[4]

Abstract—**Datasets with a balanced distribution of data are often difficult to find in real life. Although various methods have been developed and proven successful using shallow learning algorithms, handling unbalanced classes using a deep learning approach is still limited. Most of these studies only focus on image data using the Convolution Neural Network (CNN) architecture. In this study, we tried to apply several class handling techniques to three datasets of unbalanced text data. Both use a data-level approach with resampling techniques on word vectors and algorithm-level using Weighted Cross-Entropy Loss (WCEL) to handle cases of imbalanced text classification. With Bidirectional Long-Short Term Memory (BiLSTM) architecture. We tested each method using three datasets with different characteristics and levels of imbalance. Based on the experiments that have been carried out, each technique applied has a different performance on each dataset.**

Keywords—Imbalanced text classification; deep learning; resampling technique; weighted cross-entropy loss

I. INTRODUCTION

Datasets with unbalanced class conditions are usual in real life, for example, in the case of fraud detection[1], cancer diagnosis[2], and spam detection[3]. These are challenges for machine learning models to perform classification tasks because samples are not the same in each class. As a result, the classifier may have high accuracy in the majority class. However, the classifier tends to ignore the minority class, so it has poor performance for detecting the minority class[4]. On the other hand, the minority class sometimes has a more important role because it has beneficial information, for example, when diagnosing cancer where patients are in the minority class. If the learning algorithm cannot detect the minority class properly, it can endanger someone's life.

Although various methods for dealing with class imbalance problems have been developed over the last two decades and have proven successful in various domains, most of them still focus on shallow learning algorithms[5]. Several researchers express similar things, among others: [6], [7], and [8], which states that the handling of unbalanced classes in deep learning has not been studied further. Based on a survey conducted by[5], more than 80% of research related to unbalanced classroom problems in deep learning still focuses on the field of computer vision using the Convolution Neural Network (CNN) architecture. Research by[9] as well, using Convolution Neural Network (CNN) to estimate the accuracy of the head pose angle based on deep learning in image recognition.

There are three approaches used to deal with unbalanced classes: data-level methods, algorithm-level methods, and hybrid methods[10]. Data-level methods are carried out by changing the data distribution in each class or resampling to achieve the desired condition. Resampling is done by reducing data from the majority class (under sampling) or adding data to the minority class (oversampling). Although it has been proven to be effective in overcoming the problem of data imbalance based on the survey conducted, on the other hand, under sampling has the potential to eliminate data that has important information. Oversampling can result in the learning model being overfitted so that the performance of the resulting model may not necessarily improve and increase computational effort [11]. The following method is algorithm-level, namely by making direct modifications to the learning algorithm to reduce bias in the majority class. Finally, the hybrid method is a combination of data-level and algorithm-level methods[8].

II. LITERATURE REVIEW

Handling unbalanced classes using a data-level approach is done by[6], [12], and[13]. Study[6] compared the Random Under Sampling (RUS), Random Oversampling (ROS), and Two-Phase Learning methods using multiclass image datasets trained with various CNN architectures. Overall, ROS has the best performance compared to the other two methods, RUS has poor results, and two-phase learning is considered less effective in dealing with class imbalance cases. The two-phase learning method was also proposed by[12] to classify WHOI-Plankton dataset images with a high level of imbalance. Different from[6], study[12] Instead, it concludes that this approach has been proven to be effective in improving the performance of the minority class while maintaining the performance of the majority class. Still using image dataset and CNN architecture, research[13] gets better performance by implementing the ROS method to handle unbalanced classes.

The algorithm-level approach is carried out by[14], [15], and[16]. Study[14] implemented a cost-sensitive CNN to classify various image datasets. This study also compares cost-sensitive with data-level approaches such as SMOTE and RUS. The proposed method is proven to have the most superior performance. Unfortunately, the performance metric used in this study is only accuracy, in which the evaluation method is not appropriate for measuring the performance of learning models with unbalanced classes[5]. Furthermore, research[15] used a cost-sensitive deep neural network (CSDNN) to predict hospital readmission. This study also compares the proposed

---

*Corresponding Author.

method with shallow learning algorithms such as Decision Tree and Support Vector Machine. The proposed method is proven to get better performance. CNN's cost-sensitive approach was also used in this study[16] for time-series classification with unbalanced data. The proposed method provides superior performance compared to the data-level approach.

The research that applies the hybrid method is[17] by combining SMOTE technique and weighted loss function to classify various image datasets with deep neural network architecture. The proposed method can improve the learning algorithm's performance, but on the other hand, the application of SMOTE also produces noisy data.

In this study, we intend to use the Bidirectional Long-Short Term Memory (BiLSTM) architecture to classify text data with various levels of imbalance. Several techniques to overcome class imbalances will be applied and observed how they affect the learning model's performance.

III. RESEARCH METHOD

A. Data Acquisition

We use three datasets in the form of labeled text with varying levels of imbalance. The level of imbalance is using the imbalance ratio (IR) formulated in Equation 1 by dividing the maximum class size against the minimum class[18].

$$IR(\varsigma) = \frac{\max_i \varsigma_i}{\min_j \varsigma_j} \tag{1}$$

Where:

$\max_i \varsigma_i$=maximum class size, $IR(\varsigma)$=imbalance ratio.

The first dataset we use is the Customer Support Tickets Dataset, which contains complaints about problems from users of an application obtained from the Google Play Store review column. The second dataset is the News Category Dataset, containing news headlines from 2012 to 2018 obtained from HuffPost[19]. Then the last dataset is the Drug Review Dataset which contains patient reviews related to drugs from the disease[20]. This study only used a few categories or classes from the three datasets with details, as shown in Table I.

TABLE I. DETAILS OF THE DATASET

Datasets	Instances	Class	IR (ς)
Cust. Support Tickets	5,150	3	4.99
News Category	50.879	3	15.72
Drug Review	57,463	5	8.52

B. Data Preprocessing

This stage includes the case folding process, data cleaning from special characters, normalization, to data representation. The cleaned text data will be converted into numeric digits by utilizing the Tokenizer library on Keras. Then the digits will be converted into word vectors using the word embedding technique with a dimension of 300. The word embedding technique applied is different based on the language contained in the dataset. The first dataset uses the fastText technique because the Indonesian language training model is available. In

contrast, the second and third datasets use the GloVe technique because an English training model is available.

C. Unbalanced Class Handling

There are two approaches that we will use in this research, namely:

1) Data-Level: In simple terms, Fig. 1 shows the resampling technique in this study. Resampling of training data is 80% training and 20% testing. The resampled data is a word vector from the training dataset.
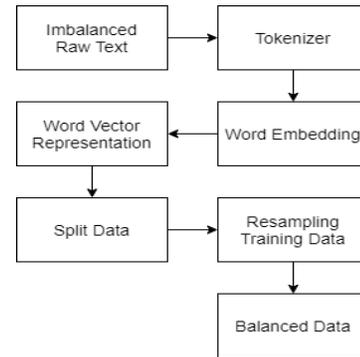
Fig. 1. Data Balancing Step.

We used RUS and Tomek Links (TL) for undersampling. RUS will delete some instances randomly in the majority class until the desired distribution is reached. In comparison, Tomek Links will delete a pair of closest neighbors but belong to a different class[21]. To perform oversampling, we used ROS and Synthetic Minority Oversampling Technique (SMOTE). ROS will randomly duplicate instances of the minority class, while SMOTE will generate synthetic data by linear interpolation between adjacent minority class samples[22]. The distribution of classes before and after resampling is shown in Fig. 2 for Customer Ticket Dataset, Fig. 3 for News Category Dataset, and Fig. 4 for the Drug Review Dataset.
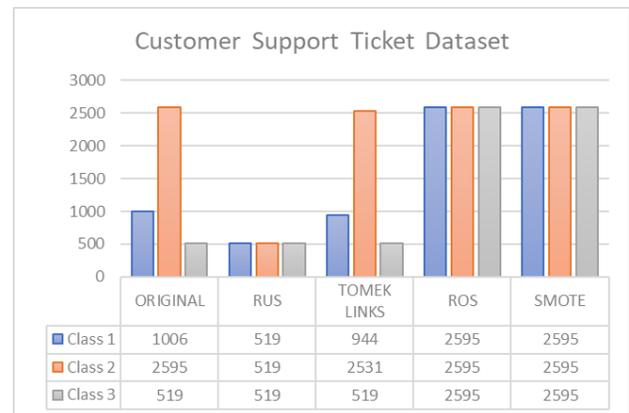
Fig. 2. Class Distribution before and after Resampling on the Customer Support Ticket Dataset.
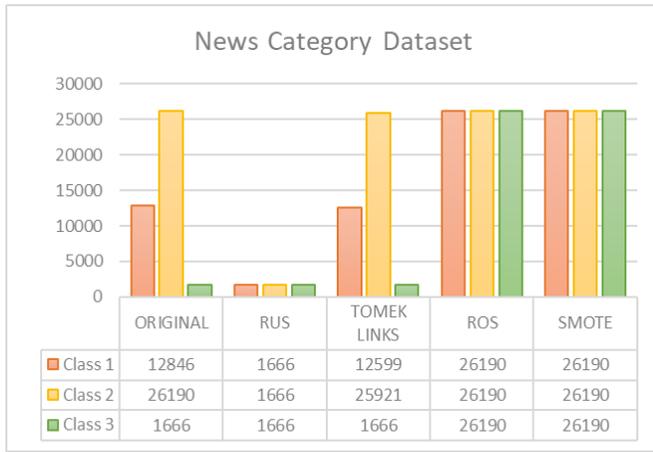
Fig. 3.   Class Distribution before and after Resampling on the News Dataset.
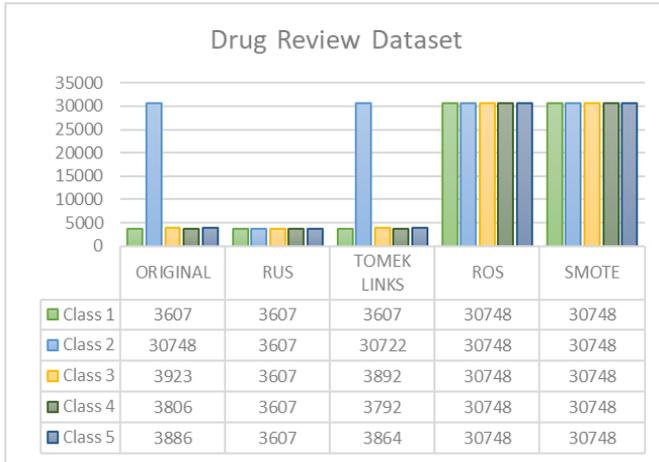
Fig. 4.   Class Distribution before and after Resampling on the Drug Review Dataset.

2) Algorithm-Level: In the algorithm-level approach, we use Weighted Cross-Entropy Loss (WCEL) to increase the sensitivity of the learning model to minority classes. For this reason, significant weight will be applied to the minority class, while the majority class will be given a smaller weight. By applying this weighting, the cross-entropy loss can be formulated in Eq. 2.

$$L(y, o) = - w \sum_{c=1}^{M} y_{o,c} \log(p_{o,c}) \qquad (2)$$

With w is the weight of the calculated class using $w = \frac{Cmax}{Ci}$ where Cmax is the size of the majority class, and Ci is the number of data samples in the class c. Then M Then represent the number of classes, y is a binary (0 or 1) indicator if the class label *c* is the correct classification for the observation o and p is the predicted probability.

D. Build Neural Network

In this study, we use Bidirectional LSTM architecture because, according to[23], the highest accuracy value is when using bidirectional LSTM to perform multiclass text classification tasks. There are 64 neurons in each hidden layer (the more neurons, the more accuracy, but the computation

time will be extended). We trained the network with a batch size of 8 based on[23], that smaller the batch size, the greater the accuracy, and applied a dropout with a probability of 0.6 to reduce the risk of overfitting. The higher the dropout value given, it will reduce overfitting, but if it is too high, it can decrease accuracy, so try n error only to determine the dropout value. The cost function we use is cross-entropy, specifically for applying the algorithm-level method in handling class imbalances. We use different weights for each class through the equations described previously. As a gradient descent optimization algorithm, we use the Adaptive Momentum Estimation (Adam) Optimizer with a learning rate of 0.001.

E. Evaluation

Evaluation with the right measuring tools is needed to compare the performance of several applied approaches. Considering that the minority class has a negligible impact on accuracy, apart from using Overall Accuracy, we also use other performance metrics, namely True Positive Rate (TPR), True Negative Rate (TNR), F1 Score, and Geometric Mean. F1 Score is a combination of Precision which calculates the positive class score classified in a positive class, and Recall, which represents how well the prediction of the positive class to be a single score. Next, by considering sensitivity (another term for Recall) and Specificity, which represents how well the prediction of the negative class is, we use the Geometric Mean (G-Mean) to combine the two into a single score. Since the deep learning algorithm is stochastic, we train five times for each method, then take the average Score of the five training results. The equations for calculating True Positive Rate (TPR) / Sensitivity / Recall / Hit Rate, True Negative Rate (TNR) / Specificity / Selectivity, Overall Accuracy, F1 Score, and G-Mean are as follows.

$$TPR = \frac{TP}{TP+FN} \qquad (3)$$

$$TNR = \frac{TN}{TN+FP} \qquad (4)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (5)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} = 2 \times \frac{\frac{TP}{TP+FP} \times \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \qquad (6)$$

$$G\text{-Mean} = \sqrt{Sensitivity \times Specificity} = = \sqrt{\frac{TP}{TP+FN} \times \frac{TN}{TN+FP}} \qquad (7)$$

IV.   RESULTS AND DISCUSSION

The results of this study are summarized in Table II. Based on our results, no method consistently provides the most superior performance using three predefined benchmarks. In the Customer Support Ticket Dataset, the highest accuracy was obtained without applying any method, while the highest F1 Score and G-Mean were obtained using WCEL. Furthermore, the highest accuracy was obtained using TL in the News Dataset, while the highest F1 Score and G-Mean were obtained using ROS. Finally, in the Drug Review Dataset, the highest accuracy and F1 Score were obtained without using any method, while the highest G Mean was obtained using WCEL.

TABLE II. COMPARISON OF THE PERFORMANCE IN EACH METHOD

Method	Cust. Support Ticket			News Category			Drug Review		
	Overall Accuracy	F1 Score (Average)	G-Mean (Average)	Overall Accuracy	F1 Score (Average)	G-Mean (Average)	Overall Accuracy	F1 Score (Average)	G-Mean (Average)
None	**0.836**	0.789	0.834	0.892	0.755	0.795	**0.930**	**0.837**	0.904
RUS	0.700	0.656	0.783	0.752	0.658	0.811	0.836	0.672	0.803
TL	0.796	0.722	0.767	**0.900**	0.787	0.825	0.878	0.697	0.801
ROS	0.826	0.784	0.828	0.892	**0.814**	**0.872**	0.916	0.832	0.904
SMOTE	0.810	0.760	0.799	0.896	0.803	0.859	0.904	0.806	0.886
WCEL	0.834	**0.798**	**0.867**	0.864	0.755	0.867	0.926	0.834	**0.906**

Table III shows the comparison of class-level performance for each method for the Customer Support Ticket Dataset. There are three classes/categories: Class 1: Account, Class 2: Customer Service, and Class 3: Transaction. Table IV for the News Category Dataset shows three classes/categories, namely Class 1: Entertainment, Class 2: Politics, Class 3: Tech. Table V the Drug Review Dataset shows five classes/categories, namely Class 1: ADHD, Class 2: Birth Control, Class 3: Tech, Class 4: Insomnia, Class 5: Weight Loss.

TABLE III. COMPARISON OF TPR, TNR, F1 SCORE, AND G-MEAN IN EACH METHOD ON THE NEWS CATEGORY DATASET

Method	True Positive Rate (TPR)			True Negative Rate (TNR)			F1 Score			G-Mean		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
None	0.746	0.892	0.722	**0.934**	0.786	0.958	0.770	0.882	0.714	0.834	0.838	0.830
RUS	0.708	0.648	**0.828**	0.810	0.866	0.864	0.618	0.752	0.598	0.756	0.750	0.844
TL	0.656	**0.896**	0.562	0.920	0.698	0.964	0.686	0.862	0.618	0.776	0.790	0.736
ROS	0.720	0.890	0.740	0.926	0.772	0.962	0.740	0.876	**0.736**	0.814	0.828	0.842
SMOTE	0.710	0.882	0.640	0.914	0.732	**0.972**	0.722	0.864	0.694	0.806	0.804	0.786
WCEL	**0.824**	0.848	0.806	0.908	**0.878**	0.942	**0.780**	**0.884**	0.730	**0.864**	**0.864**	**0.872**

In the Customer Support Ticket Dataset, WCEL has the most superior performance except for the F1 Score in Class 3. However, WCEL's performance is consistently superior to the original data or without applying any method. Although the accuracy produced by WCEL is still less than the original data, the resulting difference is not too far, which is 0.2%.

TABLE IV. COMPARISON OF TRUE POSITIVE RATE (TPR) AND TRUE NEGATVE RATE (TNR) IN EACH METHOD ON THE CUSTOMER TICKET DATASET

Method	True Positive Rate (TPR)			True Negative Rate (TNR)			F1 Score			G-Mean		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
None	0.870	0.934	0.360	0.926	0.848	**0.998**	0.856	0.924	0.484	0.898	0.890	0.598
RUS	0.788	0.736	**0.752**	0.810	0.876	0.922	0.718	0.816	0.440	0.798	0.804	**0.832**
TL	0.856	**0.948**	0.474	**0.944**	0.846	0.994	**0.864**	**0.930**	0.566	0.898	0.894	0.682
ROS	**0.886**	0.910	0.670	0.918	0.892	0.982	0.862	0.922	**0.658**	**0.904**	**0.900**	0.812
SMOTE	0.856	0.930	0.622	0.942	0.862	0.986	**0.864**	0.928	0.616	0.900	0.896	0.780
WCEL	0.864	0.874	0.718	0.918	**0.906**	0.954	0.844	0.908	0.514	0.890	0.886	0.826

In the News Dataset, SMOTE has consistently superior performance over the original data, although it is not superior to other methods. On the other hand, the TL performance is superior to the original data except for the G-Mean Score in Class 2. TL even has the highest accuracy compared to other methods.

In the Drug Review Dataset, the original dataset tends to be superior to the unbalanced class handling method. Although ROS has a reasonably high contribution to Class 4 and Class 5, ROS lowers the performance of Class 1, which is a minority class. In the case of this dataset, it seems that the minority class already has a pretty good performance. In theory, the learning algorithm will have difficulty detecting the minority class.

Based on the results we got, no method has the most superior performance over other methods. Overall, RUS had the worst performance, despite getting the highest G-Mean Score in Class 3 in the News Dataset. ROS and WCEL tend to improve the performance of the minority class, but at the same time, sometimes, these methods also reduce the performance of the majority class. Meanwhile, TL and SMOTE performed exceptionally well on the News Dataset, but not on the other two datasets that we used in this study.

Although the data-level method approach contributes to improving the performance of the minority class in some cases, the weakness of this approach is that it takes a long time to resample, except for RUS and ROS. The comparison of the time required for each data-level method is shown in Fig. 5. SMOTE and TL, which performed well on the News Dataset, took resampling time of more than 1.6 hours for SMOTE and more than 3 hours for TL. With a performance increase of <1%, this seems less applicable to the case of big data, where the amount of data available will be much larger, so the time required for resampling will also be longer.

TABLE V.   COMPARISON OF TPR, TNR, F1 SCORE, AND G-MEAN IN EACH METHOD ON THE DRUG REVIEW DATASET

Method	True Positive Rate (TPR)					True Negative Rate (TNR)					F1 Score					G-Mean				
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 1	Class 2	Class 3	Class 4	Class 5	Class 1	Class 2	Class 3	Class 4	Class 5	Class 1	Class 2	Class 3	Class 4	Class 5
None	**0.952**	**0.990**	**0.962**	0.668	0.618	**1.000**	0.990	**0.992**	0.962	**0.970**	**0.956**	**0.996**	**0.950**	0.646	0.636	**0.976**	**0.990**	**0.978**	0.800	0.774
RUS	0.860	0.912	0.896	0.376	0.602	0.972	0.980	0.980	0.960	0.918	0.790	0.950	0.848	0.336	0.436	0.914	0.944	0.934	0.526	0.698
TL	0.830	0.984	0.876	0.252	0.706	0.988	0.964	0.988	**0.976**	0.928	0.840	0.982	0.872	0.254	0.538	0.906	0.970	0.930	0.416	0.782
ROS	0.904	0.974	0.886	0.668	**0.772**	0.990	0.978	0.996	**0.976**	0.962	0.886	0.980	0.896	**0.690**	**0.706**	0.946	0.972	0.936	0.806	**0.862**
SMOTE	0.874	0.968	0.888	**0.692**	0.628	0.988	0.964	0.988	0.962	0.968	0.874	0.976	0.886	0.650	0.642	0.930	0.968	0.936	**0.814**	0.780
WCEL	0.950	0.984	0.960	0.624	0.702	0.994	**0.994**	0.992	0.970	0.964	0.936	0.990	0.944	0.632	0.668	0.972	**0.990**	0.976	0.774	0.816
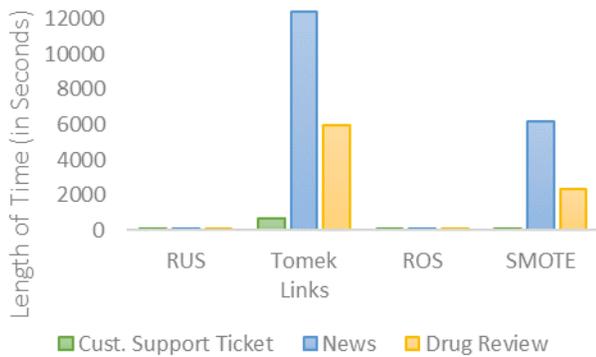
Fig. 5.   Comparison of the Length of Time for Resampling.

V.   CONCLUSION

This study discusses several testing methods for handling unbalanced classes, including RUS, TL, ROS, and SMOTE for data-level methods and WCEL for algorithm-level methods. We used three labeled text datasets with varying amounts of data and levels of balance. The neural network architecture we use is Bidirectional LSTM. We tested five times for each method. From the research results obtained, no method has the most superior performance because the quality of the data is also very influential on the learning model's performance. Both data-level and algorithm-level approaches can, in some cases, improve minority class performance. However, at the same time, The algorithm-level used in this study sometimes also reduces the performance of the majority class. Meanwhile, in the data-level approach, the resulting increase does not seem proportional to the time required for resampling.

Further research on the handling of unbalanced classes in deep learning is needed to produce consistent performance and be implemented effectively and efficiently.

ACKNOWLEDGMENT

REFERENCES

[1] S. Makki, Z. Assaghir, Y. Taher, R. Haque, MS Hacid, and H. Zeineddine, "An Experimental Study With Imbalanced Classification Approaches for Credit Card Fraud Detection," IEEE Access, vol. 7, pp. 93010–93022, 2019, doi:10.1109/ACCESS.2019.2927266.

[2] S. Fotouhi, S. Asadi, and MW Kattan, "A comprehensive data level analysis for cancer diagnosis on imbalanced data," J. Biomed. information., vol. 90, no. December 2018, p. 103089, 2019, doi:10.1016/j.jbi.2018.12.003.

[3] P. Ratadiya and R. Moorthy, "Spam filtering on forums: A synthetic oversampling based approach for imbalanced data classification," arXiv, 2019.

[4] T. G. Pratama, R. Hartanto, and N. A. Setiawan, "Machine learning algorithm for improving performance on 3 AQ-screening classification," Commun. Sci. Technol., vol. 4, no. 2, pp. 44–49, 2019, doi: 10.21924/cst.4.2.2019.118.

[5] JM Johnson and TM Khoshgoftaar, "Survey on deep learning with class imbalance," J. Big Data, vol. 6, no. 1, 2019, doi:10.1186/s40537-019-0192-5.

[6] M. Buda, A. Maki, and MA Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," Neural Networks, vol. 106, pp. 249–259, 2018, doi:10.1016/j.neunet.2018.07.011.

[7] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and PJ Kennedy, "Training deep neural networks on imbalanced data sets," Proc. int. jt. Conf. Neural Networks, vol. 2016-Octob, pp. 4368–4374, 2016, doi:10.109/IJCNN.2016.7727770.

[8] S. Pouyanfar et al., "Dynamic Sampling in Convolutional Neural Networks for Imbalanced Data Classification," Proc. - IEEE 1st Conf. Multimed. inf. Process. Retrieval, MIPR 2018, pp. 112–117, 2018, doi:10.109/MIPR.2018.00027.

[9] K. Arai, A. Yamashita, and H. Okumura, "Head Position and Pose Model and Method for Head Pose Angle Estimation based on Convolution Neural Network" International Journal of Advanced Computer Science and Applications(IJACSA), vol. 12, no. 10, pp. 42–49, 2021. Available: http://dx.doi.org/10.14569/IJACSA.2021.0121006

[10] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," prog. Arti. Intell., vol. 5, no. 4, pp. 221–232, 2016, doi:10.1007/s13748-016-0094-0.

[11] P. Branco, L. Torgo, and R. Ribeiro, "A Survey of Predictive Modeling under Imbalanced Distributions," pp. 1–48, 2015, [Online]. Available: http://arxiv.org/abs/1505.01658.

[12] H. Lee, M. Park, and J. Kim, "Plankton classification on imbalanced large scale database via convolutional neural networks with transfer

learning," in 2016 IEEE International Conference on Image Processing (ICIP), Sept. 2016, pp. 3713–3717, doi:10.109/ICIP.2016.7533053.

[13] P. Hensman and D. Masko, "The Impact of Imbalanced Training Data for Convolutional Neural Networks," Ph.D., 2015, [Online]. Available: https://www.kth.se/social/files/588617ebf2765401cfcc478c/PHensmanD Masko_dkand15.pdf.

[14] SH Khan, M. Hayat, M. Bennamoun, FA Sohel, and R. Togneri, "Cost-sensitive learning of deep feature representations from imbalanced data," IEEE Trans. Neural Networks Learn. Syst., vol. 29, no. 8, pp. 3573–3587, 2018, doi:10.1109/TNNLS.2017.2732482.

[15] H. Wang, Z. Cui, Y. Chen, M. Avidan, A. Ben Abdallah, and A. Kronzer, "Predicting Hospital Readmission via Cost-Sensitive Deep Learning," IEEE/ACM Trans. Comput. Biol. Bioinformatics., vol. 15, no. 6, pp. 1968–1978, 2018, doi:10.109/TCBB.2018.2827029.

[16] Y. Geng and X. Luo, "Cost-sensitive convolution-based neural networks for imbalanced time-series classification," arXiv, 2018.

[17] R. Harliman and K. Uchida, "Data- and algorithm-hybrid approach for imbalanced data problems in deep neural networks," int. J. Mach. Learn. Comput., vol. 8, no. 3, pp. 208–213, 2018, doi:10.18178/ijmlc.2018.8.3.689.

[18] J. Ortigosa-Hernández, I. Inza, and JA Lozano, "Measuring the class-balance extent of multi-class problems," Pattern Recognition. Lett., vol. 98, pp. 32–38, 2017, doi:10.1016/j.patrec.2017.08.002.

[19] R. Misra, "News Category Dataset," 2018. https://www.kaggle.com/rmisra/news-category-dataset.

[20] F. Gräßer, S. Kallumadi, H. Malberg, and S. Zaunseder, "Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning," in Proceedings of the 2018 International Conference on Digital Health, Apr. 2018, pp. 121–125, doi:10.1145/3194658.3194677.

[21] R. M. Pereira, Y. M. G. Costa, and C. N. Silla, "MLTL: A multi-label approach for the Tomek Link undersampling algorithm: MLTL: The Multi-Label Tomek Link," Neurocomputing, vol. 383, pp. 95–105, 2020, Available: https://doi.org/10.1016/j.neucom.2019.11.076.

[22] F. Charte, A. J. Rivera, M. J. Del Jesus, and F. Herrera, "MLSMOTE: Approaching imbalanced multilabel learning through synthetic instance generation," Knowledge-Based Syst., vol. 89, pp. 385–397, 2015, [Online]. Available: https://doi.org/10.1016/j.knosys.2015.07.019.

[23] A. Hikmah, S. Adi, and M. Sulistiyono, "The Best Parameter Tuning on RNN Layers for Indonesian Text Classification," in the 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2020, pp. 94–99.

A New Framework for Accelerating Magnetic Resonance Imaging using Deep Learning along with HPC Parallel Computing Technologies

Hani Moaiteq Aljahdali

Faculty of Computing and Information Technology
King Abdulaziz University, Jeddah, Saudi Arabia

Abstract—**MRI (magnetic resource imaging) has played a vital role in emerging technologies because of its non-invasion principle. MR equipment is traditional procedure being used for imaging biological structures. In medical domain, MRI is a most important tool being used for staging in clinical diagnosis that has ability to furnish rich physiological and functional information and radiation and non-ionizing nature. However, MRI is highly demanding in several clinical applications. In this paper, we have proposed a novel deep learning based method that accelerates MRI using a huge number of MR images. In proposed method, we used supervised learning approach that performs network training of given datasets. It determines the required network parameters that afford an accurate reconstruction of under-sampled acquisitions. We also designed offline based neural network (NN) that was trained to discover the relationship between MR images and K-space. All the experiments were performed over advanced NVIDIA GPUs (Tesla k80 and GTX Titan) based computers. It was observed that the proposed model outperformed and attained <0.2% error rate. With our best knowledge, our method is the best approach that can be considered as leading model in future.**

Keywords—Magnetic resonance imaging (MRI); segmentation; classification; acceleration; deep learning

I. INTRODUCTION

Magnetic resonance imaging (MRI) is a fundamental instrument for clinical determination, illness and furthermore in clinical exploration. Because of its solid ability they give rich useful data and non-radiation and non-ionizing nature [1]. MRI is a non-obtrusive imaging approach for acquiring organic data with high spatial goal. Compared to X-ray computed tomography [9], MRI scan times are longer due to the use of a data acquisition system [1], which is sampled by a Fourier domain [15], also known as k-space. The suggestion for this is in parallel imaging and echo imaging [5], to reduce the time required for MRI scanning. However, in clinical diagnostic procedures, imaging speed should be improved at deterioration in order to minimize active movement and burden placed on patients. MRI is a technique used to detect an error or disease in the brain, heart, knees, etc. MRI deals only with X-rays, but is completely different in the medical and biological fields of science. Every MRI (brain, heart) has same technique but different method. A computer, GPU and Graphics plays [5], an important role in this biological technique. The most important part of this technique is algorithm which tells how the machine behaves and which command applies to which time.

Deep Learning (DL) is the foundation of this strategy since basically everything done by profound learning. DL is a part of AI dependent on the utilization of numerous layers to learn information portrayals [1], and can be applied to both regulated and unaided learning. These various layers permit the machine to get familiar with different level highlights of information to accomplish its ideal capacity.

Deep learning techniques with validated neural networks have been recently incorporated into clinical imaging, where successfully used to demonstrate the effects of dividing parts of the brain, brain tissue, cardiac structures, bone and cartilage on MR images. The purpose of this study has been to develop and test the feasibility of in-depth study methods in MR imaging. The Harmonic phase (HARP) algorithm is a medical imaging image detector that is able to extract and process motion information from a magnetic resonance image (MRI) sequence. CNN based Training and Classification [3, 4, 24- 26], MRI Brain Imaging. These days, CNN plays a major role in the fine-tuning and testing of brain tumors present in Magnetic Resonance Imaging (MRI) imaging. Graphics is also important in MRI to improve image quality.

During the past few years, compressed sensing and later deep learning have remained in forefront of accelerated MRI gaining, leading to large and important improvements in terms of the time efficiency of image acquisition (time efficiency plays an important role to reduce the cost) with a hardly obvious reduction in image quality. The purpose of the study focused on the stage of image restoration and resolution [3, 14], trying to recover high quality MRI scans from reduced sets of their measurements available through partial sampling of the k-space. At the same time, recent studies have also attempted to directly increase the geometry of k-space trajectories, indicating further improvements. This method also improves image quality by reducing the movement. There is another technique like MRI called Fast Magnetic Resonance [2], Imaging (MRI) which is much needed in many medical applications. This method also improves image quality by reducing image movement. The method can reconstruct each image in 0.22ms-0.37ms [2], for real-time applications. Once the desired resolution has been selected the minimum scan time is determined by the need for sufficient data to meet the NY-Quist-Shannon [16], sampling procedure. Compression sensing (CS) vision is fully compatible with MRI scanning sequence design with very little data needed for image reconstruction. Reconstructing and reconstructing MRI images in sample data at high speed in the data acquisition process leads to deep

network-based learning. MRI is associated with the speed of detection that slowly detects the sample information which is not directly collected in the image area, but rather in the k-space because the term contains general spatial information. The speed where k-space can be detected is limited due to physical and hardware constraints. The fast MRI approach is under k-space-for-sample, which can provide a rate proportional to the under-sampling ratio. The challenge for rapid MRI is to find an algorithm that can reconstruct the image from under-sampled data [15].

A. Challenges

Some of the challenges faced are given below:

1) GPU Computation: Computational acceleration in the graphics analysis unit (GPU) can create high-resolution imaging of magnetic resonance imaging (MRI) in clinical settings, thereby improving the quality of MR images over a wide range. Because of the huge size of the dataset, it is unimaginable to expect to stack patches for all subjects in PC memory without a moment's delay. To quantify this, a Python generator is required to pick and deliver fixes independently for each group.

2) Connectivity: Useful network is by and large arranged by assuming chronicles that the transient relationship between two cerebrum districts is driven by low recurrence motions. In this work, we utilized example free planning to distinguish possibilities in the hubs of the mind organization, which are regularly examined with useful availability MRI, and these occasions are displayed to happen in short and long windows that add to estimated network availability [18].

3) Sensitivity: We further investigated the case wherein the sign can show up in one of a few areas and found that CNN spatial affectability relates to IO. Notwithstanding, CNN affectability was far underneath ideal in identifying some complicated surface examples. These estimations recommend that CNNs might have exceptionally huge execution contrasts when distinguishing the presence of spatial examples. The case wherein the sign can show up in one of a few areas and found that CNN spatial affectability compares to IO [27-33]. Nonetheless, CNN affectability was far underneath ideal in distinguishing some complicated surface examples. These estimations recommend that CNNs might have exceptionally huge execution contrasts when distinguishing the presence of spatial examples. These distinctions might majorly affect the exhibition of imaging frameworks intended to recognize low spatial examples [17].

4) Noise Reduction: Noise reduction due to lack of locally changing Russian noise in MR images. With the advent of intensive learning methods, some pre-processor steps have become less important to the performance of previous sections. For example, curvature change and quantitative-based power standardization are often effectively suppressed by z-score estimates, while serious practice-based separation shows another work improvement when applying standardization prior to the process [19].

5) Intensity Normalization: The normalization of noise is toward planning the power of all images on an ideal or reference scale, for example, somewhere in the range of 0 and 4095. Regarding the de-learning system, processing the z-score where one makes a cut is the division of pixels by the normal picture force and the standard deviation of power from all the pixels in a picture, another well-known speculation process [19].

6) Collecting Multiple Types of Data: Another test is that the events for which we have sufficient information, is typically only one sort of information, for example, picture information. In any case, just seeing pictures can tackle specific clinical issues. In case it is an issue of distinguishing an infection from a picture that identifies a disease type characterized from a pathology picture, or an irregularity out of the radiology picture, then, at that point, a solitary information type is likely adequate. In any case, numerous issues require additional information from only one mode. Specifically issues like a clinical forecast; more reference is required with regards to patients given by clinical record information and pathology [20].

7) Scanning Cost: Fast magnetic resonance imaging (MRI) has become very popular for some clinical applications, thereby reducing test costs and improving patient experience. The quality of the film can be improved by minimizing antiquity and time differences. Presumably, when selecting the image view and the appropriate target, the initial test time is usually determined by a precondition to obtain sufficient raw information in accordance with the Nyquist–Shannon test criteria [2].

8) Image Reconstruction: The process of converting the acquired raw data into image is called image reconstruction and on modern MRI devices, it is finished by committed reconstruction software that fills in as a magnet encoding gadget with inclination and radio recurrence equipment. Radio recurrence radiation is transmitted to the patient, where it animates the charge of the tissue and discharges radio recurrence signals from the tissues. The radio recurrence loop is utilized to accomplish the superposition (all out) of all tissue signals [21].

9) Point Spread Function: A significant component of the picture is the point show work, which shows how much neighbors are seeping from one another because of inflexible spatial goal and different impacts. PSF is a picture gotten by replicating a solitary point object. With the fitting point engendering capacity, the picture of the point source must be a picture with a sign of a solitary pixel. Be that as it may, explicit point publicity exercises in MRI digress altogether from this perfect structure. [21].

10) Speed of Detection: MRI is associated with slower detection speed, due to data samples it can be collected directly in the image area, but instead in the k space, which contains standard location data. Here the k-space and image and image are related to deviations, the setting of one domain is limited to another. Samples of raw data are sequentially

available in the k-space and the speed at which the limited k-limited space can be obtained due to visual and hardware constraints. When taking the necessary field view adjustments and MRI predictions, the green details k-space we need to find traditionally are determined by Nyquist's procedure - Shannon's procedure [39].

Leading to existing accelerating MRI, a list of challenges is presented in Table I as follows.

TABLE I.   CHALLENGES TOWARD ACCELERATING MAGNETIC RESONANCE IMAGING

#	Challenges	Description
1	GPU	System required to form quality image to detect disease.
2	Connectivity	Connectivity MRI, and have been shown that these events occur over short and long windows Contribute to measured network connectivity.
3	Sensitivity	CNN sensitivity was far below optimal in detecting some complex texture patterns.
4	Noise Reduction	If noise reduction is not lesser the noise observed in MRI image. However, improvement appears when applying normalization before the intensive learning-based segmentation process.
5	Intensity Normalization	Normalization of power is the way toward planning the force of all pictures to a normal/standard or reference scale.
6	Collecting Multiple Types	Another challenge is that if you have got enough data, it is usually just one type of data, such as image data
7	Scanning Cost	This can likewise conceivably build the picture quality by lessening the movement ancient rarities and difference waste of time
8	Image Reconstruction	Reconstruction networks for multi-coil data, By expanding the deep cascade of CNNs and leveraging data consistency Layer.
9	Point Spread Function	Feature of an image is the point dispersion function.
10	Speed of Detection	Desired field-of-see and spatial goals of MRI pictures are resolved, the k-space crude or raw information.

II. LITERATURE WORK

In this article, we investigated twenty relevant approaches from different research articles that are presented in this section.

Wang, Shanshan, et al., [1], proposed a deep learning method for accelerating magnetic resonance imaging (MRI). MRI is an essential tool in medical diagnostics, diagnostics and in clinical research because of its rich quality and robust dynamic properties it provides anatomical and functional information. Trying a deep learning process at the highest level of data has shown an explosive popularity with many layers of research on the availability of powerful GPUs. It also used the neural network (CNN) to find objects. The resulting off-line neural network was also designed and trained to detect the map relations between the MR images obtained from the full zero and k-space data. Experimental results in MR data have the advantage of efficient and accurate thinking. The CNN network studied the end-to-end mapping between MR images with a sample and zero saturation.

Warach, Steve, et al. [2], the author proposes a Fast Magnetic Resonance Imaging (MRI) is in high demand in many clinical applications to reduce scanning costs and improve the patient's experience and knowledge. This method improves image quality by reducing movement. Once the required resolution is selected, the minimum scan time is determined according to the need for sufficient data in accordance with the sample criteria. Sample data is not stored directly in image space, but is associated with slower acquisition speeds than K-space, because the term contains spatial-frequency information. K-space travel speed is limited due to physical and hardware limitations. The fastest MRI procedure is the under-sample K-space, which provides the acceleration rate with the under-sample ratio. The challenge for Fast MRI is to find an algorithm that can reconstruct an image from under-sampled data and change the name.

Authors said about the accelerated parallel MR image reconstruction in [3], the paper proposes a reconstruction networks for multi-coil data by extending deep cascade of CNNs. There are two articles which one is POCSENSE and the other one is calibration-less. The networks are the extensions of CNN deep cascades (DC-CNN), where the sub-networks and data consistency layers are between them. For parallel imaging, the data layer can be extending two network variants, which can be computed by using such type of algorithms. Authors presented a novel method for studying conflicting transitions from one MRI to another in [5]. Because MRI images are available for diagnostic purposes, when this happens is distributed, this information is zero. Although the data used here include only healthy subjects, future work will include pathological images of tumors. The main strength of magnetic resonance imaging (MRI) is the ability to measure different tissue differences. To evaluate our results, we compare the number of different network depths, input features, and training topics [34-38].

According to [6], Magnetic resonance is one of the most important diagnostic and therapeutic indicators, as well as the degree of physical and physical impairment of magnetic resonance scan acqustism when MRI reconstructs high-resolution imaging based on local K-cell data, the use of existing network data in which Kranselskii - Mann iteration for K-space translate algorithm is used to make a tread pattern for detailed study of low frequency sampling and Gaussian random sample and similarity. The proposed provided the reconstruction results among other CS and parallel imaging algorithm comparison. The KM method uses k-space scaling, which also improves the reconstruction efficiency.

In [7], authors said about the network acquired through read-only transmission using tens of images in the test domain which achieves almost the same performance as the network specially trained for thousands of test images. The networks were well-formed MR images with various experimental domains. Differential diagnosis of soft tissue on MRI has made it a common practice in many diagnostic applications. Due to the diversity and features of natural and MR images, the use of the Image Net test network domain is scheduled to be the end of tens of images. Network training is a distortion of the supervised learning process aimed at obtaining a set of network parameters that reach reconstruction under pre-acquisition. In

which deep architecture is used with many subnetworks. The subnetwork consists of CC and CNN blocks, and each block follows the DC block and in which each CCN block is trained sequentially so that they are recombined to synthesize images of multi-coil nature from ImageNet Can be synthesized from ImageNet under Zero-filled Fourier reconstruction [40-45].

Author investigates about the deep learning electrical properties tomography (EPT) [10], for application to various simulated and in-vivo datasets including pathology to obtain quantitative brain conductivity maps. However, from the results this concludes that networks can be restricted to data that have anatomic geometries and artifacts that are very close to those present in the training data. This emphasizes the importance of completing training information relating to the geometry of the brain and tissue components. In addition, in addition to the accuracy of the standard EPT method, the training dataset must include many types of in-vivo techniques as well.

According to Sandino CM, Dixit N, Cheng JY, and Vasanawala SS [8], aims at the deep neural network of fast dynamic magnetic resonance imaging, which can be accelerated using integrated methods of building architecture, which allows the detection of image quality under sample data. Unfortunately, CS reconstruction takes hours between dynamic MRI scan and image availability for diagnosis in this work. CNN improve rapid reconstruction. Dynamic magnetic resonance imaging the MRI organs, such as the heart and brain, must be continuously scanned over a long period of time to obtain a series of images that illustrate the anatomy and magnitude of movement. Longer scan times accelerate the exploited narrow sensory reconstruction schemes [48-51].

Repetition of dynamic MRI acquisition in space and time can be used to achieve high spatio-temporal resolution while interpreting data.

In [9], authors raise Dynamic cardiac MRI obtained using similar assumptions, in which Fourier data is obtained by multiple recipients in a variety of locations. For the purpose of CNN reconstruction, the Fourier raw data is converted into a photographic background because it is a natural environment above the photographic background because it is a more natural environment for layers of spatial exploitation spaces. It also gives CNN a hot start, as you don't have to study Fourier to convert images. This construction occurs very quickly when raw data is sampled from the Cartesian grid in the Fourier space, as is the case in the examples shown in it. We also wanted to simplify CNN installation by mixing data from multiple accepted pounds into a single installation image. This is done using the total number of non-reconstructed image groups from each recipient.

Qin, Chen, et al., [10], discussed about Neural networks have recently received interest in reconstructing MR detection under the sample. Network performance should be best done through training and testing of data from the same domain. The purpose of this study was to introduce a transfer method to solve the problem of data shortages in training complex high-speed MRI networks [46-48]. Neural networks are trained in thousands of samples from public data of natural images or MR images of the brain. The network was then configured to use only dozens of MR brain images in a separate test domain. A comparative analysis of existing techniques has been presented in Table II as follows.

TABLE II. COMPARATIVE ANALYSIS OF EXISTING DEEP LEARNING STATE-OF-THE-ART TECHNIQUES USED TO ACCELERATE MRI

Ref.	Proposed Model	Characteristics	Used Technology	Limitations
[11]	U-Net	The Refinement Module (R) puts multi-kilogram coil data into a single image, enters U-Net, and reverts back to multi-coil data.	• Sensitivity Map Estimation (SME)	Variational network with the shallow CNNs replaced with U-Nets (VNU
[12]	E2E-VN	The proposed end-to-end (E2E-VN) network model is proposed. E2E-VN demonstrates the importance of reading sensory maps as part of a network.	• SSIM	Adjustment of parameters
[13]	VGG 19	The idea of using small filter filters is popular and therefore deep networks and deep training networks using pre-configured fixed versions	• PROSTATEx	noncancerous tissues with multi-parametric MRI using data
[14]	DenseNet	This energizes include reuse and brings down the quantity of boundaries for a given profundity. Thick Nets are in this way especially appropriate for littler informational indexes.	• CPU,CNN	Expands on the thoughts of ResNet, yet as opposed to including the actuations created by one layer to future layers, they are essentially linked organised.
[15]	ResNet (18,34,50,101, 152, Upto 1001)	Increase the accuracy to maximum	• CPU, but if we go deeper layers up to 1000 then need GPU & TPU	Computational cost increase rapidly
[16-18]	Sigma Net	enhance the learning ability of the CNN by making it deeper	• NVIDIA GTX 580 GPUs	With an increase in depth system is overfitting.
[20-21]	Alex Net	enhance the learning capacity of the CNN by making it deeper	• NVIDIA GTX 580 GPUs	With an increase in depth system is overfitting
[22]	Res Next	Expands on ResNet and Google Net by utilizing beginning modules between skip associations	• Google Net	Quadratic time increases
[23]	NASNet	The control network (standard neural network) proposes a construction that aims to perform at a particular level of work, and by test and error learns to propose an improved models.	• AutoML	NASNet was based on Cifar-10, and has diffident computer requirements, but quiet has a very good facial features

III. PROPOSED METHODOLOGY

In this section we have presented the proposed architecture that shortly explains the algorithm that how the machine works efficient and fast. Our article shortly tells how the imaging works in MRI with help of GPU [48-51]. In this field, if you do have graphics, and not reduce the motion of the imaging, without these things your machine can't work efficiently and it becomes useless. If you have an algorithm but based on the big data and the if you have an another algorithm the work efficiently but in small amount data that reason of time efficient algorithm because both work same but one takes time and other one is easy to implement and time efficient, that's the reason to improve our technology. In this architecture, in offline training use simulate calibration data that is used in MRI for the measurements then use learn dataset parameters and simulate calibration data that's use database and send to data in data base to learn the data that is also a raw MRI data and use model interference that is helpful for images reconstruction and used for MRI scanner that extracts calibration data and to send data to model inference and then image reconstruction; then used graphics for image output to clear the image and improve the patient experience and scanning cost. In our model that is going to propose, adding some important functionality that improves the performance of algorithm and proposed architecture. This section provides material, the source of the brain MR image dataset, and the algorithm used to perform brain MR tissue segmentation. Fig. 1 presents a block diagram of our proposed model. The total number of pieces for all channels is 15, which leads to 200 pieces or 9 slices with a total of 135 images per patient, 1 mm inter-slice gap and 0.78 mm size tone. 0.78 mm × 0.5 mm. The proposed method is applied to real datasets containing 512 × 512 pixel size brain MR images and converted to grayscale before further processing. The following sections discuss algorithm implementation.

In the article, architecture that we compared to our architecture, this architecture works according to their own algorithm. In this architectre, there is a training phase and

Recon phase. Firstly the machine input CT image and in training phase input the reference mask if the reference or related image or mask is loss the data is in CAE and train the data filtering the image and pseudo CT and PET raw data that is not useful but sometimes that is useful is some-terms but this algorithm is not useful because some errors and this architecture can't reconstruction about the image, time efficiency, scanning cost and reduce motion of image. This Architecture is a stack of two-dimensional axial images. All composite input images were optimized using standard calibration and plotted with a 340 x 340 matrix size using pre-translation as the default Auto encoder (CAE) input. The encoder weight and the disorder wheels weight when initialized using the initialization program are defined and updated using the sliding scale domain with a constant rate of 0.01 to 0.9. The intermittent CAE network estimates growth tissue marks and analyzes them to the file veils produced by CT information. CT reference subtleties affirm that the organization comprehends the connection between MR pictures and reference names. The framework was arranged using multiclass cross-entropy setback as an objective work, where the not really settled in a more modest than ordinary bunch of four pictures in each emphasis. Network preparing is performed on 60000 cycle steps, comparing to 33 hours of preparing information to accomplish preparing changes. Preparing subtleties were changed before every mishap to make irregularity in clump preparing. Other organization boundaries incorporate an expanding number of channels from 64 to 512 from the primary layer to the furthest limit of the implanting organization. The maximum derivation layer utilized a 2 X 2 window with column 2, bringing about a custom picture decrease of 2. The high-goal decoder network test utilized close-up areas, which increment the picture size by 2 things for every layer. When the preparation stage was finished, the design of the CAE network was ready and used to record the bones, air, and delicate tissues of the new MR pictures, which were subsequently handled into pseudo CT pictures. In this Framework the in-depth MRAC framework was developed in an integrated computational environment that integrates Python, MATLAB, and C / C ++.
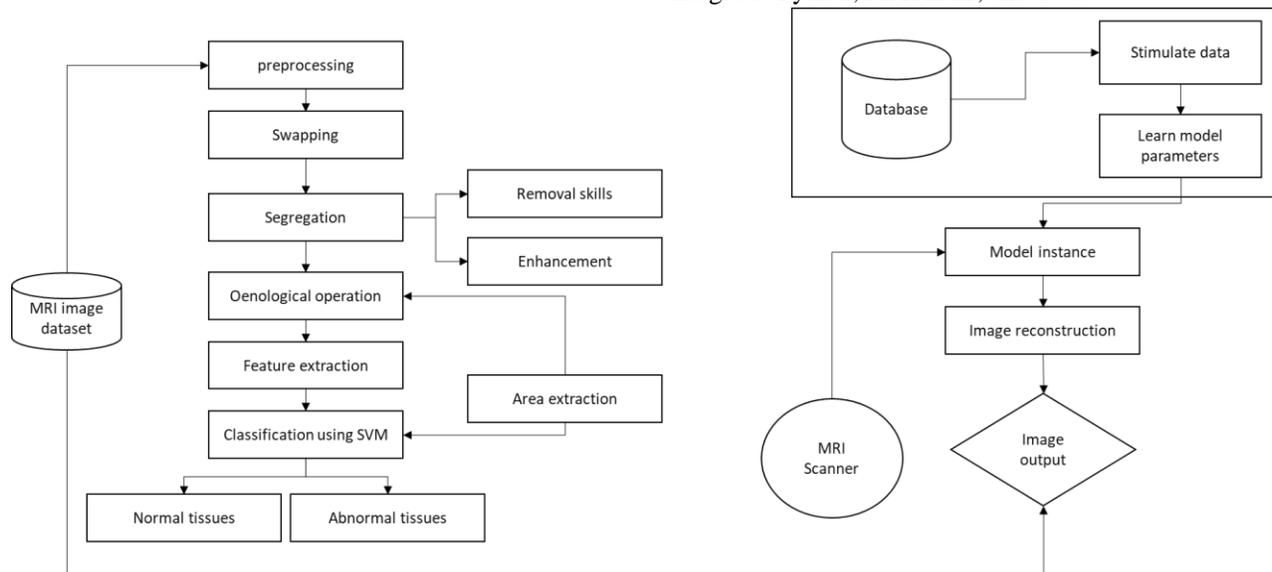
Fig. 1. Block Diagram of Proposed Method.

A. *Skull Stripping*

Head trauma is the process of separating brain tissue (cortex and cerebellum) from the nearby region (skull area and no brain). It is also a very important preparatory step that follows further analysis if there are multiple neurological MRI images (such as image registration or tissue fragmentation). Fig. 2 presents the multiple steps of tissue fragmentation.
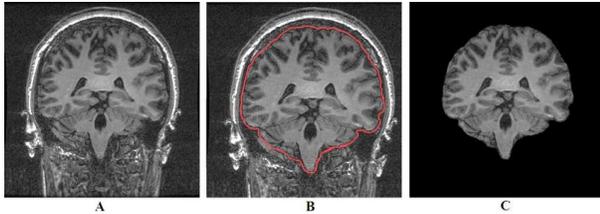
Fig. 2. Tissue Fragmentation Steps.

Skull Stripping/Head trauma is an important process in biomedical image analysis, and it is necessary to successfully diagnose brain tumor from MR images. Skull removal is the process of removing all the tissue that is not working in the brain tissue. By removing the skull, you may be able to remove excess tissue such as fat, skin, and skull from the brain. There are a few schemes available for skull scanning; other popular techniques are to disassemble the skull using a contour image, skull dissection is based on segment and morphological function, as well as the ripping of the skull based on the analysis of the histogram or number of divination. Fig. 3 presents the skull stripping algorithm sections. This study uses a skull stripping method based on the threshold function of the skull to remove skull tissue.
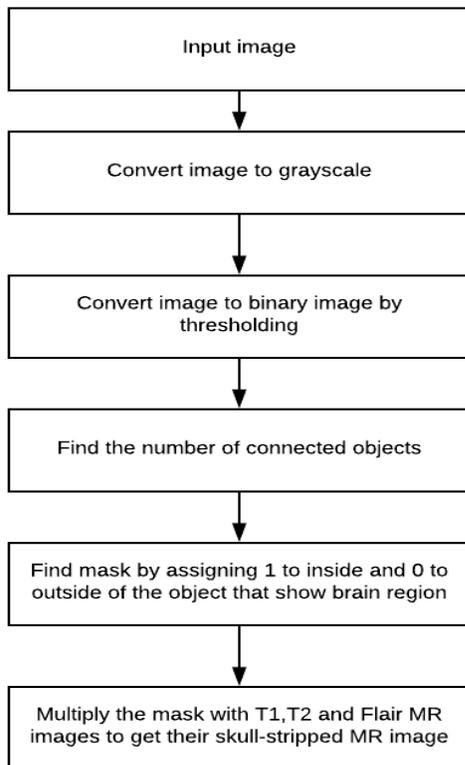
Fig. 3. The Phases Followed in Skull Stripping.

IV. EXPERIMENTS AND RESULTS

The data training comprises of north of 500 completely tested MR mind pictures we gathered from a 3T scanner (SIEMENS MAGNETOM TrioTim). The pictures are of an incredible variety counting hub, sagittal, level ones, different ones, for example, T1, T2 and PDweighted pictures and of various sizes. Informed assent was acquired from the imaging subject in consistence with the Institutional Review Board strategy. Undersampled estimations were reflectively gotten utilizing the 1D low-recurrence testing cover and the 2D Poisson circle inspecting veil. The enormous measure of ruined/ground truth subimage matches are then produced with the size of 33×33. At long last we utilize 90% of the subimage matches as the preparation dataset and the rest 10% for approving the preparation interaction.

We utilize three layers of convolution for the organization. The boundaries are individually set as n1 = 60, n2 = 30, M1 = 10, M2 = 6 and M3 = 4. The channel loads of each layers are instated by irregular qualities from a Gaussian dispersion with zero mean and standard deviation 0.001. The predisposition are completely introduced as 0. The preparation requires around three days, on a local machine with 16G memory and 24 CPU Intel Xeon processor with 2 Quado 6000 GPUs. Fig. 4 shows a bunch of recreation consequences of a cross-over mind picture.
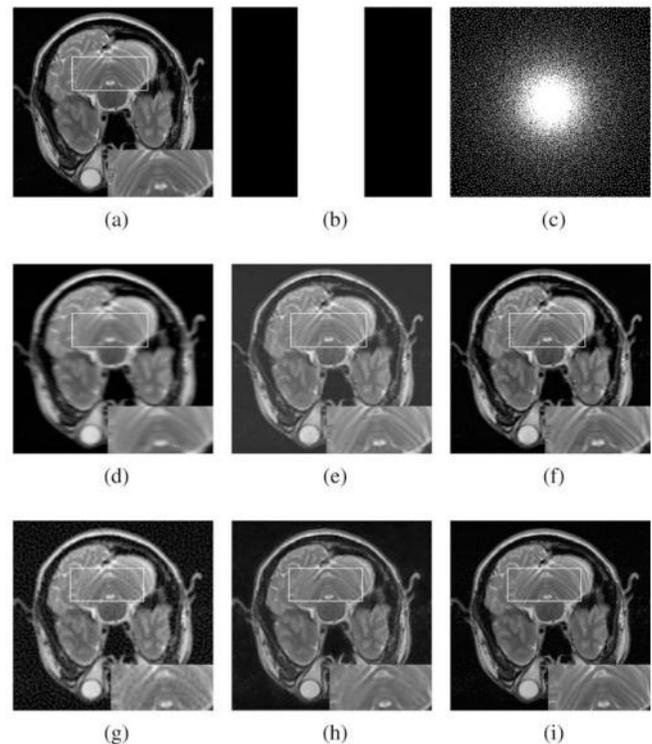
Fig. 4. Recreation Output of a Transversal Brain Image.

The cerebrum dataset was acquired completely tested with 12-channel head curl and T2-weighted super twist reverberation (TSE) succession (TE = 91.0ms, TR = 5000ms, FOV = 20×20cm, grid = 256×270, cut thickness = 3mm) by means of 3T scanner. Furthermore, the information was then undersampled reflectively with 1D low-recurrence examining

cover at a speed increase component of 3 and the 2D Poisson plate at a speed increase component of 5. We likewise tried the proposed strategy on a sagittal mind picture.

which was procured on a GE 3T scanner (GE Healthcare, Waukesha, WI) with a 32-channel head curl and 3D T1-weighted ruined angle reverberation grouping (TE=minimum full, TR= 7.5ms, FOV=24 × 24 cm, framework = 256 × 256, cut thickness=1.7mm). We can see from the pictures that there are many subtleties and designs caught by the organization. Besides, the picture created by the basic reproduction model is very near the unique picture. As indicated by Fig. 4(f), we can see the distinction picture is clamor like and comprises just the shape data. There are no conspicuous subtleties and designs lost. It exhibits that the proposed network is fit for reestablishing the subtleties and fine designs which are disposed of in the zero-filled MR picture. Moreover, albeit the disconnected preparation requires approximately three days, under similar GPU designs, it takes undeniably under 1 second for each web-based MR recreation case.

V. DISCUSSION AND RECOMMENDATION

All the previous articles that is based on the imaging, time-efficient and reducing cost but there is an error in some type of algorithms, their algorithms is not more efficient, fast and mainly highly graphical imaging to reduce the high type of motion. The challenge of fast MRI is to find an algorithm that can reconstruct and remove the image from the sample data below. The authors argue that, through time propose an analytical learning method for training generator network and stabilize training with rapid integration environment and less parameter input. In the dynamic visual field of MRI imaging such as heart and brain it should be continuously monitored to obtain a series of high-resolution images and over time. Various loss functions used for CNN training appeared in the developed models, which were quantitatively and qualitatively similar to CNN reconstructions. In the article, we want to train GAN's algorithm for other complex tasks such as perceptual loss or reconstruction. The best approach would be to train deep reconstruction in the radiologist network, which shows images in parallel scores provided by the radiologist or radiologists committee. However, this may require a large amount both data and time of radiologists. We have designed and implemented a GPU, Graphics and algorithm based fast magnetic resonance imaging (MRI) that uses deep learning neural network term to improve patient experience, scanning cost, time efficiency, image quality and reduce its motion. The algorithm matters because if the algorithm is faster and best, the machine works efficiently. Our architecture proposes a short overview of algorithm that works fast, time efficient and mainly to detect errors and disease. In imaging, Graphics plays an important role for enhance the patient experience and improve the reliability to detect more problems and disease more efficiently. Compared to CS-state of the art reconstruction techniques, our CNN achieves 150x faster reconstruction speed without any loss of image quality. Our main purpose is to maintain the image quality and its reconstruction. Connectivity as you know connectivity is the main part if your connectivity is not stable you result is in doubtful so build a strong connectivity with an algorithm and its commands. Sensitivity, is used to detect text patterns so it's

a challenge however, we use an algorithm to detect low spatial patterns.

In this study, using MR images of the brain, we isolated brain tissue into normal tissues such as white matter, gray matter, cybrospinal fluid (posterior) and infected tissue. MRI is associated with slower detection speed, due to data samples that can be collected directly in the image area, but rather in the k space, which contains general location data. Here the k-space and the image and the image are related to deviations, adjustments in one domain set limits on another. Samples of raw data are sequentially available in the k-space and the speed at which the k-limited space can be detected is limited by visual and hardware constraints. When the required field view resolution and MRI imaging is taken, the raw k-space data we need to obtain is traditionally determined by Nyquist - Shannon's sampling process. In short, we recommended all the parts in our architecture that how's our algorithm works and output the image this is main concern of GPU. If the algorithm works efficiently your data output gave efficient report to reduce the scanning cost, works on skull skipping and dataset these are the terms to improve the patient experience and also improve the image reconstruction and image quality.

VI. CONCLUSION

This study presented an accelerated magnetic imaging resonance remaking through Deep Learning. We determined that CNN can be utilized to display MR images recreation of 2-D information. Time reconstruction is accelerated by a factor of 150 in comparison to ES-PIRIT. The study experiments showed that there is two factors time scan to accelerate CNN model. In order to reconstruct MR images, several algorithms exploit the redundancy while receiving data from different receivers concurrently. Initially, we collected data from 32 receiver channels that were merged to a single channel by adopting whole-square operation. We observed that to train data over multi-channels, it is a tough and complex process; however, neural network models should be able to learn parallel imaging. The main purpose of this is to maintain the images graphics in very short period of time to maintain or lower the scanning cost with help of creating efficient algorithm. Recent developments documented show great potential for in-depth learning strategies in the field of MR brain image analysis. Even though further developed learning strategies have been utilized for brain MRI as of late. They are primarily emphasizing to move from traditional existing methodologies to mature AI/ML/DL based models. In bioinformatics, the brain MRI analyzing has been a vital challenge in computer based methods because of its complex structure and variations in appearance. MR scale does not match the scale due to differences in cognitive conventions, image retrieval, and the existence of pathology. Therefore, there is a need for familiarity strategies such as in-depth learning that can manage these differences.

REFERENCES

[1] Wang, Shanshan, et al. "Accelerating magnetic resonance imaging via deep learning." *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*. IEEE, 2016.

[2] Warach, Steve, et al. "Fast magnetic resonance diffusion-weighted imaging of acute human stroke." *Neurology* 42.9 (1992): 1717-1717.

[3] Schlemper J, Duan J, Ouyang C, Qin C, Caballero J, Hajnal JV, Rueckert D. Data consistency networks for (calibration-less) accelerated parallel MR image reconstruction. arXiv preprint arXiv:1909.11795. 2019 Sep 25.

[4] Defazio, Aaron. "Offset sampling improves deep learning based accelerated mri reconstructions by exploiting symmetry." *arXiv preprint arXiv:1912.01101* (2019).

[5] Alkan, Cagan, John Cocjin, and Andrew Weitz. "Magnetic resonance contrast prediction using deep learning." *Google Scholar* (2016).

[6] Chenevert, Thomas L., Paul E. McKeever, and Brian D. Ross. "Monitoring early response of experimental brain tumors to therapy using diffusion magnetic resonance imaging." *Clinical cancer research* 3.9 (1997): 1457-1466.

[7] Sandino, Christopher M., et al. "Deep convolutional neural networks for accelerated dynamic magnetic resonance imaging." *preprint* (2017).

[8] Sandino CM, Dixit N, Cheng JY, Vasanawala SS. Deep convolutional neural networks for accelerated dynamic magnetic resonance imaging. preprint. 2017.

[9] Sriram, Anuroop, et al. "End-to-end variational networks for accelerated MRI reconstruction." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2020.

[10] Qin, Chen, et al. "Convolutional recurrent neural networks for dynamic MR image reconstruction." *IEEE transactions on medical imaging* 38.1 (2018): 280-290.

[11] Ushinsky A, Bardis M, Glavis-Bloom J, Uchio E, Chantaduly C, Nguyentat M, Chow D, Chang PD, Houshyar R. A 3D-2D hybrid U-net convolutional neural network approach to prostate organ segmentation of multiparametric MRI. American Journal of Roentgenology. 2021 Jan 19;216(1):111-6.

[12] Sriram A, Zbontar J, Murrell T, Defazio A, Zitnick CL, Yakubova N, Knoll F, Johnson P. End-to-end variational networks for accelerated MRI reconstruction. InInternational Conference on Medical Image Computing and Computer-Assisted Intervention 2020 Oct 4 (pp. 64-73). Springer, Cham.

[13] Saba, Luca, et al. "Brain MRI-based Wilson disease tissue classification: an optimised deep transfer learning approach." *Electronics Letters* 56.25 (2020): 1395-1398.

[14] Ruiz J, Mahmud M, Modasshir M, Kaiser MS, Alzheimer's Disease Neuroimaging Initiative FT. 3D DenseNet ensemble in 4-way classification of Alzheimer's disease. InInternational Conference on Brain Informatics 2020 Sep 19 (pp. 85-96). Springer, Cham.

[15] Ebrahimi A, Luo S, Chiong R. Introducing Transfer Learning to 3D ResNet-18 for Alzheimer's Disease Detection on MRI Images. In2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ) 2020 Nov 25 (pp. 1-6). IEEE.

[16] Hammernik, Kerstin, et al. "Σ-net: Systematic Evaluation of Iterative Deep Neural Networks for Fast Parallel MR Image Reconstruction." *arXiv preprint arXiv:1912.09278* (2019).

[17] Liao, Xin, et al. "Machine-learning based radiogenomics analysis of MRI features and metagenes in glioblastoma multiforme patients with different survival time." *Journal of cellular and molecular medicine* 23.6 (2019): 4375-4385.

[18] Knoll F, Murrell T, Sriram A, Yakubova N, Zbontar J, Rabbat M, Defazio A, Muckley MJ, Sodickson DK, Zitnick CL, Recht MP. Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge. Magnetic resonance in medicine. 2020 Dec;84(6):3054-70.

[19] Shinan K, Alsubhi K, Alzahrani A, Ashraf MU. Machine Learning-Based Botnet Detection in Software-Defined Network: A Systematic Review. Symmetry. 2021 May;13(5):866.

[20] Lu, Siyuan, Zhihai Lu, and Yu-Dong Zhang. "Pathological brain detection based on AlexNet and transfer learning." *Journal of computational science* 30 (2019): 41-47.

[21] Fayyaz, Saqib, et al. "Solution of combined economic emission dispatch problem using improved and chaotic population-based polar bear optimization algorithm." *IEEE Access* 9 (2021): 56152-56167.

[22] Lundervold, Alexander Selvikvåg, and Arvid Lundervold. "An overview of deep learning in medical imaging focusing on MRI." *Zeitschrift für Medizinische Physik* 29.2 (2019): 102-127.

[23] Kim HS, Yoo KY, Kim LH. Improved Performance of Image Semantic Segmentation using NASNet. Korean Chemical Engineering Research. 2019;57(2):274-82.

[24] Hirra, Irum, et al. "Breast Cancer Classification From Histopathological Images Using Patch-Based Deep Learning Modeling." *IEEE Access* 9 (2021): 24273-24287.

[25] Bukhsh, Madiha, et al. "An Interpretation of Long Short-Term Memory Recurrent Neural Network for Approximating Roots of Polynomials." IEEE Access 10 (2022): 28194-28205.

[26] Tufail, Hina, M. Usman Ashraf, Khalid Alsubhi, and Hani Moaiteq Aljahdali. "The Effect of Fake Reviews on e-Commerce During and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection." IEEE Access 10 (2022): 25555-25564.

[27] Mumtaz, Mamoona, Naveed Ahmad, M. Usman Ashraf, Ahmed Alshaflut, Abdullah Alourani, and Hafiz Junaid Anjum. "Modeling Iteration's Perspectives in Software Engineering." IEEE Access 10 (2022): 19333-19347.

[28] Asif, Muhammad, et al. "A Novel Image Encryption Technique Based on Cyclic Codes over Galois Field." Computational Intelligence and Neuroscience 2022 (2022).

[29] Mehak, Shakra, et al. "Automated Grading of Breast Cancer Histopathology Images Using Multilayered Autoencoder." CMC-COMPUTERS MATERIALS & CONTINUA 71.2 (2022): 3407-3423.

[30] Naqvi MR, Iqbal MW, Ashraf MU, Ahmad S, Soliman AT, Khurram S, Shafiq M, Choi JG. Ontology Driven Testing Strategies for IoT Applications. CMC-Computers, Materials & Continua. 2022 Jan 1;70(3):5855-69.

[31] S. Tariq, N. Ahmad, M. U. Ashraf, A. M. Alghamdi, and A. S. Alfakeeh, "Measuring the Impact of Scope Changes on Project Plan Using EVM," vol. 8, 2020.

[32] Asif M, Mairaj S, Saeed Z, Ashraf MU, Jambi K, Zulqarnain RM. A Novel Image Encryption Technique Based on Mobius Transformation. Computational Intelligence and Neuroscience. 2021 Dec 17;2021.

[33] Shinan, Khlood, et al. "Machine learning-based botnet detection in software-defined network: a systematic review." Symmetry 13.5 (2021): 866.

[34] Hannan, Abdul, et al. "A decentralized hybrid computing consumer authentication framework for a reliable drone delivery as a service." Plos one 16.4 (2021): e0250737.

[35] Fayyaz, Saqib, et al. "Solution of combined economic emission dispatch problem using improved and chaotic population-based polar bear optimization algorithm." IEEE Access 9 (2021): 56152-56167.

[36] Hirra I, Ahmad M, Hussain A, Ashraf MU, Saeed IA, Qadri SF, Alghamdi AM, Alfakeeh AS. Breast cancer classification from histopathological images using patch-based deep learning modeling. IEEE Access. 2021 Feb 2;9:24273-87.

[37] Ashraf MU, Eassa FA, Osterweil LJ, Albeshri AA, Algarni A, Ilyas I. AAP4All: An Adaptive Auto Parallelization of Serial Code for HPC Systems. INTELLIGENT AUTOMATION AND SOFT COMPUTING. 2021 Jan 1;30(2):615-39.

[38] Hafeez T, Umar Saeed SM, Arsalan A, Anwar SM, Ashraf MU, Alsubhi K. EEG in game user analysis: A framework for expertise classification during gameplay. Plos one. 2021 Jun 18;16(6):e0246913.

[39] Siddiqui N, Yousaf F, Murtaza F, Ehatisham-ul-Haq M, Ashraf MU, Alghamdi AM, Alfakeeh AS. A highly nonlinear substitution-box (S-box) design using action of modular group on a projective line over a finite field. Plos one. 2020 Nov 12;15(11):e0241890.

[40] Alsubhi, Khalid, et al. "MEACC: an energy-efficient framework for smart devices using cloud computing systems." Frontiers of Information Technology & Electronic Engineering 21.6 (2020): 917-930.

[41] Riaz S, Ashraf MU, Siddiq A. A Comparative Study of Big Data Tools and Deployment PIatforms. In2020 International Conference on Engineering and Emerging Technologies (ICEET) 2020 Feb 22 (pp. 1-6). IEEE.

[42] Ashraf MU, Eassa FA, Ahmad A, Algarni A. Empirical investigation: performance and power-consumption based dual-level model for exascale computing systems. IET Software. 2020 Jul 27;14(4):319-27.

[43] Manzoor, Anam, et al. "Inferring Emotion Tags from Object Images Using Convolutional Neural Network." Applied Sciences 10.15 (2020): 5333.

[44] Alsubhi, Khalid, et al. "A Tool for Translating sequential source code to parallel code written in C++ and OpenACC." 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA). IEEE, 2019.

[45] Ashraf MU, Arshad A, Aslam R. Improving Performance In Hpc System Under Power Consumptions Limitations. International Journal of Advanced Research in Computer Science. 2019 Mar;10(2).

[46] Javed, Rushba, et al. "Prediction and monitoring agents using weblogs for improved disaster recovery in cloud." Int. J. Inf. Technol. Comput. Sci.(IJITCS) 11.4 (2019): 9-17.

[47] Ali, Muhammad, et al. "Prediction of Churning Behavior of Customers in Telecom Sector Using Supervised Learning Techniques." 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE). IEEE, 2018.

[48] Ashraf MU, Eassa FA, Albeshri AA, Algarni A. Performance and power efficient massive parallel computational model for HPC heterogeneous exascale systems. IEEE Access. 2018 Apr 9;6:23095-107.

[49] Ashraf MU, Eassa FA, Albeshri AA, Algarni A. Toward exascale computing systems: An energy efficient massive parallel computational model. International Journal of Advanced Computer Science and Applications. 2018 Jan;9(2).

[50] Ashraf MU, Eassa FA, Albeshri AA. Efficient Execution of Smart City's Assets Through a Massive Parallel Computational Model. InInternational Conference on Smart Cities, Infrastructure, Technologies and Applications 2017 Nov 27 (pp. 44-51). Springer, Cham.

[51] Ashraf MU, Eassa FA, Albeshri AA. High performance 2-D Laplace equation solver through massive hybrid parallelism. In2017 8th International Conference on Information Technology (ICIT) 2017 May 17 (pp. 594-598). IEEE.

Artificial Neural Network based Power Control in D2D Communication

Nethravathi H M[1]*, Dr S Akhila[2]*
Research Scholar, Department of ECE, BMSCE, VTU, Karnataka, India[1]
Professor, Department of ECE, BMSCE, VTU, Karnataka, India[2]

Abstract—As a viable technique for next-generation wireless networks, Device-to-Device (D2D) communication has attracted interest because it encourages the usage of point-to-point communications between User Equipment (UE) without passing over base stations (BS). Device-to-device (D2D) communication has been proposed in cellular networks as a supplementary paradigm to primarily increase network connection. This research takes into account a cellular network where users are trying device-to-device (D2D) connection. A D2D pair is composed of two D2D users (DUEs), a transmitter, and a receiver. To improve spectral efficiency, we use the premise that the D2D pairs only employ one communication channel. In order to minimize interference between D2D pairs and increase capacity, a power control is required. In the scenario where only typical cellular channel gains between base stations and DUEs are known and channel gains among DUEs are completely inaccessible, we address the issue of D2D power control. For each individual D2D pair, we use an artificial neural network (ANN) to calculate the transmission power. We show that the maximum aggregate capacity for the D2D pairs may be reached while anticipating the transmission power setting for D2D pairs using cellular channel gains.

Keywords—*Artificial Neural Network (ANN); Base Stations (BS); CUE; Device-to-Device (D2D); DUE; ML; User Equipment (UE)*

I. INTRODUCTION

In general, the allocation of spectral resources and power control are two processes that are not separate from one another. Power control is a way to limit interference between network users while guaranteeing the minimal quality of service requirements. Interference between cellular and D2D users occurs when shared resources are assigned; in this scenario, the transmission power of the D2D user is lowered in accordance with the minimum reference value of the cellular communication. Thus, the network can also establish a maximum transmission value for D2D devices, depending on the impact of the different power levels of the devices on the quality of the cellular links.

Since there is minimal to no chance of interference between D2D and cellular users, the transmit power may frequently be greater when dedicated spectrum resources are allocated to D2D devices rather than shared resources. Thus, the management of spectrum resources can be carried out mainly by the network or by the devices themselves. In dedicated sharing, devices must be aware of the environment and use resources adaptively, trying to cause minimal interference to other users. Leaving resource management in

the hands of the device it results in a less predictable process, but with greater flexibility and less complexity.

Power control is one of the most often used interference avoidance methods [1]. This technique allows an adjustment of maximum D2D transmission power so as not to exceed the predefined SINR limit in cellular communication that is, the transmission power level can be limited by the eNB in order to reduce potential interference to cellular receivers. In addition, the eNB can also control access to shared cellular communications resources and D2D peers, which represents greater efficiency in the use of spectrum [2].

However, this technique is simple, but not efficient, in the sense that by placing the restriction at the D2D power level, it implies that it may affect D2D communication, that is, it results that D2D communication may not always be viable.

The schematic depiction of the power control strategy is shown in Fig. 1. The distance between the D2D pairs, the distance between the D2D pairs and the eNB, or the distance between the DUE_T pairs and the cellular users (CUE) all have a significant role in how well this strategy works.
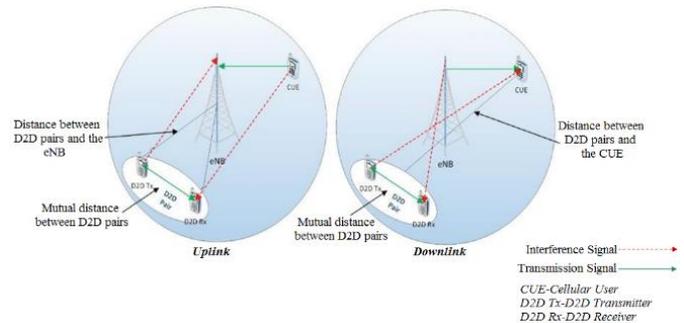
Fig. 1. D2D Power Control Technique [3].

The performance of the D2D communication is unaffected by the power reduction if the D2D pairs are close together and remote from the eNB or CUE.

The D2D power reduction can provide a very low probability of D2D communication or even completely preclude communication between the D2D pairs in the opposite scenario, when the distance between the D2D pairs is great while also being reasonably close to the eNB or CUE.

In spite of this, the power management strategy only works well for reducing D2D communication interference in cellular communication when the D2D pairs are near together and at the same time separated from the eNB and/or CUE. The

*Corresponding Author.

performance of D2D communication might be harmed by severe regulations that limit the transmission power of D2D users.

II. LITERATURE REVIEW

Machine learning has application in a wide range of fields: image processing, audio, finance, economics, social behaviour analysis, telecommunications network management, the latter being one of the most promising applications of this branch of artificial intelligence [4] [5] [6].

Telecommunications networks are now able to learn from and extract information from data thanks to the discipline of machine learning's growth. The latter is of vital importance for new wireless network standards such as 5G [4] [5] [6].

Until now, the development of machine learning and telecommunications networks have been carried out as different research fields, but the application of ML in 5G networks has demonstrated the present and future potential of this combination of paradigms and technologies. It is now clear that this is the case with mobile edge computing, location-based services, contextual networking, edge caching, network traffic monitoring, and big data analytics [4] [5] [6].

Machine learning is perfect for difficult problems that need a lot of human fine-tuning to solve or for problems for which there is no traditional answer. The above problems can be addressed by replacing conventional software that contains a large number of rules with software (containing ML routines) that is capable of automatically learning from data. Automatic feature extraction, anomaly detection, scenario prediction, environment adaption, information gathering on complicated issues with enormous volumes of data, and pattern discovery are some of the distinctions between machine learning algorithms and conventional cognitive algorithms [4] [5] [6].

Many parameters in wireless and mobile networks are estimated using heuristic techniques because there is no analytical answer or because it is impractical given its complexity and related expenses. In these situations, machine learning algorithms can aid in the resolution by forecasting and estimating variables and functions based on the existing information [4] [5] [6].

Despite the advantages of using ML, this technique has limitations that prevent its widespread use in wireless communications, including the findings' interpretability, the difficulty of locating pertinent data, the needed processing power, the complexity added, and the lengthy training cycles of some algorithms, among others, which cause the cost, time, latency and delay introduced to be incompatible with some real-time applications [4].

Despite the difficulties to be overcome, it can be said that machine learning is already widely used in the modelling of a variety of technical issues of next-generation systems, including large-scale MIMO, low-latency communications, heterogeneous networks made up of femtocells and small cells, vehicular networks, D2D networks, among others [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15].

Basically AI is any technique that allows computers to imitate human behaviour; ML are those algorithms that use statistical techniques to allow machines to "learn" through experiences; and Multi-layer neural networks are used in the deep learning subfield of machine learning to carry out algorithm execution [4] [5] [6].

D2D communications were initially found to be beneficial for increasing spectral efficiency and frequency reuse, while reducing communication delays. However, this mode of communication has introduced additional interference due to multiple simultaneous communications on the same frequency bands, which is difficult to control and mitigate. This is especially true when D2D communications take place in Inband and/or Underlay circumstances, i.e., on the same frequency as the cellular network and/or concurrently with cellular communications [16].

Numerous academics have looked at resource distribution challenges in this setting, such as connecting for D2D communications and power and frequency distribution. The performance of D2D and cellular communications in a basic wireless network was examined by the authors of [17] in relation to the effect of the QoS requirement on latency. They developed the best power allocation strategies taking into account statistical QoS provisioning in order to maximise network performance while adhering to QoS limitations.

The authors of [18] described an underlying D2D-based network that allows D2D users to serve as a relay for cellular communications in order to optimize the transfer rates that D2D users may achieve while still meeting the QoS criteria of cellular users. The best power allocation technique for the base station (BS) and the D2D transmitter is found in closed mode when there is a global power restriction. Moreover, the simulation results demonstrate the superiority of using D2D devices as complete relays compared to classical cellular communications.

In order to increase the overall throughput of cellular and D2D users while maintaining the required minimum QoS for WiFi users, the authors of [19] investigated the issue of shared channel and power distribution. They proposed a particle warm optimization (PSO) algorithm, which significantly reduces interference in both licensed and unlicensed bands and improves throughput performance. This work has focused on the derivation of optimal solutions or heuristics that are applied offline, i.e., with the a priori knowledge of the state of the entire system. For online deployments, where device locations and channel conditions are constantly shifting, these techniques would be insufficient. The research has suggested increased learning methodologies as a solution to these problems.

In [20], the authors proposed a Q-learning method that jointly assigns channels and power levels to D2D devices to maximize end-to-end traffic flow in an underlying scenario D2D. They showed that their method outperforms benchmarks in terms of average system capacity. The authors of [21] employed Q-learning with logarithmic regret to minimize total transmission power and associations to SB. Then they extended their proposal to distribute Q-learning to reduce complexity. The results obtained demonstrated the superiority

of their approaches compared to benchmarks, in terms of reduction of transmission energy consumption. Additionally, the authors of [22] looked at the issue of power distribution in D2D underlying networks. The authors suggested a D2D power control based on extreme hierarchical machine learning that, in terms of communication speed and energy efficiency, surpasses both distributed Q-learning and CART decision trees. Finally, in order to maximise the D2D sum throughput, [23] developed a deep reinforcement learning strategy for scheduling D2D connections in a cellular underlay network. The results of the simulation showed that the suggested strategy outperformed the conventional procedures.

D2D communications would offload content and traffic off the cellular network, reduce recovery and content delivery, and enhance the energy efficiency of the cellular network, all of which would considerably improve system performance in caching cellular systems. These benefits have been demonstrated in several recent works.

In [24], the authors proposed to cache content either in the cellular BS or in D2D devices. Next, they studied the co-design of the content caching and delivery policy, given their prior knowledge of user demands. The authors described in [25] the D2D cognitive communications underlying cellular transmissions in order to cache files between devices transparently. On BS and D2D devices, they then assessed the delay and queue length. Additionally, the authors in [26] suggested a caching method that reduces the typical content delivery time in a D2D-supported cellular network. They demonstrated the superiority of their proposed greedy algorithm over the naive popularity-based caching policy. In order to maximize the probability of successful transmission, which gauges the percentage of users meeting their QoS delay criteria, the authors of [27] took use of the collaboration between BS caching and D2D caching. They showed that by using the proposed block coordinate descent algorithm, they can achieve significant performance gains over conventional caching methods.

In contrast, the authors of [26] used a user behaviour learning algorithm that predicts user requests and estimates file popularity. Thus, this information is used to adjust the caching policy between D2D devices, with the aim of minimizing transmission delays for file delivery. Simulations demonstrate that the suggested strategy performs better than both probabilistic and naïve caching. In [28], the authors examined the combined cache placement and resource allocation problem in heterogeneous networks with D2D support. In terms of average content delivery times, they suggested two imperfect heuristic methods that outperform D2D-unsupported systems. Authors of [29] examined content placement in the context of a D2D-assisted Hetnet with the goal of reducing the average content delivery time. For a single content library, the optimal solution is obtained and then used to design a low-complexity heuristic for a larger content library. They demonstrated through simulations the advantages of storing popular data in D2D devices inside the cellular system that had the best connectivity to other D2D devices.

The authors of [30] investigated content caching and delivery in a D2D network. They used a deep-Q network (DQN) method to improve content distribution with respect to delays and energy restrictions, and echo-state networks for predicting user mobility and content popularity. In [31-32], the authors formulated a Multi-Agent Reinforcement Learning (MARL) D2D caching problem, aiming to improve the average latency content delivery rate and cache hit rate, without prior knowledge of content popularity. To solve it, they proposed Q-learning for independent learning devices and joint action learning devices. Through simulations, they found that joint action learning algorithms outperform individual learning algorithms and other basic approaches.

Numerous research have been conducted to improve the spectral efficiency of D2D networks using power regulation as a resource allocation problem [33-40]. A non-convex issue, sum capacity-focused power management over D2D pairs. Because of this, a number of iterative techniques with varied degrees of complexity are provided in the literature. But latency problems might arise with iterative techniques. Deep neural networks (DNN) are being employed as an alternative lately by researchers to regulate instantaneous power in D2D communication [34]. Through supervised [35]–[36] or unsupervised [37]–[40] learning based on offline training, the DNN significantly decreases the complexity of power control. Importantly, power control approaches that combine unsupervised learning and the DNN surpass the presently used iterative strategies in terms of cumulative capacity. For unsupervised learning, a DNN loss function is necessary. Examples include the overall capacity as a function of channel gains among DUEs and DUE transmission powers. The fact that all of the aforementioned approaches, both conventional and DNN-based, typically take into account full information of all the D2D channel gains is a significant disadvantage. By feeding the D2D channel gains into the neural network as an input, machine learning techniques change the transmission powers. Sometimes it is possible to reduce the need that channel state information (CSI) contain all distributed D2D channel gain values. In contrast to the signalling needed for conventional cellular communications, even a partial understanding of the benefits of D2D channels indicates that they come at a significant expense in terms of extra channel estimates and signalling.

This paper's main proposition is a unique power control technique for D2D communication based on artificial neural networks (ANNs) that requires no further understanding of the D2D channel gains. Since the channel quality to all nearby base stations is sent during a shared network operation, there is no signalling overhead. Our suggested ANN's main goal is to establish a connection between D2D and cellular channel gains. In order to increase the total capacity, the transmission power of the D2D pairs is then changed using this relation. It is vital to keep in mind that there is no known explanation for the association between cellular channel gains and the overall D2D pair capacity. Therefore, it is difficult to provide a suitable loss function for an ANN based on unsupervised learning. Because we first identify the targeted DUEs transmission powers that increase the cumulative capacity, we employ a supervised learning technique. The ANN is then

trained to create a mapping between cellular channel gains and the target transmission powers in order to arrive at the correct power setting. The whole training procedure takes place offline, and without any further training during communication, the trained ANN is used for swift power control decisions in the real network.

The remainder of the paper is laid out as follows. The proposed methodology for power control in D2D is presented in Section III. Section IV states the problem under study. Section V talks about power control for D2D pairs using ANN. Section VI discusses the MATLAB based simulation outcome of the research followed by the conclusive remarks in Section VII.

III. PROPOSED METHODOLOGY

A. System Model

We take into account a model with M D2D users (DUEs) and L base stations producing N D2D pairings inside a square area (i.e., $N = M/2$ provided M is an even number). To ensure the viability of the D2D connection, the distance D_{max} between the receiver DUE_R and transmitter DUE_T that make up the D2D pair is restricted. It is expected that the common channel is used by D2D pairs. The various D2D pairs that are using the channel interfere with one another. This definition of the nth D2D pair's capacity reads as follows [33]:

$$C_n = B \log_2 \left[1 + \frac{p_n g_{n,n}}{\sigma_0 B + \sum_{\substack{j=1 \\ j \neq n}}^{j=N} p_j g_{j,n}} \right] \qquad (1)$$

Where,

- σ_0 is symbolized for the noise power spectral density.

- p_j represents the transmission power of the jth DUE_T, and $g_{j,n}$ is the channel gain between the nth DUE_R and the jth DUE_T.

- B is symbolized for channel bandwidth.

- p_n is the symbol for the transmission power of the nth DUE_R.

Since it is challenging to predict D2D channel gains, therefore a channel between any DUE_R and DUE_T ($g_{j,n}$ and $g_{n,n}$) is thought to be unknown.

Since D2D users will continue observe the transmit channels to the server base station (for prediction, decoding, etc.), the channel quality information between each D2D user and neighbouring base stations should be periodically measured and reported to the server base station. The $G_{m,l}$ stands for the equivalent expected channel gain between the lth and the mth DUE.

IV. PROBLEM FORMULATION

To boost the overall capacity of D2D pairs, this study effort aims to optimize the transmission power p_n for every n^{th} D2D pair. Since p_{max} and p_{min} are the maximum and minimum transmission powers, respectively, the binary power

control is assumed resulting in $p_n \in \{p_{min}, p_{max}\}$. To optimize their combined capacity, the D2D pairs' transmission power must be set up as follows:

$$P = \arg\max \sum_{n=1}^{n=N} C_n \qquad (2)$$

$$p_n \in \{p_{min}, p_{max}\}, \forall_n \in \{1, 2, \dots, N\} \qquad (3)$$

Where $P = \{p_1, \dots, p_N\}$ is the vector holding all D2D pair transmission powers, maximizing the total number of D2D pair capacity, and constraint of equation (3) guarantees that each D2D pair's transmission power is set to either maximum or minimum.

Maximizing the cumulative capacity of D2D pairings is the goal of the optimization problem in Eq. (2). But as we can see from Eq. 1, C_n is reliant on the channel gains of D2D pair. We focus on the case when the developers of current systems presuppose full or at least partial knowledge of the channel gains of D2D pairs.

V. POWER CONTROL FOR D2D PAIRS USING ARTIFICIAL NEURAL NETWORKS

Based on biological neural structures, artificial neural networks are arrangements popularly used in machine learning. A Neural Network is made up of neurons, which in turn are organized into layers. Each layer can contain one or multiple neurons, and its position in the neural network dictates which class it belongs to: input layer, hidden layer, or output layer. Input layer and output layer are terms used to describe the first and last layers, respectively. Between the input and output layers, there are several layers, each of which is referred to as a hidden layer.

The cellular channel gains and D2D channel gains are related mathematically, which is the basis of the optimization issue in equation (2). For mobile networks, the relationship between cellular channel gains and D2D channel gains is unknown, and it cannot even be deduced analytically from any known mobile network characteristics. In order to automatically learn this relation and adjust the transmission strength of the D2D pairs in accordance, we advise employing an Artificial Neural Network (ANN). More specifically, the ANN may be thought of as a regulator that manages the transmission power of the D2D pair based only on the information provided by the D2D users to the base stations regarding cellular channel gains.

When using the binary power control, the goal of Eq. (2)'s optimization is to set each D2D pair's transmission power to either $p_n = p_{min}$ or $p_n = p_{max}$. Thus, the challenge of N identical binary classification is how to set the transmission power for N D2D pairs. In order to build the mapping between cellular channel gains and the ideal binary transmission power setting for every n^{th} D2D pair that maximises the cumulative capacity of D2D pairs, we propose the utilization of a fully-connected ANN.

The suggested ANN for binary classification is shown in Fig. 2. An input layer $X = (X_1, X_2, \dots X_n)$, hidden layers $H = (H_1, H_2, \dots H_n)$, and an output layer $Y = (Y_1, Y_2, \dots Y_n)$ make up the proposed ANN. The suggested ANN features an input layer that contains an input vector, and this input layer

aligns the cellular channel gains from the D2D users to the base stations. The term, $Out_{X_0} = P_{1,1}, P_{1,2}, \ldots, P_{C,D}$ with a length of $C \times D$ is a vector representing the cellular channel gains between base stations and D2D users and is produced by the input layer.

$(H_1, H_2, \ldots H_n)$

Input Layer   **Hidden Layer**
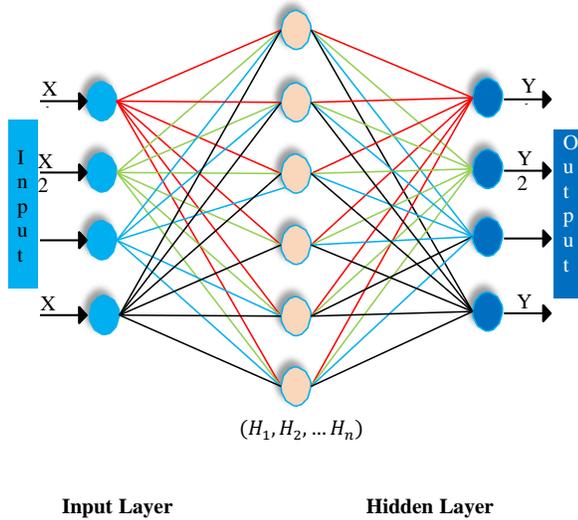
Fig. 2.   Proposed ANN Architecture for Binary Classification, Matching to a Single D2D Pair's Transmission Power.

Because the value of the sigmoid function is between 0 and 1, the output of the ANN will be $out_Y \in [0,1]$, displays the probability that $p_n = p_{max}$. Consequently, the nth D2D pair's transmission power is set to:

$$p_n = \begin{cases} p_{max} & if\ out_Y > 0.5 \\ p_{min} & otherwise \end{cases} \qquad (4)$$

A basic model of a neuron is described by applying an activation function Φ on a linear combination F of the input data, $X = (X_1, X_2, \ldots X_n)$, with their corresponding weights, $W_1, W_2, \ldots W_d$, and a bias constant b. Thus getting the output Y such that $Y = \Phi(F)$, in which $F = b + \sum_{i=1}^{d} W_i X_i$. This output is then passed along with the other outputs of this layer as input data to the next layer; this chain is perpetuated until the final output is generated by the output layer.

The best-known neuron model is called the sigmoid neuron, whose activation function Φ is the sigmoid function, that is, $\Phi(F) = \frac{1}{1 + e^{-\eta F}}$, and therefore the output Y is calculated as follows:

$$Y = \frac{1}{1 + e^{-\eta \sum_{i=1}^{d} W_i X_i - b}} \qquad (5)$$

The popularity of the sigmoid neuron is due to the ease of obtaining the partial derivative of this function, since the technique used by many to adjust the weights and bias is gradient descent. These adjustments are made by the back-propagation and feed-forward processes.

There is no simple mathematical relationship between the cumulative capacity of the D2D pairs and the cellular channel gains that can be used to calculate the transmission power of the D2D pairs. We therefore provide an offline supervised learning-based technique in which the optimum binary transmission powers are identified following a comprehensive search in order to maximise the sum capacity of D2D pairs. The transmission power of the nth D2D pair is then provided as a targeted class associated with the set of the cellular channel gains as features to the proposed artificial neural network. A single learning sample is comprised of the features and targeted class. From the collected learning samples, testing and training sets are produced.

VI.   Simulation Results

A. Simulation Parameters

TABLE I.   SIMULATION PARAMETERS

The radius of the cell	500 meters
Maximum transmit power of all devices P_{max}	0.2 Watt
Minimum data rate requirement of all D2D pairs	3 bps/Hz
The proportion of CUE's minimum rate requirement to CUE's achievable data rate	0.2
Maximum iterations of Dinkelbach's method	3
Maximum iterations of condensation method	2

B. Simulation Results

The proposed ANN is trained using samples of cellular channel gains and their associated expected transmission powers from the training set. In order to prevent overfitting, the trained ANN is next assessed on a test set of data using cellular channel gains that were not used during the training. This illustrates the set of samples' classification accuracy. The simulation results of our study are illustrated in Fig. 3 to 6 and Tables I and II.
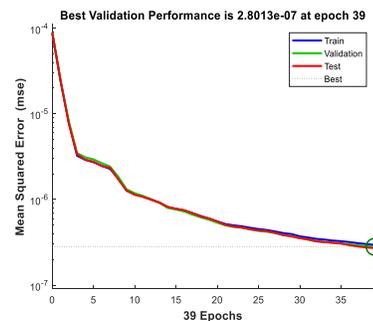
Fig. 3.   Training and Testing Performance of Proposed Approach.
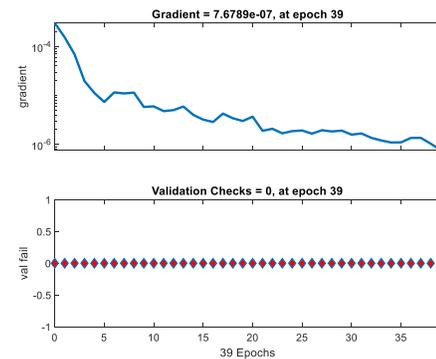
Fig. 4.   Performance Measures of Proposed Classifier.
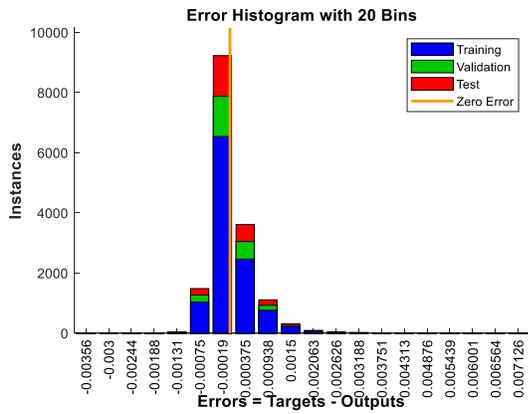
Fig. 5.   Error Histogram Graph.

TABLE II.   PERFORMANCE RESULTS

Optimal result obtained by NN (System Efficiency)		63.0430
Throughput		78.9144
Total Power Consumption		1.2518 Watt
Channel Gain of First Cell's CUEs	First CUE	2.6703
	Second CUE	1.9669
Channel Gain of Second Cell's CUEs	First CUE	2.0554
	Second CUE	2.2861
Channel Gain of First Cell's D2D Devices	First Device	1.5199
	Second Device	6.8406
	Third Device	9.6770
	Fourth Device	12.1281
Channel Gain of Second Cell's D2D Devices	First Device	4.3259
	Second Device	13.7552
	Third Device	11.8723
	Fourth Device	9.8168
Classification Accuracy	p_{min}=93.6%	**Total Accuracy = 92.25%**
	p_{max}=90.9%	
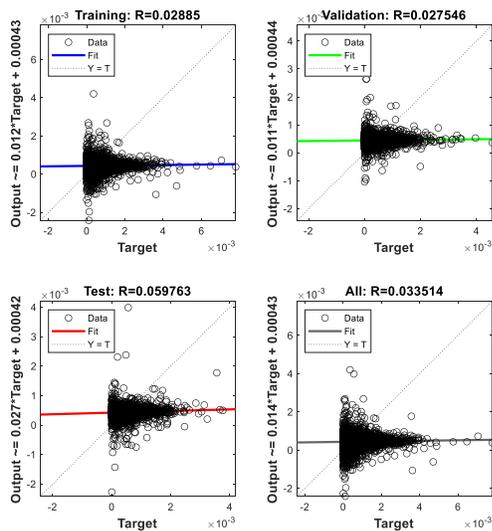
Fig. 6.   Final Output.

VII. CONCLUSION

For D2D communication, we have developed a novel power control method in this paper that doesn't require any understanding of the channel gains. In order to determine the transmission power of each D2D pair using an artificial neural network, the suggested method simply utilises the cellular channel gains between D2D users and nearby base stations. The main advantage of the suggested plan over current practises is that the network does not experience any additional signalling overhead. It is only necessary to be aware of the cellular channel gains, which are reported on a regular basis for a variety of objectives connected to conventional communication and handover. The suggested method surpasses the scenario with no power control and achieves a D2D pair sum capacity that is almost ideal with a maximum classification accuracy of 92.25%. Expanding the applicability of the suggested strategy to the estimation of D2D channel gains, which can then be applied to any radio resource management issue, should be the primary focus of future study.

VIII. CONFLICTS OF INTEREST

Declare conflicts of interest or state "The authors declare no conflict of interest."

AUTHOR CONTRIBUTIONS

All authors had contributed to this work.

ACKNOWLEDGMENT

REFERENCES

[1] Islam, T., Kwon, C. and Noh, Y., 2022. Transmission Power Control and Relay Strategy for Increasing Access Rate in Device to Device Communication. IEEE Access, 10, pp.49975-49990.

[2] Safdar, G.A., Ur Rehman, M., Muhammad, M. and Imran, M.A., 2022. Interference Mitigation in D2D Communication Underlaying LTE-A Network. Interference Mitigation in Device-to-Device Communications, pp.13-47.

[3] Fan, Z., Gu, X., Nie, S. and Chen, M., 2017, December. D2D power control based on supervised and unsupervised learning. In 2017 3rd IEEE International Conference on Computer and Communications (ICCC) (pp. 558-563). IEEE.

[4] Morocho-Cayamcela, M.E., Lee, H. and Lim, W., 2019. Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions. IEEE access, 7, pp.137184-137206.

[5] Jiang, C., Zhang, H., Ren, Y., Han, Z., Chen, K.C. and Hanzo, L., 2016. Machine learning paradigms for next-generation wireless networks. IEEE Wireless Communications, 24(2), pp.98-105.

[6] Yadav, S.S., Hiremath, S., Surisetti, P., Kumar, V. and Patra, S.K., 2022. Application of Machine Learning Framework for Next-Generation Wireless Networks: Challenges and Case Studies. Handbook of Intelligent Computing and Optimization for Sustainable Development, pp.81-99.

[7] Tayyaba, S.K., Khattak, H.A., Almogren, A., Shah, M.A., Din, I.U., Alkhalifa, I. and Guizani, M., 2020. 5G vehicular network resource management for improving radio access through machine learning. IEEE Access, 8, pp.6792-6800.

[8] Li, T., Zhao, M. and Wong, K.K.L., 2020. Machine learning based code dissemination by selection of reliability mobile vehicles in 5G networks. Computer Communications, 152, pp.109-118.

[9] Bega, D., Gramaglia, M., Fiore, M., Banchs, A. and Costa-Perez, X., 2019, April. DeepCog: Cognitive network management in sliced 5G

networks with deep learning. In IEEE INFOCOM 2019-IEEE conference on computer communications (pp. 280-288). IEEE.

[10] Asadi, A., Müller, S., Sim, G.H., Klein, A. and Hollick, M., 2018, April. FML: Fast machine learning for 5G mmWave vehicular communications. In IEEE INFOCOM 2018-IEEE Conference on Computer Communications (pp. 1961-1969). IEEE.

[11] Balevi, E. and Gitlin, R.D., 2017, December. Unsupervised machine learning in 5G networks for low latency communications. In 2017 IEEE 36th International Performance Computing and Communications Conference (IPCCC) (pp. 1-2). IEEE.

[12] Li, J., Zhao, Z. and Li, R., 2018. Machine learning-based IDS for software-defined 5G network. Iet Networks, 7(2), pp.53-60.

[13] Klautau, A., Batista, P., González-Prelcic, N., Wang, Y. and Heath, R.W., 2018, February. 5G MIMO data for machine learning: Application to beam-selection using deep learning. In 2018 Information Theory and Applications Workshop (ITA) (pp. 1-9). IEEE.

[14] Sim, G.H., Klos, S., Asadi, A., Klein, A. and Hollick, M., 2018. An online context-aware machine learning algorithm for 5G mmWave vehicular communications. IEEE/ACM Transactions on Networking, 26(6), pp.2487-2500.

[15] Hashima, S., ElHalawany, B.M., Hatano, K., Wu, K. and Mohamed, E.M., 2021. Leveraging machine-learning for D2D communications in 5G/beyond 5G networks. Electronics, 10(2), p.169.

[16] Malik, P.K., Wadhwa, D.S. and Khinda, J.S., 2020. A survey of device to device and cooperative communication for the future cellular networks. International Journal of Wireless Information Networks, 27(3), pp.411-432.

[17] Cheng, W., Zhang, X. and Zhang, H., 2015. Optimal power allocation with statistical QoS provisioning for D2D and cellular communications over underlaying wireless networks. IEEE Journal on Selected Areas in Communications, 34(1), pp.151-162.

[18] Zhang, G., Yang, K., Liu, P. and Wei, J., 2014. Power allocation for full-duplex relaying-based D2D communication underlaying cellular networks. IEEE Transactions on Vehicular Technology, 64(10), pp.4911-4916.

[19] Girmay, G.G., Pham, Q.V. and Hwang, W.J., 2019. Joint channel and power allocation for device-to-device communication on licensed and unlicensed band. IEEE Access, 7, pp.22196-22205.

[20] Luo, Y., Shi, Z., Zhou, X.I.N., Liu, Q. and Yi, Q., 2014, December. Dynamic resource allocations based on Q-learning for D2D communication in cellular networks. In 2014 11th international computer conference on wavelet actiev media technology and information processing (ICCWAMTIP) (pp. 385-388). IEEE.

[21] Huang, Y.F., Tan, T.H., Wang, N.C., Chen, Y.L. and Li, Y.L., 2018, July. Resource allocation for D2D communications with a novel distributed Q-learning algorithm in heterogeneous networks. In 2018 International Conference on Machine Learning and Cybernetics (ICMLC) (Vol. 2, pp. 533-537). IEEE.

[22] Xu, J., Gu, X. and Fan, Z., 2018, September. D2D power control based on hierarchical extreme learning machine. In 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) (pp. 1-7). IEEE.

[23] Moussaid, A., Jaafar, W., Ajib, W. and Elbiaze, H., 2018, October. Deep reinforcement learning-based data transmission for D2D communications. In 2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) (pp. 1-7). IEEE.

[24] Gregori, M., Gómez-Vilardebó, J., Matamoros, J. and Gündüz, D., 2016. Wireless content caching for small cell and D2D networks. IEEE Journal on Selected Areas in Communications, 34(5), pp.1222-1234.

[25] Yang, C., Zhao, X., Yao, Y. and Xia, B., 2016, December. Modeling and analysis for cache-enabled cognitive D2D communications in cellular networks. In 2016 IEEE Global Communications Conference (GLOBECOM) (pp. 1-6). IEEE.

[26] Li, Y., Zhong, C., Gursoy, M.C. and Velipasalar, S., 2018. Learning-based delay-aware caching in wireless D2D caching networks. IEEE Access, 6, pp.77250-77264.

[27] Wang, Y., Tao, X., Zhang, X. and Gu, Y., 2017. Cooperative caching placement in cache-enabled D2D underlaid cellular network. IEEE Communications Letters, 21(5), pp.1151-1154.

[28] Jaafar, W., Ajib, W. and Elbiaze, H., 2018, October. Joint caching and resource allocation in D2D-assisted heterogeneous networks. In 2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) (pp. 1-8). IEEE.

[29] Jaafar, W., Ajib, W. and Elbiaze, H., 2019, September. Caching optimization for D2D-assisted heterogeneous wireless networks. In 2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) (pp. 1-6). IEEE.

[30] Yin, J., Li, L., Xu, Y., Liang, W., Zhang, H. and Han, Z., 2018, November. Joint content popularity prediction and content delivery policy for cache-enabled D2D networks: A deep reinforcement learning approach. In 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (pp. 609-613). IEEE.

[31] Jiang, W., Feng, G., Qin, S. and Yum, T.S.P., 2018, April. Efficient D2D content caching using multi-agent reinforcement learning. In IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (pp. 511-516). IEEE.

[32] Jiang, W., Feng, G., Qin, S., Yum, T.S.P. and Cao, G., 2019. Multi-agent reinforcement learning for efficient content caching in mobile D2D networks. IEEE Transactions on Wireless Communications, 18(3), pp.1610-1622.

[33] Najla, M., Gesbert, D., Becvar, Z. and Mach, P., 2019, December. Machine learning for power control in D2D communication based on cellular channel gains. In 2019 IEEE Globecom Workshops (GC Wkshps) (pp. 1-6). IEEE.

[34] Gündüz, D., de Kerret, P., Sidiropoulos, N.D., Gesbert, D., Murthy, C.R. and van der Schaar, M., 2019. Machine learning in the air. IEEE Journal on Selected Areas in Communications, 37(10), pp.2184-2199.

[35] Sun, H., Chen, X., Shi, Q., Hong, M., Fu, X. and Sidiropoulos, N.D., 2017, July. Learning to optimize: Training deep neural networks for wireless resource management. In 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) (pp. 1-6). IEEE.

[36] de Kerret, P., Gesbert, D. and Filippone, M., 2018, May. Team deep neural networks for interference channels. In 2018 IEEE International Conference on Communications Workshops (ICC Workshops) (pp. 1-6). IEEE.

[37] Lee, W., Kim, M. and Cho, D.H., 2018. Deep power control: Transmit power control scheme based on convolutional neural network. IEEE Communications Letters, 22(6), pp.1276-1279.

[38] Kim, M., de Kerret, P. and Gesbert, D., 2018, October. Learning to cooperate in decentralized wireless networks. In 2018 52nd Asilomar Conference on Signals, Systems, and Computers (pp. 281-285). IEEE.

[39] Lee, W., Kim, M. and Cho, D.H., 2018. Transmit power control using deep neural network for underlay device-to-device communication. IEEE Wireless Communications Letters, 8(1), pp.141-144.

[40] Liang, F., Shen, C., Yu, W. and Wu, F., 2019. Towards optimal power control via ensembling deep neural networks. IEEE Transactions on Communications, 68(3), pp.1760-1776.

Providing a Framework for Security Management in Internet of Things

XUE Zhen*, LIU Xingyue

School of Management
China University of Mining and Technology-Beijing
Beijing, 100083, China

Abstract—**With the advent of Internet of Things technology, tremendous changes are taking place. Perhaps what humans never even imagined will come in the near future, and just as the Internet surrounds all aspects of people's daily lives, intelligent objects will autonomously take over all aspects of people's lives. So far, a lot of research and development has been done in the field of Internet of Things, but there are still many challenges in this field. One of the most important challenges is the issue of security in the Internet of Things. Therefore, in this paper, while reviewing the requirements, models and security architectures of the Internet of Things, a framework for security management in the Internet of Things is proposed, which takes into account various aspects and requirements. The proposed framework uses various ideas such as cryptography, encryption, anomaly detection, intrusion detection, and behavior pattern analysis and can be considered as a basis for future research. The purpose of this research is to determine security requirements and provide a method to improve security management in the Internet of Things. Based on the tests, the proposed method is completely 100% resistant against data modification attacks. Against impersonation attacks up to 97% and against denial of service attacks up to 89% resistant detection accuracy.**

Keywords—Internet of Things (IoT); security management; security requirement; security model; security architecture

I. INTRODUCTION

Internet of Things (IoT) is a new concept in the world of technology and communication. The term Internet of Things was first used in 1999 by Kevin Ashton. He described a world in which everything, including inanimate objects, would have a digital identity of their own, allowing computers to organize and manage them. The Internet currently connects all people, but with the IoT, all things are connected. In other words, the Internet of Things is a modern technology in which any entity (human, animal, or objects) can send data through communication networks, either the Internet or intranet [1, 2].

According to Cisco research, the birth of the Internet of Things is estimated between 2008 and 2009. This research also shows that after 2008, the number of devices connected to the Internet is more than the total number of human population and their number is increasing exponentially. Thus, it is predicted that by 2030, the number of devices connected to the Internet will reach more than 30 billion IoT devices in the world [3, 4]. In research communities, the Internet of Things has been defined from different aspects and perspectives; therefore, several definitions have been provided for it. The difference of different views is rooted in the two terms "Internet" and

"things". The first (internet-oriented) perspective moves the IoT perspective to a network-oriented perspective, and the second (object-oriented) perspective shifts the main focus to public objects that must be integrated into a common framework [5]. But in the field of information and communication, the terms "Internet" and "things" when used together, evoke a new concept. From a semantic point of view, the Internet of Things means "a wide global network, based on standard communication protocols, of interconnected objects, each of which can be uniquely addressed" [4]. Right now, the Internet of Things has evolved from a random collection of purpose-built networks. For example, today's cars have multiple networks to control engine performance, safety features, communication systems, etc. Commercial and residential buildings also have different control systems for cooling, heating and air conditioning, telephone service, security and lighting. With the progress and evolution of the Internet of Things, these networks and many other networks will be connected to each other with additional management, analysis and security capabilities. This issue will increase the power of the IoT [2]. The potential possibilities offered by the Internet of Things make it possible to use it in many applications. But currently there is only a small part of its applications in society. Nevertheless, in many different contexts and environments, the use of new applications is likely to improve people's quality of life. As an example, can refer to its various uses at home, when traveling, and also its various uses at work, when sick. These environments are now equipped with various objects that have a basic intelligence. But in most cases, these objects do not have any communication capabilities. Creating the ability to communicate with each other, for these objects, and enriching the information received from the surrounding environment, implies that there are different environments in which a very wide range of applications can be deployed [3]. Many challenges have arisen with the emergence of the Internet of Things. These challenges can be examined from different aspects, including social and technical. As an example, can mention security challenges [5] and legal challenges [6]. In order to accept and spread the Internet of Things, its challenges must be well identified and overcome [7]. One of the limitations of this study is low power and few resources, which is not always possible due to the implementation of complex security algorithms on Internet of Things systems. Security management is one of the basic security challenges in the Internet of Things, and this issue is very important and necessary for the development of this emerging technology.

Despite this issue, the activities carried out so far have not been able to provide a complete and appropriate answer to this issue. Therefore, further research in this regard can contribute to the development of the frontiers of knowledge and finally, provide new horizons for the development and applications of the Internet of Things. Therefore, in this research, an effort will be made to propose a suitable solution for security management, while reviewing the literature on the subject and determining the security requirements of the Internet of Things, which will ultimately lead to improvements in this field and also to improve the level of security in the Internet of Things. Since in this paper focus is on the security challenges of the Internet of Things. In the continuation of this paper, security requirements, security models and security architectures of the Internet of Things are examined first, and then a framework for managing security in the Internet of Things is presented.

- The following questions are raised in this research:

- What are the security requirements of the Internet of Things?

- How to meet the security requirements of the Internet of Things?

- How to provide a method to improve security management in the Internet of Things?

In the following, Paper is configured as follows. In Section II, the requirements of the Internet of Things in different environments are discussed. Sections III and IV deal with security models and architectures in the Internet of Things. In Section V, the framework of the proposed method for managing security in the Internet of Things is presented, and finally, 6 is related to the conclusion.

II. RELATED WORK

Users can ensure that they have a secure, private place with the aid of data security. The centralized administration of data security across several Internet of Things levels is the subject of this study. The compliance standard, identity management, data management, policy engine, and audit reports are the key axes of this paper. Standards for security, hardware security, adherence to certain legislation, and flaws in the IoT layer are not demonstrated in this paper [8]. The Internet of Things has numerous advantages, but it also faces significant difficulties. The most crucial of them is raising an IoT network's vital connections' susceptibility to hacker attacks. There is currently no tested access control technique for designing security frameworks with device authentication. This issue has been resolved in the source [9] of the method of expanding a security framework with powerful and transparent security protection. These issues involve creating new authentication mechanisms, an access control subsystem, and extremely precise risk indicators, as well as looking at the security requirements of three scenarios, including IoT of the body, IoT for the home, and IoT for the hotel. The sample security framework provided provides us with some workable answers to some of the security issues associated with the Internet of Things. In source [10], researchers looked into the traits of seven well-known frameworks in an effort to make choosing a good framework for an industrial application easier. The choice

of a suitable framework is complicated by the growing number of frameworks and platforms that already exist and offer varying degrees of support for the aforementioned needs. It takes effort. The goal of this study is to highlight new developments in existing research and the most recent commercial Internet of Things frameworks, as well as to provide a technical comparison of their traits. An examination of the Internet of Things' current state and security issues is provided by source [11]. (IoT). Connecting people to anything and anywhere is one of the objectives of the IoT framework. Perception, Network, and Application layers make up the three-layer architecture that characterizes the Internet of Things. Therefore, in order to fulfill Internet of Things goals that have a high reliability factor, a variety of security principles must be applied in each layer. Only if the security concerns related to the IoT framework are resolved can its future be assured. Many researchers have attempted to design appropriate countermeasures in order to solve the unique security issues of IoT layers and devices. In order to secure the Internet of Things, this paper presents an overview of security principles, technological and security concerns, possible solutions, and future perspectives. IACS will undergo architectural modifications as a result of the acceptance and deployment of Internet of Things (IoT) technologies, including improved connectivity to industrial systems. The growth of IoT technologies in industrial systems has altered the IACS architecture by adding more connections inside the chips. Industry 4.0 and physical-cyber systems have both been looked at in the paper [12]. This paper defines the IIoT and examines the associated IoT subcategories. In order to count and characterize IIoT devices when examining system design and security threats and vulnerabilities, it offers an analytical framework for IIoT. The conclusion of this research identifies certain gaps in the literature. In [13], a framework for blockchain-based GSD public access control is given, offering users a platform for comprehensive GSD administration. First, users and GSDs issue visual identities based on the World Wide Web Consortium (W3C) Decentralized Identifiers standard (VIDs). Then, they added user and device authentication to the GSD-DIDs protocol. Finally, an integrated access control system for GSD was created based on the decentralization and non-tampering characteristics of blockchain. This system comprises registering, granting, and cancelling access privileges.Using this framework, users can accomplish decentralized, lightweight, and fine-grained GSD access control, according to the findings of the experiments. The research [14] introduces the information security risks the Internet of Energy faces, classifies errors, and studies the information security defenses of distributed energy stations operating in the Internet of Energy environment. This paper analyzes the countermeasures to protect the distributed energy station's information security, analyzes the security framework of the distributed energy station by building the network security framework, and presents the system architecture of the distributed energy station in the Internet of Energy environment in accordance with its network security features. The study [15] employed a systematic literature review (SLR) technique and went through all of the basic literature to look for structures and themes. These approaches were carried out in four steps: eligibility, screening, identification, and so on.

After 568 papers from reliable journals were evaluated, 260 papers and 54 reports were analyzed. Additionally, they used MAXQDA to conduct an analysis in which nodes and themes were first discovered. A qualitative model was created using MAXQDA after classification. The proposed paradigm has literary backing, making it beneficial for IoT consumers, developers, and IT managers. In order to balance security and other services in dense IoT, an ascending authentication framework (AAF) is presented in [16]. Through integrated keying and allocation, the proposed system offers a high level of authentication and user services. A framework that includes service providers and end users defines this procedure. In this authentication, key distribution follows a discrete assignment method while service authentication uses hyperelliptic curve cryptography. Regression learning is used to determine the discontinuity and continuity in authentication in this case. For future service releases, the level of authentication and security redemption is determined by a modification in the distribution function. As a result, the regression sequence is completely altered without the inclusion of any new keys. This lowers failure rates, response latency, service authentication time, and computational complexity. In order to study the use of IoT in diabetes monitoring, critical issues, a systematic literature analysis on the adoption of MHIoT for diabetes management was undertaken in [17] Methodology comprehensive literature review. Despite the larger benefits of MHIoT in such resource-constrained environments, key studies show that underdeveloped countries are falling behind. Findings indicate that the biggest barriers to MHIoT adoption for diabetes control are infrastructure costs, security concerns, and privacy concerns. As healthcare costs decline in a context of limited resources, the opportunities afforded by MHIoT exceed the constraints. To fully reap these benefits and handle problems, more study is needed on infrastructure needs and privacy issues. The paper [18] presents trustworthy algorithms for the Internet of Things' security framework. On actual marketing data for the bank obtained from the Cloud Internet of Things, these algorithms, which include Simple Bayes (NB), Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), ID3, and C4.5, are applied (CIoT). By identifying the key factors influencing success and evaluating the effectiveness of CIoT and SDM algorithms, this study aims to develop an effective framework for improving marketing campaigns for banks. This work is anticipated to boost scientific contributions in researching marketing information capabilities by merging SDM with CIoT. There are eight factors used to calculate how well SDM algorithms perform. Precision, balance precision, accuracy, root mean absolute error, recall, F1-score, and run time. Experimental results demonstrate the success of the suggested framework, which has higher accuracy and good performance. The findings demonstrated the importance of marketing strategies and customer service to a company's success and survival.

III. INTERNET OF THINGS SECURITY REQUIREMENTS

To implement security in any system in the Internet of Things, different areas such as environment, user, software, physical device, information and network should be considered in an integrated manner and in the form of a general solution. In other words, paying attention to each of these areas alone cannot be a guarantee for creating and establishing security in the Internet of Things, but requires an answer that considers all these areas in a comprehensive and integrated manner [19, 28]. To protect communications in IoT, sufficient assurance in terms of confidentiality, integrity, authentication and non-repudiation of information flow must be provided. The security of Internet of Things communication can be considered in the context of communication protocols. Also, this work can be done with external mechanisms and on the other side of communication. Other security requirements must also be considered for the Internet of Things. In particular, by considering the communication between objects, it can be concluded that attacks such as denial of service can target data availability. Therefore, there is a need for mechanisms to protect against such threats that guarantee the proper functioning of IoT communication protocols. Other security requirements of the IoT include privacy, anonymity, responsibility, and trust, which are necessary for the social acceptance of most future applications of the IoT [20, 35, 27]. Security key management is another important security requirement in the field of Internet of Things. In particular, due to the fact that the security keys are accessible to attackers, if an attacker gets hold of these keys, he will be able to recover all the information being sent by the relevant device. In addition to key management, there is a need for appropriate methods for generating, distributing, exchanging, updating and revoking keys. Appropriate encryption protocols and mechanisms should also be used. IoT devices are expected to operate unattended and deploy in unprotected environments. Subsequently, this makes IoT devices easily accessible to attackers and increases the risk of physical attacks on them. Therefore, there is a need for mechanisms to prevent interference and attacks related to hardware elements and embedded chips in IoT devices [21, 36]. Generally, the security requirements in the IoT can be examined and analyzed from different aspects. In this research, based on the studies, a complete list of security requirements in the field of IoT has been prepared and shown in Table I. In addition, each of these requirements are also placed in a special category. This table can be used to develop security models and architectures and propose new frameworks to improve security management in the IoT, which will ultimately lead to the improvement of the security of the IoT.

TABLE I.   SECURITY REQUIREMENTS IN THE FIELD OF IOT

Row	Requirement	Description	Category
1	confidentiality	Protection from unauthorized disclosure of information	Information flows
2	Integrity	Ensuring the correctness or correctness of the information	
3	Authentication	Ensuring that only authorized entities can access information	
4	Access level	Classification of access based on permissions	
5	non-denial	Ensuring that the sender or receiver does not deny sending or receiving information	
6	Privacy	Non-disclosure of sensitive information of people	social
7	Anonymity	The possibility of activity without the need to introduce the real identity	
8	Responsibility	Accepting the legal consequences of people's activities	
9	trust	Accepting people's approval	
10	Management of security keys	Including how to use public and private keys	encryption
11	Protocols	Guidelines and standards to be followed	
12	Security of hardware elements	Protection of hardware elements against various damages	Physical care
13	Embedded chip security	Protection against penetration and modification of embedded chips	

IV.   IoT Security Models

As shown in Fig. 1, in old information systems, there are three components of communication, control and calculation. There is a link between communication and control, which is sometimes called "information" and sometimes called "sensor network", which means collecting information from the control part and then putting it in the communication part. There is another link between communication and calculation, which is called "Internet" in China and "Cyber" in America [22, 29, 37].
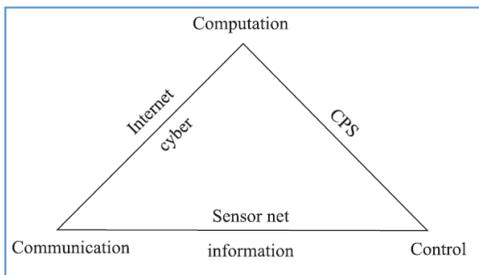
Fig. 1.   IoT Model [22,34].

Theoretically, there is no link between control and calculation because control depends on the intervention of people. With the advent of the IoT, the link between control and calculation has also been established, which has made the impact of calculation results on control possible. This link, called CPS, is a real bridge connecting the data layer to the physical layer. In a broad sense, the entire diagram of Fig. 2 is the IoT, which includes cyber. But in a smaller sense, IoT refers to the link between computation and control. With the emergence of this link, it is expected to face a new security challenge called "direct control. Although the control should be done without human intervention, a security control should be used to prevent hackers from interfering. Therefore,

considering this issue, the security model of the IoT according to Fig. 2 requires a security control between calculation and control [22, 33].

In Fig. 3, a model called U2IoT is presented, which is a heterogeneous system, including Unit IoT and Ubiquitous IoT. This model uses social factors for security layers and adds intelligence and adaptability to security requirements. At the core of this model, the IoT Unit is similar to the human neural network, which refers to a primitive cell that provides responses for specific applications. Ubiquitous IoT includes industrial, local, national and public Internet of Things, which is the result of the integration of several IoT Units and is similar to the social organization framework. Specifically, Unit IoT includes IoT networks, sensors, distributed control nodes, management and centralized data center (M&DC), and Ubiquitous IoT also includes industrial data center and management (iM&DC), local data center and management (lM&DC), and National Data and Management Center (nM&DC) [23, 30, 39].

Fig. 4 shows the security layers and requirements of the U2IoT model. In this figure, the x-axis is related to U2IoT, the y-axis is related to security requirements, and the z-axis is related to security layers.
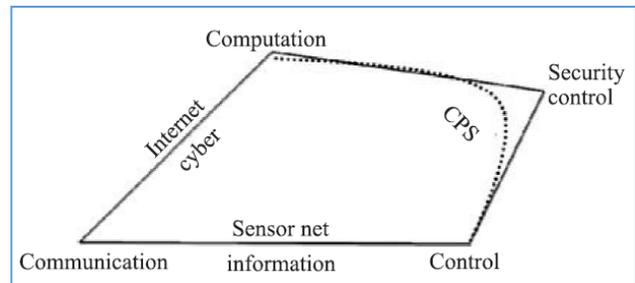
Fig. 2.   IoT Security Model.
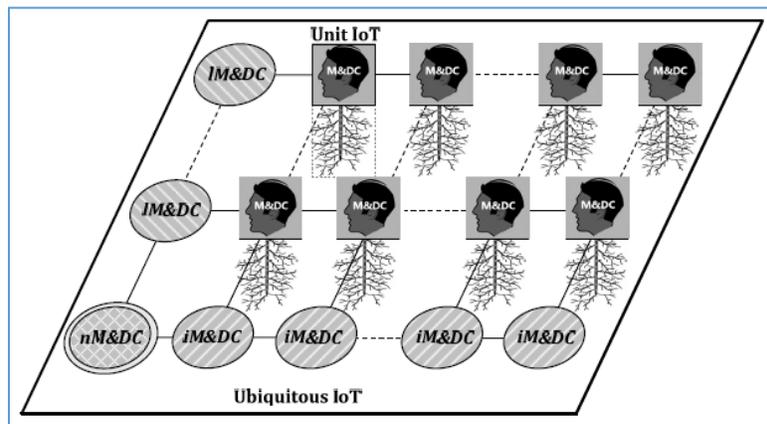
Fig. 3.   U2IoT Security Model [23].
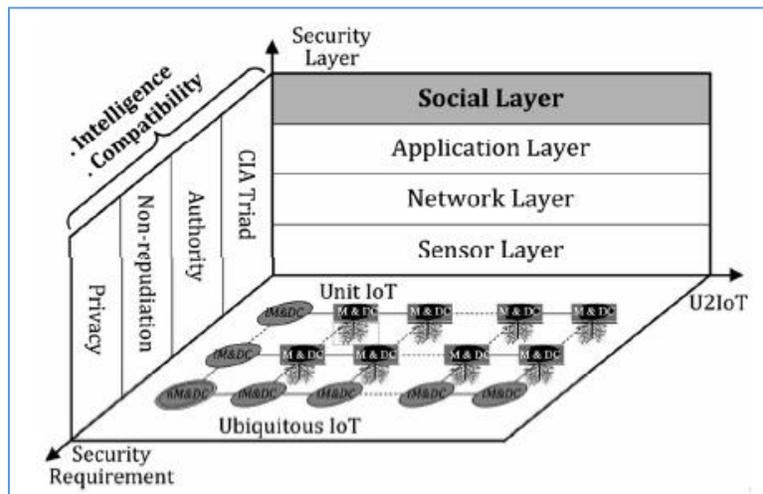
Fig. 4.   Security Layers and Requirements of the U2IoT Model[23].

In this model, security is divided into four layers: sensor, network, application, and social. In the sensor layer, entities are observed to extract information and discover semantic resources. Also, special techniques are used to detect the effective Integrity and adaptation of imprecise information interactions. The network layer includes network interfaces, communication channels, network management, information storage, and intelligent processing. In this layer, centralized, distributed, and hybrid network topologies are used to help monitor and maintain the network configuration in real-time. In this layer, the safe transfer of information is ensured by using algorithms for coding, extracting, integrating, reconstructing, exploring and aggregating data. The main function of this layer is to transfer and process the information obtained from the sensor layer and realize data exchange between high-scale heterogeneous networks.

Application layer provides functionality for specific applications and provides implicit interfaces to the infrastructure to perform testing, monitoring, and auditing applications. Standard protocols and service composition technologies have been applied to realize integration between heterogeneous distributed networks and its applications such as support monitoring, intelligent scheduling, intelligent search, and cloud computing. Such applications must be adapted for

dynamic environments. The social layer includes social features in the U2IoT model. This layer is basically provided for communication between objects and other support networks to establish correlation between cyber characteristics and its corresponding profile in social networks. Corresponding social characteristics are assigned to each entity and hierarchical management and data centers implement general security considerations. In the social layer, various interfaces are available to entities that act on their cyber existence and control their behavior. Meanwhile, in this layer, other social combinations are also considered, such as property control management, social relationship modeling, and entity behavior formulation. The security requirements considered in this model are three CIA (Confidentiality, Integrity and Availability), Authority, Non-repudiation, and Privacy, which bring reliable security along with privacy protection [23].

V.   IoT SECURITY ARCHITECTURES

So far, various security architectures have been proposed for the IoT. The main difference between the presented architectures is related to the number of layers and also the difference in performance of each layer in the IoT architecture. Fig. 5 presents a four-layer architecture including sensing, network, support, and application layers. The sensory layer

collects all the information through the physical equipment and identifies the physical world. The information collected by this layer includes object characteristics, environmental conditions, etc. Physical equipment also includes all types of sensors, including GPS and RFID, which are responsible for collecting data and information. In fact, sensors are the key component of this layer that captures the physical world and displays it in the digital world. The network layer is responsible for the reliable transmission of information from the sensory layer, primary processing, classification and combination of information. In this layer, information transmission is based on several main networks, including the mobile Internet communication network and satellite networks. Also, network infrastructure and communication protocols are also required to exchange

information between devices. The support layer establishes a platform for reliable support of the application layer. On this support platform, all intelligent computing capabilities will be organized through grid network and cloud computing. This layer plays the role of combining the application layer upwards and the network layer downwards. The application layer provides services such as personal information services, smart transfers, and environmental monitoring according to the needs of users. In fact, users will access the Internet of Things through application layer interfaces using televisions, personal computers, mobile devices, and the like. In the presented model, network security management is integrated with all the different layers of the architecture and manages them at different levels [24, 31, 32].
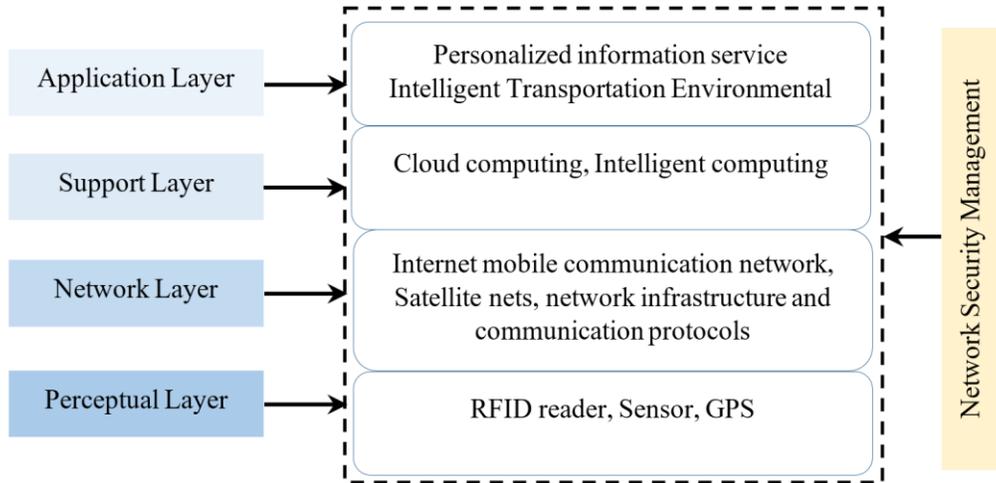
Fig. 5.   Four-Layer Security Architecture of IoT [24].

Fig. 6.   Three-Layer Security Architecture of IoT[25].

Fig. 6 shows a three-layer architecture including the sensing layer, the transmission layer, and the application layer. This architecture has divided each layer into several sub-layers. The sensory layer is divided into two sublayers of sensory nodes and sensory network. The transmission layer is divided into three sublayers: access network, central network, and local network. The application layer is also divided into two application support sub-layers and Internet of Things applications. Each layer in this architecture has its own technical support. These technologies play an irreplaceable role at all levels. These techniques are more or less related to the range of existing issues that cause insecurity, privacy and other data security issues [25].

The IoT must ensure the security of all layers. In addition, the security of the Internet of Things should include the security of the entire system. The sensory layer includes RFID security, wireless sensor network security, RSN security, and anything else related to this layer. The transport layer includes access network security, central network security, and local network security. For example, 3G access network security, Ad-Hoc network security, WiFi security, and GPRS security are related to the access network substrate. Internet security is related to the security of the central network sublayer and local network security is related to the local network sublayer in the transmission layer. The application layer also includes application support and specific applications of the Internet of Things. Security in the support layer includes middleware technology security, service support platform security, cloud computing platform security, information development platform security, and other support platform; security must ensure the security of all layers. In addition, the security of the Internet of Things should include the security of the entire system. The sensory layer includes RFID security, wireless sensor network security, RSN security, and anything else related to this layer. The transport layer includes access network security, central network security, and local network security. For example, 3G access network security, Ad-Hoc network security, WiFi security, and GPRS security are related to the access network substrate. Internet security is related to the security of the central network sublayer and local network security is related to the local network sublayer in the transmission layer. The application layer also includes application support and specific applications of the Internet of Things. Security in the support layer includes middleware technology security, service support platform security, cloud computing platform security, information development platform security, and other support platform security. IoT applications in different industries have different application requirements. For example, smart home security requirements are different from telemedicine security requirements or smart traffic security requirements, and they should be specified according to the type of application of the relevant requirements [25].

Fig. 7 shows the security architecture in three layers of sense, network and application and middleware. In this form, middleware and application are considered in one layer. In the sense layer, hashing algorithms, encryption mechanisms, anonymity methods, risk assessment and intrusion detection are placed. In the network layer, peer-to-peer encryption, routing security, data integrity and intrusion detection are considered. In the application and middleware layer, integrated identity recognition, cryptographic mechanisms, firewalls, risk assessment, and intrusion detection are placed [26].
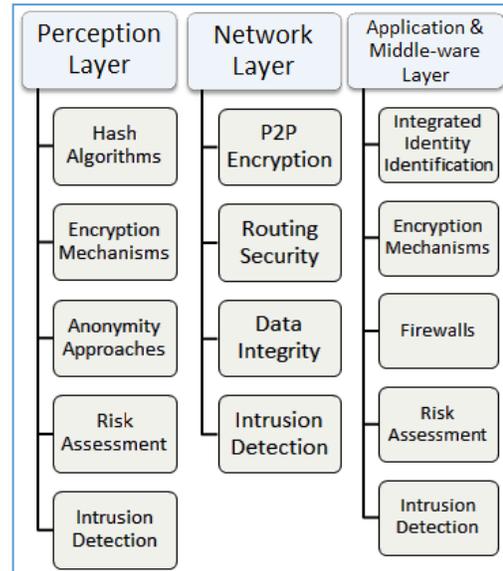
Fig. 7. IoT Security Architecture [26].

In Fig. 8, a security architecture for the U2IoT model is presented. In this architecture, three parts of information security, physical security and management security are considered. Information security covers security layer and security requirement. Physical security covers external content and core infrastructure, including intrinsic security and adaptive security. Security management includes application requirements, national/local/industry regulations, and international policy and standards. In this architecture, the connection of the U2IoT model with the security architecture at the cyber, physical and social levels is considered [38, 40].
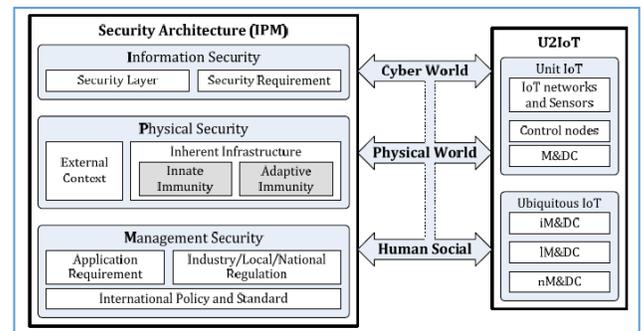
Fig. 8. Security Architecture of the U2IoT Model [23].

VI. THE PROPOSED FRAMEWORK FOR MANAGING SECURITY IN THE INTERNET OF THINGS

In order to create a comprehensive security in the Internet of Things, it is necessary to ensure the security of all the entities of the Internet of Things at different levels from different aspects. For this, it is necessary to first define the different entities of the Internet of Things. Generally, the following entities can be imagined in the Internet of Things:

- Objects: are the main components of the Internet of Things.

- Communication networks: networks that provide communication between objects.

- Internet: establishes the global connection of different communication networks.

- Users: a set of human agents who are the main users of the Internet of Things.

- Applications: Manage objects to meet users' needs.

A. Objects

There are different types of objects, and according to the type of each object, its security requirements will also be different. Generally, objects can be classified into three different categories based on how they interact with their surroundings:

- Sensor: They collect information from the surrounding environment. (example: camera).

- Stimulator: execute commands in their surroundings. (example: door lock).

- Combined: perform both sensor and actuator functions. (example: robot).

It is obvious that the security requirements of a sensor object that only collects information from its environment will have important differences with the security requirements of an actuator object that only executes commands in its environment, and a hybrid object must also meet the security requirements of two types benefited previously. For greater simplicity, a composite object that is both a sensor and a trigger is conceptually divided into two different parts, and security requirements can be entered for each part separately based on its functionality. Therefore, from now on, only the first two types will be considered. A sensor must be able to evaluate and validate the accuracy of its observations, and only if this evaluation and validation is correct, it can encrypt its observations using its private key and propagation the encrypted packet in the communication network for predefined purposes. With this, the requirements of comprehensiveness and confidentiality will be met. Fig. 9 shows the proposed model for sensor security in the Internet of Things.

According to Fig. 9, in the proposed model, two modules of physical care and intrusion detection are also used to increase the security of sensors. The physical guard module ensures the security of the hardware elements and chips embedded in the sensor, and the intrusion detection module is an intrusion detection system embedded in the sensor, which investigates the intrusion of hackers into the sensor. If a sensor has been compromised, it will be notified to other related departments using cryptography. By doing this, other entities can isolate the compromised sensor before any malicious action by the attacker and prevent any further damage. The use of cryptography in sensors has other advantages, for example, any suspicious activity can be reported. Therefore, a secret communication layer is created between different entities and the sensor, which makes other entities aware of the danger

before any serious incident and before they succeed in attacking the sensor. All the entities that need the information of a sensor can decrypt the information sent by the sensor only if they have the public key of that sensor. In addition, the communication network is responsible for checking and authenticating entities that request to use sensor information. In fact, the communication network must guarantee that only authorized entities have access to sensor information. This is done through a predefined protocol. This protocol is defined by each application separately and exclusively. This will satisfy the authentication and access level requirements .The commands that are given to an actuator for execution must also be previously encrypted using special keys, and the actuator must also decrypt the commands using the key it has. In addition, a stimulus must be able to evaluate and verify the correctness of the commands it has received for execution, and only if this evaluation and verification is correct, it can execute the received commands in the external environment. Fig. 10 shows the proposed model for actuator security in the Internet of Things. In this model, as well as the proposed model for the security of sensors, the physical surveillance module and the intrusion detection module are used. Since an actuator, unlike a sensor, cannot send information to other entities, therefore, it cannot inform other entities of the dangers that have happened to it in a normal way, like a sensor. Therefore, an innovative method is used to inform other entities of dangerous conditions and intrusion into a trigger. In this method, a certain behavior is hidden in the execution of the commands of an actuator, and other entities, by observing this hidden behavior in the executive behavior of the actuator, will notice its critical condition and take the necessary measures to minimize possible damages. Although this idea adds a hardware overhead to the system, the resulting advantage is also significant. In other words, for each actuator, one or more sensors, which call monitoring sensors, are assigned to check the behavior of the stimulus, and in case of observing an abnormal behavior or a hidden behavior that is already defined in the intermediate protocol, the necessary notification done and the required security measures will be taken.

It should be noted here that all the explanations that were previously provided about how to manage keys and predefined protocols for sensors are now also true about how to manage keys and define protocols for operators, and repeating it is avoided.

B. Communication Networks

Communication networks are networks that establish communication between objects. In order to establish secure communication between objects, it is necessary to pay attention to logical communication in addition to physical communication. As an example, suppose that in a smart home there are security cameras and electronic alarms in a communication network. Now, is it reasonable that a security camera can issue commands that lead to the opening of the electronic lock on the door or not? Therefore, a communication network should be able to perform certain security controls based on a series of pre-defined protocols and policies, in addition to establishing the ability to exchange information between different objects and creating a secure communication channel. As mentioned earlier, these protocols and policies are

defined by the respective applications. Fig. 11 shows the proposed model for a communication network in the Internet of Things.

As can be seen, in this model, an anomaly detection module is used for network management. This module can detect anomalies based on communication channel activities, as well as pre-defined applications, protocols, and policies, and report them to the network management unit. The network management unit will also prevent unnecessary and suspicious communication if necessary. In this way, a double intelligence will be created in the communication network, which will ultimately improve the security of the Internet of Things.

C. Internet

The Internet, as the main infrastructure of the Internet of Things, establishes the global connection of various communication networks. In this way, access to objects from

anywhere in the world will be possible. This issue will make the requirement of accessibility, which is one of the main requirements of the Internet of Things, be met. Since the communication of objects in the proposed model has a hierarchical structure, therefore, in the proposed model, there is a very good savings in assigning unique identifiers to objects. In other words, a unique local identifier is assigned to each object in the respective communication network, and a unique global identifier is assigned to each communication network. In this way, by using the sum of two network identifiers and the object, a global unique identifier will be obtained, which makes the addressing of each object unique in the world. In addition, more hierarchical levels can be used for this purpose. In this way, a suitable answer will be created for one of the basic challenges of the Internet of Things, which is related to assigning a unique identifier to each object.
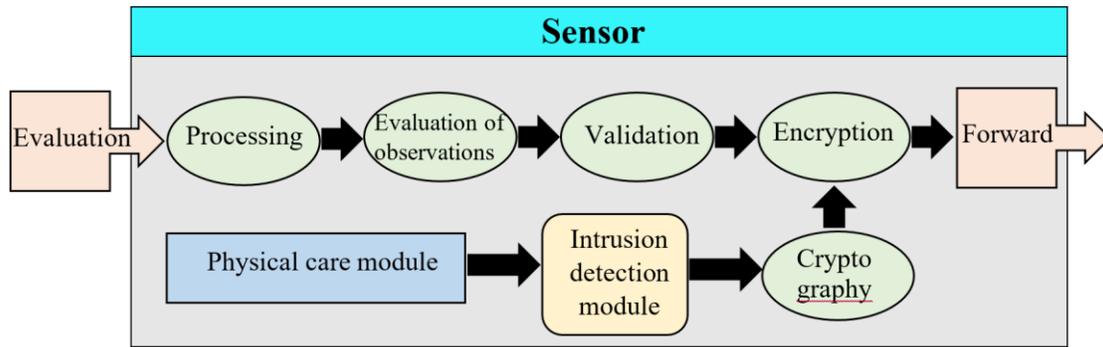
Fig. 9. Proposed Model for the Security of Sensors in the Internet of Things.
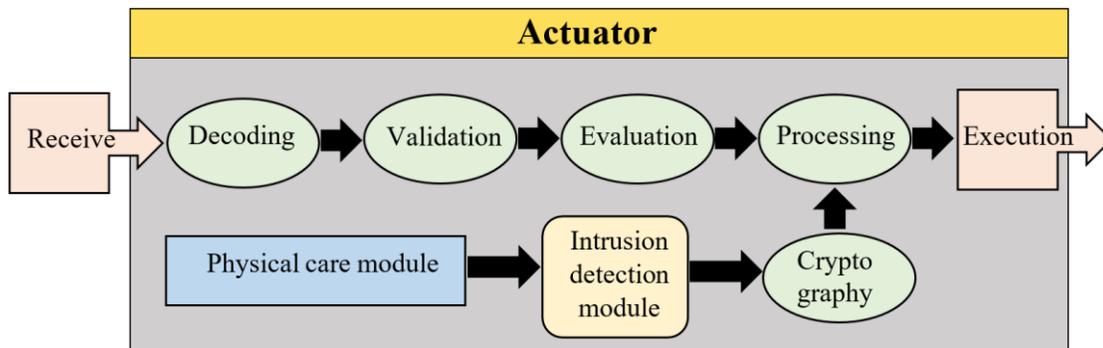
Fig. 10. Proposed Model for the Security of Actuator in the IoT.
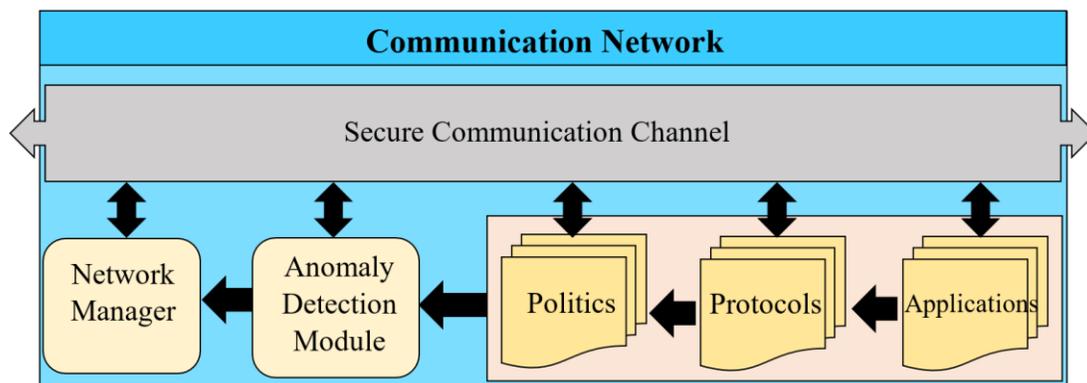
Fig. 11. Proposed Model for Communication Network in IoT.
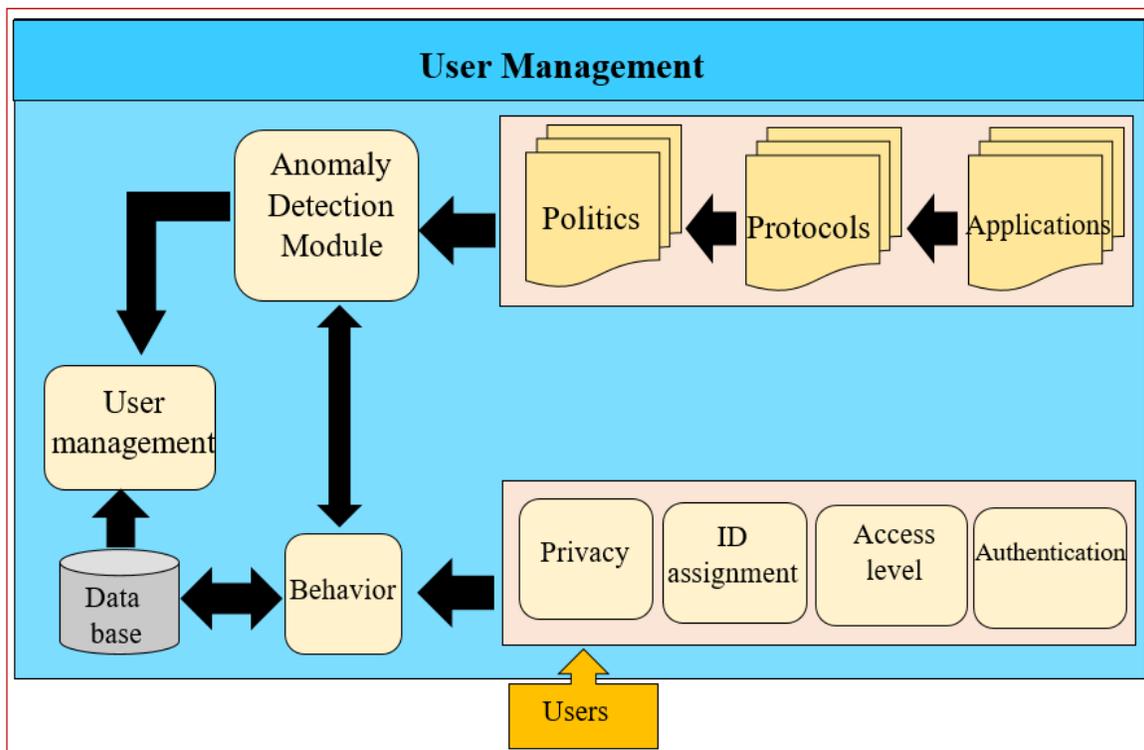
Fig. 12. Proposed Model for Managing Users in the IoT.

D. *Users*

Users are a set of human agents who use the Internet of Things to perform their desired applications. Generally, users can be divided into three different categories:

- Users: there are people who only use the Internet of Things at the user level.

- Managers: There are people who define applications and protocols at the management level.

- Attackers: unauthorized people who put themselves in the place of one of the above two groups and implement their goals.

User management is one of the most sensitive activities that must be done in the Internet of Things. In this regard, there are many security requirements such as authentication, access level determination, privacy, anonymity, responsibility and trust that must be covered. Fig. 12 shows the proposed model for user management. In this form, the components of authentication, access level determination, ID assignment, and privacy work in parallel, and an additional component called "behavior check" examines the behavior of users and stores the necessary information in the behavior database. This information is used along with applications, protocols and policies to detect anomalies in user behavior.

As shown in the Fig.12, the main idea in the proposed model is to use the behavioral characteristics of users to detect anomalies, and finally, by using the obtained information as well as the information in the database, user management operations are performed. This operation can include activating or deactivating users as well as issuing or not issuing access licenses for them.

E. *Applications*

Applications are responsible for managing objects to meet the needs of users. In fact, applications determine how objects should be related to each other and what mission they should perform. Applications also determine what activities different users can perform in the Internet of Things. In fact, applications provide a set of protocols, each of which provides a set of policies that are used in all the models described so far. As an example, consider the application of a smart air conditioning system.

In this application, the temperature and humidity sensors as well as the electronic key actuators of the cooling and heating system are defined as endpoints. The final task of this application is to adjust the temperature of the environment based on the needs of the users. Therefore, it is necessary for users to be able to determine the optimal minimum and maximum temperature. With this explanation, this particular application provides a set of protocols that describe how sensors, actuators, and users communicate. In addition, each of these protocols also provides a set of policies. For example, setting the minimum temperature by the user leads to a protocol that shows how the user interacts with the relevant data and changes it. For this purpose, a set of policies is considered. As an example, assume a policy where the user is allowed to change only one temperature data grade per step. In fact, this policy ensures that there are no drastic changes in sensitive data at once. In addition, such a policy also provides the ability to observe and track user behavior. Fig. 13 shows how applications, protocols and policies interact.
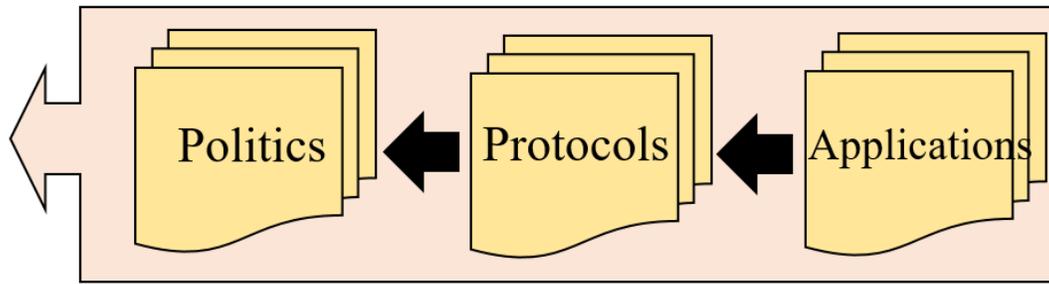
Fig. 13. Proposed Model for the Interaction of Applications, Protocols and Policies in the IoT.

VII. EVALUTION AND RESULTS

To evaluate the proposed method, after simulation, various scenarios are executed and the success or failure of the system in detecting and responding to an attack is measured as a measure of accuracy. Of course, to create diversity in the tests, a random number has been used, which generates identifiers for the objects that are the target of the attack, and the desired scenario is executed on it.

Also, to achieve stable and reliable results, according to the number of sensors and actuators, each experiment with random object ID is repeated 100 times and the average accuracy obtained is shown.

As can be seen from the above graphs (Fig. 14 and Fig.15), in general, the accuracy of the proposed method is significant. The point that should be noted here is that based on the tests, the proposed method is completely (100%) resistant against data modification attacks. It also performs well against impersonation attacks (97%). But against denial of service attacks, it has a lower detection accuracy (89%). This issue also seems quite logical because one of the prominent features of the proposed method is the use of behavioral and cryptographic parameters, which makes the resistance of the proposed method against change and impersonation attacks much higher than other attacks.
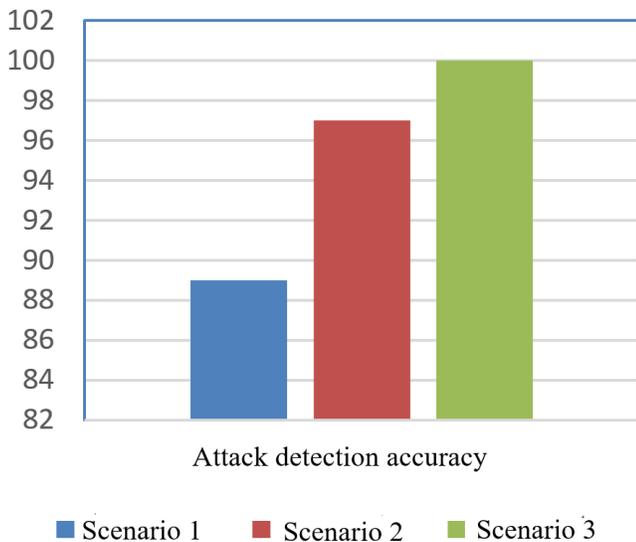
Fig. 14. Comparing the attack Detection Accuracy in Different Scenarios by Implementing the Proposed Method.
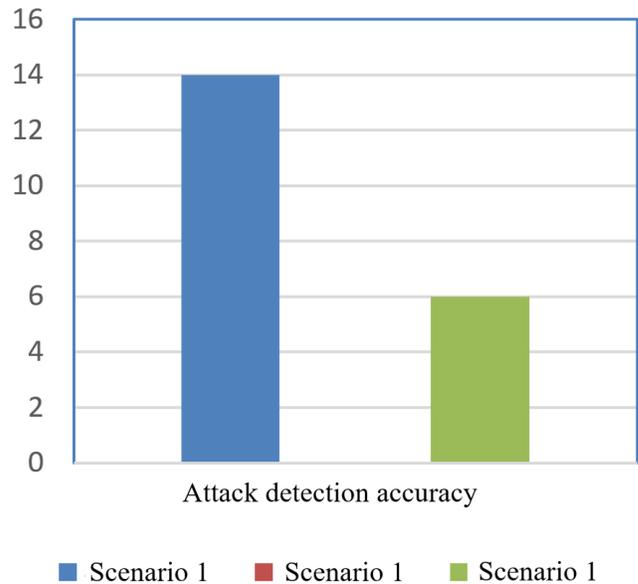
Fig. 15. Comparison of attack Detection Accuracy in Different Scenarios without Implementing the Proposed Method.

Contrary to this issue, if the proposed method is not used and relies only on the application and the centralized control system, the accuracy of attack detection will also be greatly reduced. In this case, the accuracy of detecting denial of service attacks is higher than others (14%), followed by the accuracy of detecting modified attacks (6%) and finally impersonation attacks cannot be detected at all (0%). It should be noted that these results are completely logical considering the way the mentioned system is implemented and its lack of use of impersonation detection mechanisms. Table II compares the mentioned results numerically.

TABLE II. NUMERICAL COMPARISON OF TEST RESULTS

Scenario	Attack detection accuracy without implementing the proposed method	Attack detection accuracy by implementing the proposed method
1	14%	89%
2	0%	97%
3	6%	100%

VIII. CONCLUSION

Until now, various activities have been carried out in line with the existing security challenges for the Internet of Things, but a comprehensive and integrated solution has not been provided for it, and more research is needed in this field. The

main activities carried out in the field of Internet of Things security are related to determining security requirements and providing models and security architecture. Generally, the security requirements in the field of Internet of Things can be placed in different categories, including data and information flows, encryption, social and physical care. Generally, the security models and architectures presented for the Internet of Things consider different levels and apply security issues at different levels separately. Therefore, there is a need to develop new models and architectures that cover all the security aspects of the Internet of Things in an integrated manner. For this purpose, the security requirements of the Internet of Things should be paid attention to from various aspects, and the connection between different levels should be fully seen in the security model of the Internet of Things. Although such work requires multiple activities and an optimal answer cannot be achieved at once. In this Paper, a new framework for managing security in the Internet of Things has been presented, which has presented solutions to improve security in the Internet of Things in a hierarchical manner and at different levels. The main focus of this framework has been relying on solutions for anomaly detection as well as intrusion detection, and in this regard, behavioral characteristics have been used, as well as data encryption of sensors and behavioral activities of actuator. What is specifically paid attention to in this Paper is the issue of notifying other departments by using encryption of the correctness of the operation or the possibility of detection of intrusion into the sensors and actuators, and based on the anomalies reported at the levels of the communication network and also manage users to make the necessary decisions. In other words, in the proposed model, the communication network is in charge of managing the network and it does this by using the protocols and policies specified by the applications and based on the observed anomalies. The user management level also uses the behavioral characteristics of users to detect anomalies and takes the necessary decisions if abnormal behaviors are observed. In other words, at this level, by storing user behaviors, a behavioral database will be created, which will be the main basis for subsequent decisions. This database stores user behaviors like a long-term memory and can be used in various ways. For example, this database can be used to track user activities and distinguish real users from attackers who intend to infiltrate the system.

REFERENCES

[1] Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Poor, H. V. (2021). Federated learning for internet of things: A comprehensive survey. IEEE Communications Surveys & Tutorials, 23(3), 1622-1658.

[2] Shi, V. T., & Nhg, D. R. (2022). Channel Estimation Optimization Model in Internet of Things based on MIMO/OFDM with Deep Extended Kalman Filter. *Advances in Engineering and Intelligence Systems*, *1*(02).

[3] Evans, D. (2011). The internet of things: How the next evolution of the internet is changing everything. CISCO white paper, 1, 1-11.

[4] Thilakarathne, N. N., Kagita, M. K., & Priyashan, W. D. (2022). Green internet of things: The next generation energy efficient internet of things. In Applied Information Processing Systems (pp. 391-402). Springer, Singapore.

[5] Azad, F. A., Rad, S. A., & Arashpour, M. (2022). Back-stepping control of delta parallel robots with smart dynamic model selection for construction applications. Automation in Construction, 137, 104211.

[6] Mozaffari, H., & Houmansadr, A. (2020, January). Heterogeneous private information retrieval. In Network and Distributed Systems Security (NDSS) Symposium 2020.

[7] Honar Pajooh, H., Rashid, M., Alam, F., & Demidenko, S. (2021). Multi-layer blockchain-based security architecture for internet of things. Sensors, 21(3), 772.

[8] Balachandar, S., & Chinnaiyan, R. (2019). Centralized reliability and security management of data in internet of things (IoT) with rule builder. In *International Conference on Computer Networks and Communication Technologies* (pp. 193-201). Springer, Singapore.

[9] Huang, X., Craig, P., Lin, H., & Yan, Z. (2016). SecIoT: a security framework for the Internet of Things. *Security and communication networks*, *9*(16), 3083-3094.

[10] Paniagua, C., & Delsing, J. (2020). Industrial frameworks for internet of things: A survey. *IEEE Systems Journal*, *15*(1), 1149-1159.

[11] Hosseini, S., & Khamesee, M. B. (2021). Modeling And Simulation And Imaging Of Blood Flow In The Human Body. NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO, 13235-13244.

[12] Boyes, H., Hallaq, B., Cunningham, J., & Watson, T. (2018). The industrial internet of things (IIoT): An analysis framework. *Computers in industry*, *101*, 1-12.

[13] Tan, L., Shi, N., Yu, K., Aloqaily, M., & Jararweh, Y. (2021). A blockchain-empowered access control framework for smart devices in green internet of things. *ACM Transactions on Internet Technology (TOIT)*, *21*(3), 1-20.

[14] Zhang, J. (2021). Distributed network security framework of energy internet based on internet of things. *Sustainable Energy Technologies and Assessments*, *44*, 101051.

[15] Ali, A., Mateen, A., Hanan, A., & Amin, F. (2022). Advanced Security Framework for Internet of Things (IoT). *Technologies*, *10*(3), 60.

[16] Panneerselvam, N., & Krithiga, S. (2022). A novel security framework for densely populated Internet of Things users in pervasive service access. *Computer Communications*, *184*, 86-95.

[17] Mutunhu, B., Chipangura, B., & Twinomurinzi, H. (2023). A Systematized Literature Review: Internet of Things (IoT) in the Remote Monitoring of Diabetes. In *Proceedings of Seventh International Congress on Information and Communication Technology* (pp. 649-660). Springer, Singapore.

[18] Liu, Y., Alzahrani, I. R., Jaleel, R. A., & Al Sulaie, S. (2023). An efficient smart data mining framework based cloud internet of things for developing artificial intelligence of marketing information analysis. *Information Processing & Management*, *60*(1), 103121.

[19] Weber, R. H. (2010). Internet of Things–New security and privacy challenges. Computer Law & Security Review, 26(1), 23-30.

[20] Mozaffari, H., Houmansadr, A., & Venkataramani, A. (2019, December). Blocking-Resilient Communications in Information-Centric Networks using Router Redirection. In 2019 IEEE Globecom Workshops (GC Wkshps) (pp. 1-6). IEEE.

[21] Chandrashekhar, A. M., Chaitra, K. V., & Koti, S. (2016). Security Fundamentals in Internet of Things. International Journal of Research, 3(01), 854-860.

[22] Granjal, J., Monteiro, E., & SaSilva, J. (2015). Security for the internet of things: a survey of existing protocols and open research issues. Communications Surveys & Tutorials, IEEE, 17(3), 1294-1312.

[23] Aufner, P. (2020). The IoT security gap: a look down into the valley between threat models and their implementation. International Journal of Information Security, 19(1), 3-14.

[24] Azimirad, V., Sotubadi, S. V., & Nasirlou, A. (2021, November). Vision-based Learning: a novel machine learning method based on convolutional neural networks and spiking neural networks. In *2021 9th RSI International Conference on Robotics and Mechatronics (ICRoM)* (pp. 192-197). IEEE.

[25] Wang, B., Liu, X., & Zhang, Y. (2022). Internet of things. In Internet of Things and BDS Application (pp. 71-127). Springer, Singapore.

[26] Ning, H., & Liu, H. (2012). Cyber-physical-social based security architecture for future internet of things. Advances in Internet of Things, 2(01), 1.

[27] Radhoush, S., Bahramipanah, M., Nehrir, H., & Shahooei, Z. (2022). A Review on State Estimation Techniques in Active Distribution Networks: Existing Practices and Their Challenges. *Sustainability*, *14*(5), 2520.

[28] Jing, Q., Vasilakos, A. V., Wan, J., Lu, J., & Qiu, D. (2014). Security of the internet of things: Perspectives and challenges. Wireless Networks, 20(8), 2481-2501.

[29] Fortino, G., Guerrieri, A., Savaglio, C., & Spezzano, G. (2022). A Review of Internet of Things Platforms through the IoT-A Reference Architecture. In International Symposium on Intelligent and Distributed Computing (pp. 25-34). Springer, Cham.

[30] Trik, M., Molk, A. M. N. G., Ghasemi, F., & Pouryeganeh, P. (2022). A Hybrid Selection Strategy Based on Traffic Analysis for Improving Performance in Networks on Chip. *Journal of Sensors*, *2022*.

[31] Mohamed, A. M. A., & Hamad, Y. A. M. (2020, September). IoT security: review and future directions for protection models. In 2020 International Conference on Computing and Information Technology (ICCIT-1441) (pp. 1-4). IEEE.

[32] Abiodun, O. I., Abiodun, E. O., Alawida, M., Alkhawaldeh, R. S., & Arshad, H. (2021). A review on the security of the internet of things: challenges and solutions. Wireless Personal Communications, 119(3), 2603-2637.

[33] Yao, X., Farha, F., Li, R., Psychoula, I., Chen, L., & Ning, H. (2021). Security and privacy issues of physical objects in the IoT: Challenges and opportunities. Digital Communications and Networks, 7(3), 373-384.

[34] Trik, Mohammad, Amir Massoud Bidgoli, Hossein Vashani, and Saadat Pour Mozaffari. "A new adaptive selection strategy for reducing latency in networks on chip." *Integration* (2022).

[35] Bhatt, S., & Ragiri, P. R. (2021). Security trends in Internet of Things: A survey. SN Applied Sciences, 3(1), 1-14.

[36] Vahidi Farashah, M., Etebarian, A., Azmi, R., & Ebrahimzadeh Dastjerdi, R. (2021). An analytics model for TelecoVAS customers' basket clustering using ensemble learning approach. *Journal of Big Data*, *8*(1), 1-24.

[37] Trik, M., Mozaffari, S. P., & Bidgoli, A. M. (2021). Providing an adaptive routing along with a hybrid selection strategy to increase efficiency in NoC-based neuromorphic systems. *Computational Intelligence and Neuroscience*, *2021*.

[38] Ali, A., Mateen, A., Hanan, A., & Amin, F. (2022). Advanced Security Framework for Internet of Things (IoT). Technologies 2022, 10, 60.

[39] Mozaffari, H., & Houmansadr, A. (2022). E2FL: Equal and Equitable Federated Learning. *arXiv preprint arXiv:2205.10454*.

[40] Sun, J., Zhang, Y., & Trik, M. (2022). PBPHS: A Profile-Based Predictive Handover Strategy for 5G Networks. *Cybernetics and Systems*, 1-22.

Novel Strategies Employing Deep Learning Techniques for Classifying Pathological Brain from MR Images

Mitrabinda Khuntia[1], Prabhat Kumar Sahu[2], Swagatika Devi[3]

Dept. of Computer Science and Engineering, ITER, S'O'A University, Bhubaneswar, India[1,2,3]

Abstract—**Brain tumors are the most widespread as well as disturbing sickness, among a very precise expectancy of life almost in their serious structure. As a consequence, therapy planning is a critical component in enhancing the characteristics of the patient's life. Image modalities like computed tomography (CT), magnetic resonance imaging (MRI), along with ultrasound images are commonly used to assess malignancies in the brain, breast, etc. MRI scans, in evidence, are employed in this study to identify the brain tumors. The application of excellent categorization systems on Magnetic Resonance Imaging (MRI) aids in the accurate detection of brain malignancies. The large quantity of data produced through MRI scan, on the other hand, renders physical distribution of tumor and non-tumor in a given time period impossible. It does, however, come with major obstruction. As a consequence, in order to decrease human mortality, a dependable and automated categorizing approach is necessary. The enormous geological and anatomical heterogeneity of the environment surrounding the brain tumor makes automated classification of brain tumor a difficult undertaking. This paper proposes a classification of Convolutional Neural Networks (CNN) for automated brain tumour diagnosis. To study as well as compare the findings, other convolutional neural network designs such as MobileNet V2, ResNet101, and DenseNet121 are used. Small kernels are employed to carry out the more intricate architectural design. This experiment was carried out using Python and Google Colab. The weight of a neuron is characterized as minute.**

Keywords—*Brain tumor; CT; MRI; CNN; MobileNet V2; ResNet101; DenseNet121*

I. INTRODUCTION

The process of changing an input image into an output image or image attributes is known as image processing. The primary goal of all image processing techniques is to make it simpler to visually recognize the image or item under investigation. Methods that incorporate quantitative analysis will augment the traditional visual examination of images. Anatomical segmentation of regions of interest (ROI), such as distinguishing a volume of abnormal tissue from a backdrop of normal tissue, is an important stage in the image processing pipeline. This will enable statistical examination of aspects that are not evident to human vision. Image segmentation is critical in the realm of image processing. Segmentation is vital in medical imaging for feature extraction, image measurements, and image presentation [1]. Image segmentation is the partitioning or segmentation of a digital image into comparable sections with the primary goal of simplifying the image under review into something more understandable and simpler to visually evaluate. Image segmentation techniques are classed as thresholding, region-based, supervised, or unsupervised.

The central nervous system sends sensory information and the activities that go with it all across the body [2]–[4]. Brain is most intricate organism in human body. It serves various roles and controls the operations of the body's other systems. The brain's frontal lobe aids in motor control, problem solving, and decision making. The parietal lobe regulates body posture. Memory and hearing are handled by the temporal lobe, whereas vision is handled by the occipital lobe. The cerebral cortex is a grey substance that surrounds the cerebrum and is made up of several neurons. The cerebellum regulates different voluntary movements in living animals equipped with nervous system. When brain cells grow abnormally and are not effectively managed, a brain tumor can form. A tumor is an abnormal collection of cells that is generated by uncontrolled cell division. It has been discovered that all tumor variations are not malignant. Brain tumors are the world's tenth largest cause of mortality. Cancer, in its most basic sense, refers to malignant tumors rather than benign ones. There are two categories of brain tumours: grades 1 and 2 (low) and grades 3 and 4 (high).

Computer-Aided Diagnosis (CAD) can be extremely beneficial in the investigation, prognosis, along with the remedy of brain tumors. A typical brain tumor CAD system includes three major phases: tumor ROI segmentation, feature extraction, and classification (based on the retrieved features) [5]. The most significant and time-consuming aspect of such a system is brain tumor segmentation, which may be done manually or automatically. The requirement of tracing ROIs is an evident restriction of such systems, which might pose a few issues. First of all, because of the varying nature of brain tumors greatly in shape, size, and location, tracing ROIs may be difficult and frequently not totally automated. This may result in severe segmentation mistakes that aggregate into subsequent phases, resulting in erroneous categorization. Second, tumor-surrounding tissues are thought to be distinct between tumor classifications. Third, relying entirely on ROI characteristics implies full ignorance of tumor location information, which might have a significant impact on categorization.

The aforementioned issues inspire us to offer an alternate technique for classifying brain tumor that fully eliminates the segmentation step. Brain tumour detection is critical in biological applications. The brain tumor grouping continue to help doctors in disease diagnosis. Throughout the classification process, many operations including as preprocessing, feature extraction, and classification are necessary. Preprocessing is a stage in image processing that happens before feature extraction to establish the location of an area or item. This technique

includes filtering, standardizing, and identifying things prior to the extraction stage. The method of obtaining essential numeric values from photographs in order to distinguish them is known as feature extraction.

Brain MRI image is mainly used to diagnose tumors as well as to mimic tumor growth. The data is generally employed in tumor detection and therapeutic methods. An MRI image contains more information than a CT or ultrasound image. An MRI scan can detect anomalies in brain tissue and offer extensive information on brain architecture. MRI may detect a variety of brain diseases, which include abnormalities like cyst and tumor. It can detect white and gray matter of brain. Tumor diagnosis at an early stage is crucial since tumors may be hazardous in many scenarios and, in the worst-case scenario, lead to death. As a result, tumor prediction utilizing automated algorithms may be of great help in tumor identification and is the most secure option.

Convolutional neural networks (CNNs) are a type of deep learning technique which is extensively used in image processing. It is based on basic activities of the human brain [6]. CNNs have several advantages over ordinary machine learning networks, which can be achieved by increasing the number of training data, leading in an efficient and resilient model [7]. CNN is used in image processing techniques such as MRI image segmentation, identification, and classification, together with brain cancer classification and detection. Therefore, we offer a fully automatic CNN-based MRI image segmentation along with classification system for three types of malignancies: meningioma, glioma, and pituitary tumors in this study.

The remainder of the paper is divided into the following subsections. Section II goes over the existing literature on brain MRI classification. Section III describes the convolutional neural network. Section IV elaborates a detailed survey on some popular deep learning based MRI classification techniques. Section V describes analysis of MRI images using CNN architecture. Section VI describes the proposed CNN methodology. Section VII elaborates on the experimental work and compares it to the current state-of-the-art CNN model. Section VIII describes the experimental results. Section IX describes the limitation of the existing CNN architecture. Section X discusses about the experimental results. Finally section XI concludes the research paper.

II. LITERATURE SURVEY

Segmenting the region of interest from an item is one of the most complex and time-consuming processes, and segmenting the tumor from an MRI Brain image is a major one. Several current publications divide brain images into normal (tumorous) and pathological categories (non-tumorous). Al-Baderneh et al. [8] investigated the arrangement of brain MRI applying Artificial Neural Network as well as K-Nearest Neighbour (KNN) including textural characteristics utilizing 181 images of sick brains as well as 94 images of normal brains. The study [9] describes how to use feed-forward back-propagation to classify MRI vision as abnormal or normal. The particular algorithms are supervised (classes are known) and need feature retrieval before classification. Rajesh et al. [10] provide one such strategy, in which they conducted classification using a Feed Forward Neural Network with ternary layers, 50 nodes

in the hidden layer, along with single output node. Provisional investigation can be found in [11] as well as Taie et al. [12] used Support Vector Machine (SVM) to accomplish classification.

Cutting-edge deep learning algorithms are being developed in tandem with these technologies. Many of these works do not employ brain imaging to determine if something is normal or pathological, but they were included since they were done on a variety of other sorts of distribution. Pereira et al. [13] employed CNN to identify gliomas in their article. Deep learning was utilized by Kamnitsas et al. [14] to classify ischemic stroke. The author [15] examined the Adaptive Network-based Fuzzy Inference Approach, a suggested technique for categorizing tumors into five classes (ANFIS). Another study used pre-trained AlexNet to categorize and segment tumors depending on Gray-Level Co-Occurrence Matrix (GLCM) features [16]. SVM [17], CNN [18], other studies include Recurrent Neural Network (RNN) [19], AlexNet transfer learning network of CNN [20], VGG-16, Inception V3 and ResNet50 [21], along with CNN quartet technique.

The authors suggested a strategy for glioma classification that merged SVM and KNN in [22]. The accuracy for multi-classification is 85%, but the accuracy for binary classification is 88%. Ertosun and Rubin [24] advocated using CNN to differentiate between low along with high grade gliomas as well as their grades. They achieved 71% as well as 96% accuracy, respectively. Using axial brain tumor pictures, Paul et al. [23] trained and developed two unique classification algorithms (a fully connected CNN). The CNN architecture's accuracy was 91.43%, with two convolutional layers followed by two fully linked layers. M Malathi, P Sinthia, and colleagues [25] demonstrated completely automatic brain tumour segmentation using a convolutional neural network. Brain tumours and gliomas were the most common and lethal, with little life suspense on its highest level. Manual segmentation was a time-consuming task in medical practise, and its execution was highly dependent on the operator's experience. J. Seetha and S. Selvakumar Raja et al. [26] demonstrated that the massive amount of data generated by MRI impedes manual classification of tumour vs non-tumor at an exacting time. However, it has a few limitations (i.e. precise quantitative measurements were given for a restricted number of images). The classification of brain tumours using convolutional neural networks is described in this paper.

Vipin Y. Borole, Sunil S. Nimbhore, and Dr. Seema S. Kawthekar discuss brain tumours [27]. Because of the structure of the brain, brain tumour recognition is a difficult task for MRI images. A brain tumour is an unusual development of cerebrum cells. X-ray images provide better contrast of various delicate tissues of the human body. X-ray images outperform CT, Ultrasound, and X-beam images. In this, various preprocessing, post-processing, and strategies such as (Filtering, difference improvement, Edge recognition) and post-processing systems such as (Histogram, Threshold, Segmentation, Morphological operation) are accessible in MATLAB for the location of brain tumour images via image handling (IP) apparatus (MRI-Images). Gooya et al. [28] presented GLISTR, a method for segmenting gliomas in multi-modal MR images by registering the images to a probabilistic atlas of healthy individuals. The incorporation of the tumour

growth model into the anatomy of the patient brain was the major contribution. Based on histogram analysis of temporal Magnetic Resonance Image (MRI) data, Manikis et al. [29] proposed a novel framework for assessing tumour changes. The proposed method detects tumour distribution and quantitatively models its growth or shrinkage, potentially assisting clinicians in objectively assessing subtle changes during treatment.

Bauer et al. [30] developed a novel method for converting a healthy brain atlas to MR images of tumour patients. They presented a new method for adapting a general brain atlas to an individual tumour patient image that uses sophisticated models of bio-physio mechanical tumour growth. Roy et al. [31] proposed an investigation into automated brain tumour detection and segmentation from brain MRI. Brain tumour segmentation was an important step in extracting information from complex MRI brain images. Sindhushree K.S et al. [32] created a brain tumour segmentation method and validated it using two-dimensional MRI data. In addition, detected tumours are represented in three dimensions. To detect tumour, high pass filtering, histogram equalisation, thresholding, morphological operations, and segmentation using connected component labelling were used. The extracted two-dimensional tumour images were reconstructed into three-dimensional volumetric data, and the tumour volume was calculated.

Ertosun and Rubin [24] proposed using CNN to distinguish between low and high grade gliomas and their grades. They achieved 71% and 96% accuracy, respectively. Krol et al. [33] used axial brain tumour images to train and develop two distinct classification approaches (a fully connected CNN). The accuracy of the CNN architecture, which consisted of two convolutional layers followed by two fully connected layers, was 91.43%. In the next section, we will go through the various methods for classifying MRI as normal or abnormal.

III. CONVOLUTIONAL NEURAL NETWORKS

Convolution is a convolution layer approach placed on a linear algebra operation which helps in multiplying the filter in the image [34]. CNN is a neural network that analyses information by using a grid topology. The convolution layer, is used as the first panel. It changes the input data without altering initial information. Another frequent sort of layer is the pooling layer, which is used to compute the maximum or average value of the image's pixel coordinates. By generating a feature map, CNN can learn complex features.

A. Brain Imaging Modalities

To assess the brain structure, four basic technologies (CT, PET, DWI, and MRI) are commonly employed for brain malignancies.

1) Computed Tomography (CT): A CT scan is a diagnostic imaging test that produces images of the internal body using X-ray technology. CT scans are utilized to evaluate and diagnose brain abnormalities, they are also used to assist various surgical operations. It creates a 3D image of soft tissues along with bones using X-rays and a computer. CT is a non-invasive, painless technique for your healthcare professional to diagnose issues. CT scans are commonly performed with and without contrast agents to help the radiologist detect any abnormalities.

2) Position Emission Tomography (PET): PET is used to evaluate brain metabolism as well as the distribution of radio labeled chemical agents in the brain. PET detects emissions from metabolically active compounds that have been radioactively labelled and circulated throughout the body. The brain emission data PET scans are computer-processed to provide multidimensional pictures of chemical distribution throughout the brain. PET is a sort of functional imaging which can detect metabolic anomalies such as glucose metabolism, blood flow, oxygen consumption, amino acid metabolism and liquid synthesis. These measures represent the amount of brain activity in distinct brain areas and help us to understand more about how the brain functions. When they initially became available, PET scans outperformed all other metabolic imaging modalities in terms of resolution and completion time (as low as 30 seconds).

3) Diffused Weighted Image (DWI): Diffusion weighted imaging (DWI) is a technique for producing signal contrast based on Brownian motion changes. DWI is a method used to evaluate the human body's molecular function and micro-architecture. DWI is a critical component of today's cutting-edge magnetic resonance imaging and is widely employed in neuro-imaging and cancer research. DWI is a continuously changing technological field, with new applications emerging on a regular basis.

4) Magnetic Resonance Imaging (MRI): MRI is one of the most popular imaging procedures utilized before and after surgery, with the goal of giving critical information for the treatment plan. In the medical industry, MRI is commonly used to identify and see elements in the body's internal structure. It is used to identify differences in biological tissues and is significantly superior to computed tomography (CT). Strong magnets are used in magnetic resonance imaging (MRI) scanners to polarise and excite hydrogen nuclei (one protons) in human tissue, creating a measurable signal that is spatially recorded and translated into body images. The proper interpretation of brain MRI images is vital for gathering important information that can aid in the early identification and diagnosis of illnesses.

B. Contributions

Convolutional Neural Networks (CNNs) have shown to be extremely fruitful in diagnosing a wide range of disorders and are commonly utilized in bio-medical image analysis. Those networks are particularly useful in detecting, classifying, and segmenting brain cancers in MRI datasets. Our approach's primary goals are as follows:

- Using deep learning and convolutional neural networks to detect brain cancers in MRI datasets.

- Experiment findings have shown that enlarging a dataset utilizing rotation, flipping, and translation approaches is quite beneficial for training the VGG architecture.

IV. A DETAILED SURVEY ON SOME POPULAR DEEP LEARNING BASED MRI CLASSIFICATION TECHNIQUES

A. MobileNet V2

MobileNet-V2 is a 53-layer deep convolutional neural network. It enables real-time categorization in devices with

little computing power, such as smart phones. MobileNetV2 has an inverted residual architecture, with narrow bottleneck layers functioning as the residual blocks' input and output. It also uses lightweight convolutions to filter features in the expansion layer. Finally, nonlinearities in thin layers are eliminated. MobileNetV2 enhances the up-to-date execution of mobile models on a variety of tasks along with benchmarks, as well as over a wide range of model sizes. It is a powerful component for obtaining the object recognition and segmentation. MobileNetV2 is a feature extractor with tremendous capability for object segmentation and detection.

Advantages

- Network size has been reduced to 17MB.

- The number of parameters has been reduced to 4.2 million.

- They are faster in performance and excellent for mobile apps.

- A convolutional neural network with a low latency.

- MobileNets employ depth-wise separable convolutions in a simplified design.

- MobileNet employs two simple global hyperparameters to efficiently trade off accuracy for latency.

- MobileNet might be utilized for item identification, fine-grain categorization, face recognition, large-scale geolocation, and other applications.

B. ResNet101

The residual block is the fundamental basic element of ResNet101. As we proceed further into the network with a large number of layers, the computation gets increasingly complicated. These layers are built on top of one another, with each layer attempting to learn an underlying mapping of the desired function. We try to fit a residual mapping instead of these pieces. Another benefit of this extremely deep architecture is that it enables for up to 150 layers of this, which we then stack on a regular basis. We additionally double the number of filters and use stride two to spatially down sample. Finally, only layer 1000 was entirely linked to output classes. The fundamental loss of the present network is that due to the enormous number of parameters, assessing it is fairly expensive [35]. The ResNet101 focuses largely on treating the second non-linearity as an identity mapping, which implies that the result of the addition operation between the identity mapping and the residual mapping should be sent immediately to the next block for processing. Batch Normalization is at the heart of ResNet101. Batch Normalization modifies the input layer to improve network performance. The problem of covariate shift is solved. The Identity Connection is used by ResNet101 to assist avoid the network from experiencing the vanishing gradient problem. Deep Residual Network improves network speed by leveraging bottleneck residual block design.

Advantages

- Networks with several layers (even thousands) may be trained simply without increasing the training error percentages.

- ResNets can help solve the vanishing gradient problem by using identity mapping.

C. DenseNet121

DenseNet121 is a novel category of convolutional neural network in which all previous layers are linked to the current layer. A DenseNet121 is a sort of convolutional neural network that connects all levels directly by using dense connections between layers through Dense blocks. A deep DenseNet121 is defined as a collection of DenseNets (referred to as dense blocks) with extra convolutional and pooling operations performed between each dense block. DenseNet121 connects layers together using dense blocks [36]. DenseNet connects each layer to every other layer. This is incredibly effective.

Advantages

- They address the problem of disappearing gradients.

- They enhance feature dissemination.

- They promote feature reuse; and

- They minimize the number of parameters greatly.

V. ANALYSIS OF MRI IMAGES USING CNN ARCHITECTURE

Classification is important because it arranges images into distinct groups. It is the first step in diagnosing any disease by forecasting an area or region having anomalies. The CNN model has several layers, including the convolution layer, pooling layer, flatten layer, dropout layer, and dense layer. In addition to the layers used in the CNN process, this study includes an activation function based on rule activation. An image in the shape of a number that interweaves the first convolution, with a resolution of 240x240 pixels. Kernels with a size of 3x3 and a thickness of three are used in accordance with the channel of the image data and filters. The model will perform the activation and data pooling functions after receiving the results of the operation. The Pooling layer process reduces the size of the feature map. The convolution process produces a feature map, which is then used for subsequent convolution processes. The next step is to flatten the feature map into vector form in order to perform a fully-connected layer process to produce image classification. This section briefly examines three alternative deep learning architectures (DenseNet121, MobileNet V2, and ResNet101) in addition to the recommended approach. Because of its successful performance in image classification that automatically finds key parts, CNN was employed to carry out the recommended classification strategy for brain MRI images. Fluid Attenuated Inversion Recovery (FLAIR) of MRI is the imaging method used here. It looks like a T2 picture but has a relaxation time (TR) along with longer echo (TE). The present sequencing is the disease perceptive as well as greatly simplifies the distinguishing CSF from an aberration.

VI. PROPOSED MODEL

A. MobileNet V2 Model

MobileNet-V2 is a fully convolutional network designed for mobile devices. It has an inverted residual framework,
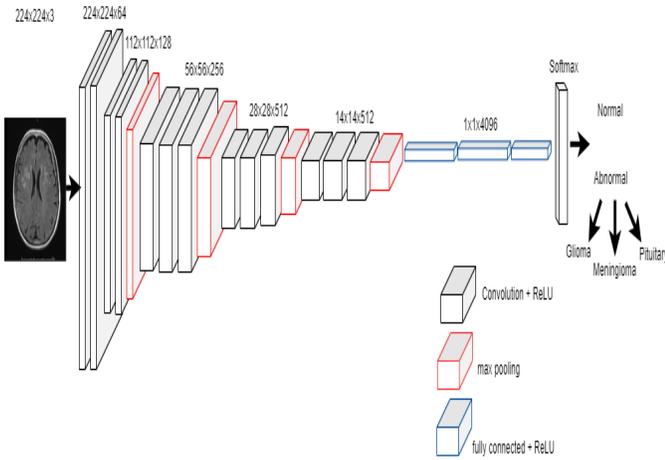
Fig. 1. MobileNet-V2 Model.
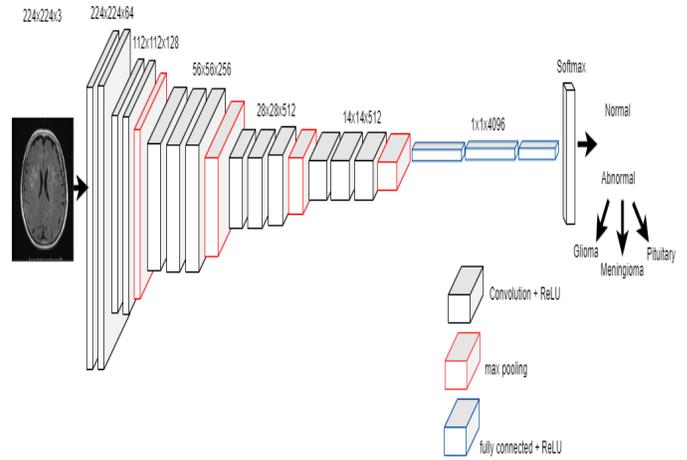
Fig. 2. ResNet Model.

along the congestion layers associated with residual connections. As a source of nonlinearity, the intermediate extension layer filters features with lightweight depth-wise convolutions. The original and improved images come in a variety of range. The first phase is via eliminating the noise from an image by using preprocessing, via median filtering. After that, the images are shrunk to a precise magnitude of 124x124x3 to ensure that they are not at all tiny; that is done to preserve the proportion along with the assists in improved training assuming that the sizes are just the same. The magnitude of 124 was selected since it is the fundamental size of accessible image. The images continue to turn into gray scale for easier understanding of characteristics. The particular images are then transferred to the convolutional layer, the most essential part of a CNN. Fig. 1 shows how the stride varies in each convolutional layer. The MobileNetV2 model has a total parameter count of 2,263,108 and consists of a first fully convolutional layer with 32 filters, followed by 17 residual bottleneck layers and the activation function Relu. It achieves promising accuracy results while using less computing memory and resources. Furthermore, it transforms them into a high speed network for image processing jobs by using dropout, dense followed by softmax activation function. MobileNetV2 is a lightweight convolutional neural network utilized in synchronous functions. First, it has 2,228,996 trainable parameters than classic convolution, which minimizes computing expenses, along with 34,112 parameters are used as non-trainable parameters.

B. ResNet101

Using ImageNet data, ResNet101 took first state in the ILSVRC 2015 classification test. Because the pre-trained representation is used straight-forwardly for image categorization, transfer learning is adaptive. The images are trimmed such that just the brain area is visible. Preprocessing is the initial stage in removing noise from a image. After median filtering, the images are shrunk to an exact size of 124x124x3 to ensure that they are not too small; this preserves the ratio and aids in better training if the sizes are all the same. Because it is the least accessible, the image size of 124 was chosen. For improved learning of characteristics, the images are turned to gray scale. These images are subsequently transmitted to the

convolutional layer, which is the most crucial aspect of a CNN. Fig. 2 shows how stride varies in each convolutional layer along with pooling layer. The network can accept input images with height and width multiples of 32 and channel width of three. For the purpose of clarity, we'll assume the input size is 224 × 224 x 3. The kernels utilised to conduct the convolution operation in all three levels of the stage 1 block are 64, 64, and 128 correspondingly. Because the convolution operation in the Residual Block is conducted with stride 2, the size of the input is decreased to half in terms of height and width, while the channel width is doubled. As we advance through the stages, the channel width doubles and the input size is cut in half. Each pooling method is repeated for five times. Flattening is an important layer following pooling since we need to turn the whole matrix representing the input images into a single column vector, which is required for processing. The data is subsequently put into the Neural Network for processing. The model is trained and updated using the dataset under consideration for our study to solve two-class problems with outputs of abnormal (class 0) as well as normal (class 1). The ResNet50 model has a total parameter count of 42,666,372 and consists of a first fully convolutional layer with 32 filters, followed by 15 residual bottleneck layers and the activation function Relu. It achieves promising accuracy results while using less computing memory and resources. Furthermore, it transforms them into a high speed network for image processing jobs by using dropout, dense followed by softmax activation function. First, it has 42,561,028 trainable parameters than classic convolution, which minimizes computing expenses, along with 105,344 parameters are used as non-trainable parameters.

C. DenseNet121

The first step is to use preprocessing to eliminate noise from a image. It is accomplished by the use of median filtering. The median filter is used to eliminate outliers from a picture while preserving the image's information. The photos are scaled to an exact size of 124x124x3 after median filtering to ensure that they are not too tiny; this retains the ratio and assists in better training if the sizes are all the same. The image size of 124 was chosen since it is the least accessible.

Fig. 3. DenseNet Model.

Fig. 4. Dataset of Brain Tumor.
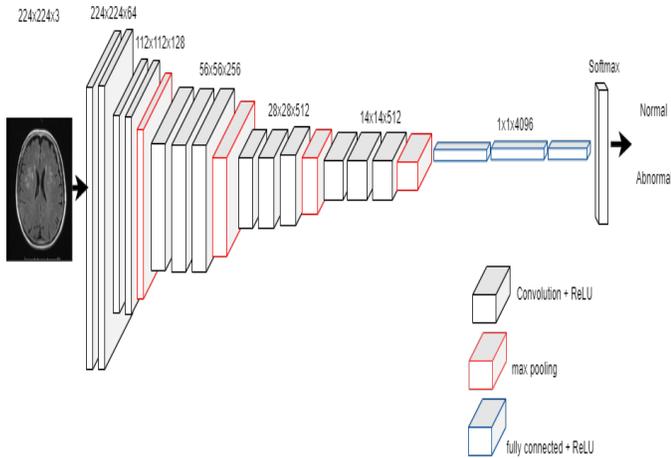
For improved feature learning, the images are transformed to gray scale. These images are subsequently sent to the convolutional layer, which is the most critical aspect of a CNN. Figure 3 shows how the stride varies in each convolutional layer. DenseNet121 employs a first fully convolutional layer with 57 filters, followed by 119 residual bottleneck layers and activation function maximum pooling with 7,047,504 total parameters. The network may accept input images with height, width multiples of 32, and channel width of three. For the sake of clarity, we will take the input size to be 224×224 x 3. The kernels utilised to execute the convolution operation in all three levels of the block of stage 1 are 64, 64, and 128 correspondingly. Because the convolution operation in the Residual Block is conducted with stride 2, the size of the input will be decreased to half in terms of height and breadth, but the channel width will be doubled. As we move through the stages, the channel width doubles and the size of the input is cut in half. Instead of using the Gradient descent (GD) technique, the Adam optimizer was used, which maintains a consistent learning rate for each weight in a network. Dropout, a regularizer, is used in our technique in completely linked layers. For this reason, a rate of 0.5 is specified. As a loss function, the binary cross-entropy loss function (log loss) was used. Finally, ReLU is used in conjunction with the Adam optimizer for classification, where 0.5 is classified as [1 0] (abnormal) and anything else as [0 1] (normal). The DenseNet model has a total parameter count of 7,041,604 and consists of a first fully convolutional layer with 32 filters, followed by 15 residual bottleneck layers and the activation function Relu. It achieves promising accuracy results while using less computing memory and resources. Furthermore, it transforms them into a high speed network for image processing jobs by using dropout, dense followed by softmax activation function. First, it has 6,957,956 trainable parameters than classic convolution, which minimizes computing expenses, along with 83,648 parameters are used as non-trainable parameters.

VII. EXPERIMENTS AND ANALYSIS OF RESULTS

The information came from the open-source Kaggle database. The collection contained X-ray photos of healthy and brain tumour patients. To extract the characteristics, a CNN model is used. Four Conv2D layers, three Maxpooling2D levels, one flatten layer, two dense layers, and a ReLu activation function comprise the model. The ReLU function is used to activate the last thick layer. The primary goal of this research is to evaluate the accuracy of the planned model to that of the pretrained model. The final layers are fully adjustable, with options such as Average Pooling, Flatten, Dense, and Dropout.
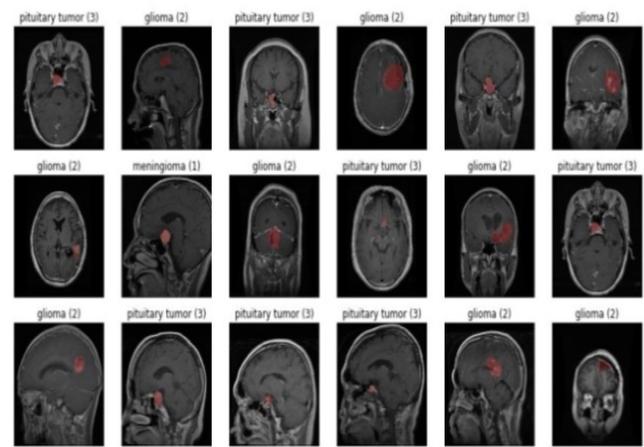
A. Dataset Used

Magnetic Resonance Imaging is the most effective tool for detecting brain cancers (MRI). The scans yield massive quantities of data pictures, which the radiologist examines. We propose a classification model in this paper that would allow us to take MRI pictures of the patient as input and calculate whether or not a tumour exists in the brain as output. The dataset of the proposed framework has been taken from the kaggle repository as shown in Fig. 4. The dataset contains 826 brain MRI images of glioma tumor, 822 images of meningioma tumor, 847 images of pituitary tumor, and 395 images with no tumor. As a result, there are 2890 images in total. 80% of the dataset has been used for training where as 20% for testing purposes. X-ray images of both healthy and brain tumor patients were included in the collection. The models learn to recognize images based on the properties extracted from the provided images. To eliminate noise, the dataset was gathered locally and pre-processed with an adaptive histogram equalizer.

B. Tools Used

To carry out the implementation, the Python programming language is employed. Keras and TensorFlow are the libraries utilized. Several Python-based packages are studied in this study to implement our techniques.

C. Preprocessing

The modified curvature diffusion equation (MCDE) is used to normalize image intensity as additional step in the preprocessing phase. The Wiener filter is used in medical imaging to increase local and spatial information. When the noise level is excessive, it is difficult to recover the edge of
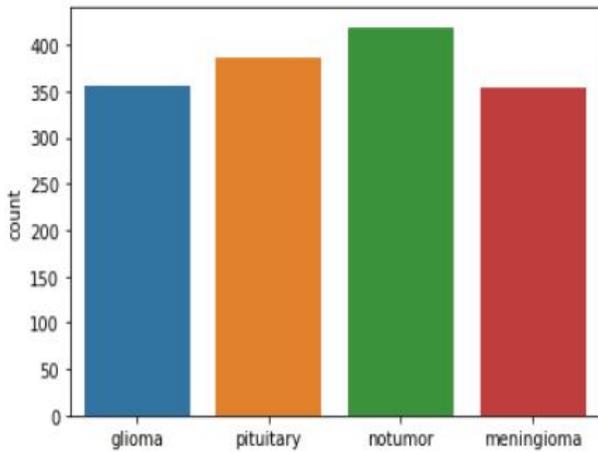
Fig. 5. Testing Simulation Result of Dataset.
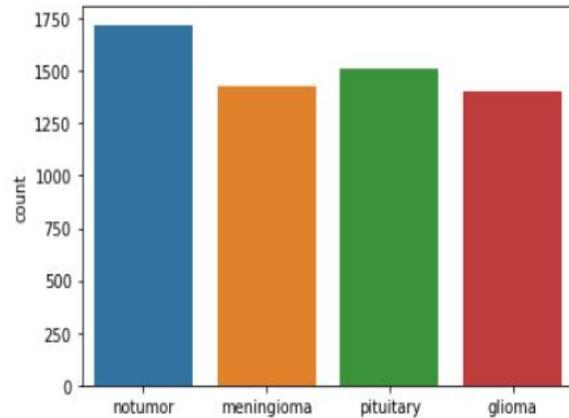
Fig. 6. Training Simulation Result of Dataset.

an image. Because images have distinct variations in intensity, contrast, and size, pre-processing is used to provide smooth training [12]. The wrapping and cropping method will be applied to the input image in the first pre-process. The input image is verified against the edge of the major object in the image during wrapping. Cropping occurs when the biggest edge of an image is set so that the item in the image stays intact. Because the picture sizes in the collection differ, resize the image to (240x240x3) = (image width x image height x number of channels). To aid in learning, normalise pixel values to the 0-1 range. The next step is to make an array out of each photograph in the collection. The image is used as a preprocessed input by MobileNetV2, DenseNet, and ResNet50. Coding is the final stage. The tagged data is converted into a numerical label, which may be understood and evaluated. The dataset is then separated into three sections: 20% for validation, 70% for training, and the remaining for testing.

Fig. 7. Classification Report of MobileNet-V2 Model.

D. Performance Metrics

Researchers evaluate many performance indicators in classification, with accuracy being the most commonly used performance parameter. Accuracy, precision, F1 Score, and support are the parameters used to validate our results.

VIII. EXPERIMENTAL RESULTS

To categorize the images as normal or abnormal, MobileNet V2, DenseNet121, and ResNet101 are used. The implementation is separated into two parts: general categorization into normal or abnormal brain tumors and specific classification into different categorises of the brain tumors. There are two approaches: k fold cross validation with k fold = 5 as well as 8 (arbitrary values), along with the generalisation strategy, which does not use training phase data in the testing phase. Both training as well as testing evidence of the used dataset are represented in Table I.

The graphical representation of result of the testing simulation is shown in Fig. 5 and the result of the training simulation is shown in Fig. 6.

A. Results and Analysis

In MobileNetV2, total parameters 2,263,108 are used. Out of which, 2,228,996 are used as a training parameters and 34,112 are used as a non-training parameters.

Fig. 7 shows the classification results for parameters such as precision, recall, f1-score as well as support for four classes of dataset used along with macro-average accuracy of 97% and weighted-average accuracy of 97%.

Fig. 8 shows the confusion matrix for the four classes of tumor dataset in MobileNet-V2 Model, which shows exponential value as 3.5e+02, that means the accurate values is 350 in decimal for glioma tumor, exponential value as 3.2e+02, that means the accurate values is 320 in decimal for meningioma tumor, exponential value as 4.2e+02, that means the accurate values is 420 in decimal for normal patients and exponential value as 3.8e+02, that means the accurate values is 380 in decimal for pituitary tumor patients.

Fig. 9 shows the graphical representation of the training and validation accuracy/loss for the MobileNet-V2 model. The testing accuracy and loss is almost constant for number of

TABLE I. TESTING-TRAINING SIMULATION RESULT OF THE DATASET

Types of tumor	Testing Simulation result	Training Simulation Result
Glioma tumor	350	1400
Pituitary tumor	375	1500
No tumor	425	1700
meningioma tumor	340	1400
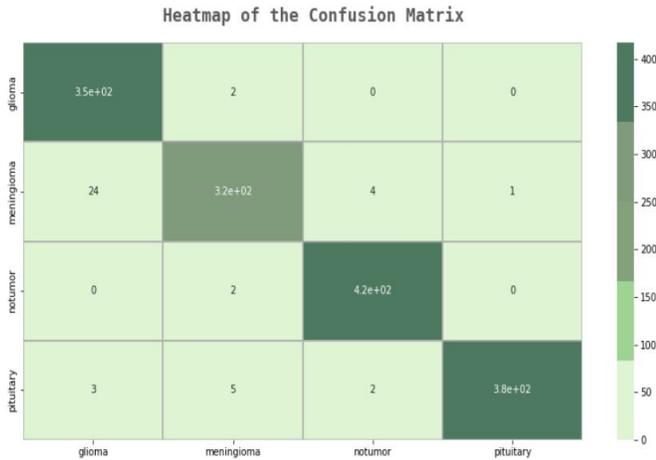
Fig. 8. Confusion Matrix of MobileNet-V2 Model.

Fig. 10. Classification Report of DenseNet Model.
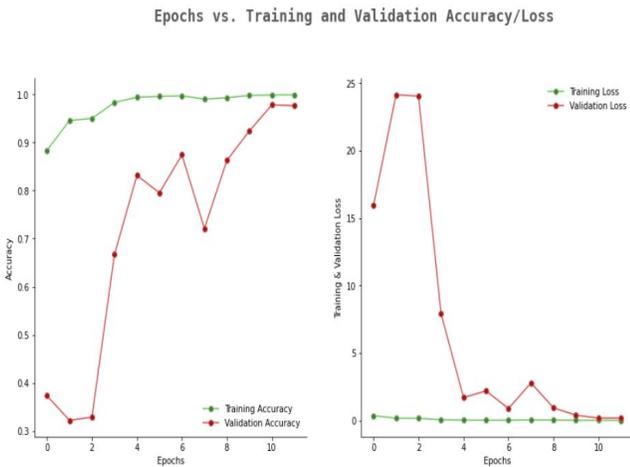
Fig. 9. Training and Validation Accuracy/Loss of MobileNet-V2 Model.
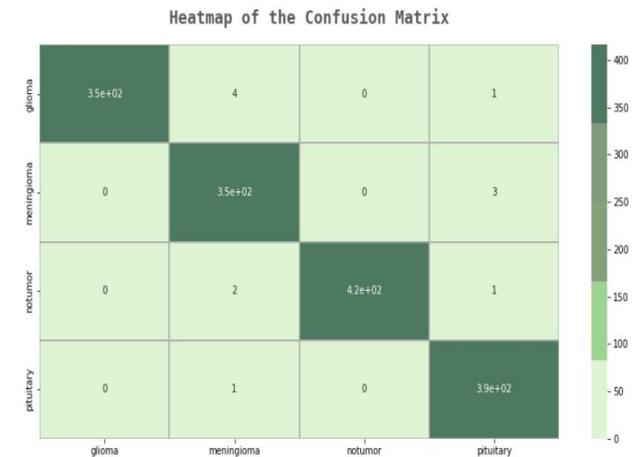
Fig. 11. Confusion Matrix of DenseNet Model.

epochs, whereas the validation accuracy and loss is varying greatly for number of epochs.

In DenseNet121, total parameters 7,047,504 are used. From which, 6,966,026 parameters are used as a training and 81478 are used as a non-training parameters.

Fig. 10 shows the classification results for parameters such as precision, recall, f1-score along with support of four classes of dataset used as well as macro-average accuracy of 99% and weighted-average accuracy of 99%.

Fig. 11 shows the confusion matrix with the measures of 3.5e+02 in exponential form, that means the accurate values is 350 in decimal for glioma tumor, the exponential value as

3.5e+02, that means the accurate values is 350 in decimal form for meningioma tumor, measures of 4.2e+02 in exponential form, that means the accurate value is 420 in decimal form for no tumor and the exponential value as 3.9e+02, that means the accurate values in decimal form is 390 for pituitary tumor patients dataset in DenseNet Model.

Fig. 12 shows the graphical representation of the training and validation accuracy/loss for the DenseNet121 model. The testing accuracy and loss is almost constant after certain number of epochs, whereas the validation accuracy and loss is varying greatly at the beginning for number of epochs but after certain number of epochs it comes to a constant value.

Total parameters used for ResNet101 model are
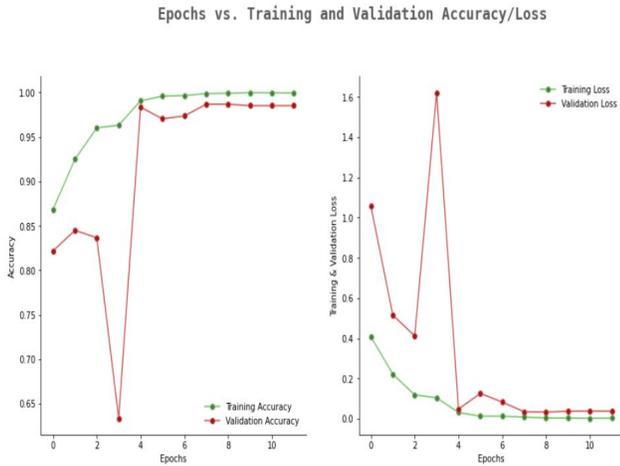
Fig. 12. Training and Validation Accuracy/Loss of DenseNet Model.
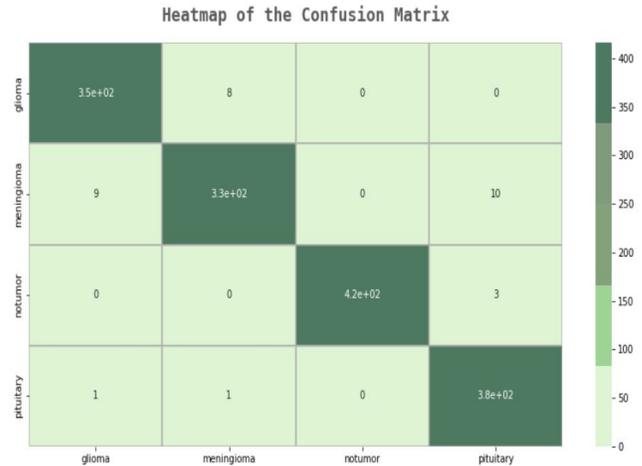
Fig. 14. Confusion Matrix of ResNet101 Model.

Fig. 13. Classification Report of ResNet101 Model.
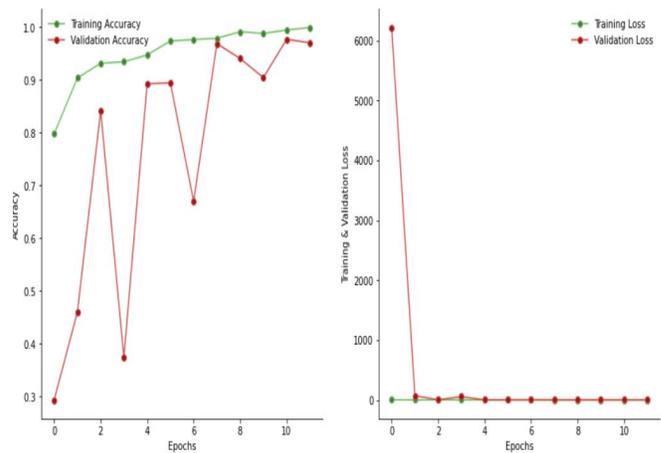
Fig. 15. Training and Validation Accuracy/Loss of ResNet101 Model.

21,963,424. Out of which, 21,961,408 are used as a training parameters and 2,016 are used as a non-training parameters.

Fig. 13 shows the classification results for parameters such as precision, recall, f1-score as well as support of four classes of dataset used along with macro-average accuracy of 98% as well as weighted-average accuracy of 98%.

Fig. 14 shows the confusion matrix for the four classes of tumor dataset, which shows 3.5e+02 as exponential value, that means the accurate values is 350 in decimal for glioma tumor, 3.3e+02 as exponential value, that means the accurate values is 330 in decimal form for meningioma tumor, the exponential value as 4.2e+02, that means the accurate value is 420 in decimal for no tumor and the exponential value as 3.8e+02, that means the accurate value is 380 in decimal for pituitary tumor in ResNet Model.

Fig. 15 depicts a graphical depiction of the ResNet101 model's training and validation accuracy/loss. The testing accuracy and loss are practically constant for the number of epochs, but the validation accuracy varies substantially with

the number of epochs, but the validation loss is constant.

IX. LIMITATIONS OF EXISTING MACHINE LEARNING METHODS

The following are the primary difficulties in detecting brain tumors:

- It is still difficult to detect a little quantity of tumour since it might be mistaken for a normal region.

- Some existing techniques are effective in one tumour site but not in another (enhanced or non-enhanced), and vice versa.

X. DISCUSSION

The Table II displays the comparison of loss along with accuracy values for test parameters of different machine learning methods. The following parameters such as precision, recall, F1 score and accuracy have been used to assess the prediction results. Obtained values of precision, recall, F1 score as well as support for classification of ML models are represented

TABLE II. COMPARISON OF LOSS AND ACCURACY VALUES FOR TEST PARAMETERS OF DIFFERENT MACHINE LEARNING METHODS

Machine Learning Methods	Test Loss	Test Accuracy
MobileNet-V2	3.8652	0.30735
DenseNet121	0.25539	0.893838
ResNet101	0.082731	0.971789

in Table III. MobileNet-V2 models are the least suited for the image classification of brain tumors, as the accuracy for both macro-average and weighted-average is 97%. It can be seen that DenseNet model shows best results, which shows the accuracy for both macro-average and weighted-average is 99% due to its improved generalization and embedded ensemble learning feature. Our experimental performance proves the highest for our proposed CNN model with an accuracy of 92% which is more than all other models trained.

Following a thorough examination of existing state-of-the-art approaches, the following challenges have been identified:

- A brain tumor grows rapidly in size. As a consequences, early tumor diagnosis is an extremely important job.

- MRI pictures are inaccurately classified due to magnetic field changes in the coil.

- Another tough procedure is the structured and best feature extraction and selection.

XI. CONCLUSION

Convolutional neural networks are still a hot topic in the field of automated tumor segmentation. It is critical for radiologists to understand convolutional neural networks in order to be prepared to use these technologies in clinical practice in the future. A thorough examination of several CNN designs was conducted along with their limitations when faced with a small dataset. One of the most significant fields of medical imaging is the hunt for a treatment for various forms of brain tumours. To avoid overfitting, this article employs a data augmentation approach prior to classification. To reach a conclusion, we examined three machine learning approaches: MobileNet-V2, DenseNet121, and ResNet101. However, this study demonstrates the significance of supervised learning approaches in the development of CAD systems to reduce the burden on radiologists. A future investigation could include collecting larger brain MR images to generalise the classifier systems. T2-weighted contrast-weighted MRI images were used in this study. The fundamental purpose of this project is to develop a high-accuracy, high-performance, and low-complexity automated brain tumour classification system. As the loss for test parameters is very high for MobileNe-V2 i.e. 3.8652 as compared to DenseNet121 i.e. 0.25539 and ResNet101 i.e. 0.082731, due to lot of errors, so this MobileNet V2 model is going to be modified in our future work.

REFERENCES

[1] Devi, S., Sahoo, M. N., Muhammad, K., Ding, W., & Bakshi, S. (2019). Hiding medical information in brain MR images without affecting accuracy of classifying pathological brain. Future Generation Computer Systems, 99, 235-246.

[2] Park, J. G., & Lee, C. (2009). Skull stripping based on region growing for magnetic resonance brain images. NeuroImage, 47(4), 1394-1407.

[3] Khan, M. A., Lali, I. U., Rehman, A., Ishaq, M., Sharif, M., Saba, T., ... & Akram, T. (2019). Brain tumor detection and classification: A framework of marker-based watershed algorithm and multilevel priority features selection. Microscopy research and technique, 82(6), 909-922.

[4] Raza, M., Sharif, M., Yasmin, M., Masood, S., & Mohsin, S. (2012). Brain image representation and rendering: A survey. Research Journal of Applied Sciences, Engineering and Technology, 4(18), 3274-3282.

[5] Watson, C., Kirkcaldie, M., & Paxinos, G. (2010). The brain: an introduction to functional neuroanatomy. Academic Press.

[6] Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Networks, 16(5-6), 555-559.

[7] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.

[8] Al-Badarneh, A., Najadat, H., & Alraziqi, A. M. (2012, August). A classifier to detect tumor disease in MRI brain images. In 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 784-787). IEEE.

[9] Mohsen, H., El-Dahshan, E. S. A., & Salem, A. B. M. (2012, May). A machine learning technique for MRI brain images. In 2012 8th International Conference on Informatics and Systems (INFOS) (pp. BIO-161). IEEE.

[10] Rajesh, T., & Malar, R. S. M. (2013, July). Rough set theory and feed forward neural network based brain tumor detection in magnetic resonance images. In International Conference on Advanced Nanomaterials & Emerging Engineering Technologies (pp. 240-244). IEEE.

[11] Balasubramanian, C., & Sudha, B. (2014). Comparative study of de-noising, segmentation, feature extraction, classification techniques for medical images. Int. J. Innov. Res. Sci. Eng. Technol, 3(3), 1194-1200.

[12] Taie, S. A., & Ghonaim, W. (2017, March). CSO-based algorithm with support vector machine for brain tumor's disease diagnosis. In 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) (pp. 183-187). IEEE.

[13] Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. IEEE transactions on medical imaging, 35(5), 1240-1251.

[14] Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., ... & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical image analysis, 36, 61-78.

[15] Roy, S., & Bandyopadhyay, S. K. (2018). Brain tumor classification and performance analysis. Int. J. Eng. Sci, 8, 18541-18545.

[16] Krishnammal, P. M., & Raja, S. S. (2019, April). Convolutional neural network based image classification and detection of abnormalities in mri brain images. In 2019 International Conference on Communication and Signal Processing (ICCSP) (pp. 0548-0553). IEEE.

[17] Hamid, M. A., & Khan, N. A. (2020). Investigation and classification of MRI brain tumors using feature extraction technique. Journal of Medical and Biological Engineering, 40(2), 307-317.

TABLE III. CAMPARATIVE ANALYSIS OF CLASSIFICATION METRICS ON THREE DIFFERENT DEEP LEARNING TECHNIQUES

Machine Learning Methods	Types of tumor	Precision	Recall	F1-Score	Support
MobileNet-V2	Glioma tumor	0.93	0.99	0.96	356
	Meningioma Tumor	0.97	0.92	0.94	353
	No Tumor	0.99	1.00	0.99	419
	Pituitary Tumor	1.00	0.97	0.99	387
DenseNet121	Glioma tumor	1.00	0.99	0.99	356
	Meningioma Tumor	0.98	0.99	0.99	353
	No Tumor	1.00	0.99	1.00	419
	Pituitary Tumor	0.99	1.00	0.99	387
ResNet101	Glioma tumor	0.97	0.98	0.97	356
	Meningioma Tumor	0.97	0.95	0.96	353
	No Tumor	1.00	0.99	1.00	419
	Pituitary Tumor	0.97	0.99	0.98	387

[18] Hanwat, S., & Jayaraman, C. (2019). Convolutional neural network for brain tumor analysis using MRI images. Int. J. Eng. Technol, 11, 67-77.

[19] Kalaiselvi, K., Karthikeyan, C., Shenbaga Devi, M., & Kalpana, C. (2020). Improved Classification of Brain Tumor in MR Images using RNN Classification Framework. Int. J. Innov. Technol. Explor. Eng.(IJITEE), 9, 1098-1101.

[20] Wahlang, I., Maji, A. K., Saha, G., Chakrabarti, P., Jasinski, M., Leonowicz, Z., & Jasinska, E. (2022). Brain Magnetic Resonance Imaging Classification Using Deep Learning Architectures with Gender and Age. Sensors, 22(5), 1766.

[21] Khan, H. A., Jue, W., Mushtaq, M., & Mushtaq, M. U. (2020). Brain tumor classification in MRI image using convolutional neural network. Math. Biosci. Eng, 17(5), 6203-6216.

[22] Zacharaki, E. I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E. R., & Davatzikos, C. (2009). Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 62(6), 1609-1618.

[23] Paul, J. S., Plassard, A. J., Landman, B. A., & Fabbri, D. (2017, March). Deep learning for brain tumor classification. In Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging (Vol. 10137, pp. 253-268). SPIE.

[24] Ertosun, M. G., & Rubin, D. L. (2015). Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks. In AMIA annual symposium proceedings (Vol. 2015, p. 1899). American Medical Informatics Association.

[25] Malathi, M., & Sinthia, P. (2019). Brain tumour segmentation using convolutional neural network with tensor flow. Asian Pacific journal of cancer prevention: APJCP, 20(7), 2095.

[26] Seetha, J., & Raja, S. S. (2018). Brain tumor classification using convolutional neural networks. Biomedical & Pharmacology Journal, 11(3), 1457.

[27] Borole, V. Y., Nimbhore, S. S., & Kawthekar, D. S. S. (2015). Image processing techniques for brain tumor detection: A review. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 4(5), 2.

[28] Gooya, A., Pohl, K. M., Bilello, M., Cirillo, L., Biros, G., Melhem, E. R., & Davatzikos, C. (2012). GLISTR: glioma image segmentation and registration. IEEE transactions on medical imaging, 31(10), 1941-1954.

[29] Manikis, G. C., Emmanouilidou, D., Sakkalis, V., Graf, N., & Marias, K. (2011, November). A fully automated image analysis framework for quantitative assessment of temporal tumor changes. In 2011 E-Health and Bioengineering Conference (EHB) (pp. 1-6). IEEE.

[30] Bauer, S., May, C., Dionysiou, D., Stamatakos, G., Buchler, P., & Reyes, M. (2011). Multiscale modeling for image analysis of brain tumor studies. IEEE transactions on biomedical engineering, 59(1), 25-29.

[31] Roy, S., Nag, S., Maitra, I. K., & Bandyopadhyay, S. K. (2013). A review on automated brain tumor detection and segmentation from MRI of brain. arXiv preprint arXiv:1312.6150.

[32] Sindhushree, K. S., Manjula, T. R., & Ramesha, K. (2013). Detection and 3D reconstruction of Brain Tumor from Brain MRI images. International Journal of Engineering Research & Technology (IJERT), 2(8), 528-534.

[33] Krol, A., & Gimi, B. (2017, March). Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging. In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series (Vol. 10137).

[34] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[35] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[36] Alsaif, H., Guesmi, R., Alshammari, B. M., Hamrouni, T., Guesmi, T., Alzamil, A., & Belguesmi, L. (2022). A Novel Data Augmentation-Based Brain Tumor Detection Using Convolutional Neural Network. Applied Sciences, 12(8), 3773.

Towards a Blockchain-based Medical Test Results Management System

Phuc Nguyen Trong, Hong Khanh Vo, Luong Hoang Huong, Khiem Huynh Gia, Khoa Tran Dang,
Hieu Le Van, Nghia Huynh Huu, Tran Nguyen Huyen, Loc Van Cao Phu, Duy Nguyen Truong Quoc,
Bang Le Khanh, Kiet Le Tuan
FPT University, Can Tho City, Viet Nam

Abstract— The role of test results in the diagnosis and treatment of patients' diseases at medical facilities cannot be ignored. Patients must have a series of tests that are related to their symptoms. This can be repeated as many times as possible, depending on the type of disease and treatment. Seriously, in the cases where the patients lose their medical test record (i.e., patient's medical history), the diagnosis is difficult due to the lack of information about the medical history as well as the symptoms/complications in the previous treatments. Storing this treatment information in medical centers can address risks related to user failure (e.g., loss of medical test records and wet/fire documents). However, users face a bit of difficulty when they change to other medical centers for medical examination and treatment since the data is stored locally, and difficult to share this with others. Current solutions focus on empowering users (i.e., patients) to share medical information related to disease treatment. However, the main barrier to these approaches is the knowledge of the users. They must embrace some background in terms of the technologies, risks, and rights they may share with treatment facilities. To solve this problem, we propose a Blockchain-based medical test result management system where all information is stored and verified by the stakeholder. The data will be stored decentralized and updated throughout the treatment process. We implement a proof-of-concept based on the Hyperledger Fabric platform. To demonstrate the effectiveness of the proposed system, we conduct evaluation methods based on three main tasks of the system: initializing, accessing, and updating data on six different scenarios (i.e., increasing in size of processing requests). The evaluation based on Hyperledger Caliper helped us to have a deeper analysis of the proposed model.

Keywords—Blood donation; blockchain; hyperledger fabric; blood products supply chain

I. Introduction

The need for health care is extremely urgent for all ages. The diagnosis process has a great influence on the treatment of the patient. To be able to make an accurate diagnosis of the condition, doctors must consider the test results as well as the patient's medical history. This information is usually compiled in a medical test result. Depending on the medical facility, medical test results are provided in the form of paper results or electronic results [1]. One of the biggest difficulties in this traditional process is the long processing time and waiting time, not to mention all the information resulted in the medical test result (i.e., information about the patient's health). In addition, the traditional medical process is not yet capable of reusing the patient's existing test results. Specifically, the current storage methods are only centrally stored in a central or hospital server in the big cities. This situation does not apply to small health facilities in the countryside [2]. For this reason, all resulted information is manually analyzed by the physicians before being resulted in the patient's medical result [3].

This situation is extremely risky for the healthcare system because i) it is easy for patients to lose the medical test results due to natural issues (e.g., flood, fire) or their failure (e.g., lost); ii) it is very difficult to back up those results because of technology and equipment limitations. To this end, the previous approaches that have contributed a lot to these risks are based on centralized personal data methods (i.e., where the user is central and is allowed to share any data with the hospital/ medical facility or third party) [4]. In addition, the methods propose a decentralized management mechanism for users, including hospital or medical center staff (e.g., doctors, nurses) and patients and their family members; [5] in special cases (e.g., emergency [6]). These approaches are based on the assessment that health care is needed and prioritized over privacy issues. However, both groups of approaches have encountered a binding mechanism that the patient must be able to use smart devices (i.e., smartphones). They argue that privacy risks are of great concern and that there must be a reasonable mechanism for healthcare-related issues where medical data is exploited/stored/processed. This argument is completely correct and can be applied to big medical institutions where the facility is available for data storage and processing. However, for the other scope (i.e., small medical centers), not many approaches provide the solution in terms of storage and processing of the medical data [7], [8].

This paper is one of the first attempts to address the above issues related to the storage of medical test results from medical/healthcare centers. We aim for a decentralized storage solution that is not bound by data storage and processing equipment in small and medium medical centers. This study opens up a potential approach where decentralized methods of storing and processing patient personal data (i.e., test results) can be applied. In addition to infrastructure barriers, we also consider issues related to data transparency, where all information stored is verified by stakeholders (e.g., patients, the patient's relatives in the emergency situation, and the doctor). This method gives transparency to all stored data. All of the above solutions lead to a Blockchain-based approach, where all data is verified by the parties involved. Furthermore, the data is stored and processed decentralized.

Blockchain technology is known for its outstanding features of transparency and immutable content. Picha Edwardsson et al. studied the possibility of using blockchain technology

to create a secure, community-facing information verification database with the goal of creating a solution that could improve the reliability of verifying information and monitoring each authenticity verification process for digital content, including images and videos. The paper indicates that blockchain is not yet ready to be directly applied to fact-checking processes in a real-world scenario. The study also shows that the application of blockchain to verify a scenario is entirely possible and highly reliable and transparent [9]. Several approaches address these problems by applying Blockchain techniques in the other environment (e.g., cash-on-delivery [10], [11], [12], healthcare [5], [4], [6], supply chain [13], [14], [15], and others [16], [17], [18]). As for the patient's/patient's ability to use technology, we assume that they have the ability to read information about their personal data usage through their phone whenever there is a phone call. access from outside the system.

Rather than emphasizing the role of patients in our system [19], [20], [21], our proposed model aims to propose a de-centralized store and process system for the patient's medical test results (see more details in Section II). In other words, our system is device-centric instead of user (i.e., patient) or service provider (i.e., app). All processing requests are guaranteed by the parties (i.e., patients and therapists) and stored on a distributed ledger. An important difference from our system is that our proposed model still uses a trusted third party to manage the encryption and decryption of data before and after processing them. Acknowledging that adopting a model of depending on a trusted third party will compromise the security of the entire system (i.e., third party trust level). However, it also brings benefits when the data is secure and has fewer burdens on users as well as the user's background requirements [22]. Responsibility for the protection of personal data is assigned to a security company (i.e., third party). As for the implementation, we exploited the Hyperledger Fabric platform for our proof-of-concept. The related evaluation to prove the effectiveness (i.e., focusing on initialization, query, and update) was analyzed by Hyperledger Caliper.

Stemming from the research problem of ensuring transparency and decentralized storage for patients' medical test results (see details in the related work section), we propose a model for information management about patients. Test results based on Blockchain technology and Smart contract. Therefore, our main contribution revolves around three aspects: (a) building a medical test result management system based on Blockchain and Smart contract; (b) building proof-of-concept on top of Hyperledger Fabric; and (c) assessing the appropriateness of the approach based on an analysis of three main scenarios (i.e., initialize, retrieve, and update) based on the Hyperledger Caliper platform.

The next section presents the state-of-the-art. Sections III and IV present our approach, processing model, and system implementation. Section V builds an environment for evaluating proposed models and makes comments on their strengths and weaknesses as well as future directions in Section VI. Finally, we summarize the study in Section VII.

II. Related Work

There are many approaches that have proposed methods for remote diagnosis and treatment of diseases, which are

data mining and other practical applications based on medical data by exploiting the strengths of the blockchain technology. For example, Chen et al. [23] proposes a model for storing and controlling personal data in a healthcare environment based on Blockchain technology. This system can collect information from IoT devices (i.e., medical devices in real time). To improve the security of the system, the authors build an anonymous data sharing environment and encrypt the patient's personal data before storing them on cloud servers. Similarly, Du et al. [1] and Son et al. [24] used medical centers (i.e., hospitals) to store data and manage access and those hospitals. Specifically, they categorize two types of medical data protection policies: global for all data shared outside of the medical center, and local, which is accessed only by individuals at the medical center. medical (i.e., doctor, nurse). However, one of the major limitations is that through this solution, patients do not have full control over their data as the data and policies are stored in the hospital. Patra et al. [25] proposes a cloud-based model to build an information system at the national level, providing a more convenient solution for patients in rural areas at the lowest cost. Specifically, instead of having to go to health care centers in large companies, they propose a solution to diagnose and treat diseases remotely. Specifically, citizens are encouraged to provide their personal healthcare information, which will be stored in the health cloud and accessed by health professionals and policymakers to provide more medical services. Similarly, Rolim et al. [26] proposes a framework that covers the process from data collection to cloud-based data delivery. Using sensors mounted on medical equipment, data can be collected and stored directly in the cloud, which can be accessed by authorized medical professionals.

Some other approaches build a user-centric (i.e., patient) model, who has full discretion to share their personal data with providers/health care facilities. economic (i.e., in a medical setting). For example, Makubalo et al. [27] has summarized the above approaches in their publication. They argue that the methods of building a user-centric health data sharing system are facing a lot of difficulties due to the limitations of the method of building centralized data system (i.e., data) stored and processed centrally in cloud servers). Yin et al. [28] introduced a patient-centric system built in the cloud with a data collection layer, data management layer, and medical service delivery layer based on medical records of the patient. To protect data privacy, many approaches have adopted attribute-based encryption (ABE), one of the most common encryption schemes used in cloud computing, to define patient data object. Depending on the context, the policy tells to lose (or not) grant the corresponding access rights. For example, Barua et al. [29] proposes an ABE-based access control model based on patience and privacy protection; Chen et al. [30] described a new framework with a cloud-based, privacy-aware Role-Based Access Control model that can be used for control, data traceability, and access allowed access to healthcare data resources. Methods for applying the Access Control model are also introduced for dynamic policies [31], [32] or protection policies for both security and privacy [33].

In addition, Madine et al. [34] has introduced a Smart Contract-based system that provides patients with reliable, traceable and secure control over their medical data (i.e., which is stored non-invasively). concentrate). To increase the security

and privacy of medical data, they used the decentralized storage feature of the interplanetary file system (IPFS) to store and share patient medical data safely. For practical applications, HealthBank has proposed a healthcare system and surrounding ecosystems that allow users (i.e., patients) to manage and control their data.[1] This solution is recommended to be able to comply with strict security and privacy regulations (e.g., GDPR) and to assist users in using their services. In addition, the system also proposes solutions for storing personal data with complex data encryption algorithms, immutability and accountability. Similarly, HealthNautica and Factom Announce Partnership have used blockchain technology to ensure the integrity of patient medical data while providing transparency based on blockchain technology and encryption of sensitive data (e.g., personal information, health status).[2] With the same approach based on Blockchain technology and IPFS, Misbhauddin et al. [35] introduced the MedAccess platform, A Scalable Architecture for Blockchain-based Health Record Management. The platform supports on-chain storage and processing allowing doctors, lab technicians and patients to securely manage medical records. However, these systems face some problems in the processing and storage of personal data. Specifically, Le et al. [12] has argued that not all data collected must be processed on-chain. Instead, Son et al. [14] argues that personal data that is either not directly related to treatment or diagnosis may be stored off-chain (i.e., offchain). Similar to the above approach, to increase the processing capacity for the whole system, Zyskind et al. [36] presented an approach based on in-chain and out-of-chain processing. Onchain processes require all entities of a typical personnel management system, where patient and medical staff information is stored; in contrast, encrypted medical data is stored on a separate centralized storage server to enable faster access and low cost. However, the above methods have major limitations including that any information that is validated must be executed on-chain instead of local processing. This only benefits storage but does not change data handling (i.e., since all information still executes on-chain) [7].

The above approaches have brought many solutions to today's traditional health care systems. However, in developing countries (e.g., Vietnam) where medical equipment and supplies are one of the barriers that directly affect people's health care process. In addition, the above approaches require a certain knowledge of information technology as well as the risks related to security and privacy. It is for the above reasons that a few case studies (i.e., applied to a specific geographical area - country, region) address the upper limits of [37]. In this article, we provide Blockchain-based support for the management of test results in medical centers.

III. The Blockchain-based Medical Test Results Management System

A. Traditional Model

Fig. 1 shows the basic steps of the traditional medical test results management process. This model describes the five main steps, excluding the risks of losing medical test
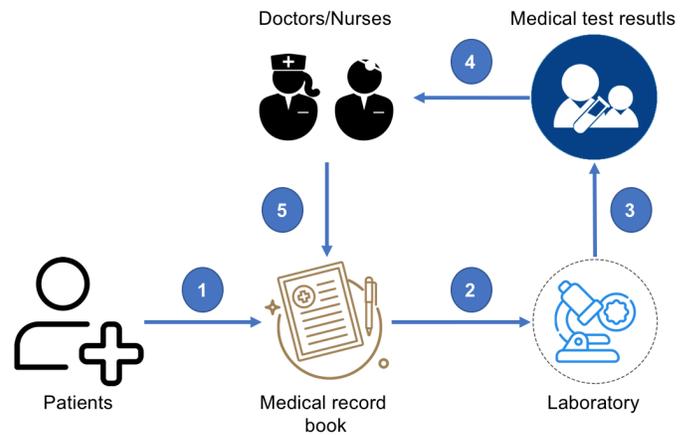
Fig. 1. The Traditional Process for Medical Test Results Management System.

results. In other words, this process will be repeated for each patient whenever they have a routine checkup or a health-related reason. Specifically, in the first step, the patient registers for a medical test result, which includes basic information about the patient, such as full name, address, phone number, or medical condition. The medical test result number is also the patient number at that hospital. In the second step, the patient brings the medical test result to a specialist at the hospital, called a laboratory (e.g., eye, blood, urine) for sampling. This procedure requires a very long wait time from the patient. The patient then receives information about this form in the third step before forwarding this information to the doctors and nurses for consultation in the fourth step. Finally, the consultation results are updated in the medical test result of the patient in the fifth one.

For the process of storing patient information, the storage of their information is completely manual. Only a few major medical centers in major cities support the storage of medical results on their centralized database. This demonstrates that it is not feasible to share a patient's medical result between different healthcare facilities. It is easy to see that there are many inconveniences for both patients and hospital staff when using the current testing/receiving process, respectively. The first limit comes from the patient, all information stored on the medical test result must be ensured carefully, and the medical test result must not be lost otherwise, all procedures will have to be repeated from the beginning with a new medical test result. Changing the place of treatment/examination is extremely difficult because the patient has to bring the medical test results issued at the previous medical facility to a new one. In addition, the loss of medical test results is extremely risky, besides the reason for having to repeat the entire sampling process, since they relate to the diagnosis process. Regarding the responsibility of physicians (i.e., doctors/nurses), they must reread a patient's entire medical history each time their patient has a follow-up visit. This is similar to the process of examining a new patient.

---

[1]https://www.healthbank.coop/2018/10/30/healthbank-creates-the-first-patient-centric-healthcare-trust-ecosystem/

[2]https://www.factom.com/company-updates/ healthnautica-factom-announce-partnership/

B. Proposed Model

To solve the above problems, we introduce a model based on Blockchain technology, where all information related to the testing process and the storage of patient's medical test results are updated and shared freely in the healthcare environment. Fig. 2 shows our proposal system based on Blockchain technology and distributed ledger (i.e., Hyperledger). As a first step, the patient initializes a global ID for not only a certain healthcare facility but also for others ones (e.g., the hospital in the same city). Unlike the traditional process, in another word, this ID will identify the user globally, which means that the patient can be examined at another medical facility without affecting the diagnosis process. Specifically, doctors/nurses can retrieve information about a patient's medical history based on their global ID (this will be covered in more detail in the next steps). From the initial global ID, users can generate more than 1 medical test result (i.e., per medical facility or healthcare service). These records store all test results and related patient information (i.e., similar to a medical test result in the paper). The data stored on the medical test result is always updated to Hyperledger (step 3). Users will then go to the respective Laboratories to take samples (step 4) before seeing a doctor in person to receive advice on their health status (step 6). This is the biggest difference between our model compared to the traditional model. Patients do not need to wait a long time at the facility; instead, an appointment is delivered to their device (e.g., smartphone) whenever their result is available. Meanwhile, the remaining steps will be executed independently at the system under the confirmation of the relevant parties. Specifically, after testing, the results are updated to the Hyperledger, and this information includes the user's corresponding medical test results and metadata about the time and location of the test as well as the doctors participating in the consultation. In case the patient goes to another medical facility, the patient's permission (or the patient's family member's/relatives in some special cases) must be obtained before accessing the patient's medical data (i.e., over-privileged permission). After receiving the request from the system, the doctors will enter the diagnostic results into the system (i.e., Hyperledger). The whole process will be confirmed by the stakeholders during the execution. The data will be encrypted when there is no request for access or update from the relevant parties (e.g., patient, nurse, doctor). The next section presents our approach based on Hyperledger Fabric.

IV. IMPLEMENTATION

A. Permission Diagram

Fig. 3 presents the working mechanism of the request authentication process in this paper. Specifically, we built two organizations with corresponding encrypted material certificates, each organization includes two users and two peers. Each peer is responsible for maintaining the version of the ledger so that the network and data can be maintained even if other peers are shut down.

When the user initiates a request and sends it to the service. The back-end service processes the data and sends it to the smart contract API. When receiving the request and the data, the smart contract sends this to the peers in the network for authentication and data interaction purposes. During the

creation, querying or updating data processes, peers check the identity of the request to decide whether to allow access to the data at the distributed ledger. If the identified user of the request is not defined in the data collection, the system denies access and sends a message to the back-end API to notify the user; the system allows access and proceeds with further processing steps.

B. Hyperledger Component

The model in this paper is implemented on the Hyperledger Fabric platform. Fabric is a permissionless blockchain platform that integrates smart contracts, the storage of data to the distributed ledger is controlled through the smart contract APIs, from which the data is simplified and easily traced. Each request that goes through the smart contract is verified with public and private key pairs. In other words, if the user does not exist in the system, the system is better protected from malicious requests outside the system.

The Fabric system in this paper includes two organizations. Each organization consists of two peers to store smart contracts, where each peer registers two users and is authenticated with public and private key pairs. The components of the model are shown in Fig. 4

When user devices access the system to initiate/query or update data for a particular transaction, requests are sent from the client to the services of the existing system. Then, these services send access information to the peers belonging to the organization located in the blockchain network. At this step, the peers conduct verification of that user's key pairs, and if the successful peer authentication process proceeds to send information to the smart contract with the transaction type declared in a smart contract requested by the user, the smart contract will go through the designed features function to access the distributed ledger to initiate/query or update specific data.

C. Our Proposed Model's Diagram

One of the most important parts of the model lies in the validation and interaction with the patient's global ID and their medical data described in Fig. 5 and 6. In particular, the main functions include initializing and querying the patient's global ID and their medical data.

Fig. 5 depicts the process of storing new record data (e.g., patients' global ID and their medical data). In step 1, when the user initializes information about a certain ID, the data is sent to the back-end service of the health center's information management system. In the next step, the back-end APIs (i.e., backend) check, authenticate, and initialize the default values, then pass the parameters to the API inside the smart contract. At this point, a smart contract transfers data and stores transactions to the distributed ledger of the blockchain network. The default values for parameters sent from the request are intended to minimize errors caused by null field data.

Fig. 6 presents the process of retrieving data of a particular (e.g., patients' global ID and their medical data). When the user sends a query request to the system, the service query data checks and confirms whether the parameter ID of their

Fig. 2. The Proposed Model for Blockchain-Based Medical Test Results Management System.

Fig. 3. Permission Diagram.

Fig. 5. Initializing and Storing the new Data.

Fig. 4. Hyperledger Fabric Component.

Fig. 6. Retrieving/Querying Data Process

medical data exists or not. Then, the smart contract's APIs are called and passed into the corresponding parameter. Next, the smart contract's APIs check for the existence of data in the request before querying. In the case that the ID does not exist, the smart contract sends an error notification to the user's device; otherwise, it returns the relevant data/record of the patient corresponding to the requested ID.

V. EVALUATION SCENARIOS

A. Environment Setting

Our paradigm is deployed on the Hyperledger Fabric network maintained inside docker containers. In this section, we measure the performance of chaincode in the two scenarios: initializing (i.e., creating data) and accessing data. The exper-

iments are deployed on Ubuntu 20.01 configuration, core i5 2.7Ghz, and 8GB RAM.

To prove the effectiveness of our model, we also define several experiments by exploiting the Hyperledger Caliper[3] that is used to design the test scenarios and collect all the information regarding the performance.

B. Results

1) Data Creation: In this scenario, the study measures the performance of the data initialization function/data created (e.g., medical record book) performed through smart contracts. The number of requests sent simultaneously from two users[4]. Table I shows the execution results of the data initialization/creation function (e.g., medical record book). The data initialization/creation script is conducted with two users concurrently making 1000 - 6000 requests to the system. We measure the parameters of command success/failure and system latency (i.e., max, min, avg). Based on the execution results in Table I, it can be seen that the number of successful and failed requests is stable (except in the case of 6000 requests). Specifically, the number of failed requests is limited to less than 7,500 (i.e., 7,458 requests - 16.24%). Meanwhile, the lowest case was with only 6.57% (2,953 requests). The highest failed request rate is in the first 1,000/s request, the system is more stable in terms of data creation with only an average of 5K errors per scenario (from 2K requests to 6K). For system-wide latency, we recorded the number of requests with response delays per 1,000 requests/second to 6,000 requests per second. Specifically, the data in Table I demonstrate that the highest latency ranges from 1,626.57 to 1,781.15 seconds. The minimum is less than 8 seconds. The average delay when creating new data is less than 900 seconds. This is acceptable because creating thousands of new records at the same time is very unlikely in medical centers. The results observed in this scenario also demonstrate that the system supports very well with the continuous generation of new profiles.

2) Data Access (Retrieving/Querying): In the second experiment, we consider the data access (e.g., medical record book). We also set up 6 scenarios from 1000 to 6000 requests which access the medical record book from 2 users. Table II shows the execution results of the data access function (e.g., medical record book). Compared with the first task (i.e., data creation), the results of 6 scenarios to evaluate the data accessibility of our proposed blockchain-based system are more balanced. Retrieval of stored data is extremely important. Indeed, considering health data retrieval time directly affects the patient's health care.[5] To solve this problem, we consider the latency of the system (i.e., the maximum/average/minimum time it takes to process the request of data accessed from the system). Specifically, the maximum time to wait for a data retrieval request is 15 seconds (Note: all of our simulation scenarios use single information retrieval/querying data - not concluding. complex access requirements, such as join and group by commands like database management systems on SQL). The minimum wait time is almost instant response

(i.e., with only 0.06 seconds). The average time for each data retrieval request is about 7.35 seconds. Given the number of successful and failed requests, we also collect the number of requests at every 1000 to 6000 requests per second. The number of successful and failed requests is fairly balanced, around 80% of the requests are successful in all 6 scenarios.

3) Data Edit/Update: Finally, we look at the user's ability to update the medical test result's data. This parameter reflects whether a doctor or nurse updates information about a patient's medical record (e.g., new symptoms, diagnoses). In this scenario, we also conduct a review of six different scenarios, each of which will require processing from 1000 to 6000 requests per second. We also measure two parameters, similar to the two scenarios above, the number of successful and failed requests and the overall latency, which is shown in Table III. In terms of time, updating data is more complex than the previous two scenarios (i.e., initialization and access). Specifically, we must determine if user information exists in Hyperledger, then we determine which information needs updating (e.g., symptoms, disease diagnoses). Because of the above requirement, the execution time for the update task is longer. Specifically, the latency of all scenarios ranges from 850 to 950 seconds in the maximum case. The minimum latency ranges from 0.5 seconds to 0.6 seconds, while the average latency required by an application to process ranges from 370 seconds to 400 seconds. Similarly, the number of failed requests was also higher than the success requests in all six scenarios (with an average of about 51%).

VI. Discussion

Comparing all three evaluation scenarios, we find that real-time is acceptable. We also describe why there is a difference between the time lag in the execution of requests from the system depending on the complexity of the query. Specifically, the most prolonged time lag was recorded in data initialization due to updating to Hyperledger. This is different from the traditional way of storing data, where the information is only stored in tables and is done by the system administrator. On the contrary, initiating a medical test result requires confirmation from all relevant parties. In addition, defining constraints in an update request is more complex than in a retrieval request. The update time clearly defines the information the requester wishes to add/update to the existing medical test results. Finally, the fastest execution time is the data retrieval request which offers more promise for a Blockchain-based system than traditional storage systems.

However, Section V provides a marked change in all three data creation, retrieval, and update scenarios regarding the number of success and failed requests. Specifically, in the update scenario, the failure rate of requests is much higher than in the data initialization scenario (with more than 50% compared to less than 20%). A similar method occurs when comparing data retrieval and initialization with more than 20% and less than 20% of failed requests, respectively. This happens because we build a system that simulates the interactions between the parties (e.g., patient, nurse, doctor). In particular, the update and retrieval request must require the data to be initialized before. Otherwise, the request is considered a failure. For the update scenario, the system also requires that the updated information be initialized before being replaced

---

[3]https://www.hyperledger.org/use/caliper

[4]We set up one organization with two users and two peers

[5]In this scenario, we do not include time for encoding and decoding. code the data stored on Hyperledger

TABLE I. DATA CREATION/INITIALIZATION (I.E., MEDICAL RECORD BOOK) FOR THE MEDICAL TEST RESULTS OF THE PATIENT

Number of requests	Max Latency (s)	Min Latency (s)	Avg Latency (s)	Success	Fail
1,000	1,627.11	7.23	817.17	38,479	7,458
2,000	1,781.15	5.21	893.18	41,962	2,953
3,000	1,626.57	5.40	815.99	38,504	5,331
4,000	1,659.02	6.94	832.98	39,375	5,417
5,000	1,744.65	5.93	875.29	39,824	6,347
6,000	1,765.50	6.67	886.09	40,136	5,261

TABLE II. DATA ACCESS (I.E., MEDICAL RECORD BOOK) FOR THE MEDICAL TEST RESULTS OF THE PATIENT

Number of requests	Max Latency (s)	Min Latency (s)	Avg Latency (s)	Success	Fail
1,000	11.78	0.03	7.43	91,127	22,307
2,000	13.99	0.01	7.45	91,307	26,322
3,000	13.35	0.02	7.37	92,325	25,519
4,000	12.54	0.01	7.34	91,674	26,785
5,000	11.56	0.02	7.32	92,047	26,622
6,000	14.35	0.04	7.33	91,408	27,044

TABLE III. DATA UPDATE (I.E., MEDICAL RECORD BOOK) FOR THE MEDICAL TEST RESULTS OF THE PATIENT

Number of requests	Max Latency (s)	Min Latency (s)	Avg Latency (s)	Success	Fail
1,000	852.24	0.56	373.31	12,971	13,835
2,000	856.32	0.59	374.80	12,991	14,099
3,000	889.55	0.55	375.58	12,984	14,083
4,000	919.85	0.60	373.73	13,039	14,129
5,000	931.84	0.51	391.16	13,027	14,178
6,000	852.14	0.59	375.52	12,896	14,052

with new data (e.g., patient information and medical history). Initializing a dummy data system according to the above requirements is extremely difficult because we do it on two separate user groups.

For the system specification, we have not included encryption and decryption times for the data stored on Hyperledger. We assume that a trusted third party will take care of this. In terms of execution time, including the user critical generation time, as well as encryption and decryption, will increase the execution time for the whole system. This is hard to meet on our simulation system. In addition, this proposed model is also the first attempt to build a blockchain-based system that aims to offer a test management model in medical centers in developing countries. We intend many potential research directions to follow after this work. One of the mandatory requirements for health systems is confidentiality (i.e., authentication and authorization). We apply the proposed models based on the dynamic data support the environment of IoT devices [38], [39]. For authorization, a model based on ABAC [32], [31] and supporting dynamic policy [40], [41] is an appropriate choice in the context of the current health system. For encryption requirements, we use a trusted authority that provides a solution to store and protect patient data on Hyperledger [42].

VII. CONCLUSION

In this work, we propose a test process management system based on Blockchain technology. The main contributions of our solution are threefold: (a) building a medical test result management system based on Blockchain and Smart contract; (b) building proof-of-concept on top of Hyperledger Fabric; and (c) assessing the appropriateness of the approach based on an analysis of three main scenarios (i.e., initialize, retrieve, and update) based on the Hyperledger Caliper platform.

Specifically, (**a**) all user-related information as well as test results, diagnoses, and patient medical records are stored on Hyperledger (distributed ledger). All this stored information is authenticated by the relevant parties (i.e., patient, nurse/-doctor). We also offer a traditional test process management system. Thereby, we compared it with our proposed model before implementing proof-of-concept implementation on Hyperleger Fabric platform (i.e., (**b**)). In evaluating the feasibility of the proposed system, we analyze 3 key tasks of a Blockchain-based system (i.e., data initialization, retrieval and update) on six scenarios requiring access from 1,000 to 6,000 requests/second. Comments and future directions are presented in the Discussion section of the paper. Based on the analysis results, we found that our proposed model works stably in the scenario of up to 6,000 incoming requests per second in a simulated environment with limited resources (i.e., (**c**)).

REFERENCES

[1] M. Du, Q. Chen, J. Xiao, H. Yang, and X. Ma, "Supply chain finance innovation using blockchain," *IEEE Transactions on Engineering Management*, vol. 67, no. 4, pp. 1045–1058, 2020.

[2] W. H. Organization *et al.*, *Increasing access to health workers in remote and rural areas through improved retention: global policy recommendations*. World Health Organization, 2010.

[3] S. S.-L. Tan and N. Goonawardene, "Internet health information seeking and the patient-physician relationship: a systematic review," *Journal of medical Internet research*, vol. 19, no. 1, p. e5729, 2017.

[4] N. Duong-Trung, H. X. Son, H. T. Le, and T. T. Phan, "Smart care: Integrating blockchain technology into the design of patient-centered healthcare systems," in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*, ser. ICCSP 2020, 2020, p. 105–109.

[5] ——, "On components of a patient-centered healthcare system using smart contract," in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*, 2020, p. 31–35.

[6] H. X. Son, T. H. Le, N. T. T. Quynh, H. N. D. Huy, N. Duong-Trung, and H. H. Luong, "Toward a blockchain-based technology in dealing with emergencies in patient-centered healthcare systems," in *International Conference on Mobile, Secure, and Programmable Networking.* Springer, 2020, pp. 44–56.

[7] N. T. T. Quynh, H. X. Son, T. H. Le, H. N. D. Huy, K. H. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, N. Duong-Trung *et al.*, "Toward a design of blood donation management by blockchain technologies," in *International Conference on Computational Science and Its Applications.* Springer, 2021, pp. 78–90.

[8] H. T. Le, T. T. L. Nguyen, T. A. Nguyen, X. S. Ha, and N. Duong-Trung, "Bloodchain: A blood donation network managed by blockchain technologies," *Network*, vol. 2, no. 1, pp. 21–35, 2022.

[9] M. Picha Edwardsson and W. Al-Saqaf, "Drivers and barriers for using blockchain technology to create a global fact-checking database," *Online Journal of Communication and Media Technologies*, vol. 12, no. 4, p. e202228, 2022.

[10] N. Duong-Trung, X. S. Ha, T. T. Phan, P. N. Trieu, Q. N. Nguyen, D. Pham, T. T. Huynh, and H. T. Le, "Multi-sessions mechanism for decentralized cash on delivery system," *Int. J. Adv. Comput. Sci. Appl*, vol. 10, no. 9, 2019.

[11] X. S. Ha, H. T. Le, N. Metoui, and N. Duong-Trung, "Dem-cod: Novel access-control-based cash on delivery mechanism for decentralized marketplace," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom).* IEEE, 2020, pp. 71–78.

[12] N. T. T. Le, Q. N. Nguyen, N. N. Phien, N. Duong-Trung, T. T. Huynh, T. P. Nguyen, and H. X. Son, "Assuring non-fraudulent transactions in cash on delivery by introducing double smart contracts," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, pp. 677–684, 2019.

[13] N. H. Tuan Khoi *et al.*, "Vblock - blockchain based traceability in medical products supply chain management: Case study in vietnam," in *International Conference on Artificial Intelligence for Smart Community*, 2020.

[14] H. T. Le, L. N. T. Thanh, H. K. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, K. H. N. Vuong, H. X. Son *et al.*, "Patient-chain: Patient-centered healthcare system a blockchain-based technology in dealing with emergencies," in *International Conference on Parallel and Distributed Computing: Applications and Technologies.* Springer, 2022, pp. 576–583.

[15] H. X. Son, M. H. Nguyen, N. N. Phien, H. T. Le, Q. N. Nguyen, V. Dinh, P. Tru, and P. Nguyen, "Towards a mechanism for protecting seller's interest of cash on delivery by using smart contract in hyperledger," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 4, pp. 45–50, 2019.

[16] H. H. Luong, T. K. N. Huynh, A. T. Dao, and H. T. Nguyen, "An approach for project management system based on blockchain," in *International Conference on Future Data and Security Engineering.* Springer, 2021, pp. 310–326.

[17] N. H. Tuan Khoi *et al.*, "Domain name system resolution system with hyperledger fabric blockchain," in *International Conference on Inventive Computation and Information Technologies*, 2022.

[18] X. S. Ha, T. H. Le, T. T. Phan, H. H. D. Nguyen, H. K. Vo, and N. Duong-Trung, "Scrutinizing trust and transparency in cash on delivery systems," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage.* Springer, 2020, pp. 214–227.

[19] P. Zhang, J. White, D. C. Schmidt, G. Lenz, and S. T. Rosenbloom, "Fhirchain: applying blockchain to securely and scalably share clinical data," *Computational and structural biotechnology journal*, vol. 16, pp. 267–278, 2018.

[20] V. Patel, "A framework for secure and decentralized sharing of medical imaging data via blockchain consensus," *Health informatics journal*, vol. 25, no. 4, pp. 1398–1411, 2019.

[21] M. Egorov, M. Wilkison, and D. Nuñez, "Nucypher kms: decentralized key management system," *arXiv preprint arXiv:1707.06140*, 2017.

[22] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," in *2010 proceedings ieee infocom.* Ieee, 2010, pp. 1–9.

[23] Z. Chen, W. Xu, B. Wang, and H. Yu, "A blockchain-based preserving and sharing system for medical data privacy," *Future Generation Computer Systems*, vol. 124, pp. 338–350, 2021.

[24] H. X. Son, M. H. Nguyen, H. K. Vo *et al.*, "Toward an privacy protection based on access control model in hybrid cloud for healthcare systems," in *International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on EUropean Transnational Education (ICEUTE 2019).* Springer, 2019, pp. 77–86.

[25] M. R. Patra, R. K. Das, and R. P. Padhy, "Crhis: cloud based rural healthcare information system," in *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance*, 2012, pp. 402–405.

[26] C. O. Rolim, F. L. Koch, C. B. Westphall, J. Werner, A. Fracalossi, and G. S. Salvador, "A cloud computing solution for patient's data collection in health care institutions," in *2010 Second International Conference on eHealth, Telemedicine, and Social Medicine.* IEEE, 2010, pp. 95–99.

[27] T. Makubalo, B. Scholtz, and T. O. Tokosi, "Blockchain technology for empowering patient-centred healthcare: A pilot study," in *Conference on e-Business, e-Services and e-Society.* Springer, 2020, pp. 15–26.

[28] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-cps: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Systems Journal*, vol. 11, no. 1, pp. 88–95, 2015.

[29] M. Barua, X. Liang, R. Lu, and X. Shen, "Espac: Enabling security and patient-centric access control for ehealth in cloud computing," *International Journal of Security and Networks*, vol. 6, no. 2-3, pp. 67–76, 2011.

[30] L. Chen and D. B. Hoang, "Novel data protection model in healthcare cloud," in *2011 IEEE International Conference on High Performance Computing and Communications.* IEEE, 2011, pp. 550–555.

[31] N. M. Hoang and H. X. Son, "A dynamic solution for fine-grained policy conflict resolution," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 116–120.

[32] H. X. Son and N. M. Hoang, "A novel attribute-based access control system for fine-grained privacy protection," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 76–80.

[33] Q. N. T. Thi, T. K. Dang, H. L. Van, and H. X. Son, "Using json to specify privacy preserving-enabled attribute-based access control policies," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage.* Springer, 2017, pp. 561–570.

[34] M. M. Madine, A. A. Battah, I. Yaqoob, K. Salah, R. Jayaraman, Y. Al-Hammadi, S. Pesic, and S. Ellahham, "Blockchain for giving patients control over their medical records," *IEEE Access*, vol. 8, pp. 193 102–193 115, 2020.

[35] M. Misbhauddin, A. AlAbdulatheam, M. Aloufi, H. Al-Hajji, and A. Al-Ghuwainem, "Medaccess: A scalable architecture for blockchain-based health record management," in *2020 2nd International Conference on Computer and Information Sciences (ICCIS).* IEEE, 2020, pp. 1–5.

[36] G. Zyskind, O. Nathan *et al.*, "Decentralizing privacy: Using blockchain to protect personal data," in *2015 IEEE Security and Privacy Workshops.* IEEE, 2015, pp. 180–184.

[37] K. L. Quoc, H. K. Vo, L. H. Huong, K. H. Gia, K. T. Dang, H. L. Van, N. H. Huu, T. N. Huyen, L. Van Cao Phu, D. N. T. Quoc *et al.*, "Sssb: An approach to insurance for cross-border exchange by using smart contracts," in *International Conference on Mobile Web and Intelligent Information Systems.* Springer, 2022, pp. 179–192.

[38] N. T. T. Lam, H. X. Son, T. H. Le, T. A. Nguyen, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, and H. V. K. Nguyen, "Bmdd: A novel approach for iot platform (broker-less and microservice architecture, decentralized identity, and dynamic transmission messages)," *International Journal of Advanced Computer Science and Applications*, 2022.

[39] H. H. Luong, T. D. Anh, K. N. H. Tuan, and H. X. Son, "Ioht-mba: An internet of healthcare things (ioht) platform based on microservice and brokerless architecture," 2021.

[40] S. H. Xuan, L. K. Tran, T. K. Dang, and Y. N. Pham, "Rew-xac: an approach to rewriting request for elastic abac enforcement with dynamic policies," in *2016 International Conference on Advanced Computing and Applications (ACOMP).* IEEE, 2016, pp. 25–31.

[41] H. X. Son, T. K. Dang, and F. Massacci, "Rew-smt: a new approach for rewriting xacml request with dynamic big data security policies," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2017, pp. 501–

515.

[42] M. Uddin, "Blockchain medledger: Hyperledger fabric enabled drug traceability system for counterfeit drugs in pharmaceutical industry," *International Journal of Pharmaceutics*, vol. 597, p. 120235, 2021.

Wheat Diseases Detection and Classification using Convolutional Neural Network (CNN)

Md Helal Hossen[1]
Student, Dept. of CSE
Comilla University
Cumilla, Bangladesh

Chowdhury Shahriar Muzammel[3]
Assistant Professor, Dept. of CSE
Comilla University
Cumilla, Bangladesh

Shuvra Acharjee[5]
Student, Dept. of CSE
Comilla University
Cumilla, Bangladesh

Md Mohibullah[2]
Assistant Professor, Dept. of CSE
Comilla University
Cumilla, Bangladesh

Tasniya Ahmed[4]
Assistant Professor
Institute of Information Technology
Noakhali Science & Technology University
Noakhali, Bangladesh

Momotaz Begum Panna[6]
Dept. of CSE
Jahangirnagar University
Dhaka, Bangladesh

Abstract—**Ever since the medieval era, the preponderance of our concentration has been concentrated upon agriculture, which is typically recognized to be one of the vital aspects of the economy in contemporary society. This focus on agriculture can be traced back to the advent of the industrial revolution. Wheat is still another type of grain that, in the same way as other types of harvests, satisfies the necessity for the essential nutrients that are required for our bodies to perform their functions correctly. On the other hand, the supply of this harvest is being limited by a variety of rather frequent ailments. This is making it difficult to meet demand. The vast majority of people who work in agriculture are illiterate, which hinders them from being able to take appropriate preventative measures whenever they are necessary to do so. As a direct consequence of this factor, there has been a reduction in the total amount of wheat that has been produced. It can be quite difficult to diagnose wheat illnesses in their early stages because there are so many various forms of environmental variables and other factors. This is because there are numerous distinct sorts of agricultural products, illiteracy of agricultural workers, and other factors. In the past, a variety of distinct models have been proposed as potential solutions for identifying illnesses in wheat harvests. This study demonstrates a two-dimensional CNN model that can identify and categorize diseases that affect wheat harvests. To identify significant aspects of the photos, the software employs models that have previously undergone training. The suggested method can then identify and categorize disease-affected wheat crops as distinct from healthy wheat crops by employing the major criteria described above. The reliability of the findings was assessed to be 98.84 percent after the collection of a total of 4800 images for this study. These images included eleven image classes of images depicting diseased crops and one image class of images depicting healthy crops. To offer the suggested model the capability to identify and classify diseases from a variety of angles, the photographs that help compensate for the collection were flipped at a variety of different perspectives. These findings provide evidence that CNN can be applied to increase the precision with which diseases in wheat crops are identified.**

Keywords—*Wheat crop diseases; artificial intelligence; convolution neural networks; image processing; feature extraction*

I. Introduction

In addition to terms of its civilization, but also terms of its cuisine, Bangladesh is a very diverse country. Bangladesh is primarily an agricultural nation, with 80% of its inhabitants dependent on the sector for their livelihood. Wheat is an excellent resource for a variety of minerals, including magnesium and selenium. Certain nutrients are essential for maintaining healthy health. Wheat leaves are particularly susceptible to damage from leaf rust. It has a high incidence of fungus illness, latent infection, and other diseases. Wheat diseases focus on leaves and can be diagnosed using Deep Learning and computer vision [1]. Wheat is the main carbohydrate source in most countries. Both wheat protein and wheat starch are readily absorbed by the body [2]. When wheat is combined with a tiny quantity of animal or legume protein, it becomes highly nutritious because it includes minerals, vitamins, and fats (lipids). One of the earliest plants that have been domesticated, wheat has long been a vital food source for many nations. The nutritional content of the wheat grain is very high. The wheat grain contains 14% protein, 14% starch, and other nutrients like fiber, vitamins, minerals, and high-quality amino acids. Wheat was previously essential to global nutrition due to its high nutritional value and excellent storage qualities. For the manufacturing of contemporary food, basic resources derived from wheat are crucial, such as wheat starch and wheat proteins. 20% of our daily caloric intake comes from bread and baked goods, which are essential to a healthy diet [3,4]. We find it fascinating that one of the oldest crops on the planet may employ modern processing techniques to offer a significant solution to fresh problems. In a variety of applications, the natural components of wheat increase quality and lower resource usage in the final product for the consumer. In the developing world, wheat accounts for more than 35% of the calories consumed from cereal, compared to 74% in the rich world and 41% globally. Almost 70% of wheat is utilized for food, whereas just 20% and 2% to 3% of it are used for industrial processing and animal feed, respectively. Between 2001/02 and 2016/17, the world's consumption of wheat increased by 25%. Two-thirds of the globe's wheat is utilized for food, 20% for animal feed, and 3% to 5% for seed,

industry, and other uses. The economy grew rapidly over a shorter time as a result of the transportation of wheat. Wheat might be used to make a variety of novel foods, including muffins, cereal, and bread. The largest producer of wheat in the world, China has produced over 2.4 billion tonnes of wheat in the past 20 years or around 17% of global production. China consumed roughly 148.5 million metric tons of wheat in the marketing year 2021–2022. In that year, the United States consumed close to 31 million metric tons of wheat. Precision agriculture, often known as artificial intelligence systems, is assisting in enhancing the overall quality and accuracy of harvests. AI technology aids in the detection of pests, plant diseases, and undernutrition on farms. Artificial intelligence (AI) sensors can identify and target weeds before deciding which herbicide to use in the area. The condition of a plant can be assessed by looking at its leaves. This effort aims to create a system based on statistical analysis and blob detection that can detect and classify different diseases. Precision agriculture, often known as artificial intelligence (AI) technologies, is assisting in enhancing harvest quality and accuracy. AI technology helps identify plant illnesses, pests, inadequate plant nutrients, etc. Additionally, it enables farmers to keep an eye on the condition of the soil and crops. After rice, wheat is Bangladesh's second-most significant crop for producing staple foods. Since independence, it has become more significant as a crop for food and nutrition security. From about 0.115 million tons in 1971–1972 to 0.73 million tons in 2005–2006, wheat production climbed considerably. Several study articles discuss the various problems associated with diagnosing wheat crop diseases, such as the frequencies of disease groups, the accurateness, and the dataset; however, there are a lot of areas that may be improved. We have attempted to classify all of the different diseases that can harm wheat crops in this article. We have decided to focus on this system due to our desire to aid our farming sector by fixing these challenges so that the rate of production of wheat may be increased. This is why we have chosen to work on this system. As a result, we are concentrating on the challenge of binary image classification, in which a picture of a leaf can either be assigned to a wheat crop that is healthy or one that is infected with a disease. Here, we used 12 classes, 11 of which were for photographs of crops that were sick, and one class was for a healthy wheat crop. The 12 varieties of wheat crops that we have access to for our research are as follows: Barley yellow dwarf, Black chaff, Common root rot, Fusarium head blight, Leaf rust, Powdery mildew, Tan spot, Wheat loose smut, Wheat soil-borne mosaic, Wheat streak mosaic, Karnal bunt with Healthy wheat crops samples. Following is how the rest of the article is organized. Section II of this study's literature review provides details on the previous study. The materials and procedures utilized to assess our strategy are covered in Section III, along with a description of our system, picture pre-processing, the Keras sequential model, and the datasets that were employed. Section IV contains a report on the experimental findings and related discussion. The conclusions are presented in Section V. Limitations were discussed in Section VI, along with suggestions for further study. In this study, a disease detection system for wheat employing a dataset of 4800 pictures and CNN models is described. The Keras-sequential model served as the foundation for the CNN model.

II. LITERATURE REVIEW

With automated wheat disease diagnostic system is what the authors of article [5] set out to demonstrate and verify. They were able to circumvent the challenges and achieve an average output of 97.95percent by employing VGG-FCNVD16. They have made use of the 2017 Wheat Disease Database (WDD2017),. There were four distinct categories of illness. K-means clustering is proposed as an automatic and successful strategy in this study. Image segmentation was utilized by the author [6] to identify diseases on wheat leaves. They have discovered that the accuracy rates are more than 90 percent for three main disorders (powdery mildew, leaf rust, and stripe rust). In [7], author developed a Computer Vision Framework for the Identification and Classification of Wheat Diseases Using Jetson GPU Infrastructure demonstrating that manually identifying and interpreting wheat illnesses requires a significant amount of time and effort. They demonstrated that the VGG19 model was accurate in identifying wheat disease 99.38% of the time. Article [8] offered an ML-based system for automatically identifying brown- and yellow-rust wheat diseases. They provided an efficient ML-based framework for wheat disease detection and classification. The suggested scheme surpassed known ML algorithms with 99.8% accuracy. PRI and ARI at different development stages were computed as all conceivable three-band combinations across such susceptible wavelengths, and their ability to predict yellow rust disease severity was examined by the author [9]. In the article [10], the author used a better deep convolutional architecture to find and classify leaf and spike wheat diseases. The study comes up with a brand-new way to group wheat diseases. A new deep learning system has been made that can appropriately put 10 different wheat diseases into the right category. The paper [11] conducted a thorough assessment of recently published data and addressed WD prediction methods for recognizing and categorizing wheat illnesses. A machine learning-based early detection strategy based on hyperspectral images was reported in [12]. This research initially extracted the normalized difference texture indices (NDTIs) and vegetation indicators to discriminate between wheat with powdery mildew and healthy wheat (VIs). Study [13] used three common diseases and an insect of winter wheat as examples to investigate the applicability of channels' reflectance and many conventional vegetation indicators (VIs) of seven high-resolution satellite sensors. A method for diagnosing five fungal diseases that harm wheat shoots was provided by the author in this article [14]. The dataset's 2414 images of wheat fungal infections were utilized (WFD2020). They discovered a 94.2% accuracy rate. In [15], the author advocated Qt-based applications. Using image processing, the identification rate is 96.2% and the accuracy is 92.3%, approximately similar to human eyesight. This approach identifies, diagnoses, and classifies crop illnesses. It's a field-inspecting agricultural robot. According to the paper, using certain wavelength ranges, reflectance measurements may be employed for disease diagnosis and discrimination in the early stages of infection. This method has been used to identify wheat powdery mildew. Article [16] describes PCR methods for detecting wheat illnesses and DNA regions for fungal identification. PCR-based approaches for detecting wheat pathogens need additional research. For diagnosing wheat leaf illnesses and their severity, Paper[17] proposed an algorithm based on Elliptical-Maximum Margin Criterion (E-

MMC) metric learning. The highest identification accuracy of the proposed method is 94.16%. Paper [18], presented an AI-based algorithm (ANN). After analyzing 300 test photos of wheat leaves, the proposed system successfully detected illness in 97% of instances. Research [19] examined the viability of employing the newly announced Sentinel-2 Multispectral Instrument (MSI) to discern between the several phases of yellow rust infection (healthy, mild, and severe) in winter wheat. In the article [20], Continuous wavelet analysis (CWA) was compared to traditional spectral features to diagnose yellow rust on leaves. Both phases exhibited perfect R2 and RMSE of 0.81 and 0.110. In the paper [21], high, medium, and low LAI values were utilized to identify wheat leaf rust in the canopy. Four machine learning (ML) techniques were developed to measure disease severity (DS) at the canopy scale: support vector regression, boosted regression trees, random forests regression, and Gaussian process regression. Author [22] developed a kernel discriminant technique (SVIKDA) for detecting yellow rust, aphid, and powdery mildew in winter wheat at the leaf and canopy level. The author of this study [23] presented a simple convolutional neural network (CNN) model named SimpleNet for detecting wheat illnesses such as glume blotch and scab in genuine field photos [24]. This research assessed the spectro-optical, photochemical reflectance index's (PRI) precision for assessing the yellow rust disease index (DI) in wheat (Triticum aestivum L.) and its utility in detecting the sickness using hyperspectral pictures. The precision of the earlier study's use of a different model is unclear. Additionally, although it is not ideal, they have used fewer datasets to train the CNN model and other models. The CNN model is currently essential in identifying plant diseases. For greater accuracy, image pre-processing is also desirable. Therefore, a CNN model based on a Keras-sequential model is suggested in this study paper to more reliably identify and categorize wheat illnesses. The following are the article's main factors that contribute:

- The far more recent identification and classification findings from the massive dataset utilizing the Keras-sequential model are given. Compared to previous CNN models, the testing model demonstrated more accuracy.

- During the model's training and testing, samples from 12 diseases that afflicted the class were taken, demonstrating how the CNN model can be useful in identifying wheat diseases.

- To allow the CNN model to recognize and categorize diseases from various angles, images were rotated at various angles.

- A total of 4800 recorded photos were used to create a dataset, demonstrating how using more datasets can improve the training accuracy of the CNN model.

III. Methodology and Design

In this day and age of advanced technology, the early diagnosis and identification of plant diseases are of the utmost importance. If diseases can be diagnosed in their early stages,

then it will be much simpler to administer the appropriate treatment. More importantly, the production of agricultural goods would not be affected in any way by any diseases. Farmers and agro-technologists would benefit more from the disease detection system if it were possible to transform it from a manual system to a machine system. The objective of this study is to create a machine-learning tool that employs a CNN model to identify and diagnose wheat illnesses early on. The proposed classification scheme was developed to identify and classify various diseases that can affect wheat. The model includes the following information:

- The various layers and the layering scheme of the model.

- The number of components for each output data dimension for every tier.

- The number of variables (strength training) for each tier.

- A list of all the design variables.

The following modules are included in the system that is being proposed: dataset collecting, image pre-processing, feature extraction, identification, and classification. These modules are currently being redesigned to be the flowchart that is seen below in Fig. 1:
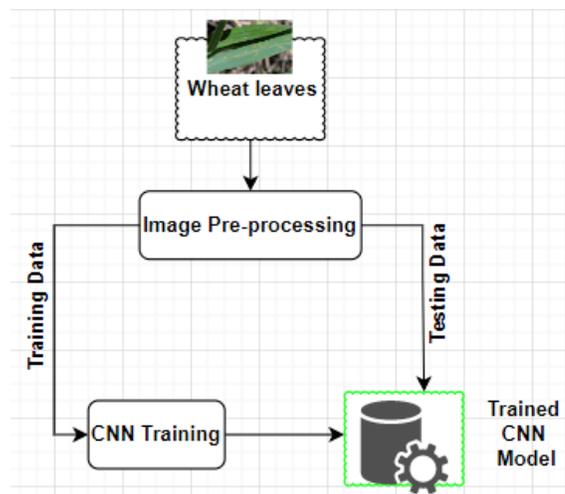
Fig. 1. Overview of the Proposed System.

The proposed system used the following algorithm:

1) Image pre-processing of the collected dataset.
2) Import the appropriate modules.
3) Load photos from the dataset.
4) Convert images to binary images and save them in a pixel array.
5) Image noise reduction
6) Define the input layer and build the model.
7) Encoded the layer as needed.
8) Train the model

TABLE I. NUMBER OF IMAGES TAKEN PER CLASS

Class Name	Number of Images
Barley yellow dwarf	400
Black chaff	400
Common root rot	400
Fusarium head blight	400
Healthy wheat	400
Leaf rust	400
Powdery mildew	400
Tan spot	400
Wheat loose smut	400
Wheat soil-borne mosaic	400
Wheat streak mosaic	400
Karnal bunt	400

9) Determine training accuracy
10) Test the model with various inputs
11) Obtain the model's testing accuracy.

The dataset utilized in the development of this suggested system was compiled using data obtained from a variety of online sources. Kaggle provided a dataset, while Github provided additional datasets, both of which were retrieved. After that, the two datasets were combined, and the result was a dataset consisting of 4,800 photos. This dataset included photographs of various areas of wheat leaves and roots that were impacted by the illness, as well as images of wheat leaves that were disease-free. These are shown in Table I below:

Pre-processing is a term used to describe activities with the lowest quality photographs. both the input and the output are intensity images at this level of abstraction. These famous pictures are of the same type as the original data that the sensor originally acquired, with a matrix of image function values typically used to depict an intensity image. (brightness). Pre-processing is intended to enhance the image. data that reduces unintentional distortions or enhances certain aspects of the image geometric changes of pictures (such as rotation, scaling, and translation) are categorized as pre-processing but are crucial for subsequent processing since similar approaches are employed here. Feature extraction tries to reduce the number of features in a dataset by extracting new ones from existing ones. This new, smaller set of features should then describe the bulk of the data included in the initial shipment of characteristics. In the object-based approach used by Feature Extraction to classify images, an object (also known as a segment) is a group of pixels with common spectral, spatial, and/or textural features. Using standard pixel-based classification approaches, the spectral data contained in each pixel is used to classify photographs. A deep learning model called CNN is used to handle data with a grid pattern, such as photographs. because of the high degree of precision it has, CNNs are utilized for picture categorization and recognition. Yann LeCun, a computer scientist, first proposed it in the late 1990s after becoming intrigued by how humans recognize objects visually. Utilizing CNNs is advantageous since they can create an internal representation of a two-dimensional image. This enables the model to pick up on location and size in different types of data structures, which is crucial when working with photos. Each plant class's diseases are taught to the CNN classifiers. The classifier, which has been trained to categorize various diseases in that plant, is called up using Level 2 results. A pre-trained CNN model called VGG16 is employed for picture classification. It has been fine-tuned to
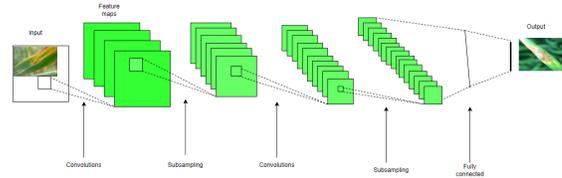
Fig. 2. Image Classification Process using CNN.

TABLE II. THE MODEL SUMMARY FOR THE PROPOSED SYSTEM

Layer (type)	Output Shape	Param #
sequential (Sequential)	(32, 256, 256, 3)	0
conv2d (Conv2D)	(32, 254, 254, 32)	896
maxpooling2d (MaxPooling2D)	(32, 127, 127, 32)	0
conv2d1 (Conv2D)	(32, 125, 125, 64)	18496
maxpooling2d1 (Max Pooling 2D)	(32, 62, 62, 64)	0
conv2d2 (Conv2D)	(32, 60, 60, 64)	36928
maxpooling2d2 (Max Pooling 2D)	(32, 30, 30, 64)	0
conv2d3 (Conv2D)	(32, 28, 28, 64)	36928
maxpooling2d3 (Max Pooling 2D)	(32, 14, 14, 64)	0
conv2d4 (Conv2D)	(32, 12, 12, 64)	36928
maxpooling2d4 (Max Pooling 2D)	(32, 6, 6, 64)	0
conv2d5 (Conv2D)	(32, 4, 4, 64)	36928
maxpooling2d5 (Max Pooling 2D)	(32, 2, 2, 64)	0
flatten (Flatten)	(32, 256)	0
dense (Dense)	(32, 64)	16448
dense1 (Dense)	(32, 12)	780

easily suit picture classification datasets after being trained on a sizable and diverse dataset. Convolutional Neural Networks have three different kinds of layers: Hidden Convolutional Layer connects each input neuron. CNN's input neurons are partially connected to the buried layer. Pooling shrinks the map. CNN has activation and pooling layers. Last is Fully Connected Tiers. The last pooling or convolutional layer output is flattened and applied. Fig. 2 shows the image classification process using CNN. The model summary for the proposed system is given in Table II.

A variety of APIs is available from Keras-sequential that can be used to define neural networks. Sequential API, Functional API, and Model Sub-classing API are all different names for APIs. Layer by layer can be built up a model using a sequential API. One input layer, one hidden layer, and one output layer make up the Keras-sequential model. Two neurons are joined to form the hidden layer.

IV. EXPERIMENTAL SETUP AND RESULT ANALYSIS

The previous section showed how to use a CNN model to classify and identify wheat illnesses. In this section, the theoretical process that was described in our previous section will be converted to a computational function so that using the digital device for image processing the diseases of wheat can be identified and classified. This experimental phase will be deployed using algorithm construction, deploying the algorithm via computational machine and output evaluation. We will be able to take the necessary measures for additional research and assess a good identical system that will work more accurately and perfectly based on the results of the performance evaluation that we conduct. All of the experimental data, as well as the experimental design and the analysis of the system, are given throughout this section. There are many different

TABLE III. TRAINING AND VALIDATION ACCURACY WITH LOSS

Epoch	Total Steps Per Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	30	0.9555	0.1023	0.9062	0.3057
50	30	0.9778	0.0628	0.9479	0.1309
100	30	0.9862	0.0403	0.9479	0.1245
150	30	0.9873	0.0333	0.9792	0.1049
200	30	0.9882	0.0118	0.9797	0.0203
250	30	0.9901	0.0099	0.9812	0.0188
300	30	0.9979	0.0021	0.9861	0.0139
350	30	1.0000	0.0000	0.9933	0.0067
400	30	0.9912	0.0088	0.9891	0.0102
450	30	0.9812	0.0188	0.9803	0.0197
500	30	0.9801	0.0199	0.9791	0.0209

Fig. 5. Comparison between Epoch vs Training Loss.

Fig. 6. Comparison between Epoch vs Validation Accuracy.

forms of accuracy in the methods used to evaluate systems, and they are all covered here in that they can be of great use in one's subsequent academic endeavors. This research project was performed in the rainy season but this is not the production season of wheat. For this reason, the dataset for this research project was collected from a variety of online resources. Then, 4800 photos of wheat crops with 11 diseases and healthy wheat crop images were mixed. 3840 photos were utilized for preparing the model, 480 for validation, and 480 for testing the system. 32-batches were used. The following graphic in Fig. 3 shows training, validating, and testing dataset ratios:
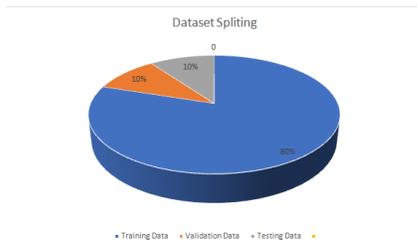
Fig. 3. Dataset Splitting Ratio for the Proposed System.

Fig. 7. Comparison between Epoch vs Validation Loss.

After pre-processing all the images, the feature vector was created using feature extraction. Then this feature vector was used for training the CNN model. A total number of 500 epochs were moved to the dataset for training and validation testing. There was a total of 30 steps per epoch. Using these feature vectors the model was trained and a validation test was performed for checking the validity of the dataset (Fig. 4to 7). The measurement shows in Table III.

Fig. 4. Comparison between Epoch vs Training Accuracy.

After completing training and validation testing on the dataset using the proposed model these scores were stored for prediction and testing of the system. Following the testing had been completed on the dataset, the F-1 score and recall value were then determined. After performing all the processes it was found that the training accuracy was 100%, the validation accuracy is 99.33%, and the testing accuracy is 98.84%. A comparison between training and validation accuracy is shown below in Fig. 8:
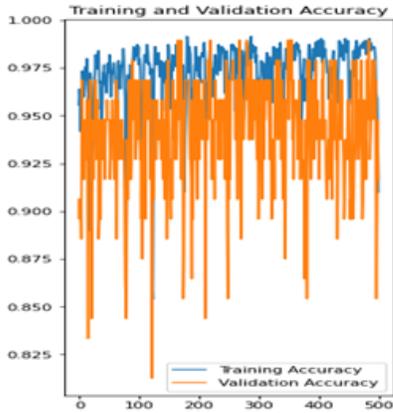
Fig. 8. Training vs. Validation Accuracy.

TABLE IV. RECALL VALUE, F1-SCORE, PRECISION VALUE PER CLASS

Class	Recall	F-1 Score	Precision
0	1.0	0.99	0.98
1	1.0	1.0	0.99
2	0.99	1.0	1.0
3	0.98	1.0	0.99
4	1.0	1.0	1.0
5	0.97	0.97	1.0
6	1.0	0.99	1.0
7	1.0	0.98	1.0
8	0.99	1.0	0.97
9	0.97	0.98	0.97
10	1.0	1.0	1.0
11	0.98	0.99	0.98
Avarage	0.99	0.99	0.99

Comparison of training and validation loss is given below in Fig. 9:
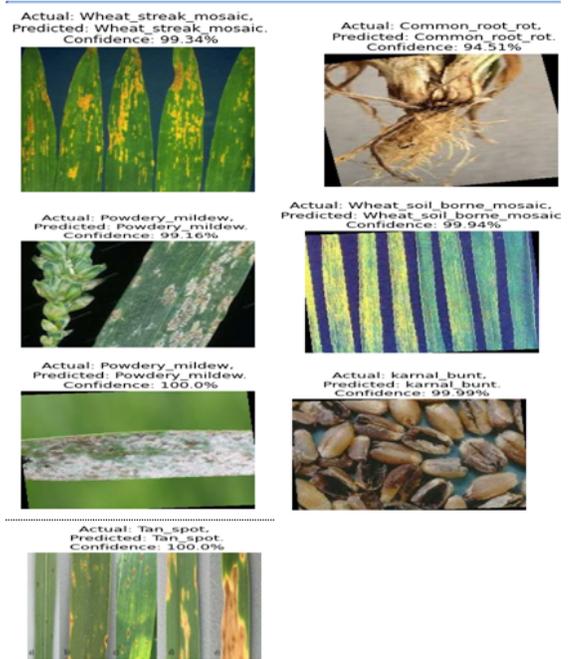
Fig. 10. Prediction Result based on Samples.

Following the Recall value, the F-1 Score, and Precision values were obtained after calculating the confusion matrix shown in Table IV.

When compared to the accuracy of the other proposed systems, the accuracy of this system is adequate to be considered satisfactory. Even though the accuracy of the research [24] was measured at 99.8 percent, the authors chose to concentrate their attention on two disease categories, namely, brown and yellow rust disease. The following Table V provides a comparison of the following:

V. CONCLUSION

This study presents a CNN-based paradigm for wheat illnesses. We gathered high-quality images of wheat leaf diseases such as barley yellow dwarf, black chaff, common root rot, fusarium head blight, leaf rust, powdery mildew, tan spot, wheat loose smut, wheat soil-borne mosaic, wheat streak mosaic, and karnal bunt from Bangladeshi agricultural websites to evaluate the proposed system. We also gathered samples

Fig. 9. Comparison of Training and Validation Loss.

Finally, some samples were tested for checking the confidence and testing accuracy of this proposed system. Average confidence for identifying and classifying varied from 93%-100%. The average F-1 score for all classes was 99%. The average recall value and precision value were 0.99 and 0.99. The following Fig. 10 is showing these prediction results based on the proposed system:

TABLE V. COMPARISON BETWEEN THE PROPOSED METHOD AND OTHER PROPOSED METHODS BY OTHER AUTHORS

Model Proposed	Number Of Classes	Dataset	Accuracy
K-Mean Clustering [25]	Three classes including powdery mildew, leaf rust, stripe rust		90%
MTMF + NDVI	Two classes including powdery mildew and leaf rust	Onfield Experiment	88.6%
Machine Learning [26]	Two classes brown- and yellow-rusted disease	1050 samples	99.8%
Photo chemical Reflectance Index + Anthocyanin Reflectance Index [27]	One class including Yellow rust	Onfield Experiment	93.2%
Deep Convolutional neural network [28]	One class including Yellow rust	Onfield Research	85%
Proposed System	Twelve classes include Barley yellow dwarf, Black chaff, Common root rot, Fusarium head blight, Healthy wheat, Leaf rust, Powdery mildew, Tan spot, Wheat loose smut, Wheat soil-borne mosaic, Wheat streak mosaic, and karnal bunt	4800 samples	98.84%

of healthy wheat crops. Techniques such as segmentation and resizing are utilized so that the pre-processing can be precise. The use of distinct feature descriptors allows for the extraction and combination of a wide variety of features, including shape, color, and texture features. Following the completion of the comparative study, the proposed CNN model was found to have higher levels of accuracy. Accuracy, precision, recall, and F1-score were used to evaluate the suggested methodology on unseen data. A comparison between our method and the other methods already in use is carried out as part of the ongoing evaluation process. As a consequence of this, it has been determined that our approach is superior to other methods in terms of the accuracy with which it recognizes and categorizes wheat diseases.

VI. FUTURE WORK

Based on the experiments, the approach has the potential to help farmers and agri-technologists measure wheat disease. As a consequence of this, they would be able to implement necessary precautions for the control and prevention of disease. A software program will also be investigated in greater depth to assist farmers in discovering wheat diseases as quickly as is practically possible. This will help agricultural workers quickly identify wheat illnesses. In addition, the suggested system will be capable to manage additional disease classes, which will allow it to manage a greater number of illness types. Farmers would benefit more from this approach if there were a greater variety of crop diseases included in it. In addition, it was observed that adopting multidimensional CNN could provide more accuracy for detecting disease. Multidimensional CNN can recognize output from a variety of inputs. Furthermore, it would be preferable if such a model could be created to acquire numerous photos and diagnose disease in each of them at the same time.

REFERENCES

[1] Shewry, Peter R., and Sandra J. Hey. "The contribution of wheat to human diet and health." Food and energy security 4.3 (2015): 178-202.

[2] Murdia, L. K., et al. "Maize utilization in India: An overview." American Journal of Food and Nutrition 4.6 (2016): 169-176.

[3] Nkhata, Smith G., et al. "Fermentation and germination improve nutritional value of cereals and legumes through activation of endogenous enzymes." Food science & nutrition 6.8 (2018): 2446-2458.

[4] Iqbal, Muhammad Javid, Naureen Shams, and Kalsoom Fatima. "Nutritional Quality of Wheat." Wheat. IntechOpen, 2022.

[5] Lu, J., Hu, J., Zhao, G., Mei, F., & Zhang, C. (2017). An in-field automatic wheat disease diagnosis system. Computers and electronics in agriculture, 142, 369-379.

[6] Niu, X., Wang, M., Chen, X., Guo, S., Zhang, H., & He, D. (2014, August). Image segmentation algorithm for disease detection of wheat leaves. In Proceedings of the 2014 International Conference on Advanced Mechatronic Systems (pp. 270-273). IEEE.

[7] Aboneh, T., Rorissa, A., Srinivasagan, R., & Gemechu, A. (2021). Computer Vision Framework for Wheat Disease Identification and Classification Using Jetson GPU Infrastructure. Technologies, 9(3), 47.

[8] Khan, H., Haq, I. U., Munsif, M., Khan, S. U., & Lee, M. Y. (2022). Automated Wheat Diseases Classification Framework Using Advanced Machine Learning Technique. Agriculture, 12(8), 1226.

[9] Zheng, Q., Huang, W., Cui, X., Dong, Y., Shi, Y., Ma, H., & Liu, L. (2018). Identification of wheat yellow rust using optimal three-band spectral indices in different growth stages. Sensors, 19(1), 35.

[10] Goyal, L., Sharma, C. M., Singh, A., & Singh, P. K. (2021). Leaf and spike wheat disease detection & classification using an improved deep convolutional architecture. Informatics in Medicine Unlocked, 25, 100642.

[11] Kumar, D., & Kukreja, V. (2022). Deep learning in wheat diseases classification: A systematic review. Multimedia Tools and Applications, 1-45.

[12] Khan, I. H., Liu, H., Li, W., Cao, A., Wang, X., Liu, H., ... & Yao, X. (2021). Early detection of powdery mildew disease and accurate quantification of its severity using hyperspectral images in wheat. Remote Sensing, 13(18), 3612.

[13] Yuan, L., Zhang, H., Zhang, Y., Xing, C., & Bao, Z. (2017). Feasibility assessment of multi-spectral satellite sensors in monitoring and discriminating wheat diseases and insects. Optik, 131, 598-608.

[14] Genaev, M. A., Skolotneva, E. S., Gultyaeva, E. I., Orlova, E. A., Bechtold, N. P., & Afonnikov, D. A. (2021). Image-based wheat fungi diseases identification by deep learning. Plants, 10(8), 1500.

[15] Xu, P., Wu, G., Guo, Y., Yang, H., & Zhang, R. (2017). Automatic wheat leaf rust detection and grading diagnosis via embedded image processing system. Procedia Computer Science, 107, 836-841.

[16] Kuzdraliński, A., Kot, A., Szczerba, H., Nowak, M., & Muszyńska, M. (2017). A review of conventional PCR assays for the detection of selected phytopathogens of wheat. Microbial Physiology, 27(3), 175-189.

[17] Bao, W., Zhao, J., Hu, G., Zhang, D., Huang, L., & Liang, D. (2021). Identification of wheat leaf diseases and their severity based on elliptical-maximum margin criterion metric learning. Sustainable Computing: Informatics and Systems, 30, 100526.

[18] Majumdar, D., Kole, D. K., Chakraborty, A., & Majumder, D. D. (2015, August). An integrated digital image analysis system for detection, recognition and diagnosis of disease in wheat leaves. In Proceedings of the Third International Symposium on Women in Computing and Informatics (pp. 400-405).

[19] Zheng, Q., Huang, W., Cui, X., Shi, Y., & Liu, L. (2018). New spectral index for detecting wheat yellow rust using Sentinel-2 multispectral imagery. Sensors, 18(3), 868.

[20] Zhang, J., Pu, R., Loraamm, R. W., Yang, G., & Wang, J. (2014). Comparison between wavelet spectral features and conventional spectral features in detecting yellow rust for winter wheat. Computers and Electronics in Agriculture, 100, 79-87.

[21] Azadbakht, M., Ashourloo, D., Aghighi, H., Radiom, S., & Alimohammadi, A. (2019). Wheat leaf rust detection at canopy scale under different LAI levels using machine learning techniques. Computers and Electronics in Agriculture, 156, 119-128.

[22] Shi, Y., Huang, W., Luo, J., Huang, L., & Zhou, X. (2017). Detection and discrimination of pests and diseases in winter wheat based on spectral indices and kernel discriminant analysis. Computers and Electronics in Agriculture, 141, 171-180.

[23] Bao, W., Yang, X., Liang, D., Hu, G., & Yang, X. (2021). Lightweight convolutional neural network model for field wheat ear disease identification. Computers and Electronics in Agriculture, 189, 106367.

[24] Wenjiang, H., Yue, S., Yingying, D., Huichun, Y., Mingquan, W., Bei, C., & Linyi, L. (2019). Progress and prospects of crop diseases and pests monitoring by remote sensing. Smart Agriculture, 1(4), 1.

[25] Niu, X., Wang, M., Chen, X., Guo, S., Zhang, H., & He, D. (2014, August). Image segmentation algorithm for disease detection of wheat leaves. In Proceedings of the 2014 International Conference on Advanced Mechatronic Systems (pp. 270-273). IEEE.

[26] Khan, H., Haq, I. U., Munsif, M., Khan, S. U., & Lee, M. Y. (2022). Automated Wheat Diseases Classification Framework Using Advanced Machine Learning Technique. Agriculture, 12(8), 1226.

[27] Zheng, Q., Huang, W., Cui, X., Dong, Y., Shi, Y., Ma, H., & Liu, L. (2018). Identification of wheat yellow rust using optimal three-band spectral indices in different growth stages. Sensors, 19(1), 35.

[28] Zhang, X., Han, L., Dong, Y., Shi, Y., Huang, W., Han, L., ... & Sobeih, T. (2019). A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral UAV images. Remote Sensing, 11(13), 1554.

An Effective Decision-Making Support for Student Academic Path Selection using Machine Learning

Pélagie HOUNGUE
Institut de Mathématiques et de
Sciences Physiques
Université d'Abomey-Calavi
Dangbo, Bénin

Michel HOUNTONDJI
Institut de Mathématiques et de
Sciences Physiques
Université d'Abomey-Calavi
Dangbo, Bénin

Théophile DAGBA
Ecole Nationale d'Économie
Appliquée et de Management
Université d'Abomey-Calavi
Cotonou, Bénin

Abstract—In Benin, after the GCSE (General Certificate of Secondary Education), learners can either enroll in a Technical and Vocational Education and Training (TVET), or further their studies in the general education. Majority of those who take the latter path enroll in Senior High School by choosing the Biology stream or field of study. However, most of them do not have the abilities required to succeed in this field. For instance, for the last edition of the Senior Secondary Education Certificate (French baccalaureate) held in June 2022 in Benin, the Biology field of study had a low success rate of 42%. Therefore, one may consider that there is a problem in the orientation of the students. In recent years, Machine Learning has been used in almost every field to optimize processes or to assist in decision-making. Improving academic performance has always been of general interest. And, good academic performance implies good academic orientation. The goal of this study is to optimally help learners who have just obtained their GCSE to select their field of study. For this purpose, two major elements are predicted: i) Scientific or Literary ability of students, ii) Literature or Mathematics and Physical Sciences (MPS) or Biology stream of learners. More precisely, the average marks in Mathematics, Physics and Chemistry Technology (PCT) and Biology from 6th to 9th grade for 325 students are used. Machine Learning algorithms such as Decision Tree, Random Forest, Linear Support Vector Classifier (SVC), K-Nearest Neighbors (KNN), and Logistic Regression are used to predict learners' ability and the stream. As a result, for learners' ability prediction, we obtained the best accuracy of 99% with the random forest algorithm for a split that reserved around 21% of the dataset for testing. As for the learners' stream prediction, we obtained the best accuracy of 95% with the Linear SVC algorithm for a split that reserved around 20% of the dataset for testing. This study contributes to Educational Data Mining (EDM) by performing academic data exploration using numerous methods. Furthermore, it provides a tool to ease students academic path selection, which may be used by educational institutes to ensure student performance. This paper presents the steps and the outputs of the study, we performed with some recommendations for future research.

Keywords—*Academic path; academic performance; machine learning; educational data mining*

I. INTRODUCTION

The improvement of academic performance has always been a concern for the educational system's actors [1], because the supposed performances in schools no longer satisfy everyone expectations. This situation requires a formula that could invert the trend. The education system is challenged to find a scientific instrument to overcome practices which continue to promote this deleterious situation [2].

According to [3], factors that impact field of study choice include: ability, experience, habit, program, instructor's role, university/school atmosphere and study culture. Authors in [4] also acknowledged that the field of study chosen by a student is in relation with prior knowledge and judgment of his/her own competence. Students often engage with peers and educational institutes through social networking to gather information about the university/school's fields of study or forthcoming courses. Therefore, deciding on a academic path can be stressful for secondary education students. Authors in [5] think that students need group guidance for fields of study and major choices.

If good or bad academic results have brought praise to some schools or tarnished the image of others, it is because of the lack of proper use of the data to direct students in the different fields of study. In fact, in Benin's high schools, students have to further their studies in general education track (Literature, Mathematics and Physical Sciences, Biology) or in Technical and Vocational Education and Training (TVET). Orientation in the TVET is not systematic because students must take an entrance exam before being admitted. Therefore, orientation problem is more acute in general education because it is done at best, on the basis of the marks obtained at the General Certificate of Secondary Education (GCSE). Otherwise, learners are oriented according to their parents' choice.

The use of Artificial Intelligence (AI) in general and more particularly, Machine Learning, in almost all fields, allows nowadays, to predict from the available data, a number of interesting elements for decision-making [6][7]. Thus, it is easy to understand the importance of using Machine Learning to improve the quality of academic performance [8][9]. This study contributes to research in Educational Data Mining (EDM) by developing a prediction model for academic orientations [10][11], in high schools in Benin. The study is articulated in two parts. We predict: (1) learner's aptitude or ability (literary or scientific), (2) learner's fields of study (Literature, Mathematics and Physical Sciences-MPS, Biology).

In the remaining of the paper, we address the background concepts of the performed study. Then, a state-of-the-art analysis is performed. Subsequently, the proposed prediction model' architecture is released, followed by the performance evaluation results. Furthermore, we highlight discussions and give an overview of the application that shows a concrete use of the optimal model. Finally, we summarize contributions and limitations of the proposed model and give the conclusions.

II. BACKGROUND

In this section, we briefly discuss the concept of Educational Data Mining (EDM) and describe the beninese education system.

A. Educational Data Mining

Nowadays, many research activities are interested in data mining, and EDM has become a promising field of research [12], [13]. EDM uses several algorithms to improve educational results and account for educational procedures in future decision-making. EDM can be defined as the techniques for finding the specific types of data coming from the education system and implementing these techniques to better understand students and the system [14] [15]. Some applications of EDM can likely be a recommender system for students and prediction of their performance.

B. The Beninese Education System

After two years in Preschool (optional), students must complete six years in Primary school to achieve and obtain the Primary School Certificate (PSC) [16]. Primary school years are: Year 1 (Introductory Courses - IC), Year 2 (Preparatory Courses - PC), Year 3 (Elementary Courses 1st year - EC1), Year 4 (Elementary Courses 2nd year - EC2), Year 5 (Middle Courses 1st year - MC1) and Year 6 (Middle Courses 2nd year - MC2). Then, it takes seven years to complete Secondary school. Indeed, Secondary school in Benin, is made up of two levels: the first known as General Junior Secondary and the second is divided in General Senior Secondary and TVET. The General Junior Secondary grades are: 6th grade, 7th grade, 8th grade and 9th grade. Thus, at the end of General Junior Secondary school, learners must pass the exam of the GCSE. Afterward, grades of General Senior Secondary or TVET are: 10th grade, 11th grade and 12th. In addition, to completing the 10th grade, learners must enroll in a specific domain of study such as: General Senior Secondary (Literature, MPS or Biology stream for example) and TVET (Mechanical Science and Technology or Electrical Science and Technology stream for example). At the end of the General Senior Secondary or Technical Secondary, learner should take the Senior Secondary Education Certificate, equivalent of the French Baccalaureate. After that, they can enroll in the Tertiary Education in the existing courses and complete the years required to obtain a bachelor degree, a master or a doctorate [17]. Table 1 presents the main subjects taught in three streams, selected for the performed research. Fig. 1 illustrates the Beninese education system.

TABLE I. MAIN SUBJECTS

Stream	Main subjects
Literature stream	French, English, German, Spanish, Philosophy
MPS stream	Math, PCT
Biology stream	Math, PCT, Biology

III. RELATED WORK AND DISCUSSIONS

Hereby, are an overview of existing work and their shortcomings compared to the model proposed in this study.

A. Related Work

In literature, several ML algorithms are used for students academic orientation in high school. Whatever the educational system, at a given moment in his/her academic career, student at junior secondary level is required to make a choice of stream.

Therefore, to better orient learners, many research studies have been undertaken. In [18], authors predicted the performance of students in bachelor's and master's degrees in computer science and telecommunications.

Some models are based on real data. Usually, authors used tools such as a chatbot to collect data [19]. In [20], authors noticed that the increase in data did not significantly improve the obtained results.

Individuals in a database are characterized by a number of variables and all of them are not necessarily relevant for learner orientation. In [21], authors insist on correctly detecting the relevant variables involved in the process and their relationships with each other. In [22], authors used learners' scores to make predictions. In addition to grades, they used the number of absence per subject of the student.

In order to have a maximum prediction accuracy, in [23], authors compared several ML algorithms such as Support Vector Machine (SVM), neural networks, regressions, random forest, k-nearest neighbors, Naive Bayes', decision trees, etc.

Several performance measures exist to determine the degree of reliability of a prediction model. Following the example of [24], in which authors used three performance scores to validate their model (accuracy, Cohen's kappa, and the ROC curve), other authors used only the accuracy to assess the performance of their models. They used data that is not very large in size (a size that varies between 100 and 250 individuals). They obtained an accuracy of 94% with the random forest. In [25], authors were faced with an explosion of data and they obtained the best accuracy of 97% with Bayes' naive.

In addition, the authors of [26] implemented a framework which predicts academic orientation using supervised machine learning. They had a dataset of 350 individuals and compared the performance of decision tree, KNN, SVM and logistic regression by cross-validation and by a split that reserved 30% of the data for testing. Their proposal is mainly based on personality types such as Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. They obtained the best AUC (Area Under Curve ROC) with the decision tree which was 0.8.

B. Discussions

From all above, we can notice three types of categories to classify the models aiming at providing academic or professional guidance for learners. There are the size of the data, the nature of the data and the nature of the prediction.

As far as data size is concerned, we can distinguish models that are based on a huge data size [25] and those based on a relatively small data size [24].

For the nature of data, some models are designed based solely on learners' grades and others are designed based
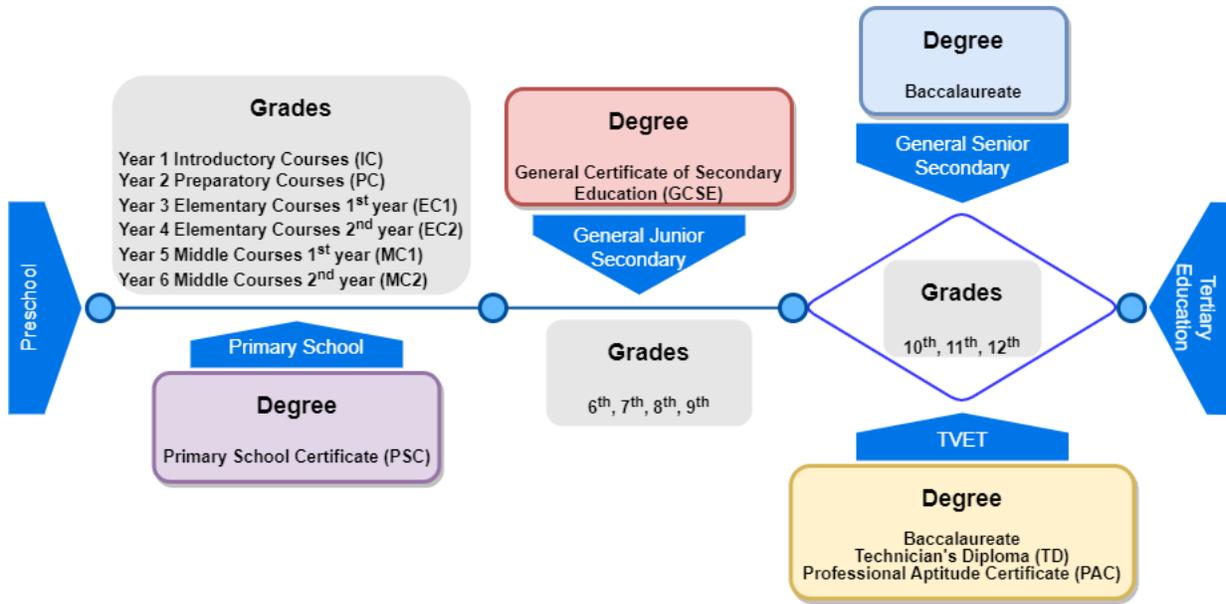
Fig. 1. The Beninese Education System.

not only on grades but also on learners' social environment variables [22].

Finally, different models are not intended to predict the same things. One set is designed to predict school dropout [21], other to predict the stream or the aptitude/ability in a given track and another to predict a score or the average mark [23].

Based on related work study, we can conclude that for academic orientation prediction, Bayes naive is the most suitable, when dealing with massive data, while random forest is the most suitable when data size is relatively small. It should also be noted that the combination of random forest and regression does not really provide good prediction accuracy.

Unlike previous works, current study focuses at first stage on learners' grades for guidance. Our educational context is different from those found in the literature and not all subjects have an impact on the learners' streams choice. Since a database of learners' digitized grades is not available, we started by collecting these grades using the students' transcripts of records. We trained the models on several ML algorithms with the scope of having maximum accuracy.

IV. OUR PREDICTION MODEL

To achieve academic orientation prediction for Benin's high schools, we propose the architecture illustrated in Fig. 2. Our architecture includes several stages: preprocessing, model creation, model evaluation, and model optimization. However, the first three are compulsory for any prediction model and are described in this section.

A. The Dataset

A learner in the MPS stream, must have a basic knowledge of mathematics and PCT, and a learner in the Biology stream must have a basic knowledge of mathematics and PCT, as well as Biology. We can conclude that these subjects make it possible to distinguish scientific learners from literary learners. Moreover, considering subjects such as French, English, Philosophy, History and Geography, would not be optimal because they are cross-cutting subjects. One can be in Literature, MPS or Biology stream and be excellent in these subjects. In Benin, evaluations are done in secondary schools on a semester or quarterly basis. A semester or quarterly average in a given subject does not reflect the actual performance of learners in that subject over the course of a year. For this purpose, we used available transcripts of records to calculate annual averages in mathematics, PCT and Biology.

The dataset used in this study contains 325 learners' instances with 13 variables. Table II shows a description of all variables of the dataset.

TABLE II. DESCRIPTION OF VARIABLES

Symbol	Meaning
Mm6	Annual average mark of the 6th grade in Mathematics
Mp6	Annual average mark of the 6th grade in PCT
Ms6	Annual average mark of the 6th grade in Biology
Mm5	Annual average mark of the 7th grade in Mathematics
Mp5	Annual average mark of the 7th grade in PCT
Ms5	Annual average mark of the 7th grade in Biology
Mm4	Annual average mark of the 8th grade in Mathematics
Mp4	Annual average mark of the 8th grade in PCT
Ms4	Annual average mark of the 8th grade in Biology
Mm3	Annual average mark of the 9th grade in Mathematics
Mp3	Annual average mark of the 9th grade in PCT
Ms3	Annual average mark of the 9th grade in Biology
S	The stream of the learners

B. Preprocessing

Since the data collection stage has been performed manually, the probability to push up some shortcomings is high.
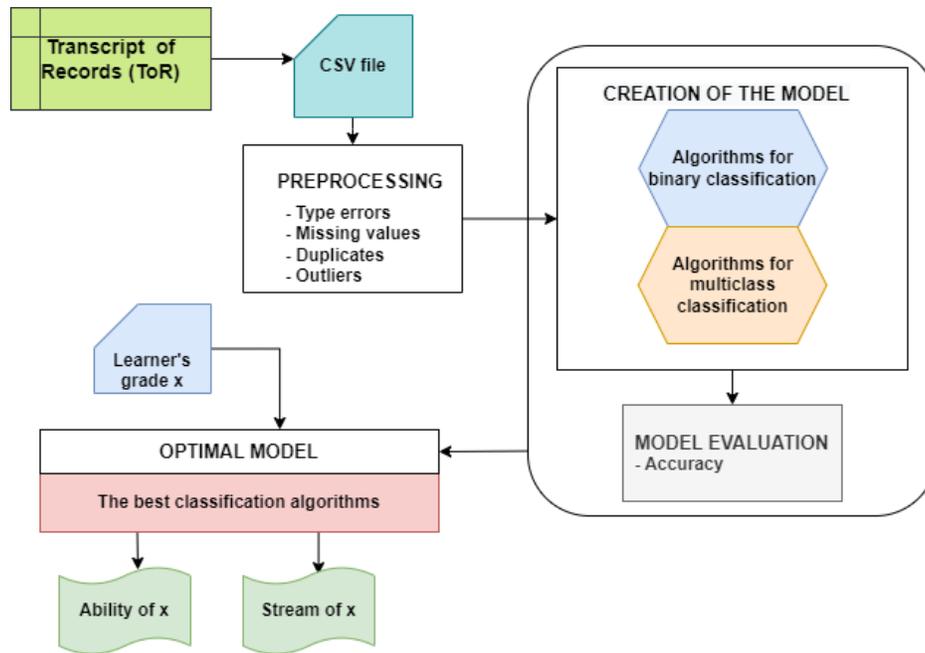
Fig. 2. Machine Learning-Based Architecture for Academic Orientation.

The most common are:

- Outliers Values
  Generally, the average mark is between 0 and 20. A typo error can lead to enter a value outside this range.

- Duplicates Values
  Two learners may have the same average marks in all considered subjects. Also, the same average marks can be entered twice by mistake.

- Missing Values
  Students may not have average marks in some subjects, for a variety of acceptable and unacceptable reasons.

The preprocessing stage allows us to clean the collected data (the average marks) to deal with missing, duplicates, and outliers values [27]. Then, the exploration of the average marks was done and allowed to notice that all the average marks of the learners will not be relevant for the prediction of their orientation. Indeed, we want to predict the optimal study path for this mass of general education learners who rush to the Biology stream because of the several opportunities it offers. Among this batch of learners who enroll in the Biology stream, some are more likely to take literature and others to choose TVET.

Taking the MPS or Biology stream, requires for learners to have good skills in Mathematics, Physics, Chemistry and Technology (PCT) and Biology. Therefore, we collected learners' yearly average marks in Mathematics, PCT and Biology from the 6th grade to 10th grade and the yearly average of the 10th grade. Fig. 3 provides an overview of the data.

It can occur that some students choose Biology stream and fail, probably because they underperformed or were not proficient. Our concern here is to orient the learners in the best possible way, that is finding the right stream. Then, we proceeded to label the data using the following assumptions:

- If the annual yearly average as well as the yearly averages in PCT, mathematics and Biology in the 10th grade are greater than or equal to 11, this instance is labeled as Biology stream.

- If an average mark of a given instance of data, in the 10th grade is greater than or equal to 14 and the annual average marks in mathematics and PCT in the 10th are greater than or equal to 15, then, the instance is labeled as Mathematics and Physical Sciences (MPS) stream.

- Otherwise, the instance is labeled as Literature stream.

C. Model Design

During this stage, ML algorithms allowed us to create prediction models. For this purpose, we use five Machine Learning algorithms and design two types of models:

- The first category of model is used to predict a learner's literary or scientific aptitude. It also allows the orientation of some learners towards TVET, since scientific aptitudes are compulsory for some streams such as Mechanical Science and Technology or Electrical Science and Technology for example.

- The second category of model predicts Literature, Biology or MPS stream of the learner.

D. Model Evaluation

In the model evaluation stage, we mainly use accuracy metric.

	Mm6	Mp6	Ms6	Mm5	Mp5	Ms5	Mm4	Mp4	Ms4	Mm3	Mp3	Ms3	MS
count	325.000000	325.000000	325.000000	325.000000	325.000000	325.000000	325.000000	325.000000	325.000000	325.000000	325.000000	325.000000	325.000000
mean	13.763077	14.132308	13.907692	12.680000	13.960000	13.590769	12.021538	12.926154	12.683077	10.316923	10.907692	11.747692	12.006585
std	2.722084	2.190968	2.602961	2.985934	2.622599	2.812325	3.241727	2.876162	2.846027	3.853152	3.434228	3.166818	2.225005
min	6.000000	7.000000	6.000000	5.000000	7.000000	7.000000	4.000000	5.000000	6.000000	2.000000	3.000000	3.000000	5.730000
25%	12.000000	13.000000	12.000000	11.000000	12.000000	12.000000	10.000000	11.000000	11.000000	7.000000	8.000000	9.000000	10.390000
50%	14.000000	14.000000	14.000000	12.000000	14.000000	14.000000	12.000000	13.000000	12.000000	10.000000	11.000000	11.000000	11.710000
75%	16.000000	16.000000	16.000000	15.000000	16.000000	16.000000	14.000000	15.000000	15.000000	13.000000	13.000000	14.000000	13.570000
max	20.000000	19.000000	19.000000	19.000000	20.000000	20.000000	20.000000	20.000000	20.000000	19.000000	19.000000	19.000000	18.270000

Fig. 3. Data Overview.

The accuracy is the metric that is often used to evaluate the performance of a classification model. It is the rate of good prediction. Therefore, the closer the accuracy is to 1, the better the model is performing.

V. PERFORMANCE EVALUATION

This section focuses on presenting the outcomes of the study in terms of model performance. Indeed, performance evaluation is performed in two steps: the first one uses 10 folds cross-validation and the second performs a specific split. The implementation code of the proposed models is available online (https://github.com/Jomamer/StudentAcademicPathSelection).

A. Overview of Labels

Before presenting the outcomes of the performed study, we present here the labels. In the dataset, there are 41% (133 instances) literary learners and 59% (192 instances) scientific learners.

Furthermore, there are 41% (133 instances) of Literature learners, 11% (37 instances) of MPS learners and 48% (192 instances) of Biology learners.

B. Cross-Validation Performances

The model is evaluated by performing a cross-validation of 10 folds.

Fig. 4 shows the mean accuracy and the std (standard deviation) of each algorithm for predicting learners' scientific or literary ability. Random Forest has the highest mean accuracy of 0.94 and the third lowest std of 0.08. It is followed by Linear SVC which gets 0.92 as mean accuracy and the second lowest std of 0.07. Logistic Regression gets the lowest mean accuracy of 0.87 and the highest std of 0.13.

Fig. 5 shows the mean accuracy and the std of each algorithm for predicting learner' Literature, MPS or Biology stream. Linear SVC and Random Forest obtain the best mean accuracy which is 0.90. They have the second and third lowest std of 0.08 and 0.1 respectively. Logistic Regression gets the lowest mean accuracy of 0.71 and the highest std of 0.15.

C. A Specific Split Performances

At this stage, we have booked for the prediction of each element, at least 20% of the dataset for validation.

In order to predict learners' scientific or literary ability, the size used for the test set by this split is 71 (more than 21% of the dataset). Table III presents the performances achieved. Random Forest has the best performance, followed by Linear SVC. KNN and Logistic Regression obtain the lowest performances.

For predicting Literature, MPS or Biology stream of the learners, the size dedicated to the test set, by this split is 66 (more than 20% of the dataset). Table IV presents the performances obtained. Linear SVC has the best accuracy.

TABLE III. A SPECIFIC SPLIT PERFORMANCES FOR ABILITY PREDICTION

Algorithms	Accuracy	Recall	f1_score
Decision Tree	0.97	0.96	0.98
Random Forest	0.99	0.98	0.99
Linear SVC	0.97	0.98	0.98
KNN	0.92	0.87	0.93
Logistic Regression	0.9	0.96	0.92

TABLE IV. A SPECIFIC SPLIT PERFORMANCES FOR STREAM PREDICTION

Algorithms	Accuracy
Decision Tree	0.86
Random Forest	0.92
Linear SVC	0.95
KNN	0.94
Logistic Regression	0.82
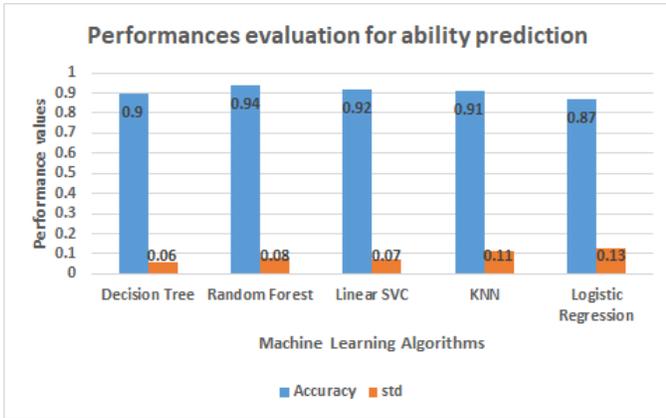
Fig. 4. Performances for Cross-Validation based Ability Prediction.

Fig. 6. Confusion Matrix for Ability Prediction.
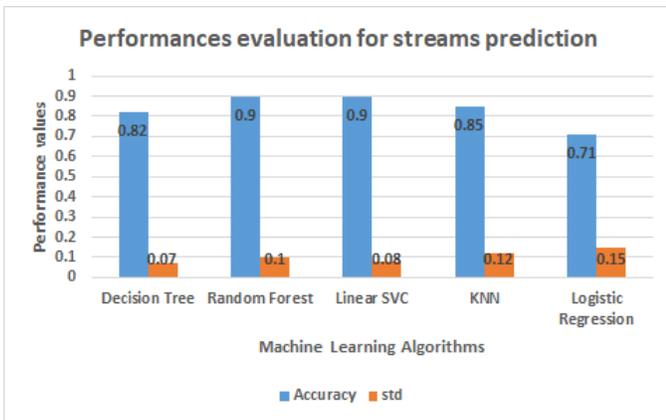
Fig. 5. Performances for Cross-Validation based Streams Prediction.
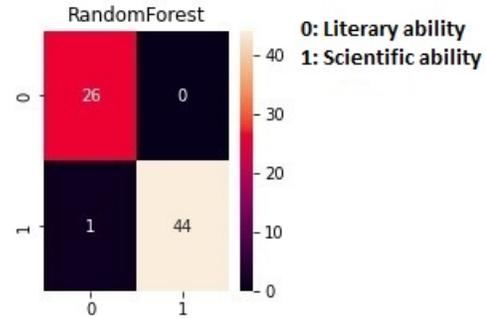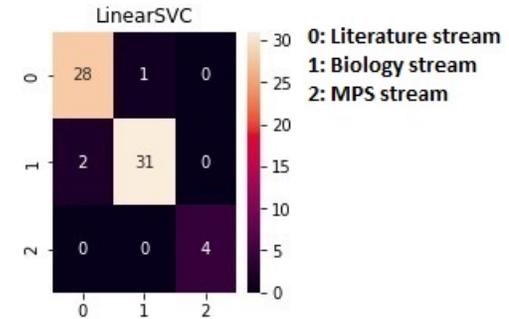
Fig. 7. Confusion Matrix for Stream Prediction.

Fig. 6 and Fig. 7 represent respectively the best confusion matrices for ability prediction and stream prediction. Indeed, Random Forest presents the best confusion matrix for ability prediction. However, it raised one error. In fact, the algorithm predicts a scientific ability for the learner instead of the literary one.

On the other hand, Linear SVC gives the best confusion matrix for Literature, MPS or Biology stream prediction. It correctly predicts the learners in MPS stream. However, it made an error, which is to predict two learners for Biology stream, whereas they are actually in Literature stream. Another prediction error, less serious than the previous one, is the classification of a learner in Literature study rather than Biology stream. This error seems to be less severe because, usually a learner who is able to attend Biology stream can also attend Literature stream.

D. Optimization of the Model

The tuning of algorithm parameters allowed us to reach the optimal model. To achieve it, we used the Gridsearch method to make cross-validation of K parameter of the KNN, the max-depth for the trees decision and the n-estimators for the random forest on the training set. A higher number of trees gives better performance but slows down the code.

In order to get better results, we tried to monitor the importance of the characteristics with Random Forest for the learners' scientific or literary ability, and the learners' stream. Taking into account the most important features, could improve the results. Fig. 8 gives an overview of the importance of the characteristics with Random Forest for the scientific or literary aptitude of the learners as well as the learners' stream.
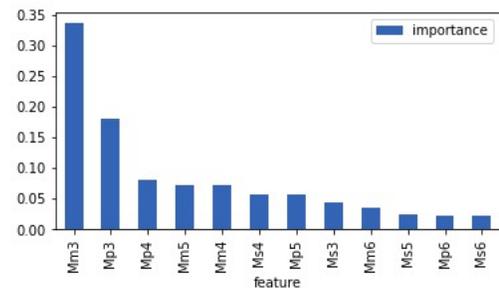
Fig. 8. Importance of Characteristics for Learner Ability and Stream with Random Forest.

Overall, the yearly average mark of mathematics and PCT of learner in 9th grade are very important. Average mark of 6th grade are less important. We tried to improve the performance obtained by considering only the characteristics that were having at least a given importance but it was not the case. We can conclude that average mark of the 9th grade are certainly more important but they are not enough to guide the learners.

The optimal model is the one that offers the best performance. Two models are chosen at the end of the evaluation stage:

- a model to predict the scientific or literary ability of learner.

- a model to predict the Literature, MPS or Biology stream of learner.

VI. Discussion

This discussion is conducted along two axes: starting with the prediction of the learners' scientific or literary ability, then approaching the prediction of the Literature, MPS or Biology stream of the learners.

A. Case 1: Predicting Learners' Scientific or Literary Ability

Random Forest has the best mean accuracy of 0.94 and the third lowest std of 0.08. Linear SVC is not far behind with 0.92 for mean accuracy and 0.07 for std. With the specific split, which reserves more than 21% of the dataset for testing, Random Forest presents the best performances, which are 0.99 for accuracy, 0.99 for f1-score and 0.98 for recall. With the confusion matrix obtained, we noticed that, it made only one error. It predicted that a learner is in a scientific stream when in fact he is in literary stream.

From the above, we can deduce that for the prediction of scientific or literary aptitude of learners, Random Forest is the most suitable. Linear SVC can also be used as its performance is very close to the Random Forest one. It even has a lower std than Random Forest in the cross-validation comparison.

B. Case 2: Predicting Learners' Literature, MPS or Biology Stream

Linear SVC and Random Forest had the best mean accuracy of 0.9 and the second and third lowest std of 0.08 and 0.1 respectively. With the specific split, reserving more than 20% of the dataset for testing, Linear SVC has the best accuracy of 0.95. With the confusion matrix obtained, we noticed that it correctly predicts the learners of the MPS stream. However, it makes one error, in the sense that it predicts two learners from the Biology stream, when they are actually from Literature stream. There is also another prediction error that is less serious than the previous one. It predicts that one learner is from Literature stream when he is actually from Biology stream.

After all, we can conclude that for predicting Literature, MPS or Biology stream of the learners, Linear SVC is the most suitable. Random Forest can also be used as its performance is very close to that of Linear SVC. It even obtains the same average accuracy as Linear SVC in the cross-validation comparison. It is true that KNN has the second best accuracy in the specific split comparison, but we do not focus on it, because it has the second lowest performance in the cross-validation comparison.

In general, performance obtained when predicting learners' scientific or literary ability is better than those obtained when predicting learners' streams. This may be related to the fact that, usually, a learner who is able to attend MPS stream can also attend Biology stream. So, if the proposed model predicts the Biology stream for a learner who is actually enrolled in MPS stream, this is not an error.

As a matter of fact, it should be recognized that a learner's average mark does not only depend on the previous performance of the learners. It could rely on many factors such as: i) *How does the teacher explain the lessons?* ii) *Are the classmates much motivated to outperform?* iii) *Did the learner change school?* iv) *Is the teacher proficient with the subject?*

VII. Application of the Optimal Model

A Web page in https://jomamer-orientation-streamlit-app-rl3b0z.streamlitapp.com/ gives an overview of the framework that we designed to allow school's authorities to use the model. As a matter of fact, to make the proposed model usable, we designed a simple ML web application with Python and Streamlit. In particular, Streamlit is a Python package compatible with most of the Python libraries that are used for ML (scikit learn, keras, seaborn, matplotlib, numpy, pandas, tensorflow, etc.). It is an open source application framework in Python language that allows to create web application for data science and machine learning in a short time.

Notebook Jupyter allowed us to pre-test a number of things in order to deduce optimal models for predicting scientific or literary ability of the learners and Literature, MPS or Biology stream of the learners. After importing Streamlit, the useful libraries have been called. Then, we imported the entire dataset and trained the identified optimal models. As a matter of fact, the split allows us to reserve a part of the dataset to evaluate the model performance. In production, it was no longer a question of testing the model, but rather of using it. Therefore, we have used the entire dataset so that the model learns more cases.

Finally, the obtained results are quite satisfactory. Indeed, the strength of the proposed model relies on two predicted indicators which are complementary and play absolutely, a major role in the effectiveness of learner's orientation.

VIII. Contributions and limitations

In this work, we developed a model for predicting academic orientations using ML techniques. To achieve this, we predicted the learners' scientific or literary ability and Literature, MPS or Biology stream. We used mathematics, PCT, and Biology average marks from the 6th grade to the 9th grade and the size of the dataset is 325. Our results are quite satisfactory for both models. The two models, together, allow achieving an optimal orientation for the learners.

However, it is worth to emphasize that in the paper [26] published in June 2022, authors went in the same direction. But, the major difference between both proposed approaches lies in the fact that they based themselves only on personality variables whereas the study performed in this paper is based on learner' average marks. Overall, the obtained performance results are better. It should be noted that both approaches may be complementary and future studies could look at how to combine them for optimal learner guidance about academic path selection.

On the other hand, with respect to the current study, authors of [24] used the same algorithms. Random forest has given an

accuracy of 94%, Decision tree 93% and Logistic regression 85%. They had bootstrapped a dataset of 101 records. One can notice that, the performances we obtained are better. Moreover, our data are real and in addition, we have also used the cross-validation which is recommended when dealing with a small size of data. It should be noted that authors of [24] have also taken into account characteristics such as age, gender and geographical area.

As a limitation, the current size of the dataset, can be underlined. Moreover, academic performance does not depend solely on the average marks obtained by learners. Several variables such as social environment factors could be taken into account.

IX. Conclusions

Everyone, no matter the position, tries to find a way to improve student's academic performances. In this work, we brought our support with a model, using Machine Learning which is very useful in almost every field nowadays. We compared the performance of five ML algorithms using a cross-validation and a specific split to predict learner's scientific or literary ability and the Literature, MPS or Biology stream of the learners. For this purpose, the Mathematics, PCT and Biology average marks from 6th grade to 9th grade of 325 instances of data are used. The best performance is pointed out by Random Forest algorithm with 99% of accuracy. Most probably, a larger data size could allow improving performance results even for the other algorithms.

The main limitation of this study resides on only considering the average marks obtained by the learners to perform the prediction. Thus, other parameters such as social environment factors may improve the performance results. The dataset size may also be increased. It would be interesting to look in these directions for future studies.

Acknowledgment

References

[1] C. Bellei, X. Vanni, J. P. Valenzuela, and D. Contreras, "School improvement trajectories: an empirical typology," *School Effectiveness and School Improvement*, vol. 27, no. 3, pp. 275–292, 2016.

[2] S. Dhawan, "Online learning: A panacea in the time of covid-19 crisis," *Journal of educational technology systems*, vol. 49, no. 1, pp. 5–22, 2020.

[3] C. E. Garcia and C. W. Yao, "The role of an online first-year seminar in higher education doctoral students' scholarly development," *The Internet and Higher Education*, vol. 42, pp. 44–52, 2019.

[4] A. Dirin, M. Nieminen, and A. Alamäki, "Social media and social bonding in students' decision-making regarding their study path," *International Journal of Information and Communication Technology Education (IJICTE)*, vol. 17, no. 1, pp. 88–104, 2021.

[5] D. Park and Y. Kim, "A study on the effects of paramedic students' major selection motivation and occupational values on employment preparation behavior," *Journal of Digital Convergence*, vol. 18, no. 8, pp. 263–270, 2020.

[6] P. Kaur and R. K. Singh, "Feature selection pipeline based on hybrid optimization approach with aggregated medical data," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022.

[7] I. Paryudi, A. Ashari, and K. Mustofa, "The performance of personality-based recommender system for fashion with demographic data-based personality prediction," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022.

[8] B. Meriem, H. Benlahmar, M. A. Naji, E. Sanaa, and K. Wijdane, "Determine the level of concentration of students in real time from their facial expressions," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022.

[9] A. S. A. Osman, "Assessing the quality of educational websites in sudan using quality model criteria through an electronic tool," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022.

[10] A. Abdelhadi, S. Zainudin, and N. S. Sani, "A regression model to predict key performance indicators in higher education enrollments," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022.

[11] A. A. Saa, "Educational data mining & students' performance prediction," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, 2016.

[12] J. Calderon-Valenzuela, K. Payihuanca-Mamani, and N. Bedregal-Alpaca, "Educational data mining to identify the patterns of use made by the university professors of the moodle platform," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022.

[13] W. Al Madhoun, "Predictive modelling of student academic performance–the case of higher education in middle east," Ph.D. dissertation, University of East London, 2020.

[14] R. S. Baker, T. Martin, and L. M. Rossi, "Educational data mining and learning analytics," *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications*, pp. 379–396, 2016.

[15] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, p. e1355, 2020.

[16] J. Alladatin, J. Bernachez, D. Bergeron *et al.*, "Overview of primary school principals' educational level and training in benin: The challenges related to the expected competencies and skills," *Annals of the University of Craiova, Series Psychology, Pedagogy*, vol. 43, no. 2, pp. 145–162, 2021.

[17] A. Assogbadjo, R. Idohou, and B. Sinsin, "Review of the higher education system in benin: Status, challenges, opportunities and strategies for improvement," *African Journal of Rural Development (AFJRD)*, vol. 1, no. 1978-2017-2051, pp. 139–149, 2016.

[18] H. Hamsa, S. Indiradevi, and J. J. Kizhakkethottam, "Student academic performance prediction model using decision tree and fuzzy genetic algorithm," *Procedia Technology*, vol. 25, pp. 326–332, 2016.

[19] O. Zahour, A. Eddaoui, H. Ouchra, O. Hourrane *et al.*, "A system for educational and vocational guidance in morocco: Chatbot e-orientation," *Procedia Computer Science*, vol. 175, pp. 554–559, 2020.

[20] R. Bertolini, S. J. Finch, and R. H. Nehm, "Testing the impact of novel assessment sources and machine learning methods on predictive outcome modeling in undergraduate biology," *Journal of Science Education and Technology*, vol. 30, no. 2, pp. 193–209, 2021.

[21] M. d. C. Nicoletti and O. L. de Oliveira, "A machine learning-based computational system proposal aiming at higher education dropout prediction." *Higher Education Studies*, vol. 10, no. 4, pp. 12–24, 2020.

[22] F. Ouatik, M. Erritali, F. Ouatik, and M. Jourhmane, "Students' orientation using machine learning and big data," 2021.

[23] A. Tarik, H. Aissa, and F. Yousef, "Artificial intelligence and machine learning to predict student performance during the covid-19," *Procedia Computer Science*, vol. 184, pp. 835–840, 2021.

[24] A. Dirin and C. A. Saballe, "Machine learning models to predict students' study path selection," *iJIM*, vol. 16, no. 01, p. 159, 2022.

[25] F. Ouatik, M. Erritali, and M. Jourhmane, "Student orientation using machine learning under mapreduce with hadoop." *J. Ubiquitous Syst. Pervasive Networks*, vol. 13, no. 1, pp. 21–26, 2020.

[26] H. El Mrabet and A. Ait Moussa, "A framework for predicting academic orientation using supervised machine learning," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–11, 2022.

[27] B. Malley, D. Ramazzotti, and J. T.-y. Wu, "Data pre-processing," *Secondary analysis of electronic health records*, pp. 115–141, 2016.

Towards an YouTube Verified Content System based on Blockchain Approach

Phuc Nguyen Trong, Hong Khanh Vo, Luong Hoang Huong, Khiem Huynh Gia,
Khoa Tran Dang, Hieu Le Van, Nghia Huynh Huu, Tran Nguyen Huyen, The Anh Nguyen,
Loc Van Cao Phu, Duy Nguyen Truong Quoc, Bang Le Khanh, Kiet Le Tuan
FPT University, Can Tho City, Viet Nam

Abstract— **YouTube connects people with each other through an online video sharing service platform. With the great development of the entertainment industry, content on YouTube is accessible to many people of different ages. However, verifying the content posted on YouTube is clean or not is a difficult problem. Dirty content is violent, pornographic and vulgar content that causes serious psychological harm to the segment of users under the age of 18, i.e., especially those of an age who are not yet aware of the harmful effects of content. Toxic will bring to the child's behavior. Agree that Google (i.e., YouTube) has developed a YouTube Kid application where the videos are only for children under the age of 13. However, cultural and educational differences between regions strongly influence the choice of children. Select content for children. Therefore, the content restrictions on the YouTube Kid application have not yet met all the requirements of parents around the world. There have been many development directions to identify videos containing malicious content based on deep learning. However, there is no method to build a tool to support parents of children to share and identify videos with objectionable content (e.g., violence, pornography, obscene words) on the YouTube platform. In this research paper, we introduce YVC, a YouTube-verified content platform by applying blockchain's distributed, public validation. This tool helps parents validate YouTube content and issue a report to reduce dirty content on YouTube. To demonstrate the effectiveness of our approach, we implement the proof-of-concept in the three most popular EVM platforms: Ethereum, Fantom, and the Binance smart chain. Compared to the YouTube Kids (i.e., the most common shared video platform for the under 13-year-old kid), our approach is able to capture the video preferences of the parents covering the difference areas/countries.**

Keywords—Blockchain technology; public authentication; Ethereum; Fantom; Binance smart chain platform; social media platform

I. Introduction

With the development of technology, many social media platforms are born and become indispensable parts of people's lives. Among them, the YouTube video-sharing platform is a prominent online presence that provides various types of content for today's society. With a large and wide user base, YouTube has opened a monetization policy for creators, and this policy becomes a potential monetization target not only for large companies but also for individual creators [1].

Dirty content appears to creep inside the content on Youtube in an uncontrollable way. As Hank Green notes, YouTube is "a bridge between creators who are making stuff and advertisers who want to make money. YouTube makes its money through these people but also has to keep both of these

groups happy" [2]. Creators all want their video products to attract many people, the ultimate purpose is to develop media channels and earn profits. To attract more people, dirty content with erotic and violent nature gradually crept into the daily entertainment content.

Other social platforms (e.g., Reddit, Twitter) have done a lot of research regarding the issue of detecting harmful content (i.e., violence, calls to violence, pornography). For example, Tseng et al. [3] and Bellini et al. [4] highlight how forums like Reddit serve as platforms for discussing methods of intimate partner violence. Similarly, abuse, cyberbullying, and online harassment on online platforms such as Twitter was also studied by [5], [6]. For work related to malicious behavior detection on the YouTube platform, Chu et al. [1] exploits the differences used to monetize illegal YouTube content that could potentially harm viewers and other users. Thereby, the authors propose methods to prevent abuse of this service (similar findings are presented in the Related Work section of the article).

YouTube has taken a remedy by censoring and removing violent and pornographic content on the platform. In addition, the age restriction mode for adults and minors is set on the admin rights of the creator. Moreover, YouTube launched [7], a version made entirely for children to prevent children from accidentally being exposed to dirty content that causes psychological effects. However, the measures taken by Youtube have not eliminated the dirty content that is being uploaded every day. A weakness of the age limit when watching videos lies in the admin rights of the creator. If the creator intentionally does not limit the age before giving the product to the user segment older than 18, the content will still be published. open, anyone can enter. Recognizing the advantages and risks of harmful content for children's education and entertainment, Google founded YouTube Kids to serve children under 13 years old. Based on the regulations for child-friendly programs, a number of violations were identified and restricted to television programs. When a child surfs the Internet, the same rules can be automatically detected and filtered. However, content filtering on YouTube for Kids currently relies on metadata attributes, where inappropriate content can pass the [8] filter.

Blockchain technology is known for its outstanding features of transparency and immutable content. Picha Edwardsson et al. studied the possibility of using blockchain technology to create a secure, community-facing information verification database with the goal of creating a solution that could improve the reliability of verifying information and monitoring each authenticity verification process for digital content, including

images and videos. The paper indicates that blockchain is not yet ready to be directly applied to fact-checking processes in a real-world scenario. The study also shows that the application of blockchain to verify a scenario is entirely possible and highly reliable and transparent [9]. Several approaches address these problems by applying Blockchain techniques in the other environment (e.g., cash-on-delivery [10], [11], [12], healthcare [13], [14], [15], supply chain [16], [17], [18], and others [19], [20], [21]).

To solve the problem of verifying dirty content through social media platforms (specifically in the research paper, YouTube). We introduce YouTube verify content system based on blockchain (YVC). YVC's purpose is to verify dirty content on youtube platform and store it on blockchain platform. The purpose of storing data on the blockchain platform is data transparency and community application. With the YVC system, we build three main user groups: Reporter, Verifier and Middleman. YVC system consists of 6 main steps from Reporter report content then to Verifier verifying dirty content and Middleman is responsible for forwarding verified data back to social media platform. Besides, to define the logical constraints in the smart contract to maintain the stable operation of the system, we also design the authorization service for the stakeholders. To design logical constraints on smart contracts, we additionally exploit Solidity language[1]. To evaluate YVC, we implemented a test model on all three of the most popular platforms that currently support EVM, including Ethereum[2], BNB Smart Chain[3] and Fantom[4].

The research problem of this article is to define the YouTube-communication channel based on blockchain and smart contract technologies, which support the parents can detect and share the violation content of the videos before reporting those videos to the Youtube. This approach also define the protocol to verify the voting from the check based on crowdsourcing approach. Therefore, the contribution of this paper is threefold: i) designing the YVC (i.e., Youtube verify content) model to support the user detect the unusual content; ii) implementing the proof-of-concept based on the blockchain and smart contract technologies; and iii) deploying the proof-of-concept on the three most common EVM-supported blockchain platform (i.e., ETH, BNB, Fantom) to select the most suitable platform.

Following this introduction, a state-of-the-art is presented to help understand the limitations and challenges of the current approaches. Then we define the architecture of the blockchain-based YVC system, and we also verify the structure of the smart contract and the database. In the next section, we describe the implementation process including the smart contract, data structure, YVC execution algorithm, and authorization. Section 6 focuses on the evaluation process based on deploying YVC in the three platforms. Finally, suggestions for future research and conclusion are made in the last section.

---

[1]Solidity is an object-oriented, high-level language for executing smart contracts. Smart contracts are programs that govern the behavior of accounts in the Ethereum state https://docs.soliditylang.org/en/v0.8.7/

[2]Ethereum https://ethereum.org/en/whitepaper/

[3]Binance Smart Chain https://github.com/bnb-chain/whitepaper/blob/master/WHITEPAPER.md

[4]Fantom https://fantom.foundation/research/wp_fantom_v1.6.pdf

II. RELATED WORK

Detect illegal activity on YouTube. The YouTube ecosystem has been studied to identify harmful content. Recent works analyzed the metadata and sharing features (i.e. keywords, hashtags) on spam YouTube videos [22], [23]. Another line of work focuses on the automatic classification of harmful activity on YouTube. Fueled by Elsagate, a controversy in which YouTube videos classified as children's content contain inappropriate themes (e.g. Elsa in the Disney movie Frozen performs the action. pique). Papadamou et al. [24] have developed a binary classifier to detect potentially annoying/harmful YouTube videos for toddlers. Similarly, several efforts have identified spam and clickbait videos by studying video metadata, comments, user activity, and video attributes [25], [26], [27]. On the contrary, we provide an overview of malicious content based on user reviews through a user-review sharing system based on blockchain technology. Our system aims to detect malicious behavior based on two aspects (i.e., content sharer and content identifier).

Based on the benefits of the blockchain-based system, more and more information verification systems are built on a blockchain platform for the purpose of transparent verification and community application. As a prime example of the application of blockchain technology in content verification, Ghimire et al propose a new method of video integrity verification method (IVM) using the blockchain technology framework. The study aims to verify data integrity and avoid tampering with video recordings in providing evidence of crime scenes or road accidents. The verification process is done by shredding the videos and comparing them with the hash value of the video previously stored in the blockchain [28].

The integrity verification method (IVM) applied by Ghimire, et al. in the blockchain platform shows that blockchain brings a new transparent and community-oriented approach to content verification in the digital age. In addition, Sathish, et al. developed Aurum [29], a new multimedia streaming platform that hosts user-generated channels with authenticated content through the use of a blockchain-driven creative process. The platform requires public content verification and the goal of community-driven development in which the research paper also mentions that the Aurum system uses effective support protocols to ensure QoS.

Aiming at community authentication and authenticity transparency, Banerjee, Prabal, et al. develop a trusted and fair platform for sharing content without a central moderator. Banerjee, Prabal, et al. have applied blockchain to manage content lists and also use smart contracts to support storage [30]. The system is built on Hyperledger Fabric and the data layer is built on Tahoe-LAFS with the goal of scalability at a lower cost than the Ethereum blockchain platform.

Besides, Zelensky, et al. also have a research paper on analyzing and verifying video data [31]. The problem of data validation to avoid interference and correct the content of violations, banking, remote management, and action confirmation. In which the algorithm will use Swype code using the movement mobile camera to verify the video content. However, the research paper by Zelensky, Aleksandr, et al. still uses a third-party intermediary to verify the content and does not really highlight the characteristics of the blockchain.

Compared with previous approaches, YouTube verified content system (YVC) builds a content authentication system applying blockchain technology on social media platforms (specifically in the research paper, YouTube). In addition, YVC aims at a transparent authentication application for the community, which also reduces the error when third parties approach.

III. YVC Architecture

YouTube verifies content based on the blockchain architecture depicted in Fig. 1. YVC consists of three main elements Reporter, Verifier, and Middleman.

- *Reporter*: The person who reports dirty content to YVC. In the YVC system, the reporter acts as an information provider, the reporter is anonymous on the system and does not need to pay gas fees when providing information.

- *Verifier*: The person who verifies the content on the YVC system. In the YVC system, the verifier acts as a content verifier, and the address information of the verifier will be public on the blockchain. However, personal information will be kept anonymous. Besides, Verifier will bear a small gas fee for verification work.

- *Middleman*: An intermediary who forwards verified and publicized reports to social media platforms (specifically in our research, Youtube).

Platforms applied in the YVC system include Website, Smart Contract, and social media platform.

- *Website*: This is the place where the information is reported and waiting for verification. In the YVC system, the website is the place where the verifier interacts with the smart contract to verify the reported content.

- *Smart Contract*: This is where the verification information of the verifier is stored. In this study, we implement smart contract deployment on the Ethereum blockchain platform and Binance smart chain platform.

- *Social Media Platforms*: These are the social networking platforms where content is stored. In the YVC system, we focus on the Youtube social networking platform.

According to Fig. 1, YVC has six steps. . In the first step, the Reporter is a viewer of the content on social media platforms (Specifically in the research paper, YouTube), if it feels that the content contains vulnerable content, the Reporter will report that content through the website of the YVC system. In the second step, the report information is displayed on the web, now everyone when accessing the website can see the reported content. In the third step, Verifier will verify content to see if the reported content contains objectionable content or not. When the verification is successful, the information about the verified content will be published on the blockchain (specifically in the research we will publish on the Ethereum blockchain platform and Binance smart chain platform). One thing to note is that, in the third step, Verifier will incur a

fee to publish verification information on the blockchain, this fee is called a gas fee. In the fourth step, when the content is successfully verified, the content is also pushed to the database of the YVC system, about the storage structure we will go into details in the following section. In the fifth step, Middleman is the database administrator, Middleman has the role of getting data from the database and reporting back to social media platforms (specifically YouTube). Note that, because the information is publicly published on the blockchain, if Middleman changes the information in the database, the real information is still public and traceable. From there, it can be seen that the application of blockchain helps to reduce the intervention of third parties. In the sixth step, after successfully retrieving data from the database, Middleman reports that data to the social media platform where the reported content originated. In this step, the social media platform based on data reported by Middleman and verified on blockchain to incinerate vulnerable content on their platform. Note that, since the information on the blockchain is public and immutable, social media platforms' reporting disregard for content is minimized.

IV. Implementation

A. Smart Contract Structure

- The address (on blockchain) of verifier (`verifier`).

- The name of channel which includes on the report content (`channelName`).

- The name of video which includes on the report content (`videoName`).

- The link of video on social media flatform which includes on the report content (`linkOfVideo`).

- Duration of the start of the video containing vulnerable content (`startVulnerableDuration`).

- Duration of the end of the video containing vulnerable content (`endVulnerableDuration`).

- The verified information will be stored in the verify content with the declared struct type (`verifyContent`). The information inside the struct will be determined by two data types: boolean for true/false verification and string for storing results.

All information in the Verify field will be stored inside the smart contract (see Fig. 2). This storage will contribute to transparency between Reporter, Verifier, and Middleman. If any changes are made, the changed data will be made public on the blockchain network.

To determine whether the verified information is true or false. Verify the information is a set of variables stored inside (`verifyContent`) with data type struct. The information to verify includes (`channelName`) verified true or false by (`isChannelName`), similar to (`videoName`) will be verified by (`isVideoName`), with (`linkOfVideo`) will be verified by (`isLinkOfVideo`), with (`startVulnerableDuration`) will be verified by (`isStartVulnerableDuration`), with (`endVulnerableDuration`) will be verified verified by (`isEndVulnerableDuration`), and finally verifyResult
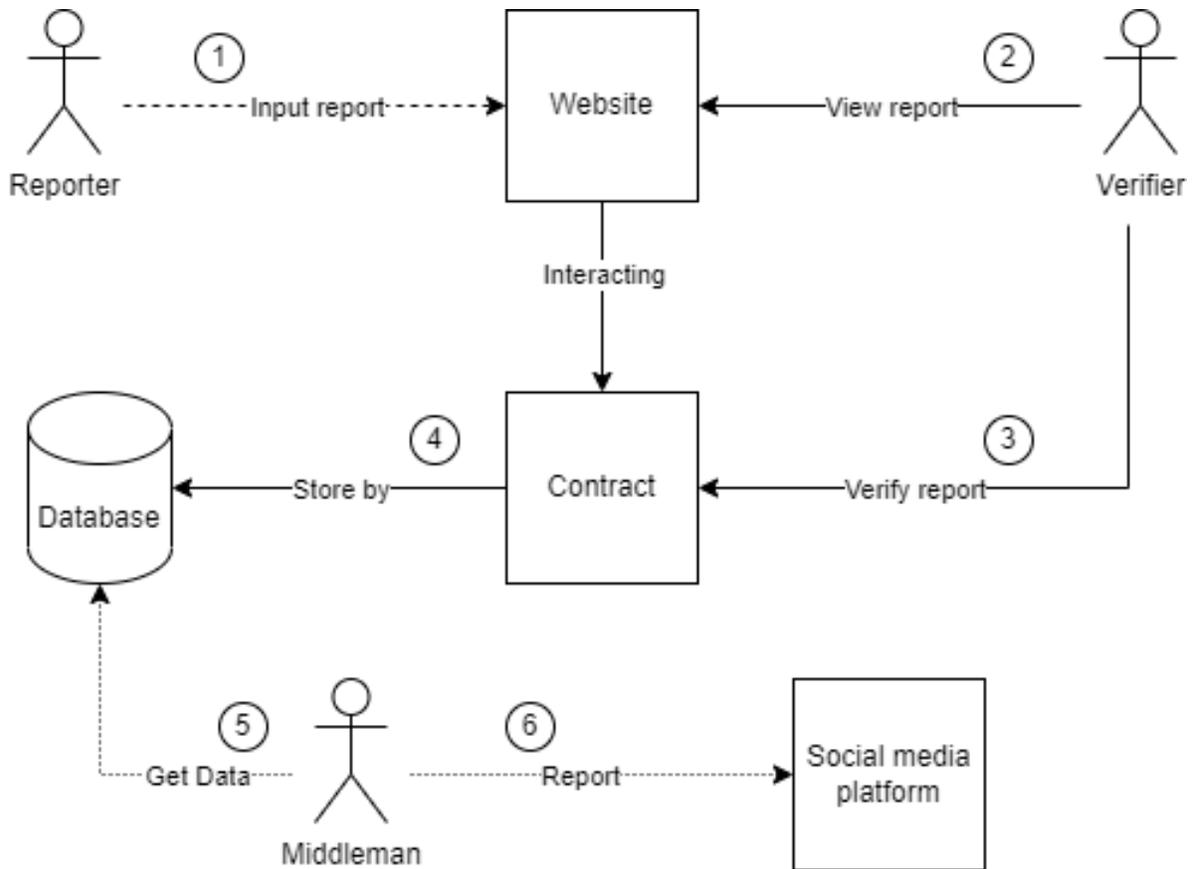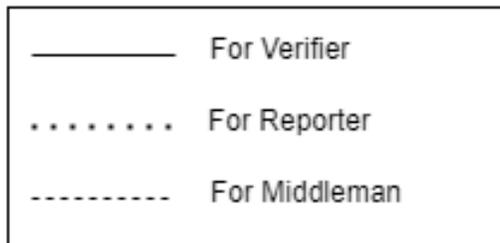
Fig. 1. Youtube Verify Content (YVC) base on Blockchain Architecture.

will save the verification result of the (`verifier`), (`verifyResult`) will be automatically programmed to display the result based on the verification result of the Verifier.

The final result received after verifying content is in the (`verifyResult`) variable. The result will be a set of true or false evaluations of the reported information and a string that evaluates the result (`verifyResult`) based on the evaluation of the (`verifier`).

B. Database Structure

Fig. 3 describes the structure of the database. The list below details the component in the structure.

- The name of channel which includes on the report content (`channelName`).

- The name of video which includes on the report content (`videoName`).

- The link of video on social media flatform which includes on the report content (`linkOfVideo`).

- Duration of the start of the video containing vulnerable content (`startVulnerableDuration`).

- Duration of the end of the video containing vulnerable content (`endVulnerableDuration`).

- The address of the contract. Is a blockchain platform where verified content is stored.
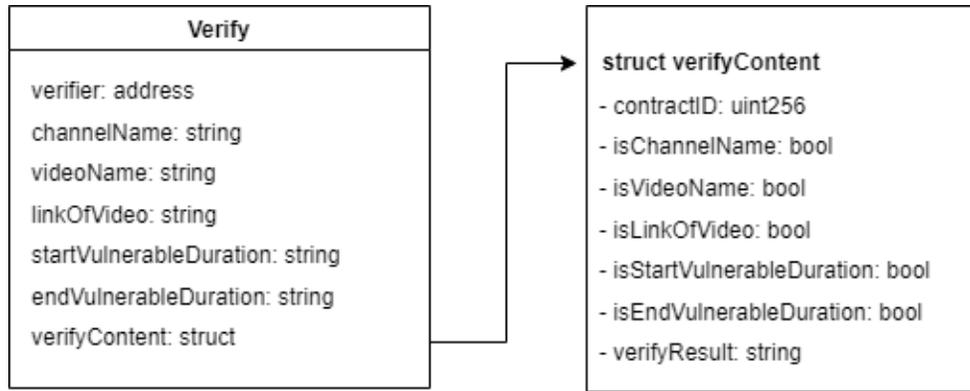
Fig. 2. Smart Contract Structure.
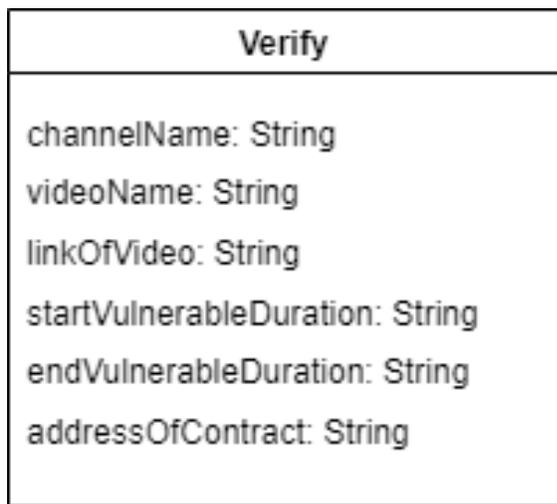
Fig. 3. Database Structure.

(`addressOfContract`).

The database is used to store information about the content present on the blockchain platform, which means that the content has been successfully verified. Information stored inside the database includes (`channelName`), (`videoName`), (`linkOfVideo`), (`startVulnerableDuration`), (`endVulnerableDuration`) and (`addressOfContract`). The storage is to help Middleman manage the database effectively and easily transfer the reported information back to social media platforms (specifically in the research paper, Youtube). Because verified content is publicly available on the blockchain, if there is a change in the database, the information about that verified content will remain unaffected and, more importantly, unchanged. This helps the YVC system to become transparent and limit Middleman's information interference.

The database will not store the evaluation information of the Verifier, but will instead store the address of the evaluated contract on the blockchain. This storage is to streamline the database, and when Middleman reports back to the social media platform, the social media platform side can easily verify the verified contract.

C. Algorithm

Algorithm 1 YVC Execution

1: Input: verifier, channelName, videoName, linkOfVideo, startVulnerableDuration, endVulnerableDuration, verify-Content
2: Output: verifyContent
3: Begin: set isChannelName = false, isVideoName = false, isLinkOfVideo = false, isStartVulnerableDuration = false, isEndVulnerableDuration = false, verifyContent = ""
4: manual update isChannelName
5: manual update isVideoName
6: manual update isLinkOfVideo
7: manual update isStartVulnerableDuration
8: manual update isEndVulnerableDuration
9: **if** isLinkOfVideo == false **then**
10:     update verifyResult = "Due to verifier, link is not correct to verify"
11:     update isStartVulnerableDuration = false
12:     update isEndVulnerableDuration = false
13:     print verifyContent struct
14: **else if** isStartVulnerableDuration == false or isEndVulnerableDuration == false **then**
15:     update verifyResult = "Due to verifier, range of duration vulnerable content is not correct"
16:     print verifyContent struct
17: **else if** isLinkOfVideo == true isStartVulnerableDuration == true isEndVulnerableDuration == true **then**
18:     update verifyResult = "Due to verifier, video contain vulnerable content"
19:     print verifyContent struct
20: **end if**

The algorithm executes sequentially from top to bottom of the YVC framework's execution. First, the system will take input information including the `verifier` as the address when the Verifier conducts content validation. `channel-Name`, `videoName`, `linkOfVideo`, `startVulnerableDuration`, `endVulnerable-Duration` are the information obtained based on the Reporter's report information. `verifyContent` is a variable that stores the Verifier's content verification value. In the next step, the system will automatically set the values to be verified to false and the resulting string is

TABLE I. THE PERMISSION OF THE AUTHORIZED IN YVC

Function/ Method	Verifier	Reporter	Middleman
constructor	Authorized	-	-
setContent	Authorized	-	-
setVerifyContent	Authorized	-	-
getChannelName	Authorized	Authorized	Authorized
getEndVulnerableDuration	Authorized	Authorized	Authorized
getLinkOfVideo	Authorized	Authorized	Authorized
getStartVulnerableDuration	Authorized	Authorized	Authorized
getVerifyContent	Authorized	Authorized	Authorized
getVideoName	Authorized	Authorized	Authorized
verifyContent	Authorized	Authorized	Authorized

empty because Verifier has not started verifying. From lines 4 to 8, Verifier verifies the content to see if the reported content is vulnerable. After Verifier verifies the content, the system will base it on the verified data to give verification results. From lines 9 to 13, if `isLinkOfVideo` is false, the result is returned and stored in the verifyResult variable as "Due to verifier, link is not correct to verify" and the values of `isStartVulnerableDuration` and `isEndVulnerableDuration` are set to false, then the system From lines 14 to 16, if `isStartVulnerableDuration` is false or `isEndVulnerableDuration` is false, the variable that stores the `verifyResult` result will be updated to "Due to verifier, range of duration vulnerable content is" not "correct", then the system prints the `verifyContent` dataset. From lines 17 to 20, if `isLinkOfVideo` is true and `isStartVulnerableDuration` is true and `isEndVulnerableDuration` is true, the variable that stores the `verifyResult` result will be updated as "Due to verifier, the video contains vulnerable content", similarly the system will print the `verifyContent` dataset. and terminate the algorithm.

D. Authorization

The YVC system provides functional authorizations to the stakeholders in the system. The list of functions is described in Table I. The three main user sets inside the system include Verifier, Reporter, and Middleman. When a user joins the system, they will fit into the above user set.

We verify the actor's authority on smart contract by account address (public key) in the Binance smart chain network. The functions and methods in smart contract are described below.

- *constructor*: The function is executed first and executes only once when verifying content on the smart contract. constructor in the YVC system will be used to validate the verifier address when performing a transaction. Therefore, only the verifier can authorize this method

- *setContent*: A function used to confirm the reported content information. This method consists of a set of variables (*channelName, videoName, linkOf- Video, startVulnerableDuration, endVulnerableDuration*) that represent the information of the reported content, respectively. When verifying content, only Verifier can authorize this method.

- *setVerifyContent*: A function used to provide content verification information. This method includes

a set of boundaries (*contractID, isChannelName, isVideoName, isLinkOfVideo, isStartVulnerableDuration, isEndVulnerableDuration, verifyResult*) these variables correspond to the information to be verified. When verifying content, only Verifier can authorize this method.

- *getChannelName*: This is the method used to extract information about the channel name. This information is authorized on all 3 user sets: Verifier, Reporter, and Middleman

- *getEndVulnerableDuration*: A method used to extract information about the end time of the vulnerable video. This information is authorized on all three user sets, Verifier, Reporter, and Middleman.

- *getStartVulnerableDuration*: A method used to extract information about the start time of the vulnerable video. This information is authorized on all three user sets, Verifier, Reporter, and Middleman.

- *getLinkOfVideo*: This is the method used to extract information about the link containing the vulnerable video. This information is authorized on all three user sets, Verifier, Reporter, and Middleman.

- *getVideoName*: This is the method used to extract the information of the video name reported. This information is authorized on all three user sets, Verifier, Reporter, and Middleman.

- *verifyContent*: This is the method used to extract verified content information. This information is authorized on all three user sets, Verifier, Reporter, and Middleman.

V. EVALUATION

A. Environment Setting

We use Solidity language and Remix IDE to program smart contracts because it is pretty popular. When implementing this solution, compiler 0.8.7+commit.e28d00a7 is stable. We also use the default EVM version and don't need optimization. The gas limit is 3000000, just enough to deploy smart contract. We also use the MIT License because we want more people to reuse our solution. Choosing a blockchain network to run smart contracts on is also an issue to be evaluated. We have measured (will be covered in the Experimental section) and decided to use the several platforms, namely, Ethereum, and BNB Smart Chain platforms.

Fantom platform contract deployed: https://testnet.ftmscan.com/address/0x9e307b1de9d4f4d3fcc3682b752f09d135490395;

Ethereum platform contract deployed: https://rinkeby.etherscan.io/address/0x51a75b578a2a64f007a728e50c2fb8b104200078;

BNB smart chain platform contract deployed: https://testnet.bscscan.com/address/0x2ad5ed1473fedf93557621dc420f7d91b39a0626.

TABLE II. THE GAS FOR THE SMART CONTRACT EXECUTION OF THE THREE PLATFORMS, I.E., FANTOM, ETHEREUM, AND BNB CHAIN

Gas for execution	Fantom	Ethereum	BNB Chain
Deploy contract	0.003927833 FTM ($0.0011)	0.00448948 ETH ($6.67)	0.01120938 BNB ($3.12)
call setContent	0.000247230422 FTM ($0.000066)	0.00022971 ETH ($0.34)	0.00773244 BNB ($2.15)
call setVerifyContent	0.000171262102 FTM ($0.000046)	0.00015912 ETH ($0.24)	0.005584608 BNB ($1.55)

B. Results

Reporters have the role of defining and submitting report content to the YVC system. Since this submission has not yet reached the validation stage on the blockchain, the Reporter will not waste gas reporting content. This is similar to Middleman, Middleman acts as the receiver of the authenticated data on the blockchain and transfers that authenticated data to the database which then reports back to the social media platform. Middleman's authentication and data forwarding does not affect smart contracts on the blockchain platform. Therefore, Middleman will not cost gas on the YVC system.

For Verifier, when the verifier verifies content in the YVC system, they will call two methods, *setContent* and *setVerifyContent*. When these two methods are called, the verifier will incur a gas fee for publishing verification on the blockchain. In Table II, we conduct verification to determine the price of each gas fee on blockchains including the Fantom blockchain platform, Ethereum blockchain platform, and Binance smart chain platform.

VI. CONCLUSION

This article introduces a content verification system for social media platforms (namely, YouTube) based on blockchain and a smart contract called YVC (YouTube Verify Content). The main aim of this study is to solve the problem of tightening dirty content that is existing on social media platforms. The highlight of the system is the application of blockchain technology to limit the human impact in changing report information. To evaluate the results of the YVC system, we deployed and analyzed it on three popular platforms, including Fantom, BNB Smart Chain, and Ethereum. Based on the evaluation results, we can confirm that Fantom's Gas is one of the lowest of the three. Regarding possible future directions, we aim to expand to apply the system to verify a lot of dirty content on multiple social media platforms. We will set up verification not only for videos but also for other digital content.

REFERENCES

[1] A. Chu, A. Arunasalam, M. O. Ozmen, and Z. B. Celik, "Behind the tube: Exploitative monetization of content on youtube," in *31st USENIX Security Symposium (USENIX Security 22). Boston, MA: USENIX Association. URL https://www. usenix. org/conference/usenixsecurity22/presentation/chu*, 2022.

[2] H. Green, "Hank green, 35 minutes on youtube demonetization. (2017)," 2017. [Online]. Available: https://www.youtube.com/watch?v=ouMeAaAWUEg

[3] E. Tseng, R. Bellini, N. McDonald, M. Danos, R. Greenstadt, D. McCoy, N. Dell, and T. Ristenpart, "The tools and tactics used in intimate partner surveillance: An analysis of online infidelity forums," in *29th USENIX Security Symposium (USENIX Security 20)*, 2020, pp. 1893–1909.

[4] R. Bellini, E. Tseng, N. McDonald, R. Greenstadt, D. McCoy, T. Ristenpart, and N. Dell, "" so-called privacy breeds evil" narrative justifications for intimate partner surveillance in online forums," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW3, pp. 1–27, 2021.

[5] D. Chatzakou, N. Kourtellis, J. Blackburn, E. De Cristofaro, G. Stringhini, and A. Vakali, "Measuring# gamergate: A tale of hate, sexism, and bullying," in *Proceedings of the 26th international conference on world wide web companion*, 2017, pp. 1285–1290.

[6] Y. Hua, M. Naaman, and T. Ristenpart, "Characterizing twitter users who engage in adversarial interactions against political candidates," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–13.

[7] Youtube, "Youtube kid," online; accessed 29 October 2022. [Online]. Available: https://www.youtube.com/intl/ALL_ie/kids/

[8] S. Alghowinem, "A safer youtube kids: An extra layer of content filtering using automated multimodal analysis," in *Proceedings of SAI Intelligent Systems Conference.* Springer, 2018, pp. 294–308.

[9] M. Picha Edwardsson and W. Al-Saqaf, "Drivers and barriers for using blockchain technology to create a global fact-checking database," *Online Journal of Communication and Media Technologies*, vol. 12, no. 4, p. e202228, 2022.

[10] N. Duong-Trung, X. S. Ha, T. T. Phan, P. N. Trieu, Q. N. Nguyen, D. Pham, T. T. Huynh, and H. T. Le, "Multi-sessions mechanism for decentralized cash on delivery system," *Int. J. Adv. Comput. Sci. Appl*, vol. 10, no. 9, 2019.

[11] X. S. Ha, H. T. Le, N. Metoui, and N. Duong-Trung, "Dem-cod: Novel access-control-based cash on delivery mechanism for decentralized marketplace," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom).* IEEE, 2020, pp. 71–78.

[12] N. T. T. Le, Q. N. Nguyen, N. N. Phien, N. Duong-Trung, T. T. Huynh, T. P. Nguyen, and H. X. Son, "Assuring non-fraudulent transactions in cash on delivery by introducing double smart contracts," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, pp. 677–684, 2019.

[13] N. Duong-Trung, H. X. Son, H. T. Le, and T. T. Phan, "On components of a patient-centered healthcare system using smart contract," in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*, 2020, p. 31–35.

[14] ——, "Smart care: Integrating blockchain technology into the design of patient-centered healthcare systems," in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*, ser. ICCSP 2020, 2020, p. 105–109.

[15] H. X. Son, T. H. Le, N. T. T. Quynh, H. N. D. Huy, N. Duong-Trung, and H. H. Luong, "Toward a blockchain-based technology in dealing with emergencies in patient-centered healthcare systems," in *International Conference on Mobile, Secure, and Programmable Networking.* Springer, 2020, pp. 44–56.

[16] N. H. Tuan Khoi *et al.*, "Vblock - blockchain based traceability in medical products supply chain management: Case study in vietnam," in *International Conference on Artificial Intelligence for Smart Community*, 2020.

[17] H. T. Le, L. N. T. Thanh, H. K. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, K. H. N. Vuong, H. X. Son *et al.*, "Patient-chain: Patient-centered healthcare system a blockchain-based technology in dealing with emergencies," in *International Conference on Parallel and Distributed Computing: Applications and Technologies.* Springer, 2022, pp. 576–583.

[18] H. X. Son, M. H. Nguyen, N. N. Phien, H. T. Le, Q. N. Nguyen, V. Dinh, P. Tru, and P. Nguyen, "Towards a mechanism for protecting seller's interest of cash on delivery by using smart contract in hyperledger," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 4, pp. 45–50, 2019.

[19] H. H. Luong, T. K. N. Huynh, A. T. Dao, and H. T. Nguyen, "An approach for project management system based on blockchain," in *International Conference on Future Data and Security Engineering*. Springer, 2021, pp. 310–326.

[20] N. H. Tuan Khoi *et al.*, "Domain name system resolution system with hyperledger fabric blockchain," in *International Conference on Inventive Computation and Information Technologies*, 2022.

[21] X. S. Ha, T. H. Le, T. T. Phan, H. H. D. Nguyen, H. K. Vo, and N. Duong-Trung, "Scrutinizing trust and transparency in cash on delivery systems," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2020, pp. 214–227.

[22] E. Bouma-Sims and B. Reaves, "A first look at scams on youtube," *arXiv preprint arXiv:2104.06515*, 2021.

[23] A. Tripathi, K. K. Bharti, and M. Ghosh, "A study on characterizing the ecosystem of monetizing video spams on youtube platform," in *Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services*, 2019, pp. 222–231.

[24] K. Papadamou, A. Papasavva, S. Zannettou, J. Blackburn, N. Kourtellis, I. Leontiadis, G. Stringhini, and M. Sirivianos, "Disturbed youtube for kids: Characterizing and detecting inappropriate videos targeting young children," in *Proceedings of the international AAAI conference on web and social media*, vol. 14, 2020, pp. 522–533.

[25] T. C. Alberto, J. V. Lochter, and T. A. Almeida, "Tubespam: Comment spam filtering on youtube," in *2015 IEEE 14th international conference on machine learning and applications (ICMLA)*. IEEE, 2015, pp. 138–143.

[26] V. Chaudhary and A. Sureka, "Contextual feature based one-class classifier approach for detecting video response spam on youtube," in *2013 Eleventh Annual Conference on Privacy, Security and Trust*. IEEE, 2013, pp. 195–204.

[27] S. Zannettou, S. Chatzis, K. Papadamou, and M. Sirivianos, "The good, the bad and the bait: Detecting and characterizing clickbait on youtube," in *2018 IEEE Security and Privacy Workshops (SPW)*. IEEE, 2018, pp. 63–69.

[28] S. Ghimire, J. Y. Choi, and B. Lee, "Using blockchain for improved video integrity verification," *IEEE Transactions on Multimedia*, vol. 22, no. 1, pp. 108–121, 2019.

[29] S. K. Sathish, A. A. Patankar, and H. Khanna, "Aurum: A blockchain based decentralized video streaming platform," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019, pp. 1–8.

[30] P. Banerjee, C. Govindarajan, P. Jayachandran, and S. Ruj, "Reliable, fair and decentralized marketplace for content sharing using blockchain," in *2020 IEEE International Conference on Blockchain (Blockchain)*. IEEE, 2020, pp. 365–370.

[31] A. Zelensky, V. Voronin, E. Semenishchev, I. Svirin, and A. Alepko, "Video content verification using blockchain technology," in *2018 IEEE International Conference on Smart Cloud (SmartCloud)*. IEEE, 2018, pp. 208–212.

Blood and Product-Chain: Blood and its Products Supply Chain Management based on Blockchain Approach

Phuc Nguyen Trong, Hong Khanh Vo, Luong Hoang Huong, Khiem Huynh Gia, Khoa Tran Dang,
Hieu Le Van, Nghia Huynh Huu, Tran Nguyen Huyen, The Anh Nguyen, Loc Van Cao Phu,
Duy Nguyen Truong Quoc, Bang Le Khanh, Kiet Le Tuan
FPT University, Can Tho City, Viet Nam

Abstract— This paper provides a novel implementation of blockchain technology, and data is stored in a decentralized distributed ledger to assist information protection in blood supply chain management and prevent data loss or identity theft. The present blood supply is used exclusively from the blood of volunteers (known as donors), making blood and its derivatives one of the significant roles in treating diseases today. In particular, depending on the type of product extracted from the blood (e.g., red blood cells, white blood cells, platelets, plasma). They require different procedures and storage environments (e.g., time, temperature, humidity). However, the current blood management processes are done manually - where all medical staff does all data entry. Additionally, data about the complete blood donation process (e.g., blood donors, blood recipients, blood inventories) is held centrally and is challenging to examine accurately. Therefore, ensuring centralized data security is extremely difficult because of stealing personal information or losing data. In this study, we present the blockchain technology-based blood management process and offer Blood and Product-Chain, a decentralized distributed ledger that stores data to address these restrictions. Specifically, we target two main contributions: i) we design the Blood and Product-Chain model to manage all relevant information about blood and its products based on blockchain technology, and ii) we implement the proof-of-concept of Blood and Product-Chain by Hyperledger Fabric and evaluate this in the two scenarios (i.e., data creation and data access).

Keywords—Blood donation; blockchain; hyperledger fabric; blood products supply chain

I. INTRODUCTION

The rising need for medical supplies and services, particularly blood, due to which it is necessary to improve blood management. Blood has many components, such as red blood cells, white blood cells, platelets, plasma, and other features [1]. Each blood component serves a specific function in the human body; for example, red blood cells help transport oxygen to cells/parts of the body, and a liter of blood can sustain the life of a premature baby for two weeks, and an accidental blood loss trauma sufferer may need 40 or more units of blood to survive. Therefore, blood is an essential medical resource, and blood management is a fundamental problem the human species needs to solve.

However, all blood collected must be analyzed before being transferred to the recipient. One important part of this work is reducing the risk of infection by transmission [2]. Specifically, the common infection mechanism is hepatitis B virus (HBV),

human immunodeficiency virus (HIV), and hepatitis C virus (HCV). Besides, blood is generally collected from volunteers; hence the supply chain for blood and its products is impacted by a lack of donors, a delay in testing, and, most importantly, the short shelf life of perishable blood products [3]. However, reducing these risks necessitates a difficult process and expensive blood supply chain management.

It can be argued that optimizing the blood supply chain management process is the key to solving the current blood shortage problem in medical facilities. The traditional model is depicted in Fig. 1. Specifically, the blood donation center will directly contact donors (i.e., in case of an emergency) or organize events calling for community blood donations. Donors must undergo a health assessment before donating blood. Then, this raw blood is converted into different components such as red blood cells, white blood cells, platelets, and plasma. The amount of blood and its products after censorship will be partially transferred to medical / health care centers (periodically), or the rest will be stored in the warehouse in case of emergency. It is also important to consider how blood is distributed and stored because different blood components have varied usage requirements and lifespans.

Besides, the information of blood, recipients, and donors is stored in a centralized server. Specifically, information about donors is divided into many areas. If those donors are not at their residence address, they cannot participate in the system. Additionally, the information stored is very private for donors and recipients (i.e., it comprises information unrelated to blood illnesses). Risks related to centralized storage include user misconduct and information loss. To solve a series of important issues mentioned above, we applied blockchain and smart contract technology (i.e., Hyperledger Fabric platform).

Blockchain technology is known for its outstanding features of transparency and immutable content. Picha Edwards-son et al. studied the possibility of using blockchain technology to create a secure, community-facing information verification database with the goal of creating a solution that could improve the reliability of verifying information and monitoring each authenticity verification process for digital content, including images and videos. The paper indicates that blockchain is not yet ready to be directly applied to fact-checking processes in a real-world scenario. The study also shows that the application of blockchain to verify a scenario is entirely possible and highly reliable and transparent [4]. Several approaches address

these problems by applying Blockchain techniques in the other environment (e.g., cash-on-delivery [5], [6], [7], healthcare [8], [9], [10], supply chain [11], [12], [13], and others [14], [15], [16]).

Therefore, this paper focuses on building Blood and Product-Chain: the management transportation of blood and its products based on blockchain technology. The main contributions of Blood and Product-Chain are two-fold. i) we design the Blood and Product-Chain model to manage all relevant information about blood and its products based on blockchain technology, ii) we implement the proof-of-concept of Blood and Product-Chain by Hyperledger Fabric and evaluate this in several scenarios.

In this paper, the structure is organized as follows. The next section of the paper presents a literature review of relevant issues. The Blood and Product-Chain overall architecture and implementation are shown in two following sections (III and IV). Section V describes the framework and benefits of the proposed system via evaluation. Finally, Section VI dedicates the conclusions and future research opportunities.

II. Related Work

A. Healthcare Systems based on Blockchain Technology

There are several blockchain-based approaches in terms of healthcare/blood donation management systems. For example, Du et al. [17] and Son et al. [18] used medical centers (i.e., hospitals) to store data and manage access and those hospitals. Specifically, they categorize two types of medical data protection policies: global for all data shared outside of the medical center, and local, which is accessed only by individuals at the medical center. medical (i.e., doctor, nurse). However, one of the major limitations is that through this solution, patients do not have full control over their data as the data and policies are stored in the hospital. Some other approaches build a user-centric (i.e., patient) model, which has full discretion to share their personal data with providers/health care facilities. economic (i.e., in a medical setting). For instance, Makubalo et al. [19] have summarized the above approaches in their publication. They argue that the methods of building a user-centric health data sharing system are facing a lot of difficulties due to the limitations of the method of building centralized data system (i.e., data) stored and processed centrally in cloud servers). Yin et al. [20] introduced a patient-centric system built in the cloud with a data collection layer, data management layer, and medical service delivery layer based on the medical records of the patient. To protect data privacy, many approaches have adopted attribute-based encryption (ABE), one of the most common encryption schemes used in cloud computing, to define patient data object. Depending on the context, the policy tells to lose (or not) grant the corresponding access rights. For example, Barua et al. [21] proposes an ABE-based access control model based on patience and privacy protection; Chen et al. [17] described a new framework with a cloud-based, privacy-aware Role-Based Access Control model that can be used for control, data traceability, and access allowed access to healthcare data resources. Methods for applying the Access Control model are also introduced for dynamic policies [22], [23] or protection policies for both security and privacy [24]. Ateniese [25]

proposed a new framework that makes it possible to re-write or compress the content of any number of blocks in decentralized services exploiting the blockchain technology.

One disadvantage of centralized storage in the above approaches is transparency [26], [27]. To address this issue, Lam et al. [28], [29] demonstrated the implementation of a micro services-oriented software architecture for middleware that collects, stores, and traces data in a centralized manner in order to provide data analysis. However, for the specific requirement (e.g., blood donation w.r.t supply chain), we need a specific model to balance the demand (i.e., blood recipients) and supply (i.e., the time for the next round of donors). To apply these advantages, many studies focus on blockchain-based approaches for the blood supply chain donation issues, which contribute to the improvement of demand and supply requirements.

B. Blood Supply Chain Management Systems based on Blockchain Technology

Trieu [30] proposed the cold chain model for blood donation based on the Hyperledger Fabric platform called BloodChain.[1] They only considered the blood data and ignored its products as well as the other storage requirement for them. The main contribution of the two above papers is only to verify the information of the recipients to the donors. These papers are very soon introduced in the large picture of this problem. Similar to BloodChain, Lakshminarayanan et al. [32] presented a blood supply chain management system based on Hyperledger Fabric to ensure the transparency requirement in terms of blood units between donors and recipients. To detect the location of the donor's blood, on the other hand, Toyoda et al. [33] provided hardware-based approaches (i.e., RFIDs) in combination with the blockchain to track the status of the blood donation. However, there are some limitations to the solutions above. For example, the verification of the system proposed in incomplete due to the lack of evaluation analysis. Furthermore, the monitoring solution proposed in [33] is limited to monitoring blood bags only, and it does not guarantee the traceability of blood components (i.e., red blood cells, platelets, white blood cells, platelets and plasma). Since different blood components have other shelf lives and storage temperatures, the order of the user preference should also be considered.

Moreover, the authors in [34] have proposed a decentralized solution based on an Ethereum-based blockchain for blood transport. In their design, certified blood donation centers (CBDCs) are the only privileged members with the right to create smart contracts to manage the entire system. Donors were recognized through an identifier such as their social security number and password. Besides, Peltoniemi et al. [35] discussed how decentralized blockchain was for plasma tracking and management. In more detail, the system stores donors' information before separating their plasma. They then maintain the origin of the plasma and identify poor blood quality. Another solution based on Hypeledger was proposed by Kim et al. [36] to build a blockchain-based blood supply chain management system. For the GDPR privacy compliance, Campanile et al. [37] presented literature solutions and design

---

[1]Their model is the extended version of [31].

implementation in the context of a traffic management system for the Internet of Vehicles based on the Pseudonymization/Cryptography solution, evaluating its viability, its GDPR compliance and its level of risk. In addition to the fact that this solution is based on a private blockchain, to protect the privacy of donors, it only focuses on tracing information related to the supply chain, but it does not track the identity of the donors as well as the part of a blood unit.

To sum up, none of the above blockchain-based solutions guarantees data privacy while providing a robust system for tracking and managing the donated blood supply chain. In addition, the above methods focus only on managing blood information rather than considering their products (i.e., red blood cells, white blood cells, platelets, and plasma). This is extremely important because each product has different management conditions and usage times. Our proposed solution captures various aspects of the blood donation system such as collection, distribution, request and delivery of blood units. It ensures that all these aspects are captured as decentralized, traceable, accountable, transparent, secure, and auditable. Our solution tracks all necessary stages of the donated blood unit cycle, from donation until consumption.

III. BLOOD AND PRODUCT-CHAIN ARCHITECTURE

To address the limitation of the traditional approach which mentioned in Section II, we first introduce the overall architecture, which provides the overview model. In this model, we provide the key components (i.e., actors) as well as the relationship among them. After that, the relationship and main steps are presented in the detailed model which highlights the Blood and Product-Chain in nine main steps.

A. Overview Model

There are several approaches that provide the traditional model for the blood supply chain (e.g., [38]). In this section, we target the overall architecture of the Blood and Product-Chain model before focusing on the two main processes that store the donor's information in the distributed ledger and use this info to contact the donor candidate for blood donation in the next round.

For the first target, Fig. 1 shows the three actors who have the main role in our proposed system, namely the Doctor/Nurse, recipients, and donors. Whereas, the remaining components represent the medical facilities as well as the blood-collection/protection centers/ways (e.g., blood transporter, blood bank), respectively. All the corresponding data of the facilities or actors are stored and processed in the distributed ledger. This process allows the stakeholders to verify their data before using it. This also increases transparency for the whole system (i.e., accessing and performing blood data/transferring the blood of the donors to the other facilities).[2] Thereby, medical facilities can retrieve and confirm data related to the treatment process.

For the second target, we present the two main blood donation process in Fig. 2 and 3 for the donor's information and

---

[2]Regarding the collected data for each donor, we store the blood type, number of blood donations, and its metadata (e.g., time, location), as well as other personal information in a decentralized approach (i.e., distributed ledger).

contacting process in medical facilities, respectively. The main contribution of this task is that we consider the decentralized system, where a certain facility is allowed to contact the other donors, who donate their blood in another hospital, rather than only their list.

B. Detailed Model

Blood and Product-Chain is presented in Fig. 4 with eight main actors including medical staff (nurses, doctors), recipients, donors, hospitals (i.e., clinic, medical clinic, mobile blood collection unit), hematology hospital, blood bank, blood transporter, and distributed ledger. Blood and Product-Chain focuses on building a Blockchain-based blood supply chain model, where the transport of blood between medical facilities plays an important role and reduces the pressure on the hematology hospital in the blood collection process from donors. Specifically, step 1 collects the treatment history of recipients through the records of nurses and doctors (i.e., medical staff). All this information will be updated to the hospital where the recipients are being treated (Step 2). In the event that the amount of blood and blood products is insufficient during treatment, the hospital will submit a request by updating the current status of blood amount on the distributed ledger. All information about blood volume, blood type and other special requirements are shared with relevant parties in Blood and Product-Chain (Step 3). At this step, the data of medical facilities is also updated. Blood volumes in a medical facility's system can be shared with other facilities to optimize blood availability. To do that, the establishment needs to update the status and send a request to the transporter about the collection and receipt locations of blood and blood products (Steps 4 and 5). After that, the transporter must always update the progress of the shipment on the distributed ledger (Step 6). All information about location, time, and order information can be verified and moderated by the stakeholders. We continue to consider another case where the requested amount of blood and blood products is not available in the system, i.e., the state of blood and blood products is not available. This request is forwarded to the hematology hospital (Step 7). Step 8 describes the procedure for checking the amount of blood and blood products in stock. Step 9 contacts the donors to schedule a blood test and donation if the requested amount of blood is unavailable. The list of donors is gathered from previous blood donation sessions and selected the candidates with the closest addresses.

IV. IMPLEMENTATION

A. Permission Diagram

Fig. 5 presents the working mechanism of the request authentication process in this paper. Specifically, we built two organizations with corresponding encrypted material certificates, each organization includes two users and two peers. Each peer is responsible for maintaining the version of the ledger so that the network and data can be maintained even if other peers are shut down.

When the user initiates a request and sends it to the service. The back-end service processes the data and sends it to the smart contract API. When receiving the request and the data, the smart contract sends this to the peers in the network
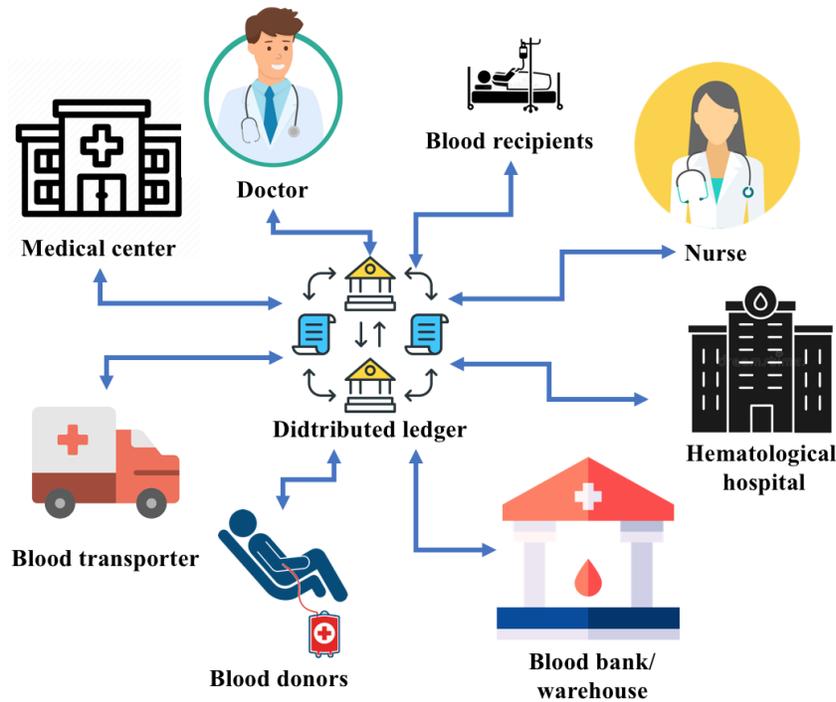
Fig. 1. The Overall Architecture of the Blood&Product-Chain.
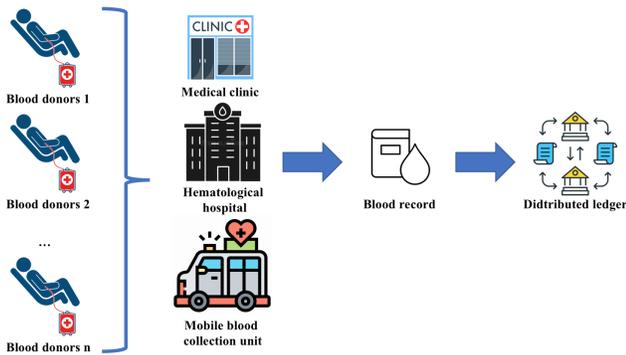
Fig. 2. The Storage Process of the Donor's Information in Distributed Ledger.
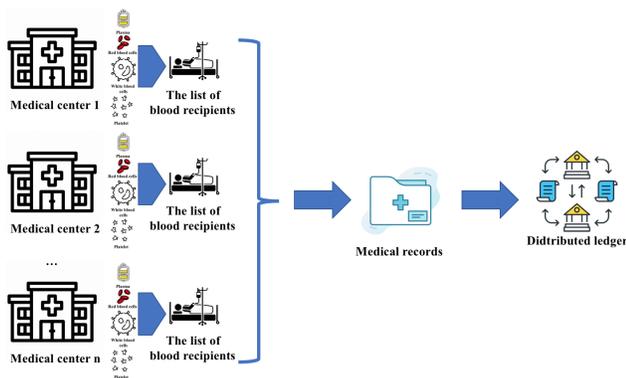
Fig. 3. The Storage Process of the Medical Center's Information in Distributed Ledger.

for authentication and data interaction purposes. During the creation, querying or updating data processes, peers check the identity of the request to decide whether to allow access to the data at the distributed ledger. If the identified user of the request is not defined in the data collection, the system denies access and sends a message to the back-end API to notify the user; the system allows access and proceeds with further processing steps.

B. Hyperledger Component

The model in this paper is implemented on the Hyperledger Fabric platform. Fabric is a permissionless blockchain platform that integrates smart contracts, the storage of data to the distributed ledger is controlled through the smart contract APIs, from which the data is simplified and easily traced. Each request that goes through the smart contract is verified with public and private key pairs. In other words, if the user does not exist in the system, the system is better protected from malicious requests outside the system.

The Fabric system in this paper includes two organizations. Each organization consists of two peers to store smart contracts, where each peer registers two users and is authenticated with public and private key pairs. The components of the model are shown in Fig. 6

When user devices access the system to initiate/query or update data for a particular transaction, requests are sent from the client to the services of the existing system. Then, these services send access information to the peers belonging to the organization located in the blockchain network. At this step, the peers conduct verification of that user's key pairs, and if the successful peer authentication process proceeds to send information to the smart contract with the transaction
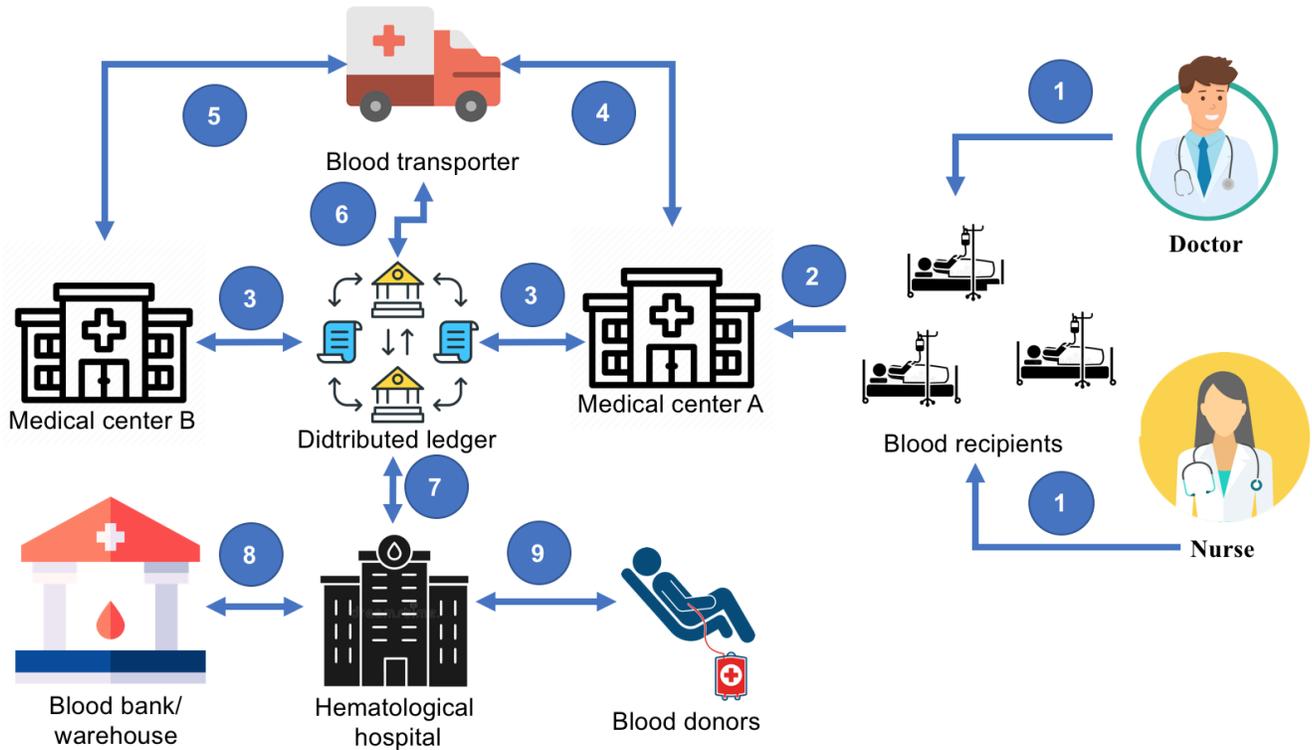
Fig. 4. The Detailed Architecture of the Blood & Product-Chain.
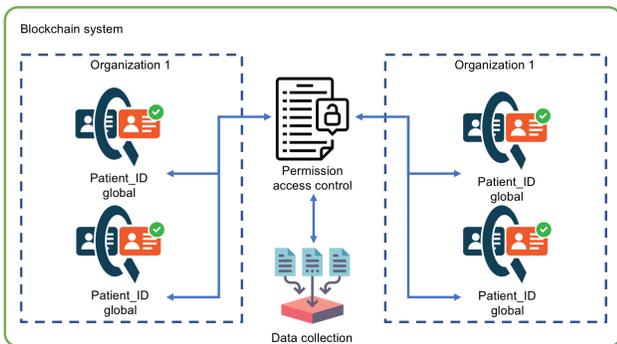
Fig. 5. Permission Diagram.
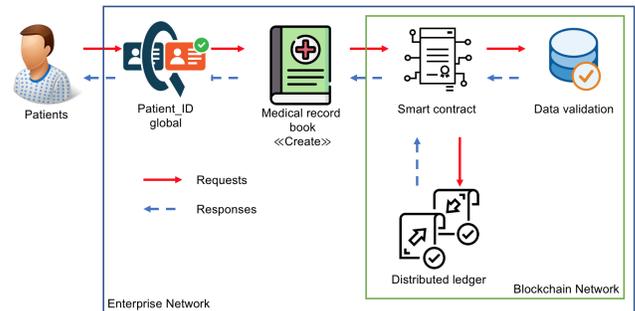
Fig. 7. Initializing and Storing the New Data.
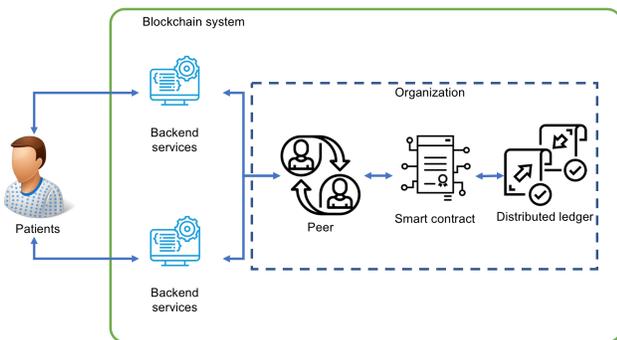
Fig. 6. Hyperledger Fabric Component.

type declared in a smart contract requested by the user, the smart contract will go through the designed features function to access the distributed ledger to initiate/query or update specific data.

C. Blood Pro-Chain Diagram

One of the most important parts of the model lies in the validation and interaction with the blood records described in Fig. 7 and 8. In particular, the main functions include initializing and querying the blood records.

Fig. 7 depicts the process of storing new record data (e.g., blood and its products). In step 1, when the user initializes information of a blood record, the data is sent to the back-end service of the health center's information management system. In the next step, the back-end APIs (i.e., backend) check, authenticate and initialize the default values, then pass

Fig. 8. Retrieving/Querying Data Process.

TABLE I. DATA CREATION/INITIALIZATION (E.G., BLOOD RECORD) RESULTS IN THE BLOOD & PRODUCT-CHAIN

#requests	Succ	Fail	Send Rate (TPS)	Max Latency(s)	Min Latency(s)	Avg Latency(s)	Throughput (TPS)
1000	30505	18607	163.7	1251.62	1.21	780.6	31.9
2000	32196	18494	169	1324.05	11.62	831.13	31.3
3000	29776	22283	173.5	1186.66	12.63	762.59	35.2
4000	28269	21524	165.9	1138.24	7.24	722.93	35.1
5000	34362	17087	171.5	1366.83	6.85	837.88	31.5
6000	32828	16485	164.4	1306.1	6.29	818.69	30.8
7000	34831	14861	165.6	1347.43	5.6	843.07	30.3
8000	27970	21070	163.5	1015.16	8.1	666.99	37.4
9000	30158	21698	172.9	1043.7	0.92	677.78	38.8
10000	23672	37894	205.2	840.33	6.17	600.35	54.1

TABLE II. DATA ACCESS (E.G., BLOOD RECORD) RESULTS IN THE BLOOD & PRODUCT-CHAIN

#requests	Succ	Fail	Send Rate (TPS)	Max Latency(s)	Min Latency(s)	Avg Latency(s)	Throughput (TPS)
1000	107654	5646	377.7	8.36	0.01	5.79	376.2
2000	108426	5639	380.2	7.9	0.01	6.1	378.8
3000	109370	5964	384.4	8.04	0.01	6.17	382.8
4000	109103	5940	383.5	7.98	0.01	6.17	382
5000	110939	5240	387.3	7.68	0.02	6.02	385.9
6000	110992	5452	388.1	7.86	0.01	6.08	386.9
7000	110769	5271	386.8	7.7	0.01	6	385.2
8000	110329	5844	387.2	8.02	0.01	6.04	385.8
9000	111384	5235	388.7	7.82	0.01	5.95	387.4
10000	85194	1643	289.5	8.09	0.01	2.89	289.1

the parameters to the API inside the smart contract. At this point, a smart contract transfers data and stores transactions to the distributed ledger of the blockchain network. The default values for parameters sent from the request are intended to minimize errors caused by null fields data.

Fig. 8 presents the process of retrieving data of a particular (e.g., blood and its products). When the user sends a query request to the system, the service query data checks and confirms whether the parameter ID of the blood record was sent or not. Then, the smart contract's APIs are called and passed into the corresponding parameter. Next, the smart contract's APIs check for the existence of data in the request before querying. In the case that the ID does not exist, the smart contract sends an error notification to the user device; otherwise, it returns the data of the blood record corresponding to the requested ID.

V. EVALUATION SCENARIOS

A. Environment Setting

Our paradigm is deployed on the Hyperledger Fabric network maintained inside docker containers. In this section, we measure the performance of chaincode in the two scenarios: initializing (i.e., creating data) and accessing data. The experiments are deployed on Ubuntu 20.01 configuration, core i5 2.7Ghz, and 8GB RAM.

To prove the effectiveness of our model, we also define several experiments by exploiting the Hyperledger Caliper[3] that is used to design the test scenarios and collect all the information regarding the performance.

B. Results

1) Data Creation: In this scenario, the study measures the performance of the data initialization function/data created (e.g., blood record) performed through smart contracts. The number of requests sent simultaneously from two users[4]. Table I shows the execution results of the data initialization/creation function (e.g., blood record). The data initialization/creation script is conducted with two users concurrently making 1000 - 10000 requests to the system. We measure the parameters of command success/failure, send rate (transaction per second), system latency (i.e., max, min, avg), and throughput (transaction per second) Based on the execution results in Table I, it

---

[3]https://www.hyperledger.org/use/caliper

[4]We set up one organization with two users and two peers

can be seen that the number of successful and failed requests is stable (except in the case of 10000 requests). In the first nine scenarios (1000 - 9000 requests), the number of failed requests ranges from 14861 - 22283, while the number of successful requests is maintained at a much higher rate, from 27970 - 34831 requests. However, in the last scenario, the failure rate is higher than the success rate, 37894 and 23672, respectively. This proves that the system works well with the scenario from 9000 requests (i.e., two users - 2 peers). In addition, we also evaluate the latency of the whole system. Specifically, the maximum delay value ranges from 840.33 to 1366.83 seconds. Besides, the value of send rate ranges from 163.5 to 173.5 (TPS) in the first nine scenarios and the highest in the tenth scenario with 205.2 (TPS). Similarly, throughput measurements ranged from 30.3 to 38.8 (TPS) in the first nine scenarios and reached a maximum value of 54.1 (TPS) in the last one.

2) Data Access (Retrieving/Querying): In the second experiment, we consider the data access (e.g., blood record). We also setup 10 scenarios from 1000 to 10000 requests which access the blood record from two users. Table II shows the execution results of the data access function (e.g., blood record). Compared with the first solution, the results of 10 scenarios to evaluate the data accessibility of Blood & Product-Chain are more balanced. Specifically, the successful data retrieval commands rate accounted for 95% (ranging from 85194 to 111384 requests). The system's latency is also minimal, with a maximum of 8.36 seconds and a minimum of 0.01 seconds. The send rate value achieves very high performance from 289.5 to 388.7 (TPS). Similarly, throughput values range from 289.1 to 387.4 (TPS).

C. Discussion

This is the first attempt to develop a system for the management and storage (i.e., supply chain of blood and its products) of products extracted from blood. Compared to other approaches (see more detail in the related work subject), we aim for adaptability and load-bearing with a large number

of requirements described in scenarios in the data creation and data access sections. Our approach can guarantee a large number of requests (i.e., 10K requests/sec) to initiate and retrieve data.

In future work, we aim to build an authorization mechanism for stakeholders [39], [40]. Specifically, users (e.g., donors) are allowed to design policies to manage their personal data (e.g., what data to share, and with whom). Moreover, this research result is only the first step toward building a system based on blockchain technology in a real environment. We, therefore, aim to implement the proposed model for export in more complex scenarios where healthcare facility processes have multiple roles and execute off-chain (i.e. out of scope for the current version) of medical facilities.

VI. CONCLUSION

The paper applies the benefits of Blockchain technology (i.e. transparency, decentralized storage) to propose Blood and Product-Chain: blood and its products transportation management process is based on the limitations of the traditional system. The paper provides a proof of concept based on the Hyperledger Fabric platform, which stores information about blood and blood products during storage and transport. Information is transparently stored for easy verification during transportation and storage. Detailed assessments of the number of successful, and failed requests, latency, send rate (TPS), as well as throughput (TPS), analyzed based on the Hyperledger Caliper platform, proved the feasibility of our approach.

REFERENCES

[1] M. Bohonek, D. Kutac, J. P. Acker, and J. Seghatchian, "Optimizing the supply of whole blood-derived bioproducts through the combined implementation of cryopreservation and pathogen reduction technologies and practices: An overview," *Transfusion and Apheresis Science*, vol. 59, no. 2, p. 102754, 2020.

[2] P. Sullivan, "Developing an administrative plan for transfusion medicine—a global perspective," *Transfusion*, vol. 45, pp. 224S–240S, 2005.

[3] M. M. Jansman and L. Hosta-Rigau, "Recent and prominent examples of nano-and microarchitectures as hemoglobin-based oxygen carriers," *Advances in colloid and interface science*, vol. 260, pp. 65–84, 2018.

[4] M. Picha Edwardsson and W. Al-Saqaf, "Drivers and barriers for using blockchain technology to create a global fact-checking database," *Online Journal of Communication and Media Technologies*, vol. 12, no. 4, p. e202228, 2022.

[5] N. Duong-Trung, X. S. Ha, T. T. Phan, P. N. Trieu, Q. N. Nguyen, D. Pham, T. T. Huynh, and H. T. Le, "Multi-sessions mechanism for decentralized cash on delivery system," *Int. J. Adv. Comput. Sci. Appl*, vol. 10, no. 9, 2019.

[6] X. S. Ha, H. T. Le, N. Metoui, and N. Duong-Trung, "Dem-cod: Novel access-control-based cash on delivery mechanism for decentralized marketplace," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 2020, pp. 71–78.

[7] N. T. T. Le, Q. N. Nguyen, N. N. Phien, N. Duong-Trung, T. T. Huynh, T. P. Nguyen, and H. X. Son, "Assuring non-fraudulent transactions in cash on delivery by introducing double smart contracts," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, pp. 677–684, 2019.

[8] N. Duong-Trung, H. X. Son, H. T. Le, and T. T. Phan, "On components of a patient-centered healthcare system using smart contract," in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*, 2020, p. 31–35.

[9] ——, "Smart care: Integrating blockchain technology into the design of patient-centered healthcare systems," in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*, ser. ICCSP 2020, 2020, p. 105–109.

[10] H. X. Son, T. H. Le, N. T. T. Quynh, H. N. D. Huy, N. Duong-Trung, and H. H. Luong, "Toward a blockchain-based technology in dealing with emergencies in patient-centered healthcare systems," in *International Conference on Mobile, Secure, and Programmable Networking*. Springer, 2020, pp. 44–56.

[11] N. H. Tuan Khoi *et al.*, "Vblock - blockchain based traceability in medical products supply chain management: Case study in vietnam," in *International Conference on Artificial Intelligence for Smart Community*, 2020.

[12] H. T. Le, L. N. T. Thanh, H. K. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, K. H. N. Vuong, H. X. Son *et al.*, "Patient-chain: Patient-centered healthcare system a blockchain-based technology in dealing with emergencies," in *International Conference on Parallel and Distributed Computing: Applications and Technologies*. Springer, 2022, pp. 576–583.

[13] H. X. Son, M. H. Nguyen, N. N. Phien, H. T. Le, Q. N. Nguyen, V. Dinh, P. Tru, and P. Nguyen, "Towards a mechanism for protecting seller's interest of cash on delivery by using smart contract in hyperledger," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 4, pp. 45–50, 2019.

[14] H. H. Luong, T. K. N. Huynh, A. T. Dao, and H. T. Nguyen, "An approach for project management system based on blockchain," in *International Conference on Future Data and Security Engineering*. Springer, 2021, pp. 310–326.

[15] N. H. Tuan Khoi *et al.*, "Domain name system resolution system with hyperledger fabric blockchain," in *International Conference on Inventive Computation and Information Technologies*, 2022.

[16] X. S. Ha, T. H. Le, T. T. Phan, H. H. D. Nguyen, H. K. Vo, and N. Duong-Trung, "Scrutinizing trust and transparency in cash on delivery systems," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2020, pp. 214–227.

[17] M. Du, Q. Chen, J. Xiao, H. Yang, and X. Ma, "Supply chain finance innovation using blockchain," *IEEE Transactions on Engineering Management*, vol. 67, no. 4, pp. 1045–1058, 2020.

[18] H. X. Son, M. H. Nguyen, H. K. Vo *et al.*, "Toward an privacy protection based on access control model in hybrid cloud for healthcare systems," in *International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on EUropean Transnational Education (ICEUTE 2019)*. Springer, 2019, pp. 77–86.

[19] T. Makubalo, B. Scholtz, and T. O. Tokosi, "Blockchain technology for empowering patient-centred healthcare: A pilot study," in *Conference on e-Business, e-Services and e-Society*. Springer, 2020, pp. 15–26.

[20] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-cps: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Systems Journal*, vol. 11, no. 1, pp. 88–95, 2015.

[21] M. Barua, X. Liang, R. Lu, and X. Shen, "Espac: Enabling security and patient-centric access control for ehealth in cloud computing," *International Journal of Security and Networks*, vol. 6, no. 2-3, pp. 67–76, 2011.

[22] N. M. Hoang and H. X. Son, "A dynamic solution for fine-grained policy conflict resolution," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 116–120.

[23] H. X. Son and N. M. Hoang, "A novel attribute-based access control system for fine-grained privacy protection," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 76–80.

[24] Q. N. T. Thi, T. K. Dang, H. L. Van, and H. X. Son, "Using json to specify privacy preserving-enabled attribute-based access control policies," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2017, pp. 561–570.

[25] G. Ateniese, B. Magri, D. Venturi, and E. Andrade, "Redactable blockchain–or–rewriting history in bitcoin and friends," in *2017 IEEE European symposium on security and privacy (EuroS&P)*. IEEE, 2017, pp. 111–126.

[26] H. H. Luong, T. D. Anh, K. N. H. Tuan, and H. X. Son, "Ioht-mba: An internet of healthcare things (ioht) platform based on microservice and brokerless architecture," 2021.

[27] F. Alharbi, "Progression towards an e-management centralized blood donation system in saudi arabia," in *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)*. IEEE, 2020, pp. 1–5.

[28] L. N. T. Thanh, N. N. Phien, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, H. X. Son *et al.*, "Sip-mba: A secure iot platform with brokerless and micro-service architecture," 2021.

[29] N. T. T. Lam, H. X. Son, T. H. Le, T. A. Nguyen, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, and H. V. K. Nguyen, "Bmdd: A novel approach for iot platform (broker-less and microservice architecture, decentralized identity, and dynamic transmission messages)," *International Journal of Advanced Computer Science and Applications*, 2022.

[30] H. T. Le, T. T. L. Nguyen, T. A. Nguyen, X. S. Ha, and N. Duong-Trung, "Bloodchain: A blood donation network managed by blockchain technologies," *Network*, vol. 2, no. 1, pp. 21–35, 2022.

[31] N. T. T. Quynh, H. X. Son, T. H. Le, H. N. D. Huy, K. H. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, N. Duong-Trung *et al.*, "Toward a design of blood donation management by blockchain technologies," in *International Conference on Computational Science and Its Applications*. Springer, 2021, pp. 78–90.

[32] S. Lakshminarayanan, P. Kumar, and N. Dhanya, "Implementation of blockchain-based blood donation framework," in *International Conference on Computational Intelligence in Data Science*. Springer, 2020, pp. 276–290.

[33] K. Toyoda, P. T. Mathiopoulos, I. Sasase, and T. Ohtsuki, "A novel blockchain-based product ownership management system (poms) for anti-counterfeits in the post supply chain," *IEEE access*, vol. 5, pp. 17465–17477, 2017.

[34] M. Çağlıyangil, S. Erdem, and G. Özdağoğlu, "A blockchain based framework for blood distribution," in *Digital Business Strategies in Blockchain Ecosystems*. Springer, 2020, pp. 63–82.

[35] T. Peltoniemi and J. Ihalainen, "Evaluating blockchain for the governance of the plasma derivatives supply chain: How distributed ledger technology can mitigate plasma supply chain risks," *Blockchain in Healthcare Today*, 2019.

[36] S. Kim and D. Kim, "Design of an innovative blood cold chain management system using blockchain technologies," *ICIC Express Letters, Part B: Applications*, vol. 9, no. 10, pp. 1067–1073, 2018.

[37] L. Campanile, P. Cantiello, M. Iacono, F. Marulli, and M. Mastroianni, "Risk analysis of a gdpr-compliant deletion technique for consortium blockchains based on pseudonymization," in *International Conference on Computational Science and Its Applications*. Springer, 2021, pp. 3–14.

[38] H. Le Van, H. K. Vo, L. H. Huong, P. N. Trong, K. T. Dang, K. H. Gia, L. V. C. Phu, D. N. T. Quoc, N. H. Tran, H. T. Nghia10 *et al.*, "Blood management system based on blockchain approach: A research solution in vietnam," *IJACSA*, vol. 13, no. 8, 2022.

[39] S. H. Xuan, L. K. Tran, T. K. Dang, and Y. N. Pham, "Rew-xac: an approach to rewriting request for elastic abac enforcement with dynamic policies," in *2016 International Conference on Advanced Computing and Applications (ACOMP)*. IEEE, 2016, pp. 25–31.

[40] H. X. Son, T. K. Dang, and F. Massacci, "Rew-smt: a new approach for rewriting xacml request with dynamic big data security policies," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2017, pp. 501–515.

360° Virtual Reality Video Tours Generation Model for Hostelry and Tourism based on the Analysis of User Profiles and Case-based Reasoning

Luis Alfaro, Claudia Rivera and Ernesto Suarez
Universidad Nacional de San Agustín de Arequipa
Arequipa, Perú

Alberto Raposo
Pontificia Universidade Católica do Río de Janeiro
Río de Janeiro, RJ - Brazil

Abstract—This paper proposes an adaptive software architecture focused on hotel marketing based on immersive virtual reality (VRI) with 360° videos, which includes a component based on Case-Based Reasoning (CBR) to provide experiences that correspond to the analysis of user profiles. For the validation of the system, considering that the use of VR can trigger experiences in several dimensions, affective, attitudinal and behavioral responses, as well as the cognitive load were evaluated using visualizations of 2D photographs contained in hotel websites , which were compared with 360° videos in a VRI environment. To test the hypotheses, a quasi-experimental study was conducted with an independent sample group, in which subjects were randomly assigned to the two types of visualizations. The contribution of the article lies in the incorporation of marketing concepts and approaches in VRI experiences with 360° videos through virtual objects that are used by the software architecture, as well as in the proposed validation of the effectiveness of the proposal.

Keywords—*Immersive virtual reality; adaptive software architecture; case-based reasoning; user profiles*

I. INTRODUCTION

Making visits to tourist resorts, hotels, museums, laboratories, factories, etc. in tourist, hotel and educational environments through virtual tours mediated by 360° immersive virtual reality (IVR) devices can become an important factor for the success of marketing strategies used by hotel, tourist, educational and other organizations.

VR employs virtual scenarios that allow the user to move freely in the virtual environment, while 360° VR offers spherical/panoramic experiences [1]. Immersive 360° video became popular in contemporary applications due to a higher perceived credibility in those scenarios, considering that 360° VR content is based on real-world photos and videos, which can be more representative of the environment the experience is focused on [2], allowing to offer a clearer and more detailed vision of the environment and a more pleasant experience [3]. Wu and Lin [4] consider that VR environment can be directly linked to 360° videos, impacting the perception-sensation of presence.

This work is focused on the proposal of a system model that performs the division of 360° videos, to then make the composition of tours, considering an intelligent component based on Case Base Reasoning (CBR) that adapts and performs the composition of tours according to the specific requirements of each user profile, which will be displayed as 360° virtual reality videos tours.

For the implementation of the CBR component, some data related to potential guests, such as the area of the hotel, the reason for the trip, whether they will travel accompanied, the services they require from the hotel and the priority they give to each of these services, are considered in the inputs for the CBR algorithm, which obtains the cases from a database based on the information of previous guest visits.

Wan, Tsaur, Chiu, and Chiou [5] suggest that when using VR as a resource for promotion, the context and specific details of the objectives should be considered, because results may vary and a one-size-fits-all approach should be avoided, as there are reported differences in the effectiveness and impacts of VR [6]. In this perspective, cognitive load, affective aspects and behavioral outcomes were identified as factors to be considered when comparing visualizations using VR resources with traditional ones [7]. In order to validate the effectiveness of the proposed model, it is investigated whether the visualization of 360° VR video tours can outperform the visualization of 2D photos used by traditional marketing resources, which is still under study.

Several studies were focused on the evaluation of the content and functions of hotel web sites, based on online satisfaction of consumers. For [8], the content of the website and its functions can directly influence the preferences and decisions of customers. Website functionalities represent the completeness of a website and its features [9], considering a list of hotel website functionalities, whose classification establishes four dimensions [10]: general hotel information, referring to the mechanism through which a hotel introduces products and services to potential consumers; reservation information, referring to the functions and information about the hotel room to be reserved, website management and surrounding information, referring to online services and features that help hotels communicate with potential consumers.

Website usability design is used to determine the experience and interaction of consumers on hotel websites [11], [12]. Au Yeung and Law [13] conducted a usability study adopting a heuristic model, which considered the following aspects: navigability, ease of use of the website, the number of languages available [14]. Consumers' online satisfaction is influenced by the ease of navigating web pages for information search and that a good user-friendly system of a website leads to a positive first impression. The users are always looking for the website quality and quality of hotel services [15].

The previous studies were focused on establishing a comprehensive evaluation of websites both of the functionality and the usability criteria to establish the general usefulness performance of a hotel website, being necessary also to explore and evaluate the influence of the adaptability of tours in hotel facilities and services to the requirements of user profiles, as well as the objective of determining the cognitive, affective and attitudinal responses, in the exploration of hotel web sites, in order to demonstrate whether for the visualization of hotel environments and services, the use of 360°VR whose technical knowledge requirements and costs are lower than those of immersive VR applications, can be more powerful in making users feel influenced in relation to booking the hotel and/or or say positive things to other potential users compared to 2D images.

The following is a description of the points discussed in the different sections. Section II reviews the literature, Section III presents the proposed software model and the use of the CBR engine for the proposed model, Section IV establishes the methodology, Section V describes the tests and discusses the results, and finally, Section VI establishes the conclusions and recommendations for future work.

II. REVIEW OF THE LITERATURE

A. Immersive Virtual Reality Technologies (VRI)

A well known definition definition of VR is: "the use of a computer-generated 3D environment in which the user can navigate and interact, resulting in a real-time simulation of one or more of the user's five senses" [16]. As [17] argue, solutions focused on tourism applications can be non-immersive, semi-immersive and fully immersive, and when a solution is more immersive, the more complicated and complex is the implementation of the technology, devices and content.

Significant improvements in VR technologies impacted more frequent use in tourism research and practice [18]. Various authors divide VR applied in tourism into five categories: planning and management, education, marketing, accessibility, heritage conservation, and entertainment. This paper is specifically focused on marketing in tourism and hospitality. Several authors focused on the similarities and differences between VR and augmented reality (AR) [19]. Other authors focused on VR as a visualization technique in which the real environment completely disappears and the client is immersed in a completely digital world [20]. Other authors state that mixed reality (MR) describes the combination of a real environment with digital content, and can vary perceptions of the real environment without complete virtual immersion [21].

From a hardware perspective, a virtual environment represents a "digital space" in which a user's movements are tracked and their environment rendered, and displayed through various senses, according to those movements [22]. A virtual environment makes it possible to replace the real-world environment by allowing users to block out the physical world of stimuli and fully immerse themselves in the virtual world [23]. The emergence of smartphones and online virtual environments allowed marketing experts to engage tourists as active participants in VR applications, where destinations can be experienced from the comfort of their home [24].

VR technologies, properly used in the proposed adaptive model, which have Artificial Intelligence components, provide an immersive interaction, which is focused on the characteristics and interests of users, which should be framed in the concepts of Experiential Marketing and Sensory Marketing, for the adaptation of experiences and virtual objects in a personalized way and according to the experiences offered by a particular hotel and in a particular environment.

B. 360° Videos in Virtual Immersion

The use of 360° videos is relevant from the use of VR technology to create immersive spaces in which the observer mimics, captures through the senses and feels what has been artificially created in the environment. According to [25], immersive videos show more interesting information within the 360° environment than normal videos and are more eye-catching, due to the greater interest of users in those visual contents, since there is a feeling of being present in the explored content; in addition, that information of interest to explore can be obtained, that is why the number of consumers in the entertainment industry using 360° videos is increasing day by day.

The system includes a perceptual component, coupled with the technological component, which generates in the viewer the illusion of being in another place, despite knowing that this is not real, as highlighted in [26]. The narrative receives particular attention because it has been shown that the deliberate manipulation of the narrative structure affects the phenomenon called "transportation" of the reader or viewer, as mentioned in [27]. The "theory of narrative transportation" [28] describes an experience by which the reader or viewer "disappears" in the narrative worlds through the commented theory.

In [29], it is proposed that immersion should generate a sensation similar to when one is immersed in the ocean: of being surrounded by a completely different reality. In [30], an emphasis is placed on the narrative with the proposal that immersion is a different mental process, an integrative mixture of images, feelings and attention, where the participant's attention and capabilities are focused on the events occurring in the narrative. In that direction, the type of narrative required for virtuality must rethink the current narrative structures, as proposed in [31] and [32]. In [26] it is argued that, in order to rethink the narrative structure, it is necessary to manage the space in such a way that, through it, the viewer perceives the sensation of being immersed in the reality proposed in the 360° video; considering the technology on which the narrative is based. A satisfactory answer to the change of narrative structure has not yet been achieved. The research for this topic focuses on improving the production of 360° videos, which is a challenging path to undertake.

In relation to these VR resources, according to [33] several authors, two dimensions are defined as the virtual presence of the observer, the media or technological characteristics of the system (media form) and the content of that system (content factors); the latter includes the objects, actors or events represented in the narrative.

C. User Profiles

Understanding what users want to get from the services they use is important, as it can facilitate the offer of different

personalized content [34], it is important to evaluate the behavior of users and generate personalized profiles, based on the data obtained, resulting in valuable knowledge for the content creator; additionally, it can allow the possibility of including advertisements or interactive advertising, for example, in the form of questionnaires; but this will not be useful only as an advertising campaign, but can also be a means to collect valuable information about a brand or product, and to support decisions in future marketing actions.

D. Case-Based Reasoning (CBR)

The basic concepts of CBR can be presented as follows: a so-called "base case" is generated, which is a problem question that will be compared with previous cases; for this purpose, a similarity function is used to recognize the possibility of adaptation between the new case and the existing cases in the case base. As a result, the most matching case is selected and this solution will be considered as a suggested solution to the new case. The three steps involved in CBR are:

1) Step 1: Build a case base. For greater accuracy, the complete case base will be accessed.

2) Step 2: Implement a similarity function. The implementation of this function will depend on the experience and logic of the developer. This function must be able to recognize the degree of similarity between the new case and the cases stored in the case base.

3) Step 3: Find the most similar cases and choose the best ones, according to the adaptation criteria.

The degree of similarity between attributes is multiplied by a weight factor, which represents how much that attribute influences the search, and then the sum of the similarity of all the attributes is calculated to obtain an overall measure of the similarity of the case [35]. In this way, we can obtain the percentage of similarity with which it is possible to sort the list of cases that have been previously stored, so we can deduce that the system will be constantly learning and improving.

In [36], it is mentioned that the CBR allows a great capacity for learning and adaptation to the context, since it resorts to past experiences to solve similar problems, as well as to generate creative solutions to new problems. Each time a new user enters to consult about tours that he/she wants to visualize, the probability of finding a visual resource that is useful to him/her increases, so that, thanks to this approach, we can prepare our system for future searches that have not been taken into account due to the great amount of variety that may exist in the preferences of the users.

E. Adaptive Systems in Marketing

For adaptive marketing systems, there are several sources of experience, such as Intelligent Learning Environments (ILE), Tutorial Systems (ITS) [37], Pedagogical Agents (PA) [38], and Adaptive e-Learning systems [39], [40], whose characteristics and adaptation techniques can contribute to achieve experiences and present virtual objects in adaptive marketing systems, according to the profiles, characteristics and expectations of potential hotel guests, improving their exploration and learning activities about hotels. In [41], it is mentioned that it is possible to generate first-person experiences related to the facilities and services provided by a hotel, through information exploration techniques and VR.

In [42] it is pointed out that these systems have as an important feature the ability to adapt the interaction process and the virtual objects, which must be flexible to the specific requirements and profiles of the users. The characteristics of the user must be considered for decision making, elements that constitute a key factor for the success of its development and implementation.

This approach was reviewed in several works [43], which recommend adaptations focused on Cognitive Styles (CS), to which it is convenient to adapt virtual experiences and objects, which resemble the contact with the hotel facilities or services. In [44] it is mentioned that the adaptation to the client's knowledge consists of focusing on the adaptation of the objects to be explored and the presentation of contents, according to the client's knowledge on the subject, about the facilities, services, experiences, landscape, environment, etc. The adaptability of marketing systems is one of the important factors to improve their quality and efficiency.

This work is focused on identifying the adaptability parameters of content and virtual objects for each particular hotel customer and then using a CBR for the adaptation of virtual experiences and objects. All these issues and elements explored have allowed the establishment of the proposed model.

Regarding the gaps and limitations of the related works, it was possible to verify that most of them focus on the usability, interactivity and functionality tests of the applications, but few studies analyze the cognitive load, which could influence the results of the tests and analyzes carried out.

III. Software Architecture Proposal

To provide a better experience to users using the proposed immersive application, a CBR engine was used to provide tour recommendations to potential users, based on user preferences and previous stored experiences, either from the user's own queries or from other users with similar profiles. In this way, if a user is interested in a certain service, they will be shown tours that have that service and fit their requirements.

Within the CBR, each case corresponds to an experience or query previously made. These cases are derived from legacy databases, converting organizational resources into exploitable knowledge. Within the proposal, the cases will be stored when a user decides to view one of the tours offered.

The attributes that make up each case, defined on the basis of the data most commonly used by hotel companies are:

- Id: tour identifier in the database, which will be used to show the recommendations to the user.

- Zone: main characteristic of the area where the hotel is located, for example, near the beach or in the city.

- Reason: represents the reason for the user's trip, e.g., vacation, work.

- Services: represents the services the hotel has; essential to show certain images or 360° videos when viewing a tour.

TABLE I. TABLE OF ATTRIBUTE WEIGHTS

Attribute	Weight
Zone	15%
Motive	24%
Services	17%
Company	20%
Priority	24%
TOTAL	100%

- Company: indicates whether the user is traveling alone or with other people (family, friends, partner, etc.).

- Priority: the most important criterion for the user when choosing a hotel, for example, price.

Six parameters are defined for the CBR search function:

- Attribute names: represents the attributes of the case to be taken into account at the time of the search.

- Search values: represents the search case, which will be the basis for searching for those cases that are most similar to it.

- Weights: importance of each attribute in the search, which allows the CBR to prioritize the attributes. These weights were obtained based on the opinions of experts in hotel management, marketing, researchers, and potential users, as shown in Table I.

- Terms: defines how each attribute will be searched, for example: equal, greater, close values, etc.

- Scales: represents how the difference between the cases and the search case will be searched mathematically, for example: linearly, logarithmically, etc.

- Search options: defines how the search results are sorted, the default option being to return the closest values.

The result of the search is a list of cases that, due to the proposed implemented configuration, will first display in the immersive application interface the cases that are most similar to the base tour, as recommendations, considering the preferences that were entered previously.

After the search, if the user enters one of the recommended tours, a new case is created in the case base, with the information stored in the user's profile, allowing to increase it, which will serve to make recommendations closer to the user's needs and preferences in future searches.

A. Proposed Architecture

For the development, the 4+1 architecture view model was selected, since it allows describing the system architecture using multiple concurrent views [45], as shown in Fig. 1. The views describe the system from the point of view of the main stakeholders, such as end users, developers, and researchers.

The proposed model consists of four components, where the intelligent component is responsible for adapting to the characteristics and needs of customers, which requires the identification and knowledge of the characteristics of the hotel,
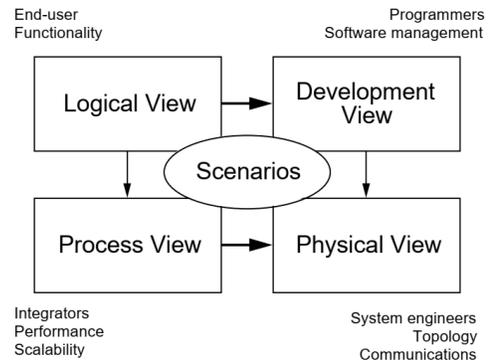
Fig. 1. Proposed Architecture. Source: Own Elaboration.
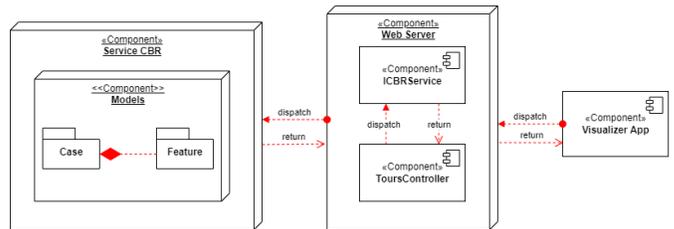
Fig. 2. Development View. Source: Own Elaboration.

its services and its environment, which will be represented in virtual objects, which must be provided to the customer, according to their characteristics and interests.

The application uses the client-server model, where the server is responsible for transmitting hotel and tour information to customers through the immersive application, and for collecting tour recommendations through the CBR.

B. Development View

The development view is shown in Fig. 2, and considers three main components: the server, the intelligent layer (CBR) and the visualization layer (immersive application). The connection between these components is based on HTTP requests that are originated by the application when the user searches for tour recommendations. The server receives the request and connects to the CBR, which returns the response of the recommendations, using the case and feature models, where a case can be composed of several features.

C. Process View

The process view in Fig. 3 presents the three most important actors: the application, the server, and the intelligent layer (CBR). The process checks if the user's preferences are saved. If they are not, or if they need to be updated, a form will be displayed and the data entered by the user will be saved. If the data is already stored and does not need to be updated, this step is skipped, and the list of recommended tours obtained from the CBR is sent.

D. Physical View

The view shows the interconnection between the web server and the immersive application, as shown in Fig. 4.
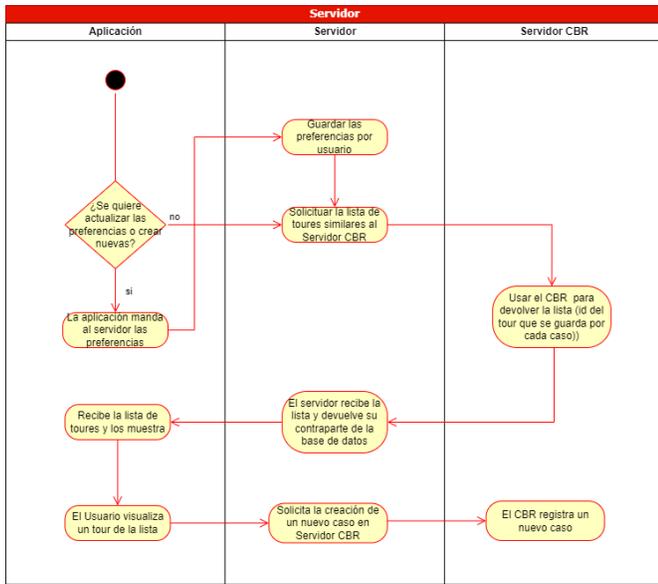
Fig. 3. Process View. Source: Own Elaboration.

Fig. 4. Physical View. Source: Own Elaboration.

The view presents five components, the web server that uses the CBR through an API and obtains the tour information through a database connection, using Entity Framework; the client that uses the connection with the web server to show the recommended hotels and tours to the user, and obtain the 360° images/videos. The intelligent layer connects to the server and only interacts with it. The application is not aware of the existence of this component.

E. Logical View

Fig. 5 shows the four main classes present in the intelligent layer: CBRController, Case, CBR, and Feature. The intelligent layer exposes a Rest API; its main controller is CBRController, whose methods expose two endpoints: getCases, to return the

Fig. 5. Logical View. Source: Own Elaboration.

list of tours related to the input tour, and addCase, to register new cases.

The CBRController class works with the Case class, which is a POJO that stores the attributes mentioned in section III. Along with this class, the CBR class is used, where the CBR cases will be converted to the Case class for its use since the CBR class works with an array of Feature instances, allowing a more comfortable handling of the CBR.

IV. METHODOLOGY

A. Model Development Methodology

For the development of the model, a methodology based on seven stages was followed:

1) Requirements and user story gathering with hotel managers.
2) Design of the customer journey map and analysis of the guest experience by management and marketing specialists.
3) Capture of 360° images and videos in test hotel establishments.
4) Design and prototyping of immersive application user interfaces.
5) Implementation of the visualization layer as an immersive application in Unity, targeting VR glasses.
6) Implementation of the intelligent layer, based on an RBC engine and a Rest API.
7) Functionality testing, based on test cases, and validation testing, based on the TLX-NASA method.

The operation of the CBR starts with the generation of a new search request for tour recommendations, based on the entry of a new case containing the following data: destination zone, reason for the trip, required services, company, and priority in the search.

The CBR configuration defines the terms, scales and weights for the search of recommendations. For this, each case is evaluated by independently comparing each of its attributes. Then, a total value (hit) will be displayed for each case, which represents the distance in the search space, and is used as an indicator of the similarity of the case. The lower the value, the greater the similarity to the search case, i.e., a hit of 0.0 translates as identical cases. At the end of the evaluation of all the cases, a list ordered by similarity will be returned, according to the hit value, eliminating the identical cases in order not to repeat results, and the information obtained from the CBR can be used.

The web server uses the list of tour IDs to retrieve the tour information to display and recommend to the user. If the user decides to view one of the recommended tours, it is considered to match the user's interests, so the search parameters and the selected tour are stored as a new case in the CBR, allowing it to be used in subsequent searches, and to obtain greater precision in future recommendations within the immersive application.

For the validation of the CBR, the case library subset test (CLST) technique, proposed by [46], is used, where the main idea is to select a subset of the case library and use this subset to evaluate the effectiveness of the recovery and adaptation functions of the system. For this purpose, the validation criterion is defined, which consists of two parameters:

- Result Acceptability Criteria (RAC): The maximum acceptable relative error of the solution compared to the standard. A RAC of 15% has been defined.

- System Validity Criteria (SVC): Threshold value to determine whether, after running and evaluating the subset of cases, the system can be considered valid. The percentage of accepted cases must be greater than this value. An SVC of 75% has been defined.

A total of 50 cases were used for the retrieval and adaptation tests under the CLST technique. The results obtained from the tests are shown in Table II.

After the tests were performed, 100% acceptance was obtained in the recovery tests, and 88% in the adaptation tests, so the proposed CBR model is considered valid, according to the SVC (75%).

B. Methodology for Validating the Effectiveness of the Model

The literature review on consumer behavior establishes that affective, cognitive and attitudinal responses are important elements in understanding consumer habits, the decision-making process and ultimately predicting user behavior. The specific research question to validate the results obtained with the architecture is:

How do traditional photo and video displays compare to 360o video displays in VR applications, in terms of (1) cognitive load, i.e., task and experience efforts associated with the technology? (2) affective responses, and (3) attitudinal/behavioral responses.

VR can trigger experiences in several dimensions, such as cognitive, affective, sensory, attitudinal and behavioral [7]:

TABLE II. CBR Test Results

N°	Recovery	Result	Adaptation	Result
1	100.00	Success	100.00	Success
2	100.00	Success	100.00	Success
3	100.00	Success	100.00	Success
4	100.00	Success	100.00	Success
5	100.00	Success	100.00	Success
6	100.00	Success	61.27	Failed
7	100.00	Success	100.00	Success
8	100.00	Success	61.27	Failed
9	100.00	Success	100.00	Success
10	100.00	Success	61.27	Failed
11	100.00	Success	100.00	Success
12	100.00	Success	100.00	Success
13	100.00	Success	100.00	Success
14	100.00	Success	100.00	Success
15	100.00	Success	100.00	Success
16	100.00	Success	100.00	Success
17	100.00	Success	100.00	Success
18	100.00	Success	100.00	Success
19	100.00	Success	100.00	Success
20	100.00	Success	100.00	Success
21	100.00	Success	100.00	Success
22	100.00	Success	100.00	Success
23	100.00	Success	100.00	Success
24	100.00	Success	100.00	Success
25	100.00	Success	100.00	Success
26	100.00	Success	86.26	Success
27	100.00	Success	86.26	Success
28	100.00	Success	86.26	Success
29	100.00	Success	86.26	Success
30	100.00	Success	86.26	Success
31	100.00	Success	100.00	Success
32	100.00	Success	100.00	Success
33	100.00	Success	100.00	Success
34	100.00	Success	100.00	Success
35	100.00	Success	100.00	Success
36	100.00	Success	100.00	Success
37	100.00	Success	56.13	Failed
38	100.00	Success	100.00	Success
39	100.00	Success	51.01	Failed
40	100.00	Success	100.00	Success
41	100.00	Success	100.00	Success
42	100.00	Success	56.13	Success
43	100.00	Success	100.00	Success
44	100.00	Success	100.00	Success
45	100.00	Success	100.00	Success
46	100.00	Success	100.00	Success
47	100.00	Success	100.00	Success
48	100.00	Success	100.00	Success
49	100.00	Success	100.00	Success
50	100.00	Success	100.00	Success

1) Cognitive CLT Response [47]: and the limited perceptual capacity model of attention [48] give a theoretical framework for comprehension the processing of Virtual Reality visualizations. Virtual Reality hotel images will be associated with greater cognitive/perceptual load than traditional forms of imagery, and the next hypotheses are established:

Hypothesis 1a: 360° VR video visualizations of hotel en-

vironments and services will not elicit higher cognitive load compared to traditional photo and video visualizations of the similar scenes.

Hypothesis 1b: 360° VR video visualizations of hotel environments and services will elicit higher cognitive load compared to traditional photo and video visualizations of the similar scenes.

2) Affective Response: Many studies have found that VR experiences provoke emotional responses or conditions, like a galvanic skin response, skin conductance, increased heart rate, blood pressure and respiration [49], because VR engages the senses at an increase intense level, it generate an intense affective response compared with the traditional imagery. The immersive nature of VR, the intensified presence and sensory stimulation become a intense emotional experience. In addition, VR tends to be conducive to evoking a wide range of affective responses, such excitement, arousal, as general emotions of frustration, and pleasure [50]. For that reason, it is logical to assume that hotel VR 360° videos would evoke affective responses, which may be stronger relative to traditional photos and videos. As a result, the next hypotheses are established:

Hypothesis 2a: 360° VR video displays of hotel environments and services will not evoke stronger affective responses compared to photo and 2D video displays of the similar scenes.

Hypothesis 2b: VR 360° video displays of hotel environments and services will evoke stronger affective responses compared to photo and 2D video displays of the similar scenes.

3) Attitudinal and Behavioral Responses: Knowledge on how 360° videos in Virtual Reality affect customer behaviors and attitudes remains insufficient, particularly in tourism and hospitality [51]. Van Kerrebroeck et al. [52] report that consumers had more positive attitudes, satisfaction, and loyalty after being exposed to the VR experience in the mall context.

Studies by Choi and Taylor [53] and Stoyanova et al. [54], propose that sense of presence and representational richness (relevant to Virtual Reality) can produce more positive consumer attitudes towards a brand, as well as higher purchasing intentions.

Based on that assumption, the next hypotheses are established:

Hypothesis 3a: 360° VR video displays of hotel environments and services will not evoke more positive attitudinal responses compared to photo and 2D video displays of the same scenes.

Hypothesis 3b: 360° VR video displays of hotel environments and services will evoke more positive attitudinal responses compared to photo and 2D video displays of the same scenes.

Hypothesis 4a: VR 360° video displays of the hotel surroundings, environments, and services will not evoke stronger behavioral responses compared to photo and 2D video displays of the same scenes.

Hypothesis 4b: VR 360° video displays of the hotel surroundings, environments, and services will evoke stronger

TABLE III. RELIABILITY STATISTICS

Cronbach's alpha	Cronbach's alpha based on standardized elements	Number of elements
.830	.832	10

behavioral responses compared to photo and 2D video displays of the same scenes.

The procedure established is the next: All measurement items keep out demographic information were adopted from another studies.

For affective responses to the visualizations received, the Pleasure, Arousal, and Dominance (PAD) scale [55], widely validated in some studies measuring affective response to colors and physical environments [56], can be used. The scale begins with a statement, "After visualizing the images of the hotel, I feel ..." followed by 18 bipolar items:

1) Pairs of pleasure-related adjectives: unhappy-happy, annoyed-pleased, dissatisfied-satisfied, depressed-content, despairing-hopeful, bored-relaxed.
2) Pairs of arousal adjectives: relaxed-stimulated, calm-excited, slow-frenetic, bored-nervous, sleepy-awake, unexcited-excited.
3) Pairs of adjectives related to dominance: controlled-controlling; influenced-influential; insignificant-important; submissive-dominant; restrained-free.

A 7-point semantic differential scale, from 1 to 7, can be used to assess each bipolar pair [56].

Attitudes and behavioral intentions toward hotels were measured through a 7-item scale adopted from Slevitch and Oh [57]. The measure includes attitudinal questions such as:

1) How likely would it be, that you would decide to book at the hotel you saw?
2) To what extent would you be satisfied with your decision to stay at that hotel?
3) Does the hotel you viewed match its price?

The behavioral questions are as follows:

1) How likely would you be to book such a hotel?
2) How likely would you be to say positive things about the hotel to others?

To measure cognitive workload, one of the most established tests is the NASA TLX, which assesses subjective mental workload [58], and is based on the cognitive demand of a task [59]. The NASA TLX assesses workload in six areas: mental demand, physical demand, temporary demand, performance, effort, and frustration, which are assessed using a differential scale. Its reliability and validity have been confirmed in several studies.

The Cronbach's alpha approach was applied to validate the consistency of the survey questions of both groups, obtaining the satisfactory results shown in the Table III.

A quasi-experimental study was conducted to test the hypotheses. The lobby of a boutique hotel and a beach hotel and various environments were captured in both 2D photos and

videos, as well as 360° videos, which can be viewed through a PC and with an immersive helmet, respectively. Subjects were randomly assigned to two types of visualizations that serve as experimental conditions. The experiment took place in various locations and environments.

For the test case of videos in VR, the VR HMD device that allows immersive experiences based solely on smartphone was used to create an immersive virtual experience [60].

Participants are individually exposed to an experience and must be asked to imagine a situation in which they are choosing a hotel, considering one of the motivations established in the study, as well as a location. Cost was not considered. Subjects were then exposed to one of the experimental conditions for five minutes. Participants in the VR group received assistance in donning the devices and verbal permission was obtained from the participants to use their images in the study.

After receiving the experimental treatment, subjects were asked to fill out a questionnaire on Google forms. The duration of the trial was completed within one month.

V. Tests and Discussion of Results

The sample consisted of 312 potential customers, randomly selected from a population of 1620 users, composed of young people, including professionals, undergraduates and graduate and postgraduate university students. Tests 1 and 2 included an equal number of participants, 156 each. The analysis of the results is as follows:

A factor analysis restricted to the use of only three components was used to (1) establish the dimensionality of the PAD scale Pleasure, Excitement and Dominance - and (2) develop composite variables for based on the sum of values for each factor. Principal component analysis was used as extraction method and varimax with Kaiser normalization was applied as rotation method and the cut-off value of the correlation was 0.5. The factor analysis was adequate, based on Bartlett's sphericity test ($p < 0.0000$), and the KMO test (Kaiser-Meyer-Olkin), which yielded a statistic of 0.838.

A. Cognitive Load Answers

Test 1 and Test 2 produced responses with some differences for most items in the cognitive load component (see 2D Cognitive Dimension Table, available at [61]). Scores were significantly lower for Test 1 compared to Test 2 by 12.5% on the mental demand scale. There was a lower physical demand in 360° VRI visualization compared to 2D by 16.6%. There was also a lower temporal demand in 360° VRI visualization compared to 2D by 13.5%. Less effort was also found in 360° VRI visualization compared to 2D by 8.3%. The frustration level was slightly lower in 360° VRI compared to 2D by 4.2%.

B. Affective Answers

Test 1 and Test 2 produced responses with some differences for most items in the affective load component (see Affective Dimension 2D Table, available online [61]), as shown in the results, and statistically significant differences were identified, equivalent to 17%, with a larger difference observed for females. There was a statistically significant difference in responses between the visual conditions on how insecure,

discouraged, irritated, stressed and annoyed respondents felt about the task. Scores were significantly lower for Test 1 compared to Test 2 by 4.9% on the Pleased-annoyed scale. The 360° VRI images scored lower compared to the scores produced in the 2D condition by -2.1% for the Animated-Abashed scale. Participants who viewed the 2D images scored higher compared to those exposed to the 360° VRI by 9.4% on the tense-relaxed scale.

C. Attitudinal and Behavioral Responses

There was a significant difference between the attitude responses for Test 2 data and a behavioral response in Test 1. Participants in Test 2D felt more positive about staying at the hotel compared to those exposed to 360° VRI images, with a difference of 29%. Participants exposed to 2D photos were more likely to book into the hotel compared to their 360° VRI counterparts by 25.9%. It can be inferred that at this point bookings would have higher preferences for 2D because the hotel's website allows access to information on all the hotel's environments and services, while 360° VRI visualizations could be accessed in a free way through hotspots, or by providing user profile data and displaying only information of interest, showing partial information on the hotel, so further studies are required to corroborate the results obtained.

D. Limitations and Further Research

The data to model and validate the architecture were collected in the city of Arequipa, considering potential hotel clients, as well as Master's in Marketing and undergraduate students from the National University of San Agustín, who may not have accurately reproduced world conditions. real and limit the external validity of the results. Therefore, the application of the results to the hotel industry must be done with caution. In future research, an effort should be made to design conditions that are similar to real world experiences.

The samples met the minimum size requirements, derived in part from the sanitary restrictions derived from the Covid-19 pandemic, but they may still be small. The study also included a limited scope of affective responses as only three dimensions were examined and other potentially relevant emotions were not considered. In the future, psychophysiological measures could be included in the studies to correct deficiencies that could have been caused by the self-reported measures and those that may not have been the best way to capture affectivity.

E. Discussion and Implications

The findings suggest that 360° VR visualizations differ from traditional 2D images only in some dimensions of cognitive load.

At the same time, VR users felt less insecure, discouraged, irritated, stressed and annoyed by the task at hand. Such findings suggest that various visual aspects and the complexity of visual details produce a variety of impacts on cognitive load and task performance.

Visuals of hotel environments and services displayed in 360° VR produced stronger affective responses compared to 2D visuals, with a larger difference for women. The possible

explanation for the observed inconsistencies may lie in the fact that previous studies examined affective responses to VR in other contexts, such as gaming, food service, retail, etc. Gibson and O'Rawe [51] and Wan et al. [5] caution that the effects of RV may vary in different contexts; therefore, the ac-tual findings may point to peculiarities of the hotel context, which merit further investigation. Furthermore, the results might have been affected by weaknesses in self-reported measures when applied to affective responses. Self-reported measures are prone to cognitive biases and socially desirable responses and, more importantly, such measures are unable to capture respondents' unconscious affective reactions. It is also possible that, as explained, access to limited information through hotspot login or filtered by user profiles may have influenced attitudinal evaluation, and further studies with more complete information on 360° VRI are required.

The current study also produced evidence that VR images of hotel service layers may be more powerful in making users feel more positive about staying at the hotel and saying positive things about the hotel to other potential users compared to 2D images. The findings were not very consistent, indicating that further research is needed.

Regarding the challenges encountered, the use of 360° videos for immersive reality represents a great opportunity, given the low cost of display devices, which is why more empirical research is needed that can be used in other fields such as of hospitality, so that investments in this field get a good return on investment.

VI. Conclusion

An adaptive software architecture model based on hotel marketing concepts and strategies was proposed from 360° VRI videos and to validate the effectiveness of experiences with the model, 360° VRI visualizations were evaluated in comparison with 2D photographs. Since traditional 360° video immersion and 2D image viewing can trigger experiences in various dimensions, cognitive, affective, attitudinal, and behavioral loads were evaluated.

The results of the study suggest that 360° VR views differ from traditional 2D images only on a few dimensions of cognitive load. Also, 360° VR users felt less insecure, discouraged, irritated, stressed, and annoyed while performing viewing tasks, suggesting that visualizations and the complexity of visual details produce a variety of impacts on cognitive load and performance. task performance. Viewing hotel environments with 360° VR produced stronger affective responses compared to 2D visuals.

The study also found that 360°VR image displays of hotel environments and services may be more powerful in making users feel positive about booking the hotel and saying positive things to other potential users compared to 2D images. The findings were not very consistent, indicating that further research is needed to establish that investments in emerging technologies in hotel marketing are made based on a framework of adequate knowledge and experience.

Acknowledgment

References

[1] C. P. Ortet, A. I. Veloso and L. Vale Costa, *Cycling through 360 Virtual Reality Tourism for Senior Citizens: Empirical Analysis of an Assistive Technology.* Sensors, 22, 6169, 2022.

[2] A. Kittel, P. Larkin, I. Cunningham and M. Spittle, *Virtual Reality: A SWOT Analysis in Comparison to Virtual Reality.* Front. Psychol, 11, 563474, 2020.

[3] A. Sheikh, A. Brown, M. Evans and Z. Watson, *Directing attention in 360-degree video.* In Proceedings of the IBC 2016 Conference, Amsterdam, The Netherlands, 8–12 September, 2016.

[4] D.-Y. Wu and J.-H. T. Lin, *Ways of Seeing Matter: The Impact of a Naturally Mapped Perceptual System on the Persuasive Effects of Immersive Virtual Reality Advertising.* Commun. Res. Rep., 35, pp. 434–444, 2018

[5] C. Wan, S. Tsaur, Y. Chiu and W. Chiou, *Is the advertising effect of virtual experience better or contingent on different travel destinations?.* Information Technology and Tourism, 9(1), pp. 45-54, 2007.

[6] T. Vyvey, E. Nunez Castellar and J. Van Looy, *Loaded with fun? The impact of enjoyment and cognitive load on brand retention in digital games.* Journal of Interactive Advertising, 18(1), pp. 72-82, 2018.

[7] L. De Gauquier, M. Brengman, K. Willems and H. Van Kerrebroeck, *Leveraging advertising to a higher dimension: Experimental research on the impact of virtual reality on brand personality impressions.* Virtual Reality, 23, pp. 235-253, 2018.

[8] P.-H. Ting, S.-T. Wang, D.-Y. Bau and M.-L. Chiang, *Website Evaluation of the Top 100 Hotels Using Advanced Content Analysis and eMICA Model.* Cornell Hospitality Quarterly, 54(3), pp. 284–293, 2013.

[9] Simon Lei and Rob Law, *Functionality evaluation of mobile hotel websites in the m-commerce era.* Journal of Travel & Tourism Marketing, 36(6), pp. 665-678, 2019.

[10] Shanshan Qi, Rob Law and Dimitrios Buhalis, *Comparative evaluation study of the websites of China-based and international luxury hotels.* Journal of China Tourism Research, 13(1), pp. 1-25, 2017.

[11] S. Schmidt, A. S. Cantallops and C. P. Santos, *The characteristics of hotel websites and their implications for website effectiveness.* Tourism Management, 28(2), pp. 504–516, 2008.

[12] G. Agrawal, A. Dumka, M. Singh and A. Bijalwan, *Assessing Usability and Accessibility of Indian Tourism Websites for Visually Impaired Gaurav.* Hindawi Journal of Sensors, Article ID 4433013, 11 pages, 2022.

[13] T. Au Yeung and R. Law, *Evaluation of usability: A study of hotel websites in Hong Kong.* Journal of Hospitality & Tourism Research, 30(4), pp. 452–473, 2006.

[14] J. Qian and R. Law, *Examination of Website Language Strategies Adopted by Five-Star Hotels in China: A Corpus Approach.* J. Theor. Appl. Electron. Commer. Res., 16, pp. 1066–1078, 2021.

[15] Anas A. Salameh, Abdullah Al Mamun, Naeem Hayat and Mohd Helmi Ali, *Modelling the significance of website quality and online reviews to predict the intention and usage of online hotel booking platforms.* Heliyon, 8(9), e10735, 2022.

[16] G. Burdea and P. Coiffet, *Virtual Reality Technology.* John Wiley & Sons, pp. 464, 2003.

[17] J. Beck, M. Rainoldi and R. Egger, *Virtual reality in tourism: a state-of-the-art review.* Tourism Review, 74(3), pp. 586–612, 2019.

[18] K. Jung, V. Nguyen, D. Piscarac and S. Yoo, *Meet the Virtual Reju Dol Harubang—The Mixed VR/AR Application for Cultural Immersion in Korea's Main Heritage.* ISPRS Int. J. Geo-Inf, 9, 367, 2020.

[19] M. La Kim, C. Lee and T. Jung, *Exploring Consumer Behavior in VR Tourism Using an Extended Stimulus-Organism-Rresponse Model.* J. Travel Res., 59, 69, 2020.

[20] V. Ramos, M. Ruiz-Pérez and B. Alorda, *A Proposal for Assessing Digital Economy Spatial Readiness at Tourism Destinations.* Sustainability, 13, 11002, 2021.

[21] I. Casaló, C. Flavián and M. Guinalíu, *Determinants of the intention to participate in firm-hosted online travel communities and effects on consumer behavioral intentions.* Tour. Manag., 31, pp. 898–911, 2010.

[22] J. Fox, D. Arena, and J. N. Bailenson, *Virtual reality: A survival guide for the social scientist*. Journal of Media Psychology, 21(3), pp. 95–113, 2009.

[23] B. G. Witmer and M. J. Singer, *Measuring presence in virtual environments: A presence questionnaire*. Presence: Teleoperators and Virtual Environments, 7(3), pp. 225–240, 1998.

[24] J. Seo, S. Kandampully and N. Liu, *Virtual reality presence as a preamble of tourism experience: The role of mental imagery*. Tourism Management, 74, pp. 55-64, 2019.

[25] S. H. H. Shah, K. Han and J. W. Lee, *Real-Time Application for Generating Multiple Experiences from 360° Panoramic Video by Tracking Arbitrary Objects and Viewer's Orientations*. Applied Sciences, 10(7), pp. 2248, 2020.

[26] M. J. Benítez de Gracia and S. Herrera Damas, *El reportaje inmersivo en video 360 en los medios periodísticos españoles*. Revista de Comunicación, 17(2), 2018.

[27] T. Gnambs, C. Schreiner, M. Appel and T. Richter, *Experiencing narrative worlds: A latent state – trait análisis*. Personality and Individual Differences, 2014.

[28] R. J. Gerrig, *Experiencing narrative worlds: On the Pyschological Activities of reading*. New Haven: Yale University Press, 1993.

[29] J. Murray, *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. New York: The Free Press, pp. 324, 1997.

[30] M. C. Green and T. C. Brock, *The role of transportation in the persuasiveness of public narratives*. Journal of personality and social psychology, 79(5), pp. 701–721, 2000.

[31] P. Lelyveld, *Virtual Reality Primer with an Emphasis on Camera-Captured VR*. Motion Imaging Journal SMPTE, 124(6), pp. 78-85, 2015.

[32] R. Aronson-Rath, J. Milward, T. Owen and F. Pitt, *Virtual reality journalism*. 2015.

[33] W. A. IJsselsteijn, H. De Ridder, J. Freeman and S. E. Avons, *Presence: concept, determinants, and measurement*. In Human vision and electronic imaging V, International Society for Optics and Photonics, 3959, pp. 520-529, 2000.

[34] F. Oliveira, A. Santos, B. Aguiar and J. Sousa, *GameFoundry: Social Gaming Platform for Digital Marketing, User Profiling and Collective Behavior*. Procedia - Social and Behavioral Sciences, 148, pp. 58–66, 2014.

[35] T. Anteneh Alemu, A. K. Tegegne and A. Nega Tarekegn, *Recommender System in Tourism Using Case based Reasoning Approach*. International Journal of Information Engineering and Electronic Business, 9(5), pp. 34–43, 2017.

[36] D. I. Tapia, J. Bajo, J. M. Corchado, S. Rodríguez, J. F. De Paz, J. M. Sánchez and A. Saavedra, *Arquitectura Multiagente para Entornos Dinámicos: Tecnología e Inteligencia Aplicadas*. In Ubiquitous Computing and Ambient Intelligence, Congreso Español de Informática (CEDI-UCAMI 2007), Thompson-Paraninfo, Zaragoza, España, 2007.

[37] L. Alfaro, C. Rivera, E. Castãneda, J. Zũniga-Cueva, M. Rivera-Chavez and F. Fialho, *A review of intelligent tutorial systems in computer and web based education*. International Journal of Advanced Computer Science and Applications, 11(2), pp. 755–763, 2020.

[38] W. Lewis Johnson and J. C. Lester, *Pedagogical agents: Back to the future*. AI Magazine, 39(2), pp. 33–44, 2018.

[39] L. Alfaro, C. Rivera, J. Luna-Urquizo, E. Castañeda, J. Zuñiga, M. Rivera-Chavez and F. Fialho, *A review of 20 years of adaptive e-learning. In : challenges and opportunities of online learning*. In Ditte Kolbæk (Ed.), New York: Published by Nova Science Publishers, pp. 275-315, 2021.

[40] A. M. Ali. Hussein and H. K. M. Al-Chalabi, *Pedagogical Agents in an Adaptive E-learning System*. SAR Journal - Science and Research, pp. 24–30, 2020.

[41] V. Shute and B. Towle, *Adaptive E-Learning*. Educational Psychologist, 38(2), pp. 105–114, 2003.

[42] A. Grubišić, S. Stankov and B. Žitko, *Adaptive courseware: A literature review*. Journal of Universal Computer Science, 21(9), pp. 1168–1209, 2015.

[43] E. Triantafillou, A. Pomportsis and E. Georgiadou, *AES-CS: adaptive educational system based on cognitive styles*. In Proceedings of AH2002 Workshop, Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Malaga, Spain, pp. 10–21, 2002.

[44] V. Esichaikul, S. Lamnoi and C. Bechter, *Student modelling in adaptive e-learning systems*. Knowledge Management and E-Learning, 3(3), pp. 342–355, 2011.

[45] P. Krutchen, *Architectural Blueprints—The "4+1" View Model of Software Architecture*. IEEE Software, 12(6), pp. 42-50, 1995.

[46] A. J. Gonzalez, Lingli Xu and U. M. Gupta, *Validation techniques for case-based reasoning systems*. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 28(4), pp. 465–477, 1998.

[47] F. Paas, P. W. M. Van Gerven and H. K. Tabbers, *The Cognitive Aging Principle in Multimedia Learning*. In R. E. Mayer (Ed.), The Cambridge handbook of multimedia learning, Cambridge University Press, pp. 339–351, 2005.

[48] N. Lavie, *Perceptual load as a necessary condition for selective attention*. Journal of Experimental Psychology: Human Perception and Performance, 21(3), pp. 451–468, 1995.

[49] M. F. Macedonio, T. D. Parsons, R. A. Diguiseppe, B. A. Weiderhold and A. A. Rizzo, *Immersiveness and physiological arousal within panoramic video-based virtual reality*. Cyber Psychology & Behavior, 10(4), pp. 508-515, 2007.

[50] L. Herrewijn and K. Poels, *Poner las marcas en juego: cómo la dificultad del juego y las experiencias de los jugadores influyen en la efectividad de la publicidad en el juego*. Revista Internacional de Publicidad, 32(1), pp. 17-44, 2013.

[51] A. Gibson and M. O'Rawe, *Virtual reality as a promotional tool: Insights from a consumer travel fair*. In T. Jung and M. C. Tom Dieck (Eds.), Augmented reality and virtual reality: Empowering human, place and business, Springer, pp. 93-107, 2017.

[52] H. Van Kerrebroeck, M. Brengman and K. Willems, *Escaping the crowd: An experimental study on the impact of a virtual reality experience in a shopping mall*. Computers in Human Behavior, 77, pp. 437-450, 2017.

[53] Y. K. Choi and C. R. Taylor, *How do 3-dimensional images promote products on the internet?*. Journal of Business Research, 67(10), pp. 2164-2170, 2014.

[54] J. Stoyanova, P. Q. Brito, P. Georgieva and M. Milanova, *Comparison of consumer purchase intention between interactive and augmented reality shopping platforms through statistical analyses*. International Symposium on Innovations in Intelligent Systems and Applications, Madrid, Spain, pp. 1-8, 2015.

[55] A. Mehrabian and J. A. Russell, *An approach to environmental psychology*. USA: The Massachusetts Institute of Technology, pp. 216–217, 1974.

[56] N. K. Park and C. A. Farr, *The effects of lighting on consumers' emotions and behavioral intentions in a retail environment: A cross-cultural comparison*. Journal of International Design, 33(1), pp. 17-32, 2007.

[57] L. Slevitch and H. Oh, *Relationship between attribute performance and customer satisfaction: A new perspective*. International Journal of Hospitality Management, 29(4), pp. 559-569, 2010.

[58] E. Galy, M. Cariou and C. Mélan, *What is the relationship between mental workload factors and cognitive load types?*. International Journal of Psychophysiology, 83(3), pp. 269-275, 2012.

[59] Z. Kablan and M. Erden, *Instructional efficiency of integrated and separated text with animated resentations in computer-based science instruction*. Computers & Education, 51(2), pp. 660-668, 2008.

[60] C. Anthes, R. J. García-Hernandez, M. Wiedemann and D. Kranzlmüller, *State of the art of virtual reality technology*. IEEE Aerospace Conference, Yellowstone, MT, United States, 2016.

[61] L. Alfaro, C. Rivera, E. Suarez and A. Raposo, *Data analysis*. [Online] Available: https://drive.google.com/drive/folders/1BHEOR5Tq2zgWBascnGji9Am817Q41Tqy?usp=sharing

A Hybrid Genetic Algorithm for Service Caching and Task Offloading in Edge-Cloud Computing

Li Li[1], Yusheng Sun[2], and Bo Wang[3]*

Guangdong Polytechnic of Science and Technology,
No.351, Kehua Street, Tianhe District, Guangzhou, China, 510640[1]
Software Engineering, Zhengzhou University of Light Industry
No. 136, Kexue Road, High-tech Industrial Development Zone, Zhengzhou, China 45000[2,3]

Abstract—Edge-cloud computing is increasingly prevalent for Internet-of-Thing (IoT) service provisioning by combining both benefits of edge and cloud computing. In this paper, we aim to improve the user satisfaction and the resource efficiency by service caching and task offloading for edge-cloud computing. We propose a hybrid heuristic method to combine the global search ability of the genetic algorithm (GA) and heuristic local search ability, to improve the number of satisfied requests and the resource utilization. The proposed method encodes the service caching strategies into chromosomes, and evolves the population by GA. Given a caching strategy from a chromosome, our method exploits a dual-stage heuristic method for the task offloading. In the first stage, the dual-stage heuristic method pre-offloads tasks to the cloud, and offloads tasks whose requirements cannot be satisfied by the cloud to edge servers, aiming at satisfying as many tasks' requirements as possible. The second stage re-offloads tasks from the cloud to edge servers, to get the utmost out of limited edge resources. Experimental results demonstrate the competitive edges of the proposed method over multiple classical and state-of-the-art techniques. Compared with five existing scheduling algorithms, our method achieves 11.3% to 23.7% more accepted tasks and 1.86% to 18.9% higher resource utilization.

Keywords—*Cloud computing; edge computing; genetic algorithm; service caching; task offloading*

I. INTRODUCTION

Internet-of-Thing (IoT) devices have become ubiquitous in our lives. IoT services are becoming increasingly needed in both the number and variety, as revealed by the Cisco Annual Internet Report [1]. Eventhough some resources are equipped on various IoT devices, they are not enough for satisfying all requirements of users, due to the limited resource of a device.

To address the issue, mobile cloud computing (MCC) exploits the cloud computing with abundent computing resources to extend the service ability of IoT devices, by offloading some tasks to the cloud [2]. But the cloud has a poor network performance as it deliveries services over wide area networks (WANs), and thus, it usully cannot process delay-sensitive request tasks. Therefore, edge computing is proposed to address the problem, by placing some resources close to user devices [3] to provide services with low latency.

By combing both benefits of edge and cloud computing, edge-cloud computing (ECC) has begun to be used more and more in both industrial and academic [4]. While the cooperation between the edge and cloud computing needs to be further enhanced for improving the resource efficiency and user satisfaction [5]. The purpose can be achieved by a joint task offloading and service caching strategy for edge-cloud computing.

unfortunately, the joint task offloading and service caching problem (TOSCP) is NP-hard [6], and there is no available method providing the optimal solution within a reasonable time for a large-scale ECC system. There mainly two kinds of ways to solve the TOSCP for providing an accepted solution in polynomial time, heuristics and meta-heuristics. Heuristics uses local search strategies for local optimal solutions with a very few time consumption. On the contrary, meta-heuristics exploit global search strategies inspired by natural and social phenomena. Meta-heuristics usually have better performances, but require much more time costs, compared with heuristics. In this paper, we design a hybrid heuristic method by combining the genetic algorithm (GA), a representative meta-heuristic method, and a dual-stage heuristic search strategy which is designed for improving the cooperation between edge and cloud computing. The introducing of the heuristic method can improve both the quality of chromosomes and the convergence velocity for GA. Different from existing related works [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], the main advantages of our work include the awareness of the heterogeneity between edge and cloud resources, providing the joint solution of the service caching and task offloading, and the integration of the dual-stage heuristic into GA.

The rest of this paper is organized as follows. Section II formulate the TOSCP concerned in this paper. Section III presents our proposed hybrid heuristic method. Section IV evaluates our proposed method by simulated experiments. And finally, Section V concludes the paper.

II. PROBLEM STATEMENT

As shown in Fig. 1, in edge-cloud computing, the service provider uses multiple edge servers (ESs) and a cloud to provide various services for its users, where the cloud provide its resources in the form of virtual machine (VM). Due to a very limited resource, each ES can be deployed only a few services at a time. On the contrary, the cloud has seemingly infinite resources, and can be provide all services all the time. In general, an ES has local area network connections to users close to it. While the cloud provides its services over wide area network, e.g., Internet. Thus, the cloud has a much poorer network performance than ESs. Briefly, the cloud has abundant computing and storage resources but a high data transmission delay, but ESs have limited resources with low network delays.
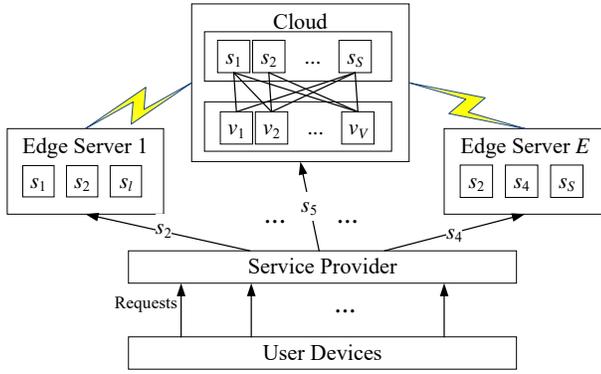
Fig. 1. The Architecture of the Edge-Cloud Computing System.

To deliver services efficiently, the service provider should carefully design the service caching and task offloading strategy, where the service caching decides which services each ES provides with its limited resources, and the task offloading strategy provide the computing node (ES or VM) where each user's request is processed. In this paper, we aim at designing a hybrid heuristic method for solving the TOSCP formulated as followings.

Given an edge-cloud computing system, there are E ESs, and the cloud provide V VMs. Without loss of generality, we use $E + V$ computing nodes, $n_1, ..., n_E$ and $n_{E+1}, ..., n_{E+V}$, to represent these ESs and VMs, respectively. For a computing node, say n_j, it has g_j computing capacity and b_j downlink bandwidth. For each ES, say n_j ($1 \le j \le E$), it has m_j storage capacity for service deployments.

The service provider provide S services, $s_1, ..., s_S$, by the edge-cloud computing system. The storage requirement of service s_k is r_k. There are T request tasks, $t_1, ..., t_T$, to be processed in the system. Each task require one of S services. We use the binary constants, $a_{i,k}$ ($1 \le i \le T, 1 \le k \le S$), to represent the service that each task requires, where $a_{i,k}$ represents that task t_i require service s_k. For its processing, task t_i needs c_i computing resource, and has h_i amount of input data to be processed. The deadline of t_i is d_i, i.e., t_i must be completed before d_i.

For the formulation of TOSCP, we define binary variables by Eq. (1) to indicate the solution, where $x_{i,j} = 1$ means that t_i is scheduled to n_j for its processing. $x_{i,j} = 0$ represents t_i is not scheduled to n_j.

$$x_{i,j} = \begin{cases} 1, & \text{if } t_i \text{ is scheduled to } n_j \text{ for its processing} \\ 0, & \text{else} \end{cases},$$
$$1 \le i \le T, 1 \le j \le E + V. \quad (1)$$

As no more than one computing node that each task can be processed,

$$\sum_{j=1}^{E+V} x_{i,j} \le 1, 1 \le i \le T. \quad (2)$$

In addition, we use binary variables $y_{j,k}$ to represent the service cache strategy for ESs, where $y_{j,k} = 1$ indicates service s_k is cached (deployed) on ES n_j ($1 \le j \le E$). Noticing that if a task is scheduled on an ES to be processed, its required

service must be cached (deployed) on the ES. Then, following equations hold.

$$y_{j,k} = \min\{1, \sum_{i=1}^{T} (x_{i,j} \cdot a_{i,k})\}, 1 \le j \le E, 1 \le k \le S. \quad (3)$$

For each ES, the total storage space of all services cached on it cannot be larger than its storage capacity, i.e.,

$$\sum_{k=1}^{S} (y_{j,k} \cdot r_k) \le m_j, 1 \le j \le E. \quad (4)$$

For tasks scheduled on a computing node, their finish time can be calculated using the pipeline execution model with the deadline decrease order [10], as shown by Eq. (5). where $ft_{i,j}$ is the finish time of t_i when it is scheduled on n_j, and $ft_{i,j} = 0$ if t_i is not scheduled on n_j. As the output data is usually much less than the input data for a task, this paper ignores the latency of the output data transfer.

$$ft_{i,j} = x_{i,j} \cdot \max\{ \max_{d_{ii} < d_i} ft_{ii,j} + \frac{c_i}{g_j},$$
$$\max_{d_{ii} < d_i} (ft_{ii,j} - \frac{h_{ii}}{b_j}) + \frac{h_i}{b_j} + \frac{c_i}{g_j}\},$$
$$1 \le i \le T, 1 \le j \le E + V. \quad (5)$$

Then, the deadline constraints can be formulated as Eq. (6).

$$ft_{i,j} \le d_i, 1 \le i \le T, 1 \le j \le E + V. \quad (6)$$

For each computing node, its usage time is the latest finish time of tasks scheduled on it, which is

$$ut_j = \max_{1 \le i \le T}\{ft_{i,j}\}, 1 \le j \le E + V, \quad (7)$$

and thus, the occupied resource of each node is

$$or_j = ut_j \cdot g_j, 1 \le j \le E + V. \quad (8)$$

For a computing node, the amount of resources used for processing tasks is the accumulated resources required by these tasks, i.e.,

$$ur_j = \sum_{i=1}^{T} (x_{i,j} \cdot c_i), 1 \le j \le E + V. \quad (9)$$

The resource utilization of each computing node and the overall resource utilization of the edge-cloud computing can be calculated by Eq. (10) and (11), respectively.

$$U_j = \frac{ur_j}{or_j}, 1 \le j \le E + V. \quad (10)$$

$$U = \frac{\sum_{j=1}^{E+V} ur_j}{\sum_{j=1}^{E+V} or_j}. \quad (11)$$

Based on the above formulations, TOSCP can be modelled as following optimization objective with constraints (1)–(11).

Maximizing $N + U$ \quad (12)

Where $N = \sum_{i=0}^{T} \sum_{j=1}^{E+V} x_{i,j}$ is the number of tasks processed by the edge-cloud computing system. Noticing that U is no more than 1, and thus, the major objective is maximizing the number of processed tasks, and the utilization maximization is the minor one. Due to the discreteness of decision variables, this optimization problem is hard to be solved exactly for large-scale problems, as its complexity is exponentially increased with the number of decision variables. Thus, we present a hybrid heuristic method for solving this problem with a reasonable time.

III. GENETIC ALGORITHM WITH DUAL-STAGE HEURISTIC

Our proposed method, DGA, exploits the global search ability of GA, where each chromosome represents a service caching solution for the edge-cloud computing, as shown in Algorithm 1. Meantime, DGA uses a dual-stage heuristic method, as shown in Algorithm 2, to solve the task offloading problem.

As shown in Algorithm 1, first, DGA initializes a population, a set of chromosomes, randomly (line 1). Then DGA repeats the evolution of the population using crossover, mutation, and selection operators (lines 5–7). After the maximum repeat time is reached, DGA provides a service caching and task offloading strategy by decoding the best chromosome with the best fitness. Where the fitness function used by DGA is the optimization objective (12), i.e., the finished task number plus the overall resource utilization.

Each chromosome corresponds to a service caching solution. The length of the chromosome, i.e., the number of genes, is the number of ESs. Genes have a one-to-one relationship with ESs, and the value of a gene indicates which services are cached on the corresponding ES. There are S services, then a gene is a binary string with length S, where binary bits have a one-to-one relationship with services to indicate whether services are cached on the ES (i.e., $y_{j,k}$ for ES n_k in Eq. 3). For example, in an edge-cloud computing, there are 5 services and 2 ESs. The chromosome (00111b, 11100b) indicates that the first three services are cached on the second ES, and the last three services are cached on the first ES.

To increase the rate of convergence, DGA see each gene value as an integer for the population evolution. To ensure the population diversity for large-scale systems, DGA uses uniform crossover, uniform mutation, and tournament selection operators to evolve the population.

Given a chromosome, DGA uses a heuristic method with a dual-stage to map the chromosome into a service caching and task offloading strategy, and calculates its fitness, as shown in Algorithm 2. There are two stages for a task. In the first stage, for each task, DGA finds whether a VM that can finish the task within its deadline. If there is such a VM, the task is pre-scheduled to the VM (lines 3–9). Otherwise, DGA searches whether there is an ES that caches the task's requested service and satisfies the task's deadline constraint. If there is such an ES, the task is scheduled to the ES (lines 10–18). Otherwise, the task cannot be processed by the edge-cloud computing for satisfying its requirements, and thus is rejected. After the first stage is completed, the next stage aims at improving the overall resource cost by re-scheduling some tasks from VMs

---

Algorithm 1 DGA: The Improved GA with a Dual-Stage Heuristic Search

Input: The information of tasks, services, ESs, and cloud VMs;
Output: A service caching and task offloading strategy;
1: Randomly initializing chromosomes;
2: **while** the maximum iteration number is not reached **do**
3:     For each chromosome, calculating its fitness value using Algorithm 2;
4:     Updating the best fitness and the best chromosome;
5:     Executing uniform crossover operator for chromosomes;
6:     Conducting uniform mutation operator on each chromosome;
7:     Selecting chromosomes by the tournament selection operator;
8:     Increasing the iteration number by one;
9: **end while**
10: **return** the service caching and task offloading strategy decoded from the best chromosome by Algorithm 2;

---

to ESs (lines 20–26). For each task pre-scheduled to VMs, the second stage looks up an ES that can satisfy requirements of the task. If there is such an ES, the task is re-scheduled to the ES. Otherwise, the task will be processed by the VM where it is pre-scheduled.

Algorithm 2 The Decoding by a Dual-Stage Heuristic Search

Input: A chromosome;
Output: The service caching and task offloading strategy, the fitness;
1: Decoding the chromosome into the service caching strategy;
2: **for** Each task **do**
3:     **for** Each VM **do**
4:         **if** The VM satisfies requirements of the task **then**
5:             Pre-scheduling the task to the VM;
6:             Accumulating the processed task (N) and the used resource amount (ur_j) based on Eq. (9);
7:             **break**;
8:         **end if**
9:     **end for**
10:     **if** The task isn't Pre-scheduled to a VM **then**
11:         **for** Each ES **do**
12:             **if** The ES satisfies requirements of the task **then**
13:                 Scheduling the task to the ES;
14:                 Accumulating the processed task and the used resource amount;
15:                 **break**;
16:             **end if**
17:         **end for**
18:     **end if**
19: **end for**
20: **for** Each task pre-scheduled to VMs **do**
21:     **for** Each ES **do**
22:         **if** The ES satisfies requirements of the task **then**
23:             Scheduling the task to the ES;
24:         **end if**
25:     **end for**
26: **end for**
27: Calculating the resource utilization (U) by Eq. (11);
28: Calculating the fitness: $N + U$;
29: **return** the service caching and task offloading strategy, the fitness;

IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of DGA based on a simulated edge-cloud system, to verify the high efficiency
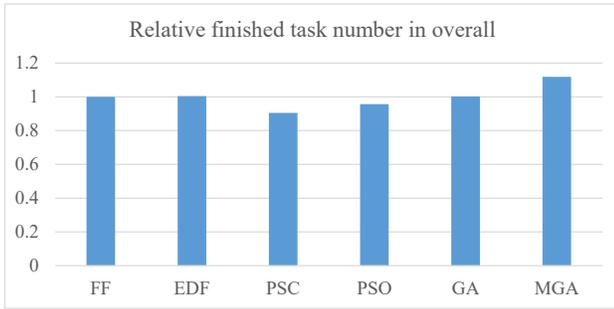
Fig. 2. The Normalized Numbers of Finished Tasks, Achieved by Various Methods.
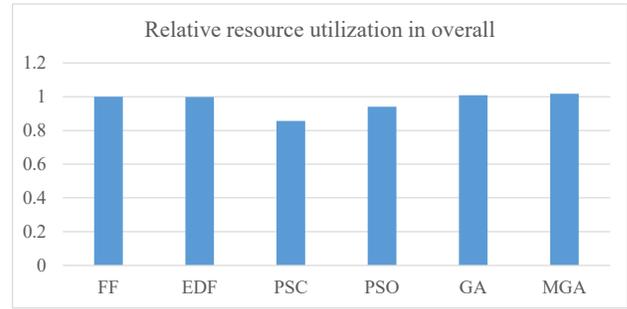
Fig. 3. The Normalized Resource Utilization Achieved by Various Methods.

of DGA in solving TOSCP.

A. Experiment Design

In the simulated edge-cloud system, we randomly generate 1000 tasks, 100 services, 10 ESs and one VM type. The computing resource required by a task is randomly set between 0.5GHz and 1.2GHz. The input data amount of each task is generated randomly in the range of [5MB, 6MB]. These two parameters are referring to [11]. The deadline is set to [1, 5] seconds for each task. Referring to [12], the storage space required by a service is in the range of [40, 80]GB, randomly. Each ES has 20GHz computing capacity. The network bandwidth between an ES and a user device is 60Mbps. The VM type is configured with 5GHz computing capacity and 15Mbps network bandwidth.

There are five benchmark methods in our experiments, First Fit (FF), Earliest Deadline First (EDF), Popularity-based Service Caching (PSC) [13], Particle Swarm Optimization (PSO) [5], and Genetic Algorithm (GA) [14]. The metrics used for the performance evaluation include the finished task number (the number of tasks whose requirements are satisfied) and the overall resource utilization, which used frequently to quantify the user satisfaction and the resource efficiency, respectively.

We repeat our experiment 100 times. In each repeat, we first generate an edge-cloud system with randomly set parameters. Then, we test the performance of our method and the five benchmark methods, respectively, in the generated system. To highlight the relative performance among these methods, we normalize each metric value of each method by dividing it by that of FF. In the following, we report the average normalized value for each metric and each method.

B. Experiment Results and Analysis

Fig. 2 shows the normalized numbers of tasks whose requirements are satisfied, when applying various methods. As shown in the figure, we can see that DGA can finish 11.3% to 23.7% more tasks than other methods. This verifies that our method has a better performance in optimizing the user satisfaction, compared with other method. This is benefited from the high population diversity of GA and the high efficiency of the dual-stage heuristic.

The normalized performances of various methods in resource utilization are shown in Fig. 3. From the figure, we
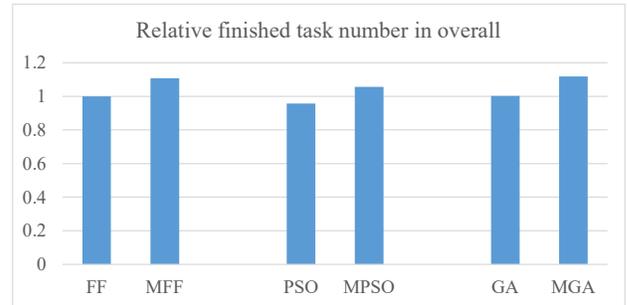
Fig. 4. The Improvement of the Dual-Stage Heuristic for Various Methods

can see that DGA achieves 1.86% to 18.9% high resource utilization than other methods. Thus, our method has a better resource efficiency.

One of the main advantages of our method is exploiting the dual-stage heuristic. The dual-stage heuristic can be applied for improving the performance of other methods, e.g., FF, PSO. Next, we conduct experiments for evaluating the improvement of the dual-stage heuristic for various methods, by comparing the performance differences between FF/PSO/GA and DFF/DPSO/DGA, where DFF/DPSO/DGA is the method of integrating FF/PSO/GA with the dual-stage heuristic. The results are shown in Fig. 4. DFF/DPSO/DGA can finish 10.8%/10.5%/11.6% more tasks than FF/PSO/GA. Thus the dual-stage heuristic can improve performance by about 10 percent for various methods.

V. CONCLUSIONS

We studied the service caching and task offloading problem for edge-cloud computing in this paper. We first formulated the problem as an binary non-linear programming with two optimization objectives, the finished task number and the resource utilization. Then, for solving the problem in polynomial time, we proposed a hybrid method which integrates GA and a dual-stage heuristic. Finally, we conducted extensive simulated experiments to evaluate the performance of our proposed method. Experiment results showed that our method can finish up to 23.7% more tasks and achieves up to 18.9% higher resource utilization, compared with five of classical and latest works.

In this paper, we focus on independent tasks without any data or logic dependency relationship. In the future, we will extent our work to support the workflow applications

consisting of interdependent tasks. In addition, we will exploit more global and local heuristic searching strategies to improve the performance of our method.

REFERENCES

[1] *Cisco Systems, Inc.*, Cisco annual internet report (2018–2023) white paper, https://www.cisco.com/c/en/us/solutions/executive-perspectives/annual-internet-report/index.html, March 2020.

[2] *X. Jin, W. Hua, Z. Wang, and Y. Chen*, A survey of research on computation offloading in mobile cloud computing, Wireless Networks, vol. **28**, 1563–1585, May 2022.

[3] *P. Cruz, N. Achir, and A. C. Viana*, On the edge of the deployment: A survey on multi-access edge computing, ACM Computing Surveys, Mar 2022, just Accepted.(DOI 10.1145/3529758)

[4] *B. Wang, C. Wang, W. Huang, Y. Song, and X. Qin*, A survey and taxonomy on task offloading for edge-cloud computing, IEEE Access, vol. **8**, pp. 186 080–186 101, 2020.

[5] *B. Wang, J. Cheng, J. Cao, C. Wang, and W. Huang*, Integer particle swarm optimization based task scheduling for device-edge-cloud cooperative computing to improve sla satisfaction, PeerJ Computer Science, vol. **8**, p. e893, 2022.

[6] *J. Xu, L. Chen, and P. Zhou*, Joint Service Caching and Task Offloading for Mobile Edge Computing in Dense Networks, IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, pp 207-215, **2018** (DOI 10.1109/INFOCOM.2018.8485977).

[7] *T. Gao, Q. Tang, J. Li, Y. Zhang, Y. Li and J. Zhang*, A Particle Swarm Optimization With Lévy Flight for Service Caching and Task Offloading in Edge-Cloud Computing, IEEE Access, vol. **10**, pp. 76636-76647, 2022.

[8] *K. Wang, W. Chen, J. Li, Y. Yang and L. Hanzo*, Joint Task Offloading and Caching for Massive MIMO-Aided Multi-Tier Computing Networks, IEEE Transactions on Communications, vol. **70**, no. 3, 1820-1833, 2022.

[9] *W. Fan, J. Han, Y. Su, X. Liu, F. Wu, B. Tang and Y. Liu*, Joint Task Offloading and Service Caching for Multi-Access Edge Computing in WiFi-Cellular Heterogeneous Networks, IEEE Transactions on Wireless Communications, vol. **21**, no. 11, pp. 9653-9667, 2022.

[10] *B. Wang, C. Wang, W. Huang, Y. Song, and X. Qin*, Security-aware task scheduling with deadline constraints on heterogeneous hybrid clouds, Journal of Parallel and Distributed Computing, vol. **153**, pp. 15-28, 2021.

[11] *Y. Dai, D. Xu, S. Maharjan, G. Qiao, and Y. Zhang*, Artificial intelligence empowered edge computing and caching for internet of vehicles, IEEE Wireless Communications, vol. **26**, no. 3: pp. 12–18, 2019.

[12] *G. Zhang, S. Zhang, W. Zhang, Z. Shen, and L. Wang*, Joint service caching, computation offloading and resource allocation in mobile edge computing systems, IEEE Transactions on Wireless Communications, vol. **20**, no. 8, pp. 5288–5300, 2021

[13] *X. Wei, J. Liu, Y. Wang, C. Tang, and Y. Hu*, Wireless edge caching based on content similarity in dynamic environments, Journal of Systems Architecture, vol. **115**, p. 102000, 2021

[14] *B. Wang, B. Lv, and Y. Song*, A hybrid genetic algorithm with integer coding for task offloading in edge-cloud cooperative computing, IAENG International Journal of Computer Science, vol. **49**, no. 2, pp. 503–510, 2022.

Aspect based Sentiment and Emotion Analysis with ROBERTa, LSTM

Uddagiri Sirisha
School of Computer Science
and Engineering
VIT-AP University, Amaravathi
Andhra Pradesh, India

Bolem Sai Chandana
School of Computer Science
and Engineering
VIT-AP University, Amaravathi
Andhra Pradesh, India

Abstract—Internet usage has increased social media over the past few years, significantly impacting public opinion on online social networks. Nowadays, these websites are considered the most appropriate place to express feelings and opinions. The popular social media site Twitter offers valuable insight into people's thoughts. Throughout the conflict between Russia and Ukraine, people from all over the world have expressed their opinions. In this study, "machine–learning" and "deep–learning" techniques are used to understand people's emotions and their views about this war are revealed. This study unveils a novel deep-learning approach that merges the best features of the sequence and transformer models while fixing their respective flaws. The model combines Roberta with ABSA (Aspect based sentiment analysis) and Long Short-Term Memory for sentiment analysis. A large dataset of geographically tagged tweets related to the Ukraine-Russia war was collected from Twitter. We analyzed this dataset using the Roberta-based sentiment model. In contrast, the Long Short-Term Memory model can effectively capture long-distance contextual semantics. The Robustly optimized BERT with ABSA approach maps words into a compact, meaningful word embedding space. The accuracy of the suggested hybrid model is 94.7%, which is higher than the accuracy of the state-of-the-art techniques.

Keywords—*Aspect based sentiment analysis, Twitter, LSTM, emotion analysis, Russia-Ukraine war, online social networks, Roberta model*

I. Introduction

People's voices are blaring out through social media in many communities. Modern society relies heavily on these platforms for sharing complete information and expressing opinions and feelings freely. Today, people spend a lot of time on social media websites [1] since the internet is so popular. Likewise, the popularity of social media platforms makes it very likely that people can positively or negatively impact society by sharing their opinions. Internet sources affect public opinion on politics, the economy, and social life. Most people are active on Twitter, sharing their views and feelings about a specific topic daily [2]. Millions of tweets are shared every day on Twitter. Sentiment analysis can mine vast amounts of text for information about people's views and responses [3,4]. People unable to express themselves in their daily lives for various reasons are more likely to express themselves on Twitter, a unique data source for sentiment analysis.

After annexing Crimea in 2014, Russia invaded Ukraine on February 24, 2022, turning the Ukraine-Russia crisis into war. Many deadly battles during the Russian invasion of

Ukraine sparked global reactions, especially on Twitter. Many journalists cover wartime events fairly and impartially [5]. Such conflicts cause long-term economic, political, and psychological issues in society. Social media is used worldwide to express political opinions. Online users are constrained to their social surroundings when examining contentious opinions. Obviously, one person cannot read all tweets about a topic. NLP analysis of tweets regarding the Ukraine-Russia war on Twitter will give us an objective view of global tendencies, providing a unique data source for press member stories and articles. Thus, a literature analysis on a Twitter debate topic may help machine and deep learning-based opinion categorization. Deep-learning-based methods automatically pick and learn features from the textual information, unlike machine learning-based methods [6]. Many degrees of deep learning view data differently. NLP approaches allow sentiment analysis and word cloud for visualization of massive Twitter data.

Deep-learning is used in various aspects such as object-detection, image-captioning [39], image segmentation [40], etc. This study uses machine and deep learning approaches to assess how twitter discussions regarding the Ukraine-Russia war affect public sentiment. Twitter API was used to evaluate the suggested approach on a large geographically tagged tweet dataset containing Ukraine-Russia war themes. Now sentiment analysis and word clouds help to discover general themes in the study. Sentiments about the Ukraine-Russia war and classify and display them into subgroups categories.

This study continues as follows: Section 2 offers a survey of relevant research. Section 3 describes the model's dataset, pre-processing, and ABSA-based Roberta-LSTM architecture for sentiment analysis. Section 4 compares the experimental results of the suggested model for the dataset with the sentiment visualisation data analysis. Finally, Section 5 concludes the study.

II. Related Work

Sentiment analysis uses data and text mining to identify textual sentiments [7]. Complex AI challenges let us determine whether a text has a favourable or bad subjective attitude. "Opinion mining" has three levels: "document", "sentence", and "aspect-based" [8–13]. Document-level sentiment analysis determines text polarity. It focuses on single-topic or entity papers. Sentence-level sentiment analysis examines positive, negative, and neutral sentences and their subjective or ob-

jective sentiment. Aspect-based sentiment analysis emphasizes "entity/object-identification", "feature-extraction", and "feature-polarity". Dictionary-based sentiment analysis models are common. A sentiment dictionary lists words having good or negative subjective orientations. In semantic orientation labelling, word types can be classed as positive or negative or given a numerical value depending on pre-established parameters. LIWC and HULIUO use common, context-free terms, while ANEW SenticNet and SentiWordNet [14] use sentiment strength and quantitative scores. Sentiment dictionaries are problematic because any human language has grammatical variances, idioms, slang, and spelling errors, which make automatic language analysis harder.

Machine learning, dictionary-based, and deep learning-based sentiment analysis are utilized nowadays [15,16]. A bag of words helps machine-learning-based approaches interpret texts into features. After that, "Naive-Bayes", "Decision-Trees", and "Support-vector-machine" classifiers are fed complex machine learning features [17]. Dictionary-based methods calculate text polarity by summing positive and negative emotion terms [18]. Machine learning-based systems may use sentiment dictionaries with positive and negative scores for words, unlike dictionary-based approaches. Thus, machine learning-based methods outperform dictionary-based ones. Hybrid machine and dictionary-based techniques coexist in the literature [19]. Deep learning-based approaches have replaced machine learning approaches, and experimental results are more promising [20, 21,37,38].

Researchers are impressed by deep-learning-based sentiment analysis. To enhance document-level sentiment analysis, the authors of this research [22] suggested a one-dimensional convolutional neural network model that incorporates temporal relations into user and product representations. The authors of [23] presented an artificial neural network-based system that suggests idioms for essay writing by analyzing similar contexts and potential phrases. In [24], the authors use feedback to improve the RNN architecture and create the LTSM model. Schuster and Paliwal utilized the "Bi-LSTM" model [25]. Deep learning methods automatically identify text emotion features. Deep learning predicts sentiment on a large Twitter dataset.

Internet use has made social media platforms popular in the past decade. Microblogs allow social media users to quickly share their daily experiences. Professionals share news and information on microblogs. Microblogs are the fastest way to report global news. As of 2021, Twitter has 330 million active users [26]. Thus, Twitter is vital for social media news dissemination. Sentiment analysis on Twitter can unlock a lot of unstructured data for researchers.

The authors of [27] sentiment analyzed tweets on coronavirus outbreaks using deep-learning classifiers. They suggested that harmful tweet content created fear, horror, and despair and that detecting such tweets would reduce the negative consequences of a pandemic on society. The authors of [28] used the BERT model to analyze attitudes by single and plural tags. The proposed method relied on emojis to convey emotions. The authors of [29] analyzed Twitter sentiment during and after the Syrian chemical attacks. The writers of [30] used the LDA algorithm to identify common debate themes from 4 million tweets regarding Covid-19 under 25 tags between "1 March 2020 and 21 April 2020". The writers of

[31] employed VADER natural language processing methods to assess coronavirus vaccine attitudes and behaviours using LSTM and Bi-LSTM-based and RNN-focused architectures. The authors of [32] employed topic modelling and sentiment analysis on tweets regarding global climate change to reveal public opinion. The authors of [36] used "IMDB", "Twitter-US-Airline-Sentiment", and "Sentiment140" datasets, and "Roberta-LSTM" outperforms state-of-the-art sentiment analysis algorithms.

The paper's contributions are threefold:

1) A hybrid ABSA-Roberta-LSTM model for sentiment analysis is proposed. Initially, aspect terms are generated, and then the Roberta model is used to tokenize words or subwords and build word embeddings. Meanwhile, the "LSTM" model encodes long-distance temporal dependencies in the word embeddings.

2) Due to the amount of Twitter data on the Ukraine-Russia war, a large spatially tagged dataset was collected from Twitter user's tweets about the war. This tweet collection was sentiment-analyzed using NLP.Twitter users often misspell and use contractions. Tweets may change tone due to spelling errors. Thus, our work performed powerful data cleansing, lemmatization, and noise removal.

3) Adjusting hyperparameters improves sentiment analysis outcomes. The empirical results reveal that ABSA-based Roberta-LSTM outperforms state-of-the-art approaches.

III. PROPOSED METHODOLOGY

This section describes the "ABSA-based-Roberta-LSTM" model for analysis. Pre-processing removes redundant tokens and symbols from the corpus. Our proposed model is then used to train and classify the dataset. The proposed methodology is shown in Fig. 1.

The "ABSA-based-Roberta -LSTM model" combines the Robustly optimised BERT method (Roberta) [33, 34] with "Long-Short-Term-Memory (LSTM)". The model efficiently maps tokens into meaningful embedding space using pre-trained Roberta weights.

A. Dataset Description

Sentiment analysis relies on data. Since the Twitter API allows real-time tweet extraction and sentiment analysis, research has expanded. For example, Twitter sentiment analysis helps to explain the Russia-Ukraine war by monitoring public opinion.

The study draws on a dataset obtained from Kaggle(https://www.kaggle.com/code/ssaisuryateja/eda-and-sentiment-analysis#EDA). The dataset is named "Ukraine Conflict Twitter Dataset". We collected the tweets from April and May of about 484221 tweets. Tweets were retrieved using hashtags. Hashtags can help with sentiment analysis, named entity recognition, and information extraction. Twitter hashtags included #ukraine, #russia,#Putin, #standwithUkraine, #Kyiv, #mariupol, #russian, #Ukrainewar, #nato, #standtogether and shown in Fig. 2.
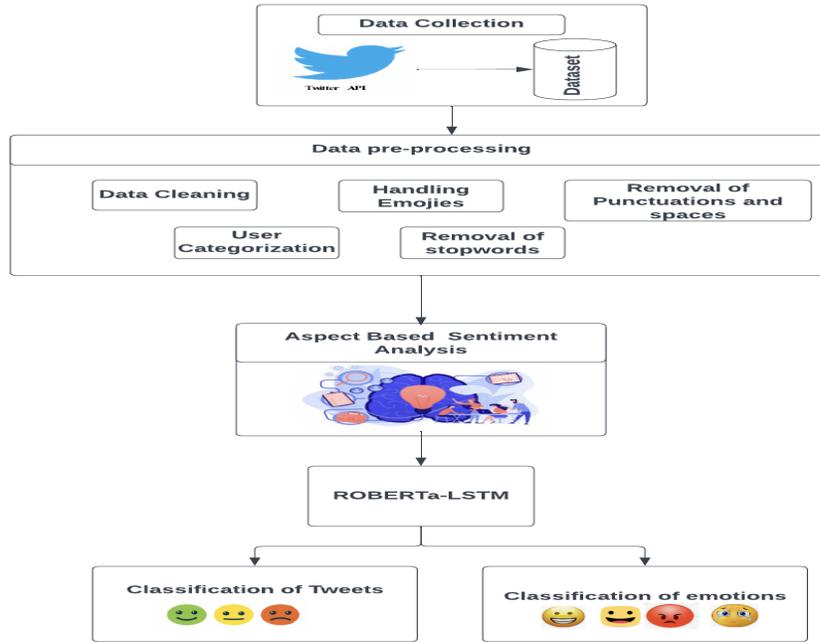
Fig. 1. Overview of Proposed Model.
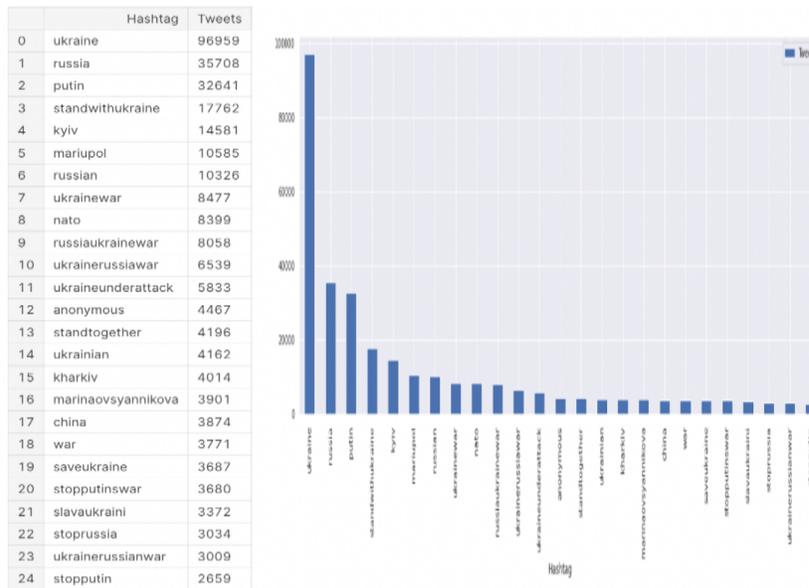
Fig. 2. Hashtag Analysis.

Overall, there were 17 distinct types of information included in the dataset, including Date, Hashtag, Followers, Friends, Location, Retweet, Username, Text, and Polarity.

B. Data Pre-processing

Data exploration helps uncover data patterns and trends. Check for missing or null values first. There are several steps involved in pre-processing data. Removal of Hashtags, removal of special symbols and punctuations, removal of stop words from data, removal of URL, stemming the words in the data, and deleting unwanted white space characters.

C. Aspect based Sentiment and Emotion Analysis

ABSA pinpoints the degree to which people's opinions differ about a specified topic.ABSA provides more detailed information than broad sentiment analysis. "Aspect Category Classification" determines an opinion text's subject and aspect pair. A dataset will receive a collection of present entities and one or more attributes based on the context of the text. Fig. 3 shows how ABSA Classification. In this example, general sentiment analysis yields "mixed" because there is one negative and one positive sentiment.

Fig. 3. ABSA based Sentiment and Emotion Analysis.

The ABSA aspect phrase "Zelensky" is positively received, whereas "Putin" is not. General sentiment analysis identifies only a tweet's predominant sentiment [34]. ABSA enables a more in-depth, sophisticated analysis by extracting sentiments from a single tweet. Since we wanted to examine how people felt about various topics across the time frame, we decided to use ABSA to glean further insight from each tweet. Using ABSA, a piece of more detailed information about the emotions in tweets is also analyzed. So in Fig. 3, the aspect term is "Woman", and the emotion of that woman illustrates "angry."

D. ROBERTa

Roberta extends Bidirectional Encoder Representation from Transformers (BERT). The Transformers [35] family developed BERT and Roberta for sequence-to-sequence modelling

to address long-range dependencies. The Transformer module implements the NLP model Twitter Roberta base-sentiment in Python. This "Roberta" based model has been pre-trained on 58M tweets and is optimised for sentiment analysis using the "TweetEval benchmark".

ABSA on pre-trained models functions extremely similarly to Roberta. The exception is that these models additionally get as input the aspect words of the sentences. Roberta tokenizes input sequences and sequence aspect terms. The suggested ABSA-based Roberta-LSTM model (Fig. 4) tokenizes the cleaned text into words or sub-words to simplify encoding word embeddings. For example, ¡s¿ is appended to the beginning and end of the text to indicate a sentence and vector length. To tokenize text at the byte level, the Roberta model uses the Byte-Pair Encoding tokenizer. The tokenizer doesn't split common words. Unusual terms are broken out.
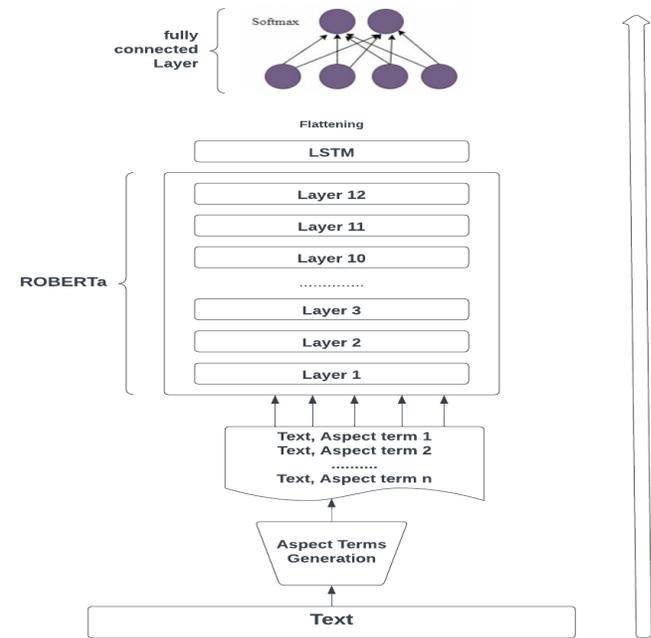
Fig. 4. ABSA based Roberta LSTM Model.

The model learns better when words are converted into numbers. Roberta uses input ids and an attention mask to tokenize text. Token indices and their numeric representations are input ids. The attention mask is not required if the sequence is to be batch processed. Tokens requiring attention are highlighted by the attention mask. The attention mask and ids are sent to the "ABSA based Roberta -LSTM" model. The "Roberta base model" consists of "12 base layers" and "125 million" parameters. The Roberta base layers aim to generate a meaningful word embedding as the feature representation, from which the more advanced layers can easily extract the required information.

E. LSTM

The results of the dropout layer are then fed into a model called "Long Short-Term Memory". Long short-term memory (LSTM) models retain data for the purpose of identifying input

dependencies that are far away in the future. LSTM has shown effective in a variety of sequence modelling applications, including text classification, sentiment analysis, time series prediction, and others.

Flatten Layer This layer flattens the input from 3D to 2D to fit inside the dense layer.

Dense Layer Dense layers are fully connected. Information is supplied to the dense layer's activation units from below. Roberta-LSTM has two layers. First dense layer hidden neurons capture the input-class relationship. Then, classification layer neurons match dataset classes. SoftMax activation function produces sentiment analysis dataset's probabilistic class distributions dataset's probabilistic class distribution in the classification layer.

F. Training Parameters

Huggingface transformer provides both Pytorch and TensorFlow implementation of the Roberta models. The "Adaptive Moment Estimation" optimizer optimizes gradient descent in "ABSA based Roberta-LSTM" model training. The "Adam optimizer" avoids local minima by using the gradient moving average, enhancing gradient descent. The Adam optimizer can deal with sparse gradients even when applied to noisy problems. The optimization approach uses the loss function to estimate the model loss in each training period. Since sentiment analysis involves multi-class, the loss function is categorical cross-entropy.

IV. EVALUATION METRICS AND EXPERIMENTAL RESULTS

A. Evaluation Metrics

Accuracy, Precision, Recall, and F1-score were used to evaluate the classification and comparison of models. In addition, train and test validation accuracy confirmed model performance. "Accuracy" refers to how well the forecast was made, and its calculation looks like this:

$$Accuracy = \left(\frac{\text{``}Correct predictions\text{''}}{\text{``}Total predictions\text{''}} \right) \qquad (1)$$

The positive and negative accuracy of the binary classification can be determined using the following formulas:

$$Accuracy = \left(\frac{\text{''}TP\text{''} + \text{''}TN\text{''}}{\text{''}TP\text{''} + \text{''}FN\text{''} + \text{''}FP\text{''} + \text{''}TN\text{''}} \right) \qquad (2)$$

"True positive", "true negative", "false positive", and "false negative" are denoted by the letters TP, TN, FP, and FN.

Precision is the percentage of clusters labelled positive that are positive. It's computed as

$$Precision = \left(\frac{\text{''}TP\text{''}}{\text{''}TP\text{''} + \text{''}FP\text{''}} \right) \qquad (3)$$

Recall, a measure of integrity, is the percentage of genuine positive predictions properly classified and calculated as

$$Recall = \left(\frac{\text{''}TP\text{''}}{\text{''}TP\text{''} + \text{''}FN\text{''}} \right) \qquad (4)$$

The lack of a balanced dataset may make the accuracy evaluation measure unreliable. The F1-score is employed. The F1-score gives target class outcomes. It considers classifier precision and recall in statistical classification study and is calculated as

$$F1 - score = (2 * Precision + Recall) \qquad (5)$$

B. Experimental Results

This section explains the experimental settings and compares the "ABSA-based Roberta-LSTM" model with other state-of-the-art approaches. The training session is limited to 100 iterations and terminated early to prevent overfitting.

According to this study, people's emotions toward the Ukraine-Russia war were mostly unfavourable. After preprocessing tweets, the tweets are passed to the "ABSA-based-Roberta-LSTM model", classified into three sentiment polarities: negative, positive, and neutral. Fig. 5 exhibits tweet sentiment polarity. Word Cloud, a sentiment analysis visualisation tool, was used to discover and show words and concepts for three sentiment studies to understand the psychological underpinnings for these reactions. Fig. 6 shows twitter dataset terms for "positive", "negative", and "neutral" attitudes.
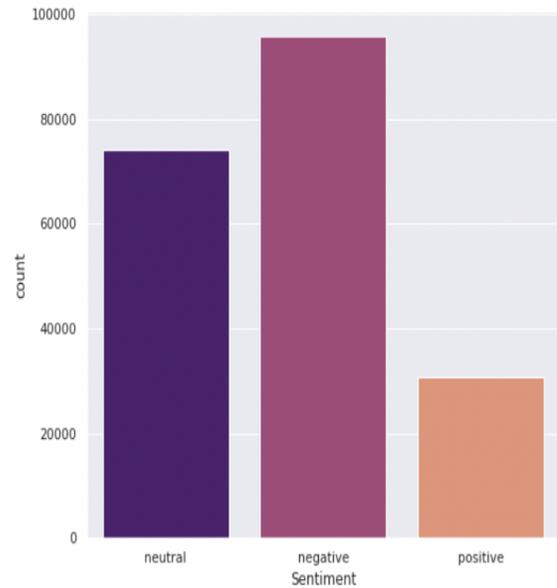
Fig. 5. The Percentages of Sentiment Polarities in the Ukraine-Russia Tweet Dataset.

Several "machine-learning" and "deep-learning" sentiment analysis approaches are tested for fairness (Table-I). The machine learning methods include "Bernoulli-distribution", "Logistic-Regression", "Decision-Tree", "K-nearest-neighbour", "Support-vector-classification" and "Random-Forest-Algorithm". The deep learning methods include "Long-Short-term-Memory" and "ABSA-based-Roberta-LSTM model". Machine learning algorithms and deep learning algorithms classify sentiments (positive,

negative, neutral) w.r.t. to precision, recall, F1score in Table I and a bar plot is plotted in Fig. 7.

Fig. 6. Word Cloud Representation of Prevalent Words in each Sentiment Polarity.

Fig. 7. Experiment Result Analysis on Ukraine-Russia Tweet Dataset (Sentiments).

TABLE I. EXPERIMENT RESULTS ON UKRAINE-RUSSIA TWEET DATASET(SENTIMENTS)

Classification Algorithms	Sentiment	Precision	Recall	F1:score
Bernoulli	NEGATIVE	0.94	0.92	0.93
	NEUTRAL	0.79	0.93	0.86
	POSITIVE	0.99	0.62	0.76
K Nearest Neighbours	NEGATIVE	1.00	0.71	0.83
	NEUTRAL	0.67	0.97	0.79
	POSITIVE	0.87	0.71	0.78
Decision Tree Classifier	NEGATIVE	0.92	0.90	0.91
	NEUTRAL	0.85	0.88	0.87
	POSITIVE	0.86	0.84	0.85
Random Forest Classifier	NEGATIVE	0.93	0.95	0.94
	NEUTRAL	0.88	0.91	0.90
	POSITIVE	0.96	0.83	0.89
Logistic Regression	NEGATIVE	0.95	0.96	0.95
	NEUTRAL	0.91	0.92	0.92
	POSITIVE	0.93	0.89	0.91
Support Vector Classification	NEGATIVE	0.95	0.96	0.96
	NEUTRAL	0.92	0.92	0.92
	POSITIVE	0.93	0.90	0.92
LSTM	NEGATIVE	0.94	0.95	0.95
	NEUTRAL	0.91	0.91	0.91
	POSITIVE	0.92	0.90	0.92
ABSA based Roberta+LSTM	NEGATIVE	0.96	0.96	0.97
(Our model)	NEUTRAL	0.93	0.93	0.93
	POSITIVE	0.94	0.92	0.94

Highly followed Twitter users can affect the audience with their emotions. The number of followers can be used to estimate outreach, but a user's interactions (mentions, responses, retweets, attributions) are a superior indicator of influence. Fig. 8 shows tweet count versus emotion analysis. According to Fig. 8, anger (1,00,445) is the biggest feeling on Twitter associated with the Ukraine-Russia war, followed by optimism (39,284), joy (25,091), and sadness (25,795). Fig. 9 shows Ukraine-Russia Tweet Dataset in terms of joy, optimism, sadness, and anger emotions.

In order to provide a reliable basis for comparison, the experiments make use of a variety of "machine-learning" and "deep-learning" approaches for emotional analysis, i.e. shown in Table II and Fig. 10.
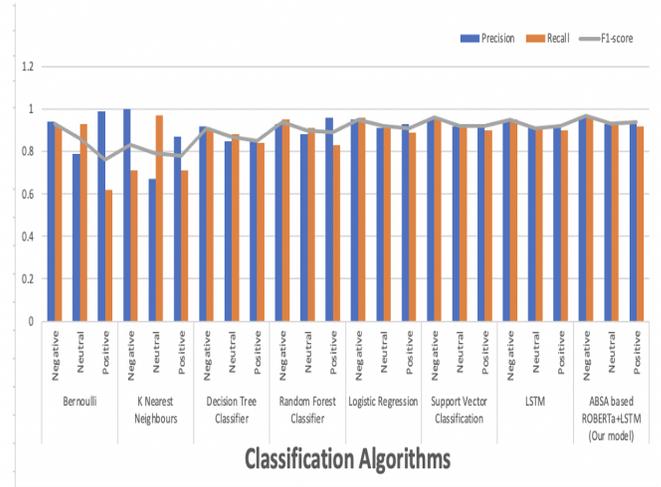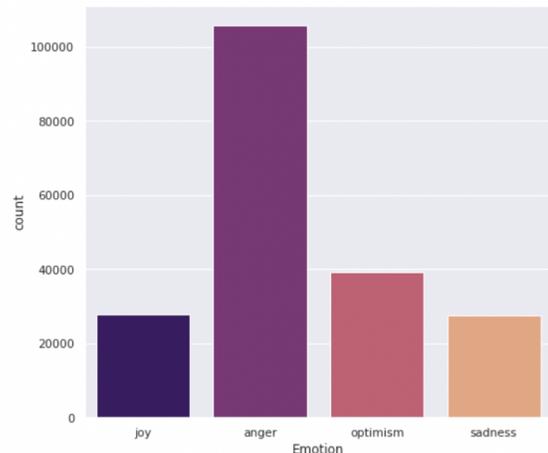
Fig. 8. The Percentages of Different Emotions in the Ukraine-Russia Tweet Dataset.

The Experiment results of the state-of-the-art approaches using the Ukraine-Russia Tweet Dataset in terms of accuracy (Sentiments) are illustrated in Table III, and a bar pot is plotted in Fig. 11.

V. CONCLUSION

Sentiment analysis plays a key role in various sectors, such as business and politics, to comprehend the public sentiments so that a strategic decision can be taken. Therefore, an effective algorithm is required to automatically determine the polarity (positive, negative or neutral) of the opinions. "Machine learning", "Deep learning", and "Recurrent neural networks" dominate sentiment analysis research. There are few works that employ transformers for sentiment analysis.

This paper provides an hybrid deep learning models for sentiment analysis i.e. "ABSA-based Roberta-LSTM" model.
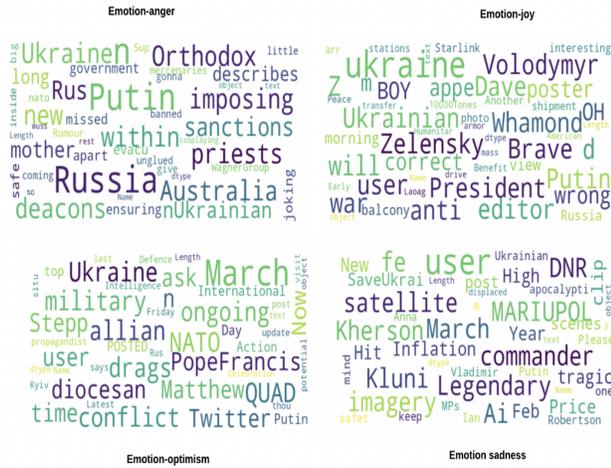
Fig. 9. Word Cloud Representation of Prevalent Words in each Emotion.

Fig. 10. Experiment Result Analysis on Ukraine-Russia Tweet Dataset (Emotions).

TABLE III. COMPARISON ANALYSIS OF VARIOUS ALGORITHMS USING UKRAINE-RUSSIA TWEET DATASET

Classification Algorithm	Precision	Recall	F1:score	Accuracy
Bernoulli	0.91	0.82	0.85	0.88
K Nearest Neighbours	0.84	0.80	0.80	0.81
Decision Tree	0.88	0.88	0.88	0.89
Random Forest Classifier	0.93	0.9	0.91	0.916
Logistic Regression	0.93	0.92	0.93	0.933
Support Vector Classification	0.94	0.93	0.93	0.937
LSTM	0.91	0.91	0.91	0.92
ABSA+Roberta+LSTM(Our model)	**0.93**	**0.93**	**0.94**	**0.947**

TABLE II. EXPERIMENT RESULTS ON UKRAINE-RUSSIA TWEET DATASET (EMOTIONS)

Classification Algorithms	Emotion	Precision	Recall	F1:Score
Bernoulli	Anger	0.78	0.99	0.88
	Joy	0.93	0.57	0.70
	Optimism	0.94	0.75	0.83
	sadness	0.98	0.69	0.81
K Nearest Neighbours	Anger	0.99	0.71	0.83
	Joy	0.36	0.98	0.53
	Optimism	0.97	0.69	0.80
	sadness	1.00	0.69	0.82
Decision Tree Classifier	Anger	0.92	0.92	0.92
	Joy	0.81	0.80	0.81
	Optimism	0.85	0.87	0.86
	sadness	0.86	0.84	0.85
Random Forest Classifier	Anger	0.89	0.99	0.94
	Joy	0.88	0.85	0.86
	Optimism	0.96	0.84	0.90
	sadness	0.99	0.81	0.89
Logisitic Regression	Anger	0.94	0.98	0.96
	Joy	0.89	0.88	0.88
	Optimism	0.92	0.89	0.91
	sadness	0.95	0.86	0.90
Support Vector Classification	Anger	0.95	0.98	0.96
	Joy	0.90	0.90	0.90
	Optimism	0.93	0.90	0.92
	sadness	0.94	0.89	0.91
LSTM	Anger	0.93	0.94	0.95
	Joy	0.91	0.91	0.91
	Optimism	0.92	0.90	0.92
	sadness	0.95	0.86	0.90
ABSA based Roberta+LSTM (Our model)	Anger	0.96	0.96	0.97
	Joy	0.93	0.93	0.93
	Optimism	0.92	0.89	0.91
	sadness	0.95	0.86	0.90

Fig. 11. Comparison Analysis of Various Algorithms w.r.t to Evaluation Metrics.

Text pre-processing normalizes and removes unimportant words. Our Proposed model is trained and analysed on the cleaned corpus. Our Proposed model uses aspect terms, Roberta, and LSTM to efficiently encode words into word embedding and capture long-distance dependencies. The "ABSA based Roberta-LSTM" model outperforms state-of-the-art sentiment analysis approaches on Ukraine-Russia twitter dataset with an accuracy of 94.7 when compared to the existing approaches.

REFERENCES

[1] Müngen, A. A., Aygün, İ., & Kaya M. (2020). Finding the Relationship Between News and Social Media Users' Emotions in the COVID-19 Process. Sakarya University Journal of Computer and Information Sciences, 3(3), 250-263.

[2] Choudhary, Nurendra, et al. "Twitter corpus of resource-scarce languages for sentiment analysis and multilingual emoji prediction." Proceedings of the 27th International Conference on Computational Linguistics. 2018. [

[3] Tam, S., Said, R. B., & Tanriöver, Ö. Ö. (2021). A ConvBiLSTM deep learning model-based approach for Twitter sentiment classification. IEEE Access, 9, 41283-41293.

[4] Bashir, S., Bano, S., Shueb, S., Gul, S., Mir, A. A., Ashraf, R., & Noor, N. (2021). Twitter chirps for Syrian people: Sentiment analysis of tweets related to Syria Chemical Attack. International Journal of Disaster Risk Reduction, 62, 102397.

[5] Graber, S. M. (2017). War of perception: a Habermasian discourse analysis of human shield newspaper reporting during the 2014 Gaza War. Critical Studies in Media Communication, 34(3), 293-307.

[6] Mehta, A., Parekh, Y., & Karamchandani, S. (2018). Performance evaluation of machine learning and deep learning techniques for sentiment analysis. In Information systems design and intelligent applications (pp. 463-471). Springer, Singapore.

[7] Farhadloo, M., & Rolland, E. (2016). Fundamentals of sentiment analysis and its applications. In Sentiment analysis and ontology engineering (pp. 1-24). Springer, Cham.

[8] Alshuwaier, Faisal, Ali Areshey, and Josiah Poon. "Applications and Enhancement of Document-Based Sentiment Analysis in Deep learning Methods: Systematic Literature Review." Intelligent Systems with Applications (2022): 200090. [

[9] Behdenna, Salima, Fatiha Barigou, and Ghalem Belalem. "Document level sentiment analysis: a survey." EAI Endorsed Transactions on Context-aware Systems and Applications 4.13 (2018): e2-e2.

[10] Aygün, İ., Kaya, B., & Kaya, M. (2021). Aspect based twitter sentiment analysis on vaccination and vaccine types in covid-19 pandemic with deep learning. IEEE Journal of Biomedical and Health Informatics, 26(5), 2360-2369.

[11] Liu, Haoyue, et al. "Aspect-based sentiment analysis: A survey of deep learning methods." IEEE Transactions on Computational Social Systems 7.6 (2020): 1358-1375 [

[12] Sun, Chi, Luyao Huang, and Xipeng Qiu. "Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence." arXiv preprint arXiv:1903.09588 (2019).

[13] CMeškelė, Donatas, and Flavius Frasincar. "ALDONAr: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model." Information Processing & Management 57.3 (2020): 102211.

[14] Abdi, A., Shamsuddin, S. M., Hasan, S., & Piran, J. (2019). Deep learning-based sentiment classification of evaluative text based on multi-feature fusion. Information Processing & Management, 56(4), 1245-1259.

[15] Hasan, Ali, et al. "Machine learning-based sentiment analysis for twitter accounts." Mathematical and Computational Applications 23.1 (2018): 11.

[16] Do, Hai Ha, et al. "Deep learning for aspect-based sentiment analysis: a comparative review." Expert systems with applications 118 (2019): 272-299

[17] Yadav, N., & Chatterjee, N. (2016, December). Text summarization using sentiment analysis for DUC data. In 2016 International Conference on Information Technology (ICIT) (pp. 229-234). IEEE.

[18] Zhao, H., Liu, Z., Yao, X., & Yang, Q. (2021). A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. Information Processing & Management, 58(5), 102656.

[19] Chen, N., & Wang, P. (2018, November). Advanced combined LSTM-CNN model for twitter sentiment analysis. In 2018 5th IEEE international conference on cloud computing and intelligence systems (CCIS) (pp. 684-687). IEEE.

[20] Abdi, Asad, et al. "Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion." Information Processing & Management 56.4 (2019): 1245-1259.

[21] Chen, T., Xu, R., He, Y., Xia, Y., & Wang, X. (2016). Learning user and product distributed representations using a sequence model for sentiment analysis. IEEE Computational Intelligence Magazine, 11(3), 34-44.

[22] Liu, Y., Liu, B., Shan, L., & Wang, X. (2018). Modelling context with neural networks for recommending idioms in essay writing. Neurocomputing, 275, 2287-2293.

[23] Tai, K. S., Socher, R., & Manning, C. D. (2015). Improved semantic representations from tree-structured long short-term memory networks. arXiv preprint arXiv:1503.00075.

[24] Meng, Wei, et al. "Aspect based sentiment analysis with feature enhanced attention CNN-BiLSTM." IEEE Access 7 (2019): 167240-167249.

[25] Hallac, I. R., Ay, B., & Aydin, G. (2021). User Representation Learning for Social Networks: An Empirical Study. Applied Sciences, 11(12), 5489.

[26] Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. (2020). Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. Applied Soft Computing, 97, 106754.

[27] Feng, Y., & Zhou, W. (2020). Is working from home the new norm? an observational study based on a large geo-tagged covid-19 twitter dataset. arXiv preprint arXiv:2006.08581.

[28] Li, Xin, et al. "Exploiting BERT for end-to-end aspect-based sentiment analysis." arXiv preprint arXiv:1910.00883 (2019).

[29] Bashir, S., Bano, S., Shueb, S., Gul, S., Mir, A. A., Ashraf, R., & Noor, N. (2021). Twitter chirps for Syrian people: Sentiment analysis of tweets related to Syria Chemical Attack. International Journal of Disaster Risk Reduction, 62, 102397.

[30] Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y., & Zhu, T. (2020). Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. Journal of medical Internet research, 22(11), e20550.

[31] Alam, K. N., Khan, M. S., Dhruba, A. R., Khan, M. M., Al-Amri, J. F., Masud, M., & Rawashdeh, M. (2021). Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data. Computational and Mathematical Methods in Medicine, 2021.

[32] Dahal, B., Kumar, S. A., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. Social network analysis and mining, 9(1), 1-20.

[33] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, arXiv:1907.11692

[34] Xu, H., Liu, B., Shu, L., Yu, P.S.: Bert post-training for review reading comprehen- sion and aspect-based sentiment analysis. arXiv preprint arXiv:1904.02232 (2019)

[35] Naseem, Usman, et al. "Transformer based deep intelligent contextual embedding for twitter sentiment analysis." Future Generation Computer Systems 113 (2020): 58-69.

[36] Tan, K. L., Lee, C. P., Anbananthen, K. S. M., & Lim, K. M. (2022). RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network. IEEE Access, 10, 21517-21525.

[37] Arava, Karuna, et al. "Sentiment Analysis using deep learning for use in recommendation systems of various public media applications." 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, 2022.

[38] Sindhura, S., et al. "Sentiment Analysis for Product Reviews Based on Weakly-Supervised Deep Embedding." 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2021.

[39] Sirisha, Uddagiri, and Bolem Sai Chandana. "Semantic interdisciplinary evaluation of image captioning models." Cogent Engineering 9.1 (2022): 2104333.

[40] Salini, Yalamanchili, and J. HariKiran. "Deepfakes on Retinal Images using GAN."

Contactless Surveillance for Preventing Wind-Borne Disease using Deep Learning Approach

Md Mania Ahmed Joy[1]
Department of CSE
East West University,
Dhaka, Bangladesh

Israt Jaben Bushra[2]
Department of CSE
East West University,
Dhaka, Bangladesh

Razoana Ayshee[3]
Department of CSE
East West University,
Dhaka, Bangladesh

Samira Hasan[4]
Department of CSE
East West University,
Dhaka, Bangladesh

Samia Binta Hassan[5]
Department of CSE
East West University,
Dhaka, Bangladesh

Md. Sawkat Ali[6]
Department of CSE
East West University,
Dhaka, Bangladesh

Prof.Dr. Omar Farrok[7]
Department of EEE
Ahsanullah University of
Science and Technology, Dhaka,Bangladesh

Mohammad Rifat Ahmmad Rashid[8]
Department of CSE
East West University,
Dhaka, Bangladesh

Maheen Islam[9]
Department of CSE
East West University,
Dhaka, Bangladesh

Abstract—**Covid-19 has been marked as a pandemic world-wide caused by the SARS-CoV-2 virus. Different studies are being conducted with a view to preventing and lessening the infections caused by covid-19. In future, many other wind-borne diseases may also appear and even emerge as "pandemic". To prevent this, various measures should be an integral part of our daily life such as wearing face masks. It is tough to manually ensure individuals safety. The goal of this paper is to automate the process of contactless surveillance so that substantial prevention can be ensured against all kinds of wind-borne diseases. For automating the process, real time analysis and object detection is a must for which deep learning is the most efficient approach. In this paper, a deep learning model is used to check if a person takes any preventive measures. In our experimental analysis, we considered real time face mask detection as a preventive measure. We proposed a new face mask detection dataset. The accuracy of detecting a face mask along with the identity of a person achieved accuracy of 99.5%. The proposed model decreases time consumption as no human intervention is needed to check an individual person. This model helps to decrease infection risk by using a contactless automation system.**

Keywords—*Computer vision; convolution neural network; COVID-19; deep learning; face mask detection; identity detection; object detection*

I. INTRODUCTION

Whenever the world faced any pandemic, history witnessed a pessimistic effect on economics, health, and national security both socially and globally [1]. Before the COVID-19, H1N1 or influenza was marked as a pandemic in the year 2009 [2] [3]. The COVID stands for "CoronaVirus Disease", it is referred to as "2019 novel coronavirus" or "2019-nCoV" as it was started in 2019 [4]. Form the beginning of this pandemic, around 228 countries have been affected by the COVID-19 to date [4]. The rapid growth of the COVID-19 infected cases has put the national healthcare capacity, modern ICU diagnostic methods, and public healthcare infrastructure to a test.

For instance, Bangladesh has less than 7,000 spaces in iso-lation units and therefore only 1622 health workers, including only 595 physicians to treat COVID-19 patients, whereas it has a population of about 165 million people [5]. On the other hand, the United States (US), which like Europe, spends about double as much per person on healthcare insurance as those other high-income counties, has drawn special criticism for its "maximum-possible-test-per-day" policy and use of its 96,596 ICU beds [5]. Overall, it became a chaotic situation which completely caught the whole world off guard. To prevent this disease, many effective vaccines have been invented [6] but these vaccines are not sufficient to give full protection to the people to prevent the COVID-19. So, people had to follow some non-therapeutic prohibitive measures like maintaining social distance, travel bans, remote office activities, country lockdown, wearing masks, etc. despite being vaccinated

Since 2019 almost all the nations of the world are strug-gling to get out of this misery. This kind of prohibition rules put a financial crisis as lockdown remains on all kinds of institutions [7]. In [8], according to the "Socio-Economic Aspects" about 60.5% of the respondents agreed that most of the low incoming people lost their jobs and around 54.8% of people also shifted to different places by leaving the city for livelihood. To ease the survival challenge of the people who lived from hand to mouth, the lockdown had to be lifted. Then there comes another challenge of ensuring that every individual wear masks in their working places. To ensure security in institutions or industries, many automation systems exist that can identify the registered person

One of them is a biometric system that works with measurements and by analyzing someone's unique features. Physical unique features like face, irises, veins, fingerprints, and behavioral characteristics such as voice, typing rhythm, or handwriting, is used for identification and authentication [9]. These systems are not contactless, and these require removing

the mask which can be one of the fatal reasons for any wind-borne disease [10]. For this, there is an increasing necessity to build a contactless surveillance system to detect facemask and person identification. So, there are a few systems that can perform both the operations. For designing such contact less automated systems, deep learning-based approaches are integrated.

The researchers tried to find out the best possible object detection model with the highest accuracy. Some of the models can even detect objects from video streams. In some cases, general convulation neural network (CNN) models are used; it cannot identify a specific object if a lot of background objects exist in a frame. Boundary box techniques and means Inter-section over Union (IoU) is mostly used for object detection.

The focus of our work is to develop an automated contactless preventive measurement system using a deep learning approach. More specifically, our system will identify if a person is wearing a mask or not, and at the same time detect the person's identity. For object detection task using face mask dataset, there are many models like CNN [12], NASNet [31], R-CNN [11], FAREC [27], CSPResNeXt-50 [23], etc., and the most recent model is YOLOv5 [19] which shows moderate inference speed and accuracy. The followings are the significant contributions of this paper:

- A new dataset is presented that consists of 177 pictures with 6 labels in total. There will be 97 photos of random people wearing masks and without masks, and 80 images of four persons, 20 images of each person wearing a mask. For each image, it contains a text file with information about the boundary boxes.

- A comparative analysis is performed using the new dataset using the YOLOv5 model. The YOLOv5 model is used for training the dataset. After 100 epochs with batch size of 8, the accuracy of 99.5% is obtained.

- This approach helps to identify any person along with detecting if he/she is wearing a mask or not in a real-life scenario. It works best for both images and video streams.

- A dedicated application is developed for contactless preventive measures using our proposed approach.

The rest of this paper is arranged in the following manner: related work that provides an overview of the relevant papers in condensed form, proposed work provides an initial overall view of the work process, materials and methods section provides an explanation of the preparation, preprocessing of the dataset, methodology, and training, result section provides the experimental analysis and conclusion provides the conclusive observations and future works.

II. RELATED WORK

There are usually two types of object detection methods based on the deep learning models: one-stage algorithms and two-stage algorithms. Two-stage methods such as R-CNN [11], Fast R-CNN [12], and Faster R-CNN [13] create bounding boxes of objects and then classify these objects. But these methods are comparatively slower and hard to implement

in real-time. So, for the efficient results, some one-stage algorithms have been developed such as You Only Look Once (YOLO) [14], RetinaNet [15], Single Shot Detector (SSD) [16] which are based on Anchor and ATSS (Adaptive Training Sample Selection) [17].

In the article [18], the authors used YOLOv5 to detect face masks, which is the most powerful object detection model at present. Their experiment has achieved a success rate of about 97.9% by using 7,959 images for training and testing. There are different forms of the YOLOv5 model. The YOLOv5x model is much heavier than the most used YOLOv5s [19].

Face Mask Recognition System has also been developed by using different versions of the YOLO algorithm [18] [20] [21]. In the article [20], the authors used improved YOLOv4 which is based on CSPDarkNet53. They used adaptive image scaling algorithms and PANet structure to get more information on the feature layer. The authors have shown a comparison between different models like YOLOv3, YOLOv4, SSD, and Faster R-CNN and have got the prominent outcome using their proposed model. Yet, there are some problems with feature extraction and false detection cases as well as low light performance has not been considered. In another the article [21], the authors have used YOLOv2 with ResNet-50 to detect medical face masks. They have used mean Intersection over Union (IoU) to estimate the best number of anchor boxes and have achieved 81% precision as a detector. For their model, they have used 1415 images as datasets. They proposed a detector model using YOLOv2 with Resnet-50 for extracting the feature and detecting different phases. Their model works better to detect medical face masks. They have also planned to use video-based deep learning models in the future. The article [22] presents the use of the YOLOv4 model which is the fourth generation of the YOLO model proposed by the authors of [23] in 2020. It is relatively faster and has better accuracy, and it is incredibly quick, easy to train, robust and also reliable. It produces promising results even for detecting small objects compared to the previous versions. It recognizes items from three types of input image/frame: unmasked faces, masked faces, and humans.

Face mask detection using YOLOv4 was introduced in [24]. As the YOLOv4 is built on the Darknet Framework and features an NVIDIA Graphics Processing Unit (GPU) with 16 gigabytes of RAM and a 2.30GHz Intel i5 CPU, its framework has created two models, one with three classes of the face veil, face concealed face, and face cover, and the other with only two classes of face cover and face cover. These models perform better for the video which creates a picture (or frame) with localized faces and a bounding box indicating whether the face is masked or not.

Multi-Stage CNN Architecture has been proposed in [25] for face mask detection. A pre-trained model that has been trained on a large dataset in order to achieve rapid generalization and reliable detection using RetinaFace [26], was selected as a single stage-1 model. Its' performance was evaluated to two other prominent face detection algorithms models, Dlib, and MTCNN [27] [28]. In stage 2, a CNN classifier was trained using three picture classification models: MobileNetV2 [29], DenseNet121 [30], and NASNet [31]. Recently, face identification along with face mask recognition using CNN approach was presented in [32]. To process a sequence of

TABLE I. COMPARISON OF RELATED WORK

Author/ref	Year	Approach/Model	Detection	Model Accuracy
K. Zhang [28]	2016	CNN	Face	95.4%
S. Sharma [27]	2016	FAREC	Face	96%
B. Zoph [31]	2018	NASNet	Scalable Image	96.2%
Kaihan [34]	2019	R-CNN (G-mask)	Face	95.97%
M. Loey [21]	2020	YOLO-v2 with ResNet-50	Face Mask	81%
A. Bochkovskiy [23]	2020	CSPResNeXt-50	Object	94.8%
G.Yang [18]	2020	YOLOV5	Face Mask	97.90%
K. Bhambani [22]	2021	YOLOv4	Real-time Face Mask	95%
S. Degadwala [24]	2021	Yolo-v4	Face Mask	96%
A. Chavda [25]	2021	Multi-Stage CNN	Face Mask	99.49%
M.S. Mazli Shahar [32]	2021	CNN, Procrustes Analysis	Face Mask, Face Identity	97.14%

pictures, the system uses the OpenCV computer vision library as well as a deep learning method for face mask identification. The receiver operating characteristic (ROC) and the calculated mean accuracy are used to estimate the system's performance (ACC).

For the face detection process, the training has been done by CNN model to recognize the owner's identity. The similarity identification approach utilizing the 68-landmark plotting and the face identification method were used for the process. To capture each frame of the video stream, a 50-neuron fully connected dense layer is flattened and fed from the output. It provides the probability for a person wearing a mask or not into 2-neuron layers in the convolutional neural network training. The CNN with Keras and Tensorflow is then implemented to train the face mask classifier and for face mask detection, the first 20 epochs, a model checkpoint trains and validates where using facial 68-landmarks is a technique that locates a human face in a picture and delivers values on the face Region of Interest (ROI) as dots. All the work was done with OpenCV, Keras and TensorFlow. Standard face recognition performance metrics were used for person identification at the time of using a face mask [33].

As it has been discussed earlier, some of the works as like [18], [20], [21] have used YOLO algorithms that function using IoU. The authors of [21] using YOLOv2 with ResNet-50 model achieved nearly 81% accuracy, in the paper [22] achieved nearly 95% accuracy. The authors of [18] achieved nearly 97.90% accuracy using YOLOv5 but all of these works only determined whether a person was wearing a mask or not. In [25], the authors have achieved more than 99% accuracy using Multi-Stage CNN. In the case of face detection, the authors of [27] and [28] have achieved up to 96% percent accuracy but they did not train their model for identifying people when they are using a face mask.

Table I presents a comparative analysis on all the related works that have been done. Although different types of object detection have been done so far, there is not a single work that matches with the objective of this paper.

III. EXPERIMENT

This section describes how the proposed methodology has been designed. At first, a dataset of images with specific formats has been generated and then classified into labels and classes. Furthermore, preprocessing on images has been done according to the need of the research model. Then the YOLO algorithm and how YOLOv5 incorporates this work is elaborated. In conclusion, all these steps lead to training that is the final step to get the desired outcome. Fig. 1 illustrates the proposed framework for using face mask dataset for predicting contactless surveillance having a machine learning backbone using deep learning.
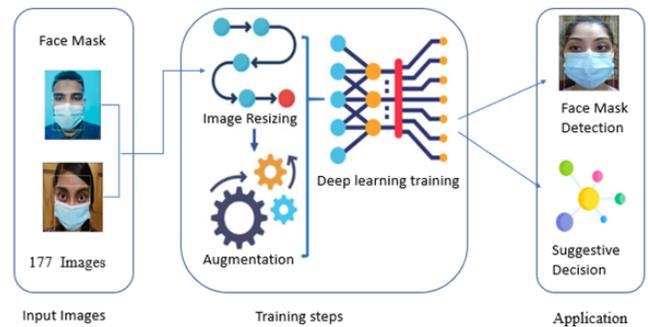
Fig. 1. Example of the Proposed Framework for Face Mask Detection and Contactless Surveillance System.

The YOLOv5 model is used to detect faces with masks and identification of that person who is wearing a mask. To train the model, at first, the images are collected. After preprocessing the images, every picture is labeled with the necessary class name. Then a dataset.YAML file is created. Then after training the model, it has been implemented on hardware. Finally, the performance has been evaluated.

A. Dataset

The images are in jpg, jpeg, png format, and the labels are in txt format. The images are of people of different ages, gender, wearing face masks and without masks. Fig. 2 shows a sample dataset with masked faces.

Fig. 2. Sample Dataset Consists of Images for a few Type Labels.

Table II shows the labels depending on the types of images and presents the number of images. A total of 177 images are used for this work and among them 97 images are of random people wearing a mask and without a mask, and among 177 images, 80 of them are of 4 people, 20 images of each person wearing a mask along with their identification. So, there are a total of 6 labels in this dataset. They are mask, no_mask, bushra, ayshee, joy, samira.

TABLE II. Details of the Images and their Labels

Types of Images	Labels	Number of Images
People wearing masks (including random people)	Mask	129
Random people wearing no mask (including random people)	No_mask	48
Person 1 wearing a mask with an identification	Bushra	20
Person 2 wearing a mask with an identification	Ayshee	20
Person 3 wearing a mask with an identification	Joy	20
Person 4 wearing a mask with an identification	Samira	20

When a dataset is created by collecting different pictures, all pictures cannot be of the same size. A model trains faster with smaller-sized images. If an image is three times as large, training the model will take six times as long, which will mount up over time. As all pictures vary in size, so if the picture is too small, it may appear unclear and if it is too big then the training time gets longer. That is why, all pictures have been resized. From the pictures, specific labeling is also needed to make classes which is the necessary step for training the model. Therefore, images are preprocessed before using the pictures as datasets.

At first, all images are resized to 1000×1000 pixels. Then all images are labeled by a software called "makesense.ai". In this software, all the labels are created first. Then on every image, the boundary box is drawn like Fig. 3 and then the labels are selected for the boundary boxes. A boundary box is drawn for each label and there are six labels used depending on the class. The class or label values are shown in Table III.

TABLE III. Class (Labels) and Values

Class (Labels)	Values
Bushra	0
Joy	1
Ayshee	2
Samira	3
Mask	4
no_mask	5

We organized boundary boxes in the following format: {Bx,By} = Coordinates of the boundary box's center

{Bh,Bw} = Width, height of the boundary box as a percent of the cell's width or height

Fig. 3. An Example of the Values from a Sample Image.

Fig. 3 shows the dot is the center of the boundary box. In Table IV, the boundary box values are shown for 1 and 4 number classes.

TABLE IV. Class (Labels) and Sample Bounding Box Values

Class	Bx	By	Bh	Bw
1	0.439	0.387	0.559	0.157
4	0.435	0.509	0.575	0.677

B. Deep Learning Framework for Training and Inference

To begin with, a dataset consisting of 177 images has been prepared where a total of 6 labels named mask, no_mask, bushra, ayshee, joy, samira are used. 97 images of random people wearing masks and without masks, and 80 images of 4 people, 20 images of each person wearing mask along with their identification are used to detect wearing mask and identify people.

After training the model, the images have split into training and validation labels, and validation prediction with confidence for the labels has been done. For training, Mosaic Data Loader [36] was used. Every training batch contains 8 mosaics, and each mosaic is a combination of four images.

The YOLOv5 [36] model works on a custom dataset that includes three cases i.e., 1st one to detect the identification of the person wearing a mask, 2nd one to detect face with a mask, and 3rd one to detect face without mask. The model is trained for predicting the label of classes for each input image.

YOLO stands for "You Only Look Once". Researcher Joseph Redmon and colleagues introduced YOLO in 2015 [35]. The General YOLO model works in three steps. In the

first step, it divides a picture into many cells with an m × m grid. Then in the second step, "s" bounding boxes are predicted in each cell. YOLO returns to the bounding boxes that exceed the minimum confidence threshold in the third step. The YOLO algorithm has various versions. In this paper, the YOLOv5 model is used. It is written in Python programming language's using Pytorch framework. This framework is user-friendly and a community driven github repository [36].
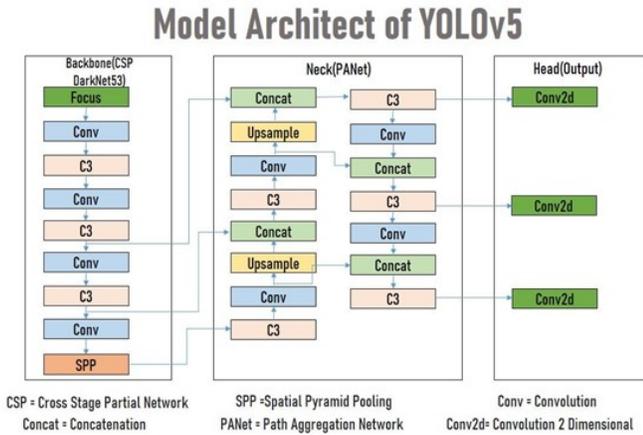
Fig. 4. YOLOv5 Model Architecture.

YOLOV5 consists of three parts: Backbone CSP darknet 53, Neck PANet, and Head Yolo V5 output layer depicted in Fig. 4. The backbone extracts feature from the inputs, Neck works to fusion features from backbone and Head generates the final results.

First part of the YOLOV5 is the Backbone CSP darknet or cross stage partial network. It focuses on the structure (Modified from Yolo V3) as well as on the CSP darknet 53 of Pytorch. CSP network has advantages like it solves repeated gradient problems, reduces model size, cuda memory layers, and increases forward, backward propagation. The CONV denotes convolution layers and C3 is composed of three convolution layers that are cascaded by various bottlenecks to find out the best fit in the network. And spatial pyramid pooling (Spp-Net) (Fig. 5) is a pooling layer that is used to extract features from the feature maps passed by previous layers. SPP works to generate fixed-size output for any type of input size. It is also used to extract necessary features by pooling that it's a multi-scale version of itself. Feature maps of inputs are duplicated to "n" numbers of versions. Using kernels of various sizes, each version is conducted by max pooling. This is how the SPP block extracts "n" number of various types of required and important features simultaneously [36].
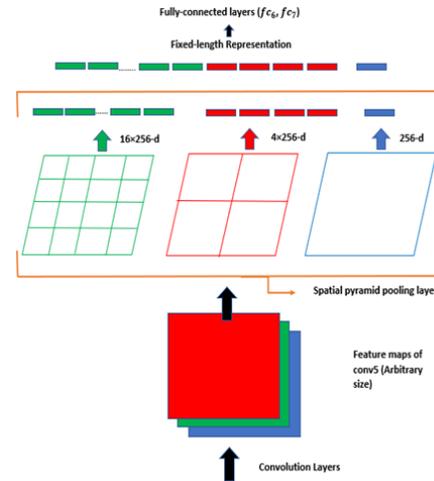
Fig. 5. Representation of SPP-NET.

Second part of this model is the Neck path aggregation network (PA-Net) [37]. It uses a modified version of feature pyramid network (FPN) structure that utilizes the information passed by the backbone. It employs bottom up (up sampling) augmentation path along with top-down It is used in FPN to directly connect with the fine-grained or learned features from low-level layers to top-level layers from backbone. For boosting information flow, prevention of missing information, and duplicate prediction, this concatenation operation is applied. It improves accuracy of the location of the object.

Last part is of the YOLOv5 model is the Head Yolo V5 output layer that uses information passed from the neck layer. The head of Yolo V5 known as Yolo layer generates three different sizes which are 18 × 18, 36 × 36, 72 × 72 of feature maps using convolutional 2-dimensional networks to achieve multi scale prediction. It enables the model to handle small, medium, and big objects and the output is a boundary box of the predicted object along with accuracy and class of the object.

C. Performance Measure

In this approach, Generalized Intersection Over Union (GIoU) loss has been used to determine the loss of the bounding box. The formula and algorithm for GIoU loss [38] is shown in Algorithm 1.

Algorithm 1 Generalized Intersection Over Union(GIoU)

Input: *Two arbitrary convex shapes*; $C, D \subseteq S \in R^n$
Output: *GIoU*
  1: *For C and D, find the smallest enclosing convex object E, where $E \subseteq S \in R^n$*
  2: $IoU = \frac{C \cap D}{C \cup D}$
  3: $GIoU = IoU - \frac{|E \backslash (C \cup D)|}{|E|}$

Here in algorithm 1, to compare similarity IoU (Intersection over Union) between two arbitrary convex shapes $C, D \subseteq S \in R^n$ is determined by, IoU = $\frac{|C \cap D|}{|C \cup D|}$. IoU has a big defect that if $|C \cap D| = 0$ and then IoU (C, D) =

0. This proves that IoU does not determine that both shapes are in proximity or not from each other. That is why GIoU (Generalized Intersection over Union) is used.

In GIoU, for both arbitrary shapes (volumes), $D \subseteq S \in R^n$, firstly the smallest arbitrary convex shapes are found $E \subseteq S \in R^n$ enclosing both C and D. E come from a similar type of geometric shape while comparing two geometric shapes. When both arbitrary ellipsoids are compared, E is smaller. They are encircled by ellipsoids. Then a ratio is computed between the area (volume) which is occupied by E subtracted from the area (volume) occupied by C and D and then divided by E's entire area (volume). This is a normalized metric that is focused on the vacant space area (volume) between C and D. Finally, GIoU has been determined. This ratio is subtracted from the value of IoU. Algorithm 1 summarizes the computations of GIoU.
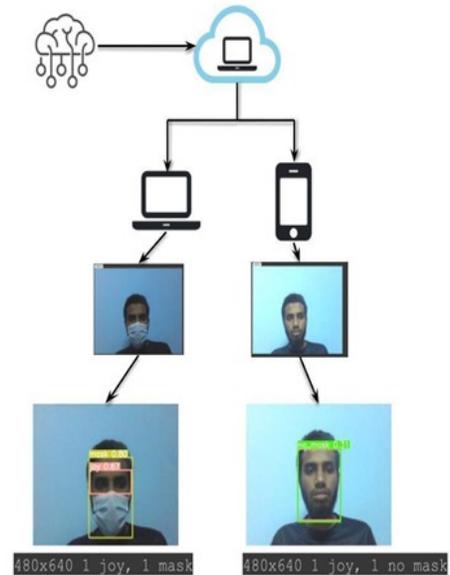
Fig. 6. Implementation Diagram.

IV. RESULTS AND EVALUATION

A. Experimental Results

The model has been implemented practically in an enclosed environment to see how it works in a real-life situation. Hardware implementation is needed to analyze the efficiency of the model. Implementation of this system provides a contactless process which ensures safety. It mitigates the spread of respiratory transmittable diseases. Implementation of this system will compel people to use masks because the system will mark people without face masks as any organization may use this system at their main entrance.

The model's dataset consists of 177 images and corresponding labels. For labeling, a total of six classes are used which are mask, no_mask, bushra, ayshee, joy and samira. A custom yaml file is created that contains the path of the dataset, number of classes and class names that are used. To train the dataset and yaml file, the YOLOv5x model is used. After training, two weights best.pt and last.pt have been obtained. YOLOv5x is a large model for processing and the weights were also large, so it needs a powerful processor. Hence, Google Colab, a cloud computing system is used to run the model. It has 12GB free graphics support and between GPU or CPU any of these can be used. Real time analysis has been performed by using the 12GB GPU of "Google Colab". For experimental purposes, the mobile and laptop devices have been used to run the model in "Google Colab". The configuration for mobile is f/2.45 aperture, SONY IMX471 sensor, and for laptop CMOS sensor technology, RGB HD Camera configuration was used. This experiment yields to a successful result to detect a person's identity wearing a face mask. Fig. 6 shows the implementation diagram of the model.

The proposed model has used GPU runtime for training the YOLOv5x model using Pytorch libraries and Stochastic Gradient Descent (SGD) optimizer. Batch size of 8 is used for each epoch and there are a total of 50 epochs for training the model, thus the experimental configuration has been obtained. The optimizer that is implemented for training the model is SGD.

SGD is used because the training time is 4 seconds which is faster than any other optimizers. SGD can be implemented on the convex problem and simple linear problems and the learning rate is determined by, $\eta/(1 + \lambda_0 t)$. Where, $\lambda_0 =$ regularization constant and η_0 is achieved by conducting initial experiments on any subsample of a dataset [39]. When the epoch increases with time, every epoch's precision (P), Recall (R), and Mean Average Precision (mAP) increases, and the training and validation of the model is also improved.

TABLE V. TRAINING AND VALIDATION RESULT BASED ON SGD

Epoch	Precision (P)	Recall (R)	mAP@0.5	mAP@[0.5:0.95]
5	0.197	0.67	0.304	0.117
10	0.711	0.718	0.67	0.386
20	0.829	0.904	0.95	0.572
30	0.816	0.963	0.949	0.528
48	0.987	0.995	0.995	0.637

In Table V, it is visible that the epoch has a proportional relation with the precision and recall, that is when the epoch is increasing the precision and recall also gets improved. For detecting the identity of a person wearing a mask, "GIoU Loss" has been used. The test results of the model are shown in Fig. 7.
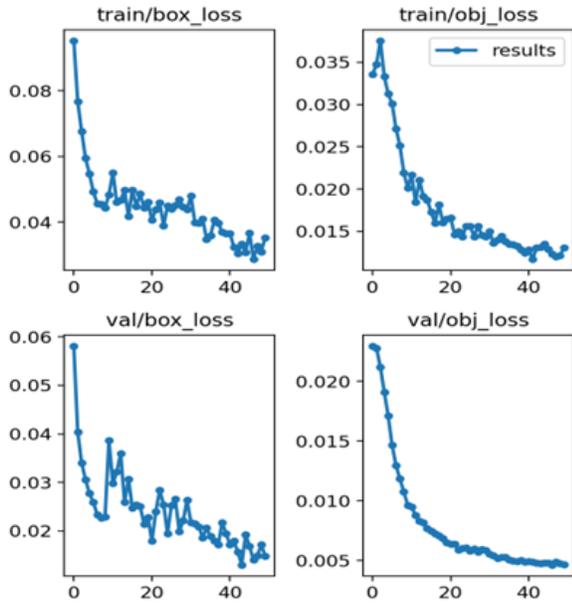
Fig. 7. The GIoU Loss Curve of Training where X Axis Represents Epoch.

After completion of the model training, the trained weights have been used for testing the model and then the performance is evaluated. The test results have classified into three types of categories.

True Positive (TP) indicates that in the test set the categories are equal to the results of the test. False Positive (FP) indicates that the sample number in the category of detected object is inconsistent. False Negative (FN) means that the resultant pattern has been detected as the inverse result or in the category of undetected. Precision refers to the ratio of TP to (TP+FP) where (TP+FP) is the number of all positive cases. Precision illustrates the sample proportions of the real positive cases from the samples detected by the model. The equation-1 shows the equation for precision.

$$Precision = \frac{TP}{TP+FP};\qquad(1)$$

Recall refers to the ratio of TP to (TP+FN) where (TP+FN) is the number of all positive test cases. The model's efficiency can be measured by recall which detects real positive cases of the test set out of the all-actual positive values. The equation that follows equation-2 shows the equation for recall.

$$Recall = \frac{TP}{TP+FN};\qquad(2)$$

Fig. 8 shows the graphs based on the precision and recall of the applied model.
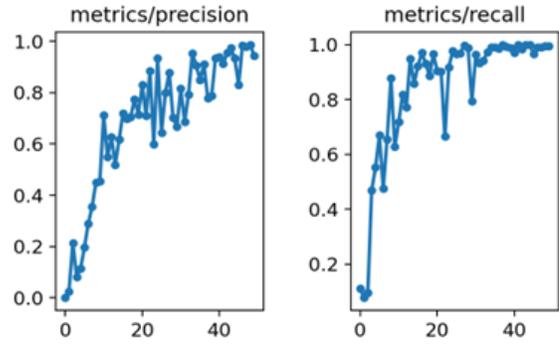
Fig. 8. Precision and Recall of the Model where X Axis Represents Epoch.

Then, to evaluate the accuracy of the model Mean Average Precision (mAP) is used. The following equation-3 shows hoe Map is determined:

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N};\qquad(3)$$

The performance evaluation from the training is shown in Table VI.

TABLE VI. PERFORMANCE EVALUATION

Class	Precision	Recall	mAP for IoU 0.5	mAP for IoU [0.5:0.95]
All	0.944	0.994	0.995	0.615
Bushra	0.719	1	0.995	0.331
Joy	0.979	1	0.995	0.639
Ayshee	1	1	0.995	0.48
Samira	0.975	1	0.995	0.733
mask	1	0.966	0.995	0.772
no_mask	0.99	1	0.995	0.736

The graphs in Fig. 9 presents the performance evaluation of the model.
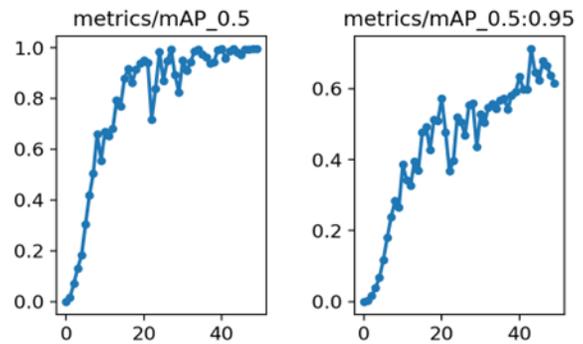
Fig. 9. mAP@0.5 and mAP@[0.5: 0.95] of the Model where X Axis Represents Epoch.

V. LIMITATIONS AND FUTURE WORK

This research includes the identification of individuals with face masks using the deep learning approach. The YOLOv5 model can detect whether any individual is maintaining the health protocols such as wearing a face mask or not. This

model is proven beneficial for preventing any airborne disease. Though it achieved the goal of the research, there are some limitations of the research. The dataset used for the model's implementation, contained only 177 images. Such small dataset does not ensure a significant accuracy if this research is applied for real-life events. Due to time constraints and inappropriate circumstances, it was not possible to conduct the research on a larger scale. A better accuracy could be expected from the model if a dataset containing more variety was used. Another limitation of the research is that the model is used only in detecting face masks, hence the user case of this model was confined into a specific object.

This research can be extended by overcoming the limitations and adding some insights to it. This model can be used while designing any other contactless surveillance systems. This model worked efficiently in face mask detection, but other object detection can also be included by implementing this model. Such systems significantly help to build any other contactless surveillance system that can highly be effective on controlling any contagious diseases as this ensures the social distancing. Another addition might be considered in this research that is, collecting enormous real-time data and implementing the model into that. Thus, ensuring the social distancing and detecting any kind of object with this model can be used to build further contactless system which can tend to prevent wind borne diseases. Enabling the unique biometric identification with the object detection system may prove as an enhanced version of this research.Such model can be utilized not only as contactless surveillance system, but also can keep the entrance record of any organization.

VI. CONCLUSION

This paper depicted the application of contact-less surveillance systems using a deep learning approach. For experimental analysis, we use YOLOv5 for object detection over a new dataset as mask detection. The ability of the YOLOv5 model to perform an efficient detection task was investigated in this research. The results demonstrate that YOLOv5 can achieve proficiency in detecting and spotting an item in a very short amount of time. In the case of surveillance of any organization at their entrance, the system not only identifies the person but also checks if he/she is wearing a mask or not. A masked face and a non-masked face dataset with no biased pictures have been developed and for the purpose of identification, the eye characteristics was examined. This system is unique in terms of operating in three different classes at the same time: a person with mask, without a mask, and the identification of a person. As the system identifies the person, it will be quite easier to detect if this specific person works in the organization or not. Using the system, if any individuals without masks are identified, they will be alerted and made conscious about wearing masks which is a preventive way to decrease COVID 19. The system also compels people to abide by the protocols of preventing this pandemic as well as helps to maintain hygiene for any wind-borne diseases. If the number of infected people decreases, the economic loss will also decrease. Manual identification of each people and checking if they are wearing masks or not is a difficult and time-consuming process. This process has been made easier and time-saving by this system. As it is a contact-less system, so social distancing can be highly maintained. This is a very useful technique for implementing in open spaces as well. An accuracy of 99.5% is achieved from this research. Some other models also have been implemented for this task but neither the accuracy was so high, nor the model could identify the features correctly in real-life scenarios whereas our chosen model successfully did the detections. Moreover, this trained model can perfectly identify features from the video streams along with the still images. This work will help to minimize the risk of air-borne viruses' spread during the time of the pandemic. Even if the COVID-19 pandemic fades away and life gets back to normal, this system can be quickly deployed to optimize the appropriate use of face masks in the workplaces so that a safe and secure work environment can be built.

REFERENCES

[1] W. Qiu, S. Rutherford, A. Mao and C. Chu, "The Pandemic and its Impacts", Health, Culture and Society, 9, pp. 1-11. 2017. DOI: 10.5195/hcs.2017.221

[2] J. Watkins, "Preventing a covid-19 pandemic," BMJ, vol. no. 368, pp. 1–2, February 2020, doi: 10.1136/bmj.m810.

[3] D. M. Morens, G. K. Folkers, and A. S. Fauci, "What is a pandemic?," The Journal of infectious diseases, vol. 200, no. 7, pp. 1018–1021, 2009, doi: 10.1086/644537.

[4] WHO, IFRC, and unicef, "Key Messages and Actions for Prevention and Control in Schools," Key Messag. Actions COVID-19 Prev. Control Sch., no. March, p. 13, 2020, [Online]. Available: https://www.who.int/docs/default-source/coronaviruse.

[5] Khan JR, Awan N, Islam MM, Muurlink O. Healthcare capacity, health expenditure, and civil society as predictors of COVID-19 case fatalities: a global analysis. Frontiers in public health. 2020 Jul 3;8:347.

[6] E. Ong, M. U. Wong, A. Huffman, and Y. He, "COVID-19 Coronavirus Vaccine Design Using Reverse Vaccinology and Machine Learning," Front. Immunol., vol. 11, p. 1581, 2020, doi: 10.3389/fimmu.2020.01581.

[7] S. Gunay, "COVID-19 Pandemic Versus Global Financial Crisis: Evidence from Currency Market," SSRN Electron. J., pp. 1–15, 2020, doi: 10.2139/ssrn.3584249.

[8] M. Bodrud-Doza, M. Shammi, L. Bahlman, A. R. M. T. Islam, and M. M. Rahman, "Psychosocial and Socio-Economic Crisis in Bangladesh Due to COVID-19 Pandemic: A Perception-Based Assessment," Front. Public Heal., vol. 8, no. June 2020, doi: 10.3389/fpubh.2020.00341.

[9] P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki, "An Introduction to evaluating biometric systems," Computer (Long. Beach. Calif)., vol. 33, no. 2, pp. 56–63, 2000, doi: 10.1109/2.820040.

[10] M. G. Garner, G. D. Hess, and X. Yang, "An integrated modelling approach to assess the risk of wind-borne spread of foot-and-mouth disease virus from infected premises," Environ. Model. Assess., vol. 11, no. 3, pp. 195–207, 2006. doi: 10.1007/s10666-005-9023-5

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

[12] R. Girshick, "Fast R-CNN," Proc. IEEE Int. Conf. Comput. Vis., vol. 2015 Inter, pp. 1440–1448, 2015, doi: 10.1109/ICCV.2015.169.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., , pp. 779–788, december 2016, doi: 10.1109/CVPR.2016.91.

[15] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," Proc. IEEE Int. Conf. Comput. Vis., pp. 2999–3007, October 2017, doi: 10.1109/ICCV.2017.324.

[16] W. Liu et al., "SSD: Single shot multibox detector," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes

Bioinformatics), vol. 9905 LNCS, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0_2.

[17] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 9756–9765, 2020, doi: 10.1109/CVPR42600.2020.00978.

[18] G. Yang et al., "Face Mask Recognition System with YOLOV5 Based on Image Recognition," 2020 IEEE 6th Int. Conf. Comput. Commun. ICCC 2020, vol. 1, no. January 2020, pp. 1398–1404, 2020, doi: 10.1109/ICCC51575.2020.9345042.

[19] . Glenn Jocher, Alex Stoken, Jirka Borovec, NanoCode012, Christopher-STAN, Liu Changyu, Laughing, tkianai, Adam Hogan, lorenzomammana, yxNONG, AlexWang1900, Laurentiu Diaconu, Marc, wanghaoyang0106, ml5ah, Doug, Francisco Ingham, Frederik, … Prashant Rai. (2020). ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements (v3.1). Zenodo. https://doi.org/10.5281/zenodo.4154370

[20] J. Yu and W. Zhang, "Face mask wearing detection algorithm based on improved yolo-v4," Sensors, vol. 21, no. 9, 2021, doi: 10.3390/s21093263.

[21] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," Sustain. Cities Soc., vol. 65, p. 102600, 2021, doi: 10.1016/j.scs.2020.102600.

[22] K. Bhambani, T. Jain, and K. A. Sultanpure, "Real-Time Face Mask and Social Distancing Violation Detection System using YOLO," Proc. B-HTC 2020 - 1st IEEE Bangalore Humanit. Technol. Conf., 2020, doi: 10.1109/B-HTC50970.2020.9297902.

[23] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020, [Online]. Available: http://arxiv.org/abs/2004.10934.

[24] S. Degadwala, D. Vyas, U. Chakraborty, A. R. Dider, and H. Biswas, "Yolo-v4 Deep Learning Model for Medical Face Mask Detection," Proc. - Int. Conf. Artif. Intell. Smart Syst. ICAIS 2021, pp. 209–213, 2021, doi: 10.1109/ICAIS50930.2021.9395857.

[25] A. Chavda, J. Dsouza, S. Badgujar, and A. Damani, "Multi-Stage CNN Architecture for Face Mask Detection," 2021 6th Int. Conf. Converg. Technol. I2CT 2021, pp. 1–8, 2021, doi: 10.1109/I2CT51068.2021.9418207.

[26] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-shot multi-level face localisation in the wild," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 5202–5211, 2020, doi: 10.1109/CVPR42600.2020.00525.

[27] S. Sharma, K. Shanmugasundaram, and S. K. Ramasamy, "FAREC - CNN based efficient face recognition technique using Dlib," Proc. 2016 Int. Conf. Adv. Commun. Control Comput. Technol. ICACCCT 2016, no. 978, pp. 192–195, 2017, doi: 10.1109/ICACCCT.2016.7831628.

[28] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," IEEE Signal Process. Lett., vol. 23, no. 10, pp. 1499–1503, 2016, doi: 10.1109/LSP.2016.2603342.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.

[30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, pp. 2261–2269, January 2017, doi: 10.1109/CVPR.2017.243.

[31] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 8697–8710, 2018, doi: 10.1109/CVPR.2018.00907.

[32] M. S. Mazli Shahar and L. Mazalan, "Face identity for face mask recognition system," ISCAIE 2021 - IEEE 11th Symp. Comput. Appl. Ind. Electron., pp. 42–47, 2021, doi: 10.1109/ISCAIE51753.2021.9431791.

[33] Z. Wang et al., "Masked Face Recognition Dataset and Application," pp. 1–3, 2020, [Online]. Available: http://arxiv.org/abs/2003.09093.

[34] K. Lin et al., "Face Detection and Segmentation Based on Improved Mask R-CNN," Discret. Dyn. Nat. Soc., vol. 2020, 2020, doi: 10.1155/2020/9242917.

[35] R. Shukla, A.K. Mahapatra and J.S.P. Peter, "Social distancing tracker using yolo v5," Turkish J. Physiother. Rehabil., vol. 32, no. 2, pp. 1785–1793, 2021.

[36] D. Thuan, "Evolution of Yolo Algorithm and Yolov5: the State-of-the-Art Object Detection Algorithm," p. 61, 2021. Available online: https://www.theseus.fi/handle/10024/452552 (accessed on 12 November 2021).

[37] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 8759–8768, 2018, doi: 10.1109/CVPR.2018.00913.

[38] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 658-666, doi: 10.1109/CVPR.2019.00075.

[39] Ketkar, N. (2017). Stochastic Gradient Descent. In: Deep Learning with Python. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-2766-4_8

Secure and Lightweight Authentication Protocol for Smart Metering System

Hind El Makhtoum
Engineering Sciences Laboratory
ENSA Kenitra, Ibn Tofail University
Kenitra, Morocco

Youssef Bentaleb
Engineering Sciences Laboratory
ENSA Kenitra, Ibn Tofail University
Kenitra, Morocco

Abstract—One of the main advantages of the new power grid over the traditional grid is the intelligent energy management by the customer and the Operator. Energy supply, demand response management, and consumption regulation are only possible with the smart metering system. Smart meters are the main component of that system. Hence, a compromised smart meter or a successful attack against this entity may cause data theft, data falsification, and server/device manipulation. Therefore, Smart grids' development and the guarantee of their services are related to the ability to avoid attacks and disasters by ensuring high security. This paper aims to provide a secure and lightweight security protocol that respects the IOT device constraints. The proposition deploys the distributed OTP calculations combined with the Blake2s hash function and the Ascon AEAD cipher to ensure authentication, confidentiality, and integrity. We propose a performance analysis, an informal and a formal security evaluation made by the AVISPA-SPAN tool. Also, we compare the proposed protocol to other similar works. The assessment proves that the proposed protocol is light, valid, secure, and robust against many attacks that threaten the NAN area of the smart metering system, namely, MITM and replay attacks.

Keywords—Internet of things; confidentiality; integrity; authentication; Ascon; Blake2; AVISPA

I. INTRODUCTION

The Internet of Things is growing rapidly and deploying more devices, systems, and other entities. These devices carry a lot of critical data, making these infrastructures a greedy target of attackers. The accessibility of these devices must be controlled so that only authorized users can access the destination server or gateway. Furthermore, any data leakage or corruption can cause serious problems for people, systems, and companies. Therefore, securing IOT must be an occupation for scientists to ensure the system's security and thus maintain its services' effectiveness.

One of the critical IoT applications is the smart grids, namely the smart metering system. Smart meters make information available to customers to manage their consumption behavior. They also provide the necessary information for operators to balance energy response and demand [1]. Smart meters are an energy revolution that will dramatically improve the efficiency and reliability of the power grid. Therefore, the metering system is a greedy target for attackers who could turn these benefits into an absolute disaster if security is not ensured.

Given such risks encountered by the grid, namely metering, a robust and secure system is a must. However, the nature of the smart metering system falls under the same limitations and constraints of the IoT but with major security risks. Indeed, smart metering systems face smart meters limitations such as computing and storage constraints [2] and those of wireless communication networks that further increase the risk of intrusions and attacks [3]. Under these conditions, deploying strong and efficient traditional security solutions is not feasible on smart meters as they involve cumbersome mathematical calculations. Therefore, the design of lightweight and strong security protocols is necessary to ensure the grid's safe operation, protecting data and users and thus maintaining customer confidence.

In light of these limitations, this article aims to propose a scheme to securely authenticate smart meters (namely in this work: SM/Device) to the neighborhood gateway (Namely in this work: SEVER/GATEWAY) while the association phase of the wireless communication. The proposed solution is based on the distributed OTP approach that lightens calculations and storage on the device side. We deploy lightweight and secure protocols for hashing and encryption. In addition, we use the blake2s as the hash function used in the OTP calculations and for the server's authentication. Ascon, the finalist of the Caesar Competition launched by NIST in 2014, is the AEAD cipher deployed in the proposed solution. Ascon cipher is a lightweight solution to ensure authenticity, integrity, and data confidentiality.

The contribution of this paper can be summarized as follows:

- We propose a secure and lightweight security protocol for the NAN area of the smart metering system to address the related security problems and the eventual limitations by using lightweight protocols and session varying parameters.

- We perform formal security analysis by the AVISPA-SPAN simulation to evaluate the proposed protocol's security, validity, and robustness against replay attacks and MITM attacks.

- We also perform informal security analysis to prove that the protocol is robust against many classical and well-known attacks.

- We perform a performance analysis of the protocol regarding computational costs, communication costs, and storage.

- According to the security and performance analysis, we compare the protocol against similar works, and we conclude that the proposed scheme achieves good security and performance results.

The rest of the paper is organized in the following way: Section 2, where we will introduce some related works that address the same issue. Section 3 will be dedicated to preliminaries about the Ascon cipher, the Blake2 hash function, and the distributed OTP approach. Then, we will present the proposition in Section 4 with the formal and informal security analysis. Performance analyses are made in Section 5. Before the conclusion, a comparison of the proposed work with other works is presented in Section 6.

II. RELATED WORKS

The authors of the work [4] provided a secure framework for IoT-based Healthcare systems that addresses four security issues in the fields. The healthcare system, an application of IoT, is also a target to malicious users that may cause dysfunctions. This paper addresses the system's access control by using AES128 with a common pre-shared key to stop the external sensors from accessing the healthcare system. The authors also ensure authenticity by using the public and private keys of the RSA-1024 to ensure that the sensors and medical persons that send messages are real and authorized to make the communication. This paper also addresses confidentiality thanks to the point-to-point encryption based on the AES128. To ensure Integrity, the authors deployed the message authentication code (MAC) based on AES-128 to ensure that the data was intact and not altered [4]. Another work that dealt with the same issue is [5]: The author proposed a protocol based on a mutual authentication based on OTP authentication for both the Device and the gateway. Once the authentication is ensured, a key is generated from an irreversible hashing function. The Key is used as an entry for the AES-GCM protocol. This work is based on AES-GCM to ensure the confidentiality and Integrity of message exchange and thus secure the channel. In the paper [6], The authors presented a security protocol in the smart home domain. The protocol is a combination of encryption algorithms (AES-GCM, RSAOAEP) implemented with SHA3-512 to ensure the confidentiality and integrity of the data communicated by the sensors. The system of [7] provides security based on X.509 certificate, RSA-based Public Key Infrastructure (PKI), hard tokens, challenge/response protocols, and operators' proxies. The system ensures confidentiality, integrity, non-repudiation, privacy, and anonymity.

The cited schemes all addressed the security issues in IoT. However, there are protocols that they deploy separated cryptographic protocols, which are slower and greedy on resource consumption. Other protocols deployed the AEAD AES-GCM, which is vulnerable to the nonce-misuse attack. This vulnerability makes the protocols vulnerable and compromises confidentiality and Integrity of the protocols.

III. PRELIMINARIES

A. Smart Metering System

The architecture of smart grids consists of three levels: the Home Area Network (HAN) with smart meters and home appliances, the middle level or Neighborhood Area Network (NAN) that connects the HAN with the WAN through gateways and concentrators, and the top level of the Wide Area Network (WAN) that is administered by a control center and MDMS servers.

At the HAN level, the smart meters collect data about the electricity consumption of the home equipment and communicate it to the Control Center through the neighborhood gateways. This data is used to make critical decisions about users and all participating parties in the grid.

The smart metering system has a potential role in the smart electrical grid. Indeed, Operator's control centers adjust the smart meters of their client remotely in order to respond adequately to the client's needs and specifications. These bidirectional operations monitor the client's equipment in the HAN area, such as consumption, device specifications, location, and pricing, thus deciding the amount of energy delivered to the client. Hence, confidentiality, privacy, and Integrity are major concerns of the grid to protect clients' consumption and their privacy to protect the grid from attacks and malware that could affect customers' privacy and cause a real disaster [8], [9].

In addition, The smart meters are accessed by multiple agents such as technicians, customers, and operators, which makes them vulnerable to physical attacks. Consequently, smart meters must be well protected in terms of authentication. Moreover, authenticity protects the Operator's system and the other customer from malicious technicians, neighbors, customers, or any other malicious intruder.

B. AEAD CIPHERS: Ascon

ENCRYPTION is a cryptographic method that encodes information to protect it and prevent unauthorized access. The encryption performs several operations(rounds, permutations.) on both Plaintext and a key. These operations produce an unreadable ciphertext that needs the previous Key to decrypt [10].

However, classical encryption does not ensure the authenticity of the message. To deal with this, authenticated Encryption with Associated Data brings up a new level of security thanks to the associated data. The associated data is data related to the sender's time and space that ensures the authenticity of the message and the sender [10].

1) Authenticated Encryption with Associated Data AEAD: Authenticated Encryption with Associated Data is a new cryptographic algorithm that ensures confidentiality, integrity, and authenticity. The technical difference between the AEAD and classical encryption is that the first one generates the ciphertext with a tag. Consequently, decryption and Integrity are performed simultaneously, and the Plaintext is readable only if the Tag is correct [6].

The AEAD encryption is defined as a set of algorithms: the key generation algorithm, the encryption algorithm, and the decryption algorithm.

The encryption algorithm has three inputs, the Nonce N, the associated data AD, and the Plaintext (message). It outputs the ciphertext and a tag : EK: N × AD × Plaintext → Ciphertext × Tag.

The decryption algorithm is the inverse algorithm with the nonce N, the associated data AD, the cipher C and the tag T as inputs. It outputs the Plaintext if C and T are valid : DK: $N \times AD \times C \times T \rightarrow M$ or Error [6].

The most popular and highly adopted lightweight AEAD algorithm is the AES-GCM, which is appropriate for the constrained devices of the IOT [6]. However, it has exposed several shortcomings. The nonce should be unique for each message and non-repeating to prevent nonce misuse attacks, which is not practical for message exchanges. Also, keys used in the hash have been found weak [11].

In order to overcome these shortcomings, NIST launched 2014 the Caesar competition to find out more efficient and lightweight encryption algorithms in terms of applicability, robustness, and security [12].

Among 54 candidates in the first round, only 29 passed to the second one, and 15 candidates were selected for the third round. The two finalist lightweight applications portfolio of the competition is Ascon and ACORN. Ascon is the selected algorithm for the proposed protocol [12].

The two variants of Ascon were selected as the first choice for lightweight applications. Several works evaluated the Ascon algorithm. All analysis supports its large security margin without practice risks, vulnerabilities, or weaknesses [13].

2) ASCON Principle: Ascon is based on the sponge and duplex construction. The algorithm of Ascon is illustrated in Fig. 1 and Fig. 2: Initialization step: An Initial Vector of 64bits (ACON-128 and ASCON-128a), the secret key K, and The nonce N are concatenated to form a 320-bit state. This state passes through the transformation "p" for "a" times/ Then, the secret Key is XORed with the transformed state [13] [14].

Associated Data: The associated data is divided into r-bit blocks and XORed with the first r-bit blocks of the state. After that, a transformation "p" is applied to the resulting state "b" times and XORed again with a 1bit constant.

After these steps, the algorithm moves either to encryption or decryption with two additional phases for each :

Encryption: The plaintext is divided into r-bit blocks, XORed with the first r bits of the state (identical to associated data). Then, it generates blocks of ciphertext that are updated with "p" transformation for "b" times.

Decryption: The ciphertext blocks are XORed with the first r bits of the state and replace the first r bits of the states. They are again updated with the p transformation b times.

The final step after encryption/decryption is the finalization which consists of XORing the secret Key with the state and applying the p transformation "a" time to the state. The secret Key K is once again XORed with the resulted state. The Tag is the least significant 128 bits of the output [13], [14].

3) Ascon Security: The best-known key recovery attack can find the Key with 2^{104} time complexity if the initialization round number is reduced to 7.

Even if the state is recovered, it is impossible to use it for key recovery or forgery attacks. Even the vulnerabilities are not useful for practical attacks.
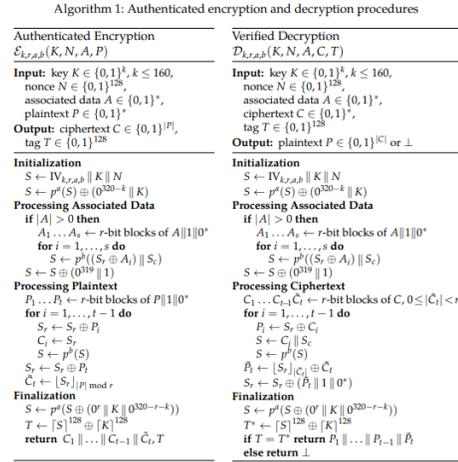
Fig. 1. Ascon Algorithm [13].

Ascon provides security against collision attacks, pre-image attacks, length extension attacks and second-preimage attacks for long messages, misuse attacks, side channel, implementation attack [13], [14], [11].

C. The Distributed OTP Concept [15]

OTP (One-time-password) is a password that changes with each session because it is based on the various counter. This technique is based on an out-of-band exchange of information, such as the keys, counter, and hash functions used in the password calculation process. Thus, OTP has considered one of the most secure authentication techniques because replay attacks are impossible (Fig. 3). The algorithm of the OTP consists of hashes and several calculations , which makes it difficult to be guessed.

The proposal uses lightweight protocols such as Blake 2, a hash function that achieves efficient security results. Blake2 generates hash digests based on stream encryption [16], [17]. The Blake function algorithm uses a 16-word constant combined with the message, salt value, and an initial vector and generates a 4x4 matrix. Then the matrix rows go through eight sets of permutations, and combinations [18], [19].
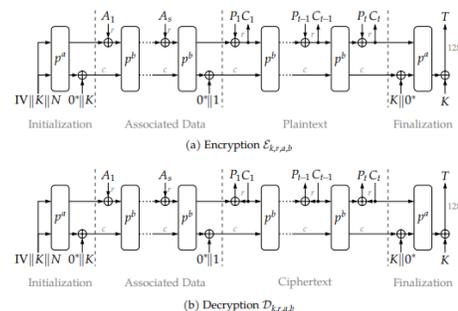
Fig. 2. Ascon Encryption and Decryption Processes [13].
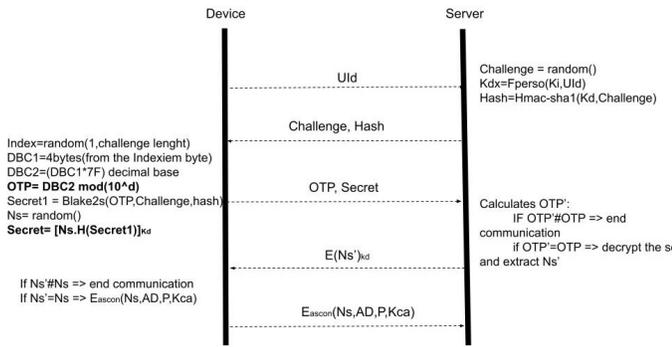
Fig. 3. Distributed OTP Authentication Protocol.

IV. PROPOSED PROTOCOL

Each wireless communication starts with an association step where the communicating objects identify each other through parameters such as identifiers or passwords. At this level, if these parameters are disclosed or intercepted by a malicious third party, the network will suffer an intrusion and lose its security. Hence, the importance of introducing and deploying secure authentication protocols for these communications, such as the proposed protocol which addresses wireless communication security at the NAN level of smart grids between the smart meters and the Gateways/concentrators of the grid.

A. Proposed Protocol Model

The proposed solution aims at designing a security protocol that ensures authentication, authenticity, integrity, and confidentiality. This proposal aims to complete the proposal of [15] while boosting its security by modifying some parameters. As shown in Fig. 4, the protocol is divided into three main parts. The communication starts with an authentication of the Device based on the OTP. Then the protocol moves on to the phase of mutual authentication by the server, based on a random value calculated by the Device. Finally, as soon as both entities are authenticated, they move on to secure the communication channel through Ascon encryption to ensure confidentiality, integrity, and authenticity.
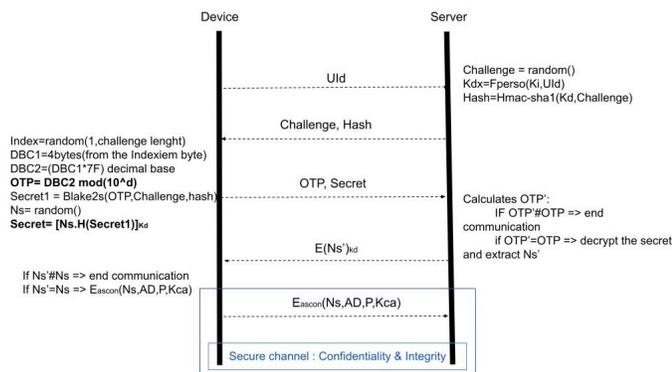
Fig. 4. Proposed Scheme.

1) Pre-Phase: Keys Exchange: This phase is an off-line phase. The symmetric key Kia is affected to the Device and the server by a trusted authority. The secret key Kd is configured on the Device in an out-of-the-line mode based on its Id. In addition, the server gets its secret Key Ki. It keeps communication with an external base where Ids and their appropriate Keys will be stored. The base will avoid attackers trying to impersonate an already authenticated Device.

2) Device Authentication :

- Step 1: The Device will start the communication request by sending its Id to the server.

- Step 2: Once the Server receives the identifier, he will check first if it is already registered. If the Device is already registered and active, it will be dropped. Else, the server will generate a random value as a challenge and calculate the corresponding Key to the Device and a hash value with the HMAC-SHA1 of the calculated key Kd and the challenge. Then he will send the Hash value and the challenges to the Device.

- Step 3: The Device will start calculating the OTP. For this purpose, he will start calculating an Index value, a random value between 1 and the received challenge length. Then, it will truncate the 4 bytes starting from the index byte of the received hash. Next, it will calculate the DBC2 as a decimal value of the multiplication of the DBC1 and 7F. OTP is finally found by applying a modulo d to force the OTP length to "d ."In addition, the Device will calculate a secret value based on Blake2s of the OTP and the last received values of Challenge and Hash.

- Step4: The Server calculates the OTP on its own and compares it to the received value. If the OTP is correct, the server will move to its authentication. If not, it will stop the communication.

3) Server Authentication :

- Step 1: This step starts simultaneously with the third step of the device authentication. While sending its OTP, the Device also sends a secret containing an Ns Value. The secret is a random value Ns encrypted and its concatenation with the blake2s hash of the OTP, the challenge, and the received hash.

- Step 2: The Server will start by authenticating the Device based on its OTP. Once the Device is authorized, the OTP will try to extract the Ns value from the received expression. If the server is right, it can decrypt the secret value. Then, since it could authenticate the Device, it will have the true value of the OTP and the keys required to do the reverse calculation and extract the correct Ns. Thus, we will gain in the number of exchanges and in time.

- Step 3: The Device will start authenticating the server by the Ns value. If the value is correct, they will move to secure the challenge; if not, the communication is stopped.

4) Channel Security: Confidentiality and Integrity: In order to ensure the appropriate security of the channel, we will use authenticated encryption with associated data. The main feature of this step is that a lightweight AEAD cipher encrypts the data. The Device will start encryption based on the Plaintext, the associated data (related to the time of transmission to avoid revealing the location in the IoT context that involve the privacy of the clients), the secret value Ns, and the symmetric key Kia (pre-shared between the Device and the server). Thus, confidentiality, authenticity, and Integrity are ensured, and the data is protected.

V. SECURITY ANALYSIS

A. Informal Analysis

1) Security Considerations:

- Efficiency:
  All used protocols in the proposal are lightweight protocols that respect the limitations of constrained devices. Blake2s and Ascon are two protocols for hashing and encryption adapted to the specifications of connected objects, which has allowed them to be outclassed in NIST competitions. In addition, OTP's computational distribution method offers more advantages to objects in terms of reduced computation and storage in objects.

- Authentication:
  The protocol consists of mutual authentication. Both the Device and the server authenticate to each other. The device authentication is based on the OTP, which is known for a high level of security. In addition, server authentication requires that it compute the OTP value and compare it to the value computed by the Device. It can also decrypt the mathematical expression to extract the received nonce that varies with each session. So, mutual authentication is ensured, safe, lightweight, and robust.

- Data Confidentiality and Integrity:
  Thanks to Ascon CIPHER's potential, it was the first candidate for the final round of the CAESAR competition. Ascon is a lightweight authenticated encryption protocol that was evaluated by many designers who prove that it is secure against many attacks on confidentiality and integrity of data and identities [13], [14].

- One Point of Failure:
  Authentication is based on the distributed calculation of OTP, which means that calculation is not concentrated. Also, both the Device and the server have secret and random parameters that are not shared clearly. Each one needs to make the distributed calculation to authenticate to the other. In other words, the knowledge of the Device and the receiver are not the same, which means there is no point in failure in the protocol.

2) Resistance against Threats and Attacks :

- Impersonation Attack:

The attacker cannot impersonate the Device because the authentication is based on the OTP that changes every session. Also, the server registers the Id devices with their correspondent OTP to avoid malicious devices that would try to impersonate the Device. The attackers cannot either impersonate the server. At the same time, the Ns value is generated in every session. Its calculation is based on ulterior parameters, particularly the secret Key Ki, which is not exchanged in the channel.

- Eavesdropping:
  In the secure exchange step, each piece of data is protected by the Ascon AEAD. The Device uses the Ns nonce shared secret with the server and the associated data related to the precise transmission time. Combining these data allows us to verify the integrity and validity of the message securely. Thus, the attacker can neither listen nor modify the messages.

- MITM (Man in the Middle):
  The proposed security scheme emphasizes strong mutual authentication, also known as two-way authentication. The Device and the server identify each other with this authentication process before starting the data exchange. Without knowing the Device's private Key, the random and unique challenge, and nonce Ns and the corresponding computations, an attacker cannot compute the authentication or encryption data and therefore cannot validate the authentication and proceed to the data exchange. Thus, the Device and the server ensure that they communicate with legitimate correspondents. Therefore, the Man in the Middle attack is impossible in this scenario.

- Forward Secrecy:
  The leakage of a secret key does not affect the rest of the communication. All used keys are combined with the session variable. Even if an attacker intercepts a key, he will not be able to compromise privacy or confidentiality because the Key alone is insufficient to make an attack. In addition, the used keys are even pre-shared or require specific calculations. Hence, a non-legitimate participant has no way to compromise the forward secrecy.

- Forgery Attack:
  An intruder will not be able to authenticate and forge an appropriate request while he has to own the Id and the secret Key Kd. In addition, as we mentioned before, the scheme resists the replay attack and provides mutual authentication. Hence, a forgery attack is prevented.

- Replay Attack:
  In the authentication phase, the OTP and the random parameters Challenge, Ns varies every session. Then, in the transmission phase, the associated data is related to timestamps, which prevents an illegitimate entity that intercept communication from reproducing the same parameters and authenticates to one of the entities.

B. Formal Analysis: Avispa

We perform the formal analysis of the proposed protocol with the AVISPA simulation tool [20] to prove that it is safe and robust against MITM and replay attacks.

The AVISPA tool uses the HLPSL language to make security simulations. It allows the designers to verify their protocols' security and robustness.

1) CAS+: CAS+ is a simple syntax based on the Alice and Bob notation. While CAS+ has a simpler syntax than the HLPSL [20], we made the basic CAS+ algorithm, presented in the Fig. 5 that will be translated later to the HLPSL. The CAS+, even if it is simpler than the HLPSL, the CAS+ is not as performant as the HLPSL language. Also, in some cases, the translator cannot translate correctly [21], [20] Consequently, we started with the CAS+ algorithm. We made the required modifications and added the additional specifications of the protocol directly to the HLPSL file.

```
protocol OurProposition;
identifiers
Device, Server : user;
Challenge,Id,OTP,X,Ns,P,AD,Ok : number;
Kd, Kca : symmetric_key;
H,E : function;
messages
1. Device -> Server : Device,Id
2. Server -> Device : {Challenge,X,Id}Kd
3. Device -> Server : {OTP,H(Challenge,X)}Kd
4. Device -> Server : {Ns,H(Challenge,X,Id)}Kd
5. Server -> Device : {Ns}Kd
6. Device -> Server : {Ns,AD,P}Kca
knowledge
Device : Device, Server, Kd, Kca, Id,P,Ns,AD,
OTP;
Server : Device, Server, Kd, Kca, OTP,X,AD,
Challenge;
session_instances
[Device:Alice,Server:bob,Id:id,Kd:kd,Kca:kca]
[Device:Alice,Server:bob,Kd:kd,
Challenge:challenge,
X:x,Kca:kca];
intruder_knowledge
Alice,bob,challenge,x,id;
goal
Device authenticates Server on Ns;
Server authenticates Device on OTP;
secrecy_of P [Server,Device] ;
```

Fig. 5. Cas+ : proposed solution

2) HLPSL: SPAN is the tool that provides the translation CAS+/HLPSL and also makes the protocol simulation and verification [21], [20] Based on the CAS+ translation made with the SPAN, we added the proposed advanced functions and defined the goals and verifications. We started by defining the two roles, Server and Device. Then, we set up the sessions, environment, verifications, and security goals. Regarding the exchanged messages, we worked on the state's expressions to make actions look alike the proposed protocol. Advanced calculations like truncate and multiplications were performed as concatenations and Xor. We set goals as the verification of the mutual authentication based on OTP and Ns. In addition, we set verification of the secrecy of the plaintext P.

3) Protocol Simulation: Once the HLPSL is executable by SPAN, we get the diagram in Fig. 6. This diagram proves that the whole protocol is readable by the verifications and that the security verifications cover the whole protocol. we moved then to launch the security verifications by the OFMC, and ATSE [22]
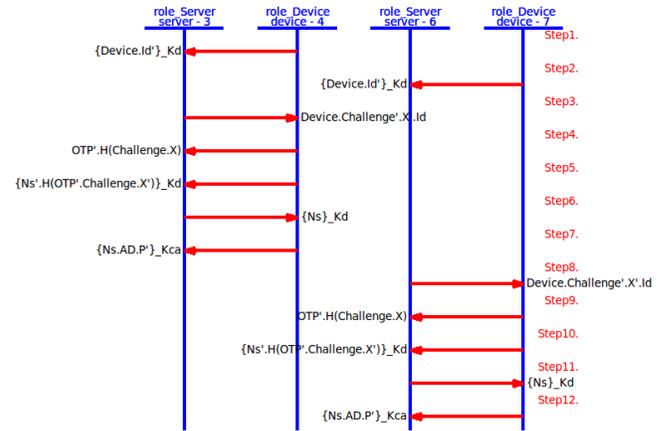
Fig. 6. AVISPA-Span Simulation.

4) Execution Results: In order to evaluate protocols and verify whether the goals set out in HLPSL's algorithm are verified, the backends execute the protocol through multiple finite iterations until the protocol is considered safe for the number of sessions or an attack is found. The OFMC and CL-AtSe backends check if a legitimate entity can execute the protocol correctly while introducing a passive attacker in Fig. 7 and Fig. 8. The four backends of the AVISPA are:

- On-the-fly Model-Checker(OFMC): This uses several symbolic techniques to represent the state-space to perform protocol falsification and verification for the boundless number of sessions in a demand-driven fashion [23].

- Constraint-Logic-based Attack Searcher (CLAtSe): This is a constraint-based approach. It uses some simplification and redundancy elimination techniques to integrate a new specification for cryptographic

```
% OFMC
% Version of 2006/02/13
SUMMARY
  SAFE
DETAILS
  BOUNDED_NUMBER_OF_SESSIONS
PROTOCOL
  /home/span/span/testsuite/results/0904f.if
GOAL
  as_specified
BACKEND
  OFMC
COMMENTS
STATISTICS
  parseTime: 0.00s
  searchTime: 0.03s
  visitedNodes: 18 nodes
  depth: 4 plies
```

Fig. 7. OFMC Verification Results.

functions [23].

```
SUMMARY
 SAFE

DETAILS
 BOUNDED_NUMBER_OF_SESSIONS
 TYPED_MODEL

PROTOCOL
 /home/span/span/testsuite/results/0904f.if

GOAL
 As Specified

BACKEND
 CL-AtSe

STATISTICS

 Analysed  : 38 states
 Reachable  : 8 states
 Translation: 0.02 seconds
 Computation: 0.00 seconds
|
```

Fig. 8. ATSE Verification Results.

- Satisfiability-based Model-Checker (SATMC) and Tree Automata based on Automatic Approximations for the Analysis of Security Protocols (TA4SP): SATMC and TA4SP results were inconclusive because they do not support Xor and modulus operators [24].

VI. Performance Analysis

we will evaluate each participant's computation and communication performance. We will separate the Device's computations and those made by the servers because of the different capacities of each.

Table I presents the different costs of the system's entities. While gateways are more powerful than devices in capacities and resources, we separate calculations to assess the protocols' weight on the constrained entity.

T refers to The time taken while (h: hashing, e: encryption, r: random, c: ascon cipher, and other: small calculations such as multiplications, truncating, and XORing). The storage is expressed by the estimated size of each parameter involved in the communication.

Table II depicts different costs generated from the three phases of the protocol in comparison with similar works that address the security of the communication between smart meters and the neighbor gateways.

As aforementioned in Table I , the proposition requires less storage than the other works. Communication cost is regular

but not the best, while the [28]'s work has better results. Regarding computation, the parameter that makes a difference is the Tc related to Ascon cipher performance. However, the difference in encryption requirement of all other works is more than the double value of the four encryptions used in this work. In addition, Ascon is a lightweight encryption cipher that is more optimized than the standard encryption algorithm.

VII. Discussion

Due to the multiple risks that threaten the smart metering system, accessible from different entities, We estimate that the risks are related to privacy and integrity authenticity. Therefore, we claim that the proposed solution is adapted to the smart metering context. The message size of smart meters is around 100bits and thus covered with the selected protocols, namely Ascon and blake2. In addition, the distributed calculation of OTP respects the constrained nature of smart meters. On the other hand, we claim the mutual authentication that requires higher costs by the side of the gateway does not affect the robustness of the protocol, while the gateway is not constrained in terms of resources. In terms of security, combining the selected protocols with the OTP distribution covers all well-known attacks against the Metering system of the smart grids.

The solution could be extended to other IoT applications. However, the obvious limitation is mainly related to energy aspects which have not been addressed in the scope of the proposed solution since smart meters are always powered by continuous power. However, it is worth considering this issue in a sustainable and ecological approach.

VIII. Conclusion and Perspectives

In this article, we have proposed a protocol that considers the three pillars of security: authentication, confidentiality, and Integrity. We started from the principle of distributing OTP for authentication, which alleviates the objects and respects their constraints in terms of performance. We have carefully chosen secure and lightweight cryptographic algorithms Ascon Cipher and Blake2, two finalists of the competitions launched by NIST, to effectively choose lightweight protocols. The informal and formal security evaluation through AVISPA revealed that the protocol is secure. The comparison has shown that the protocol is better optimized than other similar proposals. However, the costs may seem to be a drawback to the protocol's performance which needs more optimization, especially in the context of smart meters requiring real-time communications.

In addition, the generic architecture of smart metering systems can be a key for deploying innovative technologies within the overall architecture to boost performance and security more efficiently. Hence, it is worth investigating new and

TABLE I. Costs Performance Comparison
(CompC:Computational Cost; ComC: Communication Cost

Entity	CompC	ComC	Storage
Device/SM	2Th+2Te+1Tr+3To+1Tc	2	516
Server/Gateway	1Th+2Te+1Tc+4To	2	844
Total	4Th+4Te+2Tc+7o	4	1360

TABLE II. Costs Performance Comparison
(CompC:Computational Cost; ComC: Communication Cost)

works	CompC	ComC	Storage (Kb)
Proposition	4Th+4Te+2Tc+7o	4	1.3
[25]	17Th+10Te	7	NA
[26]	10Th+1Te	NA	2.6
[27]	96Th+10Te+1To	NA	4
[28]	8Th+8Te+2To	2	3.7

powerful technologies such as AI and blockchains, which have advantages in terms of time, efficiency, security, and lightning of constrained resources.

REFERENCES

[1] M. Meliani, A. E. Barkany, I. E. Abbassi, A. M. Darcherif, and M. Mahmoudi, "Energy management in the smart grid: State-of-the-art and future trends," *International Journal of Engineering Business Management*, vol. 13, p. 18479790211032920, 2021.

[2] V. Kumar, R. Kumar, and S. K. Pandey, "Lkm-ami: a lightweight key management scheme for secure two way communications between smart meters and han devices of ami system in smart grid," *Peer-to-Peer Networking and Applications*, vol. 14, no. 1, pp. 82–100, 2021.

[3] B. Nithya, J. C. Mathew, G. Kavya, N. Anutha, and A. Kumari, "An analysis on cryptographic algorithms for handling network security threats," in *2022 IEEE Delhi Section Conference (DELCON)*. IEEE, 2022, pp. 1–9.

[4] A. K. Chattopadhyay, A. Nag, D. Ghosh, and K. Chanda, "A secure framework for iot-based healthcare system," in *Proceedings of International Ethical Hacking Conference 2018*, M. Chakraborty, S. Chakrabarti, V. E. Balas, and J. K. Mandal, Eds. Singapore: Springer Singapore, 2019, pp. 383–393.

[5] M. T. Hammi, "Sécurisation de l'internet des objets," Ph.D. dissertation, Université Paris-Saclay (ComUE), 2018.

[6] M. Agrawal, J. Zhou, and D. Chang, "A survey on lightweight authenticated encryption and challenges for securing industrial iot," in *Security and Privacy Trends in the Industrial Internet of Things*. Springer, 2019, pp. 71–94.

[7] M. Mumtaz, J. Akram, and L. Ping, "An rsa based authentication system for smart iot environment," in *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 2019, pp. 758–765.

[8] A. Zambroni de Souza and M. Castilla, "Microgrids design and implementation," 2019.

[9] A. E. Ibhaze, M. U. Akpabio, T. O. Akinbulire *et al.*, "A review on smart metering infrastructure," *Int. J. Energy Technology and Policy*, vol. 16, no. 3, p. 277, 2020.

[10] G. Shay, J. Ashwin, and N. Mridul, "counter mode encryption with authentication tag nist 2020," Sep 2020. [Online]. Available: https://csrc.nist.gov/CSRC/media/Projects/Lightweight-Cryptography/documents/round-1/spec-doc/comet-spec.pdf

[11] S. Koteshwara, A. Das, and K. K. Parhi, "Fpga implementation and comparison of aes-gcm and deoxys authenticated encryption schemes," in *2017 IEEE International symposium on circuits and systems (ISCAS)*. IEEE, 2017, pp. 1–4.

[12] NIST, "Caesar: Competition for authenticated encryption: Security, applicability, and robustness"."

[13] d. christoph, e. aria, m. florian, and s. martin, "Ascon v1.2. submission to nist - csrc," Sep 2019. [Online]. Available: https://csrc.nist.gov/CSRC/media/Projects/lightweight-cryptography/documents/round-2/spec-doc-rnd2/ascon-spec-round2.pdf

[14] Z. Çamur, "A study of lightweight cryptography," Master's thesis, Middle East Technical University, 2020.

[15] H. E. Makhtoum and Y. Bentaleb, "An improved iot authentication process based on distributed otp and blake2," 2021.

[16] J.-P. Aumasson, W. Meier, R. C. Phan, and L. Henzen, "The hash function blake," 2014.

[17] H. EL Makhtoum and Y. Bentaleb, "Comparative study of keccak and blake2 hash functions," in *Networking, Intelligent Systems and Security*. Springer, 2022, pp. 343–350.

[18] V. Rao and K. Prema, "Comparative study of lightweight hashing functions for resource constrained devices of iot," in *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, vol. 4. IEEE, 2019, pp. 1–5.

[19] ——, "Light-weight hashing method for user authentication in internet-of-things," *Ad Hoc Networks*, vol. 89, pp. 97–106, 2019.

[20] A. Team *et al.*, "Hlpsl tutorial: A beginners guide to modelling and analysing internet security protocols," *Information Society Technologies*, 2006.

[21] T. Genet, "A short span+ avispa tutorial," Ph.D. dissertation, IRISA, 2015.

[22] P. R. Yogesh *et al.*, "Formal verification of secure evidence collection protocol using ban logic and avispa," *Procedia Computer Science*, vol. 167, pp. 1334–1344, 2020.

[23] M. Singh, M. Ranganathan *et al.*, "Formal verification of bootstrapping remote secure key infrastructures (brski) protocol using avispa," 2020.

[24] A. Javed, "Formal analysis of cwa 14890-1," in *International Conference on Smart Card Research and Advanced Applications*. Springer, 2011, pp. 314–335.

[25] V. Abreu, A. O. Santin, E. K. Viegas, and V. V. Cogo, "Identity and access management for IoT in smart grid," in *Advanced Information Networking and Applications*, L. Barolli, F. Amato, F. Moscato, T. Enokido, and M. Takizawa, Eds. Springer International Publishing, 2020, vol. 1151, pp. 1215–1226.

[26] N. Saxena, B. J. Choi, and R. Lu, "Authentication and authorization scheme for various user roles and devices in smart grid," vol. 11, no. 5, pp. 907–921, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7366583/

[27] Y. Liu, C. Cheng, T. Gu, T. Jiang, and X. Li, "A lightweight authenticated communication scheme for smart grid," vol. 16, no. 3, pp. 836–842, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7295548/

[28] L. Wu, J. Wang, S. Zeadally, and D. He, "Anonymous and efficient message authentication scheme for smart grid," vol. 2019, pp. 1–12, 2019. [Online]. Available: https://www.hindawi.com/journals/scn/2019/4836016/

Applying Logarithm and Russian Multiplication Protocols to Improve Paillier's Cryptosystem

Hamid El Bouabidi[1], Mohamed EL Ghmary[2], Sara Maftah[3], Mohamed Amnai[4] and Ali Ouacha[5]

Department of Computer Science, Ibn Tofaïll University

Faculty of Science, Kenitra, Morocco[1,3,4]

Department of Computer Science, Faculty of Science, Dhar El Mahraz,

Sidi Mohamed Ben Abdellah University, Fez, Morocco[2]

Department of Computer Science

Mohammed V University in Rabat, Morocco[5]

Abstract—Cloud computing provides on-demand access to a diverse set of remote IT services. It offers a number of advantages over traditional computing methods. These advantages include pay-as-you-go pricing, increased agility and on-demand scalability. It also reduces costs due to increased efficiency and better business continuity. The most significant barrier preventing many businesses from moving to the cloud is the security of crucial data maintained by the cloud provider. The cloud server must have complete access to the data to respond to a client request. That implies the decryption key must be sent to the cloud by the client, which may compromise the confidentiality of data stored in the cloud. One way to allow the cloud to use encrypted data without knowing or decrypting it is homomorphic encryption. In this paper, we focus on improving the Paillier cryptosystem, first by using two protocols that allow the cloud to perform the multiplication of encrypted data and then comparing the two protocols in terms of key size and time.

Keywords—Cloud computing; cloud security; homomorphic encryption; paillier cryptosystem; sockets

I. INTRODUCTION

Cloud computing opens up previously untapped possibilities for storage and computation outsourcing. Many people are interested in using this technology since it gives flexibility, accessibility, and cost savings [1], [2], [3]. Over the last two decades, a surge in data has been generated and stored due to the creation of the internet of things, artificial intelligence and cloud computing [4]. However many authors have proposed solutions to optimize the offloading decision and the computing resource allocation to minimize the overall tasks processing time and energy [5], [6], many users are hesitant to commit sensitive data to the cloud due to concerns about privacy and security, making cloud security a critical matter. Indeed, cloud security literature has proposed and evaluated different encryption schemes [7], [8]. Particularly intriguing is homomorphic encryption, which allows any data to remain encrypted while being processed and manipulated. The organization of this paper is described as follows: Section 2 will mention some related works to secure the cloud. Section 3 begins with a summary of Paillier's cryptosystem. Section 4 goes on to describe the two protocols that we use to perform multiplication on encrypted values, followed by our conclusion in Section 5.

II. RELATED WORKS

Research efforts are directed toward several types of homomorphic encryption to secure the cloud, including partially homomorphic encryption. These schemes allow for the execution of a single operation on encrypted data, mainly addition, as in Goldwasser-Micali [9] and DGK [10], or multiplication, as in El Gamal [11] and unpadded RSA [12]. This paper will focus on the well-known additively homomorphic Paillier scheme [13]. It enables the computation of sums on encrypted data, which is useful in a variety of applications, such as encrypted SQL databases [14], machine learning on encrypted data [15], and electronic voting [16]. The authors in [17] address issues, possibilities, and potential improvements related to homomorphic encryption. They describe how we can use homomorphic encryption to process computations in big data. The authors of [18] provide a comprehensive assessment of homomorphic encryption, highlighting current application needs and future potential in areas such as security and privacy. This paper will present two protocols to improve Paillier's encryption scheme and allow the cloud provider to perform multiplication on encrypted data.

III. PAILLIER'S CRYPTOSYSTEM

A. Background

Paillier [13] proposes a new probabilistic encryption method relying on group computations based on calculations over the group $\mathbf{Z_{n^2}}$, where n is an RSA modulus. This scheme is captivating because it is homomorphic, enables the encryption of many bits in a single operation with a constant expansion factor, and enables effective decryption. As a result, it has the potential to be suitable for a variety of cryptographic protocols, such as electronic voting and mix-nets. This approach is similar to Okamoto and Uchiyama's voting and mix-nets cryptosystem [19], in which the group $\mathbf{Z^*_{p^2 \cdot q}}$ is used, where p and q are large primes. The principal difference is that the homomorphic property of this scheme necessitates that the sum of the messages being added be less than p, which is unknown. Because the homomorphic computations in Paillier's method are simply modulo n, this problem is avoided.

B. Paillier Original Algorithm

In [13], Paillier describes two partially homomorphic cryptosystems, schemes 1 and 2. Scheme 1 is the basic Paillier scheme, while scheme 2 is a quicker decryption variant. The Paillier scheme's security relies on the n-th residues in $\mathbf{Z_{n^2}}$ and the toughness of integer factorization. Therefore, we only concisely review the fundamentals and comment on key generation and parameter selection here. Finally, we refer to the original article [13] for more information on the scheme's security. The multiplicative group $\mathbf{Z_{n^2}}$, for $n = pq$ and two prime numbers p and q serve as the setting for the Paillier's scheme. Notice that $\mathbf{Z_{n^2}}$ has $|\mathbf{Z_{n^2}}| = \phi(n^2) = n \cdot \phi(n) = (p-1)(q-1)n$ elements. The Carmichael's function on n, $\lambda(n)$, is short-handed to λ .

1) Scheme 1: In Table. I, Paillier's method is provided in it's most basic form:

TABLE I. PAILLIER'S SCHEME 1

Parameters	prime numbers $n = p.q$ $\lambda = lcm(p-1, q-1)$ $g, g \in \mathbf{Z_{n^2}}$ the order of g is a multiple of n
Public key	n, g
Private key	p, q, λ
Encryption	plaintext $m < n$ select a random $r < n$ such that $r \in \mathbf{Z^*_{n^2}}$, ciphertext $c = g^m r^n mod\, n^2$
Decryption	ciphertext $c < n^2$ plaintext $m = \frac{L(c^\lambda \, mod\, n^2)}{L(g^\lambda \, mod\, n^2)} mod\, n$

Following the notation of [14], $L(u) = \frac{u-1}{n}$, for $u = 1\, mod\, n$. This function is only used on input values u that actually satisfy $u = 1\, mod\, n$.

2) Scheme 2: This is a faster version of the original Paillier algorithm. We work in the subgroup $<g>$ generated by an element g of order αn rather than the entire group $\mathbf{Z^*_{n^2}}$. This enables exponentiation decryption using the exponent $alpha$ instead of $lambda$, which speeds up decryption depending on the size of $alpha$. Scheme 2 is described in Table II:

TABLE II. PAILLIER'S SCHEME 2

Parameters	prime numbers $n = p.q$ α divisor of $\lambda = lcm(p-1, q-1)$ $g, g \in \mathbf{Z_{n^2}}$, the order of g is a multiple of αn
Public key	n, g
Private key	p, q, α
Encryption	plaintext $m < n$ select a random $r < \alpha$ ciphertext $c = g^m (g^n)^r mod\, n^2$
Decryption	ciphertext $c < n^2$ plaintext $m = \frac{L(c^\alpha \, mod\, n^2)}{L(g^\alpha \, mod\, n^2)} mod\, n$

C. Paillier's Cryptosystem Propreties

Paillier's homomorphic encryption has the following Propreties as it is shown in Fig. 1:

- As it's a public key system, anyone with the public key can encrypt, but decryption requires the private key, which is only known to a trustworthy individual.

- It is based on probabilities. It means, an attacker cannot tell whether two ciphertexts are encryptions of the same plaintext or not.

- For addition, it includes the homomorphic properties listed below:

$$E[(m1+m2)]\, mod\, n = E[m_1] \cdot E[m_2]\, mod\, n^2 \quad (1)$$

$$E[(a \cdot m)]\, mod\, n = E[m]^a\, mod\, n^2 \quad (2)$$
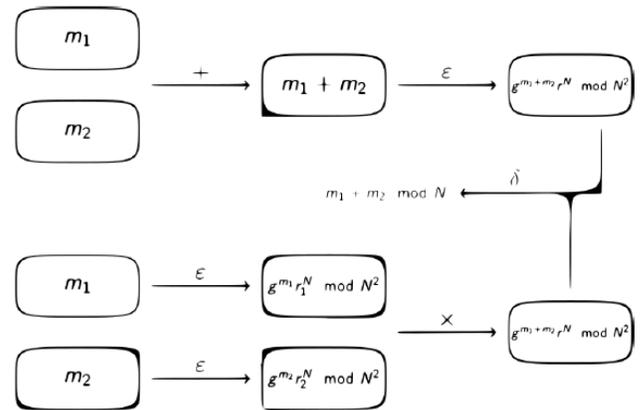
Fig. 1. Paillier Homomorphic Multiplicative Properties.

In which m is the encryption modulus and one of the public key elements. The key generation scheme is as follows:

- Choose p and q as two huge prime numbers such that:

$$gcd(p \cdot q, (p-1) \cdot (q-1)) = 1 \quad (3)$$

This condition is guaranteed if p and q have the same bit lengths.

- Calculate $n = p \cdot q$ and $\lambda = lcm(p-1, q-1)$

- Choose a random integer g from $\mathbf{Z^*_{n^2}}$

- Ensure n divides the order of g by determining whether the following modular multiplicative inverse exists: $\mu = (L(g\, mod\, n^2))^{-1}\, mod\, n$ where $L(u)$ is the quotient of the Euclidean division of $\frac{u-1}{n}$

- The public encryption key is g and n

- The private encryption key is μ and λ

The following operations can then be used to encrypt the message: $m_1 + m_2\, mod\, n$

- Let m represent a message that has to be encrypted from $\mathbf{Z_n}$.

- Choose a random r from $\mathbf{Z^*_n}$

- Calculate ciphertext as:

$$E[m] = c = g^m r^n mod\, n^2 \quad (4)$$

The decryption is basically one exponentiation modulo n^2 :

$$m = L(c^\lambda \, mod \, n^2) \cdot \mu \, mod \, n \qquad (5)$$

The decryption takes advantage of the fact that discrete logarithms are simple to compute,for example if g is chosen as $g = n + 1$ then $L(g^x) \, mod \, n^2 = x \, mod \, n$. Proof can be provided using the binomial. theorem.

IV. APPLYING RUSSIAN MULTIPLICATION AND LOGARITHM PROTOCOLS

In this section, we will describe two protocols that make the Paillier encryption scheme act like fully homomorphic encryption by allowing multiplication of two encrypted values: the Russian multiplication and the continuous logarithm protocols.

A. Paillier Encryption and the Russian Multiplication Protocol

The Russian Peasant Multiplication Method is a common practice in Russian communities. This approach substitutes the frequently used multiplication procedure and only needs the usage of the table of twos. This theorem is currently included in many number theory textbooks [20]. To proceed, multiply the partial products on the left by two and divide the partial products on the right by two. It is similar to expressing the multiplier in base two and then doing multiplications and additions by two. It's a variation of the ancient Egyptian multiplication method. This method's algorithm is as follows:

Algorithm 1 Russian Multiplication

Input: m1,m2,table tab
Output: $m1 \times m2$
1: **while** $m1 > 0$ **do**
2:   **if** ($m1\%2 = 1$) **then**
3:     $e2 = encrypt(m2, pubKey)$
4:     Add $e2$ to tab
5:   **end if**
6:   $m1 = m1//2$
7:   $m2 = m2 * 2$
8: **end while**
9: **return** tab

We used sockets that allow remote machines to communicate with each other using their IP addresses. When a client machine needs a service, it contacts a server machine. This is known as the client-server logic. One asks, the other answers, as illustrated in the sequence diagram Fig. 2:
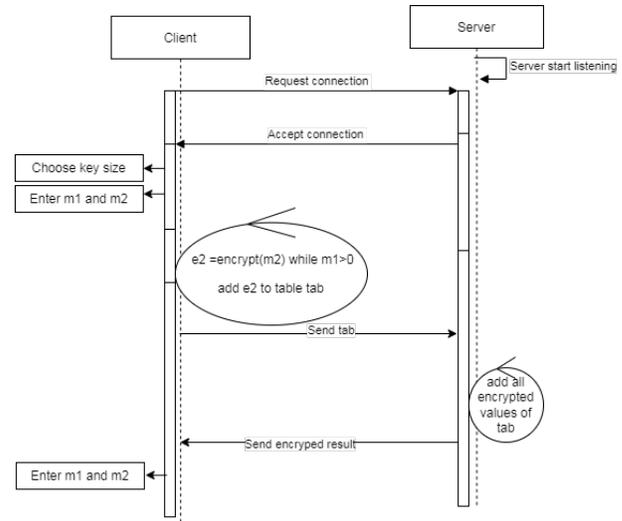
Fig. 2. Russian Multiplication Protocol Sequence Diagram.

Once the Server accepts the connexion, the client request multiplication of $m1$ times $m2$. Let's take the example where $m1 = 73$ and $m2 = 96$. The client interface will transform 73×91 by 91, 728 and 5824 using the Russian multiplication algorithm, encrypt those values and send them to the cloud provider. The cloud can easily add those encrypted values using the Paillier algorithm and then return the result to the client, who could use his private key to see the plaintext. As is shown in the following algorithm:

Algorithm 2 Russian Multiplication Protocol

Input: m1,m2,table tab
Output: $m1 \times m2$
1: {Client Side}
2: **while** $m1 > 0$ **do**
3:   **if** ($m1\%2 = 1$) **then**
4:     $e2 = encrypt(m2, pubKey)$
5:     Add $e2$ to tab
6:   **end if**
7:   $m1 = m1//2$
8:   $m2 = m2 * 2$
9: **end while**
10: send tab
11: R = socket.recieve {result send by the cloud}
12: {Cloud Side}
13: R_tab = socket.recieve(tab)
14: sum = 0
15: **for** x in R_tab **do**
16:   sum = sum + x
17: **end for**
18: socket.send(sum)

B. Paillier Encryption and Continuous Logarithm

Without logarithms, many of our modern technological advances would be nearly impossible. We take advantage of the intriguing rule that transforms multiplication into an addition.

$$log(a \times b) = log(a) + log(b) \qquad (6)$$

Algorithm 3 Continuous Logarithm Multiplication

Input: m1,m2
Output: $m1 \times m2$
  1: $l1 = log(m1)$
  2: $l2 = log(m2)$
  3: $e1 = encrypt(m1)$
  4: $e2 = encrypt(m2)$
  5: $e = e1 + e2$
  6: decrypt(e)
  7: prod=exp(m)

Lets $e1$ and $e2$ be the respective encryption of $m1$ and $m2$ in Fig. 3, by applying the continuous logarithm protocol, the client will be able to compute $m1$ times $m2$ just by sending $e1$ and $e2$ to the cloud.
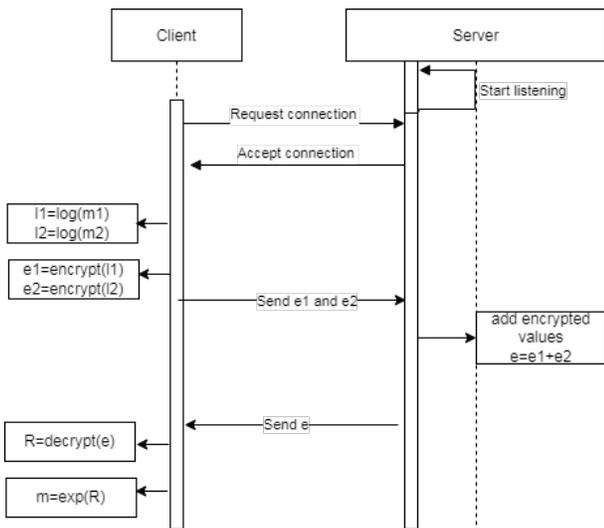
Fig. 3. Continuous Logarithm Protocol Sequence Diagram.

Once the Server accepts the connexion, the client request multiplication of $m1$ times $m2$, and the client interface will apply logarithms on both values. Then, encrypt them and send them to the cloud provider. The cloud will add those encrypted values using the Paillier algorithm and then return the result to the client interface, which could decrypt and apply exponential to display the answer for the client. For a better understanding, we provide the algorithm of the logarithm protocol that allows the cloud to compute production on encrypted values:

Algorithm 4 Continuous Logarithm Multiplication

Input: m1,m2
Output: $m1 \times m2$
  1: {Client Side}
  2: $l1 = log(m1)$
  3: $l2 = log(m2)$
  4: $le1 = encrypt(l1)$
  5: $le2 = encrypt(l2)$
  6: send(le1,le2)
  7: sum=socket.receive
  8: m=decrypt(sum)
  9: message=exp(m)
10: {Cloud Side}
11: sum=0
12: c1=socket.receive(le1)
13: c2=socket.receive(le2)
14: $sum = c1 + c2$
15: socket.send(sum)

C. Implementation and Results

In this section, we propose a description of an implementation of a desktop interface that will allow clients to encrypt a database and request the cloud provider to make calculations on the encrypted data as it is shown in Fig. 4:
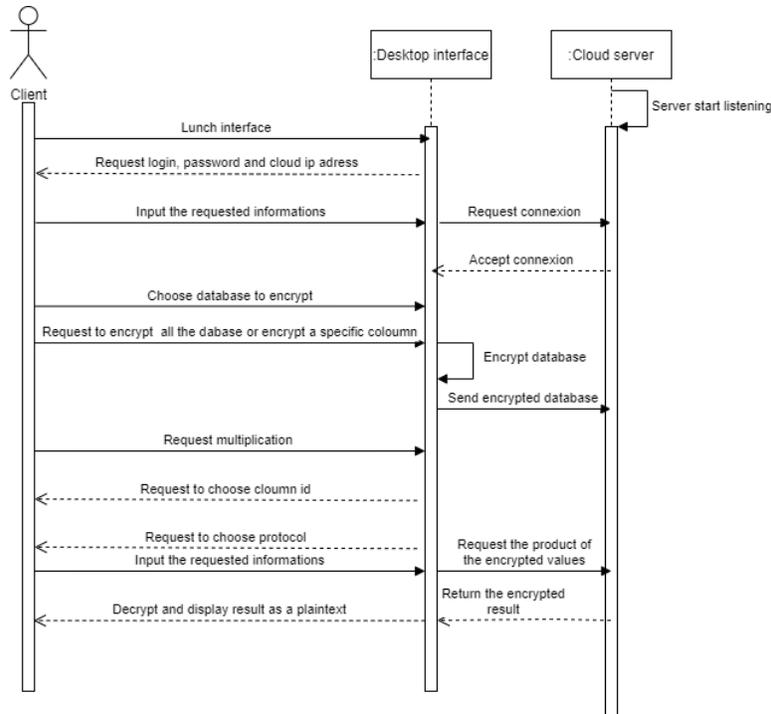
Fig. 4. The Client Desktop Interface Sequence Diagram.

The client desktop interface encrypts the database and sends it to the cloud server. Since the database is stored all encrypted in the cloud, the client sends a request to perform an operation or processing to benefit from the storage and calculation capacity of the cloud servers. Then, the client application decrypts it and returns the same result as if the operation is performed on the data in clear. We validate the applicability of our approach in different cloud solutions by implementing and managing encrypted database operations on a real cloud Hetzner. The current version of our prototype supports TinyDB databases. We chose TinyDB [21] because of the following advantages:

- Written in pure Python

- No dependencies

- Python2 and Python3 compatible

- Easy to use, very clean API Lightweight (2000 lines of code)

To improve security, we encrypted the database name, table names, and field names. This technique will allow us to do a variety of tasks without revealing any information about what we want to accomplish or the contents of our database. Which can be beneficial in many fields, such as medicine to protect the privacy of patients' information [22] or finance and Banking

[23]. To provide additional flexibility to the client so that he does not need to encrypt the whole database, especially if the vital data is on a single column. The client might use the following algorithm to encrypt just that single column:

Algorithm 5 Encrypt Column

Input: column_names table
Output: encrypted column
1: Function encryptcolumn(*id*)
2: **if** column type is String **then**
3:    encrypt column using RSA
4: **else**
5:    encrypt column using Paillier
6: **end if**

Fig. 5 illustrate the result that the cloud gets after a client decided to encrypt all the database using the Algorithm 6.
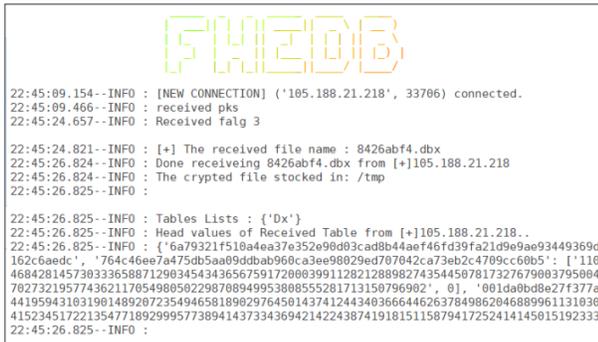
Fig. 5. Encrypted Database Received by the Cloud.

To encrypt the entire database we used the following algorithm:

Algorithm 6 Encrypt All Database

Input: Xtable table ,checked_ele String
Output: encrypted database
1: **if** Xtable exists **then**
2:   **for** non encrypted column **do**
3:     checked_ele = id_of_nonencryptedcolumn
4:     encryptcolumn(checked_ele)
5:   **end for**
6: **else**
7:   create table Xtable
8:   **for** each column in Xtable **do**
9:     **if** column type is String **then**
10:       encrypt column using RSA
11:     **else**
12:       encrypt column using Paillier
13:     **end if**
14:   **end for**
15: **end if**

Table III describes the comparison of the two protocols, using sockets, in terms of time and key size:

TABLE III. RESULT AND COMPARISON OF THE TWO PROTOCOLS USING SOCKETS

Key Length	Russian Multiplication Protocol(ms)	Logarithm Protocol(ms)
N=64	3.3	2.3
N=128	3.6	2.5
N=256	5.3	5.0
N=512	19.7	23.6
N=1024	66.4	60.4
N=2048	425.6.6	409.0
N=4096	2936.0	2830.0

With a tiny key size, we notice that both algorithms provide the same result in about the same amount of time. However, when the key size is increased, the Product encryption protocol employing Russian multiplication takes longer but performs better than the continuous logarithm method. As a result, if the operations performed on the cloud require precision, we should use the Russian multiplication protocol. Still, we can use the continuous logarithm protocol if we want speed with approximate values.

V. CONCLUSION

In this paper, we focused on improving Paillier's method by implementing two protocols that allow the cloud to conduct multiplication on encrypted data by including two protocols that transform multiplication into addition. To show the effectiveness of our approach, we created a desktop interface that enables users to benefit from the cloud while protecting the security of sensitive data held on remote servers and controlled by cloud providers. The client interface adds an extra layer of security to a database by encrypting the names of columns and tables in the database. The proposed solution would have a significant economic effect due to the assurance of data security, confidentiality, and data protection through its use. Additionally, this would encourage more businesses and financial institutions to keep their data in the cloud.

REFERENCES

[1] M. Ahmadi and N. Aslani, "Capabilities and advantages of cloud computing in the implementation of electronic health record," *Acta Informatica Medica*, vol. 26, no. 1, p. 24, 2018.

[2] A. ALI, "An overview of cloud computing for the advancement of the e-learning process," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 3, 2022.

[3] P. Modisane and O. Jokonya, "Evaluating the benefits of cloud computing in small, medium and micro-sized enterprises (smmes)," *Procedia Computer Science*, vol. 181, pp. 784–792, 2021.

[4] A. Ouacha and M. El Ghmary, "Virtual machine migration in mec based artificial intelligence technique," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 1, p. 244, 2021.

[5] M. El Ghmary, T. Chanyour, Y. Hmimz, and M. O. Cherkaoui Malki, "Processing time and computing resources optimization in a mobile edge computing node," in *Embedded Systems and Artificial Intelligence*. Springer, 2020, pp. 99–108.

[6] M. El Ghmary, Y. Hmimz, T. Chanyour, and M. O. C. Malki, "Energy and processing time efficiency for an optimal offloading in a mobile edge computing node," *International Journal of Communication Networks and Information Security*, vol. 12, no. 3, pp. 389–393, 2020.

[7] H. Pham, J. Woodworth, and M. Amini Salehi, "Survey on secure search over encrypted data on the cloud," *Concurrency and Computation: Practice and Experience*, vol. 31, no. 17, p. e5284, 2019.

[8] G. Sen Poh, J.-J. Chin, W.-C. Yau, K.-K. R. Choo, and M. S. Mohamad, "Searchable symmetric encryption: Designs and challe," *ACM Computing Surveys*, vol. 50, no. 3, 2017.

[9] F. Ízdemir, Z. Ídemiş Ízger *et al.*, "Goldwasser-micali algorithm," in *Partially Homomorphic Encryption.* Springer, 2021, pp. 43–49.

[10] I. Damgård, M. Geisler, and M. Krøigaard, "Efficient and secure comparison for on-line auctions," in *Australasian conference on information security and privacy.* Springer, 2007, pp. 416–430.

[11] E. R. Arboleda, "Secure and fast chaotic el gamal cryptosystem," *Int. J. Eng. Adv. Technol*, vol. 8, no. 5, pp. 1693–1699, 2019.

[12] S. Rubinstein-Salzedo, "The rsa cryptosystem," in *Cryptography.* Springer, 2018, pp. 113–126.

[13] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *International conference on the theory and applications of cryptographic techniques.* Springer, 1999, pp. 223–238.

[14] K. Foltz and W. R. Simpson, "Extending cryptdb to operate an erp system on encrypted data." in *ICEIS (1)*, 2018, pp. 103–110.

[15] K. Muhammad, K. A. Sugeng, and H. Murfi, "Machine learning with partially homomorphic encrypted data," in *Journal of Physics: Conference Series*, vol. 1108. IOP Publishing, 2018, p. 012112.

[16] H. Kim, K. E. Kim, S. Park, and J. Sohn, "E-voting system using homomorphic encryption and blockchain technology to encrypt voter data," *arXiv preprint arXiv:2111.05096*, 2021.

[17] B. Alaya, L. Laouamer, and N. Msilini, "Homomorphic encryption systems statement: Trends and challenges," *Computer Science Review*, vol. 36, p. 100235, 2020.

[18] M. Alloghani, M. M. Alani, D. Al-Jumeily, T. Baker, J. Mustafina, A. Hussain, and A. J. Aljaaf, "A systematic review on the status and progress of homomorphic encryption technologies," *Journal of Information Security and Applications*, vol. 48, p. 102362, 2019.

[19] R. Suwandi, S. M. Nasution, and F. Azmi, "Secure e-voting system by utilizing homomorphic properties of the encryption algorithm," *Telkomnika*, vol. 16, no. 2, pp. 862–867, 2018.

[20] B. Cevizci, "How and why does the multiplication method developed by the russian peasants work?" *Journal of Inquiry Based Activities*, vol. 8, no. 1, pp. 24–36, 2018.

[21] M. Siemens, "Tinydb," 2016.

[22] T. Oladunni and S. Sharma, "Homomorphic encryption and data security in the cloud," in *Proceedings of 28th International Conference*, vol. 64, 2019, pp. 129–138.

[23] S. Mittal, P. Jindal, and K. Ramkumar, "Data privacy and system security for banking on clouds using homomorphic encryption," in *2021 2nd International Conference for Emerging Technology (INCET).* IEEE, 2021, pp. 1–6.

Parallelizing Image Processing Algorithms for Face Recognition on Multicore Platforms

Kausar Mia[1]
Department of CSE
Daffodil International University
Dhaka, Bangladesh

Mr. Md Assaduzzaman[3]
Department of CSE
Daffodil International University
Dhaka, Bangladesh

Arnab Saha[5]
Department of CSE
Daffodil International University
Dhaka, Bangladesh

Tariqul Islam[2]
Department of CSE
Daffodil International University
Dhaka, Bangladesh

Tajim Md. Niamat Ullah Akhund[4]
Department of CSE
Daffodil International University
Dhaka, Bangladesh

Sonjoy Prosad Shaha[6]
Department of CSE
Daffodil International University
Dhaka, Bangladesh

Md. Abdur Razzak[7]
Department of CSE
Daffodil International University
Dhaka, Bangladesh

Angkur Dhar[8]
Department of CSE
Daffodil International University
Dhaka, Bangladesh

Abstract—**A good face detection system should have the ability to identify objects with varying degrees of illumination and orientation. It should also be able to respond to all the possible variations in the image. The image of the face depends on the relative camera face pose such as the nose and one eye. The appearance of a face is directly influenced by the facial expression of a person and partially occluded by objects around it. One of the most important and necessary conditions for face recognition is to exclude the background of reliable face classification techniques. However, the face can appear in complex backgrounds and different positions. The face recognition system can mistake some areas of the background for faces. This paper solves some face recognition problems including segmenting, extracting and identifying facial features that are thought to face from the background.**

Keywords—Image processing; multi-core platforms; machine learning; face recognition; parallelizing

I. Introduction

People have an excellent ability to analyse images and we can recognize the face very reliably. Indeed, humans can easily find the surrounding face despite difficult situations such as obscure parts of the face or heavy lightning. Facial recognition is considered a prerequisite for many computer vision applications such as security, surveillance, and content-based image search. So much research has been done to automatically replicate this process on machines [1]. For face detection methods, several authors have defined a vast range of methods for face detection [2]. The feature-invariant approach is primarily aimed at finding structural features that are present even when "the pose, the angle of view, or lighting conditions change". People are expected to easily recognize faces, so they need constant properties or capabilities for these fluctuations. The problem with this approach is that lighting, noise, and occlusion can severely damage image features. The template matching process defines some standard patterns for the face.

These patterns are stored in templates to describe the entire face or individual facial features. The correlation between the entered pattern and the saved pattern is calculated for recognition. In the appearance-based method in contrast to template matching, the model is trained from a series of training images aimed at capturing typical variations in facial appearance. One of the problems with this approach is that training the model can take hours or even days. Face recognition method considerations represent the boundaries between knowledge-based methods and template comparison methods. The latter is usually because it implicitly requires human knowledge to define a face template.

The aim of this work is to implement parallelize Image Processing Algorithms for Face Recognition on Multi-core Platforms. In order to reach our goal we started by analyzing the challenges associated with face detection which involves factors such as pose, presence or absence of structural components, facial expression and imaging conditions. When the face detection methods were analyzed, it was determined that the ones based on learning algorithms (appearance-based) provides better results. Because these methods eliminate the possibility of modeling error, which can occur when face knowledge is either insufficient or inaccurate.

II. Related Works

Face recognition techniques based on learning algorithms (appearance-based) have recently received a lot of attention to eliminating potential modelling errors due to incomplete or inaccurate face knowledge. A neural network-based face recognition framework developed by Rowley et al. [3] proposed, reasonably connected NN (neural network) examines small windows in the image to determine if each window contains a face. Fig. 1 shows Neural network face detection.

Authors of [1] ilustrated the understandings of computer vision and image processing. Authors of [2] mentioned many
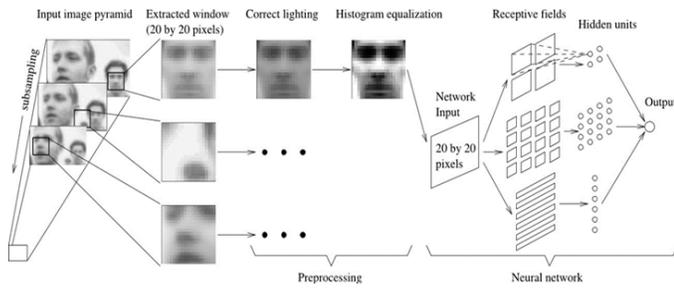
Fig. 1. Neural Network Face Detection.

Fig. 2. Pyramid of Scales.

Fig. 3. Detection Window (Sub-Window).

approach of detecting face in an image. Authors of [4] used deep learning for image recognition. Authors of [3] used neural network for face detection. Authors of [5] showed an improved algorithm in this aspect. Authors of [6] showed GPU performance in computer vision aspect. Authors of [7] worked on Parallelizing face detection. High perfomance computing for image processing is analysed by authors of [8]. Haar-like features were implemented by authors of [9] and [10]. For face detection CNN can be used [11][12] for robust face detection. Multiple faces can be detected with adaboost and Camshaft algorithms [13]. Moreover, artificial intelligence and internet of things based solution can make changes in medical sectors [14], hotel sectors [15], covid-19 patient management sectors [16], and many other sectors [17]. Remote sensing [18] makes a great change in virus-affected people monitoring [19]. IoT-based systems are also secure [20]. These systems can also help in agriculture [21], poultry farm [22], disable people management [23], electronic voting [24], gaming [25], farming [26], nursing [27], remote data sensing [28], virus affected area monitoring [29][30] from remote places [31]. IoT based irrigation systems are helping farmers [32] and farming [33]. Robotics and IoT are doing great in medical fields also [34][35]. So, in multicore platform parallelizing image processing algorithm should work better for face recognition and help mankind.
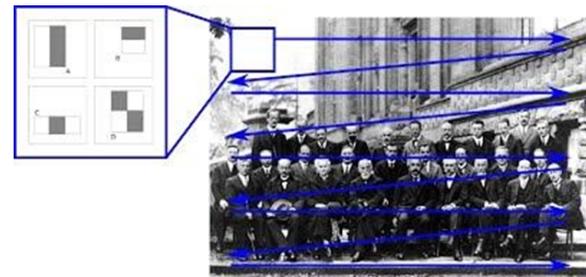
III. Face Detection Optimization

We would like to optimize the face detection task to reduce its computation time [4][3]. To do so it's important to understand how it works inside. The face recognition task first creates an image pyramid, which is a multi-scale representation of the image. This makes the face recognition scale invariant and it recognizes faces in the same recognition window. Fig. 2 shows Pyramid of scales.

On each scale it is important to notice that the size of the detection window is the same in all the scale. This process is such that smaller faces are detected on early scales while bigger faces are detected later. Fig. 3 shows Detection window (sub-window).

A. The Scale Factor

It allows the creation of the image pyramid, by re-scaling the input image we can resize a larger face to a smaller one. Each scale will reduce the size by 10% concerning the previous one. The algorithm works leisurely since a lot of scales are created. It is possible to grow 1.40 for faster detection with

the missing some faces altogether. To precisely detect the face, all the scales that return a face result are merged into one final result. In Fig. 4, 46 scales were generated. Scales (1 - 35) and (43 - 46) did not find any face, scales (36 - 42) did find a face (left image), and all the scales that found a face are merged into one final result in the right image.
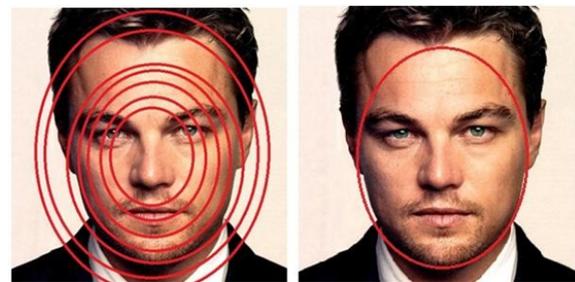
Fig. 4. 46 Scales were Generated. Scales (1 - 35) and (43 - 46) did not find any Face, Scales (36 - 42) did find a Face (Left Image), and All the Scales that found a Face are Merged into One Final Result in the Right Image.

B. Minimum and Maximum Object Size

The standard for this parameter is usually [30, 30] pixels, and for bigger objects than the size, we want this to be ignored. Now that the parameters are known, we would like to work with the low scale factor (1.1) so that all the possible faces are detected but also with a few numbers of scales. The stopping condition for the creation of scales is based on the two parameters defined above (min and max object size). Fig. 5 shows Stopping condition for the creation of scales. The size

Fig. 5. Stopping Condition for the Creation of Scales.

Fig. 8. Computation Decreased as the Min Face size Gets Closer to the Real Face Size.
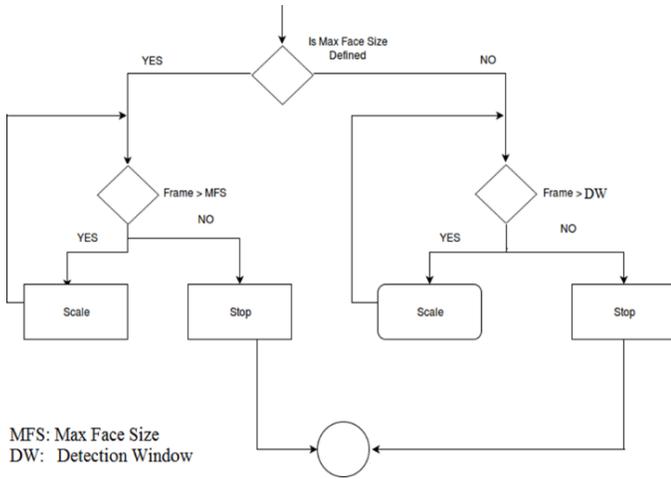
of the detection window is defined as double the min object size. So only objects that a greater than the min object size and smaller than the detection window are considered. Fig. 6 shows the detection window. By experiment was possible

Fig. 6. Detection Window.

to notice that the most important factor to optimize the face-detection is the min object size, as it directly affects the size of the detection window and the number of scales [5]. A final optimization solution for a stream of images can be seen as: first, we start by considering the worst-case scenario in which the parameter for min face size is (30, 30) and a scale factor of 10% which will generate a lot of scales once a face is found, adjust the detection window to the size of the face for the next frames (min_face_size = real_face_size ∗0.8). This process allows reducing drastically the number of scales as they are determined based on the previously detected face dimensions as it's possible to see from Fig. 7. As it is possible to notice
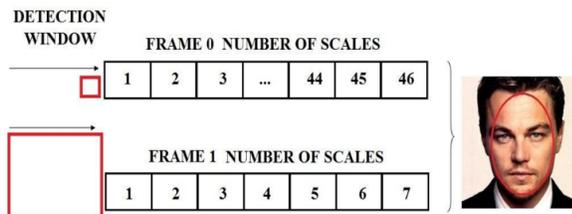
Fig. 7. Optimized Solution.

from Fig. 8, the computation time decrease drastically as the min face size gets closer to the real face size and the number of scales is reduced.

This face detection optimization should be applied only for cases where the goal is to find exactly one face or at most multiple faces that are somehow equidistant from the camera. When the algorithm is applied for the first time (left image), we do not have any a priori knowledge of the size of the faces, and almost all of them are found [36]. Because of the image pyramid of scales, the smallest faces (the ones much farther away from the camera) are the first to be found. Then the intermediate faces (the ones in the center) and finally the largest faces (the ones in the front line). In this way, the largest faces are the last ones to be found, so they will set the threshold of what size of face to look for. When the algorithm is applied for the second time (right image) only the front-line faces are found. Table (I) shows the Styles Summary.

TABLE I. STYLES SUMMARY

Paragraph Style Name (MS Word)	Font Size	Spacing Before	Spacing After
Title_text	20 (B)	0 pt.	24 pt.
Authors_name	12 (B)	0 pt.	12 pt.
Authors_aff	10	0 pt.	0 pt.
Abstract	10 (B)	0 pt.	4 pt.
Keywords	10	12 pt.	12 pt.
Body_text	10	0 pt.	0 pt.
Section_heading	10 (B)	12 pt.	4 pt.
Subsection_heading	10 (B)	12 pt.	4 pt.
Equations	Eqn	12 pt.	12 pt.
Figure	10	12 pt.	0 pt.
Figure_no	10 (I)	6 pt.	12 pt.
Table	10 pt.	0 pt.	12 pt.
Table_no	10 (I)	12 pt.	6 pt.
Reference	10	0 pt.	3 pt.

IV. PARALLEL IMPLEMENTATION

We will start by studying the parallel patterns and tools that will be applied later on to parallelize the application. The potential sources of parallelism of the Viola-Jones algorithm [37] will be also analysed followed by a detailed study of the OpenCV parallel implementation [6]. An improvement to the OpenCV parallelism will also be addressed along with the GPU implementation of the Viola-Jones algorithm. We will discuss techniques for the design of parallel computations and the tools that allow implementing such techniques [38].

A. Pipeline Paradigm

This paradigm is defined for the stream of computations only. It is applied when we have a computation that can be expressed as a composition of functions. The service time of the pipeline depends on the slowest stage. This paradigm is generally characterized by a latency (time between the input of a data and the output of its result after processing) worst than the sequential version due to the communication between the stages. The Pipeline may be used to improve the service time, however, to remove application bottlenecks with this paradigm we must ensure that the number of functions is equal to or greater than the optimal parallelism degree and the stages are somehow balanced. Fig. 9 shows the Pipeline process.

Fig. 9. Pipeline Process.

B. Embarrassingly Parallel Paradigm or Farm

This is a streaming paradigm that replicates the same function that distinct stream elements can be processed by DIM in parallel. This paradigm is composed of three modules: "Emitter, Workers, and Collector". The emitter provides to send every received input stream to one of the workers according to a certain policy. Each input element is transmitted to exactly one worker. Each worker applies the function F to the received data and sends the result to the collector which transmits each of them to the output stream. To apply the farm paradigm the function F must be a pure function, that's it without an internal state. The farm paradigm is good for applications where the order of the output is not so important as the relative speed of the workers can be different. In Fig. 10, we can see the farm as a three-stage Pipeline, in which the emitter is the first stage, the workers in the second, and the collector in the last stage.
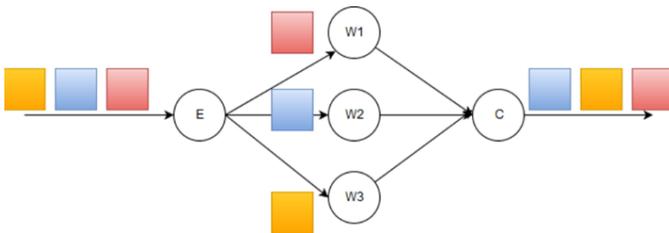
Fig. 10. Embarrassingly Parallel or Farm.

C. Map or Parallel for Paradigm

This is a data-parallel paradigm that can be applied on streams as well as on single data values. This paradigm is characterized by data partitioning and by function replication. For stream-based computations, besides service time, it also optimizes latency and memory capacity [39] which is not the case with the paradigms introduced above. During the execution of function F, each of the functions on its local data exclusively in the Map paradigm, without any collaboration

with the other workers. Parallel for is a paradigm characterized by parallelizing sequential iterative for loops with independent iterations [7]. Fig. 11 shows the Map paradigm.
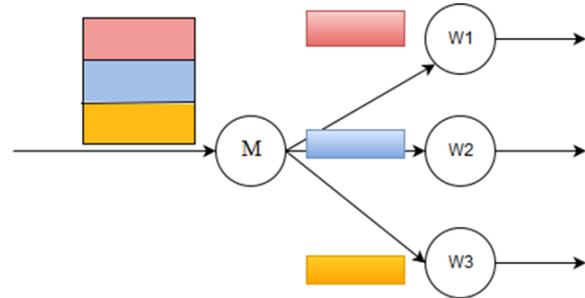
Fig. 11. Map Paradigm.

D. Fast Flow (FF)

Fast Flow is a C++ parallel programming framework that focuses on high-level, pattern-based parallel programming. It supports streaming and data parallelism. It allows expressing the combination of different patterns which increases the parallelism exploitation possibilities [8].

V. Parallel Implementation of OpenCV

OpenMP is a programming framework that greatly simplifies writing multi-threaded programs in Fortran, C, and C++. This programming framework provides loop parallelization with parallel for. Besides being able to implement parallel loops when iterations are independent, it also has the notion of task parallelism which allows assigning different dependent or independent tasks to different threads. To write efficient programs with OpenMP, we need to understand all the additional parameters of the OpenMP pragmas and also the concurrent accesses performed on shared data structures. To implement parallelization, OpenCV uses a hierarchy of tools [9]. The first tool found in the user machine is the one that is going to be applied. For our specific case, all the parallelization implemented by OpenCV was performed with OpenMP. For the object detection task, OpenCV automatically parallelizes the data in each scale with a row-wise Parallel which divides the computation space into a collection of detection windows. Each detection window is calculated independently without any need for synchronization. If the user doesn't specify the number of threads to be used, OpenCV will automatically use all of the available threads. Fig. 12 shows the OpenCV parallelization inside each scale.

A. Convert an RGB Image to a Grayscale

There is no data dependency between the three channels composing an RGB image when performing the grayscale conversion [40]. In this way, we can exploit the Map paradigm to parallelize this task. Synchronization is needed to have the equivalent grayscale of each channel. The final step is to merge the three-channel grayscale images into a single one. Fig. 13 shows the process of Converting an RGB image to a grayscale.
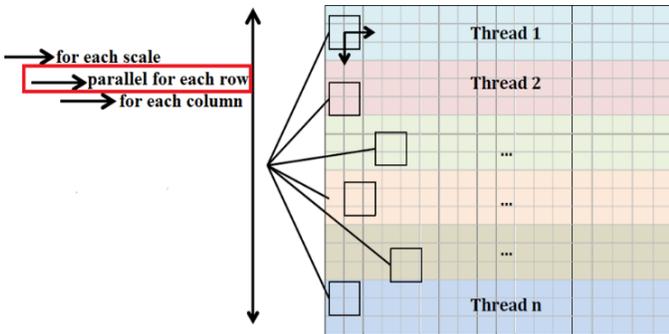
Fig. 12. OpenCV Parallelization Inside each Scale.
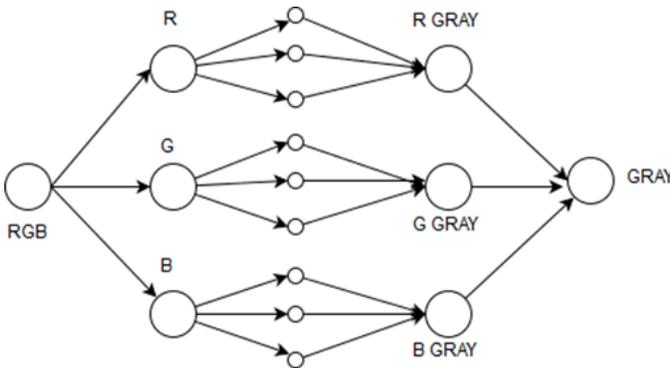
Fig. 13. Converting an RGB Image to a Grayscale.

B. Histogram Equalization

When performing histogram equalization, it's possible to exploit the Map paradigm in which the image is split to compute the histogram H of the source [11]. A synchronization point is needed to have the final histogram H. The computation of the histogram integral image is performed using a pair of recurrences. Transforming H´ as a look-up table can also be performed independently by applying the Map paradigm. Fig. 14 shows the process of Histogram equalization.
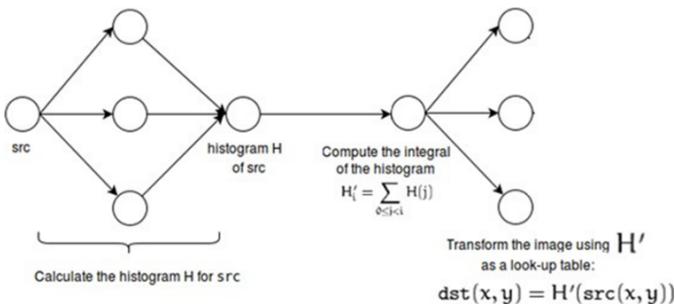
Fig. 14. Histogram Equalization.

C. Detection Window

The face detection computation space is partitioned into a set of detection windows, which can be computed independently without the need for synchronization. The schema for this case is the same as Fig. 11 showed for the Map paradigm

in which the input image is partitioned and detection windows would be applied in parallel in each of the partitions.

D. Haar-Feature Calculation and selection

Inside each detection window no data dependence on the feature value calculation and selection in such a way that each feature can be calculated and selected in a completely independent way [12][10]. All the selected features are sent to the cascade of classifiers shown in Fig. 15.
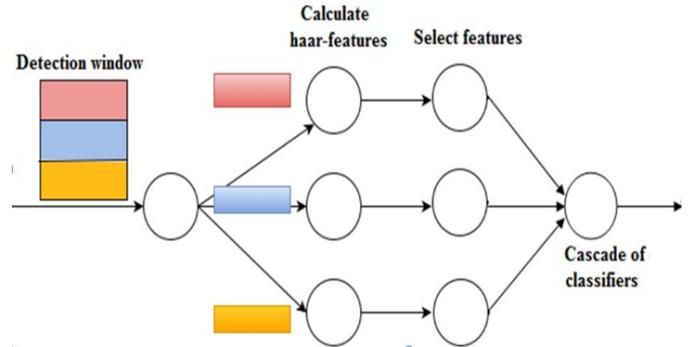
Fig. 15. Haar-Feature Calculation and Selection.

E. Feature Evaluation

All the chosen functions are evaluated through the levels of the cascade classifiers which may be parallelized to the usage of the pipeline paradigm. Map paradigm in which each feature instead of being partitioned is sent entirely to all the filters. This process is known as multicast. A synchronization point is needed to compute the outcome of the stage. The detection window is automatically discarded if the outcome of one of the stages is negative. The face detection cascade of classifiers has 22 stages. The number of filters for each stage can be seen in Fig. 16 below.

Stage	0	1	2	3	4	5	6	7	8	9	10	11	12
Filters	3	15	21	39	33	44	50	51	56	71	80	103	111

Stage	13	14	15	16	17	18	19	20	21
Filters	102	135	137	140	159	177	181	210	213

Fig. 16. Stages of Face Detection Classifier.

Fig. 17 shows Cascade of classifiers.

F. Feature Evaluation

We must scale down the input picture according to the supplied scale factor until it is the same size as the detection window to locate all potential faces of various sizes. Without the requirement for synchronization, each scale of the incoming picture may be handled in parallel. The schema for this case is the same as the figure shown on the Farm paradigm in such a way that each scale would be sent to a worker. The overall parallelization scheme can be seen in Fig. 18 below which the red circles represent.

Fig. 19. Stages.

use FastFlow. The pipeline completion time depends on the slowest stage which in this specific case is stage 2. This stage completion time can be approximated by the face detection task time, so the cost model of the pipeline can be derived as Tc = N x face_detection() where N is the size of the stream. Therefore, to fully exploit the pipeline paradigm and take into account the completion time of each task, it becomes very important to be able to parallelize the face detection procedure internally.

A. Face Detection Task Parallelization

The face detection task processes a pyramid of scales of the same frame. The image pyramid allows face detection to be scale-invariant. There is no dependency between the processing relative to the different scales, they can be executed in parallel without any need for synchronization. The computation time is different between the scales due to their different size while applying the same detection window. Fig. 20 shows Image pyramid.

Fig. 17. Cascade of Classifiers.

Fig. 18. Overall Parallelization Scheme.

Fig. 20. Image Pyramid.

VI. Additional Parallelism to the OpenCV Viola-Jones Implementation

We will focus on providing additional parallelism to the OpenCV implementation of the Viola-Jones algorithm. We will also discuss ways of improving the existing OpenCV parallel implementation. The most straightforward approach to parallelizing a stream computation made up of phases that must be executed serially is to implement a pipeline paradigm. One important aspect to take into account when using the pipeline paradigm is that the stages must be somehow balanced to exploit all the benefits of this parallelization. The proposed pipeline stages are in Fig. 19. For the results that support this pipeline configuration to implement this paradigm, we will

B. Considerations about OpenCV Parallel Implementation

OpenCV internally parallelizes the computation relative to each scale using the parallel paradigm. Externally, the scales are executed sequentially, one after the other. Even though the image scales are of different sizes, the same number of threads is assigned to compute all of them. For this specific case, the face detection task generated an image pyramid of 46 scales. We measured the time of each scale using a thread number between 1 and the maximum number of threads available (24 in this case). As evidenced in Fig. 21, the scalability achieved with 24 threads is very significant for scales of bigger size while on the other hand is negligible for scales of smaller size. If we consider the efficiency, it decreases as the number
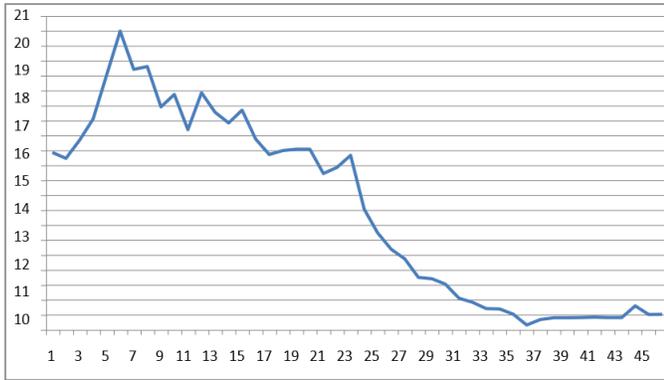
Fig. 21. Scalability Achieved with 24 Threads.

of threads increases Fig. 22. When considering the efficiency of each scale with 24 threads Fig. 23, it decreases drastically from the second half (from scale 24 to 46 of Fig. 23).
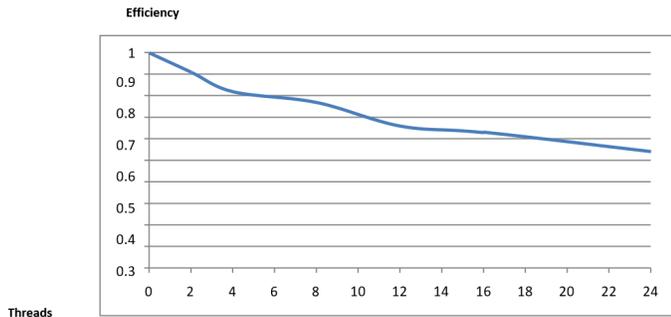
Fig. 22. Efficiency Achieved (1).
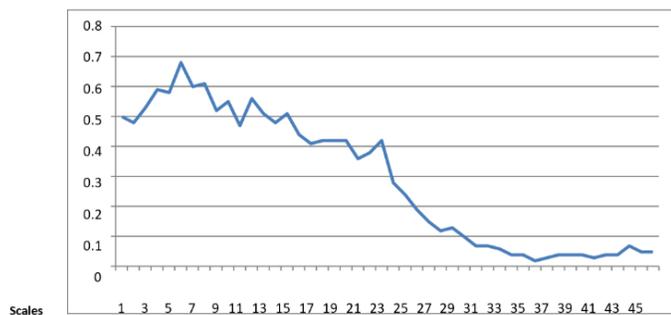
Fig. 23. Efficiency Achieved (2).

C. Improvement to the OpenCV Parallel Implementation

Analyzing deeply the results obtained from the Fig. 23 above, it's possible to realize that we can optimize the parallelization with 24 threads in the following alternative ways shown in Fig. 24.

Since in both alternatives we are parallelizing for loops, OpenMP is the most suitable parallelization tool. We can see that executing in parallel from scale 24 to 46 but inside executing serially each scale, without taking into account any
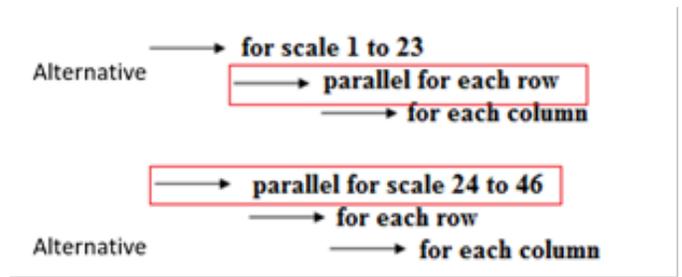
Fig. 24. Parallel Implementation.

overhead, the computation time can be roughly approximated to the time spent to execute the slowest scale (which is scale 24) with 20.62 milliseconds. But if we take the OpenCV approach which executes the scales serially while performing a row-wise parallel computation inside each scale, the execution time from scale 24 to 46 with 24 threads is roughly the sum of the times of each scale on this range which is 44.437 milliseconds. So, with the optimized approach, we can get about 2.15x speed up for the OpenCV approach. Assigning computation of different scales to different threads would fall under the category of embarrassingly parallel or farm paradigm. A dynamic load balance technique can be applied in this case such that when each scale finishes its execution, it frees the resource. In all circumstances when the exact quantity of labor of activities cannot be exactly stated before the activities take place, this strategy may provide better load balancing. Based also on other performed tests which demonstrated similar (scalability) results like the ones shown before, we divide the total number of scales by half. The heavy-sized scales in the first half would be parallelized using a row-wise parallel, while the light-sized scales in the second half would be parallelized by computing each scale in parallel. From Fig. 25 is possible to notice that the face detection parallelization is rather coarse-grained as the best scalability was achieved while the smaller one achieved the worst result.
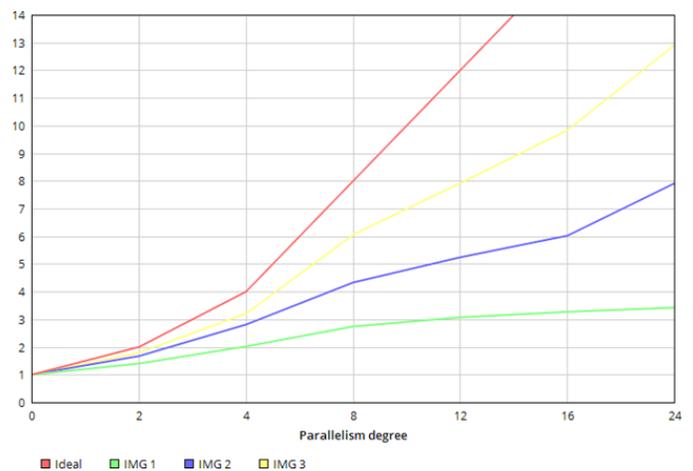
Fig. 25. Face Detection Parallelization.

D. Specific Approach: Single Face Detection

For this specific case, before finding a face the paralleliza-tion technique is the same as the one described above in the proposed heuristic solution. Once the face is found, taking advantage of the optimization, where the number of scales and the computation time is reduced drastically, as it is possible to see from the Table (II) below, the parallelization paradigm that could be applied is the farm where each scale would be assigned to a specific number of the worker. One could argue that after finding a face no further parallelization is needed taking into account that the number of scales and the computation time of each scale is reduced drastically. For our specific case, we can see from Table (II) that the sequential implementation would take as much as 4.01 milliseconds to complete while in the best-case scenario, without overheads, the parallel implementation would take 0.95 milliseconds. It is worth remembering that the results of the Tab. 2 were obtained with the largest image of our experiments (Size 2448 x 2448) which represents the worst case of the face sizes. When introducing further parallelization to this part, we need to take into account the trade-off between the gains of the parallelization and the additional overheads that it comes with as even in the best-case scenario the gains can be very negligible (around 3 ms for our case).

TABLE II. OPTIMIZED RESULTS FOR SINGLE FACE DETECTION

Scale	Size	1-thread[ms]
1	2348x2348	0.87
2	2134x2134	0.95
3	1940x1940	0.80
4	1764x1764	0.76
5	1604x1604	0.44
6	1458x1458	0.18
7	1325x1325	0.01

E. Feature Detection Parallelization

In this case, we are not interested in parallelizing each feature detection phase internally, as the optimization proposed already allows for achieving higher speed up. The important point is to run all the features (eyes, mouth, and nose) in parallel using the MISD pattern which is a variation of the Map paradigm where each worker computes in different codes. Fig. 26 shows Feature detection parallelization.

A possible approach for detecting features in parallel for multiple faces could be applying the Farm paradigm in which each face would be assigned to a certain work and inside each worker [41], the MISD paradigm would compute the features in parallel in Fig. 27.

VII. RESULTS AND DISCUSSION

Our authors saw the testing on this application, and it was determined that splitting the face did not improve performance, therefore we opted to use the four-stage pipeline. When using the Multiple Face Detection Approach to test the application, the smallest image was able to achieve real-time processing with the GPU (around 38 FPS) and quasi-real-time processing with the CPU parallelized version (around 10 FPS), while the intermediate size image was able to achieve quasi-real-time processing with GPU (Around 17 FPS) [42]. The least and intermediate photos were able to achieve real-time processing
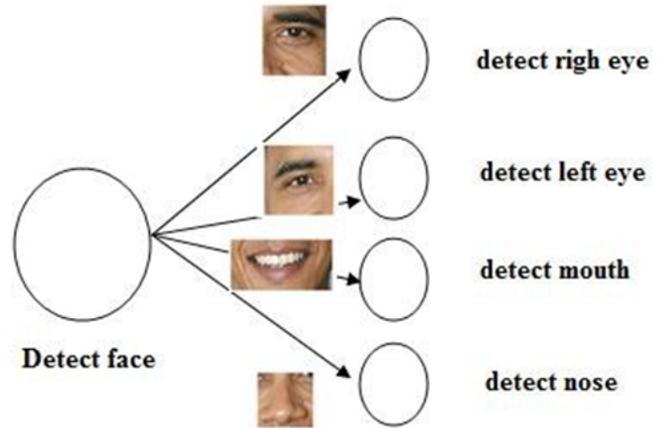
Fig. 26. Feature Detection Parallelization.
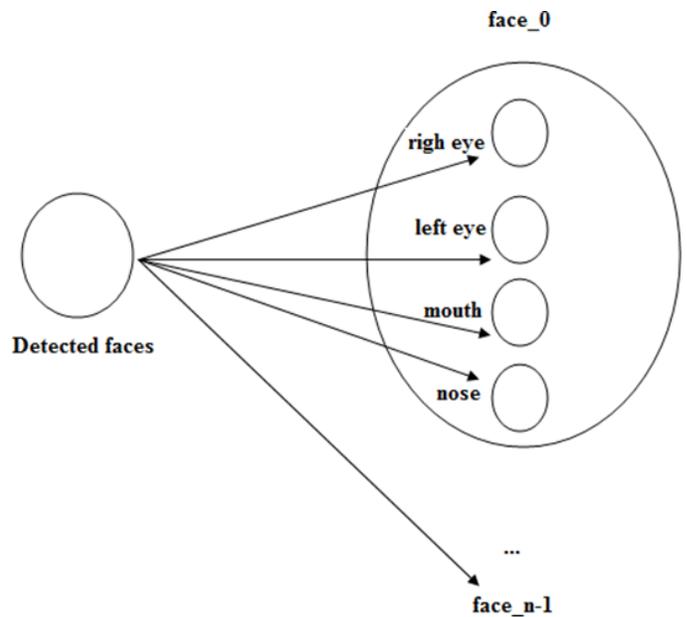
Fig. 27. Detecting Features in Parallel for Multiple Faces.

with all implementations save the originally sequential one when we ran the same trials with the Single Face Detection Approach. With the parallel CPU and GPU implementations, the largest image was able to accomplish real-time processing [43] [44]. When the application was tested on video frames, it was discovered that the size of the face had a direct impact on the sequential and parallel CPU implementations optimized for single face detection. We demonstrated that for the largest image size, the optimization introduced, allowed to decrease in the number of scales from 46 to 7 and the number of processed sub-windows from 8 443 722 to 704 which implies a reduction of 99.99% of processed sub-windows. The reliability tests showed that our application is very reliable in different scenarios such as distance, light conditions, occlusion, and rotation [45]. We demonstrated that it is possible to coexist the two parallel programming frameworks utilized in our application (Fast Flow and OpenMP) to exploit all the available resources on CPU Multi-Core Platforms.

VIII. Conclusion

This paper aimed to Parallelize Image Processing Algorithms for Face Recognition on Multicore Platforms. The heavy-sized scales in the first half would be parallelized using a row-wise parallel, while the light-sized scales in the second half would be parallelized by computing each scale in parallel. When the face detection methodologies were analyzed, was concluded that the ones based on learning algorithms (appearance-based) allow for achieving better results because eliminate the potential modeling error as a result of incomplete or inaccurate face knowledge. The state-of-the-art algorithms can process images extremely rapidly based on three innovative contributions: "The integral image, the AdaBoost learning algorithm, and a cascade of classifiers". The Viola-Jones algorithm is still not optimal as the detector becomes unreliable for faces with a certain rotation, the detector may fail when the faces are very dark while the background is relatively light and the detector fails on significantly occluded faces. To render the face detection process faster two optimizations were proposed. The first optimizes the face detection task based on the size of the face found on the previous frame. This optimization allows decreasing drastically the computation time. The drawback of this optimization is that it should be applied only for one face, or at most multiple faces that are somehow equidistant from the camera. The second optimization also allows decreasing the computation time as the size of the facial features to look for is based on a ratio to the size of the face found. To parallelize the entire face detection process, it was shown that the pipeline paradigm is the most suitable solution as the whole process is performed in a sequence of phases. To take advantage of the face detection parallelization on GPUs specifies that we need to minimize access to devise memory, transfers to and from the GPU must be timed to coincide with computation, and memory accesses to the GPU's global and shared memory must prevent bank conflicts, and branching inside kernels should be kept to a minimum. Our tests have demonstrated that our application can achieve real-time processing rates in different scenarios and it is reliable in various situations. A theoretical study on the reasons why the Viola-Jones algorithm may not allow multi-core platforms to be exploited to their full capacity. Future work includes studying in more detail the aspects and implementing them to verify their efficiency. Also, a more detailed analysis of some of the principles mentioned to take full advantage of the GPU hardware requires further investigation.

References

[1] E. Hjelmås and B. K. Low, "Face detection: A survey," *Computer vision and image understanding*, vol. 83, no. 3, pp. 236–274, 2001.

[2] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

[3] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 1, pp. 23–38, 1998.

[4] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "From facial parts responses to face detection: A deep learning approach," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3676–3684.

[5] G.-W. Kim and D.-S. Kang, "Improved camshift algorithm based on kalman filter," *Adv. Sci. Technol. Lett*, vol. 98, pp. 135–137, 2015.

[6] D. Hefenbrock, J. Oberg, N. T. N. Thanh, R. Kastner, and S. B. Baden, "Accelerating viola-jones face detection to fpga-level using

gpus," in *2010 18th IEEE Annual International Symposium on Field-Programmable Custom Computing Machines*. IEEE, 2010, pp. 11–18.

[7] M. W. Joshua Miguel, Jordan Saleh. Parallelizing face detection in software. [Online]. Available: DepartmentofElectricalandComputerEngineering,UniversityofToronto, 2013.

[8] M. Vanneschi, "High performance computing: parallel processing models and architectures," in *High performance computing*. Pisa University Press, 2014, pp. 1–552.

[9] H. Commin, "Robust real-time extraction of fiducial facial feature points using haar-like features," *arXiv preprint arXiv:1505.04286*, 2015.

[10] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Proceedings. international conference on image processing*, vol. 1. IEEE, 2002, pp. I–I.

[11] I. Kalinovskii and V. Spitsyn, "Compact convolutional neural network cascade for face detection," *arXiv preprint arXiv:1508.01292*, 2015.

[12] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[13] S. Ash with Kumar and S. Kubakaddi., "Multiple face detection and tracking using adaboost and camshaft algorithm," *Int. Journal of Research Studies in Science, Engineering and Technology*, vol. 1, no. 1, pp. 8–8, 2014.

[14] N. U. Akhund, T. Md, M. Mahi, J. Nayeen, A. Hasnat Tanvir, M. Mahmud, and M. S. Kaiser, "Adeptness: Alzheimer's disease patient management system using pervasive sensors-early prototype and preliminary results," in *International conference on brain informatics*. Springer, 2018, pp. 413–422.

[15] T. M. N. U. Akhund, W. B. Jyoty, M. A. B. Siddik, N. T. Newaz, S. A. Al Wahid, and M. M. Sarker, "Iot based low-cost robotic agent design for disabled and covid-19 virus affected people," in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*. IEEE, 2020, pp. 23–26.

[16] T. M. N. U. Akhund, M. A. B. Siddik, M. R. Hossain, M. M. Rahman, N. T. Newaz, and M. Saifuzzaman, "Iot waiter bot: a low cost iot based multi functioned robot for restaurants," in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2020, pp. 1174–1178.

[17] F. I. Suny, M. R. Fahim, M. Rahman, N. T. Newaz, T. M. Akhund, N. Ullah *et al.*, "Iot past, present, and future a literary survey," in *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*. Springer, 2021, pp. 393–402.

[18] T. M. N. U. Akhund, N. T. Newaz, and M. R. Hossain, "Low-cost remote sensing iot based smartphone controlled robot for virus affected people," 2020.

[19] T. M. Akhund, N. Ullah, N. T. Newaz, M. Rakib Hossain, and M. Shamim Kaiser, "Low-cost smartphone-controlled remote sensing iot robot," in *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*. Springer, 2021, pp. 569–576.

[20] N. T. Newaz, M. R. Haque, T. M. N. U. Akhund, T. Khatun, M. Biswas, and M. A. Yousuf, "Iot security perspectives and probable solution," in *2021 Fifth World Conference on Smart Trends in Systems Security and Sustainability (WorldS4)*. IEEE, 2021, pp. 81–86.

[21] M. Biswas, T. M. N. U. Akhund, M. J. Ferdous, S. Kar, A. Anis, and S. A. Shanto, "Biot: Blockchain based smart agriculture with internet of thing," in *2021 Fifth World Conference on Smart Trends in Systems Security and Sustainability (WorldS4)*. IEEE, 2021, pp. 75–80.

[22] T. M. Akhund, N. Ullah, S. R. Snigdha, M. Reza, N. T. Newaz, M. Saifuzzaman, M. R. Rashel *et al.*, "Self-powered iot-based design for multi-purpose smart poultry farm," in *International Conference on Information and Communication Technology for Intelligent Systems*. Springer, 2020, pp. 43–51.

[23] T. M. Akhund, N. Ullah, G. Roy, A. Adhikary, A. Alam, N. T. Newaz, M. Rana Rashel, M. Abu Yousuf *et al.*, "Snappy wheelchair: An iot-based flex controlled robotic wheel chair for disabled people," in *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*. Springer, 2021, pp. 803–812.

[24] M. M. Sarker, M. A. I. Shah, T. Akhund, and M. S. Uddin, "An approach of automated electronic voting management system for bangladesh using biometric fingerprint," *International Journal of Advanced Engineering Research and Science*, vol. 3, no. 11, p. 236907, 2016.

[25] M. Biswas, N. U. Akhund, T. Md, M. Mahbub, S. Islam, S. Md, S. Sorna, M. Shamim Kaiser *et al.*, "A survey on predicting player's performance and team recommendation in game of cricket using machine learning," in *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*. Springer, 2022, pp. 223–230.

[26] M. A. Hasan and T. Akhund, "An approach to create iot based automated smart farming system for paddy cultivation."

[27] T. Akhund, "Study and implementation of multi-purpose iot nurse-bot."

[28] T. Akhund, I. A. Sagar, and M. M. Sarker, "Remote temperature sensing line following robot with bluetooth data sending capability," in *International Conference on Recent Advances in Mathematical and Physical Sciences (ICRAMPS)*, 2018.

[29] T. Akhund, N. T. Newaz, and M. M. Sarker, "Posture recognizer robot with remote sensing for virus invaded area people," *Journal of Information Technology (JIT)*, vol. 9, pp. 1–6, 2020.

[30] T. Akhund, "Remote sensing iot based android controlled robot," *Methodology*, vol. 9, no. 11, 2018.

[31] T. M. N. U. Akhund, N. T. Newaz, and M. R. Hossain, "Low-cost remote sensing iot based smartphone controlled robot for virus affected people," 2020.

[32] T. M. Akhund, N. Ullah, N. T. Newaz, Z. Zaman, A. Sultana, A. Barros, and M. Whaiduzzaman, "Iot-based low-cost automated irrigation system for smart farming," in *Intelligent Sustainable Systems*. Springer, 2022, pp. 83–91.

[33] M. Suny, F. Islam, T. Khatun, Z. Zaman, M. Fahim, M. Roshed, M. Islam, R. Jesmin, T. M. Akhund, N. Ullah *et al.*, "Smart agricultural system using iot," in *Intelligent Sustainable Systems*. Springer, 2022, pp. 73–82.

[34] A. H. Himel, F. A. Boby, S. Saba, T. M. Akhund, N. Ullah, and K. Ali, "Contribution of robotics in medical applications a literary survey," in *Intelligent Sustainable Systems*. Springer, 2022, pp. 247–255.

[35] T. M. N. U. Akhund, M. Hossain, K. Kubra, Nurjahan, A. Barros, and M. Whaiduzzaman, "Iot based low-cost posture and bluetooth controlled robot for disabled and virus affected people," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 8, 2022. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2022.0130879

[36] O. T. C++. Histogram equalization of grayscale or color image. Accessed: April, 2022. [Online]. Available: http://opencv-srf.blogspot.it/2013/08/histogram-equalization.html

[37] R. Ladu, "Gpu: A formal introduction."

[38] O. T. C++. Histograms - 2: Histogram equalization. Accessed: April, 2022. [Online]. Available: http://docs.opencv.org/3.1.0/d5/daf/tutorial_py_histogram_equalization.html{\#}gs

[39] C. U. C. for Advanced Computing. Memory coalescing. Accessed: May, 2022. [Online]. Available: https://cvw.cac.cornell.edu/gpu/coalesced?AspxAutoDetectCookieSupport=1

[40] OpenCV. The opencv reference manual.

[41] M. Danelutto. Distributed systems paradigms and models. [Online]. Available: TeachingmaterialofLaureaMagistraleinComputerScienceandNetworking

[42] NVidia. What is gpu computing. Accessed May, 2022. [Online]. Available: http://www.nvidia.com/object/what-is-GPU-computing.html

[43] G. C. Buttazzo, *Hard real-time computing systems: predictable scheduling algorithms and applications*. Springer Science & Business Media, 2011, vol. 24.

[44] N. Prabhakar, V. Vaithiyanathan, A. P. Sharma, A. Singh, and P. Singhal, "Object tracking using frame differencing and template matching," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 4, no. 24, pp. 5497–5501, 2012.

[45] A. Jain, J. Bharti, and M. Gupta, "Improvements in opencv's viola jones algorithm in face detection-tilted face detection," *International Journal on Signal and Image Processing*, vol. 5, no. 2, p. 21, 2014.

A Hybrid Protection Method to Enhance Data Utility while Preserving the Privacy of Medical Patients Data Publishing

Shermina Jeba[1]
Department of Computing
Muscat College
Muscat, Oman

Mohd Arfian Ismail*[3]
Faculty of Computing
Universiti Malaysia Pahang
Kuantan, Pahang, Malaysia

Sarachandran Nair[5]
Department of Computing
Muscat College
Muscat, Oman

Mohammed BinJubier[2]
Faculty of Computing
Universiti Malaysia Pahang
Kuantan, Pahang, Malaysia

Reshmy Krishnan[4]
Department of Computing
Muscat College
Muscat, Oman

Girija Narasimhan[6]
Information Technology Department
University of Technology and Applied Sciences
Muscat, Oman

Abstract—Medical patient data need to be published and made available to researchers so that they can use, analyse, and evaluate the data effectively. However, publishing medical patient data raises privacy concerns regarding protecting sensitive data while preserving the utility of the released data. The privacy-preserving data publishing (PPDP) process attempts to keep public data useful without risking the medical patients' privacy. Through protection methods like perturbing, suppressing, or generalizing values, which lead to uncertainty in identity inference or sensitive value estimation, the PPDP aims to reduce the risks of patient data being disclosed and to preserve the potential use of published data. Although this method is helpful, information loss is inevitable when attempting to achieve a high level of privacy using protection methods. In addition, the privacy-preserving techniques may affect the use of data, resulting in imprecise or even impractical knowledge extraction. Thus, balancing privacy and utility in medical patient data is essential. This study proposed an innovative technique that used a hybrid protection method for utility enhancement while preserving medical patients' data privacy. The utilized technique could partition information horizontally and vertically, resulting in data being grouped into columns and equivalence classes. Then, the attributes assumed to be easily known by any attacker are determined by upper and lower protection levels (UPL and LPL). This work also depends on making the false matches and value swapping to make sure that the attribute disclosure is less likely to happen. The innovative technique makes data more useful. According to the results, the innovative technique delivers about 93.4% data utility when the percentage of exchange level is 5% using LPL and 95% using UPL with a 4.5K medical patient dataset. In conclusion, the innovative technique has minimized risk disclosure compared to other existing works.

Keywords—*Medical patients data publishing; anonymization; protection method for preserving the privacy*

I. INTRODUCTION

Data publication is the simplest data-sharing method, allowing research institutions to conduct data mining operations on published medical-patient databases for knowledge extraction. This knowledge represents, interprets, or discovers new patterns [1] [2]. However, the potential of published data has yet to be explored because scholars face several challenges when extracting information from published medical-patient data. One of these challenges is the patient data privacy, which results in the exposure of individuals' identities, unauthorized access to information and private data, and use of personal information for unintended purposes [1] [3] [17]. Even if the identity attributes (IAs) (such as names and social security numbers) that identify users from the patient table are removed based on data protection, the remaining patient data can be used to re-identify the person in most cases. Furthermore, sensitive attributes (SA) may continue to flow due to linking attacks, in which sensitive data are revealed by linking the remaining attributes, such as those in published patient data, with other available data sources. This is referred to as a composition or intersection attack [3] [17]. Several anonymization techniques in PPDP have addressed data privacy concerns while preserving data utility. The goal of data anonymization is to reduce the threat of revealing personal information while preserving the possibility of using published data and causing uncertainty in identity inference or sensitive value estimation [50] [3]. However, information loss is inevitable when attempting to achieve a high level of privacy. In addition, the anonymization techniques may affect the use of patient data, resulting in imprecise or even impractical knowledge extraction via data mining. In data applications, balancing privacy and utility is crucial. In addition, nearly every technique leaves a question unanswered about whether anonymized data can be used effectively for data mining [1] [3].

The main contribution of this paper study is to propose an innovative technique that utilized a hybrid protection method to increase the utility of medical patients' data publishing while preserving privacy. This study aims to solve the problem of identity disclosure individuals or disclosing the sensitive value in medical patient's tables whilst preserving data utility. This research's accomplishments are summarised in the following points:

- The design of an innovative technique based on the UL technique in order to prevent attackers from

identifying individuals or disclosing sensitive values in medical patient tables. In addition, the proposed technique stroked a better balance between utility, information loss, and privacy. The utilized technique could partition information into horizontal and vertical partitions. In the vertical partition, firstly, the attributes were separated into more than one. Then, the similar attributes were further grouped in a subset in a fashion that designated every attribute to a subset. Therefore, the subset of each attribute was called a cell (a pair of attributes), and the combination of these yielded the column. In the horizontal partition, the table was divided into different subsets so that each tuple could only be assigned to a single subset. Every subset of these tuples was referred to as a bucket or an equivalence class.

- A hybrid protection method can address the deficit in other existing works in determining the amount of protection required to prevent personal information disclosure. Instead of the random procedure employed in other works to break the correlations between the attribute values, the lower and upper protection levels (LPL and UPL) are then utilized in each equivalence class to determine the values of the unique and identical attributes for data privacy protection while keeping data utility. The LPL and UPL determine the level of protection surrounding the attribute values, ensuring that an attacker cannot get the sensitive information required to identify the record owner during such periods. This work also relied on value swapping within UPL and increasing the number of fake tuples within LPL as a safeguard against attribute exposure to any attack. The innovative technique protects published data from disclosure while increasing data value.

This paper falls into six sections: Section I provides the introduction and highlights the crux of the issue. Background and highlights the axes of privacy-preserving data publication are reviewed in Section II. Section III reviews relevant studies, Section IV describes the flow of research procedures in this paper, and Section V presents the evaluation of performance analysis. Finally, Section VI provide the conclusion and discussion of the study.

II. BACKGROUND

One of the advantages of pervasive computing (ubiquitous computing) is the generation of large information volumes known as "big data". The explosion of information has made retrieving private and public information on individuals a major part of everyday life [1] [30]. Typically, companies, such as health care, maps, and education acquire data to meet legitimate needs. Most data can be unstructured or intricate; a considerable portion of this data has been generated by a number of sources, like records of business sales, sensors in the use of Internet of Things, medical records of patients in hospitals, social media, and images and video archives. Big data processing using traditional data processing applications is becoming difficult [2]. Nevertheless, considering the Internet co-dependency and the IS, i.e., information systems, this data can be susceptible to corruption, theft or individual privacy

violations [3].The ability to own data and extract new knowledge, which is known as data mining, is now considered a key competitive advantage [4].

Apart from the importance of extracting new knowledge (data mining), which is significant in many applications, an increasing concern has been focused on the privacy threats that emerge during data publishing for data mining operations, where numerous establishments need to publish data in different formats to extract new knowledge [4]. As a result, there has been a lot of focus on potential data privacy infractions as well as data exploitation; hence, effective data protection must be assured because failure may result in scenarios that can harm individuals and organizations [5] [6]. As a result, many organizations must choose between providing information and safeguarding their privacy in order to obtain this essential information [1]. This situation has motivated and prompted researchers to create a new research area known as privacy-preserving data publishing (PPDP) [7] [8]. This new research area is a sub-field of data mining that has gaining traction, which began as an encouraging model for providing first-hand solutions to address the current dilemma. Besides that, various methods have been developed for protecting information privacy, or wide-ranging policies have been imposed to safeguard sensitive data in PPDP [1] [3]. The primary goal of PPDP is to make a portion of these data available to all, from which it is utilized effectively for various tasks of data publication like application in future research and, at the same time, achieving the privacy of individuals' information [3]. Although it is helpful, information loss through protection methods is inevitable when attempting to achieve a high level of privacy. In addition, these protection methods in PPDP may affect the use of data, resulting in imprecise or even impractical knowledge extraction. Furthermore, the privacy-preserving data publication (PPDP) is dependent on three axes: 1) data forms 2) Privacy Vs. Utility, and 3) methods of adversary knowledge (Fig. 1). These three axes will be described in the subsequent subsections in detail.
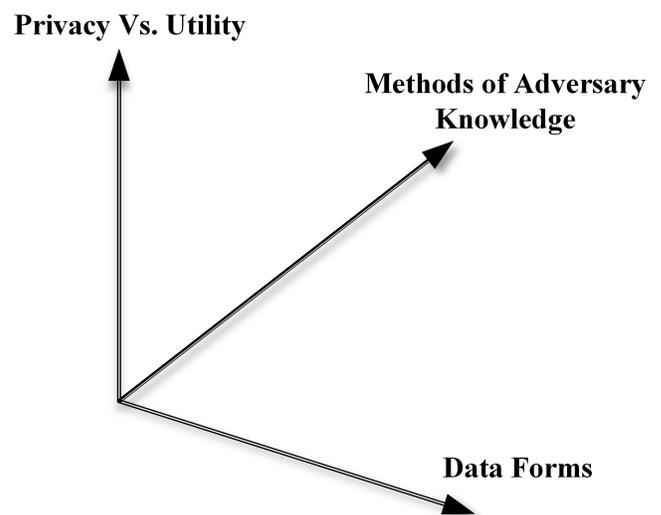
Fig. 1. Overview of Research Axes in PPDP.

A. Data Forms

Nowadays, the fast-growing number of Internet users and linked devices to ubiquitous computing have led to the generation of massive data volumes. It is essential to transform this gigantic, generated data into various forms for extracting valuable information embedded in massive data and providing many opportunities for advantages in different fields [2] [9]. Data forms, from which the types of data to be handled are selected, are seen to be a crucial component of the PPDP. The medical-patient database is one field where the deployment of big data can result in substantial changes. It can significantly raise the standard of patient care and provide insightful data that will help enhance patient outcomes, lower healthcare delivery costs, lower preventable diseases, and increase the overall quality of life.

Medical-patient databases are comprised of a wide range of records. Every record (i.e., row) signifies one client and includes a number of attributes specific to the selected client as shown in Table I [10] [11]. Such attributes are categorized into three categories [12] [13]: Identifying Attributes (IA), which distinguish explicitly the owner's records like his/her name, mobile number or the number of the driving license; Quasi-Identifier (QI) attributes, which denote the non-explicit attributes' sequence of individuals, including his/her age, gender, race, ZIP code, and date of birth as these attributes identify the record of the owner when these attributes are combined; and Sensitive Attributes (SA), which contain confidential data of individuals, such as diseases [14].

TABLE I. MEDICAL PATIENT DATABASE

Identifier (IAs)	Quasi-Identifier (QI)			Sensitive (SA)
Name	Age	Gender	Zipcode	Disease
Carl's	29	Female	462350	Disease
Abraham	22	Male	462351	Cancer
William	27	Male	150352	Flu
Linda	43	Female	462350	Heart Disease
Camila	52	Female	462350	Cancer

Much of the work done in PPDP is related to the static data publication form, with the idea of one record per person, and these records are presumed to be independent of various data publishers [10] [15] [16] [17]. However, the data form may be dynamic, i.e., it may be published consecutively while being changed by the same data publisher [18] [19] [20] [21]. The problem with static data publication forms is that it is possible for multiple organizations to publish a person's information at once [22]. In this situation, an attacker can use the published dataset and then, use a composition attack [23] [17] to change the dataset's privacy. An intruder commits a composition attack when he or she tries to find out who someone is by linking the different available attributes (QIs) in published available data to an external database to get sensitive information [23]. Therefore, the only way to protect people's privacy is to change these attributes in the available published data to hide the connection between a person and a specific value. This will stop these kinds of attacks and keep the public data usable [17].

B. Privacy vs. Utility

This aspect might be the most significant and fascinating aspect of the PPDP paradigm because any sanitized release of published data needs to address the trade-off between privacy and utility [1]. The common practice of sanitized data primarily depends upon certain guidelines and policies for restricting the publishable data types to achieve privacy. This common practice of sanitizing data is limited because it excessively falsifies data or demands a high level of trust in different scenarios of data-sharing impractically [24]. However, data privacy involves protecting private data from people who must not access this information and the individuals' capability to determine or interfere with the identity of individuals who can have access to their personal information. Developing protection tools and mechanisms for data publication is of is a major challenge so that the published data are practically usable and valid while the individuals' privacy is protected. A privacy policy is, therefore, a must to ensure that the sanitized release is safe from any attempts of intrusion while still being beneficial to the end-users. This means that there should be a balance conceived between these two notions, including utility and privacy [1].

One of the most prevalent and commonly utilised procedures in PPDP for providing privacy for individuals while retaining data utility is the anonymization technique of data before publication, which has previously been defined according to [25] as a group of certain protection methods, which aim to mitigate the risks of information disclosure for individuals, organizations or other businesses. The anonymization techniques use different protection methods and may be combined within the same technique, with the goal of causing uncertainty in identity inference or sensitive value estimation [25]. Some of these existing and most used protection methods are suppressing or generalizing [21] and perturbation [26] [27] [28] [29].

C. Methods of Adversary Knowledge

The assumption regarding the methods of adversary knowledge has led to the creation of alternative techniques and innovative protection methods. Although these techniques with protection methods provide a certain level of privacy preservation, it is crucial to examine the existing resources on adversaries regarding externally accessible data and different potential inferences for privacy-preserving of data. Privacy-preserving is classified into three categories based on the methods of adversary knowledge [24]:

The first category is that a certain privacy threat can occur when the attacker links the QI attributes within the published information with other available data sources. The attacker relies on the intersection of the datasets to exploit sensitive information since datasets are rarely isolated. This situation is known as a composition attack (i.e., the linking attacks), or an intersection attack [3] [17].

The second category is known as background knowledge attack. The attacker knows that the victim's record is in the released table or that the attacker has knowledge of some QIs for victims because the attacker is a close relative of the patient. The aim of the attacker is to link this knowledge for disclosure of sensitive attribute [1] [3]. The publishing table

is regarded as privacy-preserving when it effectively prevents the attacker from performing linkages successfully.

An uninformative principle is the aim of the third category. This published table must give the attacker very little extra information outside the specified background knowledge. When the attacker possesses a larger distinction between prior and posterior beliefs, this attack is called a probabilistic attack. The inability to categorize many QIs attributes into a published table and to keep it without change (or modification) leads to a probabilistic attack and the possibility of accurately extracting the attributes of a person [1] [17] [24] [29].

III. RELATED WORK

Data collectors collect a large amount of information from data owners and publish it so that data mining can provide a wide range of unprecedented potentials deemed necessary for providing meaningful information about data and improving the quality of medical patient data [2]. The data collected possibly hold the sensitive personal data of individuals. Thus, the goal of privacy-preserving data publishing (PPDP) is to release the data without publishing the private information about the owner of the data whilst preserving data utility in order to use them in any kind of medical or social analysis, etc. [48] [51] [52]. Differences in their concepts lead to differences in their protection methods. Binjubeir et al. [1] asserted that no general solutions could handle all privacy issues related to keeping sensitive information from unwanted disclosure. Hence, many techniques are used to provide privacy [53].

Despite the overlap between privacy and confidentiality in some contexts, when it comes to protecting people's privacy, there are certain ways they differ, especially related to their concepts and methods of protection. Confidentiality is seen as data-related, meaning that it is more about the data themselves, and it aims at protecting data from unauthorized access, alteration, or loss when transferred over a network [1] [54]. On the contrary, privacy has an additional "data owner-oriented" concept as it deals mainly with the data owners and aims to protect the private information of the data owners [55]. Hence, the current study uses PPDP as a way to keep sensitive data from being used illegally and to keep it safe from any threats. Consequently, there has been a lot of research on PPDP techniques over the past few decades.

The PPDP began with k-anonymity by Samarati and Sweeney [31]. Their work has been extended to cover various anonymization techniques like t-closeness [10], the l-diversity [32], (α, k)-anonymity [33], and Mondrian [34]. However, the aforementioned techniques have been susceptible to the composition attack [2] [23] because if two separately published tables satisfy a certain privacy principle, there will be no assurance that each pair's intersection for the equivalence classes of these two tables will be satisfying this same principle. Besides that, a higher dimensionality renders these techniques useless because the main record holders' identities are disclosed by combining the data with a public data (i.e., composition attack) [17] [35]. Readers can refer to [1] [36] [37] [38] for comprehensive studies on these techniques. Table II presents a summary of the PPDP protection techniques by offering the advantages and limitations of each technique.

Moreover, there are some techniques based on the dynamic data publication form [18] [20] [21] in addition to the static data publication form [10] [15] [16] [17]. The introduced techniques tackle composition attacks explicitly by utilizing an intersection of two or further sets of published data for uncovering individuals' sensitive information. In this anonymization process, which manages the dynamic data publication form, the data owner recognizes the entire versions of these published data and uses this information in these published versions for anonymizing the existing data set. For this study, the data owner has no knowledge of other datasets, which might be manipulated for the composition attack. Consequently, the data publications of various data custodians will be independent. However, the problem can be more challenging.

The most recent famous techniques for static and dynamic data publication form are the hybrid technique [35], e-differential privacy technique ($e-DP$) [39], slicing technique [29], merging [17], and UL method [3]. All techniques have endeavoured to create privacy-preserving by using different protection methods.

As a protection method against sensitive value disclosure, the hybrid technique in [35] uses a combination of sampling, generalization, and perturbation. While the generalized data have been used as a method of protection in the $e-DP$ technique. The $e-DP$ technique in [39] probabilistically constructs a generalised contingency table and then adds noise to the counts. By using differential privacy-based data anonymization, the $e-DP$ can give robust privacy assurance for statistical query response and protect against the composition attack [23]. According to [17] [35] [40] [41], utilising the $e-DP$ for composition attack protection can result in significant data utility loss during anonymization.

The disclosed slicing technique [29], which is regarded as an innovative data anonymization technique in PPDP, has garnered a lot of attention. The authors [17] [29] [3] presented a non-generalizable risk disclosure preventative protection method. As a result, because the attribute values are not generalised, slicing protects data privacy while also preserving data utilities. These techniques [17] [29] [3] partition the data horizontally and vertically. In the vertical partitions (i.e., attribute grouping), the extremely associated attributes can be grouped into specific columns with every resulting column containing the attributes' subset. In horizontal partitions (i.e., the tuple partition), these tuples can be grouped in specified buckets or equivalent classes (Table III). The adoption of various protective methods causes the relationship between separate columns to be broken. This solution protects the privacy of public records from the hazards of attribute and membership exposure. Furthermore, slicing is preferred for high-dimensional data anonymisation since it preserves more data utility than attribute value generalisation. It is thus preferable to formalise slicing for a fuller comprehension. Consequently, the slicing formulation has been followed as suggested by Li et al. [29].

A. Attribute Grouping

The microdata table T consists of a set of t, $t \in T$ and n the number of a attributes, where t is a tuple of T and t is represented as $t = (t[a]_1, t[a]_2...t[a]_n)$, where

TABLE II. A SUMMARY OF PROTECTION METHODS FOR PPDP TECHNIQUES

Techniques	Protection Methods	Strength	Weakness
K-Anonymization	When the values of the QI attributes are modified, it is harder for an attacker to figure out who a person is. At the same time, the released data remains as helpful as possible, and the K value is used as a measure of privacy	This technique protects an individual's identity while releasing sensitive information	Indirect attacks on k-anonymity, like homogeneity attacks, background knowledge attacks, and composition attacks, make it possible to figure out exactly what a person looks like. Also, the high dimensionality renders this technique ineffective
L-diversity	This technique works to treat the values of a specific attribute similarly, regardless of its distribution in the data, thereby resulting in the sufficient representation of SAs within each equivalence class	This technique attempts to preserve privacy by a sufficient representation of SAs within each equivalence class	This technique is subjected to skewness attacks, similarity attacks, and composition attacks. Also, the high dimensionality renders this technique ineffective
T-closeness	The SA distribution in any equivalence class should be similar to the distribution of the attribute in an overall table	This technique works to distribute SAs in any equivalence class similar to the distribution of the attribute in an overall table, which leads to preserving privacy	This technique can't protect the critical values of the records from a composition attack reliably and constantly. In addition, The high dimensionality renders this technique ineffective
Mondrian	Partitioning the domain space recursively into several regions, each of which contains at least k records. A set of QI values are generalized in each equivalence class	Getting an anonymous dataset	This technique can't reliably and always protect the important values of the records from a composition attack. Also, most classification tools don't work well with overlapping intervals because they make it hard to classify things
(α, k)-anonymity	This technique integrated two novel concepts: (α, k)- anonymization by sampling and generalization for independent datasets to protect against composition attack	This technique effectively protects privacy and preserves data utility	There is still more data loss

TABLE III. A PUBLISHED DATA BY SLICING

(Age, Gender)	(Zipcode, Disease)
(30, F)	(130350, ovarian cancer)
(23, M)	(130350, heart disease)
(28, F)	(130352, Flu)
(53, F)	(130350, heart disease)
(39, F)	(130352, Flu)
(60, M)	(130351, heart disease)

$t[a]_i \leq i \leq n$. In attribute partitioning, first, the attributes are separated into more than one; then, relevant attributes can be arranged in a specified subset, whereby each set can belong to a single subset only. Hence, the subset of each attribute is called a cell, and the combination of these yields the column. In the microdata table T, there are col columns, including $col_1, col_2...., colc$ satisfying $\bigcup_{i=1}^{c} col_i = a$ and for any $1 \leq i_1 \neq i_2 \leq col$, $col_{i,1} \cap col_{i,2} = \emptyset$. In these sensitive attributes, the sensitive attribute can be placed into the last-place position for an easy representation.

Definition 1 (cell): A cell represents one pair of attributes like (Gender, Age), where any cell $C_{col,E}$ is identified by the number of columns col_i and number of an equivalence class E_e. For example, in Table III, any cell in column (Age,

Gender) is identified by col_i and E_j, where $1 \leq i \leq col$ and $1 \leq i \leq E$ and the first equivalence class is consisting of tuples $t = t_1, t_2, t_3, t_4$.

B. Tuple Partition

The goal of tuple partition is to generate different subsets of T in a manner that each tuple can only be assigned to only one subset. Each subset of tuples is known as a bucket or an equivalence class. Assume there are E equivalence classes, $E_1, E_2, ...E_e$ then, $\bigcup_{i=1}^{e} E_i = T$ for any $1 \leq i_1 \neq i_2 \leq e$, $E_{i,1} \cap E_{i,2} = \emptyset$.

C. Problems of Slicing

Slicing depends on the application of attribute grouping and tuple partition T. In Table III, by measuring the associations (similarity) among the attributes, the attribute group is applied, where very associated attribute values can be sorted into specified columns and uncorrelated attributes can be aggregated into other columns. The attribute partition can be characterized by Gender, Age, Disease, Zip Code, whereas the tuple partition is applied by grouping tuples into an equivalence class $\{t_1, t_2, t_3\}$, $\{t_4, t_5, t_6\}$. The central part of this tuple partition involves grouping all tuples that contain identical values in a similar equivalence class or it can be close to one

another, thereby making it easier to breakdown uncorrelated attributes, and check whether an equivalence class satisfies I-diversity [1] [3] [34].

For slicing, the values of attributes are randomly permutated between the uncorrelated attributes for breaking the association among distinctive columns, whereas the attributes in columns that are highly correlated remain unchanged. However, the aspect of it remains an open question, i.e., "Does randomness always protect the identities of individuals from disclosure?" Slicing can have an impact on data utility and privacy, such as randomly permuting attribute values in each bucket, which increases the likelihood of creating bogus tuples, which reduces the utility of the released microdata. Furthermore, bogus tuples can easily cause various problems and wrong results in data mining process challenges. An attacker can learn about the implemented anonymization technique by analysing the spurious tuples in the published table, potentially breaching the privacy of public data [42] [3].

Hasan et al.[17] developed the merging technique to secure personal identification from disclosure. This been regarded as an extension of slicing. Merging's primary purpose is to preserve privacy in many separate data releases by employing cell generalisation and random attribute value permutation to seperate connection between various columns. Regarding privacy risks and data utility, the merging technique conserved data usefulness while posing minor privacy hazards because of increased false matches in the released datasets. Nonetheless, the merging technique's significant weaknesses are the randomised permutation way for the attribute values to breach the relationship between the columns and the increase in false matches for unique attributes. However, there will be a large number of matching buckets (more than the initial tuples), resulting in utility data loss, and could generate inaccurate and infeasible knowledge acquisition from data mining operations cite43 [44]. As a result, the main reasons for revealing people's identities are unique attributes or the ability of some cells in the tuple to match with cells in other tuples in the same equivalence class, allowing precise extraction of a person's attributes [1] [17] [29].

BinJubeir et al. [3] developed the UL technique as an efficient means of identifying the level of data protection required and selecting the best way to accomplish that level while keeping data utility. A lower level of protection, i.e., (LPL) and an upper level of protection, i.e., (UPL) can be employed to overcome these unique attributes with an identical data presence for data privacy protection. The unique attribute values are overcome by LPL, whilst the high identical attribute values are overcome by UPL. The LPL and UPL variables determine the level of protection surrounding the attribute values, ensuring that an attacker cannot access the sensitive information required to identify the record owner within such a time frame. The UL technique also makes use of value swapping to reduce the danger of attribute disclosure and increasing l-diverse slicing. Table IV illustrates the previous works which has been discussed of PPDP techniques for multiple independent data publishing. Some note that PPDP techniques are typically used to determine the level of privacy protection and information loss [56]. The two essential principles discussed here are privacy protection from any attack and data loss. The privacy preservation level refers to the degree of

difficulty of estimating original data from perturbed data [57]. On the other hand, the information loss is a situation in which a significant portion of information of the original data set is lost after data anonymization.

The main contribution of this paper is to suggest an innovative technique that uses a hybrid protection method to enhance utility while protecting the privacy of medical patients' data. The new technique proposed in this work is expected to keep data private while making patient data publishing more useful.

IV. Flow of Research Procedures

The flow of the research procedure is described in this section. It talks about the stages and methods that were used in this study to reach the research goals. As mentioned in the related work, many techniques have been proposed to address all privacy issues concerning protecting sensitive information from uninvited disclosure while preserving the utility of the data. However, there are still ways to enhance the utility of data while preserving user privacy. In essence, this study focuses on designing an innovative technique-based on UL technique that uses a hybrid protection method to enhance utility while protecting the privacy of medical patients' data. Besides, this work used the UPL and LPL methods for anonymisation, which is more effective in determining the amount of protection required. UPL and LPL are choosing cell values that help identify disclosure and break the link between them by using a hybrid protection method (see Stage 3: protection methods) to keep data private while making patient data publishing more useful. This study aims to get a certain level of privacy while ensuring that as little information as possible is lost during data mining. So, this study aims to ensure that composition attacks and background knowledge attacks are less likely to happen when different independent hospitals release anonymous patient data while keeping the data intact. The flow of the research procedure consists of three main components, as depicted in Fig. 2. These three components are described in depth in the subsections that follow.

A. Research Gap

In related work, an analysis of how published data can be kept private is given. The main problem that has been found is when hospitals share patient information that could help them improve their efficiency and achieve their goals for the future. Sensitive Attribute (SAs) may still flow due to linking attacks wherein sensitive data may be revealed by linking the QI attribute in the published data with other available data sources. This situation is known as a "composition attack" or "intersection attack". Also, many anonymization techniques fail to show a better balance of usefulness and privacy before any data product is made public. The criterion for evaluating the efficiency of the anonymization technique is the capability of data privacy preservation by decreasing the vulnerability of revealing people's data and protecting the likelihood of the published data being used [3] [50] [17]. Previous techniques [29] [17] [34] [35] [39] recurrently resort to using protection methods, such as suppression and generalization, randomization, and/or combined. This work proposes designing an innovative technique based on the UL technique that uses a hybrid protection method to enhance utility while protecting

TABLE IV. A SUMMARY OF PPDP TECHNIQUES

Techniques	Protection Methods	Strength	Weakness
hybrid	This technique combines sampling, generalization, and perturbation by adding the Laplacian noise to the count of every SA value in each equivalence class	The proposed work reduces the risk of composition attacks and preserves data utility	There is still more data loss
$(e-DP)$	First probabilistically generates a generalized contingency table and then adds noise to the counts	$e-DP$ provides a strong privacy guarantee for statistical query answering and protection against a composition attack by differential privacy-based data anonymization	This technique is subjected to skewness attacks, similarity attacks, and composition attacks. Also, the high dimensionality renders this technique ineffective
Slicing	This technique uses vertical partitioning (attribute grouping), horizontal partitioning (tuple partition), and its sliced table should be randomly permutated	Slicing provides data privacy by randomly permutated of data and preserves data utilities that is devoid of generalization	Random permutate for attribute values are led to creating invalid tuples which will negatively affect the utility of the published microdata
Merging	The primary aim of merging approach is to preserve privacy by increasing the false matches in the published datasets and it uses vertical partitioning, horizontal partitioning, and its sliced table should be randomly permutated	Getting an anonymous dataset and preserves data utilities	The major drawback of merging is the random permutation procedure and increasing the false matches in the published datasets
UL method	The (LPL) and (UPL) can be used to determine the level of data protection needed and employed to overcome unique attributes and an identical data presence for data privacy protection whilst preserving data utility	This technique effectively protects privacy and preserves data utility	There are alternative protection strategies that may preserve data utility and privacy.

the privacy of medical patients' data. The goal of the hybrid protection method is to discover the peculiar features to swap between them rather than a random way of separating the relationship amongst the attribute values used in other existing works to enhance the privacy of published patients' data and keep more data utility.

B. Research Methodology (Design The Innovative Technique)

The overall methodology for designing the innovative technique, as shown in Fig. 2, comprises three stages for protecting the published patient data from unsolicited disclosure. Meanwhile, published patient data remains as useful as possible. Fig. 2 illustrates the proposed innovative technique for patient data protection, at the same time preserving the utility of the data. The following is the discussion of these three stages.

Stage 1 is preparing the dataset. The datasets stage aims to initialize the dataset and measure the correlation between attributes. To evaluate the experiments with other existing works, a medical patient database was used for the experiments [45]. This dataset is the standard machine learning dataset known as the "Adult" [46] has been changed and added one new column called disease. The main reason for adding this column is to mimic the medical patient dataset. Ronny Kohavi, together with Barry Becker extracted and congregated this dataset from the 1994 United States Census Bureau. Accordingly, this dataset is made up of fifteen QI attribute with 48,842 tuples.

In the dataset initialization process, independent patient datasets were required for the simulation of the existent medical patient data publishing case, particularly in a case in which such datasets are separately published by various medical organizations that have similar records. However, the pitfall of this proposition lies in the fact that the data of an individual is often published by many medical organizations [22]. Under such conditions, any intruder can initiate a composition attack or background knowledge attack [23] [17] on such published patient datasets just to alter the privacy of the dataset. In the process of dataset initialization, this independent dataset is taken from a database of medical patients.

After the initialization process, the correlation between attributes is measured, where the initialization of dataset generates various medical patient datasets to simulate the actual independent medical patient data publishing scenario. Each medical patient dataset should be treated as microdata table T. In a case in which the microdata table T applied possess a_i attributes, where $i = 1, 2, ...$. The strength of the correlations between pairs of attributes can be computed using several methods [17] [3]. Because most attributes are categorical, the most suitable method for the estimation of the correlations between pairs of attributes is the mean square contingency coefficient (MSCC). The MSCC is a chi-square useful measure of the correlation between two categorical attributes. The value of this coefficient r ranges from $[0, 1]$. If there is a perfect relationship between the two attributes, it would be preferable to have the measure of the association have a value of 1.
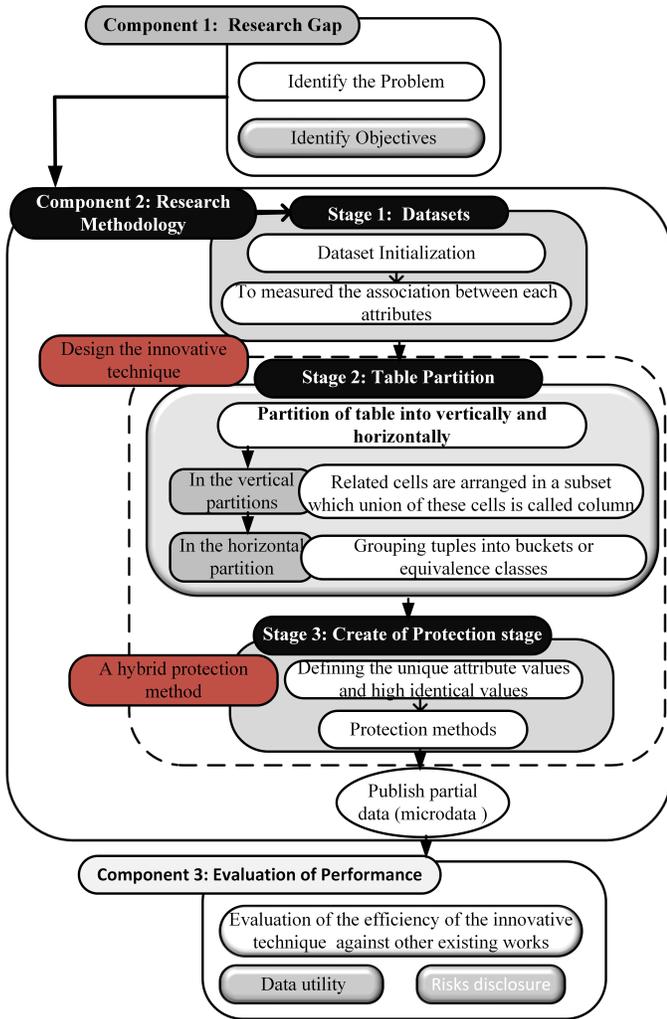
Fig. 2. Flow of Research Procedures.

Otherwise, these measures differ in their maximum value. In case of no relationship between the two attributes, the measure of association has a value of 0. The MSCC between a_1 with value domain $\{v_{11}, v_{12}, ...v_{1d1}\}$ and a_2 with value domain $\{v_{21}, v_{22}, ...v_{2d2}\}$, and their domain sizes are d_1 and d_2, respectively. The MSCC between a_1 and a_2 is defined as follows [17] [3]:

$$r^2(a_1, a_2) = \frac{1}{min\{d_1, d_2\}} \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} \frac{f_{ij} - f_i * f_j}{f_i * f_j} \quad (1)$$

, where $r^2(a_1, a_2)$ is the MSCC between a_1 and a_2 attributes; $f_{i.}$ and $f_{.j}$ refer to the occurrence fractions of the v_{1i} and v_{2j} in the data, respectively; and f_{ij} is the fraction of cooccurrence of v_{1i} and v_{2j} in these data. Therefore $f_{i.}$ and $f_{.j}$ are the marginal totals of $f_{ij} : f_{i.} = \sum_{j=1}^{d_2} f_{ij}$ and $f_{.j} = \sum_{i=1}^{d_1} f_{ij}$. $0 \leq r^2(a_1, a_2) \geq 1$.

Stage 2 deals with vertical and horizontal fractionalization of table. The dataset in the table is vertically and horizontally

fractionalized, depending on the computation of correlation r for respective attributes pairs. This phase is aimed at categorizing similar attributes according to the degree of their inter-attribute connections that are suitable for utility and privacy. Concerning the utility of data, closely connected attributes are categorized in order to make warrant that their inter-attribute connections are maintained. Notwithstanding, as regards privacy, detection of vulnerabilities is greater as a result of the categorization of unrelated attributes in comparison with the categorization of more connected attribute values, which makes them highly identifiable. In order to guarantee a higher level of protection, it is important to dissolve the connection that exists among the unrelated attributes [29]. The microdata table T consists of a set of t, $t \in T$ and n the number of a attributes, where t is a tuple of T and t is represented as $t = (t[a]_1, t[a]_2...t[a]_n)$, where $t[a]_i \leq i \leq n$. In the vertical partition, firstly, the attributes are separated into more than one, then, the similar attributes are further presented in a subset in a fashion that designates every attribute to a subset. Therefore, the subset of each attribute is called a cell (a pair of attributes), and the combination of these yields the column. In the microdata table T, there are col columns, including $col_1, col_2...., col^c$ satisfying $\bigcup_{i=1}^{c} col_i = a$ and for any $1 \leq i_1 \neq i_2 \leq col$, $col_{i,1} \cap col_{i,2} = \emptyset$. Furthermore, QIs, SAs, and all other attributes presented in columns $col_i, 1 \leq i \leq n$ They are clustered in n columns denoted by col_n, upon which the size of sensitive column col^c is not dependent. In some cases, the number of attributes a_s in the sensitive column col^c may be predetermined to be c. The col^c is determined in size by the use of parameter c, mathematically presented as $|col^c| = c$, in a case where $c = 1$, $col^c = 1$ also. That is, $col^c = \{S\}$. In a scenario where $c = 2$, the procedure is said to be equal to bucketization. In the case where $c > 1$, $|col^c| > 1$. QI attributes are contained in the sensitive column col^c. So as to ease the discussion in this study, the sensitive attribute a_s is focused on as one. Assuming the column in which a_s is contained is last column col^c. The column is as well referred to as sensitive column, as presented in Table V. In a case where several sensitive attributes are contained in the data, their individual or collective distribution may be employed [14]. Attributes (cells) that are highly related are put together in a column in vertical partitions, while unrelated are as well put in separate columns in a way that individual attributes a_i becomes designated to one subset. As shown in Table V, col_i columns $\{col_1, col_2...., coln\}$ contain all attributes a_i. In Table V, the three partitions for the Col_i columns are presented:

(1) T^* contains all columns with highly correlated attributes col^*, where $col^* = \{col_1^*, col_2^*, ...col_i^*\}$, and $col^* \in T^*$.

(2) T^{**} contains all columns with uncorrelated attributes col^{**}, where $col^{**} = \{col_1^{**}, col_2^{**}, ...col_i^{**}\}$, where $col^{**} \in T^{**}$.

(3) T^c contains columns with SA col^c when a single SA exists, and its SA is placed in the last position for easy representation, where $col^c \in T^c$ and $(col^* \cap col^{**}) \cap T^c = T$.

K-medoid clustering algorithm that is otherwise called partitioning around medoids algorithm (PAM) [47], is employed in the presentation of similar attributes in columns in a manner that designates each attribute to a column is used to organize the similar attributes into columns for each attribute to be part of a column. This algorithm guarantees the resolution of

TABLE V. EXAMPLE OF PARTITION THE TABLE T INTO THREE COLUMNS

T^* contains all columns with highly correlated attributes.		T^{**} contains all columns with uncorrelated attributes.		T^c contains a column with sensitive attributes.
col_1^*	col_2^*	col_1^{**}	col_2^{**}	col^c
(a_1, a_2)	(a_3, a_4)	(a_5, a_6)	(a_7, a_8)	(a_s)

every attribute in the form of a point in the cluster space, while the inter-attribute disparity in the clustered space is represented thus: $d(a_1, a_2) = 1 - r^2(a_1, a_2)$, that falls within the range of 0-1. However, the disparity amongst affiliated data points becomes less within the clustered space if two attributes are firmly correlated. After determining the disparity amongst affiliated between the related data points, the k-medoid method arranges related attributes in a subset called a cell, and the combination of these yields the column $(T^*, T^{**}$ and $T^c)$.

In horizontal Partition, the table is divided into different subsets so that each tuple can only be assigned to a single subset. Every subset of these tuples is referred to as a bucket or an equivalence class. Assume there are E equivalence classes, $E_1, E_2, ... E_e$ then, $\bigcup_{i=1}^{e} E_i = T$ for any $1 \leq i_1 \neq i_2 \leq e$, $E_{i,1} \cap E_{i,2} = \emptyset$. In addition, all tuples containing similar values are categorized into categories referred to as bucket or equivalence classes. Here, every individual is joined to 1 specific sensitive value in a way that makes it impossible for an attacker to penetrate the SA values of an individual where the probability is greater than 1/l. The tuples were categorized by the Mondrian [34] algorithm. They are separated in the equivalence classes, in the absence of generalization attributes, according to the top-down technique.

Stage 3 is the protection. The table partition stage generated the partition of microdata table T into partitions vertically and horizontally partitioned. The aim of the table partition stage is for all attributes to be clustered into columns (including both QIs and SAs) to prevent the unauthorized disclosure of an individual's identity by altering attributes (QI values) so that the connection that exist among the individual and specific values are hidden while ensuring that the data published is used through the application of protection methods. In this stage, the hybrid data protection method will provide robust patient data privacy while increasing medical data publishing utility for microdata table T partitioned using the innovative technique based on the UL technique via two steps, namely defining the unique attribute values, high identical values, and protection method.

1) Defining the Unique Attribute Values and High Identical Values: When it comes to a hybrid protection method, the magnitude of the connection among attributes is chiefly measured by the correlation coefficient r. Formula 2 illustrates the manner in which UPL and LPL attempt to define the unique attribute values and high identical values through the extraction of two kinds of cell values: (1) the values of the exclusive (unique) cells and (2) high identical cell values in T^{**}. Each cell that possesses unique values and such values are within the range of $0.0 < LPL \leq \Theta$ are determined by the LPL. For such attributes, the r value is usually hovering around 0 but not equivalent to 0. In a similar vein, the UPL determines those cells that possess numerous similar attributes with values that are in the range of $\Theta \leq UPL < 1.0$. r

value for such attributes usually hover around 1 but is not equivalent to 1. Assuming those cells possess high r value in T^{**}, that will imply that the possibility of cells are in the same equivalence class. A hostile party has a higher degree of certainty around the SA when such cells are linked to other cells in T^*, thereby resulting in violation of privacy. In the remaining cells, the attributes and membership remain secured since they appear in multiple equivalence classes. In addition, the remaining cell categories need to be greater than a given limit, i.e., it should have a diversity value that is greater than or equal to 2 in respective equivalence classes, for the proposed privacy goal to be achieved.

$$(UPL \text{ and } LPL) = \begin{cases} \overline{C_{col,E}} = \Theta \leq UPL < 1.0 \\ \underline{C_{col,E}} = 0.0 < LPL \leq \Theta \end{cases} \quad (2)$$

It is the goal of UPL and LPL to discover the collection of unique cells and high identical values for cells from T^{**}, that are assumed to be known to intruders into these periods: $0.0 < LPL \leq \Theta$ and $\Theta \leq UPL < 1.0$. The attributes that are for discovering within this period are referred to as the protection attributes. Protection rate, represented by $theta$, refers to values that have been initially tagged to be discovered. Typically $theta$ is in the range of $1\%-10\%$, implying that the fraction of protection qualities will be smaller than one.

Definition 2 (Matching Buckets): Assuming col^{**} represents the columns, and $col^{**} = \{col_1^{**}, col_2^{**}, ... col_n^{**}\}$, and $col^{**} \in T^{**}$. Let t^{**} represent a tuple, and $t^{**}|col_i^{**}|$ represent the col_i^{**} value of t^{**}, then let E^{**} represent an equivalence class in microdata table T^{**}, and $E^{**}|col_i^{**}|$ stand for the multiset of col^{**} values in equivalence class E^{**}. E^{**} denotes a matching bucket of t^{**} if for all $1 \leq i \leq col^{**}$, $t^{**}|col_i^{**}| \in E^{**}|col_i^{**}|$.

2) Protection Method: This study's hybrid protection method guarantees the privacy criteria in every equivalence class. In order to enhance utility as well as individual privacy in the suggested innovative technique, the connection amongst unique attributes as well as cells that possess similar identical values are eliminated through two steps.

- Creation of Fake Tuples
  During the protection method step, a random permutation in a bucket may not be shielded from attribute or membership disclosure because permutation increases the risk of attribute disclosure rather than ensuring privacy [42]. In addition, increasing the false matches as a protection method in both unique attributes and high identical values, this method may generate a slight fraction of these fake tuples; however, this can result in countless matching buckets (i.e., more than the number of original tuples), leading to a huge data utility loss, producing erroneous or unfeasible extraction of the knowledge via operations of data mining [43] [44] [3]. Therefore, the fake tuples are used in a bucket as a protection method for all unique attributes (LPL) only. A few fake tuples do not change how useful the published patient dataset is, but they make it more likely that false matches will be found during a composition attack on the published

table [17] [29]. As a protection method, n fake tuples with similar QI values are made and given sensitive values based on how sensitive values are spread out in the original dataset. The main goal of creation of fake tuples is to obtain the anonymized table T.

- Swapping or Generalization of Attributes
  The protection method for UPL attributes is attribute swapping or generalisation. This study's innovative technique ensures the privacy criterion in each equivalence class. To boost diversity and personal privacy, rank swapping is utilised to break the relationship between attributes with high identical values. Attribute swapping modifies tuple data with high identical values (UPL) by switching the values of these characteristics across record pairs in a defined proportion of the original data. When it is not possible to swap attributes, the attributes must be generalised. The primary purpose of swapping or generalising attribute values is to create the anonymized table T, which has no nonsensical combinations in the record (invalid tuples) while satisfying the l-diverse slicing.
  Definition 3 (Attribute Generalisation): Let T^{**} be part of microdata table T, and a_i^{**} be a QI attribute set in T^{**}. Generalisation replaces the QI attribute values with their generalised version. Let d_i^{**} and d_j^{**} be two domains with dimensional regions $\{d_{i1}^{**}, d_{i2}^{**}, ...d_{in}^{**}\}$ and $\{d_{j1}^{**}, d_{j2}^{**}, ...d_{jn}^{**}\}$, respectively, where $\bigcup_{d_{in}^{**}} = d_i^{**}$ and $d_i^{**} \cap d_j^{**} = \emptyset$. If the values of d_j^{**} are the generalisation of the values in domain d_i^{**}, we denote $d_i^{**} < d_j^{**}$ (a many-to-one value generalisation procedure). Generalisation works according to a domain generalisation hierarchy, which refers to a collection of domains that is ordered according to relationship $d_i^{**} < d_j^{**}$ (see Fig. 3).
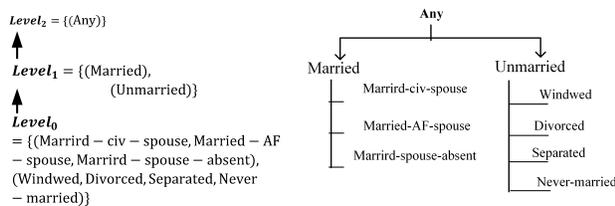
Fig. 3. Example of the Domain (Left) and Value (Right) Generalization Hierarchies for the Marital-Status(MS) Attributes.

In Fig. 3 (right), the likely domain generalisation hierarchy for marital-status (MS) attributes is described. At lower levels in the generalisation hierarchy for (MS) attributes, generalisation is not used. Nonetheless, at the top of the hierarchy levels, the MS tends to be more general. A singleton is a maximal domain level element that denotes the likelihood of values to be generalised in every domain to a single value.

C. Evaluation of Performance

This study describes maintaining privacy as minimizing disclosure of information on individuals. The usefulness of the data means to what extent we can use the sterile medical patient dataset for intensive analyses. For instance, by suppressing each QI, a medical patient dataset can be generalized. Maximum privacy is provided in this manner, but the information obtained is useless. Finding a good balance between privacy and utility is necessary because the published datasets (sanitized) must permit tasks related to data mining operations for search and analysis. As a result, the usefulness of data in the published medical patient dataset is assessed by how well statistical and aggregate data are used. The ability to protect data privacy by lowering the risk of disclosing personal information and maintaining the potential use of published data is the criterion for judging the effectiveness of the anonymization technique [3] [17] [1].

V. COMPARISON OF EVALUATION

Python was used to implement this experiment. The independent medical patient datasets were the experimental prerequisite to complete the experimentation of the actual independent data publishing setup. The independent datasets, known as the medical-patient-dataset, were pulled from the medical patient dataset [45], which contained eight QI attribute values: marital status (categorical, 7), relationship (categorical, 6), gender (categorical, 2), age (continuous, 74), work class (categorical, 8), salary (categorical, 2), disease (categorical, 16), as well as occupation (categorical, 14).

Each dataset contains 4K tuples chosen at random, with the remaining 8K tuples being used to generate an overlapping tuple pool and to check for potential composition attacks. By injecting 100, 200, 300, 400, and 500 tuples into the medical-patient-dataset, five copies of the remaining tuple pool were created, yielding datasets of the following sizes: (4.1K), (4.2K), (4.3K), (4.4K), and (4.5K) for the medical-patient-dataset.

In the experimental analysis, the static data publication form might be employed. There are two basic approaches for making the published medical patient dataset available. The first way is an interactive setting in which the data collector computes a function on the medical patient dataset in order to answer the data analyser's inquiries. The second way is a non-interactive setup in which the medical patient dataset is sanitized and then released [3]. The experiment was designed to evaluate non-interactive privacy settings; however, the majority of the work was done on differential privacy [39], which was consistent with interactive settings because medical patient datasets were commonly known to be made public. As a result, for the experiment on differential privacy, which was noted in [39], the noninteractive mode was chosen. Furthermore, the merging, $e-DP$, hybrid, UL, and Mondrian techniques yielded the quasi-identifier equivalence class as k-anonymity [16].k = 6 was chosen to build an equivalence class, where L-diversity was also provided as 6. The primary goal of L-diversity is to protect privacy by increasing the diversity of sensitive values. The Laplacian noise in a differential privacy equivalence class was added to the sensitive values' count [49] with e= 0.3 for the e-differential privacy budget.

The experiments on the medical patient database were performed in two parts. The first part was designed to evaluate the effectiveness of the innovative technique in data utility preservation by comparing it to other existing works. The second part was designed to assess the innovative technique

to see how well it can fight and prevent composition attack occurrence. The innovative technique's effectiveness was tested by relating it to the effectiveness of similar techniques, like hybrid [35], merging [17], $e - DP$ [39], UL method [3] and Mondrian [34] techniques, in the non-interactive privacy settings. The experimental results showed that the innovative technique provided privacy protections against the considered attacks while preserving data utility.

A. Data Utility Comparison

This experiment measured the data utility obtained from the distortion ratio (DR). The DR in published medical patient dataset can be evaluated by different methodologies [21] to quantify the anonymisation outcome on the overall distortion data. The generalised distortion ratio (GDR) is a suitable measure for estimating the DR [44] used to quantify the anonymisation outcome on the overall distortion data.

The swap and generalise method was used to break the association of the attributes because the majority of these attributes were categorical. For any two categorical attributes $(a_1^{**}, a_2^{**} \in T)$, where t is its taxonomy tree and a node p in t is used to swap or generalise the attributes, the DR with p is defined as follows:

$$DR(a_1^{**}, a_2^{**}) = \begin{cases} 0, a_1^{**} = a_2^{**} \\ \frac{|common(a_1^{**}, a_2^{**})|}{|N|}, a_1^{**} \neq a_2^{**} \end{cases} \quad (3)$$

, where $|N|$ denotes the set of all the leaf nodes in t and $|common(a_1^{**}, a_2^{**})|$ is the set of leaf nodes in the lowest common tree of a_1^{**} and a_2^{**} in t.

Fig 3 denotes the taxonomy of the marital-status (MS) attribute; if the values of a_1^{**} and a_2^{**} are in the same rank group and have no nonsensical combinations, then their swap values are equal, and the DR is 0. Moreover, if the values of a_1^{**} and a_2^{**} are not in the same rank group or have any nonsensical combinations, then, their generalized values are equal to $\frac{|common(a_1^{**}, a_2^{**})|}{|N|}$, and the DR is equal to $\sum_{j=1,k=1}^{n,m} d_{j,k}$, where $d_{j,k}$ represents the attribute's distortion of a_j^{**} of the tuple t_k.

The distortion ratio (DR), also known as data utility, is a proportional measure that compares the amount of distortion in a generalised medical patient dataset to the amount of distortion in a fully generalised medical patient dataset. It is possible to determine the value of the data by subtracting the DR from Equation 4 shown below [21]:

$$Datautility = (100 - DR)\% \quad (4)$$

Fig. 4 displays the data utility experimental results based on data loss on the medical-patient database. The innovative technique in Fig. 4 had a protection rate (θ) of 5% using LPL and 95% using UPL. The assessment of the innovative technique, done through its comparison with hybrid [35], merging [17], $e - DP$ [39], UL method [3], and Mondrian

[34] techniques revealed that the data utility obtained by the innovative technique was higher than that of all the known works. The merging technique had N fake tuples with the same QI values as in the original table. The sensitive values were assigned to them based on the sensitivity value distribution in the initial dataset. Hybrid, $e - DP$ and Mondrian techniques used the generalization procedure as a protection method. Therefore, these techniques resulted in more data loss than the innovative technique.The innovative technique employs selective generalization within the cell when satisfying the privacy requirements is essential; hence, more data utility is preserved.
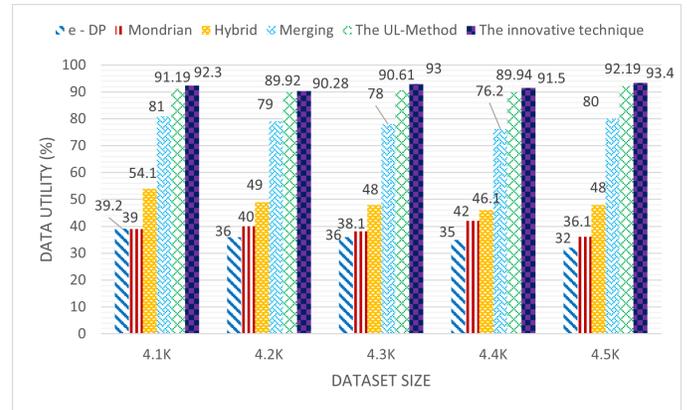
Fig. 4. Data Utility on the Medical-Patient Dataset (Protect Rate (θ) of 5% using LPL and 95% using UPL).

B. Measuring Risks

The measurement of disclosure risk in microdata during a composition attack is covered in this section. A composition attack occurs when an intrusive party, especially one knowledgeable about some of the QI values, attempts to identify a specific person in the microdata by linking several readily accessible records to an external database to disclose restricted information [8] [3]. As a result, gauging disclosure risk is essentially measuring the rareness of a cell in microdata publishing.

Medical patient dataset publishers should strive to measure the risk disclosure of PPDP outputs to ensure privacy preservation. This step is key in defining the level of protection needed. Therefore, differentiating the risk disclosure measures is important because the quantity to be measured must not depend on how the data representation method is selected. According to the works done previously, the risk disclosure can be quantified by determining the proportion of the true matches to the total matches, as expressed in Equation 5.

$$Disclosure\ risk\ ratio\ (DRR) = \frac{Matched\ records}{Total\ records} X100\%$$
$$(5)$$

Fig. 5 shows the experimental result for the disclosure risk ratio (DRR). DRR defines the adversary confidence level followed to elucidate the sensitive values in the medical-patient database. The $e - DP$ technique [39] revealed the least privacy risks compared to the innovative technique and other available

approaches. The $e - DP$ technique achieved approximately 0.63% disclosure risk ratio privacy risk for medical-patient database when k = 6, l = 6 for size of 4.5K. Based on the proposed solution [39], it probabilistically generated a generalised possibility table and added noise to the total. The $e - DP$ offered high privacy assurance and protection opposed to composition attack by differential privacy grounded data anonymization [23] [3], as shown in the results. It was observed by [35] [40] [41] [17] [3] that using $e-DP$ to protect against composition attacks generates a significant amount of data utility losses during anonymization, confirming the result discussed in Fig. 4.

The hybrid technique [35] generated a lower probability of sourcing the end-user's private data than Mondrian technique [34] and merging technique [17]. The hybrid technique achieved approximately 1.55% disclosure risk ratio (privacy risk) for medical-patient database when the K= 6, l = 6 for size of 4.5K.

Compared to the innovative technique, the UL method and the innovative technique decreased the likelihood of composition attacks on the released medical patient datasets than the hybrid [35], merging [17], and Mondrian [34] techniques by disabling the unique attributes and high identical attribute values by UPL and LPL, and providing multiple matching cells in each equivalence class, leading to protection against identity disclosure. The UL method and the innovative techniques achieve approximately 1.5% disclosure risk ratio when the K= 6, l = 6 for size of 4.5K. Meanwhile, the innovative technique enhanced the data utility (Fig. 4) through the use of hybrid protection method. Increasing the false matches for a unique attribute in LPL and value swapping for high identical values in UPL helped to enhance the data utility and guarantee a lower risk of attribute disclosure.
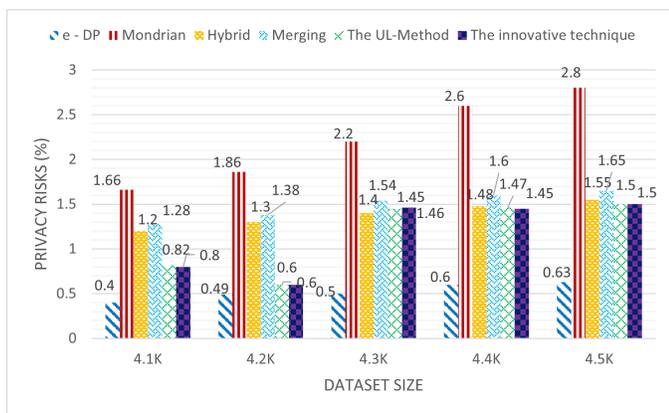
Fig. 5. Privacy Risk Medical-Patient-Database (k = 6, l = 6).

VI. CONCLUSION AND DISCUSSION

This paper presented an innovative technique using a hybrid protection method for utility enhancement while preserving data privacy for medical patient dataset; and for limiting the prospect of popular composition attack when the independent medical organizations cannot coordinate prior to medical patient dataset publication. The experiment showed that the innovative technique could satisfy the requirements of privacy after intersecting the separately published medical patient datasets. By contrast, many existing techniques reduced the published medical patient data utility due to the protection method used such as generalization and perturbation. This work, however, introduced a robust hybrid protection method by finding the unique attribute values (LPL) and high identical attribute values (UPL), then creating a fake tuple for the unique attribute and swapping values for the high identical attribute to decrease the attribute disclosure risk and ensure attainment of l-diverse in the published microdata table against the composition attacks. The model's effectiveness lies in the selection of specific attributes to enhance the privacy of published data and maintain data utility. The experimental findings revealed that the introduced method in this study could provide greater data utility compared with the state-of-the-art techniques. The achieved performance using our innovative technique helps researchers, decision-makers, and experts benefit from the published medical patient dataset to extract knowledge that may be used for disease prevention, medical decision-making, and many other areas of medical organizations. Similar to other scholarly research, this study leaves ample room for additional works to address its limitations and to expand upon its foundation to focus on adding or replacing another new protection methods to the innovative technique or extending some stages of the innovative technique to increase data utility and decrease risk disclosure. Moreover, the effectiveness of the innovative technique has been tested against composition attacks and background knowledge attacks, and by using the medical patient dataset; thus, it is important to test its performance against different attacks and by using different types of datasets.

REFERENCES

[1] M. Binjubeir, A. A. Ahmed, M. A. Bin Ismail, A. S. Sadiq, and M. Khurram Khan, "Comprehensive Survey on Big Data Privacy Protection," IEEE Access, vol. 8, pp. 20067–20079, 2020, doi: 10.1109/ACCESS.2019.2962368.

[2] M. Binjubeir, M. A. Ismail, S. Kasim, H. Amnur, and Defni, "Big healthcare data: Survey of challenges and privacy," International Journal on Informatics Visualization, vol. 4, no. 4, pp. 184–190, 2020, doi: 10.30630/joiv.4.4.246.

[3] M. BinJubier, M. Arfian Ismail, A. Ali Ahmed, and A. Safaa Sadiq, "Slicing-Based Enhanced Method for Privacy-Preserving in Publishing Big Data," Computers, Materials & Continua, vol. 72, no. 2, pp. 3665–3686, 2022, doi: 10.32604/cmc.2022.024663.

[4] J. M. Cavanillas, E. Curry, and W. Wahlster, New Horizons for a Data-Driven Economy. Cham: Springer International Publishing, 2016.

[5] Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan, and Yong Ren, "Information Security in Big Data: Privacy and Data Mining," IEEE Access, vol. 2, pp. 1149–1176, 2014, doi: 10.1109/access.2014.2362522.

[6] T. Yu and S. Jajodia, "Secure Data Management in Decentralized Systems," Sushi1 Jajodia George Mason University US, 2007.

[7] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, vol. 349, no. 6245, pp. 255–260, 2015.

[8] B.-C. Chen, D. Kifer, K. LeFevre, and A. Machanavajjhala, "Privacy-Preserving Data Publishing," Foundations and Trends® in Databases, vol. 2, no. 1–2, pp. 1–167, Jun. 2009, doi: 10.1561/1900000008.

[9] C. L. Philip Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," Information Sciences, vol. 275, pp. 314–347, 2014, doi: 10.1016/j.ins.2014.01.015.

[10] N. Li, T. Li, and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," in 2007 IEEE 23rd International Conference on Data Engineering, 2007, pp. 106–115, doi: 10.1109/ICDE.2007.367856.

[11] Y. A. A. S. Aldeen, M. Salleh, and M. A. Razzaque, "A comprehensive review on privacy preserving data mining," SpringerPlus, vol. 4, no. 1, pp. 1–36, 2015, doi: 10.1186/s40064-015-1481-x.

[12] A. Sharma and N. Badal, "Literature Survey of Privacy Preserving Data Publishing (PPDP) Techniques," International Journal Of Engineering And Computer Science, vol. 6, no. 5, pp. 1–12, 2017, doi: 10.18535/ijecs/v6i4.12.

[13] P. R. Bhaladhare and D. C. Jinwala, "Novel approaches for privacy preserving data mining in k-anonymity model," Journal of Information Science and Engineering, vol. 32, no. 1, pp. 63–78, 2016.

[14] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkatasubramaniam, "l-diversity: Privacy beyond k-anonymity," in Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on, 2006, p. 24.

[15] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkatasubramaniam, "L -diversity," ACM Transactions on Knowledge Discovery from Data, vol. 1, no. 1, p. 3, Mar. 2007, doi: 10.1145/1217299.1217302.

[16] L. Sweeney, "k-anonymity: A model for protecting privacy," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 05, pp. 557–570, 2002.

[17] A. Hasan, Q. Jiang, H. Chen, and S. Wang, "A New Approach to Privacy-Preserving Multiple Independent Data Publishing," Applied Sciences, vol. 8, no. 5, p. 783, May 2018, doi: 10.3390/app8050783.

[18] A. Gkoulalas-Divanis and G. Loukides, Medical Data Privacy Handbook. Cham: Springer International Publishing, 2015.

[19] B. C. M. Fung, K. Wang, A. W.-C. Fu, and J. Pei, "Anonymity for continuous data publishing," in Proceedings of the 11th international conference on Extending database technology: Advances in database technology, 2008, pp. 264–275.

[20] R. C.-W. Wong, A. W.-C. Fu, J. Liu, K. Wang, and Y. Xu, "Global privacy guarantee in serial data publishing," in 2010 IEEE 26th International Conference on Data Engineering (ICDE 2010), 2010, pp. 956–959.

[21] R. C.-W. Wong and A. W.-C. Fu, "Privacy-preserving data publishing: An overview," Synthesis Lectures on Data Management, vol. 2, no. 1, pp. 1–138, 2010, doi: https://doi.org/10.2200/S00237ED1V01Y201003DTM002.

[22] B. Malin and L. Sweeney, "How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems," Journal of Biomedical Informatics, vol. 37, no. 3, pp. 179–192, Jun. 2004, doi: 10.1016/j.jbi.2004.04.005.

[23] S. R. Ganta, S. P. Kasiviswanathan, and A. Smith, "Composition attacks and auxiliary information in data privacy," in Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08, 2008, p. 265, doi: 10.1145/1401890.1401926.

[24] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing: A survey of recent developments," ACM Computing Surveys (Csur), vol. 42, no. 4, pp. 1–53, 2010.

[25] T. A. Lasko and S. A. Vinterbo, "Spectral Anonymization of Data," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 3, pp. 437–446, Mar. 2010, doi: 10.1109/TKDE.2009.88.

[26] S. E. Fienberg and J. McIntyre, "Data swapping: Variations on a theme by dalenius and reiss," in International Workshop on Privacy in Statistical Databases, 2004, pp. 14–29.

[27] R. Brand, "Microdata Protection through Noise Addition," in Inference control in statistical databases, Springer, 2002, pp. 97–116.

[28] C. K. Liew, U. J. Choi, and C. J. Liew, "A data distortion by probability distribution," ACM Transactions on Database Systems (TODS), vol. 10, no. 3, pp. 395–411, 1985.

[29] T. Li, N. Li, J. Zhang, and I. Molloy, "Slicing: A New Approach for Privacy Preserving Data Publishing," IEEE Transactions on Knowledge

and Data Engineering, vol. 24, no. 3, pp. 561–574, Mar. 2012, doi: 10.1109/TKDE.2010.236.

[30] P. Mikalef, J. Krogstie, I. O. Pappas, and P. Pavlou, "Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities," Information & Management, vol. 57, no. 2, p. 103169, Mar. 2020, doi: 10.1016/j.im.2019.05.004.

[31] P. Samarati and L. Sweeney, "Generalizing data to provide anonymity when disclosing information," in PODS, 1998, vol. 98, p. 188, doi: 10.1145/275487.275508.

[32] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-Diversity: Privacy beyond k-Anonymity," ACM Trans. Knowl. Discov. Data, vol. 1, no. 1, pp. 3–es, Mar. 2007, doi: 10.1145/1217299.1217302.

[33] R. C.-W. Wong, J. Li, A. W.-C. Fu, and K. Wang, "(α, k)-anonymity: an enhanced k-anonymity model for privacy preserving data publishing," in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006, pp. 754–759, doi: https://doi.org/10.1145/1150402.1150499.

[34] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Mondrian Multidimensional K-Anonymity," in 22nd International Conference on Data Engineering (ICDE'06), 2006, pp. 25–25, doi: 10.1109/ICDE.2006.101.

[35] J. Li, M. M. Baig, A. H. M. Sarowar Sattar, X. Ding, J. Liu, and M. W. Vincent, "A hybrid approach to prevent composition attacks for independent data releases," Information Sciences, vol. 367–368, pp. 324–336, Nov. 2016, doi: 10.1016/j.ins.2016.05.009.

[36] R. Mendes and J. P. Vilela, "Privacy-Preserving Data Mining: Methods, Metrics, and Applications," IEEE Access, vol. 5, pp. 10562–10582, 2017, doi: 10.1109/ACCESS.2017.2706947.

[37] V. S. Verykios, E. Bertino, I. N. Fovino, L. P. Provenza, Y. Saygin, and Y. Theodoridis, "State-of-the-art in privacy preserving data mining," in SIGMOD Record, 2004, vol. 33, no. 1, pp. 50–57, doi: 10.1145/974121.974131.

[38] N. Zhang and W. Zhao, "Privacy-Preserving Data Mining Systems," ieee, vol. 40, no. 4, pp. 52–58, Apr. 2007, doi: 10.1109/MC.2007.142.

[39] N. Mohammed, R. Chen, B. C. M. Fung, and P. S. Yu, "Differentially private data release for data mining," in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11, 2011, p. 493, doi: 10.1145/2020408.2020487.

[40] G. Cormode, C. M. Procopiuc, Entong Shen, D. Srivastava, and Ting Yu, "Empirical privacy and empirical utility of anonymized data," in 2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW), 2013, pp. 77–82, doi: 10.1109/ICDEW.2013.6547431.

[41] R. Sarathy and K. Muralidhar, "Evaluating Laplace noise addition to satisfy differential privacy for numeric data.," Trans. Data Priv., vol. 4, no. 1, pp. 1–17, 2011.

[42] A. S. M. T. Hasan, Q. Jiang, J. Luo, C. Li, and L. Chen, "An effective value swapping method for privacy preserving data publishing," Security and Communication Networks, vol. 9, no. 16, pp. 3219–3228, Nov. 2016, doi: 10.1002/sec.1527.

[43] A. Sharma, G. Singh, and S. Rehman, "A Review of Big Data Challenges and Preserving Privacy in Big Data," in Advances in Data and Information Sciences, Springer Nature Switzerland, 2020, pp. 57–65.

[44] S. Rohilla, "Privacy Preserving Data Publishing through Slicing," American Journal of Networks and Communications, vol. 4, no. 3, p. 45, 2015, doi: 10.11648/j.ajnc.s.2015040301.18.

[45] "mohd-akaber/Medical-Patient-Database." [Online]. Available: https://github.com/mohd-akaber/Medical-Patient-Database. [Accessed: 14-Jun-2022].

[46] R. Kohavi and B. Becker, "UMI Machine Learning Repository: Adult Data Set," Irvine, CA: University of California, School of Information and Computer Science., 2019. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Adult. [Accessed: 04-May-2020].

[47] L. Kaufman and P. J. Rousseeuw, Finding Groups in Data, vol. 344. Hoboken, NJ, USA: John Wiley & Sons, Inc., 1990.

[48] M.-Q. Tran et al., "Reliable Deep Learning and IoT-Based Monitoring System for Secure Computer Numerical Control Machines Against Cyber-Attacks With Experimental Verification," IEEE Access, vol. 10, pp. 23186–23197, 2022, doi: 10.1109/ACCESS.2022.3153471.

[49] A. H. M. S. Sattar, J. Li, J. Liu, R. Heatherly, and B. Malin, "A probabilistic approach to mitigate composition attacks on privacy in non-coordinated environments," Knowledge-Based Systems, vol. 67, pp. 361–372, Sep. 2014, doi: 10.1016/j.knosys.2014.04.019.

[50] A. Majeed and S. Lee, "Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey," IEEE Access, vol. 9, pp. 8512–8545, 2021, doi: 10.1109/ACCESS.2020.3045700.

[51] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," Journal of Business Research, vol. 70, pp. 263–286, 2017, doi: 10.1016/j.jbusres.2016.08.001.

[52] C. Pu, H. Zerkle, A. Wall, S. Lim, K.-K. R. Choo, and I. Ahmed, "A Lightweight and Anonymous Authentication and Key Agreement Protocol for Wireless Body Area Networks," IEEE Internet of Things Journal, pp. 1–1, 2022, doi: 10.1109/JIOT.2022.3175756.

[53] A. Shah and R. Gulati, "Privacy Preserving Data Mining: Techniques, Classification and Implications-A Survey," International Journal of Computer Applications, vol. 137, no. 12, 2016.

[54] A. Senosi and G. Sibiya, "Classification and evaluation of privacy preserving data mining: a review," in 2017 IEEE AFRICON, 2017, pp. 849–855.

[55] Wang, Tao, Zhigao Zheng, Mubashir Husain Rehmani, Shihong Yao, and Zheng Huo. 2019. "Privacy Preservation in Big Data From the Communication Perspective—A Survey." IEEE Communications Surveys & Tutorials 21(1):753–78.

[56] C. C. Aggarwal and P. S. Yu, "A General Survey of Privacy-Preserving Data Mining Models and Algorithms," in Privacy-preserving data mining, Springer US, 2008, pp. 11–52.

[57] K. Chen and L. Liu, "Privacy preserving data classification with rotation perturbation," in Proceedings - IEEE International Conference on Data Mining, ICDM, 2005, pp. 589–592, doi: 10.1109/ICDM.2005.121.

Swarm Intelligence-based Hierarchical Clustering for Identification of ncRNA using Covariance Search Model

Lustiana Pratiwi[1], Yun-Huoy Choo[2] and Azah Kamilah Muda[3]
Computational Intelligence and Technologies Lab Research Group
Fakulti Teknologi dan Maklumat Komunikasi
Universiti Teknikal Malaysia Melaka
Melaka, Malaysia

Satrya Fajri Pratama[4]
Department of Computing
College of Business, Technology and Engineering
Sheffield Hallam University
Sheffield, United Kingdom

Abstract—**Covariance Model (CM) has been quite effective in finding potential members of existing families of non-coding Ribonucleic Acid (ncRNA) identification and has provided excellent accuracy in genome sequence database. However, it has significant drawbacks with family-specific search. An existing Hierarchical Agglomerative Clustering (HAC) technique merged overlapping sequences which is known as combined CM (CCM). However, the structural information will be discarded, and the sequence features of each family will be significantly diluted as the number of original structures increases. Additionally, it can only find members of the existing families and is not useful in finding potential members of novel ncRNA families. Furthermore, it is also important to construct generic sequence models which can be used to recognize new potential members of novel ncRNA families and define unknown ncRNA sequence as the potential members for known families. To achieve these objectives, this study proposes to implement Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) to ensure the CCMs have the best quality for every level of dendrogram hierarchy. This study will also apply distance matrix as the criteria to measure the compatibility between two CMs. The proposed techniques will be using five gene families with fifty sequences from each family from Rfam database which will be divided into training and testing dataset to test CMs combination method. The proposed techniques will be compared to the existing HAC in terms of identification accuracy, sum of bit-scores, and processing time, where each of these performance measurements will be statistically validated.**

Keywords—Covariance model; ncRNA identification; swarm intelligence; hierarchical clustering

I. INTRODUCTION

Distinguishing many different classes of noncoding (nc) ribonucleic acids (RNA) according to the performance of its variety roles has been an prevalent area in bio-computational technology [1], [2], [3], [4], [5]. For example, after the relationship between RNA structure and its function is discovered, it is desirable to know the common structure of homologous RNAs to find out the functional signatures. It is also desirable to scan a genome looking for ncRNAs. Strategies employed in protein coding gene identification are not commonly applicable for ncRNAs. Therefore, the identification of ncRNA remains an open problem in bioinformatics. However, two-bases are not necessarily covary, since some point mutations, such as G–C to G–U, are still considered as base pairing [6].

Thus, methods searching for covariation may miss valuable information. The main drawback of covariance model (CM) is the use of information of a specific gene family to gain in accuracy [7]. In areas of ncRNA identification, CM has been quite effective in finding potential members of existing families and has provided excellent accuracy in genome sequence database [8], [9], [10]. Representation of multiple secondary structure alignment using a hairpin loop based on an ordered tree are constructed automatically from existing sequence alignments or even from unaligned example sequence [11].

However, it also has considerable disadvantage, such as computationally expensive and thus hindering its application in practice [6], [12]. Apart from having problems with family-specific search that includes large processing requirements, ambiguity in defining which sequences form a family and insufficient numbers of known sequences to properly estimate model parameters are known to be a big challenge in identification of ncRNA. To improve CM performance, hierarchical clustering, as the most frequently used mathematical technique, tries to group genes into small clusters and to group clusters into higher-level systems [13]. Hierarchical clustering provides a series of nested partitions of the dataset. It splits the data into a nested tree structure, where the levels of the tree show similarity or dissimilarity among the clusters at different levels [5], [7].

In regards to this issue, hierarchical clustering has been known as a efficient and useful technique for analyzing genome data and can be applied to group known ncRNA gene families [13], [14], [15], [16]. Past researchers have applied hierarchical clustering to support the identification process in combining and clustering group of families [15]. Lessening the computational cost imposed by covariance model (CM) based noncoding RNA (ncRNA) gene finding is desirable to search the sequence data using a large number of ncRNA families [14], [17], [18]. The main consideration of CM is searching a gigabyte database of sequences for all known ncRNA gene families, which will take quite a long time, and thus is not practical [19], [22].

Hierarchical clustering has successfully reduced the search time to find members from all original ncRNA families using dot-bracket notations [13]. However, its performance continuously declining when additional families of CMs are

introduced into the CCM. This is because more structural information will be excluded, and the sequence features of each family will be significantly less apparent as the number of original CMs increase [5]. Furthermore, it is not sufficiently covering all the problem spaces. A swarm intelligence-based hierarchical algorithm is proposed in this study to select base pairs from two or more CMs of ncRNA families, such as Particle Swarm Optimization (PSO) and Genetic Algorithm (GA). The general idea is to select base pairs from one sequence feature that has fewest conflicts with the base pairs in another structure and construct new CCM structures, which will increase the discernibility performance of the CCMs. Thus, the paper is organized as follows: the next section further describes the problem which motivated this study, while Sections III and IV elaborate the proposed method and present the results, respectively, followed by the discussion of the results in Section V, and the last section concludes this work.

II. HIERARCHICAL CLUSTERING FOR COVARIANCE MODEL

In the previous study conducted by [14], a technique was introduced to reduce the search complexity when a target genome is searched for more than one known ncRNA gene family. The basic construct of the technique is to combine different CMs into one single CM which captures part of both sequence and structure features of each CM by selecting randomly three sequences, known as cluster, from each family by using hierarchical clustering. Hierarchical clustering is a powerful and useful technique for analyzing genome data because hierarchical clustering can generate a dendrogram which helps organize the CCMs. Each non-leaf node of the dendrogram is a CCM of its child nodes, and each leaf node represents a single CM of a ncRNA gene family [5]. Fig. 1 shows an example of possible structure of a dendrogram.

The existing technique applied the agglomerative approach to cluster ncRNA genes [5]. After the clustering of ncRNA gene families, the existing technique built the CCM for any two CMs in the dendrogram that share a common parent which combined the two selected CMs such that the new CCM captured features of both original CMs. CMs are built from multiple sequence alignment with annotated secondary structure. Thus, the general idea of combining two CMs of the existing technique was to select part of the multiple sequence alignment columns of each CM and connect them together to create a new multiple sequence alignment, which the new CM is built from [5].

Since the multiple sequence alignment is annotated with secondary structure, and each column of multiple alignment corresponds to a secondary structure symbol, either a dot or a bracket, indicating whether it is a paired base or not, combining the multiple sequence alignment is equivalent to combining the annotated secondary structure [18]. To combine secondary structures, the key point is to determine how to select base pairs from the two original secondary structures and put them into the new structure. It is obvious that the algorithm cannot just select all the base pairs from both original structures and simply connect them together, since it will significantly increase the complexity of the CM and make no difference to searching with the two original CMs separately [4].
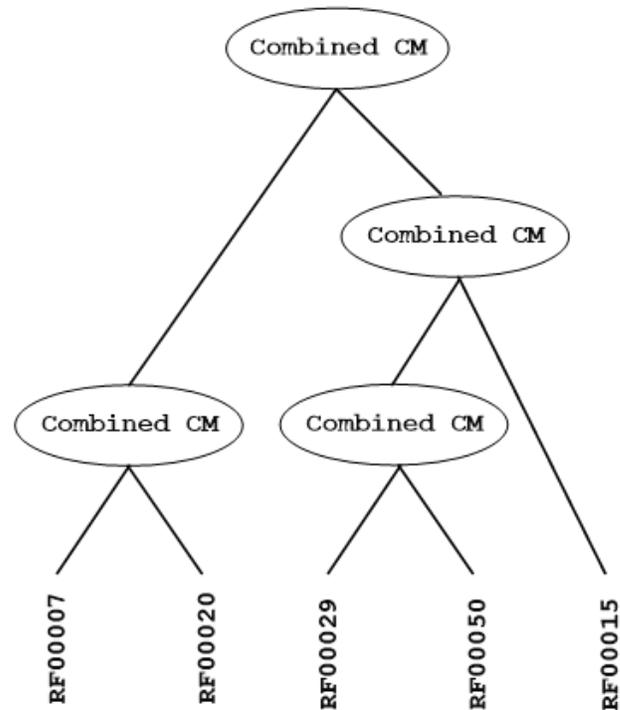
Fig. 1. A Sample of Hierarchical Clustering Result of Five ncRNA Families from Rfam Database.

In most methods of hierarchical agglomerative clustering (HAC), a measure of dissimilarity between clusters is required to decide which clusters should be combined [20], [21]. In the existing technique [23], the members of the cluster are ncRNA gene families, each of which is represented by its secondary structure in dot-bracket notation. Dot-bracket notation is widely used in describing RNA secondary structure. It uses matching brackets to indicate paired bases and dots to denote unpaired bases. This study defines a different distance function that is particularly suitable to deal with the secondary structure data in dot-bracket notation. However, there are two definitions that need to be introduced [5].

Definition 1. *Base pair conflict: Given two RNA secondary structures, a base pair (m, n) from one secondary structure is called base pair conflict if there exists a base pair (i, j) in the other secondary structure such that $m < i < n < j$, or $i < m < j < n$.*

Definition 2. *Structure distance: Given two RNA secondary structures, the structure distance between them is the average of the number of base pair conflicts in each secondary structure.*

The definition of base pair conflict is similar to the definition of pseudoknots, but the difference is that base pair conflict is for base pairs in two different structures while pseudoknots is for those in one structure [14], [24], [25]. Since CM cannot deal with pseudoknots which should be avoided in the combined structure, only one of the two conflict base pairs can be retained in the combined structure [5], [23].

Unlike the distance functions used in most clustering algorithms that measure dissimilarity between observations,

structure distance, the distance function applied in the existing technique, is a measurement of compatibility of two RNA secondary structures [5]. The smaller this value is, the more compatible the two secondary structures are. The compatibility between two secondary structures shows how much structural information can be retained when they are combined. Since the objective is to build a CCM that can capture as much information as possible about both original CMs, two CMs with more compatible secondary structure components would be perfect to be combined. The basic process of HAC [5] is as shown in Algorithm 1.

Algorithm 1 HAC Process

1: Start
2: Assign each ncRNA gene family to its own cluster, then build the distance matrix by computing the structure distance between each family.
3: Find the closest pair of clusters (the minimal element in distance matrix) and combine their secondary structures to create a new family, which corresponds to a non-leaf node in the dendrogram, and then remove the cluster pair from dendrogram.
4: Compute structure distance between the new cluster and each of the old clusters.
5: Repeat steps 2 and 3 until there is only one cluster, which corresponds to the root of the dendrogram.
6: End

Ref. [5] proposed three criteria that should be followed for selecting base pairs from two secondary structures. First, the set of selected base pairs should contain as many base pairs as possible. Since the greater number of base pairs are selected, the more secondary structure components are retained, which also means the more likely a target sequence will be found when searching the genome database. Second, the base pairs selected from one CM should not conflict with those from the other CM. This means there are no pseudoknots in the combined secondary structure due to the reason that CM cannot deal with pseudoknots. Third, each CM should have roughly the same number of base pairs selected, which means this study wants to make a balance between the two original secondary structures.

A greedy algorithm to select base pairs from two secondary structures to form a new secondary structure that satisfies the above three criteria is outlined in Algorithm 2, which is termed as Hierarchical Agglomerative Clustered Covariance Model (HACCM) by [5]. The general idea is to select a base pair from one structure that has fewest conflicts (pseudoknots) with the base pairs in the other structure, which means selecting this base pair will cause fewest deletions of base pairs in the other structure, and such selection takes turns between the two structures.

This existing technique of [5] performs rather well, in most cases, when not too many CMs are combined, where the CCM can successfully represent members from all original families and hardly provide any false alarms. However, its performance gradually deteriorate as more families of CMs are added into the CCM, since more structural information will be discarded, and the sequence features of each family will be significantly diluted as the number of original CMs increase

---

Algorithm 2 Pseudocode to Construct CCM using HACCM [5]

1: Start
2: Generate a list of N_f gene families (each family consists of five randomly selected clusters, one cluster from each original family)
3: Calibrate the gene families using *cmcalibrate* tool from Infernal package
4: **while** $N_f > 0$ **do**
5:     Generate distance matrix between gene families
6:     Select two nearest gene families
7:     Construct CCM from two nearest families by using base pair conflicts
8:     Exclude the two families from future distance matrix calculation
9:     Calculate the TP, TN, FP, and FN from the bit-score of the CCM using members of gene families
10: **end while**
11: End

---

[5]. Furthermore, the original technique is not sufficiently covering all the problem spaces, which means the solution provided may not be the best solution. The frequently used strategy to explore wider problem spaces is by employing evolutionary and swarm intelligence, where each member of the swarm represents one possible solution. By adjusting the number of unique members based on the resource availability, the problem spaces can be further explored.

To achieve this objective, this study proposes to implement swarm intelligence to ensure the CCMs have the best quality for every level of dendrogram hierarchy. On the initialization phase of the swarm intelligence technique, several possible combinations of the CMs will be generated and considered as a single member of swarm intelligence. The selected number of sequences will be unique and the number of sequences for these particles will vary to ensure the maximum coverage. This study will also apply distance matrix as the criteria to measure the compatibility between two CMs.

However, the difference between this study and the previous one is that the selected CCMs will be directly evaluated by CM scoring model, which will be assigned as one of the components of fitness value of the member of swarm intelligence instead of producing the complete dendrogram before evaluating the CM scoring model. The CM from the best member in the current iteration will be then combined with other CMs to generate the new CMs in the subsequent iterations. Furthermore, this study aims to generate reliable cluster pool by having proper selection approach to reduce sequence features dilution in CCMs and to optimize the CCMs selection models which is based on Hierarchical Agglomerative Clustering using Swarm Optimization in finding potential members of novel ncRNA. By the end of the optimization process, this study will construct generic sequence models which represent multiple families to identify unknown family members of ncRNA.

III. Proposed Swarm Intelligence-Based HACCM

This section discusses the tasks undertaken to develop the proposed techniques by hybridizing them with swarm intelli-

gence (SI) techniques, such as Particle Swarm Optimization (PSO) and Genetic Algorithm (GA). In Algorithm 3, the hybridization with SI techniques is primarily to prevent the existing HACCM technique [5] selects the local optima and forces it to reevaluate the candidates with the same merit in every iteration to find the global optima.

Algorithm 3 Optimization Process using SI for HACCM

1: Start
2: Step 1
3:    Generate balanced cluster pool based on minimum number
4: Step 2
5:    Generate similarity matrix between clusters
6:    Find the closest pair of clusters using greedy algorithm and combine it
7:    Construct tree-based CCM structures
8: Step 3
9:    Measure the fitness value of each combined clusters using CM search score and confusion matrix
10: Step 4
11:    Compare and get best hierarchical CM structure with highest fitness value
12:    Update clusters combination with PSO's velocity composite values
13: Step 5
14:    Set threshold Fit = 1 or stagnant fitness for 5 iterations
15:    Proceed to Step 2 if threshold not fulfilled
16: End

The fitness function for the PSO and GA should be identified beforehand. Although there are several possible performance measurements available, such as sensitivity or recall, precision, F_1-score, specificity, and accuracy, this study prefer to use accuracy as the parameter of fitness function because they do not solely rely on the true and false positives. However, in case of the similar fitness value between two or more PSO particles or GA individuals, other criteria should be taken into consideration, which is the sum of bit-scores, which is obtained by evaluating the CM using the *cmsearch* program from Infernal package [3]. In this study, the sum of literal value of bit-scores or similarity score of the families will be used in the fitness function as the tiebreaker parameter. Thus, the proposed fitness function of this study is given in Eq. 1.

$$F_i = \alpha \times \frac{Acc_i(t) - Acc_i(0)}{Acc_i(0)} + \beta \times \frac{SS_i(t) - SS_i(0)}{SS_i(0)}, \quad (1)$$

where Acc_i is the accuracy of ith member (particle in case of PSO and individual in case of GA) in the tth iteration ($t = 0, 1, ...$), and SS_i is the sum of bit-scores from the families of ith member, with $Acc_i(0)$ and $SS_i(0)$ are the accuracy and sum of bit-scores of the original HACCM, which is inspired from the fitness function of HelixPSO for finding RNA secondary structure [28], given in Eq. 2.

$$F = \alpha \times \frac{|S| \cap |C|}{|C|} + \beta \times \min\left(0, \frac{E(S)}{mfE}\right), \quad (2)$$

where α and β in Eqs. 1 and 2 are the parameters used to determine the importance of classification accuracy and the
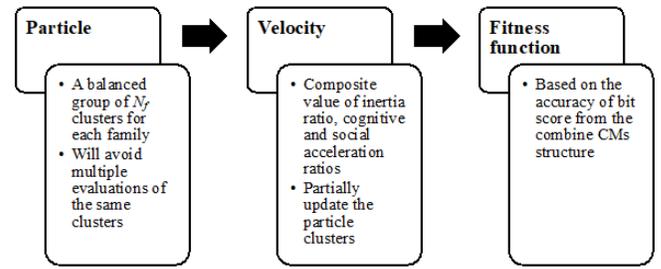
Fig. 2. Optimization Process using PSO.

subset size, where $\alpha \in [0, 1]$ and $\beta = 1 - \alpha$. In this study, the α is set to 0.9, while β is set to 0.1. By setting the accuracy and sum of bit-scores of the current PSO particle or GA individual relative to the original HACCM, the values of fitness value can be guaranteed to not deviate too much to the original HACCM performance.

A. PSO and HACCM Hybridization

In this study, Particle Swarm Optimization (PSO) [26], [27] is selected as one of swarm intelligence techniques to optimize HACCM. One of the main considerations taken for selecting PSO is due to its simple yet effective implementation. The main idea of Particle Swarm Optimized HACCM (PSO-HACCM) is that fitness function of PSO is modified by implementing confusion matrix-based performance measurement techniques calculated from the results of bit-score obtained from *cmsearch* of Infernal package. This is to allow the most optimal interaction between PSO and HACCM, and thus allow for wider search space exploration. Furthermore, there are multiple instances of HACCM executed concurrently; each of them is executed in PSO particle.

This study uses the sum of bit-scores value to be assigned as the particle position. Meanwhile, the fitness value must be identified beforehand since there are several possible confusion matrix-based performance measurement techniques, however this study found the accuracy. Each particle will examine diverse set of CM family clusters, and thus produce unique results, this is because the examined CM family clusters set and its results are recorded, to prevent different particles from examining the same set multiple times.

Apart from the modification to the fitness function, the particle velocity update strategy is also modified which is shown in Fig. 2. Like original PSO, the velocity of the PSO v_i in the $(t+1)$th iteration is affected by inertia ratio I_i, cognitive acceleration ratio C_i, and social acceleration ratio S_i, although is it slightly modified such that

$$v_i(t + 1) = I_r + C_r + S_r, \quad (3)$$

$$I_i = I \times v_i(t), \quad (4)$$

$$C_i = C \times rand() \times (p_i - x_i(t)), \quad (5)$$

$$S_i = S \times rand() \times (p_{best} - x_i(t)), \quad (6)$$

where I, C, and S are the inertia weightage, cognitive acceleration coefficient, and social acceleration coefficient, respectively. In this study, the constants I is set to 0.729844

while C and S is set to 1.49618 [29]. Since the particle is made up N_f clusters, the implementation of these ratios is that there will be maximum $\lfloor \frac{I_r \times N_f}{v_i(t)} \rfloor$ numbers of clusters are reselected from clusters pool, maximum $\lfloor \frac{C_r \times N_f}{v_i(t)} \rfloor$ numbers of clusters are reselected from the particle's personal best, and maximum $\lfloor \frac{S_r \times N_f}{v_i(t)} \rfloor$ numbers of clusters are reselected from the global best particle. The algorithm of PSO-HACCM is illustrated in Algorithm 4.

Fig. 3. Optimization Process using GA.

Algorithm 4 The Pseudocode to Construct Hybrid PSO and HACCM

1: Start
2: Set number of particles $N_p = 3$
3: Set number of iterations $N_t = 100$
4: **for** $i = 1...N_p$ **do**
5:     Generate particle P_i, which is the CCM
6:     Calculate initial sum of bit-scores of CCM as position X_i
7:     Calculate fitness value F_i
8:     Calculate inertia ratio I_i, cognitive acceleration ratio C_i, and social acceleration ratio S_i
9:     Initialize velocity $V_i = I_i + C_i + S_i$
10:     Select global best G_{best} based on maximum fitness value
11: **end for**
12: **for** $t = 1...N_t$ **do**
13:     **for** $i = 1...N_p$ where $P_i! = G_{best}$ **do**
14:         $NumInert = I_i * N_f/V_i$
15:         $NumCog = C_i \times N_f/V_i$
16:         $NumSoc = S_i \times N_f/V_i$
17:         Randomly select $n1 \leq NumInert$ clusters from cluster pool
18:         Randomly select $n2 \leq NumCog$ clusters from P_{best}
19:         Randomly select $n3 \leq NumSoc$ clusters from G_{best}
20:         Generate hierarchical CCM from new clusters
21:         Update X_i, F_i, I_i, C_i, S_i, and V_i
22:     **end for**
23:     Select new Gbest
24:     **if** $F_{G_{best}} == 1$ OR is stagnant after 5 iterations **then**
25:         Stop iteration
26:     **end if**
27: **end for**
28: End

B. GA and HACCM Hybridization

In this study, Genetic Algorithm (GA) [31] is selected as another swarm intelligence techniques to optimize HACCM. The main idea of GA-based HACCM (GA-HACCM) is similar to PSO-HACCM, where the fitness function of GA is modified by implementing well-known performance measurement techniques based on the results of bit-score calculation. This is also to allow the most optimal interaction between GA and HACCM, and thus allow for wider search space exploration based on the modified parameters of the optimization process in Fig. 3.

Similarly, there are multiple instances of HACCM executed concurrently; each of it is executed in GA individual. Each individual I_i will examine different set of CM family clusters, and thus produce unique results, this is because the examined CM family clusters set and its results are recorded, to prevent different individuals examine the same set multiple times. However, unlike PSO-HACCM, there are not much changes modification required to hybridize GA and HACCM. The algorithm of GA-HACCM is illustrated in Algorithm 5.
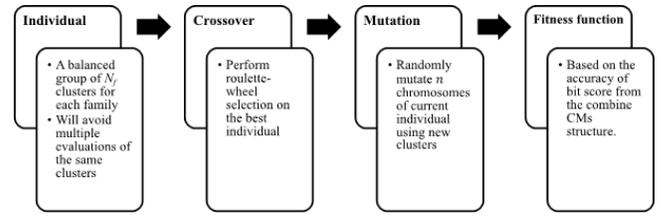
Algorithm 5 The Pseudocode to Construct Hybrid GA and HACCM

1: Start
2: Set number of individuals $N_i = 3$
3: Set number of iterations $N_t = 100$
4: **for** $i = 1...N_p$ **do**
5:     Generate individual I_i, which is the CCM
6:     Calculate initial sum of bit-scores of CCM as X_i
7:     Calculate fitness value F_i
8:     Select global best individual G_{best} based on fitness value
9: **end for**
10: **for** $t = 1...N_t$ **do**
11:     **for** $i = 1...N_i$ where $I_i! = G_{best}$ **do**
12:         **if** Randomly selected for crossover **then**
13:             Crossover with G_{best} using roulette-wheel selection
14:             Randomly mutate n chromosomes of I_i using new clusters from pool
15:             Generate hierarchical CCM from new clusters
16:             Update X_i and F_i
17:         **end if**
18:     **end for**
19:     Select new Gbest
20:     **if** $F_{G_{best}} == 1$ OR is stagnant after 5 iterations **then**
21:         Stop iteration
22:     **end if**
23: **end for**
24: End

IV. EXPERIMENTAL SETUP

With the goal stated in the section above, an extensive and rigorous empirical comparative study is designed and conducted. In this section, a detailed description of the experimental method is provided.

A. Dataset Collection and Preparation

This study obtained the dataset from Rfam database, the most commonly used database that store ncRNA gene
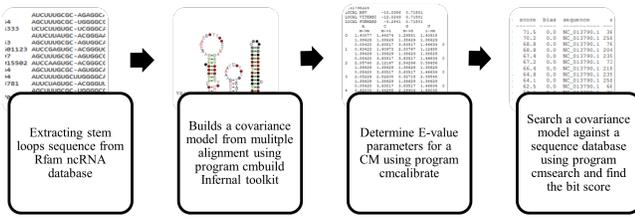
Fig. 4. Steps of Data Collection and Preparation.

TABLE I. SUMMARY OF TRAINING AND TESTING DATASET PREPARATION

Family	Number of Sequences	Number of Selected Sequences	Number of Unselected Sequences
RF00007	51	51	0
RF00015	137	51	86
RF00020	131	51	80
RF00029	86	51	35
RF00050	144	51	93

families from various species [30]. This database is a collection of ncRNA families represented by manually curated sequence alignments, consensus secondary structures and annotation gathered from corresponding taxonomy and ontology resources. The summary of the process of collecting and preparing the dataset is as shown in Fig. 4.

In this study, five sets of ncRNA gene families were selected to test CMs combination method. The selected gene families have roughly similar average length so that their CM combination would not have bias towards either of them. Thus, the number of selected sequences will be set to the minimum number of sequences available, which is fifty-one sequences. This study proposes to randomly divide the selected sequences within one family into groups of three sequences for each gene families by following steps of dataset collection and preparation in Fig. 4 then generate a balanced cluster pool from selected sequences for each family, while the rest of the unselected sequences will be used to form the testing dataset and validate the proposed technique. The process to construct from the generic CCM using existing and proposed techniques is as shown in Algorithm 6, while the summary of the training and testing dataset is shown in Table I.

Algorithm 6 The Pseudocode for the Dataset Preparation

1: Start
2: Set number of families $N_f = 5$
3: Identify minimum number sequences N_s for each family
4: Set number of sequences per cluster $N_{sc} = 3$
5: Set number of clusters per family $N_c = FLOOR(N_s/N_{sc})$
6: Generate cluster pool containing N_c clusters by randomly selecting sequences for each family
7: Set unselected sequences as testing dataset
8: End

Each family obtained from Rfam database is stored in one plain-text file. Each file consists of multiple lines, where each line represents the known ncRNA sequence from that family. There are two data contained in each line separated with tab

space: the first data is the name of the sequence, while the second data is the ncRNA sequence. The unique pattern of each family is stored in the last line of the plain-text file. Apart from the plain-text file, it is also possible to visualize the ncRNA sequence in the Rfam database.

After the ncRNA family dataset has been successfully obtained from Rfam database, the dataset must be prepared and processed so that it can be used by the tools in the Infernal package. Prior to processing by Infernal package tools, the dataset must be stored in .CM file format, which is constructed from .STO file format. The data format in .STO file format is almost similar to the format obtained from Rfam database, with only minor differences, such as the first line of the file in .STO file format must be annotated with *# STOCKHOLM 1.0* syntax and the gene family pattern of .STO file format is only limited to the dot-bracket notation, which consists of left-angle bracket ($<$), right-angle bracket ($>$), and dot symbol (.).

As mentioned earlier, the dataset must be stored in .CM file format. To generate .CM file format, *cmbuild* program from Infernal package must be used. After the .CM file is generated, it is also necessary to calibrate the data using *cmcalibrate* program so that the dataset can be used by other programs in Infernal package. However, calibrating the dataset to .CM format takes a rather long time; thus, this study needs to adjust the parameters of *cmcalibrate* program to reduce the calibration time. The *cmcalibrate* is executed without the forecasting capability, which estimates the execution time by running a small sample of data, reduced its tail loss probability to 10-2 instead of the default 10-15, and the sample size to 0.5 MB instead of the default 1.6 MB.

B. Experimental Design

The quality of the proposed technique must be validated using various performance measurement criteria, because evaluating the performance of learning algorithms is a fundamental aspect of machine learning. In this study, the performance measurements for evaluating the performance of the existing and proposed techniques are the sum of bit-scores, the processing time, and the value of confusion matrix-based performance measurement technique. Despite there are several confusion matrix-based performance measurement techniques available, such as sensitivity, specificity, precision, F_1-score, accuracy, prevalence, phi coefficient, Fowlkes–Mallows index, informedness, markedness, and a few others, this study will only focus on the accuracy as the primary performance measurement by comparing the most commonly used techniques. The sum of bit-scores and processing time are considered as supplementary performance measurements to support the results of the primary measurement.

The bit-score indicates the performance measurement for selected ncRNA sequence. The bit-score is obtained from the *cmsearch* program from Infernal package, and it is used to measure the probability of similar sequence to be found on a given set on covariance model gene family. The processing time can be measured directly while executing the existing and proposed techniques in the clean room environment.

The proposed ncRNA search model techniques are developed to construct generic sequence models which can be used to recognize new potential members of novel ncRNA
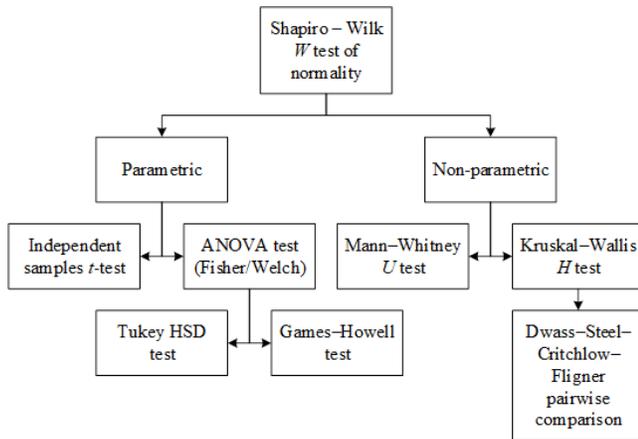
Fig. 5. Summary of Statistical Validation

families. Therefore, it is important to construct a performance measurement to assess this capability. However, since it is practically difficult to obtain the novel and unknown sequences of ncRNA to assess the existing and proposed techniques, this study instead proposed a scenario to simulate this situation.

The scenario of identifying the unknown ncRNA sequences is similar to identifying a set of testing instances by using supervised learning method. By using the datasets mentioned earlier, a portion of the data from the datasets must be set aside to form the testing dataset and the remaining data will be used to form the training dataset, which will be used to construct the CCM. After the CCM has been completely constructed, the testing dataset will be used to validate the CCM, where the result is either the sequence can be found or not found in the combined covariance model. The result will be used to validate the generic sequence models.

The performance measurements are conducted fifty times to ensure the statistical consistency of the results using the datasets mentioned earlier. These performance measurements are validated using various statistical validation techniques available in jamovi statistical package software. The summary of the statistical validation is depicted in Fig. 5.

V. RESULTS AND DISCUSSION

This section provides the results of the comparative study used to identify the most suitable technique to improve the quality of original HACCM proposed by [5]. Table II presents the descriptive test results for the accuracies and the sum of bit-scores of the best member from 50 executions using 3 members for each technique, as well as the total processing time to complete these 50 executions.

Based on the results shown in Table II, the PSO-HACCM produces the best accuracy, sum of bit-scores, and processing time compared to the original HACCM and GA-HACCM. The percentage of the differences between the PSO-HACCM compared to the GA-HACCM and original HACCM is shown in Table III.

The PSO-HACCM results shown in Table III are higher than other techniques are due to its cognitive and social accelerations capabilities, which are not present in the original

TABLE II. DESCRIPTIVE TEST RESULTS

Descriptive	Technique	Accuracy	Sum of Bit-Score	Processing Time
Valid cases	HACCM	50	50	50
	PSO-HACCM	50	50	50
	GA-HACCM	50	50	50
Missing cases	HACCM	0	0	0
	PSO-HACCM	0	0	0
	GA-HACCM	0	0	0
Mean	HACCM	82.8%	316	391
	PSO-HACCM	91.0%	359	357
	GA-HACCM	86.2%	330	392
Median	HACCM	80.0%	313	391
	PSO-HACCM	90.0%	362	355
	GA-HACCM	85.0%	325	391
Mode	HACCM	80.0%	302	328
	PSO-HACCM	95.0%	306	314
	GA-HACCM	85.0%	243	310
Standard deviation	HACCM	4.97%	37.4	31.4
	PSO-HACCM	4.95%	45.8	21.1
	GA-HACCM	5.11%	37.7	42.2
Variance	HACCM	0.247%	1400	987
	PSO-HACCM	0.245%	2098	447
	GA-HACCM	0.261%	1422	1780
Range	HACCM	25.0%	178	148
	PSO-HACCM	15.0%	173	109
	GA-HACCM	25.0%	200	229
Skewness	HACCM	1.26	0.177	0.142
	PSO-HACCM	-0.0263	-0.224	0.847
	GA-HACCM	0.329	0.394	0.735
Kurtosis	HACCM	2.65	0.126	-0.151
	PSO-HACCM	-1.38	-0.546	1.49
	GA-HACCM	0.544	0.867	1.76

TABLE III. DIFFERENCES BETWEEN TECHNIQUES

Compared Technique	Accuracy		Sum of Bit-Score		Processing Time	
	Value	%	Value	%	Value	%
HACCM	4.8%	10.2% higher	43	13.6% higher	34	8.7% faster
GA-HACCM	8.2%	5.6% higher	29	8.8% higher	35	8.9% faster

TABLE IV. PRELIMINARY SHAPIRO–WILK W TESTS OF NORMALITY

Criteria	Technique	Statistic	df	Sig.
Accuracy	HACCM	0.823	50	< 0.001
	PSO-HACCM	0.816	50	< 0.001
	GA-HACCM	0.897	50	< 0.001
Sum of bit-scores	HACCM	0.975	50	0.364
	PSO-HACCM	0.966	50	0.159
	GA-HACCM	0.979	50	0.522
Processing time	HACCM	0.981	50	0.574
	PSO-HACCM	0.957	50	0.064
	GA-HACCM	0.959	50	0.083

HACCM and GA-HACCM. However, to validate whether the difference between PSO-HACCM, GA-HACCM, and original HACCM is statistically significant, in-depth comparison of these techniques must be conducted. The results summarized in Table II are then validated to determine the significance of PSO-HACCM and GA-HACCM compared to original HACCM using either ANOVA or Kruskal–Wallis H test, because it is suitable for comparing the means from more than two distinct groups of data. However, before either ANOVA or Kruskal–Wallis H test is conducted, all techniques must be tested for normality to determine whether its data is normally distributed. Table IV presents the result of Shapiro–Wilk W tests of normality.

Table IV presents the results from Shapiro–Wilk W test of normality. It is concluded that the accuracies of PSO-HACCM,

TABLE V. KRUSKAL–WALLIS H TEST RESULTS FOR THE ACCURACY

Chi-Square	df	Asymp. Sig.
50.9	2	< 0.001

TABLE VI. POST-HOC TEST RESULTS FOR THE ACCURACY

Compared Measurements	W	p
HACCM vs PSO-HACCM	9.42	< 0.001
HACCM vs GA-HACCM	5.29	< 0.001
PSO-HACCM vs GA-HACCM	-6.07	< 0.001

TABLE VII. SHAPIRO–WILK W TESTS OF NORMALITY FOR ASSUMPTION CHECKS OF THE SUM OF BIT-SCORES AND PROCESSING TIME

Criteria	W	Sig.
Sum of bit-scores	0.993	0.655
Processing time	0.972	0.004

TABLE VIII. TEST OF HOMOGENEITY OF VARIANCES OF THE SUM OF BIT-SCORES AND PROCESSING TIME

Criteria	Levene Statistic	$df1$	$df2$	Sig.
Sum of bit-scores	1.87	2	147	0.157
Processing time	9.80	2	147	< 0.001

TABLE IX. FISHER'S TEST ANOVA RESULTS OF THE SUM OF BIT-SCORES

F	$df1$	$df2$	Sig.
15.0	2	147	< 0.001

TABLE X. POST-HOC TEST RESULTS USING TUKEY HSD TEST FOR THE SUM OF BIT-SCORES

Compared Techniques	Mean Difference	t-value	df	Sig.
HACCM vs PSO-HACCM	-43.3	-5.35	147	< 0.001
HACCM vs GA-HACCM	-13.5	-.167	147	0.222
PSO-HACCM vs GA-HACCM	29.8	3.68	147	< 0.001

TABLE XI. KRUSKAL–WALLIS H TEST RESULTS FOR THE PROCESSING TIME

Chi-Square	df	Asymp. Sig.
35.3	2	< 0.001

TABLE XII. POST-HOC TEST RESULTS FOR THE PROCESSING TIME

Compared Measurements	W	p
HACCM vs PSO-HACCM	-7.72	< 0.001
HACCM vs GA-HACCM	-0.04	0.999
PSO-HACCM vs GA-HACCM	6.80	< 0.001

GA-HACCM, and original HACCM techniques are not normally distributed since $p < 0.05$, therefore non-parametric test such as Kruskal–Wallis H test should be conducted instead. On the other hand, the sum of bit-scores and processing time criteria for these techniques are normally distributed, and thus the ANOVA can be conducted for testing the score of these techniques.

The Kruskal–Wallis H test results for the accuracy of PSO-HACCM, GA-HACCM, and original HACCM are shown in Table V, with the post-hoc results using Dwass–Steel–Critchlow–Fligner pairwise comparison is shown in Table VI. Meanwhile, the assumption checks and test of homogeneity of variances for the sum of bit-scores and processing time of PSO-HACCM, GA-HACCM, and original HACCM are shown in Tables VII and VIII, respectively.

Based on the results shown in Table V, there are a statistically significant effect of three techniques $[H(2) = 50.9, p < 0.001]$. Post-hoc comparisons using Dwass–Steel–Critchlow–Fligner pairwise comparison on each pair of groups which is shown in Table VI indicated that there is a statistically significant difference between the accuracy of HACCM and PSO-HACCM ($W = 9.42, p < 0.001$), HACCM and GA-HACCM ($W = 5.29, p < 0.001$), and PSO-HACCM and GA-HACCM ($W = -6.07, p < 0.001$).

On the other hand, based on the results shown in Tables VII and VIII, the sum of bit-scores is normally distributed and there is homogeneity of variances for the sum of bit-scores between groups of techniques, therefore the assumption of ANOVA has been validated, and thus, the Fisher's test must be conducted and Tukey HSD post-hoc tests must be used consequently. On the contrary, the processing time is not normally distributed and there are no there are homogeneity of variances for the processing time between groups of techniques, therefore the assumption of ANOVA has been violated, and the Kruskal–Wallis H test must be conducted instead.

Thus, the ANOVA for the sum of bit-scores of PSO-HACCM, GA-HACCM, and original HACCM are shown in Table IX with the post-hoc results using Tukey HSD and Games–Howell are shown in Table X for the sum of bit-scores, while the Kruskal–Wallis H test results for the processing time of PSO-HACCM, GA-HACCM, and original HACCM are shown in Table XI, with the post-hoc results using Dwass–Steel–Critchlow–Fligner pairwise comparison is shown in Table XII.

Based on the results shown in Table IX, there is a statistically significant effect of the sum of bit-scores between PSO-HACCM, GA-HACCM, and original HACCM techniques $[F(2, 147) = 15.0, p < 0.001]$ at the $p < 0.05$ level. Post-hoc comparisons using the Tukey HSD test shown in Table X indicated that the mean score for the sum of bit-scores of PSO-HACCM is statistically significantly better than HACCM $[t(147) = -5.35, p < 0.001]$ and GA-HACCM $[t(147) = 3.68, p < 0.001]$, while there is no statistically significant difference between HACCM and GA-HACCM $[t(147) = -0.167, p = 0.222]$.

Meanwhile, based on the results shown in Table XI, there are a statistically significant effect of three techniques $[H(2) = 35.3, p < 0.001]$. Post-hoc comparisons using Dwass–Steel–Critchlow–Fligner pairwise comparison on each pair of groups which is shown in Table XII indicated that the mean score for the processing time of PSO-HACCM is statistically significantly better than HACCM ($W = -7.72, p < 0.001$), and GA-HACCM ($W = 6.80, p < 0.001$). These statistical test results of accuracy, sum of bit-scores, and processing time confirm that PSO-HACCM technique is indeed improving the performance of the original HACCM.

To further validate the performance of the proposed techniques, these techniques should be validated by simulating the identification of unknown ncRNA sequences using the combined covariance model constructed by these techniques.

TABLE XIII. DESCRIPTIVE TEST RESULTS OF THE IDENTIFICATION ACCURACY

Descriptive	Technique	Accuracy
Valid cases	HACCM	50
	PSO-HACCM	50
	GA-HACCM	50
Missing cases	HACCM	0
	PSO-HACCM	0
	GA-HACCM	0
Minimum	HACCM	40.0%
	PSO-HACCM	66.67%
	GA-HACCM	32.16%
Maximum	HACCM	99.61%
	PSO-HACCM	91.76%
	GA-HACCM	90.20%
Mean	HACCM	67.6%
	PSO-HACCM	76.6%
	GA-HACCM	66.%
Median	HACCM	67.4%
	PSO-HACCM	75.3%
	GA-HACCM	67.4%
Mode	HACCM	67.4%
	PSO-HACCM	73.7%
	GA-HACCM	60.4%
Standard deviation	HACCM	12.3%
	PSO-HACCM	6.21%
	GA-HACCM	11.9%
Variance	HACCM	1.52%
	PSO-HACCM	0.39%
	GA-HACCM	1.41%
Range	HACCM	59.6%
	PSO-HACCM	25.1%
	GA-HACCM	58.0%
Skewness	HACCM	0.0769
	PSO-HACCM	0.465
	GA-HACCM	-0.379
Kurtosis	HACCM	0.226
	PSO-HACCM	-0.658
	GA-HACCM	0.526

TABLE XIV. PRELIMINARY SHAPIRO–WILK W TESTS OF NORMALITY OF THE IDENTIFICATION ACCURACY

Technique	Statistic	df	Sig.
HACCM	0.989	50	0.917
PSO-HACCM	0.955	50	0.055
GA-HACCM	0.980	50	0.534

The identification will be using the testing dataset discussed in Section IV which has been set aside prior to the construction of the generic sequence models that simulates the identification of unknown ncRNA sequences. The result of this identification simulation is summarized in Table XIII.

Based on the summarized results in Table XIII, it can be seen that the mean accuracy of PSO-HACCM is higher compared to the GA-HACCM and original HACCM. This is due to PSO-HACCM successfully constructed the generic sequence models which can be used to recognize potential members of novel ncRNA families. To determine the significance of the PSO-HACCM results presented in Table XIII compared to GA-HACCM and original HACCM, another statistical validation using ANOVA or Kruskal–Wallis H test must be conducted. As mentioned earlier, all techniques must be tested for normality to determine whether its data is normally distributed. Table XIV presents the result of Shapiro–Wilk W tests of normality for these techniques.

Based on the results presented in Table XIV, it is concluded that the identification accuracies of PSO-HACCM, GA-HACCM, and original HACCM techniques are normally

TABLE XV. SHAPIRO–WILK W TESTS OF NORMALITY FOR ASSUMPTION CHECKS OF THE IDENTIFICATION ACCURACY

W	Sig.
0.988	0.222

TABLE XVI. TEST OF HOMOGENEITY OF VARIANCES RESULTS OF THE IDENTIFICATION ACCURACY

Levene Statistic	$df1$	$df2$	Sig.
6.82	2	147	0.001

TABLE XVII. ANOVA RESULTS OF THE IDENTIFICATION ACCURACY

Test	F	$df1$	$df2$	Sig.
Fisher's	13.6	2	147	< 0.001

TABLE XVIII. POST-HOC TEST RESULTS OF THE IDENTIFICATION ACCURACY

Compared Techniques	Mean Difference	t-value	df	Sig.
HACCM vs PSO-HACCM	-9.07	-4.31	147	< 0.001
HACCM vs GA-HACCM	0.784	0.373	147	0.926
PSO-HACCM vs GA-HACCM	9.851	4.685	147	< 0.001

distributed since $p \geq 0.05$, and thus the ANOVA can be conducted for validating the identification accuracy of these techniques. The assumption checks and test of homogeneity of variances for the identification accuracy of PSO-HACCM, GA-HACCM, and original HACCM are shown in Tables XV and XVI, respectively.

Based on the results shown in Tables XV and XVI, the identification accuracy is normally distributed and there is homogeneity of variances for the identification accuracy between groups of techniques, therefore the assumption of ANOVA has been validated and the Fisher's test results must be considered and Tukey HSD post-hoc tests must be used consequently. The ANOVA for the identification accuracy of PSO-HACCM, GA-HACCM, and original HACCM is shown in Table XVII with the post-hoc results using Tukey HSD is shown in Table XVIII.

Based on the results shown in Table XVII, there is a statistically significant effect of the identification accuracy between PSO-HACCM, GA-HACCM, and original HACCM techniques $[F(2, 147) = 13.6, p < 0.001]$ at the $p < 0.05$ level. Post-hoc comparisons using the Tukey HSD test shown in Table XVIII indicated that the mean score for the identification accuracy of PSO-HACCM is statistically significantly better than HACCM $[t(147) = -4.31, p < 0.001]$ and GA-HACCM $[t(147) = 4.685, p < 0.001]$, while there is no statistically significant difference between HACCM and GA-HACCM $[t(147) = 0.373, p = 0.926]$.

From the comparative performance measurements and statistical validations conducted throughout this section, it can be concluded that PSO-HACCM is indeed performing better compared to the original HACCM as well as GA-HACCM, thanks to its generic sequence model construction.

VI. CONCLUSION

The comparison of accuracy, sum of bit-scores, and processing time have been conducted between original HACCM,

PSO-HACCM, and GA-HACCM to demonstrate the capability of the proposed techniques in constructing generic sequence models, which can be used to recognize potential members of novel ncRNA families. Performance measurement results show that proposed PSO-HACCM performs better than the original HACCM and the proposed GA-HACCM technique, in terms of optimization performance, identification accuracy, sum of bit-scores, and processing time. This finding opens up the possibilities of future works, such as leveraging Graphical Processing Unit (GPU) to speed the computation process, explorations of other swarm intelligence techniques, and inclusion of more gene families for the dataset.

ACKNOWLEDGMENT

REFERENCES

[1] C. Biology and T. B. Laboratories, *Introns and the RNA World*, RNA World, pp. 221–232, 1999.

[2] H. -H. Tseng, Z. Weinberg, J. Gore, R. R. Breaker, and W. L. Ruzzo, *Finding non-coding RNAs through genome-scale clustering*, J. Bioinform. Comput. Biol., vol. 7, no. 2, pp. 373–88, 2009.

[3] S. R. Eddy and R. Durbin, *RNA sequence analysis using covariance models*, Nucleic Acids Res., vol. 22, no. 11, pp. 2079–2088, 1994.

[4] S. F. Smith, *Covariance searches for ncRNA gene finding*, Proc. 2006 IEEE Symp. Comput. Intell. Bioinforma. Comput. Biol. CIBCB'06, pp. 320–326, 2006.

[5] W. Jiang and K. C. Wiese, *Combined covariance model for non-coding RNA gene finding*, IEEE SSCI 2011 - Symp. Ser. Comput. Intell. - CIBCB 2011 2011 IEEE Symp. Comput. Intell. Bioinforma. Comput. Biol., pp. 22–26, 2011.

[6] S. Will, K. Reiche, I. L. Hofacker, P. F. Stadler, and R. Backofen, *Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering*, PLoS Comput. Biol., vol. 3, no. 4, pp. 680–691, 2007.

[7] Y. Saito, K. Sato, and Y. Sakakibara, *Fast and accurate clustering of noncoding RNAs using ensembles of sequence alignments and secondary structures*, BMC Bioinformatics, vol. 12 Suppl 1, p. S48, 2011.

[8] T. Hermann and E. Westhof, *Non-Watson-Crick base pairs in RNA-protein recognition*, Chemistry and Biology, vol. 6, no. 12. 1999.

[9] A. MacHado-Lima, H. A. Del Portillo, and A. M. Durham, *Computational methods in noncoding RNA research*, J. Math. Biol., vol. 56, no. 1–2, pp. 15–49, 2008.

[10] S. Zhang, I. Borovok, Y. Aharonowitz, R. Sharan, and V. Bafna, *A sequence-based filtering method for ncRNA identification and its application to searching for riboswitch elements*, Bioinformatics, vol. 22, no. 14, pp. 1–11, 2006.

[11] S. E. Butcher and A. M. Pyle, *The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks*, Acc. Chem. Res., vol. 44, no. 12, pp. 1302–1311, 2011.

[12] S. Crowder, J. Holton, and T. Alber, *Covariance analysis of RNA recognition motifs identifies functionally linked amino acids*, J. Mol. Biol., vol. 310, no. 4, pp. 793–800, 2001.

[13] Z. Yao, Z. Weinberg, and W. L. Ruzzo, *CMfinder - A covariance model based RNA motif finding algorithm*, Bioinformatics, vol. 22, no. 4, pp. 445–452, 2006.

[14] S. R. Eddy, *A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure*, BMC Bioinformatics, vol. 3, p. 18, 2002.

[15] S. C. Johnson, *Hierarchical clustering schemes*, Psychometrika, vol. 32, no. 3, pp. 241–254, 1967.

[16] S. Alam, G. Dobbie, P. Riddle, and M. A. Naeem, *Particle Swarm Optimization Based Hierarchical Agglomerative Clustering*, Web Intell. Intell. Agent Technol. (WI-IAT), 2010 IEEE/WIC/ACM Int. Conf., vol. 2, pp. 64–68, 2010.

[17] G. Nowak and R. Tibshirani, *Complementary hierarchical clustering*, Biostatistics, vol. 9, no. 3, pp. 467–483, 2008.

[18] J. A. Smith, *RNA search with decision trees and partial covariance models*, IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 6, no. 3, pp. 517–527, 2009.

[19] F. Murtagh and P. Contreras, *Methods of Hierarchical Clustering*, Computer (Long. Beach. Calif)., vol. 38, no. 2, pp. 1–21, 2011.

[20] S. Savaresi, D. Boley, S. Bittanti, and G. Gazzaniga, *Cluster Selection in Divisive Clustering Algorithms*, 2nd SIAM International Conference on Data Mining, 2002.

[21] M. L. Zepeda-Mendoza and O. Resendis-Antonio, *Hierarchical Agglomerative Clustering*, Encyclopedia of Systems Biology, Springer, New York, NY, 2013.

[22] J. Augen, *Bioinformatics and Transcription*, in Bioinformatics in the Post-Genomic Era: Genome, Transcriptome, Proteome, and Information-Based Medicine, 2005, p. 408.

[23] S. Wang, S. Hou, J. Wu, and J. Wei, *Clustering of ncRNA based on structural and semantic similarity*, J. Bionanoscience, vol. 7, no. 1, pp. 20–25, 2013.

[24] D. Li et al., *Experimental RNomics and genomic comparative analysis reveal a large group of species-specific small non-message RNAs in the silkworm Bombyx mori*, Nucleic Acids Res., vol. 39, no. 9, pp. 3792–3805, 2011.

[25] S. Griffiths-Jones, A. Bateman, M. Marshall, A. Khanna, and S. R. Eddy, *Rfam: An RNA family database*, Nucleic Acids Research, vol. 31, no. 1. pp. 439–441, 2003.

[26] R. C. Eberhart and J. Kennedy, *A New Optimizer using Particle Swarm Theory*, In: Proceedings of 6th International Symposium on Micro Machine and Human Science, 1995.

[27] J. Kennedy and R. Eberhart, *Particle swarm optimization*, Proceedings of IEEE International Conference on Neural Networks, pp. 1942–1948 vol. 4, 1995.

[28] M. Geis and M. Middendorf, *Particle swarm optimization for finding RNA secondary structures*, International Journal of Intelligent Computing and Cybernetics, vol. 4 no. 2, pp. 160-186, 2011.

[29] R. C. Eberhart and Y. Shi, *Comparing inertia weights and constriction factors in particle swarm optimization*, Proceedings of the 2000 Congress on Evolutionary Computation, vol. 1, pp. 84-88, 2000.

[30] I. Kalvari, E. P. Nawrocki, N. Ontiveros-Palacios, J. Argasinska, K. Lamkiewicz, M. Marz, S. Griffiths-Jones, C. Toffano-Nioche, D. Gautheret, Z. Weinberg, E. Rivas, S. R. Eddy, R. D. Finn, A. Bateman, and A. I. Petrov, *Rfam 14: expanded coverage of metagenomic, viral and microRNA families*, Nucleic Acids Research, vol. 49(D1), pp. D192–D200, 2021.

[31] J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, 1975.

COVIDnet: An Efficient Deep Learning Model for COVID-19 Diagnosis on Chest CT Images

Briskline Kiruba S
Department of Computer Science
and Engineering
Manonmaniam Sundaranar University
Tirunelveli, Tamilnadu, India 627 012

Murugan D
Department of Computer Science
and Engineering
Manonmaniam Sundaranar University
Tirunelveli, Tamilnadu, India 627 012

Petchiammal A
Department of Computer Science
and Engineering
Manonmaniam Sundaranar University
Tirunelveli, Tamilnadu, India 627 012

Abstract—A novel coronavirus disease (COVID-19) has been a severe world threat to humans since December 2020. The virus mainly affects the human respiratory system, making breathing difficult. Early detection and Diagnosis are essential to controlling the disease. Radiological imaging, like Computed Tomography (CT) scans, produces clear, high-quality chest images and helps quickly diagnoses lung abnormalities. The recent advancements in Artificial intelligence enable accurate and fast detection of COVID-19 symptoms on chest CT images. This paper presents COVIDnet, an improved and efficient deep learning Model for COVID-19 diagnosis on chest CT images. We developed a chest CT dataset from 220 CT studies from Tamil Nadu, India, to evaluate the proposed model. The final dataset contains 5191 CT images (3820 COVID-infected and 1371 normal CT images). The proposed COVIDnet model aims to produce accurate diagnostics for classifying these two classes. Our experimental result shows that COVIDnet achieved a superior accuracy of 98.98% when compared with three contemporary deep learning models.

Keywords—Coronavirus disease; reverse transcription polymerase chain reaction; computed tomography; deep learning

I. INTRODUCTION

The COVID-19 pandemic was the outcome of the Corona virus's rapid global spread. By 2022, there were 575 million COVID-19 patients, 6 million fatalities, and 12 billion vaccination doses [1]. The symptoms of COVID-19 include fever, chills, cough, headache, exhaustion, loss of smell, sore throat, runny nose, vomiting, and diarrhea. Affected individuals may also be asymptomatic. Affected individuals may also be asymptomatic. At least one-third of infected individuals don't have any symptoms at all. Identifying and reducing the spread of COVID-19 is a challenge task for medical experts. Therefore, there is an essential need to identify a faster and more accurate diagnostic methods for COVID-19.

Reverse Transcriptase and Polymerase Chain Reaction (RT-PCR) is a highly specific test that uses a sample taken from a human nose or throat. There have been reports of low COVID-19 sensitivity and it takes 3 to 4 days [2]. Patients with COVID-19 who undergo chest radiography (X-ray) [3] are defined as having bilateral air space consolidation due to inaccuracies in laboratory methods.

Chest radiography is crucial in the early detection and treatment of individuals with suspected or confirmed chest infections during the COVID-19 pandemic. It is entirely accessible in urgent care centers and emergency rooms. The combination of multi-focal peripheral lung abnormalities, such as GGO and/or consolidation, is described by X-ray [4]. These are frequently bilateral, and the COVID-19 diagnosis is challenging since other reasons of respiratory problems prevent consistent detection in X-rays.

Chest CT, a high-sensitivity alternative method for accurate COVID-19 diagnosis, is utilized in conjunction with X-ray imaging[5]. When COVID-19 is detected, the chest CT's main function is to diagnose the disease early [6]. Several advantageous COVID-19-related chest CT scenarios: There is no distinct test for COVID-19 and there is no additional expense to discover COVID; the chest CT examination is performed for other purposes such as cancer screening [7], operations [8], and neurological examinations [9]. People with poorer respiratory conditions underwent simply a chest CT and no other testing right away to figure out their COVID-19. Close contact with the infected patients results in probable COVID-19 [10], but the viral testing come out negative in this instance. To confirm COVID-19 confirms the abnormalities displayed inside the lungs, chest CT imaging is the sole treatment option.

Physicians and radiologists can assist COVID patients with a quick diagnosis by using an automatic diagnostic tool for COVID-19. For the detection of COVID-19 using chest CT, numerous investigations have been carried out. To create an automated technique for detecting COVID-19 [11], computer scientists have thoroughly researched the deep learning models. Despite these efforts, there is a growing need to make these models more reliable and realistically useful by improving their diagnostic performance. This work presents *COVIDnet*, an enhanced and effective deep learning model architecture. To increase COVID detection accuracy on chest CT images, the proposed COVIDnet is built on CNN architecture and has 22 layers. We created a dataset of 5191 CT images (3820 COVID-infected and 1371 normal CT images) from 220 CT scans from hospitals in Tamilnadu, India, in order to evaluate our model. When compared against three current deep learning models, including CNN, VGG16, and MobileNet, our COVIDnet model had the greatest accuracy of 98.98%.

The paper is organized as follows: Section II summarizes the related studies; Section III presents the proposed methodology. Section IV describes the experimental setup and results, and Section V concludes the paper.

II. RELATED WORK

The COVID-19 pandemic has turned into a significant worldwide problem as researchers try to find a way to stop the virus' spread and infection rate. Several research projects have started based on the AI systems recommended for COVID-19 detection CT images. Studies have shown the efficiency of CT scan image processing for diagnosing COVID-19 while also demonstrating the degree of infection, which can be significant for improved treatment options. Several researchers have presented AI-based COVID-19 identification methods. For the most part, deep learning-based approaches have generated very encouraging detection accuracy for COVID-19. We have reviewed a few studies and compared them to our proposed model.

Several studies are cited when discussing chest CT imaging. 746 chest CT scan pictures are analyzed using the LeNet-5 CNN model [12], which forecasts COVID and normal. Here, the dataset has been upgraded and increased utilizing augmentation techniques, making it more useful and accurate to 86.06%. The 746 chest CT scans were trained using 147 million parameters and three distinct learning techniques using the EfficientNet model [13], which achieved an accuracy of 89.7%. For COVID-19 feature extraction and binary classification, an optimized DL network Whale Optimization Algorithm (WOANet) [14] was used. They employed CCT pictures to diagnose COVID-19 using the ResNet-50 CNN network. They optimized the hyperparameters to achieve the best performance using the backpropagation and WOA algorithms. Preprocessing and Region of Interest (ROI) extraction are not required for the suggested technique. They applied COVID-CT dataset.

A deep learning-based system for the automated diagnosis of COVID-19 patients was built via GAN architecture [15]. The suggested framework offered a strong performance at many phases and an efficient feature extraction. To produce additional Chest Computed Tomography (CCT) images for DL network training, the initial step is augmenting the data using a generative adversarial network (GAN) architecture. In the second stage, the Generative Adversarial Network (GAN) hyperparameters are optimized using the WOA optimization. The major goal was to prevent overfitting and instability problems. Finally, COVID-19 patients were automatically classified during the classification step using a pretrained InceptionV3 DL model. They made use of the 2,482 CCT scan images from the SARS-CoV-2 CT-Scan dataset. The experimental investigation demonstrated that the suggested model performed better than other cutting-edge models.

A deep learning-based automated screening strategy was developed by the local attention classification model [16] to separate CT samples found to have COVID-19 or influenza-A viral pneumonia from the sample of patients with healthy lungs. The experimental result showed an overall classification accuracy of 86.7% for the 618 lung CT images dataset. In order to extract the best attributes for identifying COVID-19 and differentiating it from healthy cases and pneumonia infections, a completely automated system was presented using CNN [17]. The 400 CT scans in the used dataset resulted in an overall COVID-19 case detection accuracy of 96%.

A better version of DRENet [18] is an autonomous deep-learning identification method to help medical professionals

TABLE I. TN-COVID DATASET

S.No.	Category	No. of chest CT studies	No. of images
1	COVID	166	3820
2	Normal	54	1371
	Total	220	5191

find and identify COVID-19-infected individuals. The obtained dataset included CT scans from 100 bacterial pneumonia samples, 88 COVID-19 samples, and 86 healthy samples. With an accuracy of 95%, this DRENet deep learning network could identify and categorize samples that were contaminated with COVID-19 and bacterial pneumonia. A system to provide a clinical identification of the pathogenic examination utilizing a CNN-based [19] automated technique in detecting distinctive COVID-19 characteristics. With 217 cases in the dataset, the accuracy was 82.9%.

Using a dataset with 1495 patients with COVID-19 and 1027 patients with CAP, an adaptive feature selection guided deep forest (AFSDF) algorithm [20] was used to classify COVID-19 from chest CT images. This model was developed using a high-level feature representation. A feature selection method was applied to lessen the redundancy of the features based on the trained deep forest model. The COVID-19 classification model was adaptively updated to include this feature selection technique. There were 91.79%, 93.05%, 89.95%, and 96.35% values for accuracy, sensitivity, specificity, and area under the ROC curve, respectively.

Developing an automated approach for classifying lung CT scans is still challenging due to the complexity of diagnosing inflammatory and infectious lung illness via visual assessment. An appropriate norm as a visual examination is subject to errors due to the vast number of patients requiring diagnosis. A method that uses a CNN-based [21] automated technique to identify specific COVID-19 traits to provide a clinical identification of the pathogenic evaluation. With 217 examples in the sample, the accuracy was 82.9%.

The aforementioned deep learning models for COVID-19 detection use chest CT images to address the working models. We performed tests on the independently gathered raw TN-COVID dataset and contrasted its performance with that of the publicly integrated dataset using accepted assessment metrics. The effectiveness of the current deep learning models to be further enhanced in order to create new models for future development and to offer visualizations for better model explanation.

III. MATERIALS AND METHODOLOGY

Fig. 1 shows the proposed methodology for COVID-19 detection using deep learning and chest CT images. Chest computed tomography (CT) images from different hospitals in Tamilnadu were gathered to create the dataset. The chest CT images are cleaned, annotated, and expanded using image augmentation techniques. Then, the final chest CT dataset is split into training, testing, and validation sets to evaluate the efficacy of the proposed COVIDnet model using five performance metrics. The final model is implemented into a
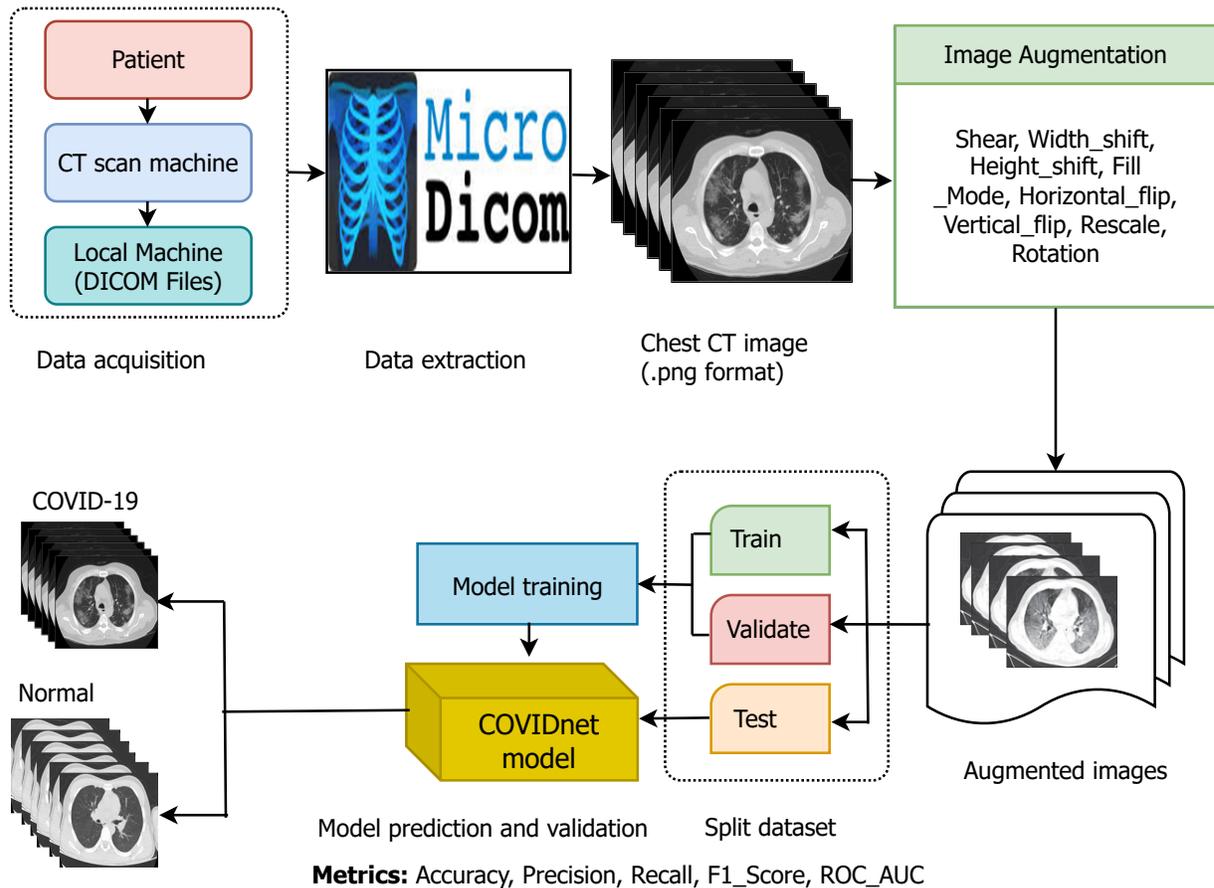
Fig. 1. The Proposed COVID-19 Detection Methodology using COVIDnet Model.

web-based application to identify whether the given chest CT image is from COVID-19 or a normal patient.

A. Data Collection

The chest CT dataset used in this study is directly collected from a hospital in Tamilnadu, India. The CT scanner (Aquilion Lightning Model Tsx 035a), which scans the chest and records the information on a local device, is used to acquire the data. The *MicroDicom viewer* application is used to extract the chest CT images from the DICOM (Digital Imaging and Communications in Medicine) format files that the CT scan machine stores internally. The extracted 16-bit grayscale images in .PNG format have a resolution of 512 by 512 pixels. With the assistance of radiologists and medical professionals, the extracted chest CT images are cleaned. After cleaning, there were 1371 and 3820 images belonging to 54 healthy and 166 COVID-infected patients, respectively. The final dataset, named the *TN-COVID dataset*, consists of 5191 images from 220 CT studies. Additional metadata and anonymized patient information, including patient history and RT-PCR report, are collected. Table I displays an overview of the dataset.

B. Image Augmentation

Data Augmentation techniques are usually employed to increase the size of a training set in order to improve the performance while reducing the overfitting of deep learning models. A number of of common image transformation operations are used to generate more training samples. They are, random rotation and zoom, height and width shift, horizontal and vertical flip, and shear. These operations generate multiple images from each image and help increases the generalization of the trained models.

C. COVIDnet Architecture

For accurate COVID diagnosis in chest CT scans, the proposed COVIDnet model is built with 22-layers. Convolutional layer, pooling layer, flattening layer, dense layer, and activation function are the five key components of the proposed model. Fig. 2 shows the COVIDnet model architecture employing these layers. Below is a description of each of these layers.

Convolutional Layer The convolution operation is applied to the input in the convolution neural network and the output is passed to the following layer. Convolution reduces the size
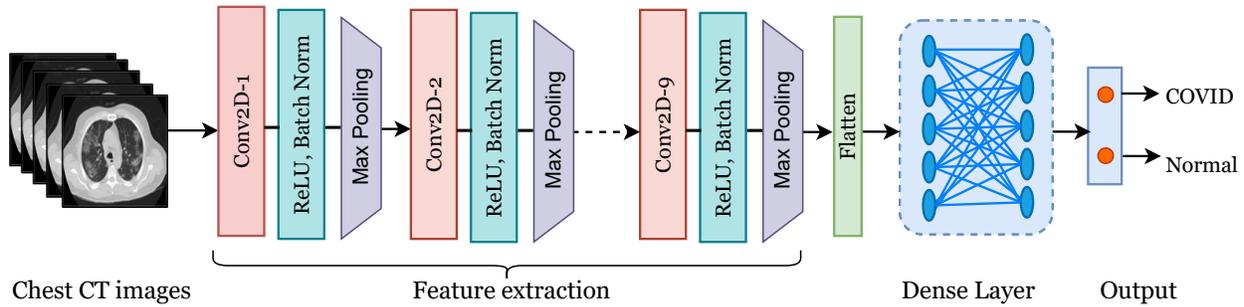
Fig. 2. The Proposed COVIDnet Architecture with 22-Layers for COVID-19 Detection on Chest CT Images.

of the image by combining all of its pixels into a single value. The parameters F, m, B, and W_j can be used to conduct convolution operations. F gives the kernel filter size, m sets feature mappings, B stands for bias, and W_j assigns the kernel weight. By employing a dot product to create one or more matrices, it extracts the feature from the input photos. To remove the features from left to right, convolutional filters are employed. Equation (1) denotes the output of the convolutional layer operation.

Pooling Layer A pooling layer is included when the image size is too huge, to shrink the feature map's size. It helps in speeding up computation, lowering the number of training parameters, and avoid overfitting. The pooling layer is usually added after convolutional layer. There are two categories of pooling: average pooling and maximum pooling. The average value for each component on the feature map is computed using average pooling, while the highest value for each component is generated using maximum pooling. The two hyper-parameters for the pooling layer are Filter and Stride. Equations (2) and (3) denotes the pooling operations. The step that kernels take to shift on the input image is marked by S, while the kernel size is denoted by F.

Flatten Layer It reduces the feature map matrix in its entirety to a single column. The fully linked layer received this input from the single column matrix to categorize the chest CT image.

Dense Layer The dense layer's function is to categorize the output from the convolutional layer for classification. Which features match mostly in a given class are determined by the preceding layer output. It outlines the high-level feature's weights. The input is the weighted average, which is then processed by a nonlinear function acting as an activation function. To calculate the weights and the previous layer, the fully linked layer supplies the accurate probabilities from various classes. The output is classified using the activation function.

Activation Function The activation function defines the sum of the weighted inputs transformed into outputs in the network. Three common activation functions are ReLU, Sigmoid, and Leaky ReLU. The performance of MaxPooling2D is proved by these three activation functions, as shown in equations (4) and (5).

It shows that x is a real number. A sigmoid function has a 0 to 1 range and resembles an S-shaped curve. To change an

TABLE II. LIST OF SYMBOLS

Parameter	Description
F	Specifies the kernel (filter) size
B	Bias
W_j	Weight of the kernel
S	Stride size
$f(z)$	Softmax function
Y_i^1	Output of a convolutional layer, where i indicates the ith feautre map
W_1, H_1	Width and Height of input
$l-1$	Fully connected layers
$m_1(l-1)$	Feautre maps

actual value into one that is easier to understand and can be translated into a probability score. Because sigmoid function networks take longer to train than other types of networks, the vanishing gradient problem quickly converges to 0. ReLU is a linear function that resolves the vanishing gradient issue by setting zero on all negative inputs. Based on the ReLU activation function, a "leaky ReLU" model leaks some positive values to 0 while defining the small slopes with negative values. The slope of the small negative function is represented by the value of x, which is less than zero.

Fig. 2 shows the proposed 22-layer CNN model. Moreover, Table III shows the list of steps used in the development and validation of the proposed COVIDnet model. Nine convolutional, max-pooling, and leaky ReLU layers make up the proposed COVIDnet model. Batch normalization speeds up and improves the stability of the model processing by normalizing, re-centering, and scaling the input. The neural network can be trained more quickly and effectively using the Adam optimizer by adjusting the parameters, first and second gradient moments to modify the learning rate and lower the loss. To discover the optimum accuracy for our suggested model, the deep learning model construction process combines all the layers, activation functions, and optimizer values. We changed the activation function and optimizer settings for every parameter in the extra layers of the proposed model. Our basic layer model is enhanced with 2D Convolutional and MaxPooling layers to produce improved results for two class
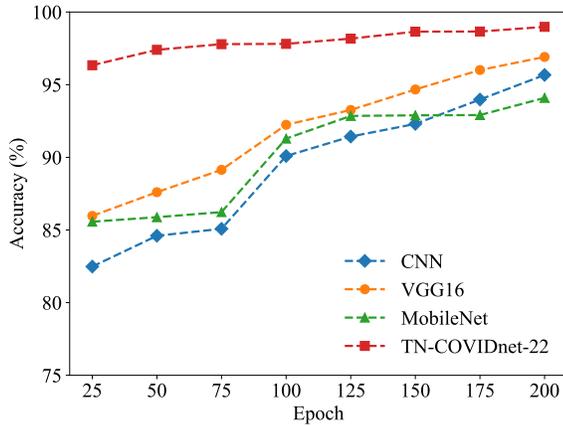
Fig. 3. Comparison of Classification Accuracy of COVIDnet with Three State-of-the-Art Models (CNN, VGG16, and MobileNet). The Proposed COVIDnet Model Obtained the Highest Classification Accuracy of 98.98% with Epoch 200.
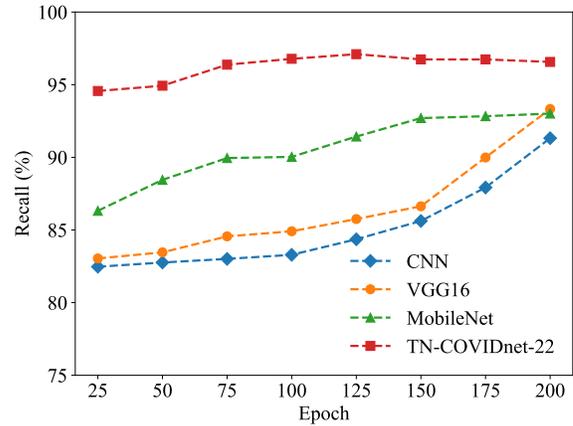
Fig. 5. Comparison of Recall Score of COVIDnet with Three State-of-the-Art Models (CNN, VGG16, and MobileNet). The Proposed COVIDnet Model Obtained the Highest Recall Score of 96.57% with Epoch 200.
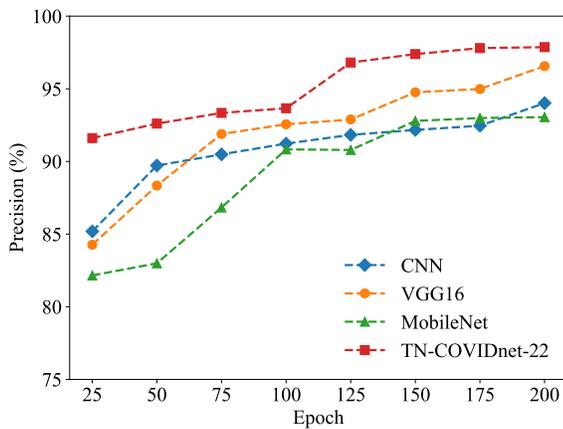
Fig. 4. Comparison of Precision Score of COVIDnet with Three State-of-the-Art Models (CNN, VGG16, and MobileNet). The Proposed COVIDnet Model Obtained the Highest Precision Score of 97.87% with Epoch 200.
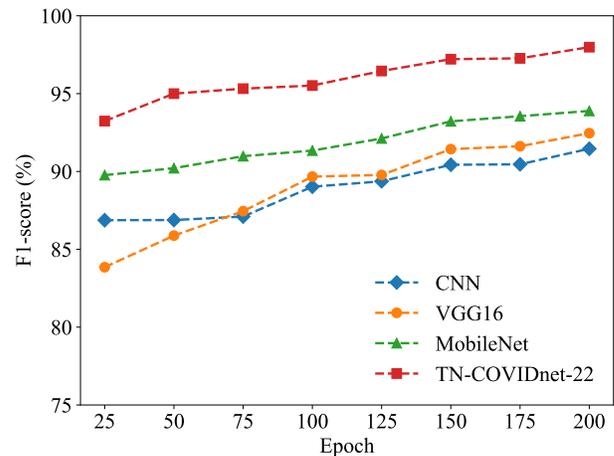
Fig. 6. Comparison of F1-Score of COVIDnet with Three State-of-the-Art Models (CNN, VGG16, and MobileNet). The Proposed COVIDnet Model Obtained the Highest F1-Score of 97.99% with Epoch 200.

classifications.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Setup

Python 3 was used to implement the proposed COVIDnet model utilizing the Keras and TensorFlow frameworks. To increase the training pace, we ran all of our tests on the GPU-powered Kaggle platform. Adam optimizer is used to train the proposed 22-layer COVIDnet model. The batch size is 32, and the learning rate is 0.0001. Different models were trained using various epochs (25, 50, 75, 100, 125, 150, 175, and 200), and we compared the model performance to prevent over-fitting and increase the model's generalizability. The model evaluation is generalized using a 5-fold cross-validation method. Additionally, accuracy, precision, recall,

F1-Score, and ROC-AUC, five common performance metrics, were utilized to compare the performance of the proposed COVIDnet with three conventional deep learning models: simple CNN, VGG16, and MobileNet. Based on ImageNet, the weights of the VGG16 and MobileNet models were initialized.

B. Results

Table IV to Table VIII and Fig. 3 to Fig. 7 compare the experimental findings. As shown in Table IV and Fig. 3, the COVIDnet model obtained the highest classification accuracy of 98.98% with epoch 200. The VGG16 model comes in second with a 96.91% accuracy rate. The accuracy rate for the MobileNet is the lowest (94.10%) as shown in Table IV and Fig. 3. As shown in Table V and Fig. 4, the COVIDnet model also obtained the highest precision of 97.87% with epoch 200. The VGG16 model comes in second with a precision score of

TABLE III. COVIDNET MODEL DEVELOPMENT AND VALIDATION

Input:	TN-COVID chest CT image dataset (5191 images)
Output:	**COVIDnet** model and prediction results (accuracy, precision, recall, F1-Score, and ROC-AUC)
Step 1:	Import sklearn, tensorflow, keras, and other libraries
Step 2:	Load the TN-COVID dataset
Step 3:	Perform image augmentation using ImageDataGenerator() by Rotation, shear, width_shift, height_shift, channel_shift, fill_mode, horizontal_flip, vertical_flip, and rescale.
Step 4:	Split the augmented images into train(60%), validation (20%) and test (20%) sets and load them into different variables.
Step 5:	Initialize *COVIDnet* model and add nine Conv2d layers and nine Maxpool2d layers. The convolution operation is performed by the below equation:

$$Y_i^1 = B_i^1 + \sum_{j=1}^{m_1(I-1)} F_i, j * W_j^{(l-1)} \tag{1}$$

Pooling operations are performed by adjusting the width and height of feature maps which is performed by the below equations:

$$W_2 = \frac{W_1 + F}{S} + 1 \tag{2}$$

$$H_2 = \frac{H_1 + F}{S} + 1 \tag{3}$$

Step 6:	Add batch normalization and activation (ReLU) function to the model. The ReLU function is shown below.

$$f(z) = \max(0, x) \tag{4}$$

$$f'(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases} \tag{5}$$

Step 7:	Add a dense layer using dense() with an activation function of Softmax

$$f(z)_i = \frac{e^{z_j}}{\sum_{k=1}^{k} e^{z_k}} for j - 1, ..., k \tag{6}$$

Step 8:	Train the *COVIDnet* on the training dataset using fit_generator(). The validation set is used to evaluate the model performance during training.
Step 9:	Test the accuracy of the trained *COVIDnet* model on the test set and report the final results using five performance measures.

96.56%. The model from MobileNet has the lowest precision at 93.05%. Aditionally, the COVIDnet model obtained the highest recall rate of 96.57% with epoch 200, as shown in Table VI and Fig. 5. The VGG16 model, which has a recall rate of 93.33%, comes next. The CNN model's 91.32% recall rate is the lowest.

Additionally, as shown in Table VII and Fig. 6, the COVID-net model had the highest F1-score of 97.99% with epoch 200. With an F1-score of 93.89%, the MobileNet model comes in second. With an F1 score of 91.46%, CNN and VGG16

TABLE IV. COMPARISON OF ACCURACY OF COVIDNET WITH THREE STATE-OF-THE-ART MODELS (CNN, VGG16, AND MOBILENET).

S.No.	Epoch	Models			
		CNN	VGG16	MobileNet	COVIDnet
1	25	82.48	85.97	85.57	96.34
2	50	84.60	87.60	85.88	97.40
3	75	85.08	89.14	86.23	97.79
4	100	90.09	92.24	91.29	97.81
5	125	91.43	93.26	92.85	98.17
6	150	92.3	94.67	92.89	98.65
7	175	93.98	96.01	92.91	98.66
8	200	95.67	96.91	94.10	98.98

TABLE V. COMPARISON OF PRECISION SCORE OF COVIDNET WITH THREE STATE-OF-THE-ART MODELS (CNN, VGG16, AND MOBILENET)

S.No.	Epoch	Models			
		CNN	VGG16	MobileNet	COVIDnet
1	25	85.19	84.27	82.16	91.61
2	50	89.72	88.34	82.99	92.60
3	75	90.49	91.89	86.83	93.34
4	100	91.23	92.56	90.83	93.66
5	125	91.83	92.89	90.79	96.81
6	150	92.17	94.76	92.79	97.39
7	175	92.46	94.99	92.99	97.80
8	200	94.02	96.56	93.05	97.87

TABLE VI. COMPARISON OF RECALL SCORE OF COVIDNET WITH THREE STATE-OF-THE-ART MODELS (CNN, VGG16, AND MOBILENET)

S.No.	Epoch	Models			
		CNN	VGG16	MobileNet	COVIDnet
1	25	82.47	83.04	86.32	94.57
2	50	82.76	83.46	88.45	94.93
3	75	83.01	84.56	89.95	96.38
4	100	83.29	84.91	90.03	96.78
5	125	84.36	85.75	91.43	97.10
6	150	85.62	86.63	92.70	96.74
7	175	87.92	89.99	92.83	96.74
8	200	91.32	93.33	93.01	96.57

models are tied for last. As shown in Table VIII and Fig. 7, the COVIDnet model also obtained the highest ROC-AUC score of 98.11% with epoch 200. With a ROC-AUC score of 94.66%, the MobileNet model comes in second. The CNN model's ROC-AUC score of 90.00% is the lowest. Table IX and in Fig. 8 compare each of the five performance measures for four models with 200. It should be highlighted that the proposed COVIDnet model outperformed all other models on the TN-COVID dataset in every statistic. These enhanced outcomes show that the COVIDnet model may aid automated diagnosis of COVID in hospitals.

V. CONCLUSION

The human society is severely impacted by COVID-19. Several automated methods for precise and speedy COVID-
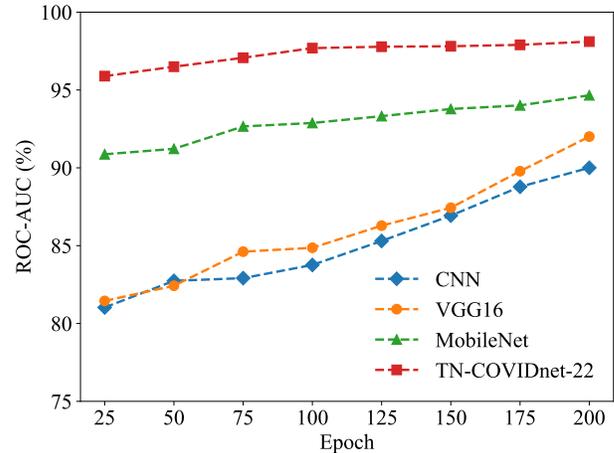

Fig. 7. Comparison of ROC-AUC Score of COVIDnet with Three State-of-the-Art Models (CNN, VGG16, and MobileNet). The Proposed COVIDnet Model Obtained the Highest ROC-AUC of 98.11% with Epoch 200.

TABLE VII. COMPARISON OF F1-SCORE OF COVIDNET WITH THREE STATE-OF-THE-ART MODELS (CNN, VGG16, AND MOBILENET)

S.No.	Epoch	Models			
		CNN	VGG16	MobileNet	COVIDnet
1	25	86.87	83.85	89.77	93.24
2	50	86.88	85.88	90.21	95.00
3	75	87.10	87.45	90.98	95.32
4	100	89.02	89.67	91.34	95.52
5	125	89.37	89.78	92.12	96.46
6	150	90.43	91.44	93.22	97.21
7	175	90.46	91.62	93.55	97.27
8	200	91.46	92.46	93.89	97.99
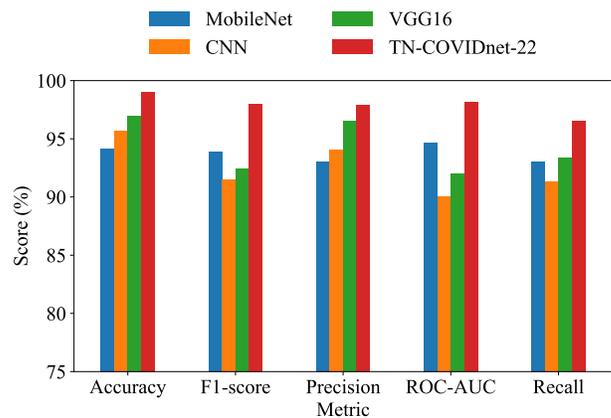
Fig. 8. Comparison of Accuracy, Precision, Recall, F1-Score, and ROC-AUC Scores of Four Models with Epoch 200.

19 diagnosis based on Chest CT images have been proposed in recent years with the advent of AI techniques. In order

TABLE VIII. COMPARISON OF ROC-AUC SCORE OF COVIDNET WITH THREE STATE-OF-THE-ART MODELS (CNN, VGG16, AND MOBILENET)

S.No.	Epoch	Models			
		CNN	VGG16	MobileNet	COVIDnet
1	25	81.03	81.45	90.87	95.89
2	50	82.74	82.43	91.22	96.50
3	75	82.91	84.61	92.66	97.07
4	100	83.76	84.86	92.88	97.69
5	125	85.30	86.29	93.32	97.78
6	150	86.93	87.44	93.78	97.81
7	175	88.78	89.78	94.01	97.90
8	200	90.00	92.00	94.66	98.11

TABLE IX. COMPARISON OF ACCURACY, PRECISION, RECALL, F1-SCORE, AND ROC-AUC SCORES OF FOUR MODELS WITH EPOCH 200

Metric	Models			
	CNN	VGG16	MobileNet	COVIDnet
Accuracy	95.67	96.91	94.09	98.98
Precision	94.02	96.56	93.05	97.87
Recall	91.32	93.33	93.01	96.57
F1-score	91.46	92.46	93.89	97.99
ROC-AUC	90.00	92.00	94.66	98.11

to identify COVID-19 using chest CT images, this research introduces a new deep learning-based model called COVIDnet. For precise COVID-19 detection on chest CT scans, a 22-layer binary classifier (normal or COVID) is proposed. On the TN-COVID dataset, the proposed COVIDnet outperformed three current deep learning models (basic CNN, VGG16, and MobileNet) with an accuracy of 98.98%, precision of 97.97%, recall score of 96.57%, F1-score of 97.99%, and ROC-AUC of 98.11%. These excellent outcomes demonstrate the path in which the COVIDnet model is being implemented into COVID-19 diagnostic tools in hospitals.

ACKNOWLEDGMENTS

REFERENCES

[1] M. Alexandridi, J. Mazej, E. Palermo, and J. Hiscott, "The coronavirus pandemic–2022: viruses, variants & vaccines," *Cytokine & growth factor reviews*, 2022.

[2] W. F. Sule and D. O. Oluwayelu, "Real-time rt-pcr for covid-19 diagnosis: challenges and prospects," *The Pan African Medical Journal*, vol. 35, no. Suppl 2, 2020.

[3] A. M. Ismael and A. Şengür, "Deep learning approaches for covid-19 detection based on chest x-ray images," *Expert Systems with Applications*, vol. 164, p. 114054, 2021.

[4] R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla Jr, and Y. M. Costa, "Covid-19 identification in chest x-ray images on flat and hierarchical classification scenarios," *Computer methods and programs in biomedicine*, vol. 194, p. 105532, 2020.

[5] B. Xu, Y. Xing, J. Peng, Z. Zheng, W. Tang, Y. Sun, C. Xu, and F. Peng, "Chest ct for detecting covid-19: a systematic review and meta-analysis of diagnostic accuracy," *European radiology*, vol. 30, no. 10, pp. 5720–5727, 2020.

[6] J. Pu, J. K. Leader, A. Bandos, S. Ke, J. Wang, J. Shi, P. Du, Y. Guo, S. E. Wenzel, C. R. Fuhrman *et al.*, "Automated quantification of covid-19 severity and progression using chest ct images," *European radiology*, vol. 31, no. 1, pp. 436–446, 2021.

[7] O. M. Al-Quteimat and A. M. Amer, "The impact of the covid-19 pandemic on cancer patients," *American journal of clinical oncology*, 2020.

[8] M. Chetan, M. Tsakok, R. Shaw, C. Xie, R. Watson, L. Wing, H. Peschl, R. Benamore, F. MacLeod, and F. Gleeson, "Chest ct screening for covid-19 in elective and emergency surgical patients: experience from a uk tertiary centre," *Clinical Radiology*, vol. 75, no. 8, pp. 599–605, 2020.

[9] D. P. Stein, R. J. Lederman, D. P. Vogt, W. D. Carey, and T. Broughan, "Neurological complications following liver transplantation," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 31, no. 6, pp. 644–649, 1992.

[10] A. Palmisano, G. M. Scotti, D. Ippolito, M. J. Morelli, D. Vignale, D. Gandola, S. Sironi, F. De Cobelli, L. Ferrante, M. Spessot *et al.*, "Chest ct in the emergency department for suspected covid-19 pneumonia," *La radiologia medica*, vol. 126, no. 3, pp. 498–502, 2021.

[11] D. Javor, H. Kaplan, A. Kaplan, S. Puchner, C. Krestan, and P. Baltzer, "Deep learning analysis provides accurate covid-19 diagnosis on chest computed tomography," *European journal of radiology*, vol. 133, p. 109402, 2020.

[12] M. R. Islam and A. Matin, "Detection of covid 19 from ct image by the novel lenet-5 cnn architecture," in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2020, pp. 1–5.

[13] H. Alshazly, C. Linse, E. Barth, and T. Martinetz, "Explainable covid-19 detection using chest ct scans and deep learning," *Sensors*, vol. 21, no. 2, p. 455, 2021.

[14] R. Murugan, T. Goel, S. Mirjalili, and D. K. Chakrabartty, "Woanet: Whale optimized deep neural network for the classification of covid-19 from radiography images," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 4, pp. 1702–1718, 2021.

[15] T. Goel, R. Murugan, S. Mirjalili, and D. K. Chakrabartty, "Automatic screening of covid-19 using an optimized generative adversarial network," *Cognitive computation*, pp. 1–16, 2021.

[16] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Q. Ni, Y. Chen, J. Su *et al.*, "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, vol. 6, no. 10, pp. 1122–1129, 2020.

[17] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song *et al.*, "Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct," *Radiology*, 2020.

[18] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, R. Wang, H. Zhao, Y. Chong *et al.*, "Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 18, no. 6, pp. 2775–2780, 2021.

[19] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng *et al.*, "A deep learning algorithm using ct images to screen for corona virus disease (covid-19)," *European radiology*, vol. 31, no. 8, pp. 6096–6104, 2021.

[20] L. Sun, Z. Mo, F. Yan, L. Xia, F. Shan, Z. Ding, B. Song, W. Gao, W. Shao, F. Shi *et al.*, "Adaptive feature selection guided deep forest for covid-19 classification with chest ct," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2798–2805, 2020.

[21] A. M. Hasan, M. M. Al-Jawad, H. A. Jalab, H. Shaiba, R. W. Ibrahim, and A. R. AL-Shamasneh, "Classification of covid-19 coronavirus, pneumonia and healthy lungs in ct scans using q-deformed entropy and deep learning features," *Entropy*, vol. 22, no. 5, p. 517, 2020.

Rao-Blackwellized Particle Filter with Neural Network using Low-cost Range Sensor in Indoor Environment

Norhidayah Mohamad Yatim[1], Amirul Jamaludin[2], Zarina Mohd Noh[3]
Centre for Telecommunication Reseasrch & Innovation (CeTRI),
Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer (FKEKK)
Universiti Teknikal Malaysia Melaka (UTeM)
76100 Melaka, Malaysia

Norlida Buniyamin[4]
Faculty of Electrical Engineering (FKE)
Universiti Teknologi MARA (UiTM)
Shah Alam, 40450
Selangor, Malaysia

Abstract—**Implementation of Rao-blackwellized Particle Filter (RBPF) in grid-based Simultaneous Localization And Mapping (SLAM) algorithm with range sensors commonly developed by using sensor with dense measurements such as laser rangefinder. In this paper, a more cost convenient solution was explored where implementation of array of infrared sensors equipped on a mobile robot platform was used. The observation from array of infrared sensors are noisy and sparse. This adds more uncertainty in the implementation of SLAM algorithm. To compensate for the high uncertainties from robot's observations, neural network was integrated with the grid-based SLAM algorithm. The result shows that the grid-based SLAM algorithm with neural network has better accuracy compared to the grid-based SLAM algorithm without neural network for the aforementioned mobile robot implementation. The algorithm improves the map accuracy by 21% and reduce robot's state estimate error significantly. The better performance is due to the improvement in accuracy of grid cells' occupancy value. This affects the importance weight computation in RBPF algorithm hence resulting a better map accuracy and robots state estimate. This finding shows that a promising grid-based SLAM algorithm can be obtained by using merely array of infrared sensors as robot's observation.**

Keywords—*Simultaneous Localization And Mapping (SLAM); occupancy grid map; Neural Network; Rao-Blackwellized Particle Filter; infrared sensor*

I. INTRODUCTION

Sensor is an important element in mobile robot. Sensor such as vision sensor, range sensor and depth sensor are commonly serves as robot's observation of the world. Depending on the environment and task of the robot, different sensor have different advantages. For example, for underwater applications, acoustic based sensor such as sonar sensor is widely applied [1] while for drone navigation, vision sensor is more suitable [2]. Similar to other fields, a cost-effective solution is also a fundamental requirement in mobile robot implementation. Although, current sensors are very reliable for autonomous navigation, but to implement these fast and high accuracy sensor comes with significant cost. Hence, there is research that focuses on developing low-cost navigational method [3]. To achieve this objective, sparse and noisy sensors such as ultrasonic sensor or infrared sensor modules is an alternative to reduce the overall cost [4], [5]. Infrared sensor is commonly used in array configuration on robot's platform

to avoid obstacles using obstacles avoidance algorithm such as Braitenberg algorithm [6]. This configuration gives sparse measurements as the measurements only available at certain angles. In contrast to range sensor such as laser ranger finder or ratating infrared sensor that gives more dense measurements at much higher cost.

There are previous works in robot's simultaneous localization and mapping or SLAM that uses sparse and noisy sensors with neural network as a method to evaluate cells' occupancy in grid map [7], [8], [9]. Using neural network to interpret sensor measurements into occupancy grid cell gives multiple advantages on noisy and sparse sensors implementation. The first one is that from the training data, neural network can generally distinguish erroneous sensor measurements or also called "maximum range" readings. These readings can be caused by poor reflecting surfaces such as glass materials [10] and in situations where no obstacle is within the measurement range of the sensor. Neural network is a simpler approach compared to identify the properties of surface's reflection and compute angle of incidence of sensor's beam to differentiate erroneous measurements [11]. Another advantage is that, the noise caused by injective and non-linear sensor model of infrared sensor can be reduced based on training data [9]. Other than that, neural network can interpret multiple sensors reading concurrently rather than interpreting range measurements independently. This way, the correlation between adjacent sensors can be exploited. This can gain more information for sensors in ring or array formation [7], [12].

To implement grid-based SLAM with array of infrared sensors, the proposed method is to counterbalance the limitation of sparse measurement produce from array of infrared sensor by expanding or extrapolating the sensor measurements. Thus, in this research, the integration of neural network with grid-based SLAM algorithm were developed and experimented with array of infrared sensors.

Overall in this paper, Section 2 describes the method used to carry out the experiments, which consists of the framework for Rao-blackwellized Particle Filter algorithm and neural network configuration. Next, Section 3 reports the results of neural network training, robot's state estimate and map estimate of the SLAM algorithm described. The analysis of the algorithm performance is included as well. Lastly, Section

4 concludes the results of this paper.

II. METHOD

In this section, the Rao-blackwellized Particle Filter (RBPF) algorithm with neural network are explained and elaborated. This consists of the method applied in each component of the RBPF framework.

A. Rao-Blackwellized Particle Filter Algorithm

The Rao-Blackwellized Particle Filter uses particles to estimate robot's state, like the particle filter algorithm. Here, each particle maintains a hypothesis of robot pose that assume its position is correct. Each particle build a map based on its trajectory which is also their own version of the map [13]. The algorithm of learning grid maps using RBPF or grid-based SLAM algorithm is described as follows:

1) **Sampling**: Sample i^{th} particle, $x_t^{[i]}$ to generate current set of particles, X_t by sampling from a proposal distribution.
2) **Importance weighting**: Calculate the importance weight of ith particle, $w_t^{[i]}$, which accounts for how the robot's observation, z_t match with the current map estimate, $m^{[i]}$.
3) **Map update**: Update the particles' map estimate, $m^{[i]}$, conditioned on robot's state, $x_t^{[i]}$ and robot's observation, z_t.
4) **Density extraction**: Calculate the density of X_t by using weighted mean method. The density of X_t, is the single state that represent the robot's position state estimate, x_{est}. The map update for x_{est} is computed as well.
5) **Resampling**: Resample the particles proportionally to the particles' weight. Particles with higher weight will be most likely to be resampled for the next generation of particles.

B. Sampling from Motion Model

In the sampling step, the current set of particles, X_t is sampled from a proposal distribution. In this paper, the particles, X_t is sampled from motion model proposal distribution to predict robot's state. Looking at the ideal case, if particles can be sampled from a true target distribution, the weight of all particles will be equal and the same set of particles, X_t can be maintained without the resampling step [14]. However, a closed form of this posterior is not available in general. As a result, typical method is to sample from motion model, $p(x_t|a_t, x_{t-1})$ as the proposal distribution and use the observation model, $p(z_t|m, x_t)$ as the importance weight.

The motion model makes used of robot's transition model which can be modelled either using velocity information or odometry measurements [15]. For this research, odometry-based model is used by utilizing the wheel encoders' measurements in the robot system. A transition model for two wheels robot from [16] is adapted.

Equation (1) describes the transition from previous state, $x_{t-1} = (x, y, \theta)^T$ to current state, $\bar{x}_t = (\bar{x}, \bar{y}, \bar{\theta})^T$. The notation $\bar{x}_t$ is used (instead of x_t) to denote that this is the

predicted value of current robot's state. The robot movement is illustrated in Fig. 1. Here, l and r are the traveled distance by left wheel and right wheel respectively. While b is the distance between two wheels. $\Delta\theta$ is the change of robot orientation where the calculation is $\Delta\theta = (r - l)/b$ [16].

$$\begin{pmatrix} \bar{x} \\ \bar{y} \\ \bar{\theta} \end{pmatrix} = \begin{pmatrix} x \\ y \\ \theta \end{pmatrix} + \begin{pmatrix} \frac{r+l}{2}\cos(\theta + \frac{r-l}{2b}) \\ \frac{r+l}{2}\sin(\theta + \frac{r-l}{2b}) \\ \frac{r-l}{b} \end{pmatrix} \qquad (1)$$

The essence of probabilistic SLAM is to model measurements by taking into account the uncertainty that comes with the measurements. The odometry measurements were described using probabilistic motion model $p(x_t|x_{t-1}, a_t)$. The motion model defines uncertainties or noise from transition model described earlier. $p(x_t \mid x_{t-1}, a_t)$ gives the likelihood of a robot current position, x_t given that previously it was at x_{t-1}. Subsequently, it moves and odometry data, a_t is measured. The odometry motion model makes used of robot's odometry data. Odometry motion model sampling is an approach where the relative odometry at time $t-1$ to t is divided into three consecutive actions; initial turn, δ_{rot1}, a translation, δ_{trans}, and a second turn, δ_{rot2} as illustrated in Fig. 2 [15].

Algorithm 1 shows the complete algorithm to sample from motion model distribution. The relative odometry is calculated using trigonometry (see line 4 to 6). At line 7 to 9, noise terms; ϵ_{rot1}, ϵ_{trans}, and ϵ_{rot2} are sampled from Gaussian probability distribution with mean zero and the standard deviation of robot-specific parameters, α_1 to α_4. These parameters specify the noise in robot motion where α_1 and α_4 are the rotational errors or angular error, and α_2 and α_3 are the translational errors. These errors come with the motion of relative odometry, δ_{rot1}, δ_{trans}, and δ_{rot2}. In the next subsection the computation of importance weight of each particle is explained.
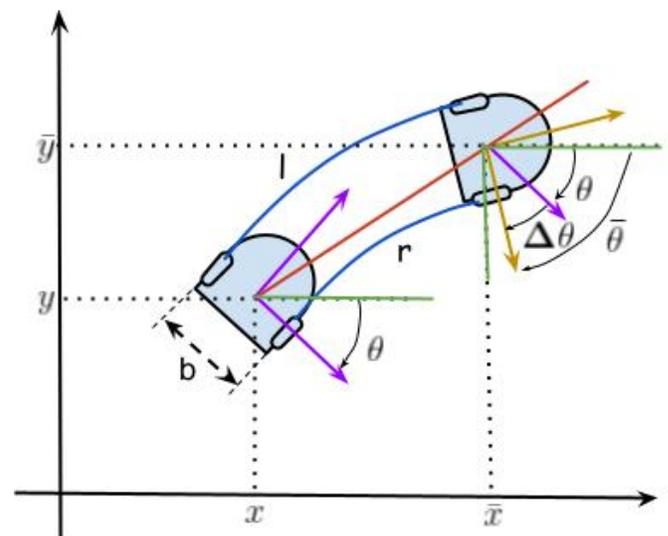
Fig. 1. Robot Moves from $(x, y, \theta)^T$ to $(\bar{x}, \bar{y}, \bar{\theta})^T$.
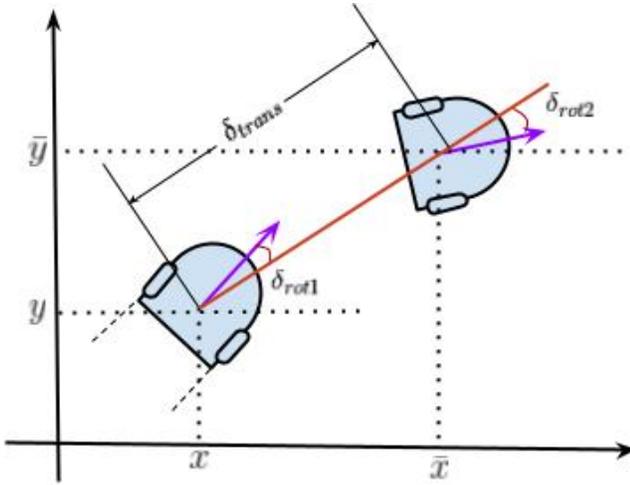
Fig. 2. Relative Odometry shows Initial Turn, δ_{rot1}, a Translation, δ_{trans}, and a Second Turn, δ_{rot2}.

Algorithm 1 Sampling with Odometry Motion Model

Require: x_{t-1}, a_t

1: $\bar{x}_t = g(x_{t-1}, a_t)$
2: $\langle x, y, \theta \rangle = x_{t-1}$
3: $\langle \bar{x}, \bar{y}, \bar{\theta} \rangle = \bar{x}_t$
4: $\delta_{rot1} = atan2(\bar{y} - y, \bar{x} - x) - \bar{\theta}$
5: $\delta_{trans} = \sqrt{(\bar{x} - x)^2 + (\bar{y} - y)^2}$
6: $\delta_{rot2} = \bar{\theta} - \theta - \delta_{rot1}$
7: $\epsilon_{rot1} \sim \mathcal{N}(0, \alpha_1|\delta_{rot1}| + \alpha_2\delta_{trans})$
8: $\epsilon_{trans} \sim \mathcal{N}(0, \alpha_3\delta_{trans} + \alpha_4|\delta_{rot1}| + \alpha_4|\delta_{rot2}|)$
9: $\epsilon_{rot2} \sim \mathcal{N}(0, \alpha_1|\delta_{rot2}| + \alpha_2\delta_{trans})$
10: $\hat{\delta}_{rot1} = \delta_{rot1} + \epsilon_{rot1}$
11: $\hat{\delta}_{trans} = \delta_{trans} + \epsilon_{trans}$
12: $\hat{\delta}_{rot2} = \delta_{rot2} + \epsilon_{rot2}$
13: $x' = x + \hat{\delta}_{trans} \cos(\theta + \hat{\delta}_{rot1})$
14: $y' = y + \hat{\delta}_{trans} \sin(\theta + \hat{\delta}_{rot1})$
15: $\theta' = \theta + \hat{\delta}_{rot1} + \hat{\delta}_{rot2}$
16: **return** $x_t = (x', y', \theta')^T$

C. Importance Weight with Map Matching

Map-matching algorithm is used to obtain the importance weight of ith particle, $w_t^{[i]}$ at current time step, t. The map matching method uses similarities of local and global map of a particle. In this research, the map matching approach uses a match function introduced in [17]. The function compares the local map and the global map of each ith particle and recorded the number of match values. The match number, $match$ is then applied to the function in equation 2. In this equation, f is a parameter that will influence the distribution of particles' weight.

$$w_t^{[i]} = w_{t-1}^{[i]} e^{\frac{match^{[i]}}{f}} \qquad (2)$$

The if and else cases in (3) shows how $match$ is calculated. A local map denotes by m_l and a global map denotes as m_g. $p(m_{g,j})$ and $p(m_{l,j})$ are the probability of occupancy for global map and local map at jth cell respectively. While occ and $free$ are the threshold values that a cell is considered

occupied or free respectively. The $match$ value is added by one if the jth cell of both local and global map exceed the occ threshold value. While for a contradict value of jth cell, the $match$ value is deducted by one.

$$match = \begin{cases} +1, p(m_{g,j}) \geq occ \land p(m_{l,j}) \geq occ \\ -1, p(m_{g,j}) \geq occ \land p(m_{l,j}) \leq free \\ -1, p(m_{g,j}) \leq free \land p(m_{l,j}) \geq occ \end{cases} \qquad (3)$$

D. Selective Resampling

After the computation of weight for each particle, particles are resampled proportionally to their weight. The resampling step is executed based on a heuristic measure of how well the proposal distribution approximates the target distribution called number of effective sample size, N_{eff}. N_{eff} measures the variance of particles' weights. The N_{eff} value is calculated using equation (4), which is the reciprocal of the sum of squares of all particles' weight. Here, N is the number of particles or sample size. Practically, the normalized value of particles weight is used in the N_{eff} calculation.

$$N_{eff} = \frac{1}{\sum_{i=1}^{N}(w^i)^2} \qquad (4)$$

A lower N_{eff} value reflects that there is high variance in particle's weight distribution. This signals that some of the particles have significantly higher weight compared to the others. Algorithm (2) shows integration of selective resampling in particle filter algorithm whereas particles only resampled when N_{eff} value falls below a certain threshold. In this research, the threshold is set to $N/2$ [18]. Thus, particles are only resampled if N_{eff} value falls below half of the number of particles. This shows that the variance of particles' weight is significantly large and particles need to be resampled in order not to waste the particles on particles with low weights. If N_{eff} value is still within the threshold, all the particles are kept, with their normalized weights (see line 1 to 3). Normalization step gives more significant weight values to all particles.

Algorithm 2 Selective Resampling Algorithm

Require: $w_t^{[i]}, i \in \{1, \ldots, N\}$

1: **while** $i \in N$ **do**
2: $\quad w_t^{[i]} = w_t^{[i]} \times [\sum_i^N w_T^{[i]}]^{-1}$
3: **end while**
4: $N_{eff} = [\sum_{i=1}^{N}(w^i)^2]^{-1}$
5: **if** $N_{eff} < threshold$ **then**
6: $\quad X_t = \{\}$
7: $\quad$ **while** $i \in N$ **do**
8: $\quad\quad$ draw $x_t^{[i]}$ with probability $\propto w_t^{[i]}$
9: $\quad\quad X_t = X_t \cup \{\langle x_t^{[i]}, 1 \rangle\}$
10: $\quad$ **end while**
11: **end if**
12: **return** X_t

E. Particle Density

As particle filter maintains a set of hypotheses of robot's state, the estimates of robot's state need to be represented. In other words, from the set of particles, X_t, the density of the particles, x_{est} is computed to evaluate the performance of the SLAM algorithm. There are multiple methods that can be implemented to estimate particles' density. One of the method is to calculate the weighted mean of all particles as the overall robot's state estimate, x_{est}. For the x and y state of x_{est}, equation (5) is used, where w^i is the weight of each particle. As for robot's heading, θ_{est}, a separate equation (6) is used to prevent discretization error in the mean heading.

$$x_{est} = \sum_{i=1}^{N} w^i x^i \qquad (5)$$

$$\theta_{est} = \arctan \frac{\sum_{i=1}^{N} \cos(\theta^i) w^i}{\sum_{i=1}^{N} \sin(\theta^i) w^i} \qquad (6)$$

F. Neural Network Configuration

In this research, neural network served as the inverse sensor model, $p(m_t|x_t, z_t)$ to evaluate the cell's occupancy value. $p(m_t|x_t, z_t)$ is termed as inverse sensor model because the model reasons from effects to causes, where the model provides world information (m_t) from sensor measurements (z_t) instead of the other way around, $p(z_t|m_t, x_t)$. The cell's occupancy value is then converted into log odd notation in the occupancy grid map algorithm which is part of the map update step. The overall map update step implemented is described in our previous work [19]. To interpret cell's occupancy, selected sensor measurements with encoded cell's position were used as the neural network inputs. The following describes the input configuration of the neural network, $\mathcal{N}$:

- Four sensors measurements, $\acute{z}_t^k$, $k \in \{1, 2, 3, 4\}$ that are closest to the cell, m_j.

- Encoded m_j's position using the distance, d_{m_j} and angle, θ_{m_j}.

Fig. 3 illustrates the measurements of d_{m_j} and θ_{m_j}. In the figure, array of infrared sensor on a mobile robot named Khepera III was used. This array configuration has nine infrared sensors, labelled as one until nine. The distance, d_{m_j} and angle, θ_{m_j} are determined by the sensor that is closest to cell m_j. In this illustration, sensor three has the closest distance. Thus, the position of cell m_j is encoded using the distance between m_j and sensor three and the angle of m_j to the same sensor. The four closest sensor measurements in this case will be sensor one, two, three and four, which are selected to be the input of neural network, $\mathcal{N}$.

At each time step, the sensor measurements and the predicted robot's state from motion model are preprocessed to determine the four closest sensor measurements, d_{m_j}, and θ_{m_j}. Hence, follows the configuration of the neural network, $\mathcal{N}$ input. The output of $\mathcal{N}$ is interpreted as probability of cell's occupancy $p(m_j|x_t, z_t)$. This process is repeated for all cells surrounding the Khepera III mobile robot, in order to build a local occupancy grid map, $\mathcal{G}$, with $n \times n$ cells as depicted in Fig. 3.

Fig. 4 shows part of the configuration of the neural network, $\mathcal{N}$. The input layer consists of input nodes that received the sensor measurements, encoded cell's position, and one bias node. There is no computation done on the input nodes but rather just to pass the inputs to the hidden layer. In the hidden layer, there are three neurons and one bias neuron. Bias neuron is used to provide every node in the adjacent layer with a constant value apart from the normal inputs that the node receives. Following normal practice, the bias for neural network, $\mathcal{N}$ is set to value 1. As for the output layer, since there is only one output which is the cell's occupancy value, thus there is only one neuron is in the output layer.

In neural network, each neuron in connected to all neurons in the adjacent layer as showed in Fig. 4. For each connection that connect from a neuron in one layer to the adjacent layer, there is a weight associated with the edge. Thus, in neural network, $\mathcal{N}$, there are $3 \times 7 = 21$ weights for the edges that connect the input layer to the hidden layer and $1 \times 4 = 4$ weights that connect the hidden layer to the output layer. Within these weights, four of them are associated with the bias neuron. Initially all the edges weights are random values.
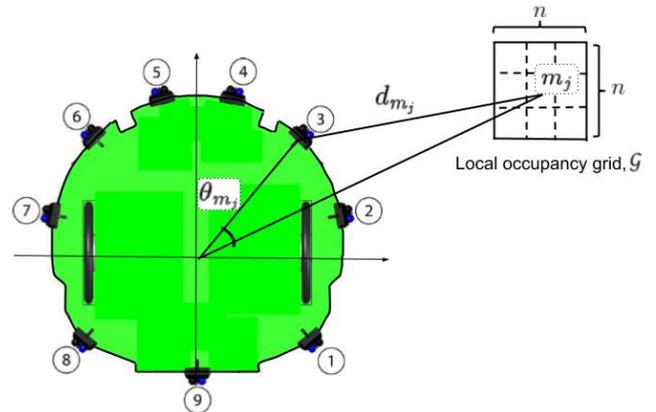
Fig. 3. The Position of Cell m_j is Encoded using d_{m_j} and θ_{m_j} Relative to the Robot.
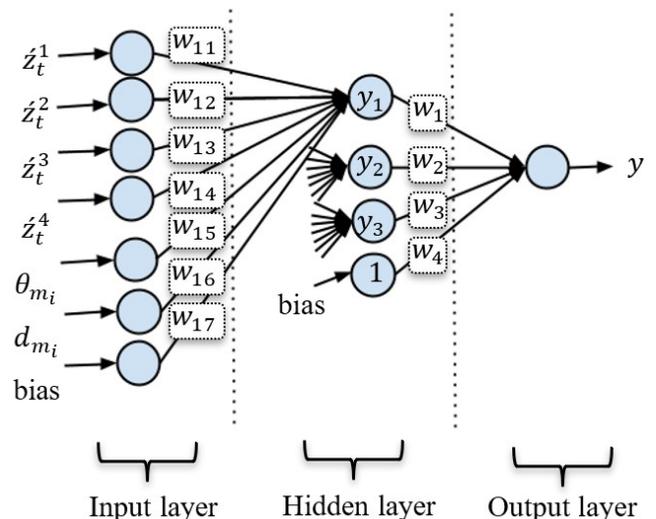
Fig. 4. Part of the Neural Network $\mathcal{N}$ Configuration.

Then, during the training process, for every set of input in the training dataset, each neuron in the hidden layer and the output layer in the neural network, $\mathcal{N}$ (excluding the bias neuron) is activated via an activation function. For example, for the first neuron in the hidden layer, the output of the neuron is obtained via equation (7).

$$
\begin{aligned}
y_1 = F(w_{11}\dot{z}_t^1 + w_{12}\dot{z}_t^2 + w_{13}\dot{z}_t^3 + w_{14}\dot{z}_t^4 \\
+ w_{15}\theta_{m_j} + w_{16}d_{m_j} + w_{17})
\end{aligned}
\tag{7}
$$

Here, w_{xx} is referred to the weight associated with the input neuron as illustrated in Fig. 4. While F is an activation function that compute the output of that neuron, y_1. The same equation with respective weights is applied on all neurons in the hidden layer which will produce the rest of the outputs, y_2 and y_3. Then y_1, y_2, and y_3, becomes the input to the activation function for the neuron in the output layer using equation (8). Here, w_x, referred the weights on edges that connects the hidden layer to the output layer as seen in Fig. 4 as well.

$$
y = F(w_1 y_1 + w_2 y_2 + w_3 y_3 + w_4)
\tag{8}
$$

The output, y, is then compared with the output in the training data for that inputs. Then, the error is propagated back to the previous layer. From the error, the weights for the edges are adjusted. This process is known as backpropagation technique. To analyze the performance of the RBPF SLAM algorithm integrated with neural network $\mathcal{N}$, implementation of inverse sensor model with $\mathcal{N}$ will be tested and compared with the RBPF SLAM algorithm without neural network in simulation experiments.

G. Overall Grid-based SLAM Algorithm with Neural Network

Fig. 5 shows the overall algorithm for the RBPF SLAM algorithm with neural network, $\mathcal{N}$ implemented. In summary, the importance weight of the particles is the observation likelihood $p(z_t|x_t)$. The process and entities in red color describes the integration of neural network, $\mathcal{N}$. Initially, measurements from infrared sensors and sampled particles are preprocessed to build the inputs for $\mathcal{N}$. In the preprocess step, encoded position of cell, m_j, and the measurement of four closest sensors, $\dot{z}_t^k$, $k \in \{1, 2, 3, 4\}$, are calculated. Then, these values are fed into $\mathcal{N}$ to build the local grid map, $\mathcal{G}$. Note that global occupancy grid map is maintained by each particle. Local map and global map are then used to calculate the importance weight of the particles by using the map matching algorithm.

The green boxes indicate selective resampling method described in section II-C. The dotted arrow lines represent the map update step which is executed after the particle's importance weight has been calculated. The final output is the density of robot's state estimate, x_{est} and its map, M, showed at the bottom part of the flowchart. The particles' density, x_{est} then updates the overall robot's trajectory, $x_{1:t}$.

III. RESULT AND ANALYSIS

The experiment was divided into two main parts. In the first part of the experiment, neural network, $\mathcal{N}$ was trained

with the input configuration described in section II-F. Then, in the second part of the experiment, the resulting map and robot's state estimate are analyzed by comparing the results to the ground truth occupancy grid map and the ground truth robot's trajectory.

A. Training Neural Network, $\mathcal{N}$

A neural network tool in MATLAB named patternet was used to train the neural network, $\mathcal{N}$. At the end of the training, the MSE value obtained was 0.0965. Table I shows the weights of edges that connecting the input layer to the hidden layer where y_1, y_2, and y_3 are the neuron in the hidden layer. Table II shows the weights for the edges connecting the hidden layer to the output layer.

Next, the resulting map and robot's state estimate from the RBPF algorithm integrated with the neural network, $\mathcal{N}$ is analyzed by comparing the result to the ground truth data. The ground truth was obtained from the robot simulator platform used in the simulation experiment named Webots robot simulator [20].

B. Map Score and RMSE

Table III shows the performance of the RBPF with neural network, $\mathcal{N}$ SLAM algorithm in terms of map's accuracy and
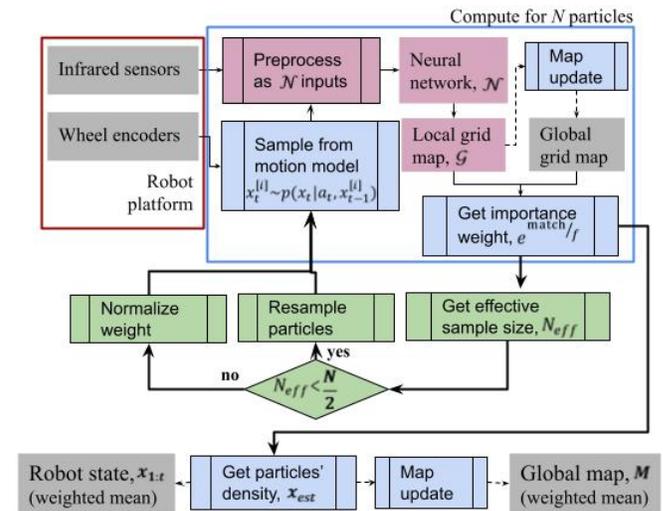
Fig. 5. Overall Algorithm for RBPF SLAM Algorithm with Neural Network, $\mathcal{N}$.

TABLE I. WEIGHTS FOR EDGES CONNECTING THE INPUT LAYER TO THE HIDDEN LAYER.

	$\dot{z}_1^t$	$\dot{z}_2^t$	$\dot{z}_3^t$	$\dot{z}_4^t$	θ_{m_j}	d_{m_j}	bias
y_1	-6.427	0.325	-0.207	0.166	0.529	0.003	1.608
y_2	-0.847	-0.615	-0.611	-0.23	1.992	-0.08	2.170
y_3	0.303	-5.525	-0.206	0.519	0.802	0.054	1.076

TABLE II. WEIGHTS FOR EDGES CONNECTING THE HIDDEN LAYER TO THE OUTPUT LAYER.

	y_1	y_2	y_3	bias
y	0.859	1.558	0.546	-1.325

TABLE III.   MAP ACCURACY AND ROBOT'S STATE ESTIMATE ERROR OF RBPF SLAM ALGORITHM.

RBPF-SLAM	f_{occ}	f_{map}	RMSE(cm)
Without NN	0.2311	0.7012	34.84
With NN	0.7365	0.8491	9.42

robot's state accuracy. For the map state estimate, two methods of analysis were adapted; a fitness score for occupied cells, f_{occ}, and fitness score for all cells, f_{map}. While for robot's state accuracy, the root mean squared (RMSE) value was used as performance measure. This analysis methods is described in our previous work [19].

From Table III, it can be observed that the performance of robot's state estimate and map estimate are in agreement. Both analysis show that the accuracy of RBPF algorithm with neural network, $\mathcal{N}$ is better compared to RBPF algorithm without neural network in terms of robot's state estimate and map estimate accuracy. By adding neural network integration, the score of overall map estimate increases from 70.1% to 84.9%. While the RMSE of the robot's state estimate decreases from 34.84 cm to 9.42 cm.

C. Closing the Loop

Fig. 6 shows the test environment that was used in the simulation experiment with the Khepera III mobile robot in the middle. The rays of infrared sensors and ultrasonic sensors are also shown in the simulator. Fig. 7 shows the resulting map obtained from the simulation experiment. In the resulting map, three trajectories are shown. The robot's ground truth trajectory is the green line. The trajectory of robot state estimate from RBPF with neural network, $\mathcal{N}$ is shown in blue line, while the red line is the trajectory obtained from robot's odometry.

The visual observations of the resulting maps shows that when the robot revisited a known place (i.e. closing the loop) at approximately time step 1550, robot state estimate was able to have better estimate compared to odometry's trajectory. Fig. 7 shows robot's state estimate of RBPF with $\mathcal{N}$ algorithm in red rectangle managed to close the loop despite of robot's odometry begin to diverge from the actual path. This is a significant improvement because by using high variance sensor such as infrared sensor with sparse measurements, a grid-based SLAM algorithm with ability to correct robot's pose in indoor environment with static condition is obtained.

D. Effect of Resampling

As mention in section II-D of the methodology section, the resampling step was executed when the efficiency of particles, N_{eff} has become lower than half of number of particles, N (i.e. $N_{eff} < N/2$). In this experiment $N = 20$ was used. Thus, whenever N_{eff} has become lower than 10, the particles will be resampled and particles' weight will be equal again with value 1. When resampling occurred, the particles that has lower weights are eliminated probabilistically, and the higher weight particles are duplicated. Hence, in some cases, good particles (i.e. particles with better robot's state estimate) are deleted as well. This will cause the spikes in pose error of robots' state estimate.
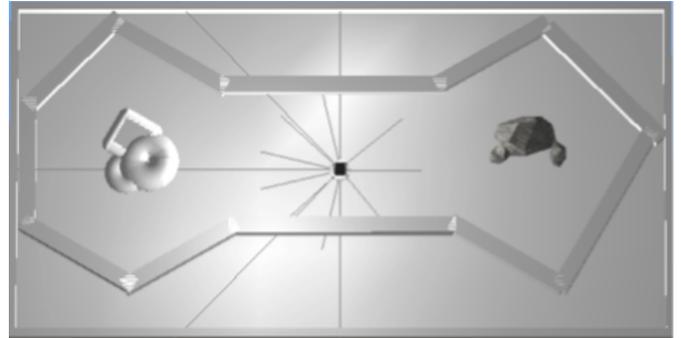
Fig. 6.   Test Environment Created in Webots Simulator.
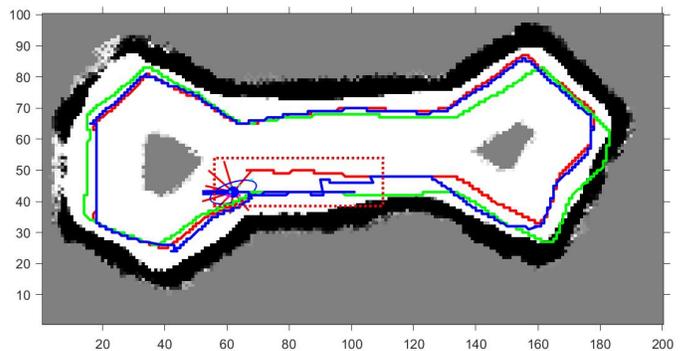
Fig. 7.   Map and Trajectory of Robot's State of RBPF with $\mathcal{N}$ in Blue Line.

Fig. 8 shows the error of robot's state estimate of the RBPF algorithm with $\mathcal{N}$ along with the N_{eff} value at each time step throughout the robot's exploration. In this figure, some of the resampling steps are marked with black dashed line. From the figure, it can be observed that, whenever N_{eff} value drop below 10 and increased abruptly to 20, this is when the resampling steps occurred. If good particles were deleted, the pose error will increase. However, if bad particles were deleted, the robot's state estimate is better, hence the pose error decreases.
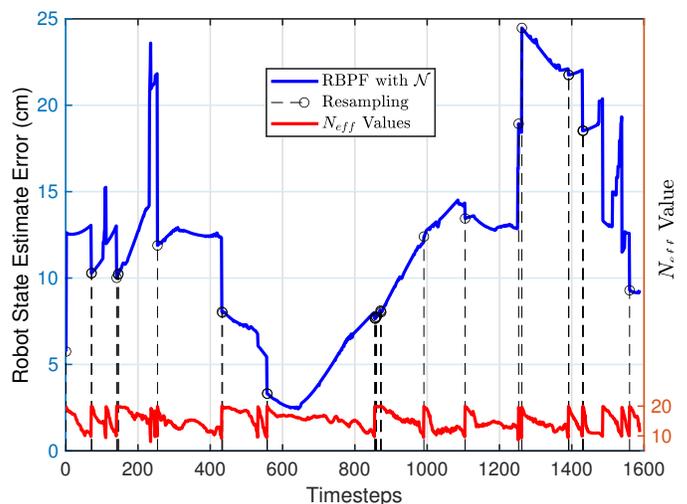
Fig. 8.   The Pose Error of RBPF with $\mathcal{N}$ Align to the Values of N_{eff} at Each Time Step.

E. Number of Particles, N

The experiment is repeated multiple times and it is observed that by using 20 particles the consistency of a converge solution is not achieved. To achieve a more stable and consistent result, a higher number of particles were used. In the subsequent experiment, 100 particles were used with a longer travelled path, where the robot was set to navigate the same route twice in the test environment. Two methods of density extraction is compared which is weighted mean, as described in Section II-E and particle with the highest weight to represent the particles' density.

Fig. 9 shows the robot state estimate error for x_{est} using weighted mean, denotes as x_{mean} and the highest weighted particle, denoted as x_{max}. The robot finished navigate the environment in the first round at approximately time step = 200 and continue on the same path afterwards. The value of the time steps is lower in this experiment, as we increased the time interval between each sensor and odometry measurements taken. It can be observed that after time step 200, the state estimate of both x_{mean} and x_{max} gradually decrease. At this point, robot is basically computing localization using particle filter algorithm with known map. It is also noted that, after time step 200, x_{mean} and x_{max} has maintained below the error of robot state from odometry, x_{odom}. This is because, robot's state from odometry x_{odom} is increasing and as it diverges from the ground truth path.

Fig. 10 shows the resulting map obtained from particle with the highest weighted, x_{max}. The robot's trajectory uses the same color notation, which are green, blue and red lines referring to the ground truth trajectory, RBPF with $\mathcal{N}$ state estimate, and robot's odometry respectively. As the robot travelled further, the error in robot's state estimate accumulates. From the figure, it is observed that in the second loop of the robot's trajectory, the robot state estimate from RBPF with $\mathcal{N}$ (i.e. blue line) is able to localize itself within the map as it travelled through the revisited area. This has contained the error from accumulating further compared to odometry's trajectory (i.e. red line) that begin to diverge from the actual path. As mentioned before, this outcome is also reflected in
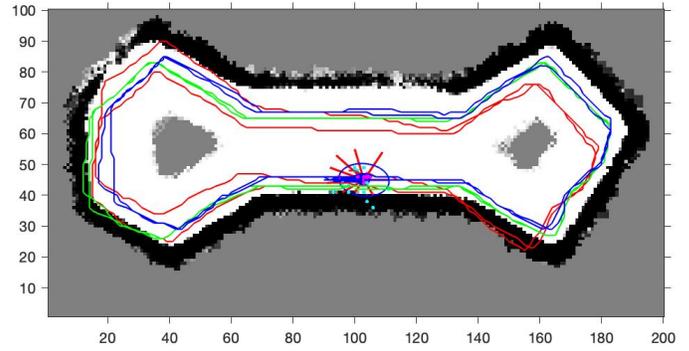
Fig. 10. Map of x_{max} of RBPF with $\mathcal{N}$ Algorithm with Robot's State Estimate and Actual Trajectory.

the pose error graph in Fig. 9, where after half way of the trajectory (i.e. approximately time step = 200) the robot state estimate error of both x_{mean} and x_{est} are below the error of odometry, x_{odom}.

The average RMSE of x_{mean}, x_{max}, and x_{odom} after completed the exploration is shown in the bar graph in Fig. 11. The bar graph shows the average RMSE for the distance error, denoted as x, y state, the x state, y state, and θ state. The x, y state uses the calculation of pose error of robot's From the bar graph, it can be observed that both density; x_{mean} and x_{max} have significantly lower average RMSE error from x_{odom} in red bar for all element of robot's state computed.

To test the consistency of the result, the experiment was repeated 25 times by using the same data set. The average RMSE of state estimate of all 25 trials is computed and shown in Fig. 12. From the bar graph it can be seen that the error of state estimate, x_{mean} and x_{max} are greater than odometry, x_{odom} for all states, except the θ state. This shows a rather different result compared to the one obtained in Fig. 11. It is concluded that, although RBPF with neural network $\mathcal{N}$ can achieve a lower error compared to odometry, x_{odom}, but it is not an entirely consistent result even with 100 particles. By using noisy sensor such as array of infrared sensors, a good
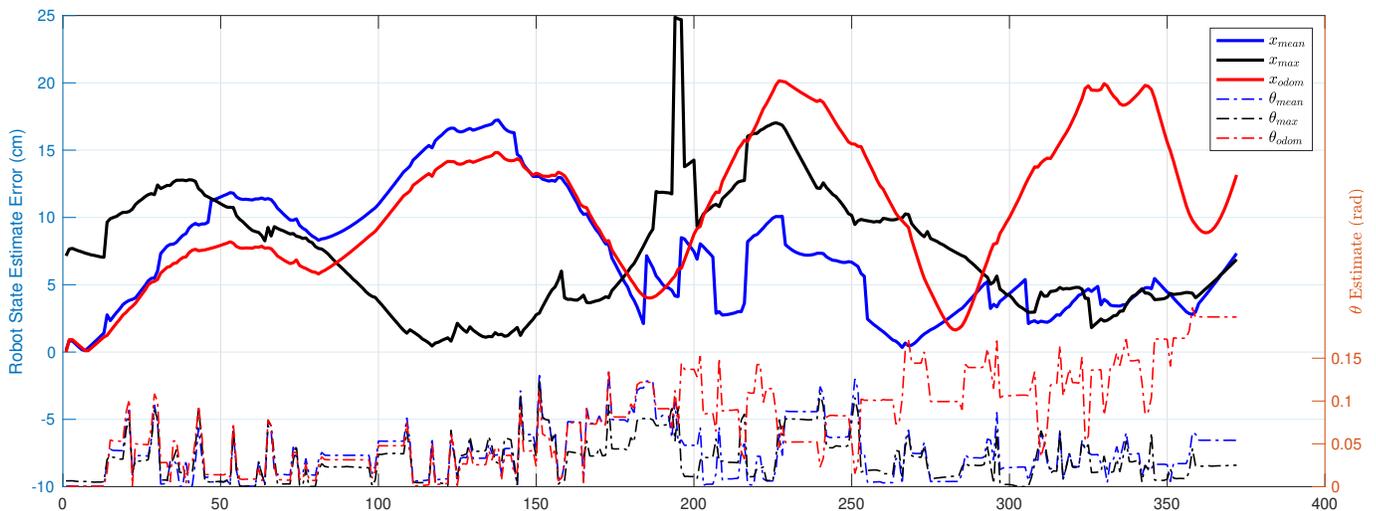
Fig. 9. The Pose Error of x_{mean}, x_{max}, and x_{odom} Align to the θ_{est} of Each Density.
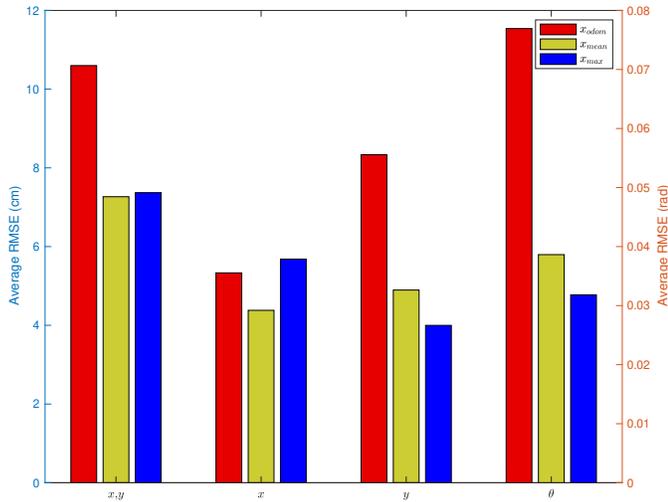
Fig. 11. Comparison of Average RMSE of x_{odom}, and Particles' Density x_{mean} and x_{max}.
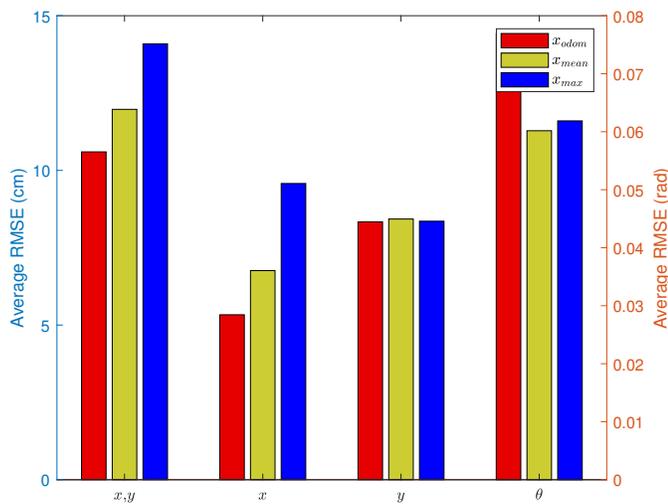
Fig. 12. Average RMSE of Robot's State Estimate after the Robot has Completed Navigation for 25 Trials of Experiment Conducted.

result can be obtained provided multiple trials were conducted.

IV. CONCLUSION

The first objective of this research was to develop grid-based SLAM algorithm with Rao-Blackwellized Particle Filter (RBPF) integrated with neural network for indoor environment. The algorithm was developed with motion model proposal distribution, a map matching function for particles' weight assignment and selective resampling method. The results show that integration of neural network has managed to increase the accuracy of map estimate and decrease the error of robot state estimate.

By using selected sensor measurements and encoded cell's position as neural network input, the neural network configuration is more robust in multiple ways. The neural network can be applied on robot platform with different sensors configuration. This is because the configuration does not require for sensors to be distributed equally apart or any specified

angle for each sensor. Furthermore, by using cells encoded value (or polar coordinate) rather than Cartesian coordinate, local occupancy grid can be resized according to the sensor's maximum range without having to retrain neural network. This is because cell's position is encoded independent of the size of local map.

The consistency of the RBPF algorithm with neural network $\mathcal{N}$ is tested with multiple trials of the experiments. It is shown that the average RMSE of robot's state estimate exceed the RMSE of state computed from odometry even after the number of particles N is increased from 20 particles to 100 particles. This concludes to obtain an adequate map and state estimate from noisy sensors such as array of infrared sensors using the RBPF with neural network algorithm multiple trials should be conducted.

REFERENCES

[1] Yang Cong, Changjun Gu, Tao Zhang, and Yajun Gao. Underwater robot sensing technology: A survey. *Fundamental Research*, 1(3):337–345, 2021.

[2] Abhishek Gupta and Xavier Fernando. Simultaneous localization and mapping (slam) and data fusion in unmanned aerial vehicles: Recent advances and challenges. *Drones*, 6(4):85, 2022.

[3] Ravinder Singh and Kuldeep Singh Nagla. Comparative analysis of range sensors for the robust autonomous navigation–a review. *Sensor Review*, 40(1):17–41, 2019.

[4] Mubariz Zaffar, Shoaib Ehsan, Rustam Stolkin, and Klaus McDonald Maier. Sensors, slam and long-term autonomy: a review. In *2018 NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, pages 285–290. IEEE, 2018.

[5] Stelian-Emilian Oltean. Mobile robot platform with arduino uno and raspberry pi for autonomous navigation. *Procedia Manufacturing*, 32:572–577, 2019.

[6] Francisco Quiroga, Gabriel Hermosilla, Gonzalo Farias, Ernesto Fabregas, and Guelis Montenegro. Position control of a mobile robot through deep reinforcement learning. *Applied Sciences*, 12(14):7194, 2022.

[7] Sebastian Thrun. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence*, 99(1):21–71, 1998.

[8] Behzad Moshiri, Mohammad Reza Asharif, and Reza HoseinNezhad. Pseudo information measure: A new concept for extension of Bayesian fusion in robotic map building. *Information Fusion*, 3(1):51–68, 2002.

[9] Yun-Su Ha and Heon-Hui Kim. Environmental map building for a mobile robot using infrared range-finder sensors. *Advanced Robotics*, 18(4):437–450, 2004.

[10] Andres J Barreto-Cubero, Alfonso Gómez-Espinosa, Jesús Arturo Escobedo Cabello, Enrique Cuan-Urquizo, and Sergio R Cruz-Ramírez. Sensor data fusion for a mobile robot using neural networks. *Sensors*, 22(1):305, 2021.

[11] Maren Bennewitz, Cyrill Stachniss, Sven Behnke, and Wolfram Burgard. Utilizing reflection properties of surfaces to improve mobile robot localization. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 4287–4292, 2009.

[12] Norhidayah Mohamad Yatim and Norlida Buniyamin. Development of Rao-Blackwellized Particle Filter (RBPF) SLAM algorithm using Low Proximity Infrared Sensors. In *The 9th International Conference on Robotics, Vision, Signal Processing & Power Applications*, 2016.

[13] Guillem Vallicrosa and Pere Ridao. H-slam: Rao-blackwellized particle filter slam using hilbert maps. *Sensors*, 18(5):1386, 2018.

[14] Cyrill Stachniss. *Exploration and mapping with mobile robots.* PhD thesis, 2006.

[15] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic robotics*. MIT press, 2005.

[16] Roland Siegwart and Illah R Nourbakhsh. *Introduction to Autonomous Mobile Robots*, volume 23. 2004.

[17] Christof Schröter, Hans-Joachim Böhme, and Horst-Michael Gross. Memory-Efficient Gridmaps in Rao-Blackwellized Particle Filters for SLAM using Sonar Range Sensors. In *EMCR*, 2007.

[18] G Grisetti. Improving grid-based slam with rao-blackwellized particle filters by adaptive proposals and selective resampling. In *International Conference on Robotics and Automation*, number April, pages 32–37, Barcelona, Spain, 2005.

[19] N A Yatim, N Buniyamin, Z M Noh, and N A Othman. Occupancy grid map algorithm with neural network using array of infrared sensors. In *Journal of Physics: Conference Series*, volume 1502, page 12053. IOP Publishing, 2020.

[20] Olivier Michel. WebotsTM: Professional mobile robot simulation. *arXiv preprint cs/0412052*, 2004.

Multi-Scale ConvLSTM Attention-Based Brain Tumor Segmentation

Brahim AIT SKOURT
Laboratory of Intelligent Systems
and Applications
University of Sidi Mohammed Ben Abdellah
Fez, Morocco

Nikola S. Nikolov
Computer Science and
Information Systems Department
University of Limerick
Limerick, Ireland

Aicha MAJDA
University of Moulay Ismail
Networks and Computer Systems
research team
Meknes, Morocco

Ahlame BEGDOURI
Laboratory of Intelligent Systems and Applications
University of Sidi Mohammed Ben Abdellah
Fez, Morocco

Abstract—In computer vision, there are various machine learning algorithms that have proven to be very effective. Convolutional Neural Networks (CNNs) are a kind of deep learning algorithms that became mostly used in image processing with a remarkable success rate compared to conventional machine learning algorithms. CNNs are widely used in different computer vision fields, especially in the medical domain. In this study, we perform a semantic brain tumor segmentation using a novel deep learning architecture we called multi-scale ConvLSTM Attention Neural Network, that resides in Convolutional Long-Short-Term-Memory (ConvLSTM) and Attention units with the use of multiple feature extraction blocks such as Inception, Squeeze-Excitation and Residual Network block. The use of such blocks separately is known to boost the performance of the model, in our case we show that their combination has also a beneficial effect on the accuracy. Experimental results show that our model performs brain tumor segmentation effectively compared to standard U-Net, Attention U-net and Fully Connected Network (FCN), with 79.78 Dice score using our method compared to 78.61, 73.65 and 72.89 using Attention U-net, standard U-net and FCN respectively.

Keywords—*Convolutional neural networks; image processing; semantic brain tumor segmentation; convolutional long short term memory; inception; squeeze-excitation; residual-network; attention units*

I. INTRODUCTION

In recent years, there have been a large number of contributions to the field of deep learning. Year after year, deep learning proves its superiority by surpassing state-of-the-art solutions in various domains such as computer vision [1], natural language processing [2], speech recognition [3] and many other application domains. In particular, in the field of computer vision, which is the main focus of this paper, deep learning has been enormously successful for tasks such as image classification [4], face recognition [5], object detection [1] and image segmentation [6]. The significant improvement in the field of deep learning is due to multiple factors. High-performance computational resources (GPU, TPU) have become more easily available. At the same time, the investments made in research as well as the amounts of collected data have also increased.

Deep learning algorithms, like other machine learning algorithms are categorized into two main categories: supervised and unsupervised algorithms. Unsupervised algorithms, an example of which is Deep Belief Network (DBN), work with unlabeled data. They use a greedy layer-wise learning strategy to fine-tune the network's parameters. This learning strategy, which is based on a contrastive version of the wake-sleep algorithm [7], performs quickly and can find a good set of parameters, even with relatively very deep architectures. Supervised algorithms, an example of which is Convolutional Neural Network (CNN), work with labeled data. CNNs have been particularly successful for solving computer vision tasks. The fine-tuning phase of a CNN is composed of consecutive convolution and pooling operations for extracting fine-tuned features, which are then used in the discriminative phase of the training process. This automated process of extracting features is what made deep learning algorithms powerful, as opposed to conventional machine learning algorithms that use hand-crafted features.

Nowadays, the use of CNNs is widespread across industries and businesses. In healthcare, CNNs achieve very promising results due to their robust feature extraction capabilities. For example, in medical image segmentation, they have achieved state-of-the-art performance [8] with a significant margin compared to conventional machine learning models, which makes them the most popular choice in different medical imaging fields. They also dominate the health informatics literature on brain [9], lung nodule [10], spleen [11], and cardiac [12] medical imaging issues, to mention a few.

In this work, we perform brain tumor semantic segmentation using a novel deep learning architecture. Brain tumors are considered one of the deadliest cancers in the world. There are various brain tumor types, but gliomas are the most common ones among adults. Furthermore, gliomas can be present with different degrees of aggressiveness with an average survival time for patients diagnosed with glioma lesser

than 14 months [13]. Therefore, time is a critical factor for doctors to act regarding gliomas. To diagnose a brain tumor, there are different types of medical image acquisition involved, such as MRI, CT scans and X-Ray, each having its pros and cons. For example, CT scans have the advantage of speed of tissue acquisition at the cost of lower quality of tissue contrast and higher radiation risk. On the other hand, MRIs are slow compared to CT scans but they are best suitable for capturing abnormal tissues with more details due to their accuracy in acquiring different types of contrasts. After the acquisition of the brain region, radiologists perform a manual segmentation of brain tumors from MRI images, which is time-consuming. Therefore, designing an automatic brain tumor segmentation is mostly desirable.

In this paper, we propose a novel deep neural network, called Multi-Scale ConvLSTM Attention Neural Network (MSConvLSTM-Att), to automatize brain tumor semantic segmentation. Our architecture is multi-scale-attention based with each level using Convolutional Long Short Term Memory (ConvLSTM) [14], Squeeze and Excitation-inception (SE-inception) [15] and Squeeze and Excitation-Residual-Network (SE-ResNet) [15]. The motivation behind using such architecture is to gather state-of-the-art feature extraction methods, the LSTM and the attention mechanism in one multi-scale architecture and perform brain tumor semantic segmentation efficiently compared to conventional deep learning based architectures.

The use of such multi-scale architecture, which is composed of multiple stages, is to generate multiple versions of the same image with different resolutions, each containing diverse semantics. The first low-level stage serves to model the spatially sequential relationship between different parts of each MRI modality (FLAIR, T1w,T1gd,T2w)[1], while the next stage manages the extraction of local features in addition to decreasing the size of the images for computational optimization. Finally, the third high-level stage captures the global representations. Thereafter, at each level, we introduce a stack of attention modules to gradually emphasize the regions that contain a large number of semantic features.

The integration of attention mechanism in the image segmentation of natural scenes has been widely adopted [16], [17], [18], [19]. However, in medical imaging, the inclusion of attention mechanism is rare [20], [21], [22], [23]. For this reason, we investigate the impact of a simple attention module in boosting the performance of standard deep networks for brain tumor semantic segmentation. Experimental results show that our proposed method improves the segmentation performance by modeling a combination of rich contextual features with local features.

The remainder of this paper is organized as follows. Next section presents related works. In Section III we introduce our proposed method in detail. Thereafter, we present and discuss the obtained results in section IV. Finally, we conclude our paper in section V.

II. Related Work

Most of the state-of-the-art deep learning architectures used for automatic medical image segmentation are inspired from

Fully Convolutional Networks (FCN) [24] or U-Net [25]. Many variants of these architectures have been proposed to perform semantic segmentation in different application domains [26], [27], [28], [29].

FCN is an architecture in which fully connected layers are replaced by deconvolution layers to generate segmentation masks [24]. Jesson *et al.* [30] proposed a variant of the standard FCN with a multi-scale loss function. With this approach it is possible to model the context in both the input and output domains. A limitation of this approach is that FCN is not able not explicitly model the context in the label domain. Compared to U-Net, FCN does not use skip connections between the contracting (i.e feature extraction path) and the expanding paths (i.e data reconstruction path).

The U-Net architecture was introduced by Ronneberger *et al.* in 2015 [25]. It overcomes the limitations of FCN by including features from the contracting path. In order to obtain the missing feature-contexts, multi-scale features are concatenated in a mirroring way. Many works have adopted this architecture to perform medical image segmentation over different parts of the human body. In a previous work of ours [26], we also adopted the U-Net architecture to perform lung CT image segmentation.

A limitation of both FCN and U-Net is that they both do not perform very well in multi-class segmentation tasks [31]. To overcome this issue, cascaded architectures can be used. They have the beneficial effect of decomposing a multi-class segmentation problem into multiple binary segmentation problems. This approach is also used in various medical image segmentation works. For example, Chen *et al.* [32] adopted a cascaded classifier to perform a multi-class segmentation. Furthermore, in [33] authors proposed a cascaded architecture to merge different feature extraction methods. Nonetheless, these models still face a problem of focusing on pixel level classification while ignoring adjacent pixels' connections. To overcome this issue, Generative models were adopted. A widely used variant of generative models is Generative Adversarial Network (GAN) [34]. GANs are employed for semantic segmentation in the following way: a convolutional semantic segmentation network is trained along with an adversarial network to discriminate segmentation maps [35]. That is, two models are trained; the first captures data distribution, while the second is used for a discriminative purpose.

To capture sequence patterns in medical imaging, Recurrent Neural Networks (RNNs) are typically used as they they are well suited for handling sequential data. Specifically in medical image segmentation, RNNs are used to keep track of features in previous image slices in order to better generate the corresponding segmentation maps. There are various RNN architectures mentioned in the literature, and amongst them Gated Recurrent Units (GRU) [36] and Long-Short Term Memory (LSTM) [37] are likely the most robust and widely used. GRU is memory efficient, nonetheless not very suitable for keeping track of long-term features. LSTM is better adapted to such tasks due to the forget gate that preserves features from previous sequences to use in upcoming sequences. [38], [39], [40] are some examples of employing RNNs for performing image segmentation for sclerosis lesions and brain tumors respectively.

---

[1]https://case.edu/med/neurology/NR/MRI Basics.htm

In the last few years, a new concept called *attention mechanism* was introduced into computer vision tasks. Attention mechanism was introduced first in neural machine translation [41] to help remember long range context from long source sentences. The added value brought by attention modules is the creation of shortcuts between the input sentence and the context vector. Attention in deep learning can be interpreted as a vector of weights that represent the importance of an element within a context. The attention vector is used to estimate how strongly is an element related to other elements (elements in this context are image pixels), it takes the sum of these elements' values weighted by the attention vector as the approximation of the target context.

The success of the attention mechanism for neural machine translation has encouraged its application to computer vision immediately [42]. In medical image segmentation, the attention mechanism was adopted in many works and various variants of attention modules have been introduced. In [43], authors propose a combination of FCN with a Squeeze and Excitation (SE) attention-based module to perform whole-brain and whole-body segmentation. They integrate the SE block in three ways: channel SE (cSE), spatial SE (sSE) and concurrent spatial-channel SE (csSE). In [20], Wang et al. perform prostate segmentation in ultrasound images using deep attentional features. They use an attention module to extract refined features at each layer, eliminate non-prostate noise and focus on more prostate details at deep layers. Furthermore, Li et al. propose an auto-encoder CNN-based architecture, called hierarchical aggregation network (HAANet) [21], which combines the attention mechanism and hierarchical aggregation to perform 3D left atrial segmentation. In another work, Oktay *et al.* propose an attention U-net [44] which extends the U-Net architecture by incorporating an attention gate in the expanding path in order to accurately segment the pancreas area.

III. METHOD

In this section, we describe our proposed architecture for brain tumor segmentation. Our method combines different techniques in order to extract relevant features and keep track of them during the entire process of segmentation.

We combine Inception, ResNet and Squeeze-Excitation blocks in one part of our architecture for relevant feature extraction, and attention modules in another part to perform brain tumor segmentation. The combination of Inception, ResNet and Squeeze-Excitation is known as the most successful architecture in the ImageNet challenge. With this combination, the team Trimps-Soushen achieved 2.99% error rate in object classification in the ImageNet challenge[2].

We first feed our network different modalities of brain MRI images (FLAIR, T1w, T1gd, T2w) to include various intensities and to better perform the semantic segmentation. Each modality is split into four patches, then for each modality, three scales of feature extraction are performed. The motivation behind this multi-scale mixture is to best separate each tumor label (enhancing tumor, tumor core, whole tumor and background).

At the first scale, ConvLSTM is used over each of the four patches to preserve the correlation among features. ConvL-STM are best suitable for catching spatiotemporal information without any much redundancy [14]. At the second scale, an SE-inception [45] module is used over the output of the first scale to extract low level features and decrease the computation cost. Fig. 1 shows the inception module [46] and Fig. 2 shows the SE-inception block.
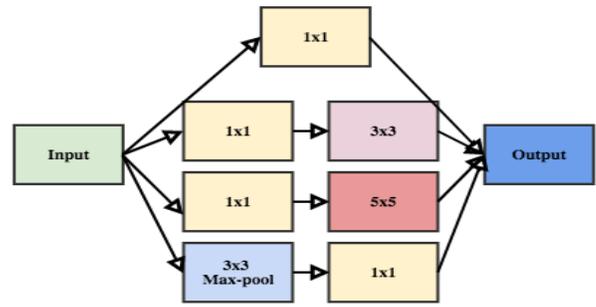
Fig. 1. Inception Block.
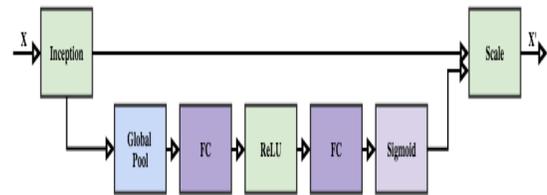
Fig. 2. SE-Inception Block.

At the third scale, we extract high level features by integrating an SE-ResNet module [45]. The use of such block increases computational complexity with a thin margin but in exchange of increasing the accuracy [45]. The ResNet block [47] and SE-ResNet are described in Fig. 3 and Fig. 4 respectively.

At each scale (different scales are highlighted by green color in Fig. 6), we combine the four outputs to form what we call *single-scale features* as stated in Fig. 6. These three *single-scale features* are then concatenated and convolved to form *multi-scale features* as mentioned in the same figure. We then take the *multi-scale features* and we combine them with each single-scale feature.

At this stage, our model holds general context feature-maps that contain different levels of features, from low to high level features. Thereafter, we add a convolution layer to refine these features.

Furthermore, in order to explore more global contextual characteristics by building connections among features, we

---

[2]https://image-net.org/challenges/beyond_ilsvrc
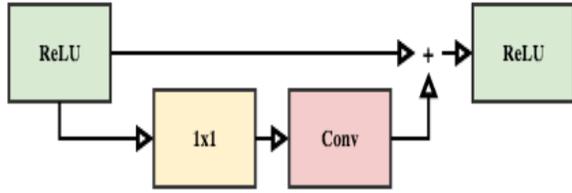
Fig. 3. ResNet Block.
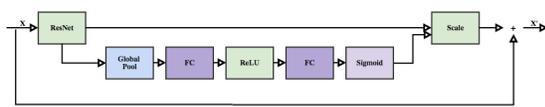
Fig. 4. SE-ResNet Block.

include attention mechanism in the form of a location-based attention module, we call it Spatial Attention Module (SAM). The attention mechanism is presented in the Fig. 5.
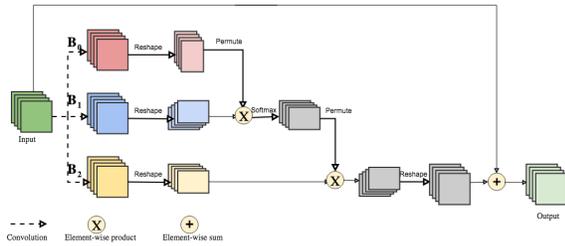
Fig. 5. Spatial Attention Module.

In Fig. 5, we assume the input to the SAM module is V, which is a 3D shaped input (W, H, C), here W, H and C represent the width, height and depth respectively. In the red branch, we perform a convolution operation, resulting in a feature map B_0 with same width and height but with depth equals to $C/8$. B_0 is then reshaped to (W, H, C). The same operation is applied to the blue branch B_1. Thereafter, we perform a matrix multiplication $B_0 * B_1^T$ and apply a softmax operation to calculate the spatial attention map following the formula in (1), where $S_{i,j}$ represents the impact of the pixel in the ith position on the pixel in the jth position.

$$S_{i,j} = \frac{exp(B_0 * B_1)}{\sum\limits_{i=1}^{W*H} exp(B_0 * B_1)} \qquad (1)$$

The yellow branch performs a convolution and results in B_2 with the same shape as V. B_2 is then reshaped to (C, W, H)

then it is multiplied by the transpose of the spatial attention map S. Furthermore, the output R is reshaped to $C \times (W \times H)$ and multiplied by a parameter λ and then an element-wise sum with input V is performed to obtain the output O as expressed in (2).

$$O = \lambda \sum\limits_{i=1}^{W*H} exp(S_{i,j} * B_2) + V \qquad (2)$$

In (2), λ is initialized by 0 and gradually updated to give more weight to the spatial attention map, as adopted in [17].

At the last level, we perform a convolution operation to generate the final prediction map for each scale and then average all these maps to output the segmentation map. Fig. 6 presents an overview of our proposed architecture.

IV. EXPERIMENTS AND RESULTS

To evaluate our architecture, we are using BRATS'18 data set for brain tumor segmentation, provided in the Medical Segmentation Decathlon Challenge[3]. This data set contains multimodal MRI data (FLAIR, T1w,T1gd,T2w)[4]. Furthermore, it contains 210 High Grade Glioma (HGG) scans and 75 Low Grade Glioma (LGG) scans. In this data set, the focus is mainly on the segmentation of different sub-regions of the glioma. First, the enhancing tumor (ET), the tumor core (TC) and finally the whole tumor (WT) as can be seen in Fig. 7. Each one of these sub-regions have some specific characteristics regarding their intensities, hence different modalities are responsible for capturing different characteristics. For example, the ET is described by areas that are hyper-intense in T1gd. The appearance of the non-enhancing tumor (NET) (solid parts) and the necrotic (NCR) (fluid-filled) is represented by areas that show hypo-intensity in T1gd when compared to T1. The WT describes the whole disease and it contains the TC and the peritumoral edema (ED), which is characterized by hyper-intensity in the FLAIR modality. The provided labels in this data set are as follows: 1 for NCR and NET, 2 for ED, 3 for ET and finally 0 for other parts of the brain. The annotations were created by domain experts and approved by other domain experts as described in [48].

Given the presence of different features related to gliomas in different modalities, we feed the four modalities as input to our architecture, then we get the semantic segmentation that belongs to these inputs. The loss function we use is the dice loss optimized using the Adam optimizer [49]. The learning rate is initially set to 0.001 and then multiplied by 0.5 after each 30 epochs. We used 500 epochs to train our network. Due to limitations in computational resources, we reduced the input size to 190×190 by cropping some of the background area and we only took from the 30th slice to the 120th given that most of the brain information is present in that interval. Furthermore, we normalize the inputs to have zero mean and unit standard deviation.

In addition, given that each session of the notebook used for training has 12 hours lifetime, we use the following strategy to train our network. We save our model and its weights after each 50 epochs and we reload it and continue training with

---

[3]http://medicaldecathlon.com
[4]https://case.edu/med/neurology/NR/MRI Basics.htm
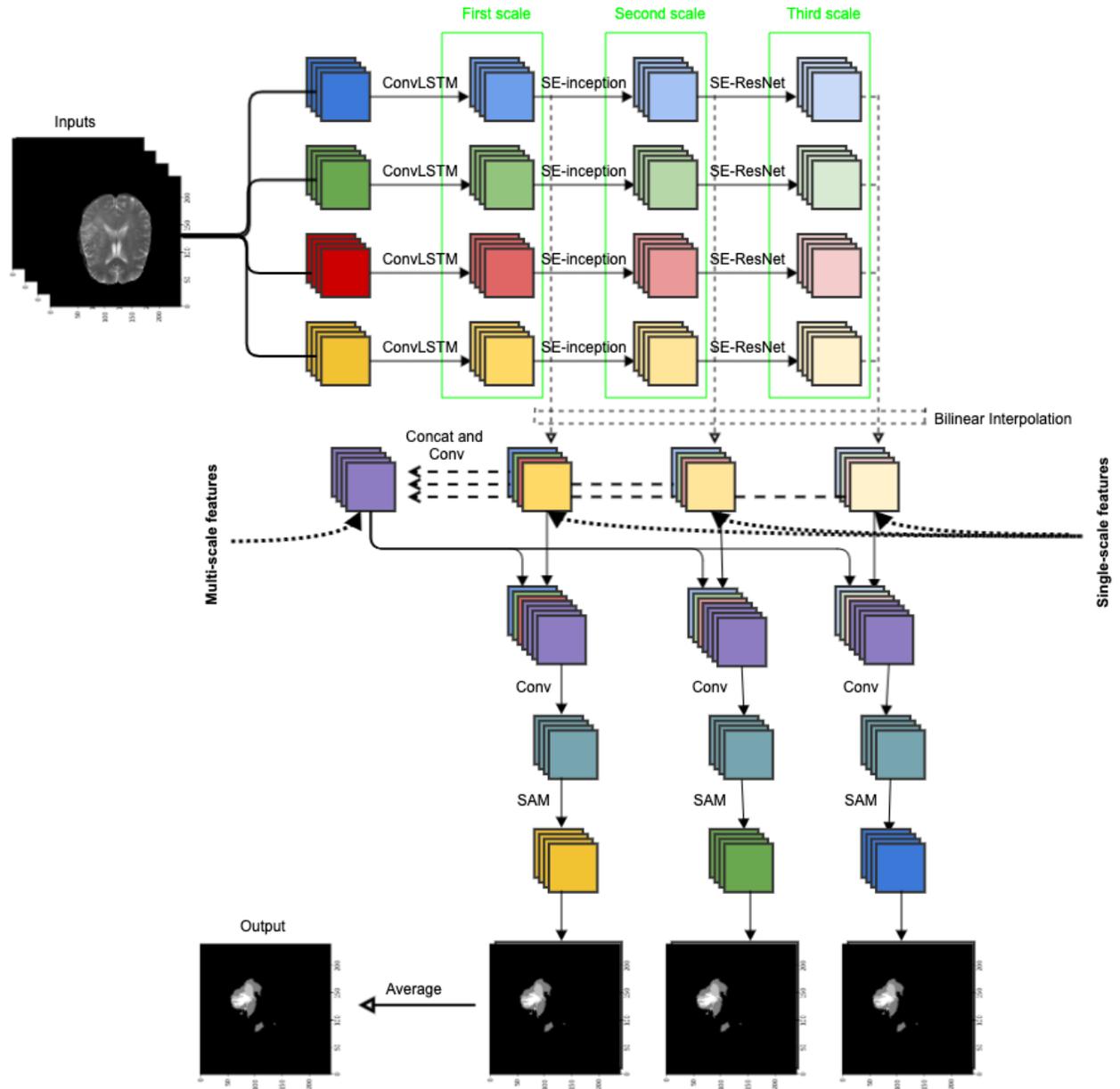
Fig. 6. Overview of our Proposed Architecture.

new data. For development, we shuffled and randomly split the images into training (225 patients), validation (30 patients), and test (30 patients). Experiments were performed in a server equipped with a single 12GB NVIDIA Tesla K80 GPU.

We compare our method with the standard UNet [25], standard FCN [24] and the Attention U-Net [44] architectures. And we evaluate their performance using the dice coefficient (DSC) as a comparison metric. Table I contains experimental results obtained using the different segmentation methods described above, and compared regarding their DSC score. Our proposed architecture achieved the best score with 64.95, 88.16 and 86.50 in ET, WT, and TC respectively and a mean score

of 79.87.

It can be observed that both our method and AttUnet, which also includes attention modules, perform better than the other ones without attention modules. This proves that adding attention modules surely enhances the segmentation procedure by putting more attention into the tumor location. Oktay *et al.* have reported the same observation in their work with MSConvLSTM-Att [44].

Our architecture outperforms AttUNet with a significant margin, this is mainly due to the focus on location attention modules, besides the use of powerful feature extraction modules (ConvLSTM, SE-Inception and SE-ResNet) in the
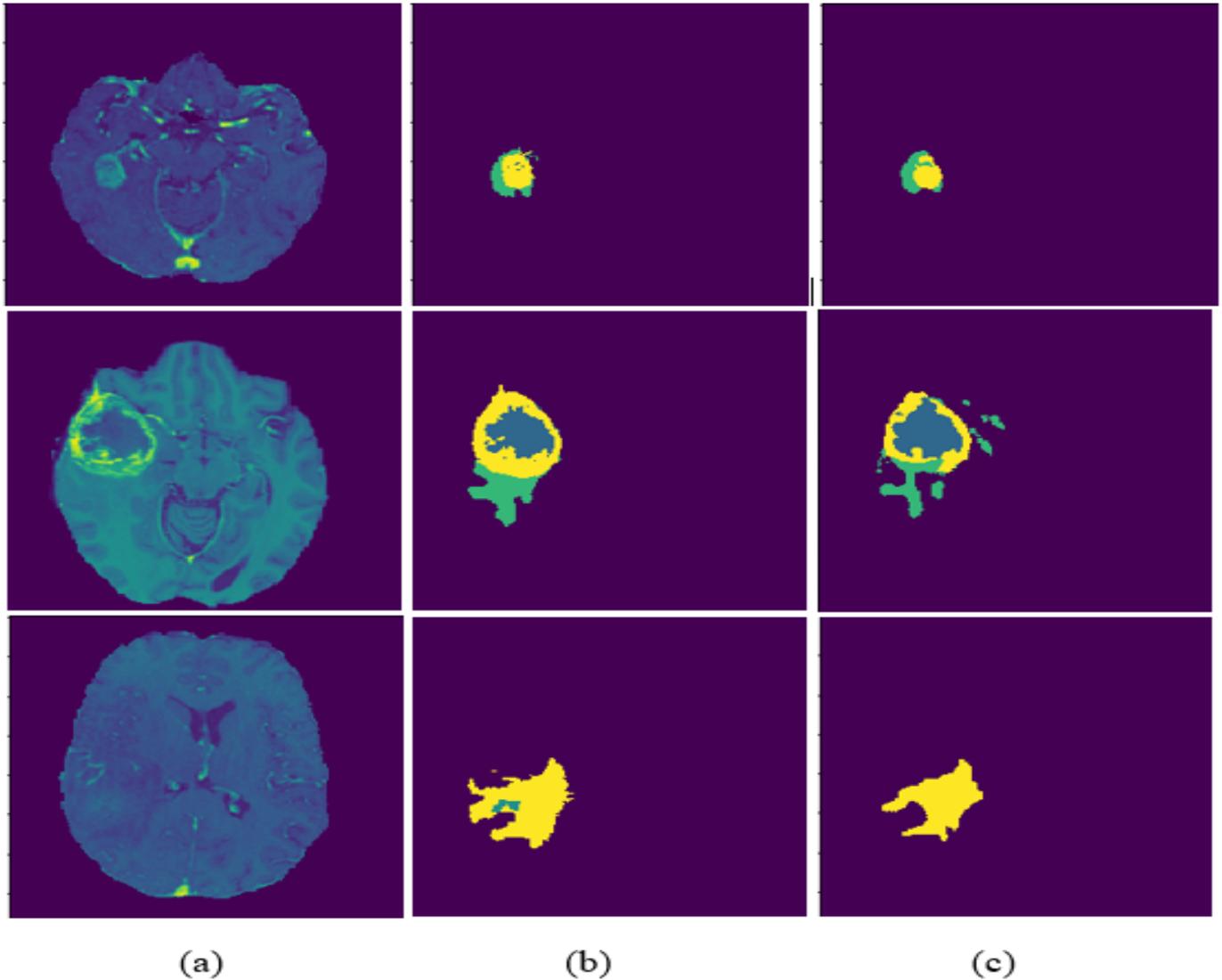
Fig. 7. Segmentation Results Sample: (a) is the Input MRI Images, (b) is the Ground Truth and (c) is the Segmentation Results using our Proposed Architecture.

TABLE I. PROPOSED METHOD'S DSC SCORE COMPARED TO THOSE OF U-NET, ATT-UNET AND FCN

Labels	ET	WT	CT	Mean
U-Net	0.563	0.848	0.797	**0.736**
Att-UNet	0.637	0.875	0.845	**0.786**
FCN	0.551	0.853	0.781	**0.728**
ours	0.649	0.881	0.865	**0.798**

first part of the architecture, which is beneficial in eliminating irrelevant features. Our proposed architecture can be implicitly considered as a cascaded architecture even though we do not explicitly use multiple cascaded architectures.

Fig. 7 displays a sample of the input MRI images, ground truth and the segmentation results using our proposed architecture. As seen in Table I, the ET segmentation has the smallest DSC value. It can be seen also in Fig. 7, where the ET region is not well detected especially in the first and third row.

It has to be mentioned that our method is slightly slower compared to the other methods, which is normal given the fact that complex building blocks has been used in order to ensure a better segmentation result.

V. CONCLUSION

In this paper, we propose a novel deep learning architecture for brain tumor segmentation we call multi-scale ConvLSTM Attention Neural Network, and we compare its performance to various deep learning architectures that are tailored to such kind of tasks. Our proposed method is built as a multi-scale

architecture composed of different state-of-the-art feature extraction blocks such as Inception, Squeeze-Excitation, Residual Network, ConvLSTM and finally Attention units. We compare the performance of our architecture to standard U-net, AttU-net and FCN that have shown effective results in semantic segmentation. Experimental results show that our proposed model outperforms standard U-net, AttU-net and FCN in terms of dice score. Our model reached 79.78 as a mean dice score for the three parts of the brain tumor, while Attention U-net, standard U-net and FCN reached 78.61, 73.65 and 72.89 respectively. We observe that both our method and the AttU-net perform better than the other ones, which can be explained that the integration of attention modules enhances the segmentation procedure. Besides, our method outperforms the AttU-net, and this is due to the use ConvLSTM, SE-Inception and SE-ResNet.

REFERENCES

[1] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186.

[3] S. Hourri, N. S. Nikolov, and J. Kharroubi, "Convolutional neural network vectors for speaker recognition," *International Journal of Speech Technology*, pp. 1–12, 2021.

[4] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[5] J. Wei, "Video face recognition of virtual currency trading system based on deep learning algorithms," *IEEE Access*, vol. 9, pp. 32 760–32 773, 2021.

[6] Y. Jalali, M. Fateh, M. Rezvani, V. Abolghasemi, and M. H. Anisi, "Resbcdu-net: A deep learning framework for lung ct image segmentation," *Sensors*, vol. 21, no. 1, p. 268, 2021.

[7] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[8] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[9] A. Myronenko, "3d mri brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 311–320.

[10] A. E. HASSANI, B. A. SKOURT, and A. MAJDA, "Efficient lung nodule classification method using convolutional neural network and discrete cosine transform," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 2, 2021. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2021.0120296

[11] H. R. Roth, H. Oda, Y. Hayashi, M. Oda, N. Shimizu, M. Fujiwara, K. Misawa, and K. Mori, "Hierarchical 3d fully convolutional networks for multi-organ segmentation," *arXiv preprint arXiv:1704.06382*, 2017.

[12] R.-R. Galea, L. Diosan, A. Andreica, L. Popa, S. Manole, and Z. Bálint, "Region-of-interest-based cardiac image segmentation with deep learning," *Applied Sciences*, vol. 11, no. 4, p. 1965, 2021.

[13] M. S. Walid, "Prognostic factors for long-term survival after glioblastoma," *The Permanente Journal*, vol. 12, no. 4, p. 45, 2008.

[14] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 2015, pp. 802–810, 2015.

[15] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[16] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid attention network for semantic segmentation," *arXiv preprint arXiv:1805.10180*, 2018.

[17] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3146–3154.

[18] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, "Attention to scale: Scale-aware semantic image segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3640–3649.

[19] H. Zhao, Y. Zhang, S. Liu, J. Shi, C. C. Loy, D. Lin, and J. Jia, "Psanet: Point-wise spatial attention network for scene parsing," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 267–283.

[20] Y. Wang, Z. Deng, X. Hu, L. Zhu, X. Yang, X. Xu, P.-A. Heng, and D. Ni, "Deep attentional features for prostate segmentation in ultrasound," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 523–530.

[21] C. Li, Q. Tong, X. Liao, W. Si, Y. Sun, Q. Wang, and P.-A. Heng, "Attention based hierarchical aggregation network for 3d left atrial segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2018, pp. 255–264.

[22] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical image analysis*, vol. 53, pp. 197–207, 2019.

[23] D. Nie, Y. Gao, L. Wang, and D. Shen, "Asdnet: attention based semi-supervised deep networks for medical image segmentation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2018, pp. 370–378.

[24] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[26] B. A. Skourt, A. El Hassani, and A. Majda, "Lung ct image segmentation using deep neural networks," *Procedia Computer Science*, vol. 127, pp. 109–113, 2018.

[27] J. Dolz, X. Xu, J. Rony, J. Yuan, Y. Liu, E. Granger, C. Desrosiers, X. Zhang, I. Ben Ayed, and H. Lu, "Multiregion segmentation of bladder cancer structures in mri with progressive dilated convolutional networks," *Medical physics*, vol. 45, no. 12, pp. 5482–5493, 2018.

[28] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.

[29] M. P. Heinrich, O. Oktay, and N. Bouteldja, "Obelisk-net: Fewer layers to solve 3d multi-organ segmentation with sparse deformable convolutions," *Medical image analysis*, vol. 54, pp. 1–9, 2019.

[30] A. Jesson and T. Arbel, "Brain tumor segmentation using a 3d fcn with multi-scale loss," in *International MICCAI Brainlesion Workshop*. Springer, 2017, pp. 392–402.

[31] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18–31, 2017.

[32] X. Chen, J. H. Liew, W. Xiong, C.-K. Chui, and S.-H. Ong, "Focus, segment and erase: an efficient network for multi-label brain tumor segmentation," in *Proceedings of the european conference on computer vision (ECCV)*, 2018, pp. 654–669.

[33] J. Liu, F. Chen, C. Pan, M. Zhu, X. Zhang, L. Zhang, and H. Liao, "A cascaded deep convolutional neural network for joint segmentation and genotype prediction of brainstem gliomas," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1943–1952, 2018.

[34] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[35] N. Souly, C. Spampinato, and M. Shah, "Semi supervised semantic segmentation using generative adversarial network," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[36] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.

[37] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[38] S. Andermatt, S. Pezold, and P. C. Cattin, "Automated segmentation of multiple sclerosis lesions using multi-dimensional gated recurrent units," in *International MICCAI Brainlesion Workshop*. Springer, 2017, pp. 31–42.

[39] T. H. N. Le, R. Gummadi, and M. Savvides, "Deep recurrent level set for segmenting brain tumors," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 646–653.

[40] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating fcnns and crfs for brain tumor segmentation," *Medical image analysis*, vol. 43, pp. 98–111, 2018.

[41] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[42] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *International conference on machine learning*. PMLR, 2015, pp. 2048–2057.

[43] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2018, pp. 421–429.

[44] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[45] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[46] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[48] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge," *arXiv preprint arXiv:1811.02629*, 2018.

[49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

NGram Approach for Semantic Similarity on Arabic Short Text

Rana Husni Al-Mahmoud
Faculty of Information Technology
Applied Science Private University
Amman, Jordan

Ahmad Sharieh
Computer Science
Department King Abdullah II School of Information Technology
The University of Jordan
Amman, Jordan

Abstract—Measuring the semantic similarity between words requires a method that can simulate human thought. The use of computers to quantify and compare semantic similarities has become an important research area in various fields, including artificial intelligence, knowledge management, information retrieval, and natural language processing. Computational semantics require efficient measures for computing concept similarity, which still need to be developed. Several computational measures quantify semantic similarity based on knowledge resources such as the WordNet taxonomy. Several measures based on taxonomical parameters have been applied to optimize the expression for content semantics. This paper presents a new similarity measure for quantifying the semantic similarity between concepts, words, sentences, short text, and long text based on NGram features and Synonyms of NGram related to the same domain. The proposed algorithm was tested on 700 tweets, and the semantic similarity values were compared with cosine similarity on the same dataset. The results were analyzed manually by a domain expert who concluded that the values provided by the proposed algorithm were better than the cosine similarity values within the selected domain regarding the semantic similarity between the datasets' short texts.

Keywords—*Arabic text; Ngram; semantic sentences similarity; short text; ALMaany; natural language; semantic similarity of words; corpus-based measures*

I. INTRODUCTION

In this paper, semantic similarity is estimated by considering the similarity between bigrams synonyms related to the same domain. Paraphrase identification can detect different linguistic expressions with the same meaning or similar meanings [1]. Analyzing the similarity of meanings is part of the semantic text similarity task.

Recent advances have made social media a major source of news, with users flooded with news about similar events. Paraphrasing news articles and recognizing semantic similarities between them is a useful practice both in many general natural language processing applications and in new event detection (first story detection) on specific events [1].

For a long time, word semantic similarity has been essential to the processing of natural language and information retrieval (IR) [2]. For instance, in academic and industrial communities alike, semantic similarity has become a vital aspect of various applications in various fields. Word sense disambiguation, information retrieval, semantic searches, and explorations of biological macromolecules are prominent examples of semantic similarity applications [3]. Furthermore, it is possible to understand and categorize documents and obtain informative knowledge using semantic annotation [2].

Semantic text similarity measures the semantic similarity between two texts (documents, paragraphs, sentences, or a combination thereof). Most of the work to date on such measures has been done at the document level (that is, comparing two long texts, or one long text and one short text). Sentence-level analysis has received a lot of attention recently. As a result, training and test data were provided in multiple languages, and different approaches were developed for detecting sentence similarity. These approaches are generally classified into three types: vector space approaches, registration approaches, and other approaches such as: B. Use topic modeling for feature extraction [1].

Typically, the process of detecting two text segments' level of similarity involves, first, employing a straightforward lexical matching method and then detecting how many lexical units are contained in both input segments to calculate the similarity score [4]. This method can be improved by employing various techniques (e.g., stemming, part-of-speech tagging) or by considering different weighting and normalization factors. While these lexical similarity methods have been somewhat successful, they sometimes fail to adequately identify the semantic similarity between two texts. For instance, even though the phrases "I own a dog" and "I have an animal" are clearly similar, most contemporary similarity detection techniques do not recognize this. Often, knowledge-based or corpus-based approaches are used to detect semantic similarity at the word level [4]. These approaches have shown some success, particularly when applied to language processing tasks. Two of the most popular text-based semantic similarity approaches are to use approximations generated by query expansion and to employ the latent semantic analysis method. The former is often used for information retrieval tasks, while the latter is used to detect the similarity between texts by automatically acquiring second-order word relations from extensive collections of texts [4]. Other noteworthy methods for detecting semantic similarity are listed below [5]:

1) Structure-based measures, which use a function that computes the semantic similarity measures on ontology hierarchy structures.

2) Information content measures, which are based on the frequency of terms in a given document.

3) Feature-based measures, by which each term is described by a set of features and the similarity measure between two terms is defined as a function of their

properties.

4) Hybrid measures, which combine the structural characteristics of the previous methods to compute semantic similarity.

The following are basics and backgrounds knowledge:

WordNet created as part of a research project at Princeton University [6]. This is an extensive English vocabulary database. In WordNet, nouns, verbs, adverbs, and adjectives are organized by semantic relationships into synsets, each representing a concept.

Semantic similarity (or topological similarity) detects the similarity between terms, sentences, and documents. Similarity between sentences and documents is calculated by considering terms that describe internal concepts. Similarity at the sentence level is detected using syntactical and lexical measures [7].

The syntactic approach primarily uses syntactic dependencies to recognize semantic similarities and build a more comprehensive picture of the meaning of the compared texts. In this way, these approaches identify whether a noun is the subject or object of a verb. Lexical-based similarity approaches, on the other hand, measure similarity between texts based on character matching.

Three problems with the existing semantic measurements are the primary motivations for this work. The first issue relates to how text is represented and similarity calculations are performed. Text representation mainly concerns converting text to vectors by using lexical representation or word embedding representation. The drawback of the first one is that it depends on the occurrences of words in the text, either it occurred in the same order of occurrence or not. And it is a critical point in the semantic similarity measure and the limitation of word embedding. Most word embedding models are trained on corpora in different domains and the semantic similarity degree of the keywords depending on the domain concepts. The second issue relates to the external dictionary and ontologies. Most of them were static and not concerned about the same topic. For example, in Arabic, there is a need for more presence of these dictionaries. The third issue is the need to represent the two texts. In some applications, like in plagiarism, there is a need to find similarities between a fragment of sentences.

These problems motivate us to current work. This work depends on an updated dictionary ALMaany [8], and all needed synonyms extracted depending on the same domain. In addition, the proposed algorithm can be applied to various varieties of text length, and consider the order of keywords by taking NGram words from all texts. The present work describes a new method for measuring semantic similarity between words and concepts that uses Ngram synonyms connected to the same domain.

The first step of the proposed algorithm is crawling articles from sites to collect the most frequent words for the same domain. The extracted keywords used in the following steps:

1) Searching for tweets depending on extracted words.
2) Extracting synonyms from ALMaany [8] and concentrate on the selected domain or topic.

The proposed algorithm depends on the synonyms, and NGrams was evaluated based on 700 tweets and compared the proposed algorithm's output with the cosine similarity values. An expert assessed the results and determined that the proposed algorithm detected the semantic similarity between the dataset's short sentences more accurately than the cosine similarity values.

The remainder of the paper is organized as follows. Section II discusses previous works related to semantic similarity measures. The proposed method is described in Section III. The experimental results are then discussed in Section IV. Finally, the paper is concluded and directions for future studies are recommended in Section V.

II. RELATED WORKS

Semantic text similarity is a measure of the degree of semantic similarity between two texts, such as documents, paragraphs, sentences, words, or a combination of them [1]. Various semantic similarity approaches have been described and summarized in many surveys. For example, [9] presented the fundamental aspects of the theoretical and practical backgrounds of semantic similarity assessments of texts. They also discussed the general technology used for sophisticated text analyses (i.e., text mining), alongside discussions of relevant methodologies, architectures, and challenges. In other work, [10] explored the development of semantic similarity methods. They classified different methods as knowledge-based, corpus-based, deep neural network-based, and hybrid methods, according to their underlying principles. It starts with traditional NLP techniques (e.g., kernel-based methods) and progresses to the most recent research on transformer-based models while examining each approach's merits and disadvantages.

[11] reviewed existing approaches to measuring semantic similarity at either the document, sentence, or word level, focusing on Arabic text. The approach utilized by [1], [12] employs a set of extracted features based on lexical, syntactic, and semantic computations to detect the similarity between tweet pairs. One approach uses knowledge and corpora to express the meanings of terms to solve the issue of polysemy and includes a constituency parse tree to capture the syntactic structures of short texts. The approach also uses word alignment features to detect the similarity between tweet pairs.

Semantics is an essential aspect of studies on natural language processing. Previously, in [13], they surveyed various deep learning approaches that have been used to detect the semantics of words, sentences, and documents. However, most previous studies have considered the semantics only of documents (i.e., either two long texts or one long text and one short text have been compared). Recently, though, comparisons between individual sentences have become more common [1]. Previous researchers have also used measured semantic similarity to compare words or concepts. However, such measures are rarely used to compare multi-word phrases [4]. Three broad categories of semantic similarity detection methods are used to determine the level of similarity between words: Dictionary/ontology-based methods consider knowledge bases to gather the semantic information that is compared when determining semantic similarity [14]. Meanwhile, corpus-based methods primarily use word frequencies to determine semantic similarity. This is done based on statistics taken from extensive corpora. Finally, hybrid methods consider more than one information source to determine semantic similarity [14].

The statistical methods employed by corpus-based approaches have recently evolved. Thus, such approaches can follow one of two principal orientations [14]:

- The first is the unsupervised orientation, which involves the use of training sets and unannotated corpora. Approaches that follow this orientation can be further divided depending on the method of discrimination used (type-based or token-based). When type-based discrimination is employed, the similarity is measured by an algorithm after the contexts have been represented, which is done via high-dimensional spaces, which are defined by word co-occurrences. Meanwhile, when token-based discrimination is used, all contexts containing the target word are clustered together. Each resultant cluster comprises contexts that contain similar usages of the target word.

- The second orientation includes supervised and semi-supervised approaches, by which an annotated training corpus with the appropriate classification models is applied. Supervised methods include probabilistic methods, and they typically employ the naive Bayes algorithm and follow the maximum entropy approach [14]. Which methods are followed when using such an approach depends on how similar the evaluated examples are. These methods compare sets of learned vector prototypes using a similarity metric. This is done for each word sense. Meanwhile, other methods consider discriminating rules to make comparisons. Such methods rely on specific rules that apply to each word sense. In turn, methods based on these specific rules merge heterogeneous learning modules [14].

The Arabic language is an official language used by the United Nations, and more than 450 million people in the world speak Arabic as their first language [1]. The vocabulary of this language is rich, and its morphology is complex. It is also a synthetic language, meaning that a given morpheme can comprise a stem and affixes, which can indicate different aspects (e.g., tense, gender, and what word class a word belongs to). Moreover, different parts of speech can be affixed to each other. Arabic is a derivational, flexional, and agglutinative language. These characteristics make it difficult to conduct research on language processing and text mining, as special tools and resources are needed. An additional problem arises from the fact that the lexical and morphological features of Arabic have a profound effect on sentence analyses. If the research question addressed in this study is to be adequately answered, such challenges must be overcome. Much research has focused on various problems related to analyses of semantic similarity and developed methods for overcoming these problems. However, most of these methods are either domain-dependent or language-dependent. Moreover, little research on this matter has examined Arabic [1]. Another shortcoming of previous studies focusing on the Arabic language is that the semantic similarity analyses employed have not utilized enough resources (e.g., tools and benchmark data) due to a lack of availability. One such research work was conducted by [15], who determined semantic similarity at the sentence level using supervised learning. Specifically, their method analyzed semantic, lexical, and syntactic-semantic features, which were extracted using an Arabic dictionary, a lexical

markup framework, and a learning corpus. After the method was used, its outcomes were assessed by Weka; the assessment showed that the proposed model produced highly accurate results [1]. However, the results were not as favorable when the method was used to detect semantic similarity between phrases and sentences. This is because, compared to word-level estimations, sentence-level estimations are substantially more challenging to perform since sentence-level semantics are noncompositional and involve many more possible interpretations.

When considering the Arabic language, similarity approaches face several significant challenges [16]:

- Arabic is a complicated (and often ambiguous) language.

- Arabic WordNet is a multilingual concept dictionary that maps Arabic word senses with their equivalents in English WordNet [17]. However, the Arabic database was built manually and does not contain sufficient essential information. It also contains many fewer concepts than English WordNet, and it is lacking several important semantic relations between synsets.

- Few Arabic corpora consider all possible domains and words. This is because each Arabic corpus focuses on only one domain; thus, these corpora do not contain all essential information.

Based on the above points, the cosine similarity measurement has been employed in many Arabic systems. Results show that this measure outperforms other lexical measurements.

Lexical similarity methods are unreliable when assessing the Arabic language because of the language's unique features, such as its morphology. Furthermore, the semantic similarity approach is undesirable when considering Arabic because of the aforementioned shortcomings of Arabic WordNet and Arabic corpora. Recently, the hybrid similarity approach has been considered potentially useful for examining semantic similarity in Arabic since it utilizes multiple measurement methods, thus providing more robust analyses than other techniques [16].

Twitter is a fast-growing social media tool with which people can connect and share microblog posts called tweets [18]. This tool also produces vast amounts of information. We have considered tweets in our research because tweets are limited to 280 characters. Thus, compared to the text posted on other social media platforms like Facebook (which has no post length limitations), tweets are brief yet tell complete stories that can be compared relatively easily.

Different methods for semantic similarity approaches have been recently proposed based on the aforementioned algorithms. Most of recent works based on word embedding techniques. Authors in [19] applied Word2Vec model on an English corpus to represent words in vector form. Then a Cosine Similarity method was used to calculate the similarity value. Authors in [20] presented an approach that combining LDA topic model approach with BERT word embedding for pairwise semantic similarity detection. A hybrid approach based on Word Embedding and External Knowledge Sources was used to find the semantic similarity value between two

short text. Another hybrid approach based on a WordNet proposed in [21] to measure concept semantic similarity.

This is clearly evident from previous works that most of suggested approaches applied on English datasets. Therefore, more research effort is needed for computing semantic similarity for Arabic Language. In summary, a good amount of work has been invested to calculate semantic similarity either depending on external static external knowledge or by using word-embedding models that created on large corpus that is not related with the tested datasets. The reduction in problem dimension at the expense of the real values' interpretability is one of the key drawbacks of word embedding that form the vector representations [22]. And, due to Arabic WordNet limitation in keyword synonyms [23], [24], we chose ALMaany [8] to extract synonyms related to Arabic keywords within specific domain. We depended on ALMaany because it is one of the most recent dictionary and continuously updated. In addition, ALMaany is fast, free, electronic and easy to use [25], [26].

Most previous related works have considered a single corpus or a dataset when detecting the semantic similarity between sentences or documents. Differently, this work proposes a new method that considers n-gram synonyms within a single domain to detect the semantic similarity between concepts and words. Furthermore, the contributions of this work are relevant to any sentence, paragraph, or document.

III. PROPOSED WORK

Due to its nature, the Arabic language may allow more than one meaning (and sometimes opposite meanings) to be assigned to the same word. Therefore, semantic similarity detection methods should find similarities between words related to the same domain. Because Arabic WordNet is limited to extract synonyms for Arabic terms within a particular domain [16], we choose ALMaany's [8]. ALMaany is essential to us because it is one of the most modern dictionaries and is regularly updated it is quick to access, free, computerized, and simple to use.

In this work, we considered a common news topic, namely the current relationship between Qatar and the UAE during the last quarter of 2017.

The following steps were conducted to find synonyms of the most frequently used keywords. First, sources were searched for relevant articles through Google. This was done because Twitter has a short-text format, and this step ensured that we would consider all keywords that could be found in tweets mentioning the news topic of interest. The data sources considered in this work were the websites of news agencies such as Reuters, news channels such as Aljazeera, and online versions of printed newspapers such as the Middle East. The sources used to obtain articles are presented in Table II. We searched for the main keywords, such as those found in Table I. Initially, we found almost 10,000 articles from online sources. After the main keywords were used to filter the articles, around 3000 remained. Table III shows a sample of the search results.

Second, we found the keywords most frequently used in the articles after removing stop words. Table V shows some of these words.

Third, we manually extracted synonyms related to the topic under investigation from ALMaany [8]. Table VI presents some of these synonyms.

The dataset of articles was used to find the most frequent words that needed to be used to extract tweets from Twitter. In an initial step, we need to prove our algorithm on short Arabic text; and then, in future works, it will be applied in paragraphs, and after that, on long articles

REST APIs and Streaming APIs make up most of the Twitter APIs[1]. Use the RESTful state transfer (REST) search API to search tweets from Twitter's search index. The REST API offers historical results going back as long as the search index allows (usually last seven days). The streaming API, however, returns information from the query's starting point. Real-time monitoring of a particular query is possible using streaming API. According to their website, Twitter's search API contains several restrictions.[2]. We developed a number of searches that include all combinations of the most frequent keywords extracted from websites' articles. Table IV shows a sample of the tweets. The results were filtered, and only tweets containing "إمارات" and "قطر" and "امارات" or "يمن" were kept.

TABLE I. LIST OF MAIN KEYWORDS THAT ARE USED IN SEARCHING FOR ARTICLES

امارات	الامارات	إمارات	الإمارات	قطر
السعودية	الرياض	دبي	أبوظبي	الدوحة

TABLE II. LIST OF SITES THAT ARE USED IN SEARCHING FOR ARTICLES

Site Url
www.alarabiya.net
www.skynewsarabia.com
www.dw.com/ar
www.bbc.com/arabic
www.france24.com/ar
www.alhurra.com
ara.reuters.com/
www.trt.net.tr/arabic
www.anb-tv.net/Arabian
www.arab48.com
www.arabi21.com
www.thenewkhalij.net
www.alhayat.com
www.alkhaleej.ae
www.aawsat.com
www.alarab.qa
arabic.rt.com
www.afp.com/ar
alkhaleejonline.net
www.cnbcarabia.com
www.middle-east-online.com
www.moheet.com
www.anntv.tv
www.huffpostarabi.com
arabic.cnn.com
www.aljazeera.net

In general, the proposed algorithm was employed to estimate the semantic similarity value of two short texts via the following process:

1) Take bigrams and trigrams.

---

[1] https://developer.twitter.com/en/docs/basics/getting-started
[2] https://developer.twitter.com/en/docs/basics/rate-limiting

TABLE III. SAMPLE OF ARTICLES

Article URL	Article Title	Article Content
http://www.huffpostarabi.com/mohammed-jamea/story_b_10885308.html	جولة نتنياهو الإفريقية والغياب العربي	اكتست الجولة التي قام بها رئيس الوزراء الإسرائيلي بنيامين نتنياهو خلال اليومين الماضيين إلى أربع دول إفريقية، أهمية بالغة لعدة أسباب، حيث إنها الزيارة الأولى لأرفع مسؤول إسرائيلي للمنطقة منذ فترة طويلة، إضافة إلى أن وفد نتنياهو يضم 80 رجل أعمال يمثلون 50 شركة إسرائيلية، وهذا مما يدل على سعيه لتعزيز التبادل التجاري والمزيد من التغلغل في القارة السمراء.
http://www.huffpostarabi.com/2017/01/19/story_n_14265640.html	اكتشاف خلل تقني في نسخة iOS 10 يسبب انهيار هواتف آيفون	اكتشف أحد مستخدمي موقع يوتيوب وجودَ خلل غريب في النسخة العاشرة من نظام تشغيل الأجهزة المحمولة iOS 10، حيث يسمح ذلك الخلل للمخربين بتعطيل أي هاتف آيفون أو حاسب آيباد، عبر إرسال رسالة نصية تحتوي على الرموز التعبيرية العلم وقوس القزح.
http://www.huffpostarabi.com/2017/04/04/story_n_15808882.html	ليس العداء للإرهاب فقط سرّ الكيمياء الشخصية بين ترامب والسيسي.. إليك نقاط التشابه بينهما	الاستقبال الحميم الذي تلقاه الرئيس المصري عبد الفتاح السيسي في البيت الأبيض من قبل نظيره الأميركي دونالد ترامب جذب انتباه وسائل الإعلام الغربية والتي لفتت إلى أنه جاء بعد أسبوعين فقط مما بدا أنه رفض من قبِل ترامب لمصافحة المستشارة الألمانية المرموقة أنغيلا ميركل.
http://www.middle-east-online.com/?id=256805	السيسي يوسع جهوده الدبلماسية لإحياء السلام	الامم المتحدة (الولايات المتحدة) - حض الرئيس المصري عبدالفتاح السيسي الفلسطينيين في خطاب له امام الجمعية العامة للأمم المتحدة على "الاتحاد"، وان يكونوا مستعدين "لقبول التعايش" بسلام مع الإسرائيليين".
http://www.middle-east-online.com/?id=256121	إما حرب اقتصادية عالمية على بيونغيانغ أو يتفرق مجلس الأمن	الامم المتحدة (الولايات المتحدة) - دعت واشنطن مجلس الأمن الدولي إلى البت الاثنين بشأن عقوبات جديدة مشددة ضد كوريا الشمالية المتهمة بتهديد السلام من خلال برامجها للأسلحة النووية والتقليدية.
http://www.middle-east-online.com/?id=256064	واشنطن تطلب أقسى العقوبات على بيونغيانغ	الامم المتحدة (الولايات المتحدة) - طلبت واشنطن رسميا التصويت الاثنين في مجلس الامن على مشروع قرار بفرض عقوبات جديدة ومشددة ضد كوريا الشمالية على الرغم من معارضة الصين وروسيا، وسط دعوات الاعلام الرسمي الكوري الشمالي لتطوير قدرات البلاد النووية.
http://www.huffpostarabi.com/2017/08/05/story_n_17685688.html	واشنطن بوست: لهذه الأسباب ترى الولايات المتحدة الإمارات حليفاً مزعجاً	الإمارات حليف مهم للولايات المتحدة، لكنه في نفس الوقت سبب لها صداعاً وإزعاجاً، بسبب عدد من الممارسات الضارة لمصالح واشنطن خاصة فيما يتعلق بالأوضاع في اليمن هكذا ترى صحيفة "واشنطن بوست" الأميركية في تقرير لها الخميس 3 أغسطس/آب 2017.
https://arabic.cnn.com/health/2017/05/05/ime-050517-eman-abdel-atti	"أسمن" امرأة في العالم تصل إلى أبوظبي" لاستكمال علاجها	الإمارات العربية المتحدة السمنة أمراض أمراض وأدوية صحة وحياة قد يعجبك أيضا عصائر "مضغوطة" يصل سعرها إلى 10 دولارات.
http://www.huffpostarabi.com/gamal-nassar/post_15480_b_17567306.html	مآلات الأزمة الخليجية على المنطقة العربية	الأزمة الخليجية أتت بظلالها على اضطراب الأوضاع والاستقرار في المنطقة، فالبرغم من المساعي الإقليمية المتمثلة في الكويت وتركيا، والجهود الدولية المتمثلة في أميركا وبريطانيا وألمانيا وفرنسا، وغيرها من الدول، فإن الأزمة لا تزال تراوح مكانها، ومرشحة للاستمرار لفترة طويلة.
http://www.huffpostarabi.com/2015/11/15/story_n_8558504.html	علامة تدل على أنك الطفل الأصغر في 16 عائلتك	الأشقاء الأصغر سنا زمرةٌ خاصة جداً، فالطفل الأصغر لم يعرف عالماً كان فيه الطفل الوحيد.

TABLE IV. SAMPLE OF TWEETS

Tweet Id	Tweet Text	Search Keywords
94155135639282892	دبلوماسي أميركي: السعودية والإمارات ارتكبتا خطأ بافتعال الأزمة الحالية مع #قطر https://t.co/WQPr6mxRoP	السعودية
94155135639282892	دبلوماسي أميركي: السعودية والإمارات ارتكبتا خطأ بافتعال الأزمة الحالية مع #قطر https://t.co/WQPr6mxRoP	قطر
94064821343822233	#الإمارات تُقيل مسؤولاً رياضياً بارزاً لمصافحته قطرياً! https://t.co/SpwrXbnlZ1	الامارات
94124963979320115	"القطرية_للتأمين" تدير ظهرها لـ #دبي وتعلن الخروج من #الإمارات" https://t.co/nvE5FQHfq4	الامارات
94213725979252327	وقائع رياضية تكشف صداقة #الإمارات لـ"إسرائيل" والعداء لـ #قطر https://t.co/PTDpMrCrE2	الامارات
94235116669167206	وقائع رياضية تكشف صداقة #الإمارات لـ"إسرائيل" والعداء لقطر https://t.co/PTDpMrU2vA	الامارات
94223791969935360	وقائع رياضية تكشف صداقة #الإمارات لـ"#إسرائيل" والعداء لقطر https://t.co/PTDpMrU2vA	الامارات
94197843956063027	وقائع رياضية تكشف صداقة #الإمارات لـ"إسرائيل" والعداء لـ #قطر https://t.co/PTDpMrCrE2	الامارات
94205373357121945	إقالة السركال كشفت "المستور" #يوسف_السركال #الإمارات #حصار_قطر		الامارات
94076320381390028	#الإمارات تُقيل مسؤولاً رياضياً بارزاً لمصافحته قطرياً! https://t.co/SpwrXbnlZ1	الامارات

TABLE V. SAMPLE OF MOST FREQUENTLY KEYWORDS FOUND IN THE ARTICLES

رئيس	خلال	سعوديه	عام	دولة	دول
اول	والإمارات	يوم	منطقة	قطر	يوم
حيث	قبل	مجلس	دبي	ذلك	اخرى
وقت	اكثر	والبحرين	خليج	ماضي	انها
امارات	وقال	محمد	عربية	متحده	عالم

TABLE VI. SAMPLE OF SYNONYMS FOR MOST FREQUENTLY WORDS

Keyword	Synonyms			
اخر	عرقل	مختلف	ابطأ	نهاية
مثل	شبيه	حكمة	عذب	مساو
عده	ادوات	متعدد	عدد	فترة من الزمن
امام	قبل	زعيم	رئيس	عالم
حول	سنة	عكس	ابدل	احاط
امن	سلام	اقر	هدأ	
ماضي	قديم	آنف	هالك	
عبد	رقيق	اصلح	مهد	
عمل	فعل	صنع	ممارسة	
تعاون	تضامن	تآزر	مشاركة	
مركز	رتبة	وسط	مسكن	
غير	آخر	سوى	مختلف	
ازمه	شدة	ضيق	محنة	
شهر	فضح	نشر	مجموع الايام	
كبير	ضخم	عظيم	كهل	
جميع	عامة	سواء	كل	
دون	اقل	رذيل	كتب	
اعلى	ارفع	اكرم	قمة	
بعض	قليل	جزء	قسم	
قرار	ادنى	حكم	قاع	
اولى	اجدر	اناب	فوض	
بيان	تصريح	منشور	فصيح	
جديده	حديثة	طازج	عصرية	
مجال	نطاق	موضع	شأن	
شيخ	استاذ	كهل	زعيم	
سابق	بارز	متقدم	زاحم	
قال	تحدث	تكلم	روى	
قبل	أمام	اقتنع	رضي	
امر	فرض	طلب	رأس	
وأضاف	اتبع	الف	دمج	
اضافه	ايواء	اكمال	جمع	
يمكن	ربما	لعل	توقع	
وكاله	انابة	تفويض	تخويل	
اكثر	معظم	أسرف	او ضح	
دعم	ساند	عون	اغاث	
اكبر	اضخم	اعظم	اعرض	
وقالت	تحدثت	روت	اخبرت	

2) Extract all synonyms of the most frequent unigrams.
3) Estimate the number of similarities between the bigram and trigram of two texts.

In this work, only unigrams and bigrams are used. The n-gram comparison can be increased if the degree of similarity between two sentences in terms of their meaning and structure is known. Algorithm 1 describes the steps taken to estimate semantic similarity.

TABLE VII. ABBREVIATIONS

Value	Abbreviation
Semantic Similarity for S1 UniGram	SS_S1U
Semantic Similarity for S2 UniGram	SS_S2U
Number if similar unigrams between S1 and S1	UniS1S2
Number of words in S1	S1L
Number of words in S2	S2L
Semantic Similarity for S1 BiGram	SS_S1B
Semantic Similarity for S2 BiGram	SS_S2B
Number if similar bigrams between S1 and S1	BiS1S2

$$SSV = (0.75 * B) + (0.25 * A) \qquad (1)$$

UniSemilarity (A)

$$SS\_S1U = UniS1S2/S1L \qquad (2)$$

$$SS\_S2U = UniS1S2/S2L \qquad (3)$$

Algorithm 1 Semantic Related Words Extraction

Require: S1, S2 two Arabic short complete text (as Tweets), with lengths n and m, respectively
Ensure: Semantic similarity value (SSV)
1: Take only nouns and verbs as features from S1, S2
2: Apply the following preprocessing on S1 and S2
3:     Remove non-Arabic characters
4:     Remove stop words
5:     Remove low-frequency tokens
6:     Determine the stem of the remaining text
7:     Take bigrams and trigrams of the two texts S1 and S2
8: Estimate the number of similarities between the bigrams and trigrams of the two texts (if the bigrams have the same token, give a higher value than if the two words are synonyms)
9: Estimate the SSV using Equations (1)–(5)(Depend on Table VII.

BiSimilarity (B)

$$SS\_S1B = BiS1S2/(S1L/2) \qquad (4)$$

$$SS\_S2B = BiS1S2/(S2L/2) \qquad (5)$$

Comparing two texts using bigrams provides an improved indication of their similarity because considering the meanings of two words gives a more accurate similarity value than considering the meaning of a single word.

Example

Assume the two texts given below are text1 and text2:

text1= وزير الخارجية البحريني: إقامة علاقات طيبة مع إيران مرهون بعدم تدخلها في الشؤون الداخلية للدول ووقف دعمها للإرهاب

text2= وزير خارجية البحرين: النظام الإيراني يدعم عددا من التنظيمات الإرهابية منها حزب الله اللبناني والميليشيات الانقلابية في اليمن

Starting from step 1 in Algorithm 1. Extracting only nouns and verbs causes text1 to become nVtext1 and text2 to become nVtext2, as follows:

nVtext1= وزير الخارجية إقامة علاقات طيبة مع إيران مرهون بعدم تدخلها في الشؤون الداخلية للدول ووقف دعمها للإرهاب

nVtext2 = وزير خارجية البحرين النظام يدعم عددا من التنظيمات منها حزب الله في اليمن

After completing all pre-processing steps(2-4) in Algorithm 1, nVtext1 becomes Text1Processed, and nVtext2 becomes Text2Processed, as follows: .

Text1Processed= وزر خرج قوم علق طوب رين رهن عدم دخل شون دخل دول وقف دعم رهب

Text2Processed= وزر خرج بحر نظم دعم عدد نظم منها حزب له يمن

Table VIII presents the results obtained from extracting unigrams, bigrams, and trigrams from Text1Processed and Text2Processed. By referring to ALMaany synonyms. The result will be as the following:

1) Number Of similarities using stemmed_Unigrams by computing how many similar words between the two

lists UniGramListText1Stemmed and UniGramList-Text1Stemmed and their synonyms equals 6.

2) Number Of similarities using stemmed_Bigrams by computing how many similar phrases between the two lists biGramListText1Stemmed and biGramList-Text2Stemmed and their synonyms equals 3.

Depending on Equation 1, SSV for text1 and text2 was 0.42 SSV=0.75 * (3/(14/2)) + 0.25*(6/14)= 0.42

Applying the cosine similarity measure to the same texts (text1 and text2) and considering term frequency as the features of words generated a similarity value between the two texts of 0.3.

In other research, they used Arabic WordNet to find the extent to which two concepts are related [3]. However, the similarity values they calculated did not depend on a specific domain, so the obtained values may have been substantially different from the actual values. For example, when extracting synonyms of an Arabic word (قوات), the related words and synonyms from Arabic WordNet, such as قوات المارينز الأمريكية, أسطول, and وحدة (see Table IX) were presented. Such comparisons between ambiguous words yield misleading values.

TABLE VIII. NGram of Text1Processed and Text2Processed

Text1Processed UniGram	Text2Processed UniGram	Text1Processed BiGram	Text2Processed BiGram	Text1Processed TriGram	Text2Processed TriGram
قوم	منها	وزر خرج	وزر خرج	علق طوب رين	وزر خرج بحر
رهن	عدد	رين رهن	نظم دعم	رين رهن عدم	بحر نظم دعم
شون	له	رهن عدم	دعم عدد	وزر خرج قوم	نظم دعم عدد
رهب	يمن	دعم رهب	عدد نظم	دخل شون دخل	عدد نظم دعم
دعم	حزب	شون دخل	خرج بحر	دول وقف دعم	نظم منها حزب
علق	بحر	طوب رين	له يمن	دخل دول وقف	خرج بحر نظم
طوب	نظم	علق طوب	نظم منها	وقف دعم رهب	دعم عدد نظم
دخل	دعم	قوم علق	حزب له	قوم علق طوب	منها حزب له
رين	خرج	دول وقف	منها حزب	خرج قوم علق	حزب له يمن
عدم	وزر	عدم دخل	بحر نظم	شون دخل دول	
دول		خرج قوم		طوب رين رهن	
خرج		دخل شون		رهن عدم دخل	
وزر		دخل دول		عدم دخل شون	
وقف		وقف دعم			

TABLE IX. Synonyms of Arabic Word from Arabic WordNet

Keyword	Synonyms
حَشْد	جيش
	حَشْد القوَات
	تجمع
فَرْد	حافِظ للسلام
	فَرْد من قوَات حفْظ السِّلام
	جيش
	جندي
نَشَرَ	نشر
	نشَر القوَات
	لعِب
	وضع
شُرْطَة	شرطة
	الأمن
	بوليس
	قوَّة الشرطة
	رجال الشرطة
	شرطي
	مجموع
قوَّات	قوَّات المارينز الأمريكية
	أسطُول
	وحدة
جَيْش	ضابِط جيش
	جيش نظَامي
	قوَات مسلحة
	عقيد
	لِواء
	مُشير
	مُلازم
	نقيب
	ضابِط
	جيش

IV. Experimental Result

Java was utilized in implementing the proposed algorithm, and the implementation was run to collect full articles from websites (see Table II) using Google Search API [27]. We used the same keywords in Table I to search Twitter accounts. We utilized Twitter4J, an unauthorized Java tool for the Twitter API, to extract tweets. The proposed algorithm's ability to detect the semantic similarity between 700 tweets was tested.

The semantic similarity measure used most often in previous work is cosine similarity. We applied cosine similarity to the same data set. Table X illustrates sample of the comparison between the SSV values and cosine similarity values. The results were analyzed manually by a domain expert who concluded that the values provided by the proposed algorithm were better than the cosine similarity values within the selected domain regarding the semantic similarity between the datasets' short texts. Trigrams and more n-grams can be considered to search for more equivalent documents. NGram can be increases as long as the length of the text increases. For examples For very short text, unigrams can be used. And for short text, bigrams can be used. So, as long as text length increases, n can be increased fot example triGram and FourGram.

Now, let us take a closer look at the values of the comparison results from examples in Table X. The semantic similarity values were enhanced based on the following:

1) We took into consideration synonyms of NGram words from the updated dictionary and concentrated on the synonyms from the same domain.
2) BiGrams similarity value increases the indication of similarity between the sentences. In our approach, BiGrams similarity value was given more weight over UniGram similarity.

As an initial step, the proposed algorithm was tested on short text. This algorithm can be applied to long text like documents and articles. Also, it can be applied to paragraphs, sentences,...etc. This algorithm has some limitations. One of these limitations it is based on an external dictionary. This problem can be solved by automatically extracting semantically related words depending on the same corpus.

V. Conclusion and Future Work

Freely available semantic similarity measurements are essential for advancing many NLP research areas, especially for under-resourced languages such as Arabic. The lack of a commonly used, trustworthy, comprehensive dictionary and ontology of semantic similar words and phrases is recognized as one of the most challenging and exciting problems facing Arabic NLP applications. However, manually computing the similarity degree of two texts is costly and nearly impossible. In this study, we have sought to tackle the phenomenon of computing the degree of semantic similarity of Arabic short texts.

This work introduced a novel similarity measure based on n-gram synonyms connected to the same domain designed to quantify the semantic similarity between concepts and words. The proposed algorithm was evaluated on a dataset of 700 tweets, and the semantic similarity values and cosine

TABLE X. COMPARISON RESULTS OF PROPOSED APPROACH AND COSINE SIMILARITY

Text1	Text2	SSV	Time (MS)	Cosine	Time (MS)
بدء التصويت في حوار rtarabic# انفتاح الإمارات على حزب الإصلاح اليمني، هل سيغير موازين القوى في الميدان؟	السعودية والإمارات تسعيان للتحالف مع حزب الإصلاح وهما المتهمتان بإسقاط #اليمن بيد# الحوثيين للتخلص منه عام 2014... https://t.co/6lzxxZjotJ ف	1	14	0.2	31
الإصلاح و #الإمارات.. تقارب يرسم خارطة التحالفات الجديدة في #اليمن https://t.co/AWKp9Zmyic	بن سلمان وبن زايد يلتقيان رئيس حزب الإصلاح اليمني https://t.co/9p7LwqFGKA https://t.co/gvXcUUe1ZU #السعودية #الإمارات #اليمن|	1	16	0.3	39
بدء التصويت في حوار rtarabic# انفتاح الإمارات على حزب الإصلاح اليمني، هل سيغير موازين القوى في الميدان؟	بن سلمان وبن زايد يلتقيان رئيس حزب الإصلاح اليمني https://t.co/9p7LwqFGKA https://t.co/gvXcUUe1ZU #السعودية #الإمارات #اليمن|	1	12	0.3	33
بن سلمان وبن زايد يلتقيان رئيس حزب الإصلاح اليمني https://t.co/9p7LwqFGKA https://t.co/gvXcUUe1ZU #السعودية #الإمارات #اليمن|	بدء التصويت في حوار rtarabic# انفتاح الإمارات على حزب الإصلاح اليمني، هل سيغير موازين القوى في الميدان؟	1	13	0.3	43
بن سلمان وبن زايد يلتقيان رئيس حزب الإصلاح اليمني https://t.co/9p7LwqFGKA https://t.co/gvXcUUe1ZU #السعودية #الإمارات #اليمن|	السعودية والإمارات تسعيان للتحالف مع حزب الإصلاح وهما المتهمتان بإسقاط #اليمن بيد# الحوثيين للتخلص منه عام 2014... https://t.co/6lzxxZjotJ ف	1	19	0.4	52
السعودية والإمارات تسعيان للتحالف مع حزب الإصلاح وهما المتهمتان بإسقاط# اليمن بيد الحوثيين للتخلص منه عام 2014.. ف# https://t.co/6lzxxZjotJ	بن سلمان وبن زايد يلتقيان رئيس حزب الإصلاح اليمني https://t.co/9p7LwqFGKA https://t.co/gvXcUUe1ZU #السعودية #الإمارات #اليمن|	1	20	0.4	49
الإمارات تُقيل مسؤولاً رياضياً بارزاً لمصافحته قطرياً!# https://t.co/SpwrXbnlZ1 رياضة# قطر# https://t.co/kU0i6zDii7	الإمارات تُقيل مسؤولاً رياضياً بارزاً لمصافحته قطرياً!# https://t.co/SpwrXbnlZ1 نبض الخليج العربي #قطر #رياضة# https://t.co/SJ2iC2fvMv	1	11	0.8	34
السعودية والإمارات تسعيان للتحالف مع حزب الإصلاح وهما المتهمتان# بإسقاط #اليمن بيد الحوثيين للتخلص منه عام 2014.. ف https://t.co/6lzxxZjotJ	بدء التصويت في حوار rtarabic# انفتاح الإمارات على حزب الإصلاح اليمني، هل سيغير موازين القوى في الميدان؟	0.88	19	0.2	32

similarity were compared. A domain expert carefully examined the results and concluded that, considering the semantic similarity between the short sentences of the datasets, the values produced by the proposed algorithm were more accurate than the cosine similarity values within the chosen domain. This work can be a basis for other works investigating the same semantic similarity problems. Semantic similarity between texts written in the Arabic language can help determine, for example, who originally published a piece of news, who rephrased a previously published news article and claimed to be the original source, and how to extend it to solve problems related to plagiarism. Further research is a domain ontology that includes all relations between words in the domain. This domain will help determine how some words are related and how they are different. This knowledge can then be used for other purposes, such as to perform sentiment analyses of the Arabic language in this domain.

REFERENCES

[1] A.-S. Mohammad, Z. Jaradat, A.-A. Mahmoud, and Y. Jararweh, "Paraphrase identification and semantic text similarity analysis in arabic news tweets using lexical, syntactic, and semantic features," *Information Processing & Management*, vol. 53, no. 3, pp. 640–652, 2017.

[2] A. Faaza, D. James, A. Zuhair, A. Keeley *et al.*, "Arabic word semantic similarity," *International Journal of Cognitive and Language Sciences*, vol. 6, no. 10, pp. 2497–2505, 2012.

[3] F. A. Almarsoomi, J. D. OShea, Z. Bandar, and K. Crockett, "Awss: An algorithm for measuring arabic word semantic similarity," in *2013 IEEE international conference on systems, man, and cybernetics*. IEEE, 2013, pp. 504–509.

[4] R. Mihalcea, C. Corley, C. Strapparava *et al.*, "Corpus-based and knowledge-based measures of text semantic similarity," in *Aaai*, vol. 6, no. 2006, 2006, pp. 775–780.

[5] T. Slimani, "Description and evaluation of semantic similarity measures approaches," *arXiv preprint arXiv:1310.8059*, 2013.

[6] G. A. Miller, "Wordnet: a lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[7] Y. Li, D. McLean, Z. A. Bandar, J. D. O'shea, and K. Crockett, "Sentence similarity based on semantic nets and corpus statistics," *IEEE transactions on knowledge and data engineering*, vol. 18, no. 8, pp. 1138–1150, 2006.

[8] almaany, "Translation and meaning in almaany," https://www.almaany.com/en/dict/ar-en/, July 2022.

[9] A. Rozeva and S. Zerkova, "Assessing semantic similarity of texts– methods and algorithms," in *AIP Conference Proceedings*, vol. 1910, no. 1. AIP Publishing LLC, 2017, p. 060012.

[10] D. Chandrasekaran and V. Mago, "Evolution of semantic similarity—a survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–37, 2021.

[11] M. Alian and A. Awajan, "Arabic semantic similarity approaches-review," in *2018 International Arab Conference on Information Technology (ACIT)*. IEEE, 2018, pp. 1–6.

[12] J. Yang, Y. Li, C. Gao, and Y. Zhang, "Measuring the short text similarity based on semantic and syntactic information," *Future Generation Computer Systems*, vol. 114, pp. 169–180, 2021.

[13] S. Zad, M. Heidari, P. Hajibabaee, and M. Malekzadeh, "A survey of deep learning methods on semantic similarity and sentence modeling," in *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2021, pp. 0466–0472.

[14] A. Zouaghi, M. Zrigui, G. Antoniadis, and L. Merhbene, "Contribution to semantic analysis of arabic language," *Advances in Artificial Intelligence*, vol. 2012, 2012.

[15] W. Wali, B. Gargouri *et al.*, "Supervised learning to measure the semantic similarity between arabic sentences," in *Computational collective intelligence*. Springer, 2015, pp. 158–167.

[16] S. S. Aljameel, J. D. O'Shea, K. A. Crockett, and A. Latham, "Survey of string similarity approaches and the challenging faced by the arabic language," in *2016 11th International Conference on Computer Engineering & Systems (ICCES)*. IEEE, 2016, pp. 241–247.

[17] H. M. Alghamdi, A. Selamat, and N. S. A. Karim, "Arabic web pages clustering and annotation using semantic class features," *Journal of King Saud University-Computer and Information Sciences*, vol. 26, no. 4, pp. 388–397, 2014.

[18] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," *Journal of the American society for information science and technology*, vol. 60, no. 11, pp. 2169–2188, 2009.

[19] D. Jatnika, M. A. Bijaksana, and A. A. Suryani, "Word2vec model

[20] analysis for semantic similarities in english words," *Procedia Computer Science*, vol. 157, pp. 160–167, 2019.

[20] N. Peinelt, D. Nguyen, and M. Liakata, "tbert: Topic models and bert joining forces for semantic similarity detection," in *Proceedings of the 58th annual meeting of the association for computational linguistics*, 2020, pp. 7047–7055.

[21] Y. Cai, Q. Zhang, W. Lu, and X. Che, "A hybrid approach for measuring semantic similarity based on ic-weighted path distance in wordnet," *Journal of intelligent information systems*, vol. 51, no. 1, pp. 23–47, 2018.

[22] L. Gutiérrez and B. Keith, "A systematic literature review on word embeddings," in *International Conference on Software Process Improvement*. Springer, 2018, pp. 132–141.

[23] A. Saif, N. Omar, U. Z. Zainodin, and M. J. Ab Aziz, "Building sense tagged corpus using wikipedia for supervised word sense disambiguation," *Procedia Computer Science*, vol. 123, pp. 403–412, 2018.

[24] H. A. Abdeljaber, "Automatic arabic short answers scoring using longest common subsequence and arabic wordnet," *IEEE Access*, vol. 9, pp. 76 433–76 445, 2021.

[25] N. Altuwairesh, "Successful translation students' use of dictionaries," *International Journal of English Linguistics*, vol. 12, no. 2, 2022.

[26] N. Sabbah and R. Alsalem, "Female translation students' knowledge and use of online dictionaries and terminology data banks: A case study," *AWEJ for Translation & Literary Studies*, vol. 2, no. 2, 2018.

[27] Google, "Google," https://developers.google.com/custom-search/, July 2022.

An Efficient Meta-Heuristic-Feature Fusion Model using Deep Neuro-Fuzzy Classifier

Sri Laxmi Kuna[1]

Research Scholar, Department of CSE
Koneru Lakshmaiah Educational Foundation
Vaddeswaram, Guntur, A.P, India

Dr. A.V. Krishna Prasad[2]

Supervisor, Department of CSE
Koneru Lakshmaiah Educational Foundation
Vaddeswaram, Guntur, A.P., India

Abstract—**Diabetic Retinopathy (DR) is the major cause of the loss of vision among adults worldwide. DR patients generally do not have any symptoms till they reach the final stage. The categorization of retinal images is a remarkable application in detecting DR. Due to the level of sugar available in the blood, the categorization of DR severity becomes complicated to determine the grading level of the damages caused in the retina. To rectify these challenges, a new DR severity classification model is proposed for detecting and treating the DR. The main objective of the proposed model is to classify the severity grades that occurred in the retinal region of the human eye. Initially, gathered retinal images are enhanced and the blood vessel segmentations are done by utilizing the optic disc removal and active contouring model. The abnormalities such as "microaneurysms, hemorrhages, and exudates" are segmented by utilizing Fuzzy C-Means Clustering (FCM) and adaptive thresholding. Then, the segmented images are given to "VGG16 and ResNet", in which the two different feature sets are acquired. Then, these features are added to obtain the second set of features as F2. Again, the enhanced images act as an input to the "VGG16 and ResNet", which are attained as the first feature set as F1. In the feature concatenation phase, the resultant of two features is used for feature fusion with the aid of weights parameter that is optimized by Modified Mating Probability-based Water Strider Algorithm (MMP-WSA), where the feature fusion is carried out using the mathematical expression. Finally, the multi-class severity classifications are done by using the Optimized Deep Neuro-Fuzzy Classifier (ODNFC), where the optimization of hyper-parameters is done by the proposed MMP-WSA. Thus, the experimental results of the proposed model have been acquired by the precise segment of the abnormalities and better classification results regarding the grade level.**

Keywords—Multi-class severity classification; diabetic retinopathy; modified mating probability-based water strider algorithm; optimized deep neuro-fuzzy classifier; fuzzy clustering model; adaptive thresholding; optic disc removal; image enhancement

I. INTRODUCTION

The retinal image evaluation is a diligent investigation platform in Diabetic Retinopathy (DR). It is the major cause of serious eye difficulties or even loss of sight in developed countries [9]. The retina mainly comprises of the optic disc, fovea, and blood vessels. A few basic symptoms of DR are "sudden changes in vision, eye pain, double vision, Eye Floaters, blurred vision/fluctuating vision and Shadow in Field of View (FOV) and Spots" [11]. The DR is considered one of the deep-rooted visual problems that tend to vision loss if it is

not identified on time [15]. Early detection by utilizing regular periodic evaluations is a determinative factor in declining the hazard of serious visual deterioration [16]. The major lesions present in the DR are referred to as exudates, cotton wools, hemorrhages, and Microaneurysms. Hence, the identification of "Microaneurysms and hemorrhages" plays a significant role in the automatic detection of retinopathy [19]. Furthermore, the determination of this disease needs a huge amount of experience and adequate knowledge from doctors. Also, because of the presence of inadequate ophthalmologists and the lack of resources, the screening of patients with Diabetes Mellitus (DM) is considered to be complicated [20], which leads to loss of vision [12]. The DR is broadly categorized into a couple of sections known as proliferative and non-proliferative. The non-proliferative refers to the weakening of blood vessels inside the retina, which causes "leakage of fluids and blood on the retinal surface" due to DR [13]. This leakage minimizes the retinal sensitivity, while it is wet and swollen. The proliferative refers to the developed levels of DR, which creates neovascularization, a creation of bloody vessels newly and naturally in the variety of microvascular network forms that originate on the retina internally [14].

In automated DR diagnosis and segmentation, soft computing models play a remarkable role. Models for DR diagnosis and classifications are genetic algorithms, neural networks, and evolutionary algorithms [21]. Early detection of diabetic retinopathy can prevent the loss of vision in future [10]. Other than automated DR categorization, the fuzzy classifiers are modeled to be translucent with steps of categorization and logic descriptions that are detectable and understandable [22]. However, it needs a huge amount of labeled data for monitoring the training networks. The categorization of retinopathy images is considered a difficult task and so many experienced clinicians are required for the manual interpretation of a high number of retinopathy images. Because of this, Fuzzy C-Means Clustering (FCM) is considered [23]. The fuzzy methodology supports providing enhanced image representations and also in enhancing the analysis performance and providing a more dependable system for screening [24]. The screening of the DR disease at the early stage itself is complicated and so the "Neuro-Fuzzy Hybridization" provides a synergistic intellectual system that joins the human-like reasoning model of a fuzzy system with a neural network's learning structure. The Neuro-Fuzzy hybrids are broadly termed "Neuro-Fuzzy Systems (NFS) or Fuzzy Neural Networks (FNNs)" [25]. Thus, the latest investigators

aim on establishing a new DR diagnosis model through deep learning approaches.

The major offerings of this paper are as follows:

- To develop a novel DR severity classification approach based on deep learning approaches for categorizing severities caused by DR, to enhance the early disease diagnosis rate, and to reduce the fatality of patients regarding vision loss.

- The image enhancement process utilizing RGB channels and processing segmentation in two stages known as "blood vessel segmentation and abnormality segmentation" by utilizing the "optic disc removal, adaptive thresholding, FCM and active contouring" model to improve the quality of segmented images.

- To establish the weighted optimal feature selection process in two models by choosing the remarkable features utilized with weight function by developing a Modified Mating Probability-based Water Strider Algorithm (MMP-WSA) for limiting the features and enhancing the performance of categorization.

- To develop an ensemble known as an Optimized Deep Neuro-Fuzzy Classifier (ODNFC), for classifying the DR severity with parameter optimization such as "hidden neurons of DNN, the learning rate of DNN, EXP-bound in fuzzy and weights" are optimized with the developed MMP-WSA for obtaining improved validity in DR categorization.

- To establish a new meta-heuristic algorithm known as MMP-WSA for mounting the weighted optimal feature selection by optimizing the weight function and enhancing the classification accuracy by optimizing the hyper-parameters of developed ODNFC.

The contribution of this research work is as follows. In Section II, the relevant works and their related difficulties are briefed. In Section III, the proposed DR diagnosis model by utilizing ODNFC and image enhancement processing is elaborated. In Section IV, the blood vessel and abnormality segmentation are discussed. In Section V, feature extraction and optimal weighted feature selection for DR diagnosis are explained. In Section VI, the developed MMP-WSA and the developed ODNFC are briefed. In Section VII, the experimental results are discussed. In Section VIII, the proposed DR diagnosis is concluded.

II. LITERATURE SURVEY

A. Related Works

In 2020, Vaishnavi *et al.* [1] established a new classification approach based on segmentation to categorize the 'DR images", effectively. The proposed approach comprised three major processing known segmentation, feature selection, preprocessing, and categorization. The proposed model has undergone pre-processing utilizing "Contrast-Limited Adaptive Histogram Equalization (CLAHE)". The "AlexNet" model was utilized for feature extraction to collect the most important features. Finally, the categorization of DR images was done by utilizing the SoftMax layer. The experimentation results

represented that the established approaches have achieved the maximum categorization rate.

In 2021, Bhardwaj *et al.* [2] proposed a system known as Hierarchical Severity Level Grading (HSG) for the identification and categorization of disorders in DR. The images of the retinal fundus available in the developed HCG system were classified into grades 0, 1, 2, 3 based on the count of anomalies, hemorrhages, and microaneurysms in the images of the fundus. The complications of the "landmark segmentation of retinal images, DR severity grading and DR retinal discrimination" have been notified in this paper contributing to the proposed model. For the categorization of DR and non-DR images, the developed systems have attained greater accuracy with SVM [17] and KNN [18] classifiers. The hierarchical inequities into more grades of harmfulness in disease have resulted in a certain accuracy range for the KNN classifier. The HCG system attained a remarkably minimum time for computation when compared with other state-of-art models.

In 2022, Vasireddi *et al.* [3] established an automated DR identification screening model that was needed to enhance the detection speed and accuracy of the diagnosis. Proper treatments have been given to patients to prevent blindness when the complication levels of DR were detected in earlier stages accurately. The utilization of the optimization algorithm along with the required parameters for tuning enhanced the model's performance. The experimentation results have indicated that the proposed model attains superior accuracy.

In 2020, Gharaibeh *et al.* [4] proposed soft computing models for the diagnosis of Hemorrhages and microaneurysms in fundus images. This process comprised of several steps "such as i) pre-processing, ii) blood vessel segmentation, iii) blood vessel removal, iv) fovea localization, v) fovea elimination, vi) feature extraction, vii) feature selection and finally viii) detection of Diabetic Retinopathy disease" known as Hemorrhages and microaneurysms. It enhanced the performance of precision, accuracy, specificity, and sensitivity and the diagnosis of the DR easier and faster and the results were efficient and provided effective outcomes in categorization models.

In 2022, Kuna Sri Laxmi et al [5] employed various CNN architectures used to classify the severity levels of Diabetic Retinopathy. In 2022, Das and Saha [6] proposed a technique on the basis of a genetic algorithm to detect the CNN parameter automatically and further classification of DR. This approach has comprised a series of "pooling and convolution layers" for the feature extraction. The hyper-parameters in the pooling layer, convolution layer, kernel count, and size of the kernel were analyzed by utilizing the genetic algorithm. The proposed model was validated on a normally available dataset known as "The Messidor dataset". Finally, the SVM utilized for DR categorization has attained greater accuracy than other available models on the basis of a genetic algorithm.

In 2020, Luo *et al.* [7] aimed on embedding a self-monitored model into an un-monitored deep learning approach. Remarkably, it has developed a "Self-supervised Fuzzy Clustering Network (SFCN)" along with a module for feature learning, fuzzy self-monitoring, and reconstruction. The

reconstruction approach and feature learning have assured the capacity of the network and the fuzzy self-monitoring module was in charge of giving more training paths for the entire network. To analyze the efficiency of the proposed model, it has established the network on three datasets, in which the results described the good performance of the un-monitored image categorization task.

In 2020, Wang *et al.* [8] proposed a hierarchical architecture to consolidate the common inter-relationship in-between the features and severity levels of DR. The proposed model was analyzed on two individual testing elements by utilizing the "receiver operating characteristic analysis, quadratic weighted Cohen's kappa coefficient and precision-recall analysis". The experimentation outcomes have demonstrated that the proposed model has enhanced the performance of DR severity in comparison with other deep-learning approaches and attained a performance close to the experienced ophthalmologists during the detection of DR severity level.

B. Problem Statement

Diabetic Retinopathy is one of the major diseases in humans, which affects the retinal region of the eye. In the medical industry, there is no early predicted symptom for retinopathy. This disease is caused in two ways: the initial stage is NPDR and the most affected phase is PDR. The remarkable symptoms are leaking blood vessels, swelling of the retina, exudates, damage to nerves, and so on. Hence, the detection and diagnosis of retinopathy become the most challenging task. The screening is targeted at an affected individual that is highly unpredictable. CLAHE and AlexNet [1] enhance the performance in terms of accuracy, sensitivity, and specificity. But, the last layer of AlexNet produces a multiclass classification problem. Gray Level Co-Occurrence Matrix and Support Vector Machine [2] yield the accurate classification of the severity grade level and increase the accuracy. However, it does not consider blood vessel bleeding and neovascularisation problem. Also, due to retinal detachment, acute blindness occurs. Deep Fuzzy Neural Network (DFNN) and Lion Optimization Algorithm (LOA) [3] provide a better severity classification. But it has more time computation complexity. PSO and fuzzy membership [4] obtains higher accuracy and extracts the blood vessels precisely. However, it mitigates the performance since it has fewer quality data. Fuzzy Neural Network [5] improves the classification rate. Moreover, the limitation is to need more numbers of rules to perform in a better way. The Genetic Algorithm and SVM [6] provide higher accuracy for different datasets. But it possesses structural and computational burdens. Fuzzy Clustering [7] yields a higher accuracy value and does not contain most qualitative criteria. However, it has an imprecision dependency on the model. Multi-task DNN [8] obtains less error and more accuracy. Moreover, due to various confidence levels, it may get inaccurate results. To overcome the mentioned challenges, it is provoked to establish a new approach for diagnosing diabetic retinopathy disease.

III. ARCHITECTURAL VIEW OF MULTI-CLASS SEVERITY CLASSIFICATION OF DIABETIC RETINOPATHY

A. Proposed DR Classification Model and Description

A new DR severity classification model is developed with a weighted feature fusion model and deep learning approach depicted in Fig. 1. In this model, the DR images data set is considered from the standard data set DIARETDB1 as input. These images are preprocessed by four steps enhancement process such as R, G, and B channel separation, entropy-based spatial filtering, fuzzy-based weight adjustment, and concatenation of R, G, and B channels. Then, the blood vessel segmentation is carried out by utilizing optic disc removal and active control using enhanced images. The abnormality segmentation is done with segmented images by FCM and adaptive thresholding. Then, the feature extraction phase is done by using a couple of models. In model 1, the obtained enhanced images are provided as the input to VGG 16 and ResNet, and the required set of features is extracted as feature set 1. In model 2, the abnormality-segmented images data set is provided as the input to "VGG 16 and ResNet" and the required set of features is extracted as feature set 2. For both feature set1 and 2, the feature fusion is carried out for obtaining "fused features with optimized weight function using the developed MMP-WSA" to obtain optimal weighted fused features. Finally, the DR severity classification is done by using developed ODNFC using the weighted fused features, in which the parameters such as "hidden neurons of DNN, the learning rate of DNN, EXP-bound in fuzzy and weights" are optimized by utilizing the developed MMP-WSA. At last, the DR severity-based outcome is obtained as four categories of DR known as normal, earlier, moderate and severe for the proposed DR severity classification model.
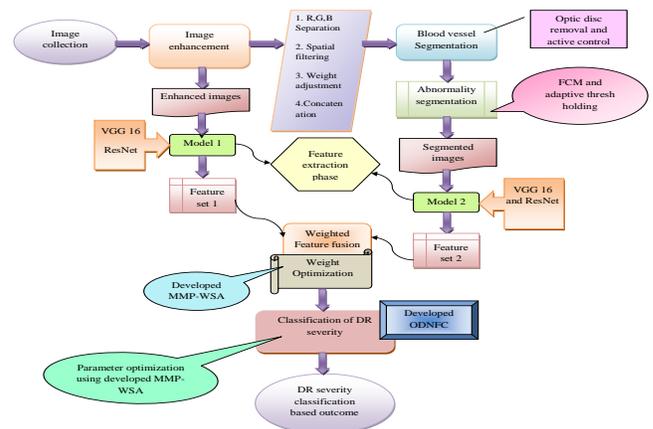
Fig. 1. Developed Multi-Class DR Severity Classification Model with Adaptive Weighted Feature Fusion and Deep Learning Approach.

B. Novel Image Enhancement Process

The images for input are taken from "https://www.it.lut.fi/project/imageret/diaretdb1/". It is a public database for the process of benchmarking the DR severity classification from digital images. the collected database images are denoted as IMG_m^{COL}, where $m = 1,2,\cdots,M$ and the total image count is represented as M. The IMG_m^{COL} is provided as input for the process of image enhancement in the

proposed DR severity classification. The image enhancement process is carried out as follows:

- Separation of RGB channels: The input images IMG_m^{COL} are pre-processed for enhancing the quality of the image by removing unwanted elements. The separation of the R, G, and B channels is utilized for the evaluation of elements of primary colors of each region of an image. Hence, the enhanced image obtained after the separation of R, G, and B channels is denoted as ENH_c^{sep}.

- Applying entropy-based spatial filtering: In this step, the RGB-separated images ENH_c^{sep} are given as input. The spatial filtering on the basis of entropy is created for global contrast enhancement of the images. Hence, a single entropy function is applied to the entire image. The algorithm gives contrast enhancement without remarkable deformation on the output image. Thus, the enhanced image obtained from the spatial filtering is represented as ENH_c^{spa}.

- Concatenation of RGB channels: Finally, the separated images are concatenated after removing all kinds of unwanted elements from the image channels. Thus, the quality of the images is enhanced for further processing. Hence, the enhanced image obtained is represented as ENH_c^{img}.

IV. ENHANCED DIABETIC RETINOPATHY SEGMENTATION MODEL THROUGH NEW ABNORMALITY SEGMENTATION

A. Optic Disc Removal

The optical disc is removed from enhanced images ENH_c^{img} of the data set for further processing. In the process of optic disc removal for the developed DR severity classification, the enhanced images ENH_c^{img} are provided as the input, and further processing is made. For removing it, the edge-enhanced image on the basis of the curvelet is opened by using a disk-shaped element, which is eliminated from the inverted equalization image. The matching filtering will intensify the visibility of blood vessels present in the enhanced image. The morphological operation is carried out for the processing of optic disc removal. Thus, the segmented image obtained after the process of optic disc removal is represented by SEG_c^{opt}

B. Active Contour-based Blood Vessels Segmentation

In the process of blood vessel segmentation, the optic disc removed image SEG_c^{opt} is provided as input to segment the blood vessel using active contour. The active contour is a technique of segmentation that utilizes energy forces and difficulties to split the pixels' interest from images for later evaluation and processing. Active contouring is the methodology of collecting the degradable structures of an image with problems and energy forces for the process of segmentation of blood vessels. The contour model determines

the borders of the image elements to create a contour. The curvature of the structure is identified by utilizing various methodologies that comprise internal and external forces. The energy functions are always interrelated to the curve of images. The contour blood vessel segmentation constraints for available images are identified on the basis of requirements. The needed structure is obtained by determining the energy function. A gathering of points that position a contour is utilized to define deformations in contour. Thus, the blood vessel segmented image obtained from active contouring is represented as SEG_c^{act}.

C. Abnormality Segmentation by Fuzzy Logic-based Adaptive Thresholding

The abnormality segmentation of the developed DR severity classification is carried out on the basis of two models known as FCM and Adaptive Thresholding on the basis of fuzzy logic and it is briefed below.

FCM [31]: Here, the blood vessel segmented images are provided as input and that is segmented into five classes, which are remarked as the blood vessel, the outer area of the eyeball, the healthy area in the background of the retina, cotton wools and hard exudates. In the processing of clustering, the input images are subdivided into 3×3 non-overlapping regions. Then, for each region, the average intensity is computed. There exists p total number of 3×3 regions present in the image.

Suppose we have a collection of N data (xi; i = 1, 2, 3, ... , N) to be segmented into five fuzzy sets. There will be a membership matrix *em* of size 5XN

Then, the value of entries of the matrix *em* is initiated randomly between 0 and 1 while keeping the sum of each column of *em* to 1 and computed on the basis of the following Eq. (6)

$$\sum_{a=1}^{h} em_{ab} = 1 \ \forall b = 1,2,\ldots,p \tag{6}$$

The FCM divides the data into five categories as h and then determines the center of cluster for every group iteratively during the minimization of cost function on the dissimilarity. Thus, the function of dissimilarity is determined on the basis of the following Eq. (7).

$$B = \sum_{a=1}^{h} \sum_{b=1}^{k} (em_{ab})^{we} ed_{ab}^2 \tag{7}$$

Here, the term em_{ab} denotes the entries of *em*, ed_{ab} indicates the Euclidean distance between a^{th} cluster h_a, b^{th} is the data point, and we **represents the weighting exponent.** Then, the cluster center and the entries are upgraded by utilizing the following Eq. (8) and Eq. (9)

$$Ci = \frac{\sum_{b=1}^{em} em_{ab} Y_{ab}}{\sum_{b=1}^{p} (em_{ab})^{we}}, \forall \ b = 1,2,\ldots,C \tag{8}$$

$$em_{ab} = \frac{1}{\sum_{c=1}^{C} \left(\frac{ed_{ab}}{ed_{cb}}\right)^{\frac{2}{we-1}}} \tag{9}$$

The output from the FCM is obtained in the form of non-overlapping regions indicating five clusters. The abnormality segmented image obtained from FCM is denoted as SEG_c^{fcm}.

Adaptive Thresholding [32]: The segmented image SEG_F^{fcm} obtained from FCM is provided as input and further processing is made. This method generates local thresholds for various areas of images. This is also called the dynamic threshold. By utilizing thresholding, the value of pixels of the images can be categorized from the background. The "adaptive thresholding algorithm" is processed by determining the weighted averages available locally in the image by utilizing the recursive filters. Thus, by using adaptive thresholding, the objects can be separated from the background and also the boundaries are separated. This is attained by generating the surface of the threshold so that the threshold value will be available for each pixel. Hence, the adaptive thresholding is computed based on Eq. (10).

$$Th_{gaussian} = \frac{1}{\sqrt{2\pi \sigma_{a,b}}} e^{\frac{pix(a,b)\mu_{a,b}}{2\sigma_{a,b}^2}} \qquad (10)$$

Here, the term a denotes the column pixel index, and b indicates the row pixel index σ denotes the standard deviation and μ represents the averages. The abnormality segmented image obtained from adaptive thresholding is indicated as SEG_c^{abn}. The process of abnormality segmentation based on FCM and adaptive thresholding is indicated in Fig. 2
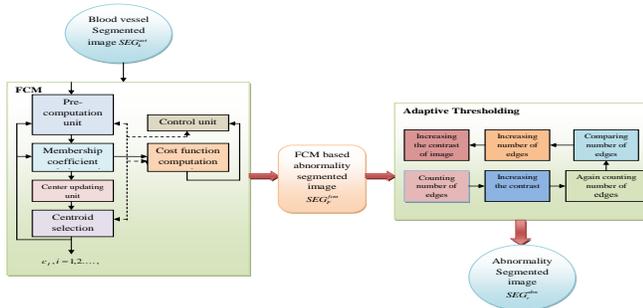
Fig. 2. Abnormality Segmentation based on FCM and Adaptive Thresholding.

V. DEEP FEATURE EXTRACTION AND OPTIMAL FEATURE SELECTION PROCESS IN TWO STAGE

A. Deep Feature Extraction

The deep feature extraction process for severity classification is determined in "model 1 and model 2". In model 1, the enhanced image obtained from pre-processing is provided as input to VGG16 and ResNet models for extracting the first set of features. In model 2, the abnormality-segmented images are fed as the input for VGG 16 and ResNet models, and the second set of features is extracted as feature set 2. Then both sets of features are optimized by utilizing the proposed MMP-WSA to obtain the optimal deep features, which are elaborated below.

VGG 16 [29]: In this, the enhanced images ENH_c^{img} and segmented images SEG_z^{abn} are provided as input. It is composed of three divisions named "convolutional layer, totally connected layer, and pooling layer". It also consists of a total of 10 layers that are different from the pooling layers. It has a developed network design. For this, the input size of an image is fixed as "224×224 pixels" and the size of the filter in the obtained image is fixed at "3×3 pixels". The output division in the VGG 16 is named "SoftMax". The gathered deep feature element from the VGG-16 is represented as $FESEG_d^{deepVGG16}$ and $FEEN_f^{deepVGG16}$ from ENH_c^{img} and SEG_z^{abn}, respectively.

ResNet-150 [29]: For this, the selected features of enhanced images ENH_c^{img} and segmented images SEG_z^{abn} are provided as input for the ResNet model. It is a developed form of the CNN model. The ResNet gives the shortest route between the divisions to resolve the problems. It also eliminates the level of dispersion, which occurs at the time of complex generation of the network. Moreover, the "bottleneck blocks" are used to fasten the learning process in the ResNet mode. The extracted features from the ResNet model are denoted as $FESEG_e^{deepRN}$ and $FEEN_g^{deepRN}$ from ENH_c^{img} and SEG_z^{abn}, respectively.

B. Optimal Feature Selection: Model 1

In model 1, the enhanced images ENH_c^{img} are provided as input for the feature extraction with VGG 16 and ResNet. Then, by utilizing developed MMP-WSA, the extracted features such as $FEEN_g^{deepRN}$ and $FEEN_f^{deepVGG16}$ are optimized for obtaining optimal deep features denoted as F_1. Total counts of 10 features are extracted from model 1.

C. Optimal Feature Selection: Model 2

In model 2, the abnormality-segmented form of images is fed as input and the feature element extraction is done by utilizing two models known as VGG 16 and ResNet. Then, by utilizing the developed MMP-WSA, the selected segmented features such as $FESEG_e^{deepRN}$ and $FESEG_d^{deepVGG16}$ are optimized for obtaining optimal deep features denoted as F_2. Total counts of 10 features are extracted in model 2.

D. Proposed Feature Fusion

The obtained optimal features F_1 and F_2 from model 1 and model 2 are then fused for obtaining optimal fused features. The weight function is considered during the process of feature fusion, which is optimized using the developed MMP-WSA, and the fusion process takes place as shown in the following Eq. (11).

$$F_{fu} = \left[F_1 * W + (1 - W) * F_2 \right] \qquad (11)$$

Here, the terms F_1 and F_2 indicate the optimal features sets from model 1 and model 2, represent the optimized weight function and denote the feature fusion.

VI. DIABETIC RETINOPATHY SEVERITY CLASSIFICATION BY OPTIMIZED DEEP NEURO FUSSY CLASSIFIER

A. Proposed MMP-WSA

The developed MMP-WSA algorithm is utilized and implemented in the DR severity classification approach for optimization of parameters such as "hidden neurons of DNN, the learning rate of DNN, EXP-bound in fuzzy and weights" in the feature fusion phase for enhancing the DR classification performance. WSA [26] is selected in this model since it attains efficient performance in optimization issues. In the stage of mating, the probability of attraction is transferred to adaptive or dynamic instead of static, which gives effective convergence performance, but there exist certain limitations in WSA such as the evaluation count is not an acceptance function of the internal elements and so there exists potential complexity in programming languages, which cannot help the global variable to control the total function and evaluation count. Due to the presence of such difficulties in conventional WSA, it is essential to propose an enhanced WSA named MMP-WSA for improving the DR severity classification. In this MMP-WSA, the probability of mating is updated with an adaptive concept, whereas in the traditional algorithm, it is updated with a random parameter.

WSA is an algorithm designed on the basis of feeding mechanism, mating style, territorial behavior, and succession of water strider bugs. The step involved in the algorithm is briefed below.

Initial Birth: The creation of candidate solutions randomly in the searching platform as indicated in Eq. (12) below.

$$Ws_j^0 = l_b + rand.(u_b - l_b): \quad j = 1, 2, \ldots, NWs \tag{12}$$

Here, the initial positions *Ws* are analyzed by utilizing an objective function to compute the fitness. The term u_b and l_b indicates the lower and upper bound of variables, Ws_j^0 denotes the initial positions of the jth *Ws* in the searching space, *NWs* indicates the population size and *rand* represents the random number that lies between $[0,1]$.

Territory Establishment: For introducing *ni* count of territories, the ith member of every group is allocated to the ith territory $(i = 1, 2, \ldots, ni)$. Hence, the number of Ws_r nesting in every territory is equal to $\frac{NWs}{ni}$ and $\frac{NWs}{ni}$ count of groups is generated orderly and Ws_r are sorted on the basis of fitness. The territory locations with good and bad fitness are obtained as female and male, respectively.

Mating: The male *Ws* gives a "ripple to the female" *Ws* for the process of mating. Since the female side response is not clear, the mating probability Pr is considered for attracting the female. In the conventional algorithm, Pr is fixed in the constant range of 0.5, which degrades the performance of an algorithm. Hence, in the developed MMP-WSA, the mating probability P_r is updated on the basis of upper-bound and lower bound based concepts as shown in Eq. (13) below.

$$Pr = [(me(R) - me(l_b))/(me(u_b) - me(l_b))] * (1 - 0) + 0 \tag{13}$$

Here, the term $me(l_b)$ denotes the mean of the lower bound and $me(u_b)$ indicates the mean of the upper bound. Then, the value of *R* is computed by the following Eq. (14).

$$R = Ws_j^s - Ws_j^{s+1} \tag{14}$$

Then, the male location is upgraded in Eq. (15).

$$Ws_j^{s+1} = \begin{cases} Ws_j^s + T.rand & \begin{array}{l} if\ mating\ happens \\ (with\ probablity\ of\ \mathrm{Pr}) \end{array} \\ Ws_j^s + T.(1 + rand) & otherwise \end{cases} \tag{15}$$

The length *T* is computed by Eq. (16).

$$T = Ws_f^{s-1} - Ws_j^{s-1} \tag{16}$$

Here, the term Ws_j^{s-1} and Ws_f^{s-1} indicates the male and female *Ws* in (s-1)th cycle.

Feeding: The process of mating requires a huge amount of energy for male water sliders *Ws* searches the food later than the process of mating. The access of objective function for the presence of food is carried out. If the value of fitness is better than the prior condition, the male *Ws* has identified food in a new location, or else it was not. In the further circumstance, the male *Ws* transport to the superlative *Ws* of the pond $\lfloor Ws_{bp} \rfloor$ for founding the food, which is computed in Eq. (17).

$$Ws_j^{s+1} = Ws_j^s + 2rand * (Ws_{bp}^s - Ws_j^s) \tag{17}$$

Fatality and Progression: If the chap *Ws* could not identify food in a new location, then it will die and another male *Ws* will replace the position of the old one. This process is expressed in Eq. (18)

$$Ws_j^{s+1} = l_{b_i}^s + rand * (u_{b_i}^s - l_{b_i}^s) \tag{18}$$

Here, the term $u_{b_i}^s$ and $l_{b_i}^s$ indicates *Ws* location for the maximum and minimum values inside the ith territory.

WSA Execution: If the circumstance of the termination is not satisfied, the baseline approach will send back to the step of mating again for creating a new disk. Here, the "Maximum Number of Function Evaluation (MaxNfes)" is represented as the execution condition. The pseudo-code of the developed MMP-WSA is indicated in Algorithm 1.

Algorithm 1: Pseudo Code of Developed MMP-WSA

Initialize random population
Compute the value of fitness *Ws*
While (execution circumstance is not satisfied) do
    Introduce *ni* region count and allow the *Ws*
    For (every region) do
    Update the mating probability Pr
      The male gives "mating ripples" and the chosen female predicts the reply to attract the female by Eq. (13)
      Update the location on the basis of the female response with Eq.

---

(15)
Analysis of the new location to search for foodstuff for compensate the obsessive power during the process of mate
    If (the male could not identify the foodstuff) then
        Searching for foodstuff sources and reaching the territory of the food-rich region by Eq. (16)
           If (male could not identify the foodstuff again)
                The starving male will be died due to malnourishment
                Another male is replaced in the place of died one by Eq. (18)
           End
        End
    End
Return Ws_{opt}

The flowchart of the projected MMP-WSA is represented in Fig. 3
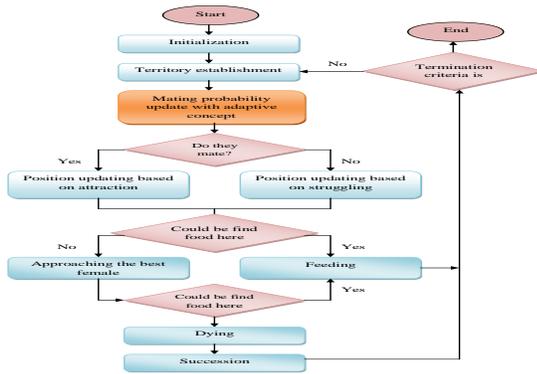
Fig. 3. Flowchart of Proposed MMP-WSA.

B. Optimized Deep Neuro-Fuzzy Classifier

The obtained optimal fused features are given as input for the process of "DR severity classification" by utilizing the developed ODNFC. For the process of image analysis, a rule-based system is modeled by fuzzy network operation. Fuzzy logic is a type of many-valued logic whose variables' truth values can be any real number between 0 and 1. To enhance this operation, the computation of "Single-Input-Multi-Output (SIMO)" is carried out based on Eq. (19).

$$if \ y \ is \ A_n \ then \begin{pmatrix} f_{1n}(y) \\ f_{2n}(y) \\ f_{3n}(y) \end{pmatrix}, n = 1,2,\dots,N \qquad (19)$$

Here, the term y represents the input variable from the discourse universe *(Y), N* is denoted as count of rules, n indicates the output count, A_n denotes the fuzzy set for n^{th} rule determined over *(Y)* and f_{in} implies the output of nth rule, the outputs are considered as non-linear functions of the inputs. For the process of analysis of images, the universe of discourse is considered as the sub-region set of the considered image. Also, the set of fuzzy is considered as the remarkable pattern. Lastly, the membership grade is considered as the similarity in-between the provided pattern and the sub-regions. Every fuzzy rule present in the rule set is entrapped in various patterns present in the image. Hence, the rewrite equation is obtained as shown in Eq. (20).

$$\begin{pmatrix} f_{1n}(y) \\ f_{2n}(y) \\ f_{3n}(y) \end{pmatrix}, n = 1,2,\dots,N \qquad (20)$$

Thus, the proposed model considers SIMO-based system, where the provided image is given as the individual input, the nth rule attracts the pattern in the images of sub-regions, every output represents a non-linear function against the sub-regions for n^{th} pattern, and the output of de-fuzzification is affected by joining of the specific patterns by fuzzy set in every image rules. The operation of fuzzy interference involves the following steps:

1) Computing of the membership: For every rule, the membership grade (M_n) of the matrix is computed, where every element present in the matrix indicates identity in between the sub-regions present in the image and fuzzy set (A_n). The membership values are allotted at the range of $[0,1]$.

$$M_n = [m_{in}] = [a_i, A_n] \qquad (21)$$

Here, the term a_i is the ith image sub-region and m_{in} is the item in M_n.

2) Computing firing strength: The firing strength of every rule is determined by membership grade normalization matrix grade as represented in Eq. (22).

$$\overline{M}_n = \frac{M_n}{\sum_{x=1}^{N} M_x} \qquad (22)$$

3) Computing final output: The final outputs of the system rules are provided as shown in Eq. (23) below.

$$w_i = \sum_{n=1}^{N} w_{in} = \sum_{n=1}^{N} \overline{M}_n f_{in}(y) \qquad (23)$$

Here, the term w_i denotes the output while w_{in} indicating the output of nth rule. From the categorization by developed ODNFC, the output is obtained in the form of severe, moderate, normal, and abnormal DR.

The proposed approach focuses on the main objective function for parameter optimization in developed DR severity classification. The parameters considered for the optimization are "hidden neurons of DNN, learning rate of DNN, EXP-bound in fuzzy and weights of fused features". The objective function is computed in Eq. (24).

$$OBFN = \underset{\{HN_{DVV}, LR_{DNN}, EX_{FUZZ}, WE\}}{\arg\max} (ACCU) \qquad (24)$$

Here, the term HN_{DVV} indicates the hidden neurons of DNN, LR_{DNN} denotes the learning rate of DNN, EX_{FUZZ} indicates the Exp-bound in fuzzy, and *WE* represents the weight. The hidden neurons count for DNN is fixed in the "range of $[5,255]$ and the learning rate count for DNN is fixed in the range of $[0.01,0.99]$ and the weight range is fixed in-

between $[0.01, 0.99]$", respectively. The term *ACCU* represents the accuracy "closeness of measuring the distance to a specific value", which is computed in the following Eq. (25).

$$ACCU = \frac{(tru_p + tru_n)}{(tru_p + tru_n + fal_p + fal_n)} \tag{25}$$

Here, the term tru_p and tru_n denotes" the true positive and true negative values" and the terms tru_p and fal_n indicates the "false negative and false positive values". The developed ODNFC-based DR severity classification is diagrammatically depicted in Fig. 4.
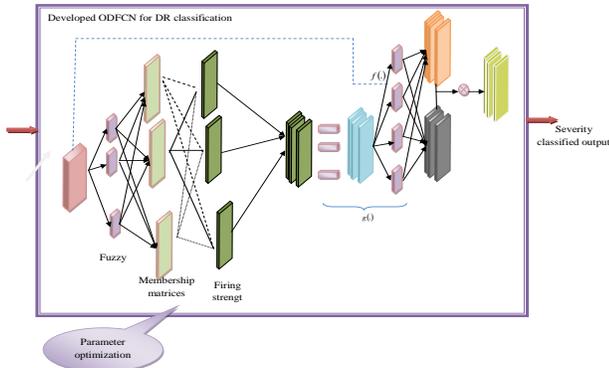
Fig. 4. Flowchart of Proposed MMP-WSA.

VII. RESULTS AND DISCUSSIONS

A. Experimental Setup

The developed DR severity categorization sculpt was established in python and the experimentation analysis is carried out. The concert evaluation of the projected approach was done by the relative examination with traditional models using several quantitative procedures. These quantitative procedures were further divided into two divisions named "positive measures and negative measures". "Positive measures or Type I includes Negative Predictive Value (NPV), Specificity, Accuracy, MCC, Precision, Sensitivity, and F1Score. Negative measures were taken as Type II measures such as False Discovery Rate (FDR), False Negative Rate (FNR), and False Positive Rate (FPR)". The experimentation was carried out on "10 counts of population and the total count of iterations was considered as 25". The proposed MMP-WSA was compared with various algorithms like "Particle Swarm Optimization Algorithm (PSO) [4], Grey Wolf Optimization (GWO) Algorithm [28], Butterfly Optimization Algorithm (BOA) [27], Water Strider Algorithm (WSA) [26] and machine learning algorithms like Deep Neural Network (DNN) [8], Resnet-VGG16 [29], Fuzzy [7], and ODNFC [30]".

B. Performance Metrics

The concert of the established DR severity classification model was tested for quality evaluation of the enhanced approach with several quantitative measures that are briefed as follows.

a) MCC *Mcc* is "a measure of the quality of binary classifications of testing" as Eq. (26)

$$Mcc = \frac{tru_p \times tru_n - fal_p \times fal_n}{\sqrt{(tru_p + fal_p)(tru_p + fal_n)(tru_n + fal_p)(tru_n + fal_n)}} \tag{26}$$

b) Specificity *Spe* is "the proportion of negatives that are correctly identified" as Eq. (27)

$$Spe = \frac{tru_n}{tru_n + fal_p} \tag{27}$$

c) NPV *Npv* is "the sum of all persons without disease in testing" as denoted in Eq. (28)

$$Npv = \frac{tru_n}{tru_n + fal_n} \tag{28}$$

d) F1-score *Fsco* is "the measurement of the accuracy in the conducted test" as Eq. (29)

$$Fsco = 2 \times \frac{2tru_p}{2tru_p + fal_p + fal_n} \tag{29}$$

e) FNR *Fnr* is "the proportion of positives which yield negative test outcomes with the test" as Eq. (30)

$$Fnr = \frac{fal_n}{fal_n + tru_p} \tag{30}$$

f) Sensitivity *Sen* is "the proportion of positives that are correctly identified" as Eq. (31)

$$Sen = \frac{tru_p}{tru_p + fal_n} \tag{31}$$

g) FPR *Fpr* is "the ratio between the numbers of negative events wrongly categorized as positive (false positives) and the total number of actual negative events" as Eq. (32).

$$Fpr = \frac{fal_p}{fal_p + tru_n} \tag{32}$$

h) FDR is "a method of conceptualizing the rate of errors in testing when conducting multiple comparisons" as denoted in Eq. (33)

$$Fdr = \frac{fal_p}{fal_p + tru_p} \tag{33}$$

C. Performance Evaluation on Several Baseline Approaches

The concert evaluation of the developed DR severity classification is carried out by comparison with several algorithms as shown in Fig. 5 at different learning percentages. By taking into consideration, the precision of the proposed MMP-WSA, the performance is 0.3%, 0.42%, 0.38%, and 0.49% enhanced than the PSO-ODNFC, GWO-ODNFC, BOA-ODNFC, and WSA-ODNFC, respectively at the learning percentage of 75. Likewise, for every performance measure, the proposed MM-WSA outperformed in terms of the developed DR severity classification model than the traditional approaches.

(a)   (b)

(c)   (d)

(e)   (f)

(g)   (h)

(i)   (j)
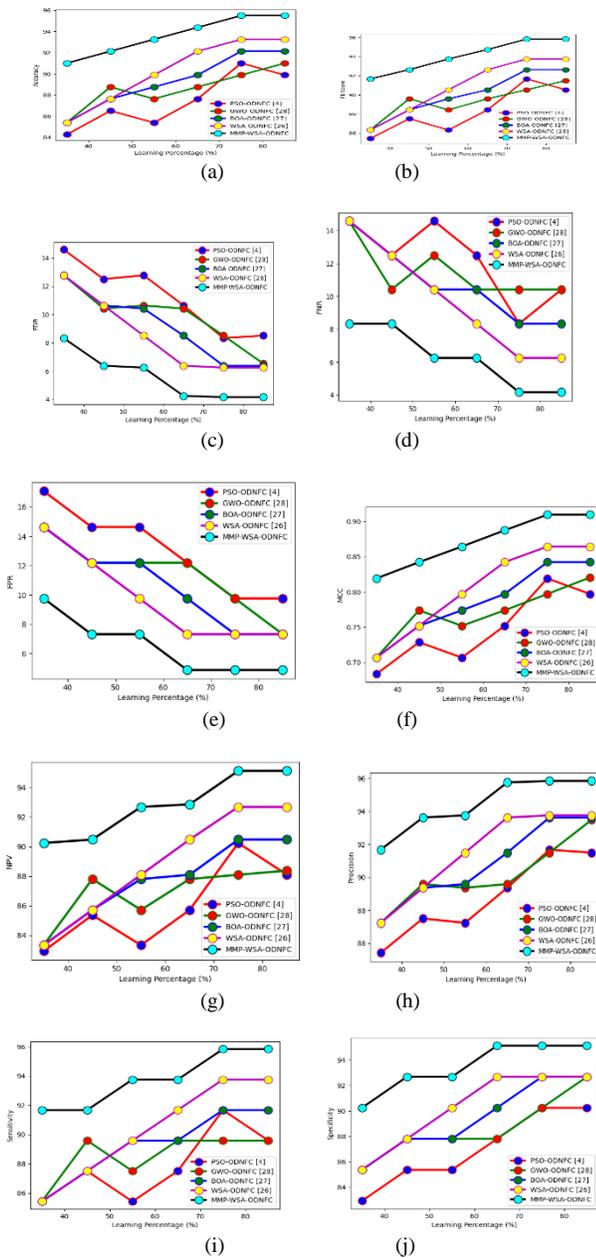
Fig. 5. Performance Analysis of Developed DR Severity Classification Model with Various Algorithms in Terms of "(a) Accuracy, (b) F1-Score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity, (j) Specificity".

D. Evaluation of the Proposed DR Diagnosis Model for Different Optimized Algorithms

The performance of the proposed DR severity classification model is carried out on the basis of developed ODNFC on comparison with different conventional classifiers as shown in Fig. 5 at varying "learning percentages". The DR severity classification model is compared with several meta-heuristic baseline approaches as indicated in Table I for analysing the performance. The comparative analysis of the proposed algorithm with other models denotes the improved performance of the proposed MMP-WSA by minimizing the errors. The performance of the proposed MMP-WSA is

0.0.9%, 0.1%, 0.08% and 0.045% enhanced than the PSO-ODNFC, GWO-ODNFC, BOA-ODNFC, and WSA-ODNFC, respectively. Hence, the proposed DR severity classification model improves its performance among several available models.

TABLE I.   COMPARATIVE NALYSIS ON DR SEVERITY CLASSIFICATION MODEL BASED ON EXISTING META-HEURISTIC ALGORITHMS

TERMS	PSO-ODNFC [4]	GWO-ODNFC [28]	BOA-ODNFC [27]	WSA-ODNFC [26]	MMP-WSA-ODNFC
"ACCURACY"	91.0112	89.8876	92.1348	93.2584	95.5056
"SENSITIVITY"	91.6667	89.5833	91.6667	93.75	95.8333
"SPECIFICITY"	90.2439	90.2439	92.6829	92.6829	095.122
"PRECISION"	91.6667	91.4894	93.617	93.75	95.8333
"FPR"	09.7561	09.7561	07.3171	07.3171	04.878
"FNR"	08.3333	10.4167	08.3333	06.25	04.1667
"NPV"	90.2439	88.0952	90.4762	92.6829	95.122
"FDR"	08.3333	08.5106	06.383	06.25	04.1667
"F1-SCORE"	91.6667	90.5263	92.6316	93.75	95.8333
"MCC"	81.9106	79.7058	84.2213	86.4329	90.9553

E. Evaluation of the Proposed DR Diagnosis Model for Various Classifiers

The complete evaluation is made on the proposed DR severity classification model for performance estimation among several classifiers as portrayed in Table II. By considering the precision of the developed ODNFC, the performance improves O.047%, 0.035%, 0.037% and 0.045% than the DNN, RESNET-VGG16, FUZZY, and DNFC, respectively. Therefore, the proposed DR severity classification model improves its performance than other available models.

TABLE II.   COMPARATIVE ANALYSIS OF DR SEVERITY CLASSIFICATION MODEL BASED ON EXISTING CLASSIFIERS

CLASSIFIERS	DNN [8]	RESNET-VGG16 [29]	FUZZY [7]	DNFC [30]	MMP-WSA-ODNFC
"Accuracy"	87.6404	87.6404	88.764	93.2584	95.5056
"Sensitivity"	85.4167	87.5	89.5833	93.75	95.8333
"Specificity"	90.2439	87.8049	87.8049	92.6829	95.122
"Precision"	91.1111	89.3617	89.5833	93.75	95.8333
"FPR"	09.7561	12.1951	12.1951	073.171	04.878
"FNR"	14.5833	12.5	10.4167	06.25	04.1667
"NPV"	84.0909	85.7143	87.8049	92.6829	95.122
"FDR"	08.8889	10.6383	10.4167	06.25	04.1667
"F1-Score"	88.172	88.4211	89.5833	93.75	95.8333

VIII. CONCLUSION

This paper has developed a novel DR severity classification model with enhanced meta-heuristic-based feature fusion and fuzzy categorization approaches. It includes a number of procedures, together with image enhancement, optic disc

removal, image segmentation, and two sets of "model 1 and model 2" characteristics. The acquired fused features were then obtained after performing feature fusion between two feature sets. Where the weighted utility was optimized by utilizing the developed MMP-WSA in fused features and the optimal weighted fused features were obtained, then the classification of DR severity was carried out by utilizing developed ODNFC, in which its hyperparameters such as hidden neurons and learning percentage of DNN, EXP-bound in fuzzy were optimized by utilizing the developed MMP-WSA. The developed MMP-WSA-ODNFC model has provided higher accuracy as 0.047% than DNN, 0.035% than ResNet-VGG 16, 0.037% than fuzzy, and 0.045% than ODNFC. Hence, the developed DR severity classification model with ODNFC-based categorization using MMP-WSA has provided improved performance than other available DR severity classification models. It also acts as a potential path for clinicians for evaluating huge data and for fastening up DR severity classification.

REFERENCES

[1] J. Vaishnavi, S. Ravi, and A. Anbarasi, "An efficient adaptive histogram-based segmentation and extraction model for the classification of severities on diabetic retinopathy", Multimedia Tools and Applications, Vol. 79, pp. 30439–30452, 2020.

[2] Charu Bhardwaj, Shruti Jain and Meenakshi Sood, "Hierarchical severity grade classification of non-proliferative diabetic retinopathy", Journal of Ambient Intelligence and Humanized Computing, Vol. 12, pp. 2649–2670, 2021.

[3] Hemanth Kumar Vasireddi, Suganya Devi K and Raja Reddy G N V, "Deep feed-forward neural network–based screening system for diabetic retinopathy severity classification using the lion optimization algorithm", Graefe's Archive for Clinical and Experimental Ophthalmology, Volume 260, pp. 1245–1263, 2022.

[4] Nasr Y.Gharaibeh, "Detection of diabetic retinopathy using partial swarm optimization (PSO) and Gaussian interval type-2 fuzzy membership functions (GIT2FMFS)", Materials Today: Proceedings, 2020.

[5] Kuna, Sri Laxmi, and Prasad, Dr. A. V. Krishna, Deep Learning Models for Classification of Diabetic Retinopathy Color Fundus Images (September 14, 2022).http://dx.doi.org/10.2139/ssrn.4218649.

[6] Sayan Das and Sanjoy Kumar Saha, "Diabetic retinopathy detection and classification using CNN tuned by genetic algorithm", Multimedia Tools and Applications, Vol. 81, pp. 8007–8020, 2022.

[7] Yueguo Luo, Jing Pan, Shaoshuai Fan, Zeyu Du and Guanghua Zhang, "Retinal Image Classification by Self-Supervised Fuzzy Clustering Network", IEEE Access, Vol. 8, pp. 92352 - 92362, 2020.

[8] Juan Wang, Yujing Bai and Bin Xia, "Simultaneous Diagnosis of Severity and Features of Diabetic Retinopathy in Fundus Photography Using Deep Learning", IEEE Journal of Biomedical and Health Informatics, Vol. 24, Issue. 12, pp. 3397 - 3407, 2020.

[9] Nitigya Sambyal, Poonam Saini, Rupali Syal and Varun Gupta, "Modified U-Net architecture for semantic segmentation of diabetic retinopathy images", Biocybernetics and Biomedical Engineering, Vol. 40, Issue. 3, pp. 1094-1109, 2020.

[10] Qianjin Li, Shanshan Fan and Changsheng Chen, "An Intelligent Segmentation and Diagnosis Method for Diabetic Retinopathy Based on Improved U-NET Network", Journal of Medical Systems, Vol. 43, No. 304, 2019.

[11] A. Shanthini, Gunasekaran Manogaran, G. Vadivu, K. Kottilingam, P. Nithyakani and C. Fancy, "Threshold segmentation based multi-layer analysis for detecting diabetic retinopathy using convolution neural network", Journal of Ambient Intelligence and Humanized Computing, 2021.

[12] J. Pradeep Kandhasamy, S. Balamurali, Seifedine Kadry and Lakshmana Kumar Ramasamy, "Diagnosis of diabetic retinopathy using multi level

set segmentation algorithm with feature extraction using SVM with selective features", Multimedia Tools and Applications, Vol 79, pp. 10581–10596, 2020.

[13] Lifeng Qiao, Ying Zhu and Hui Zhou, "Diabetic Retinopathy Detection Using Prognosis of Microaneurysm and Early Diagnosis System for Non-Proliferative Diabetic Retinopathy Based on Deep Learning Algorithms", IEEE Access, Vol. 8, pp. 104292-104302, 2020.

[14] Sehrish Qummar, Fiaz Gul Khan, Sajid Shah, Ahmad Khan, Shahaboddin Shamshirband and Zia Ur Rehman, "A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection", IEEE Access, Vol. 7, pp. 150530 - 150539, 2019.

[15] P. Saranya, S. Prabakaran, Rahul Kumar and Eshani Das, "Blood vessel segmentation in retinal fundus images for proliferative diabetic retinopathy screening using deep learning", The Visual Computer, Vol. 38, pp. 977–992, 2022.

[16] V. Desika Vinayaki and R. Kalaiselvi, "Multithreshold Image Segmentation Technique Using Remora Optimization Algorithm for Diabetic Retinopathy Detection from Fundus Images", Neural Processing Letters, 2022.

[17] Gule Saman, Neelam Gohar, Salma Noor, Ambreen Shahnaz, Shakira Idress, Neelam Jehan, Reena Rashid and Sheema Shuja Khattak, "Automatic detection and severity classification of diabetic retinopathy", Multimedia Tools and Applications, Vol 79, pp. 31803–31817, 2020.

[18] Charu Bhardwaj, Shruti Jain and Meenakshi Sood, "Deep Learning–Based Diabetic Retinopathy Severity Grading System Employing Quadrant Ensemble Model", Journal of Digital Imaging, Vol. 34, pp. 440–457, 2021.

[19] Jyostna Devi Bodapati, Nagur Shareef Shaik and Veeranjaneyulu Naralasetti, "Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification", Journal of Ambient Intelligence and Humanized Computing, Vol. 12, pp. 9825–9839, 2021.

[20] Shailesh Kumar, Abhinav Adarsh, Basant Kumar and Amit Kumar Singh, "An automated early diabetic retinopathy detection through the improved blood vessel and optic disc segmentation", Optics & Laser Technology, Vol. 121, 2020.

[21] G. Bhuvaneswari and G. Manikandan, "A novel machine learning framework for diagnosing the type 2 diabetics using temporal fuzzy ant miner decision tree classifier with temporal weighted genetic algorithm", Computing, Vol. 100, pp. 759–772, 2018.

[22] Sulaimon Ibrahim, Pradeep Chowriappa, Sumeet Dua, U. Rajendra Acharya, Kevin Noronha, Sulatha Bhandary and Hatwib Mugasa, "Classification of diabetes maculopathy images using data-adaptive neuro-fuzzy inference classifier", Medical & Biological Engineering & Computing, Vol. 53, pp. 1345–1360, 2015.

[23] Nogol Memari, Abd Rahman Ramli, M. Iqbal Bin Saripan, Syamsiah Mashohor and Mehrdad Moghbel, "Retinal Blood Vessel Segmentation by Using Matched Filtering and Fuzzy C-means Clustering with Integrated Level Set Method for Diabetic Retinopathy Assessment", Journal of Medical and Biological Engineering, Vol. 39, pp. 713–731, 2019.

[24] Sarni Suhaila Rahim, Vasile Palade, James Shuttleworth and Chrisina Jayne, "Automatic screening and classification of diabetic retinopathy and maculopathy using fuzzy image processing", Brain Informatics, Vol. 3, pp. 249–267, 2016.

[25] T. Senthil Kumar and D. Kumutha, "Comparative Analysis of the Fuzzy C-Means and Neuro-Fuzzy Systems for Detecting Retinal Disease", Circuits, Systems, and Signal Processing, Vol. 39, pp. 698–720, 2020.

[26] A.Kaveh and A.Dadras Eslamlou, "Water strider algorithm: A new metaheuristic and applications", Structures, Vol. 25, pp. 520-541, 2020.

[27] Y. Zhu and J. Teng, "Beetle swarm optimization for solving investment portfolio problems," in The Journal of Engineering, vol. 2018, no. 16, pp. 1600-1605, 11 2018.

[28] Seyedali Mirjalili, Seyed Mohammad Mirjalili and Andrew Lewis, " Grey Wolf Optimizer", Advances in Engineering Software, vol.69, pp.46–61, 2014.

[29] M.Toğaçar, B.Ergen, Z.Cömert and F.Özyurt, "A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR

Feature Selection and Machine Learning Models", IRBM, Vol.41, Issue.4, pp.212-222, August 2020.

[30] M. Hanmandlu and D. Jha, "An Optimal Fuzzy System for Color Image Enhancement," IEEE Transactions on Image Processing, vol. 15, no. 10, pp. 2956-2966, Oct. 2006.

[31] Haniza Yazid, Hamzah Arof and Hazlita Mohd Isa, "Exudates segmentation using inverse surface adaptive thresholding", Measurement, vol. 45, pp.1599–1608, 2012.

[32] L. Atikah, N. A. Hasanah, R. Sarno, A. Fajar and D. Rahmawati, "Brain Segmentation using Adaptive Thresholding, K-Means Clustering and Mathematical Morphology in MRI Data," 2020 International Seminar on Application for Technology of Information and Communication (iSemantic), 2020

A Comprehensive Insight into Blockchain Technology: Past Development, Present Impact and Future Considerations

Farhat Anwar[1], Burhan Ul Islam Khan[2]*, Miss Laiha Mat Kiah[3], Nor Aniza Abdullah[4], Khang Wen Goh[5]*

Dept. of ECE, Kulliyyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia[1, 2]
Dept. of Comp. Sys. & Tech., Faculty of CS & IT, Universiti Malaya, Kuala Lumpur, Malaysia[3, 4]
Faculty of Data Science and Information Technology, INTI International University, Nilai, Malaysia[5]

Abstract—**Blockchain technology is based on the idea of a distributed, consensus ledger, which it employs to create a secure, immutable data storage and management system. It is a publicly accessible and collectively managed ledger enabling unprecedented levels of trust and transparency between business and individual collaborations. It has both robust cryptographic security and a transparent design. The immutability feature of blockchain data has the potential to transform numerous industries. People have begun to view blockchain as a revolutionary technology capable of identifying "The Best Possible Solution" in various real-world scenarios. This paper provides a comprehensive insight into blockchains, fostering an objectual understanding of this cutting-edge technology by focusing on the theoretical fundamentals, operating principles, evolution, architecture, taxonomy, and diverse application-based manifestations. It investigates the need for decentralisation, smart contracts, permissioned and permissionless consensus mechanisms, and numerous blockchain development frameworks, tools, and platforms. Furthermore, the paper presents a novel compendium of existing and emerging blockchain technologies by examining the most recent advancements and challenges in blockchain-enabled solutions for a variety of application domains. This survey bridges multiple domains and blockchain technology, discussing how embracing blockchain technology is reshaping society's most important sectors. Finally, the paper delves into potential future blockchain ecosystems providing a clear picture of open research challenges and opportunities for academics, researchers, and companies with a strong fundamental and technical grounding.**

Keywords—Blockchain; blockchain applications; consensus algorithms; distributed ledger; smart contract

I. Introduction

According to the National Institute of Standards and Technology (NIST), blockchains are "tamper-evident and tamper-resistant digital ledgers executed in a distributed form, i.e., without a single repository, and usually without a central authority," i.e., a government, bank, or company. In their most basic form, blockchain technologies provide a platform for the secure movement of data involved in any transaction, including contracts and financial transactions [1]. Cryptography is at the core of blockchain technology, ensuring that the data being exchanged has not been tampered with and gives integrity and authenticity. The transactions involved in blockchain are just a transfer of assets, and the assets are essentially data, which may represent financial information, healthcare information, or even company information [2]. Blockchain is a buzzword that will likely be heard more in the future. Bitcoin and blockchain are gaining technical insight and becoming the chosen technology for implementing a wide range of commercial solutions in the current technological age. Most businesses worldwide are considering using blockchain technology, and even the government is laying the groundwork for the future. In general, individuals may become perplexed by the terms bitcoin and blockchain. Bitcoin is a digital cryptocurrency that may be used to make online payments without relying on a third party. In contrast, blockchain is the platform and structure that ensures every transaction is visible and unchangeable [3].

Blockchain is a public distributed ledger accessible to everyone, and anyone can become a member of this network. Bitcoin is considered the first step in developing blockchain technology [4]. Satoshi Nakamoto, the creator of bitcoin, originally announced the cryptocurrency in 2008 [2]. Satoshi Nakamoto was an unknown individual or group of individuals who began working on the bitcoin concept in 2007 under the name Satoshi Nakamoto. On the 18th of August, they registered the domain name bitcoin.org. Soon after, on the 31st of October, they issued a whitepaper detailing bitcoin, the transaction procedures, Proof-of-Work (PoW), and other aspects of the cryptocurrency. On the 9th of November, 2008, the sourceforge.net website registered the first bitcoin project. At 18:15:05 GMT on the 3rd of January, 2009, the Genesis Block (also known as block 0) was established. Blockchain is not only about tokens and coins. Blockchain is more than bitcoin [5]. The information on a blockchain is stored in blocks, which are bits of data that have been cryptographically encrypted [6]. To build a chain, each consecutive block must contain information about the previous block. As a result, the word blockchain was coined. Cryptographic hash functions and public key cryptography are used to ensure the anonymity of the blockchain [7]. It also aids in the achievement of transparency. New transactions are added to the current information based on the miners' agreement in the network. As a result of existing economic processes such as PoW, Proof-of-Stake (PoS), and others [8], the rules for validating

transactions are codified in the form of algorithms applied by miners who are also paid with a native coin.

Moreover, because the ledger operates on a distributed network, all nodes participating in the network receive a duplicate of the original information [9]. Depending on the scenario, every node in the network serves as a client and a server. Blockchain technology is available in many forms as it has undoubtedly advanced over the last decade. There are several types of blockchain technologies, each with its purpose and set of difficulties. The two most frequent types are public and private, widely utilized by bitcoin networks and private businesses. Also gaining popularity are hybrid blockchains. Besides ensuring the secure movement of currency, the technology creates a permanent historical record of all transactions and a single version of events. This condition is entirely transparent and shown in real-time for the convenience of all participants. However, blockchain technology, irrespective of the type of blockchain protocol implemented, will significantly impact the transformation of

centuries-old business practices, the establishment of greater levels of legitimacy in government, and the creation of new avenues of economic opportunity for common citizens. Fig. 1 gives a brief overview of blockchain technology.

The primary objective of this manuscript is to provide an overview of the many blockchain-related technologies and developments currently in progress. For this purpose, the proposed scheme is based on the desk research methodology considering all the major implementations from the most reputable publishers (such as IEEE, ScienceDirect, MDPI, etc.). Overall, 223 manuscripts were read and evaluated, but only 144 were chosen to be included in this paper. This review only considers papers that primarily discuss implementation strategy and results. As shown in Fig. 2, leading academic publishers like Wiley, Taylor & Francis, ScienceDirect, MDPI, Springer, and IEEE have all published increasing numbers of articles about blockchain technology in the last five years. The outcome shows a spontaneous increase in publication in the last three years.

| A digital ledger that keeps a record of all transactions taking place on a peer-to-peer network | All information transferred via blockchain is encrypted & every occurrence recorded, meaning it cannot be altered | It is decentralised, so there's no need for any central, certifying authority | It can be used for much more than the transfer of currency; contracts, records other kinds of data can be shared | Encrypted information can be shared across multiple providers without risk of a privacy breach |

Fig. 1.   Overview of Blockchain Technology.


(a)

(b)

Fig. 2.   Research Publication Trends Towards Blockchain Technology. (a) Total Publication (2018-2022); (b) % of Increase (PI) in Publication (2018-2022).

A. Organization of the Study

This study is organized as follows: Section I briefs about the significance of using blockchain in current times with a brief on its usage and formation in the industry. Section II discusses the history of blockchain with respect to multiple generations and the essential characteristic of its evolution. It also presents a clear idea about the crucial blockchain features required for confirming the proper functioning of blockchain technologies irrespective of any application. Further, this section discusses three essential types of blockchain: public, private, and hybrid forms, with compact information about its features and associated issues. This section also illustrates all the reported research attempts and highlights their advantages and limitations. Section III reviews existing architectures and components of blockchain along with its working. This section also highlights scientific developments in consensus algorithms. Further, it also presents the rationale behind migrating to SHA-3 encryption and a compact-and-illustrative discussion of approaches of consensus approach in the blockchain. The blockchain framework and platform are illustrated with respect to its advantages and shortcomings in Section IV. Section V examines emerging blockchain applications concerning their utilization and prominent issues. Section VI highlights some of the study's essential findings that offer a contributory inference from the viewpoint of analyzing the strengths and weaknesses of existing approaches in the blockchain. It also presents a briefing about the novelty of the proposed review with some existing notable reviews. A thorough examination of the challenges and unsolved research problems linked to blockchain security is included in Section VII. Further, the characteristic of a unified blockchain and its associated concerns are highlighted. Section VIII discusses the open research questions from the literature review that should be pursued in the future, and Section IX provides conclusions. References are listed at the end.

II. HISTORY, FEATURES AND TYPES OF BLOCKCHAIN

A. History of Blockchain: A Brief Overview of Three Generations

Blockchain technology began as an infrastructure for the bitcoin cryptocurrency and has since evolved into a true game-changer. Unfortunately, in addition to regularly providing new possibilities and applications, this technology has also attracted its fair share of hype and fraud. As a result, many entrepreneurs and engineers remain confused about the actual economic implications of blockchain technology. They are still unsure about its impact on businesses and whether or not they should invest in its development in the first place. Albeit, going over a few data shown in Fig. 3 shows the actual state of the technology and how the world responds to the most essential yet cliche topic of all: how to the technology. This will help better understand blockchain technology's impact on the global economy.
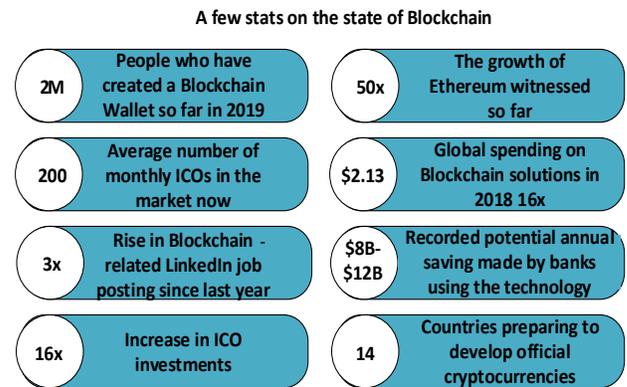
Fig. 3. A Few Statistics Illustrating the Current Development of Blockchain Technology [8].

1) Blockchain 1.0: Blockchain's first generation was intended to improve the traditional monetary system. During this time, bitcoin and other cryptocurrencies were introduced, most of which were written in C++ and followed the PoW consensus paradigm. Although blockchain-based cryptocurrencies improved transactional experience, developers discovered that the technology has far-reaching applications beyond cryptocurrencies. This became the impetus for the creation of the second generation.

2) Blockchain 2.0: As bitcoins remained a hot topic in the market, Ethereum and intelligent contracts became famous in the second blockchain generation. As a result, Ethereum's developers began addressing the coin as a cryptocurrency and platform for delivering a scalable experience and developing decentralized applications, DAPPs. They have also started to investigate the concept of smart contracts, which may be used to make agreements more secure, automated, and practical [7].

3) Blockchain 3.0: In theory, the third generation of blockchain 3.0 is the generation that showcases the most effective blockchain-based solutions currently available. As a result of this generation's focus on the economic and market implications of blockchain technology, several issues related to the creation of blockchain solutions, such as the inability to conduct cross-chain transactions, have been successfully addressed [8]. Furthermore, when it comes to developing blockchain applications, professional blockchain software developers have taken cognizance of the opportunity to use a variety of consensus algorithms other than PoW. Additional concepts such as DLT (Decentralized Ledger Technology), Information for Operational and Tactical Analysis (IOTA), and Currency of the Internet (COTI) have been developed in this generation. They have emerged as the best explanation for how blockchain will transform the world.

4) Blockchain 4.0: Business leaders are lining up to adopt blockchain technology and make it the focal point of their technological solutions. Blockchain 4.0 delivers ideas and solutions for industries that make it functional for the demands of modern business. Automation, integration of execution frameworks, and enterprise resource planning are the focal points of industry 4.0. Blockchain contributes to this modern industrial revolution by giving an increased level of privacy and security insurance to participants in it. Data collecting, asset management, supply chain management, healthcare, and financial transactions are just a few sectors where blockchain has proven beneficial. This means that blockchain 4.0 is improving the suitability of blockchain 3.0 for use in real-world business applications.

Blockchain technology is continually evolving and improving in terms of efficiency and reliability [10][11]. Blockchain technology is, thus, continually evolving and improving in terms of efficiency and reliability. There are various recent advancements towards the utilization of the blockchain-based technology. One such field is to ensure the participation of vehicles in a smart grid environment [12]. One of the critical problems in a smart grid environment is an intruder's theft of user information. Another point of vulnerability in the smart grid is the usage of central nodes for storing and sharing data. There are more malicious activities and vulnerabilities towards such a centralized storage system.

Over time, cryptography's relevance has started to increase in blockchain deployments. The primary role of cryptography in the blockchain is to protect data consistency and user privacy through symmetric or asymmetric encryption [13]. Aside from that, the adoption of digital signatures in the blockchain is more noticeable, which facilitates signing up the data block with transactional records. To date, multiple cryptographic-based approaches have been applied to secure blockchain technology [14]. The main reasons why cryptographic techniques are used in the blockchain are their ability to encrypt data, their inability to change, their ability to be scaled up, and their failure to be disputed [14].

The work carried out by Li et al. [15] used a signature-based blockchain protocol to resist key leakage attacks. The authors have used a puncturable signature scheme where an adaptive signing is carried out based on the bloom filter and Diffie-Hellman structure. Another signature-based scheme was introduced by Shahid et al. [16], where hashing is used for performing one-time signatures. The outcome offers a significantly reduced size of signatures and keys. Cai et al. [17] have constructed a protocol for quantum blind signatures and a blockchain smart contract system. Zhang and Lee [18] have presented a blockchain-based security system using group signatures to authenticate the blocks. The model is claimed to resist multiple attacks related to consensus algorithm vulnerabilities. Xiao et al. [19] have used a multi-signature approach to secure the blockchain platform. The model is claimed to maximize the efficiency of transactions. Existing schemes have also introduced secret-sharing methods to strengthen blockchain operations' security features. Private key distribution protocols that rely on secret sharing have been developed to secure financial applications, as demonstrated by the work of Xiong et al. [20]. Zheng et al. [21] have used generative adversarial networks to develop a unique secret-sharing scheme to mitigate the lost key issue and reduce the communication efficiency using blockchain. Kim et al. [22] have developed a distributed blockchain operation using a local secret-sharing scheme to enhance the cost of communication and storage. Yin et al. [23] have used a unique cryptographic deployment where a decentralized attribute key is used for encryption. A script interpreter is used for implementing ciphering processes to ensure the secret sharing of private data. Lyu et al. [24] have used a combined regulation scheme using threshold secret sharing over blockchain miners and regulatory authority. Further steganography-based blockchain schemes are also noted to be implemented in the existing system. According to this process, the embedding capacity of the host data is initially computed, followed by hiding the secret information and then formulating a blockchain network. The work by Mohsin et al. [25] used Particle Swarm Optimization for carrying out the image steganography process over a blockchain. Sarkar et al. [26] have developed a decentralized network of steganography blockchains. The secret shared key is used for further encrypting the stego-image. The outcome of the study is known to offer a higher payload. The work carried out by Giron et al. [27] has investigated steganographic structures over blockchains using Least Significant Bits over bitcoin. However, the adoption of cryptographic techniques is also associated with various limitations, as follows:

- Expensive Nature: Various studies, for example [27]-[29], have used sophisticated forms of public key encryption, which not only necessitate higher maintenance costs for key infrastructure management but also result in delays.

- Attack-Specific Solution: Most cryptographic-based approaches are highly effective at identifying and stopping predefined attacks [29]-[31], but they have received little attention regarding new evolving threats or dynamic attacks.

- Challenges in Accessibility: In the event of a network attack, data or services may be difficult to access due to the use of digital signatures or potential encryption in a critical time frame.

B. Blockchain Features

Blockchain technology has the following critical characteristics:

1) Security: Due to the usage of asymmetric cryptography, blockchain systems are intrinsically safe. Asymmetric cryptography consists of a series of public keys visible to everyone and private keys only visible to the system's owner. Using these keys, you can confirm that you own the transaction and cannot be tampered with [32]-[34]. The decentralized structure of blockchain systems, which utilize peer-to-peer (P2P) consensus processes, removes single points of failure for data compared to centrally kept data and

hence is substantially more susceptible to being compromised [34].

2) Disintermediation / Decentralization: Decentralization gives users data control and reduces the need for strong central authorities, making the system more equitable and secure. The blockchain system is more reliable and robust because each distributed node is mainly autonomous and has equal obligations and rights. If one distributed node fails, the entire network is not affected [33]. The blockchain's distributed information eliminates data loss or destruction due to reliance on a centralized place and eliminates misuse of information. Decentralized transaction execution and validation reduce intermediary costs and improve performance at central servers [33].

3) Transparency: The blockchain ledger allows anybody to see the history and specifics of each transaction. This amount of transparency is unheard of in massive financial systems. This kind of openness is achieved using a blockchain network with numerous validating peer nodes and no central authority [35]. Aside from the economic suitability of large corporations, the transparency feature has found application in healthcare and clinical trial data disclosure. Individual patients can utilize blockchain technology to readily examine their claims, medical history, transactions, and past-due payments. Researchers, doctors, and patients have long kept clinical trial data secret, causing a lack of trust in findings [36]. The transparency of blockchain has been found to assist supply chain management malpractices and product history obscurity [36]. Transparency may also help ensure fair elections and increase voter confidence [29].

4) Autonomy: Trust is established between the parties in most transactions, ensuring mutual commitment. Trust is no longer an issue with blockchain technology. The blockchain can work as a P2P system without a responsible third party. Participating nodes on the blockchain system using advanced distributed consensus algorithms [33] to handle the owner confirmation problem in transaction processes while maintaining system integrity. These blockchain transactions are executed without the intervention of a third party due to failsafe consensus mechanisms [36].

5) Immutability: It is also referred to as unforgeability, untameability, immutability, and persistency. Immutability indicates that it cannot be changed or interfered with once data is added [34]. To prevent tampering, data blocks in a blockchain structure are timestamped and encrypted with a hash algorithm [33][36]. However, immutability poses its own difficulties and obstacles for blockchain technology, and some are now questioning its benefits [37].

6) Traceability: Transparency is the capacity to track data's origin, destination, and update sequence between nodes. While data traceability is required to ensure data integrity and trust, it also improves data governance, compliance with legislation, and understanding of the impact of change [37]. Data traceability is supported by blockchain technology due to the time stamping of data updates and additions.

7) Anonymity: Blockchain's anonymity protects privacy against unlawful entry or surveillance. To maintain anonymity, transactions must be authorized without revealing personally identifiable information about the parties involved. The data is shared between nodes utilizing a trust-based method. Thus, node information is not disclosed or validated, and the data transfer is anonymous [38]. Users can communicate with produced blockchain addresses to hide their real identities. However, due to the public and distributed nature of the blockchain, total privacy cannot be guaranteed.

8) Democratized: P2P voting allows all nodes in a blockchain system to make choices democratically. Decentralized nodes employ consensus processes to enable particular nodes to introduce additional blocks to an existing blockchain, ensuring the bitcoin is correctly recorded to the shared data ledger, and its duplicates are perfectly synchronized. Nodes can accept blocks by extending them and reject incorrect blocks by not extending them [39].

9) Integrity: Blockchains are designed to resist data modification. Data integrity ensures data accuracy and consistency throughout its life cycle [35]. To do this, the blockchain network uses decentralized, immutable shared ledgers. Once a data block is approved to be included in a blockchain, it cannot be modified or amended. As a result, data reliability and integrity are essentially guaranteed [33].

10) Programmability: A standard application programming interface, API, allows users to design apps using blockchain technology. Smart contracts and decentralized apps can be created using the flexible scriptwriting framework. A programming interface for network administration is provided by the node software-defined networking (SDN) controllers [32]. Users, data query layers, infrastructure layers, and existing database layers are all data structuring provenance proposed by authors [40]. All blockchain systems provide a scripting language [41]. Authors in [32] recommend a user-friendly API. Ethereum, Tron, and Cardano are examples of programmable blockchains.

11) Fault tolerance: Blockchain is designed to be redundant and inefficient to enable excellent fault tolerance and immutability [42]. The P2P architecture of the network allows each node to act as both a client and a server, giving the network an extremely high error margin for node failures and network transit issues [42]. Blockchain is supposed to be Byzantine Fault-Tolerant, meaning that even if some nodes are down or functioning incorrectly, the network will still reach a consensus.

12) Automatic: Using specialized consensus protocols, all nodes in the system may validate and transact data automatically. Blockchain is managed and confirmed without manual involvement by a protocol [41].

C. Types of Blockchain

Blockchain can be divided into public and private categories: A public blockchain is a permissionless distributed ledger. Anyone with an internet connection must sign up on a blockchain platform to join the network. As it is primarily

used for cryptocurrency mining and trading, a public blockchain user or node can access past and present data, validate transactions, and mine cryptocurrencies. The most common public blockchain is bitcoin and Litecoin. A private blockchain is typically utilized within an organization or industry where only a small number of individuals are allowed to participate in a blockchain network instead of a public blockchain. Both public and private blockchains are further divided into permissioned and permissionless blockchains. Table I compare and contrast the permissions of various blockchain types:

Another form of blockchain is called a hybrid blockchain which integrates private and public blockchains. Users can

control who has access to the blockchain data. Encrypted data can be made public while the remainder of the blockchain remains secret. Due to the flexibility of the hybrid system, users can quickly join a private and public blockchain. An example is the Dragonchain. The similarities and differences between different types of blockchains are illustrated in Table II.

Currently, various scientific research contributions are made toward the above-mentioned different forms of blockchain. Table III highlights the frequently encountered problems being addressed by notable researchers using multiple blockchains.

TABLE I. PERMISSIONS OF VARIOUS BLOCKCHAINS

Public and Permissioned	Public and Permissionless	Private and Permissioned	Private and Permissionless
Restricted and open	Transparent and open	Restricted (hybrid methodology)	Read transparent but restricted
Read restricted and write all	Read all and write all	Read restricted and write restricted	Read all and write restricted
All can connect and transact, but read and audit is limited to permissioned users only	All can connect, transact, audit, and read	No one is allowed to connect, transact, audit, and read	All can connect, none is allowed to transact, and all can audit and read
Full write equity	Strictly democratic owing to total equity	Restricted	Full read equity
Example: Ethereum	Examples: Litecoin, Ethereum, and bitcoin	Example: Corda, Hyperledger Fabric, and R3	Example: Hyperledger Fabric

TABLE II. PERMISSIONS PERSPECTIVE OF SIMILARITIES AND DIFFERENCES OF BLOCKCHAINS

	Public	Private	Hybrid	Consortium
Membership	Unidentified Permissionless Can be malicious	Identified Permissioned Trustworthy	A mixture of Permissionless and Permissioned Identified Trustworthy	Permissioned Identified Trustworthy
Consensus Methodologies	PoW, PoS, Proof-of-Authority, Proof-of-Elapsed time, etc.	Voting or multi-party consensus Algorithm	Private Sidechains Consensus Mechanisms	Multi-party consensus algorithm or voting
Speed of Transaction	Time-consuming	Faster and lighter	Faster and lighter	Lighter and faster
Consumption of Energy	Huge energy consumption	Small energy consumption	Small energy consumption	Small energy consumption
Data in Blockchain	No finality 51% attack	Enable finality	Enable finality	Enable finality
Network	Decentralization	Partial decentralization	Partial decentralization	Partial decentralization (Hybrid between Private and blockchain)
Description	Anyone can write and read on the network, irrespective of its location. Data is validated by each network member ("node").	Permission to write and read data onto the blockchain is controlled by a "highly trusted" organization - the blockchain owner.	Controlled Permission on Read and Write.	Permissions to validate, write and read on the blockchain are regulated by some predetermined nodes, which might not be the same for every blockchain entity.
Benefits	- Secure - Transparent	- Efficient - Private	- Access control - Better performance - Scalability	- Efficient - Private
Challenges	- Inefficient -Performance -Scalability -Security	- Controlling (power is confined to one organization only) - Challenging to align several organizations to use a single blockchain - Trust - Auditability	- Difficult to align many organizations (Especially Competitors) & Set Consensus rules to join the blockchain as board members & share their data.	
Use Case	Cryptocurrency Document validation	Supply chain Asset Ownership	Medical Records Real Estate	Banking Research Supply chain

TABLE III.   EXISTING RESEARCH CONTRIBUTION FROM VARIOUS TYPES OF BLOCKCHAIN

	Authors	Problems	Methodology	Advantage	Limitation
Public Blockchain	Lee et al., 2019 [43]	Immutability	Hashing, sidechains	Higher scalability in design	Model not benchmarked
	Guo et al., 2019 [44]	Storage optimization	Redundant Residual Number System	Highly fault-tolerant	No comparative analysis
	Baza et al., 2021 [45]	Security in ride-sharing applications, fair payment	Analytical modeling based on trust, reputation	Ensures privacy preservation	Lacks comparative analysis, no extensive analysis towards security
	Asheralieva and Niyato, 2021 [46]	Resource management	Stochastic Stackelberg game, deep learning	Effective convergence performance	Success depends on a singular environment
	Mohammadzadeh et al., 2021 [47]	Invoice factoring	Smart contract management, Diffie-Hellman Key exchange	Highly simplified implementation scheme	Involves extensive time for signature verification
	Cai et al., 2021 [48]	Data privacy in crowdsensing	Open service system, private blockchain, SHA-256	Offers better verifiability, trust, and robustness	Assumption depends on the discretion of the client and could narrow down to the centralized crowd
	Bai et al., 2022 [49]	Participation of public	Consensus algorithm	Address scalability issues	Couldn't resist other forms of attacks apart from Denial of Service (DoS)
Private blockchain	Wu and Tsai, 2019 [50]	Securing agriculture network	Dark web technology	Effective privacy preservation	Resistive against only Distributed Denial of Service (DDoS) attack
	Hou et al., 2021 [51]	Data tampering in Internet of Things (IoT)	Secure consensus network	Better latency performance	Demands predefined information of the incoming malicious transaction
	Toyoda et al., 2020 [52]	Bottleneck analysis of private blockchain	Analytical framework on Ethereum	Better throughput performance	Doesn't exhibit extensive analysis with comparison
	Xu et al., 2021 [53]	Privacy on data sharing	Publishing protocol using histogram	Ensure data anonymity	Highly complete encryption scheme for large network
	Huang et al., 2020 [54]	Consensus problem	Analytical approach using the probability of network splitting, time out period of election, and rate of packet loss	Optimizes consensus algorithm	No comparison with the existing consensus-based approach
	Chattaraj et al., 2022 [55]	Access control in Software-Defined Network	Access control using a certificate, attribute-based encryption	Applicable for next-generation network	Not resistive against dynamic attackers
	Baucas et al., 2021 [56]	Monitoring of smart homes	Trilateration using the received signal strength indicator	Highly reduced execution time	Not benchmarked
	Shah et al., 2021 [57]	Resource optimization in a power grid	Heuristic approach	Reduced computational time	Demands extensive resources to process
Hybrid blockchain	Kim et al., 2022 [58]	Interoperability in heterogeneous hybrid blockchain	Analyzing multiple blockchain architecture	Identified tradeoffs in existing blockchains	Study applicable to a specific form of blockchain
	Liu et al., 2020 [59]	Energy management	Stackelberg game, blockchain	Optimal energy scheduling	Utility allocation is based on limited constraints
	Akkaoui et al., 2020 [60]	Data exchange in healthcare	Blockchain and edge computing	Significant reduction in execution time	No comparative analysis to prove its effectiveness
	Polge et al., 2021 [61]	Extensibility of blockchain operation	Emphasizing upper layers with real network infrastructure	Enhanced rate of propagation	Complex deployment cost, specific to the bitcoin application
	Zhu et al., 2020 [62]	Leveraging Crowdsensing operation	Blockchain and crowdsensing, dual consensus protocol	Extensive privacy protection	Cannot resist dynamic attacks
	Subramanian et al., 2021 [63]	Supply chain in pharmacy	Cryptocurrency, mobile application	Offers reliable suggestions for critical illness	Not assessed for a large number of operations
	Fan et al., 2021 [64]	Privacy in IoT	Federated learning, reverse auction	Offers better computational efficiency	Not assessed for a large number of operations
	Zhang et al., 2021 [65]	Controllable decentralization	Unique account scheme, an unspent transaction output	Enhance transaction processing	Higher maintenance cost
	Subramanian and Thamppy, 2021 [66]	Tracking automobiles	Ethereum blockchain, Truffle platform, Astute contract	Transaction transparency	Not assessed for a large number of operations
	Cui et al., 2020 [67]	Authentication in sensor network	Local and public chain, mutual authentication	Better security performance	Doesn't address resource demands

From the above table, it is noted that various significant research work is being evolved to improvise the performance and address multiple problems on different taxonomies of blockchain. Irrespective of various methodologies being adopted, it is observed that almost all the existing approaches toward different types of blockchains are witnessed with limiting factors. The most prominent limiting factor is narrowed implementation scope, where the system model's performance is not practically or mathematically proven for its efficiency over a large-scale deployment environment. Apart from this, the lack of benchmarking is another questionable factor towards reliable implementation over ground reality.

III. BLOCKCHAIN ARCHITECTURE

Blockchain is a decentralized database organized in a linked list of blocks, as shown in Fig. 4. Block: A block is a set of valid transactions [68]. In a blockchain system, each node can initiate an exchange of information and broadcast it to all other nodes in the network. The network nodes use past transactions to validate the transaction, and after the transaction is validated, it is added to the current blockchain. The number of transactions is aggregated and inserted into the blockchain block depending on the time window. "A block may include more than 500 transactions on average," according to bitcoin, "and the average block size is roughly 1 MB, an upper bound proposed by Satoshi Nakamoto in 2010" [69]. The type of blockchain used determines the type of data recorded within a block. In the case of bitcoin, a block comprises sender information, receiver information, and the number of bitcoins to be transmitted. The Genesis Block, often known as block 0, is the foundation for all subsequent blocks in a blockchain. In the chain of blocks, each new block is linked to the previous block by a connecting link, as shown in Fig. 4.

As illustrated in Fig. 5, each block has a header and a body. The block header (Fig. 5) contains metadata about the block, such as the previous block's hash, the nonce, the Merkle tree root, and the date.

- Previous block hash: Every block in the blockchain derives from the block before it in the inheritance chain, as shown in Fig. 6. The blockchain system is tamper-proof by using past blocks' hash to generate the new ones' hash.

- Mining statistics used in the block construction: The technique must be complex enough to prevent tampering with the blockchain.

- Bitcoin Mining: Using the previous block hash value, the transaction root hash value T, and the nonce value $nonce$ obtained by solving the consensus method, the current block hash value is calculated using (1).

$$H_k = Hash(H_k - 1|T| \,|nonce|) \tag{1}$$

- Merkle Tree Root: A Merkle tree structure organizes the transactions. The Merkle tree's root is a verification of all transactions. The Merkle root illustrated in Fig. 7 is utilized to build the block hash.
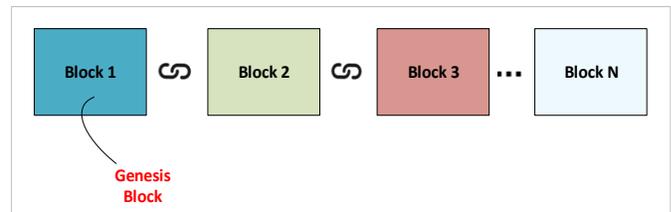
Fig. 4. Blockchain: a Chain of Linked Data Blocks.
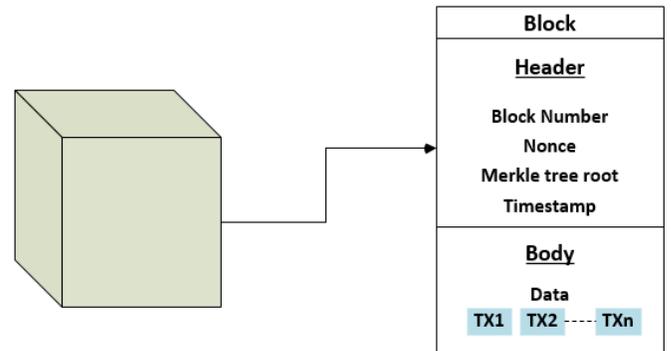
Fig. 5. Structure of a Block.
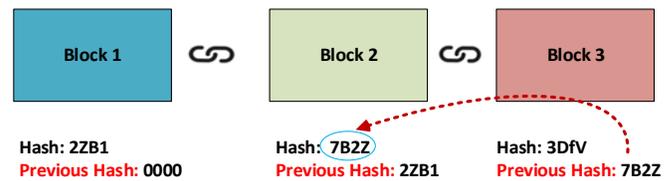
Fig. 6. Blocks Linked by Reference to the Previous Block Header Hash.

Blocks are linked together in a "chain," as depicted in Fig. 8, by incorporating the previous block header hash. This preserves the chain's integrity because modifying one last block will necessitate updating every subsequent block.

Blockchain technology is based on three fundamental principles: transparency, authenticity, and auditing. Its transparent and encrypted chain design allows everyone to witness the information flow between numerous blocks with all the details. It is like a well-connected distributed network. At the start of every new transaction, every computer database competes with all other databases to produce a new block with the necessary features. The database that successfully solves the mathematical coding and creates the block wins. This block is audited and authenticated by the other blocks. After all the blocks confirm the transaction, the data is encrypted but transparently stored in the database. Auditing and authentication processes are automatically initiated and run in the background for every block. Blocks are connected in an appropriate linear, chronological sequence, with every block holding a hash of the previous block. The transition history is still available in other blocks even if the initial original traction block is removed or hacked. As illustrated in Fig. 8 and Fig. 9, the blockchain architecture comprises six significant components integrated into the many layers of the blockchain.
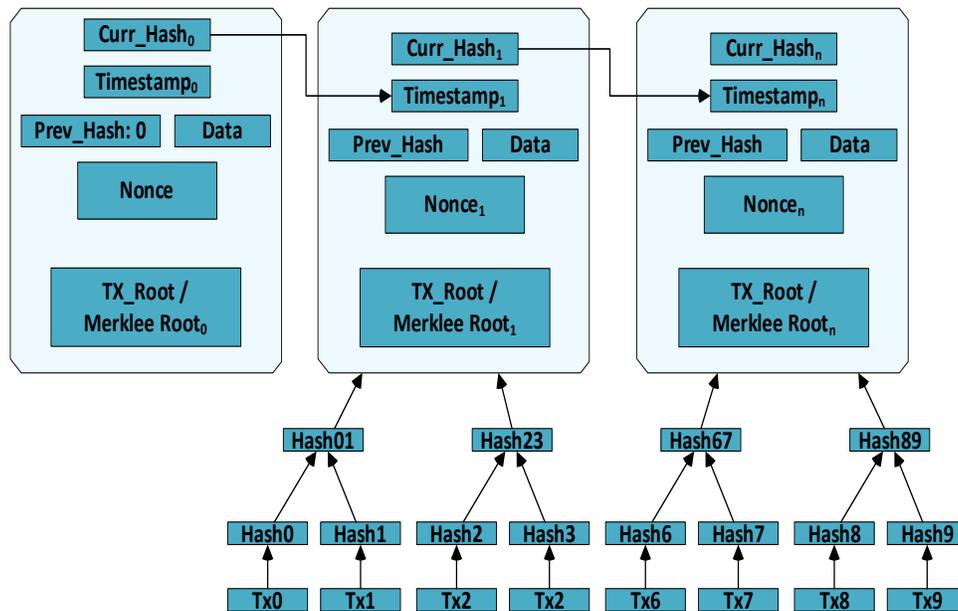
Fig. 7.   A Merkle Tree Example (Binary Hash Tree).
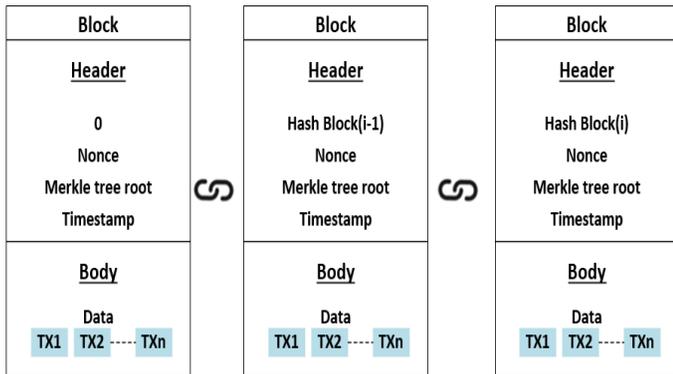
Fig. 8.   The Architecture of a Data Chain in a Blockchain Network.

Fig. 9.   Blockchain Layered Architecture.

1) Data layer: This layer comprises the chained data blocks and the techniques that go with them, like hash algorithms (e.g., SHA256 [75]), timestamp technologies, asymmetric encryption (e.g., Elliptic Curve Digital Signature Algorithm (ECDSA) [76]), and Merkle trees [77]. Each node employs various strategies to encapsulate transactions and received code in new blocks, including the hash function and Merkle tree. Each block is given a timestamp showing when it was created. After that, this new block will join the chain by connecting to the original block [70].

2) Network layer: The network or the P2P layer handles inter-node communication. It handles block propagation discovery and transactions. Also known as propagation, nodes cannot discover each other and synchronize without this P2P layer.

3) Consensus layer: This layer contains several consensus techniques (such as Proof-of-Stake (PoS) and PoW [76]) for maintaining data consistency and fault tolerance in distributed networks.

4) Incentive layer: With this layer, the blockchain's security verification can be strengthened by offering specific incentives to nodes cooperating in security verification. This layer incorporates economic rewards into the blockchain, the

currency distribution system, and the currency issue mechanism to ensure that individuals who help generate the next block receive a possible benefit [70]. Miners (the creators of new blocks) are rewarded with incentives (some coins) to encourage the network to continue working on data verification and ensuring the chain's security [71].

5) Contract layer: In this layer, there are a variety of scripts and algorithms, as well as smart contracts, which are self-executing code that is saved and secured by the blockchain.

6) Application layer: This layer consists of potential applications, scenarios, and use cases [70], among other things.

A. Components of Blockchain

A blockchain comprises many different components, each serving a specific purpose in its creation. These components are [72]:

1) Ledger: A blockchain ledger is a distributed, immutable record of its history used for decentralized data storage.

2) Peer network: A P2P network is a distributed network architecture that allows participants to share resources. The participants make their resources (such as storage capacity, processing power, link capacity, printers, etc.) accessible to other participants at their leisure. Each peer (participant node) in such a network performs the functions of both (server and client).

3) Membership services: Some blockchain types require prior authorization to participate. Indeed, membership services on the blockchain are responsible for authorizing, authenticating, and managing users' identities.

4) Smart contract: It is a self-executing contract in which the agreement's terms are encoded in lines of code that both parties can read and comprehend. Put another way. It is the digital equivalent of a traditional paper contract. Nick Szabo, an American computer scientist and digital currency researcher, suggested smart contracts in 1994. The smart agreement is executed over a blockchain network, duplicating its code throughout the network's machines. This makes it possible to be secure and transparent and facilitate contractual obligations.

*5) Wallet:* Stores key pairs and users' credentials to initiate and sign digital transactions.

6) Events: Refers to notifications of blockchain updates and activities; examples of events include adding a new block to the blockchain, alerts from smart contracts on the blockchain that enable such contracts, and the creation and propagation of a transaction made across the peer network.

7) Systems management: Blockchain is a constantly growing system that must evolve to suit its participants' needs. Indeed, systems administration provides the capability to update, monitor, and build blockchain components to suit users' needs.

B. Consensus Algorithms

A consensus technique is used in the blockchain network to reach a consensus on the status of the distributed ledger [73]. With no central authority, all network participants should collaborate to make the best choice for the network. In a context where strangers do not have faith in one another, achieving consensus among participating nodes was perhaps the most significant development that opened the way for blockchain and resulted in a set of methods to reach an agreement among participating nodes. Consensus mechanisms are communication protocols that allow robots or humans to operate together decentralized. Fig. 10 depicts our classification of the blockchain consensus model, developed after surveying the literature. As shown in Fig. 10, consensus models may be divided into nine broad groups with some different varieties based on subtle changes in how they operate.
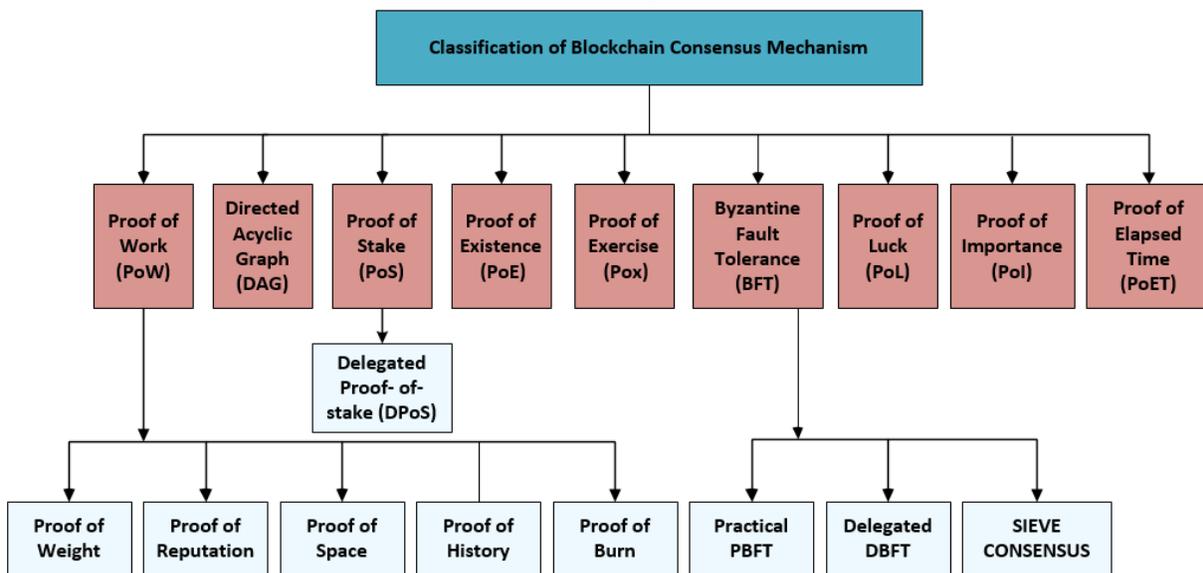

Fig. 10. Blockchain Consensus Algorithms.

TABLE IV.   COMPARATIVE ANALYSIS OF CONSENSUS ALGORITHMS

Consensus Algorithm	Chief Characteristic	Fault Tolerance	Scalability	Power Consumption
Proof-of-Work (PoW)	Power of computation	Poor	High	Enormous
Directed Acyclic Graph (DAG)	Consensus for IoT blockchain	NA	Very high	High
Proof-of-Stake (PoS)	Stake in terms of the number of coins	Less than 51 per cent stake	High	Poor
Proof-of-Existence (PoE)	Use the timestamp of the transaction to validate document existence	NA	Unknown	Unknown
Proof-of-eXercise (PoX)	Miner shall solve a matrix-based problem	NA	Unknown	Fractional energy saving
Delegated Proof-of-Stake (DPoS)	Voting to choose the witness node	Less than 51 per cent of validators	High	Poor
Proof-of-Elapsed Time (PoET)	Lottery based election	Yes		Massive
Byzantine Fault Tolerance (BFT)	Reach consensus despite the failure of some nodes to respond	33 per cent of nodes are Faulty	Poor	Poor
Delegated Byzantine Fault Tolerance (DBFT)	Reach a consensus with Untrustworthy participants	Less than 33 per cent of replicas	Low	Medium
Proof-of-Importance	Estimate importance counts coins that have been in an account for a fixed period	Unidentified	Moderate	Low energy saving
Proof-of-Luck	Cumulative luck value	NA	It does not scale well	Reduced power of computation

Table IV compares consensus algorithms based on essential criteria such as consensus model prominence, scalability, and fault tolerance.

1) Proof-of-Work: PoW is widely considered the blockchain's fundamental consensus model. PoW is built on the idea of competing for computational power to create new blocks in the blockchain. This approach requires a miner to compute and output a value. The winning value is lower than the value set by the network. PoW through the nodes is used to deal with forking (i.e., two nodes provide the winning value). The research community has developed various PoW techniques [73], such as:

a) Proof-of-Weight: Based on algorithm consensus, Proof-of-Weight consensus adds "weight" to the essential principle of PoW. These weights are proportionate to the values generated by the nodes in the network. The goal is to avoid "double-spending," where a consumer can use the same digital token twice by adding relative weight.

b) Proof-of-Reputation: A node's reputation is built by participating in transactions and holding assets. Blockchain validates a new block created by the highest reputation node. This technique prevents nodes from a reputation for prior misbehavior and adds to the blockchain's security.

c) Proof-of-Space: It is a PoW where a node requesting service must allocate more disc space than usual. This data is given to the verifier node to prove that enough space is allocated for a service request.

d) Proof-of-History: This consensus approach demands nodes to offer evidence of history. It establishes a historical record to prove an event that happened at a particular time. This is an alternative to trusting the transaction's timestamp.

e) Proof-of-Burn: This consensus method relies on burning coins to mine the next blockchain block and sending bitcoin to an unrecoverable address. Nodes burn more coins to increase lottery odds. Compared to PoW, Direct Acyclic Graph (DAG) is advocated as a viable blockchain consensus method for the IoT blockchain framework (IOTA). Scalability is an advantage of DAG since data is uploaded to the blockchain in parallel. It adds a block to the ledger as the previous transaction is processed. DAG also combats "double-spending" using powerful algorithms.

2) Proof-of-Stack: This method avoids the need for expensive mining equipment. A PoS node can execute mining or block validating based on its stake [74]. PoS recommend buying cryptocurrencies to increase block formation chances.

a) Delegated Proof-of-Stack (DPoS): It is a PoS variant where participants are encouraged to vote for the witness node, which will generate a block in the blockchain and serve as a witness [94]. The witness node is paid for producing blocks but is barred from future voting if it cannot create blocks.

3) Proof-of-Existence (POE): It is a system for authenticating the presence of documents at a specific time. Without releasing the data itself, data ownership information might be revealed. This POE paradigm helps show the existence of intellectual property documents such as patents.

4) Proof-of-Exercise (PoX): It offers a Proof-of-Exercise alternative to the PoW [95]. An exercise in PoX is a matrix based on a real-world scientific issue. It works on matrices created by system employees. Ribonucleic acid (RNA) and Deoxyribonucleic acid (DNA) sequencing and data comparison are excellent examples of matrix problem-solving.

5) Byzantine Fault Tolerance (BFT): System failure due to Byzantine fault in blockchain requires consensus [73]. Even if some nodes fail to reply, the network should reach an agreement and maintain data consistency. The distribution system makes it difficult.

a) Practical Byzantine Fault Tolerance (PBFT): Authors in [87] suggest PBFT in their work. Authors in [75] offer the first state-machine replication mechanism for asynchronous networks - a distributed file system with BFT.

b) Delegate Byzantine Fault Tolerance (DBFT): The DBFT described in the NEO whitepaper is a variation of the traditional BFT [76]. NEO blockchain's core library is currently utilizing this technology. According to them, a novel mathematical model may be used to test consensus behavior when combined with a discrete model. When compared to other algorithms, this method is more capable of coping with participants who are not trustworthy.

c) Sieve consensus: Sieve is a form of PBFT consensus for non-deterministic chain code execution [77]. Replicates can create non-deterministic chain code output. The sieve can assess the output if a small number of replicas show minimal divergence.

6) Proof-of-Luck (PoL): A new consensus approach reduces the necessary processing power and increases transaction throughput [78]. This algorithm is based on Trusted Execution Environment (TEE). PollRound and PollMine are the primary functions where each block is given a luck value between 0 and 1. The total of all luck values included within each blockchain block is combined to provide a cumulative luck value. A miner will choose to join the luckiest chain.

7) Proof-of-Importance (PoI): New Economy Movement (NEM/ XEM) uses this algorithm. This system assigns a rating to accounts based on the number of vested and unvested coins. PoI measures "importance" by how long a node's coins have been in their account. Every day, 10% of the current unvested sum becomes vested. It also depends on the account's position within the network and the number of coins vested.

8) Proof-of-Elapsed Time (PoET): For each block, a new leader is chosen by a lottery system under this consensus option. TEE is used to ensure the integrity of the electoral process. The basic steps in choosing a leader are as follows:

- It runs on miner and validator nodes.

- The validator node with the least wait time wins the election as a leader node.

The fundamental disadvantage of this consensus technique is that it requires specialized hardware [74]. A good consensus algorithm is efficient, safe, and easy to implement. Recently, efforts have been made to improve the consensus algorithms used in blockchain technology. New consensus mechanisms are being developed to solve some challenges associated with the technology. The central concept of PeerCensus [79] is to isolate block generation from transaction confirmation to boost the consensus speed dramatically. An improved consensus algorithm for ensuring blocks is produced orderly and predictable has been suggested by authors in [80]. Bitcoin's security is known to be compromised by fast block production. The Greedy Heaviest Observed Subtree (GHOST) chain selection rule [81] was devised to address this. It weighs the branches so that miners can choose the best one. Authors

in [82] proposed a consensus protocol that rewards non-interactive retrievability evidence for P2P blockchain networks. Miners save storage space by storing old block headers instead of complete blocks.

IV. BLOCKCHAIN FRAMEWORK AND PLATFORM

A blockchain framework is a software system that facilitates the deployment and creation of blockchain applications. The blockchain frameworks provide infrastructure and libraries for application development. Nodes and the software they run make up the network infrastructure. The node can be physical, virtual, or containerized. Blockchain identity management is controlled by software that provides capabilities and features such as transaction data, user identification, and the consensus mechanism for blockchain. A smart contract application is composed of code that operates within its architecture. The client application communicates with the infrastructure. The latter serves as an external interface to the application. Developing applications outside the blockchain network should be possible with a good blockchain platform. As long as the program works on a small network, it should work on a more extensive infrastructure [83]. Before developing the software, there is no need to set up the network infrastructure. Choosing an enterprise blockchain framework is difficult because no single framework has all the functionalities. As demonstrated in Fig. 11, the primary blockchain framework challenges are storage, processing power, and scalability.

Fig. 11. Blockchain Computational Complexity.

Stakeholders must be highly cautious before, during, and after using blockchain in the enterprise. Many factors must be addressed while selecting a blockchain framework. Table V shows some of the selection criteria for blockchain frameworks.

TABLE V. STANDARDS OF SELECTION FOR BLOCKCHAIN FRAMEWORKS

Criterion	Explanation
License	Features and kind of license (paid or free)
Activity	Framework upgradability support
Support model	Framework support, popularity, longevity
Roadmap	Framework roadmap and vision
Reliable backing	Corporate or open-source community
Ease of use	Intuitive and widely adopted

This section describes the most popular blockchain frameworks and platforms and is as under:

1) Bitcoin - (programming language: C++): It is the world's first and most popular cryptocurrency. Someone or a group using the pseudonym Satoshi Nakamoto established it in 2009. Cryptocurrencies like Ethereum, Litecoin, Dash, and

bitcoin cash are the blockchain's descendants. Bitcoin is a P2P electronic payment system that does not require a central clearing agency. Cryptocurrency users execute transactions by sending electronic instructions that specify who should be credited and debited and where the change should be deposited (if any). The pros and cons of bitcoin are depicted in Table VI.

TABLE VI. ADVANTAGES AND DRAWBACKS OF BITCOIN

Advantages	Drawbacks
Payment flexibility	Volatility and risk
Security and control	The size of the blockchain is about 242.2 GB
High capitalization	Time-consuming (7 transactions per second)
Minimal commission fee	Complete lack of fully-fledged smart contracts
Large-scale trading	Lack of understanding and awareness

2) Ethereum – (programming language: Go, Solidity): Ethereum is a framework with an open code for developing and launching almost any decentralized online services for a blockchain (DApps), whose work is based on smart contracts. Ethereum is a cryptocurrency that is built on the Ethereum blockchain. Vitalik Buterin made an offer in 2013 that was accepted. There are four chief components in Ethereum:

- Instead of designing different frameworks for each application or language, EVM allows all of them to be used on one blockchain. This helps create DApps in lotteries, for example.

- Smart Contracts are computer algorithms for trading gold, cryptocurrency, and other assets without a third party. If the smart contract's code's criteria are met, it executes automatically.

- DApps use smart contracts for market predictions, digital signatures, and asset transfer assurances. Most existing DApps utilize Ethereum.

- Network performance is improved via a collection of programming solutions. Ethereum now leverages Merkle trees to improve transaction hashing and scalability [108].

The pros and cons of Ethereum are depicted in Table VII.

TABLE VII. ADVANTAGES AND DRAWBACKS OF ETHEREUM

Advantages	Drawbacks
Capitalizes on blockchain	Slow transaction processing
With a $9.7 billion market worth, it's backed by huge names such as JPMorgan Chase, Amazon, Microsoft, IBM, etc.	Centralization - The Data Access Object (DAO) attack emphasized the importance of developers' word over community votes
Daily traffic	PoS update
A powerful team	The downside of being the first mover
Reliable	Scams in the marketplace

3) Tezos - (programming language: OCaml): It is an innovative contract platform created in 2014. In Leased Proof-of-Stake (LPoS), a participant makes a new block, which thirty-two other participants accept. Tezos has certain unique features that set it apart from other platforms:

- Formal verification of smart contract code.

- It is possible to upgrade the network without branching (fork).

The pros and cons of Tezos are depicted in Table VIII.

TABLE VIII. ADVANTAGES AND DRAWBACKS OF TEZOS

Advantages	Drawbacks
Community Governance (The protocol rewards its community for conducting needed or desired improvements)	In-Fighting (After its successful ICO, Tezos experienced a lot of conflict between the Tezos Foundation and the Breitmans' company in the US)
Supportive Community	Token Issuance Delays
Formal Verification	Lawsuits

4) Hyperledger (Sawtooth, Fabric, Burrow, Indy, Iroha, Cello): It is not a coin, a blockchain, or a firm, but a complex project of the Linux Foundation (Fabric, Sawtooth, and Burrow). It is a kind of open-source platform for developing enterprise blockchain applications which began in 2015. The pros and cons of Hyperledger are depicted in Table IX.

TABLE IX. ADVANTAGES AND DRAWBACKS OF HYPERLEDGER

Advantages	Drawbacks
Modularity in architecture	Intricate fabric architecture
Performance optimization	A shortage of verified use cases
Hybrid model	Fault tolerance in the network
Membership with permission	A lack of expertise among programmers
Query capability resembling SQL	

5) Hedera Hashgraph: The platform is still in the early stages of development. Histogram consensus is used by the Hedera PoS open network to ensure maximum security (ABFT) while consuming little bandwidth. The pros and cons of the Hedera Hashgraph are depicted in Table X.

TABLE X. ADVANTAGES AND DRAWBACKS OF HEDERA HASHGRAPH

Advantages	Drawbacks
Highly secure	Not open-sourced but patented
Faster: Transactions being processed in parallel	There are only 19 governors, which brings decentralization into question
The platform supports the same object-orient programmed language, Solidity, which is used for smart contracts	Lack of proven use cases

6) Ripple - (programming language: C++): Its unique feature is the absence of blockchain. Instead, it "runs a network of independent checking nodes." People, banks, organizations, and states can check nodes. A distributional

registry is created every second, and XRP is the cryptocurrency token. The pros and cons of Ripple are depicted in Table XI.

TABLE XI.   ADVANTAGES AND DRAWBACKS OF RIPPLE

Advantages	Drawbacks
Transactions with XRP are quick and inexpensive	The consensus protocol appears to be potentially less secure
The use of Ripple's payment network has already begun in financial institutions	Numerous Ripple's financial partners exclusively use RippleNet, not the company's XRP cryptocurrency
Small company owners and consumers can utilize XRP to conduct safe money transactions	Ripple has sparked controversy due to its private ownership and the Securities and Exchange Commission (SEC) lawsuit
Currency transfers on a global scale are possible	Purchasing XRP is challenging in the United States

7) Quorum - (programming language: Java): It is an Ethereum fork that maximizes transaction and contract anonymity in banking and related fields. JP Morgan created it to solve critical financial issues with smart contracts and distributional registries. It facilitates institutional volume deals and can restrict access to transaction history while maintaining system openness. When it comes to transaction verification and controlling communication between nodes, they rely on the BFT and Raft transaction algorithms to keep operations private. This solution's flaw is centralization. The pros and cons of Quorum are depicted in Table XII.

TABLE XII.   ADVANTAGES AND DRAWBACKS OF QUORUM

Advantages	Drawbacks
Exceptional performance	Framework features are only partially utilized
Consensus mechanisms based on voting	Message overheads
Privacy of transactions and contracts has been improved	Multiple "chosen" nodes confirm transactions.
Blockchain with permission	Anonymit.
	Block size restrictions

8) Corda - (programming language: Kotlin): This is a private distribution platform using Java Virtual Machine (JVM) based smart contract. Corda was created by R3 (R3CEV LLC) to record, monitor, and synchronize financial agreements. For this, a new consensus mechanism was developed, which checks and signs contracts using notarial nodes. Nodes confirming valid interest in the transaction's assets can access the data. The pros and cons of Corda are depicted in Table XIII.

TABLE XIII.   ADVANTAGES AND DRAWBACKS OF CORDA

Advantages	Drawbacks
The ability to reach a consensus on individual agreements and contracts	Using oracles (humans) to verify information and papers decreases their credibility
Restricted access	Financial sector-specific
Based on the R3 industry standard	Only financial sector users

9) Electro-Optical System (EOS) - (programming language: C++): EOS is a platform for decentralized applications; Platform rights are shared according to interest. Buying 20% of EOS blockchain firm tokens gives you 20% of the project's revenues, property, copyright, and reputation. The pros and cons of EOS are depicted in Table XIV.

TABLE XIV.   ADVANTAGES AND DRAWBACKS OF EOS

Advantages	Drawbacks
Throughput is 1200 transactions per second	There isn't even a Graphical User Interface (GUI) wallet for this project, which is still in development
New crowdfunding model	Crowdfunding conditions might frighten DApp developers
The delegate PoS consensus algorithm has a high scalability potential	The EOS team is slow to report to investors
DApps toolbox for developers	Cryptocurrency is the most criticized

10) Stellar Smart Contracts (SSC) - (programming language: JavaScript, Python, Golang, PHP): The smart contracts used are not Ethereum smart contracts. It is a transactional turing, not a thorough one. SSC can be written in any Stellar language. The pros and cons of SSC are depicted in Table XV.

TABLE XV.   ADVANTAGES AND DRAWBACKS OF SSC

Advantages	Drawbacks
Fast Transaction Speed	Less Literature
Multi-currency Support	Smart Contracts issues
Extremely Low Transaction Fees	

11) iOlite - (programming language: iOlite): To create smart contracts using natural language, iOlite provides an easy-to-use engine that understands natural language and compiles smart contract code. Regarding people who do not want to spend time learning, iOlite is a great option - it creates smart contracts. The pros and cons of iOlite are depicted in Table XVI.

TABLE XVI.   ADVANTAGES AND DRAWBACKS OF IOLITE

Advantages	Drawbacks
No battery life restrictions or need to recharge	iOlite Vaporizer is more significant than other portable models
iOlite makes it very convenient for use on the go	The vaporizer can become too hot during long vaping sessions
A vaporizer is very compact	Vapor is less dense than other portables
Stylish design that comes in many colors	It does not diminish the effectiveness of the vaporizer
The vapor produced by the model has a fresh taste	

12) Neblio: Incorporating blockchain into existing businesses is the goal of Neblio, which has an easy-to-use API in eight popular programming languages. Neblio's primary

goal is to help established enterprises use blockchain. The Neblio blockchain technology should be accessible to developers. The pros and cons of Neblio are depicted in Table XVII.

TABLE XVII. ADVANTAGES AND DRAWBACKS OF NEBLIO

Advantages	Drawbacks
Easy-to-use	Use of private keys, which are questionable at times
Single token protocol for unique tokens	Immutable Data storage means there is no going back
Secured from third-party intervention	
Users can code in eight different programming languages	

13)Lisk - (programming language: JavaScript): This flexible blockchain technology creates a simple user interface and platform for everyone to use. Lisk begins each app on its independent side chain, the development of which can be complex. The pros and cons of Lisk are depicted in Table XVIII.

TABLE XVIII.ADVANTAGES AND DRAWBACKS OF LISK

Advantages	Drawbacks
Lisk is a decentralized DApp development platform that makes the creation of apps easy and readily accessible	Lisk is in direct competition with the leading smart contract and DApp platform in the world, Ethereum
Lisk makes use of advanced Software development kit (SDK) technology, allowing developers to produce DApps and separate blockchain	
Lisk and the process involved with creating DApps make the entire platform truly decentralized and independent	

14)Dragonchain - (programming language: Java, Python, Node, C#, Go): Disney created Dragonchain in 2014. After a year of development, the firm released the first open-source code in 2016. DragonChain provides clients with "ecosystems" without central servers. The platform's multilayer security architecture aims to carve a unique position in the cryptocurrency market. The pros and cons of Dragonchain are depicted in Table XIX.

TABLE XIX. ADVANTAGES AND DRAWBACKS OF DRAGONCHAIN

Advantages	Drawbacks
Regulated platform	Poor market performance
Great functionality	
Ease of Use	

15)NEO - (programming language: Python, JavaScript, Java, C#, Go): It is an open-source platform that uses the consensus technique DBFT (DBFT) and smart contracts and cross-platform capabilities. Smart Contracts can access external resources through Oracle's built-in component. The pros and cons of NEO are depicted in Table XX.

TABLE XX. ADVANTAGES AND DRAWBACKS OF NEO

Advantages	Drawbacks
Energy-Efficient Consensus Mechanism	Unpopularity among Westerners
Ease of Developers	To compete with Ethereum, NEO has to outperform it in every way
NEO can handle around 1000 transactions per second	

16)IOTA: IOTA is a platform designed specifically for the IoT. It is a global network of connected devices. In terms of construction, IOTA is dissimilar to Ethereum, bitcoin, and other famous blockchains because it lacks the traditional linear structure of blockchain. Such a structure limits network scalability, so IOTA uses Tangle instead. A user must confirm two other users' transactions to gain confirmation for a transaction. The pros and cons of IOTA are depicted in Table XXI.

TABLE XXI. ADVANTAGES AND DRAWBACKS OF IOTA

Advantages	Drawbacks
Micro-payments	Ternary logic
Lightweight	IoT-specific framework functions
Scalable	The code has cryptic regions
Quantum-secure	No smart contract support

17)Cosmos SDK/Tendermint - (programming language: Golang): A modular execution stack at the system's heart lets apps mix and match components as needed. All modules are also sandboxed for increased application security. Consensus algorithms include PoS and BFT. The pros and cons of Cosmos SDK/Tendermint are depicted in Table XXII.

TABLE XXII. ADVANTAGES AND DRAWBACKS OF COSMOS SDK/TENDERMINT

Advantages	Drawbacks
Fully open-source project	Delegated staking freezes ATOM while validators work
Cosmos protocol allows separate blockchains to connect effortlessly	Staked ATOM are locked in a minimum of 3-weeks
ATOM can be staked to receive interest	
The interoperability does not compromise the security that Cosmos offers	
Cosmos is a community governance system that uses keys to identify users	

18)Waves: It is a Russian platform for ICO, crowdfunding, exchanges, and payment gateways. It is a software platform that includes many useful utilities and tools for developers. Users pay for the apps with the Waves tokens. The pros and cons of Waves are depicted in Table XXIII.

TABLE XXIII. ADVANTAGES AND DRAWBACKS OF WAVES

Advantages	Drawbacks
Perfect for crowdsourcing	Unclear legal status
Fair transfers	Support issue
Scalable, fast, and low-cost	Performance issue
Extremely easily accessible	Exchange is not yet mature

19)NEM (XEM): This cryptocurrency may be used to build trade, philanthropic, and banking applications. The Proof-of-Importance (PoI) consensus generation approach distinguishes this framework from others. PoI grants block development privileges to participants with the best reputation to ensure system integrity. The pros and cons of NEM are depicted in Table XXIV.

TABLE XXIV. ADVANTAGES AND DRAWBACKS OF NEM

Advantages	Drawbacks
Less energy usage	Low NEM consumer
Good capacity	Weak debut
Low commission	PoI centralized
Anti-counterfeiting	Fewer pieces of literature

20)OpenChain - (programming language: C#): It is a general ownership register that anyone may use. It may be modeled to work with virtually any use case. This application leverages partitioned consensus. There is no miner, and transactions are free and quick. Smart Contracts are autonomous actors that receive and send transactions based on business logic. The amount of privacy is adjustable from transparent to private. It suits enterprises that want to manage and issue digital assets reliably, securely, and scalable.

21)Multichain - (programming language: C++): It is a private P2P network that uses enhanced blockchain technology for financial operations. The pros and cons of Multichain are depicted in Table XXV.

TABLE XXV. ADVANTAGES AND DRAWBACKS OF MULTICHAIN

Advantages	Drawbacks
Flexible asset metadata	No smart contracts support
Permitted follow-up issue	Since they are 100% centralized, Private blockchains are beneficial as sandboxes, but not for production as they are meant only for a specific task
Multi-asset atomic payments	
Multi-way atomic asset swaps	
Multi-signature security and escrow	
Easy to setup and configure	

22)Monax - (programming language: Go): It is an open-source framework enabling business ecosystem developers to create, send, and release blockchain-based applications. The pros and cons of Monax are depicted in Table XXVI.

TABLE XXVI. ADVANTAGES AND DRAWBACKS OF MONAX

Advantages	Drawbacks
Only Monax's contract management platform can notarize electronic signatures on the blockchain	Legal Issues
Simple integration	User acceptance
Simple, smart, and Secure	
A database of digital contracts to track your commitments in real-time	

23)NodesPlus – (programming language: C++): This Russian project uses a programmed consensus to guard against server-side assaults and assure block authenticity. It also features built-in web applications for data, safes, and software protection. The framework includes a built-in smart contract system based on JavaScript and multi-currency wallets and payments. The pros and cons of Nodesplus are depicted in Table XXVII.

TABLE XXVII. ADVANTAGES AND DRAWBACKS OF NODESPLUS

Advantages	Drawbacks
When authorized nodes have access to certain transactions, the nodes Plus software platform allows you to deploy data clustering technologies	Updates and patches
Provides the strictly irreversible transaction required by government organizations, thus equating electronic data with paper documents	User acceptance

24)Blockchain Service Network (BSN) - (programming language: DAML): The China State Data Center created the BSN, unveiled in October 2019 as a blockchain platform for individuals and small businesses. It provides its developers with tools to construct blockchain applications in major business networks to assist the digital economy and smart city efforts. The pros and cons of BSN are depicted in Table XXVIII.

TABLE XXVIII. ADVANTAGES AND DRAWBACKS OF BSN

Advantages	Drawbacks
Reduced development, deployment, and operation costs	However, faith in Chinese technology has been eroding over the last year, with many countries opposing China's dominance of the 5G rollout
Access to blockchain apps made simpler	User privacy issues; Currently, the BSN Dev Alliance has no control over user activity
Access modes are flexible	
Expanding rapidly	

25)Exonum - (programming language: Rust): It is a blockchain platform developed by the Bitfury Group specifically for enterprise projects. The source code for Exonum and its application programming interfaces (APIs) are available for anybody to view. Besides providing access to the whole codebase, Exonum also provides blockchain administration tools. The pros and cons of Exonum are depicted in Table XXIX.

TABLE XXIX.   ADVANTAGES AND DRAWBACKS OF EXONUM

Advantages	Drawbacks
Developed in Rust which is among the most secure programming languages, with extensive predictable resource utilization and execution safety	Networking improvements needed
Restrict data visibility in blockchain, protecting user privacy without sacrificing security	This is a difficult time to use the network. So, even if they aren't supposed to, third parties can obtain access to blockchain data
Provides auditability	

26)Masterchain: It is a nationwide Russian blockchain network launched in 2016. It uses the Ethereum blockchain network as its foundation, but it is created with respect to Russian encryption requirements, user identification procedures, and secure scalability, among other things. The pros and cons of Masterchain are depicted in Table XXX.

TABLE XXX. ADVANTAGES AND DRAWBACKS OF MASTERCHAIN

Advantages	Drawbacks
Enables "prompt confirmation of data actuality" to a transacting customer	Closed network, Controlled by the security monitor
Instant communication	Less decentralized than blockchain

V.   EMERGING APPLICATIONS OF BLOCKCHAIN

Blockchain has been suggested to be used in many applications and cases following the successful application of technology in bitcoin because of its distinguishing characteristics. The following section provides a high-level summary of each domain, shown in Fig. 12.

A. Financial Applications

Blockchain technology is now being applied in several financial sectors, including prediction markets, asset settlement, economic transactions, and business services [113]. Blockchain is expected to be essential to the global economy's long-term growth, providing benefits to the current financial system, consumers, and society [42]. The adoption of blockchain technology in financial services can provide various benefits that can help alter the financial industry. As Klynveld Peat Marwick Goerdele (KPMG) reported, blockchain technology may minimize errors by up to 95%, enhance efficiency by 40%, and cut capital expenditure by up to 75% [84]. Blockchain technology in finance is a revolutionary idea that might transform the financial services industry. Here are a few examples of blockchain financial services applications that are gaining popularity in the industry:
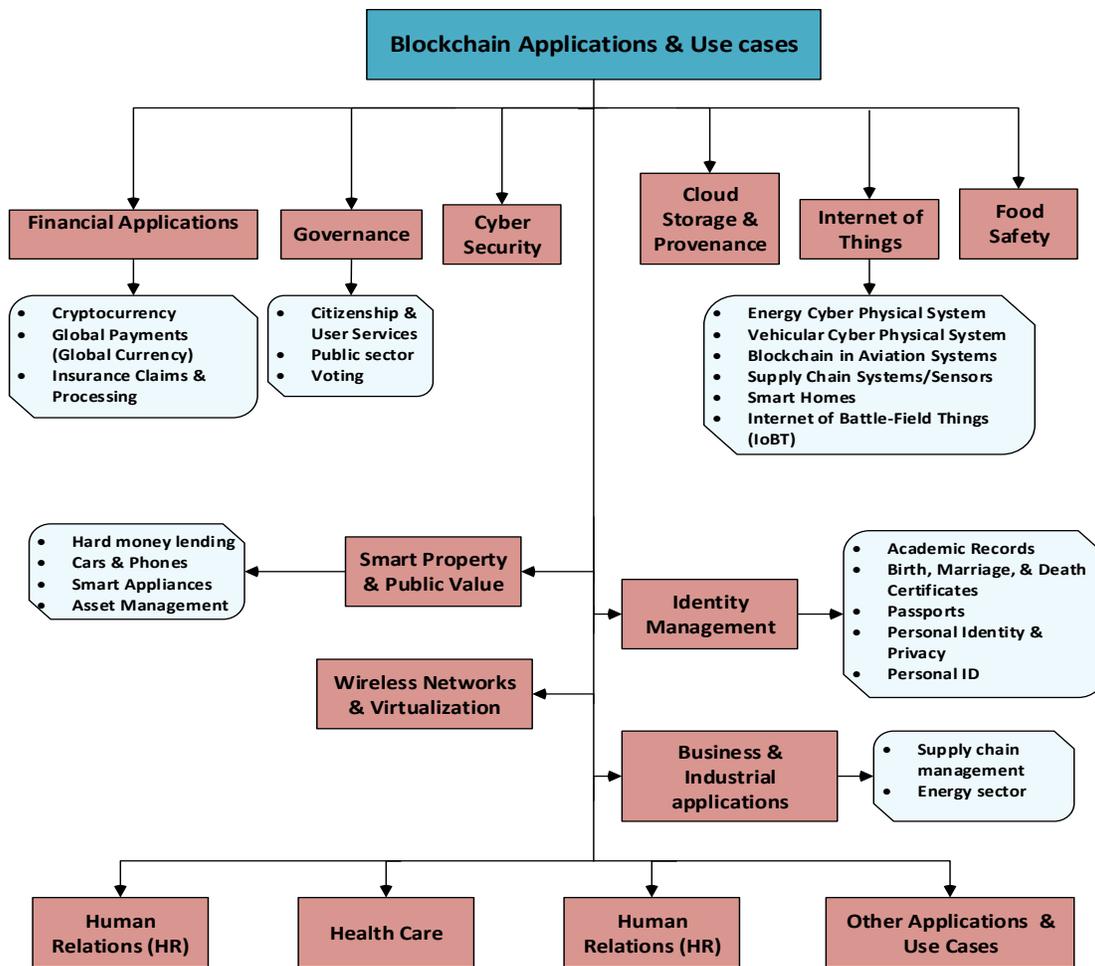

Fig. 12. Blockchain Applications.

1) Cryptocurrency: The earliest application of blockchain technology was to record digital currency transactions, and it remains the most popular usage to date. Bitcoin was the first widely adopted cryptocurrency [85]. Since then, the numerous versions of blockchain technologies have resulted in a wide range of cryptocurrencies and blockchain applications in markets and economics. Hundreds of cryptocurrencies are presently trading, according to 2017 worldwide cryptocurrency benchmarking research [86]. While many of these cryptocurrencies are labeled as "altcoins" and are essentially copies of bitcoin or other coins with different parameter settings, a small proportion of the various offerings have made significant progress.

*2) Global payments (global currenc*y): Numerous intermediaries are engaged in verifying transactions, making international payments more complicated and time-consuming to complete. The entire process has the potential to be error-prone and expensive. In large part, these problems arise because financial transactions have been centralized; entities like banks and other financial institutions prescribe methods while also being accountable for verifying the transactions. Blockchain technology, which introduces a decentralized public ledger and a robust verification technique to validate transactions, minimizes the complications associated with these transactions. Global payments are made more verifiable, immutable, secure, and faster when made through this peer-to-peer network. It is currently being used by various remittance firms [87], like Abra and Bitspark, to provide remittance services using blockchain technology.

3) Insurance claims and processing: Numerous false claims have plagued the insurance industry. Thus, to efficiently handle an insurance claim, it is vital to have up-to-date data and policies associated with each claim, which is challenging with conventional methods. Blockchain technology allows for a speedy and secure approach. Similarly, because several participants or miners must consent to each transaction's legality, any false transactions or claims may be easily identified and destroyed. This ensures that the insurers receive the compensation as fast and efficiently as possible.

The open-end issues in this regard are as follows: Irrespective of some of the notable designs in recent times [88], there are still open-ended problems in applying blockchain to financial applications. Adopting blockchain in financial applications renders mainly bitcoin usage using different forms of cryptocurrencies without any dependencies on the centralized authorization. The advantage of such a scheme is highly reduced transaction fees and risks; however, deploying a secure blockchain architecture over a heterogeneous and integrated form of financial transactions is not free from the potential risk of attacks.

B. Governance

In the government sector, the usage of blockchain technology has been demonstrated in work [89]. The authors highlight better service quality, accessibility, and transparency. Blockchain technology is immune to internet attacks; transactions are public and cannot be changed or deleted once posted, making all data transactions open, safe, and secure. According to the authors [90][91], blockchain technology is suitable for the government since it is secure, P2P, decentralized, and accessible to anyone. Some of the use cases are Citizenship and User Services [92], Public Sector [93] [94] and voting.

The open-end issues in this regard are as follows: A closer look into the applications of governance using bitcoin, as mentioned above, reveals a higher dependency on consensus from the validating nodes and the users over the network. This is challenging to confirm as a regular network is usually populated by various unknown attackers and variable traffic. While using blockchain on such an application, it is challenging to conform to the robust model to promote voter participation based on common interests. Apart from this, the usage of bitcoin still doesn't witness any universally accepted governance model.

C. Internet of Things

The number of devices with internet access continues to rise annually. Connecting numerous devices together results in the Internet of Things. The IoT will revolutionise people's lives by making smart homes a reality. While having more and more of your devices online sounds like it would make life easier, the truth is that you run a significant risk of experiencing privacy and security breaches. When there are millions of interconnected devices, it is crucial to protect data and ensure accountability. The IoT could be secured by blockchain. Some of the use cases are Energy Cyber-Physical Systems [95], Vehicular Cyber-Physical Systems [96], unmanned aerial vehicle (UAV) [97], blockchain in Aviation Systems [98], Supply Chain Systems/Sensors: [99], Smart Homes [100], Internet of Battlefield Things (IoBT) [101].

The open-end issues in this regard are as follows: With its autonomous and decentralized capabilities, blockchain technology has the potential to resolve the privacy concerns surrounding the IoT by addressing the gaps and requirements of existing solutions. However, when used in an IoT setting, they still have some problems [102]. The primary issue is that if blockchain is used to improve IoT's scalability, it could eventually lead to centralization. When dealing with IoT networks, which typically have limited computational resources, the second issue is the need for ample processing time and power to carry out ciphering. The third issue is related to the storage concern brought on by the ledger's ever-increasing size, which necessitates storing all crucial data within the node. Finally, it is challenging to demonstrate the applicability of the current research work due to potential impending factors such as compliance and legal issues in various settings.

D. Cybersecurity

Blockchain technology also has potential applications in the realm of cybersecurity. When harmful data is shared among participants/organizations using blockchain, cyberattacks can be avoided in the future [103]. Because competitors could use it unilaterally if it contains identifying information, companies and governments are hesitant to discuss cyberattacks and threat intelligence. But with

blockchain and a private and public key pair, data can be sent anonymously (such as bitcoin [104]). This alleviates concerns about disseminating sensitive company or government data to potential competitors. Here are some unanswered questions: Blockchain has been touted for its purported security benefits, but if implemented incorrectly, it could expose significant vulnerabilities in the cyber security infrastructure. As a result, the blockchain is open to attacks that could slow down the formation of new chains, steal users' private keys, reverse transactions, and more. In addition, security researchers have been focusing on DoS attacks rather than developing a foolproof solution to prevent Sybil attacks, blockchain endpoint threats, routing attacks, or phishing.

E. Cloud Storage and Provenance

Blockchain can store and share metadata that documents the history of actions and creation, particularly data or file access. Data management is crucial for forensics and accountability [105]. For example, when users view and edit collaborative documents like Google Docs, the modifications are kept in the blockchain. The blockchain saves all revisions and changes made. Again, using blockchain and provenance, cloud storage, and processing can ensure accountability and integrity. Similarly, several people can easily track, modify, and alter data in the cloud for accountability and integrity. The open-end issues are as follows: Storage and provenance in the cloud by blockchain is not a simplified task. It involves a more significant deal of complexity. There is a dependency on adopting a highly sophisticated mathematical model in blockchain that can perform secure processing and transferring of data to the cloud storage system. Consequently, this viewpoint has a more significant imbalance between computational efficiency and security effectiveness.

F. Healthcare and Food Safety

Healthcare will likely be one of the most aggressive industries to adopt or drive blockchain technology. Blockchain, the future supply chain business paradigm, can be applied to the healthcare value chain. Everything from medical records to payments to processing and analytics will be automated, benefiting all stakeholders, from patients and consumers to administrators, providers, and healthcare organizations. Using blockchain in healthcare will help achieve [3]:

- Interoperability: A single format will be used to store and share data.

- Decentralized data storage: A single technology to handle all patient data.

- Power to patients: Patients who wish to be data owners will be able to choose with whom their health records are shared.

Over 0.6 billion (equal to one in every ten people) people worldwide become unwell every year due to eating contaminated food [106], making it one of the most pressing challenges to address. Every day, approximately 1167 individuals perish [106]. Due to the transparency provided by blockchain technology, users may receive information about food, such as its composition, origins, expiry dates, and so on,

in seconds to help avoid food fraud. For food safety, food customers will have greater control over their information, which will be accurate and transparent.

G. Smart Property and Public Value

Blockchain technology may be utilized to maintain track of all actions and property records for any entity or property, including but not limited to: land, house, stocks, autos, and other investments. The information included in the blockchain is distributed to all parties interested or involved, and it can be used to make contracts and verify them. The lost data can be quickly recovered by copying it from the network [107]. Some use cases are hard money lending, cars and phones, smart appliances, and asset management. The efficiency and cost of conducting business are increased by removing the requirement for an intermediary to validate the transactions.

H. Wireless Networks and Virtualization

Wireless networks and virtualization are two topics that have been discussed recently. The increasing expansion of IoT and cyber-physical system (CPS) applications is putting a strain on wireless networks, and several approaches have been investigated to increase network capacity and coverage [108]. By preventing double-spending, in which multiple parties sublease the same wireless resource [109], blockchain can help network service providers like Mobile Virtual Network Operators (MVNOs) reliably maintain control over their users' quality of experience [110]. The open issues in deploying blockchain in wireless networks are mainly associated with data security.

I. Identity Management

This subsection discusses many identity management applications and their supposed effects on blockchain technology. Some of the essential use cases are Academic Records [111], Birth, Marriage, And Death Certificates [112], Passports [113], and Personal Identity and Privacy [114].

The open-end issues in this regard are as follows: Irrespective of essential studies towards identity management by blockchain, there are still open-ended issues to confirm the claimed identity's uniqueness and legitimacy. Apart from this, it is also challenging to verify the legitimacy of the owner at the same time. A specific form of threat is called synthetic identity proofing, where the social identification information of one individual is combined with different identity information (e.g., address, date of birth, etc.) from another person to generate a new form of a counterfeited identity. Unfortunately, blockchain cannot mitigate such problems as its prime task is to maintain a record being a ledger and not to perform identity verification.

J. Business and Industrial Applications

Blockchain can significantly improve, optimize, and automate commercial operations [115]. It is becoming increasingly common to see IoT and blockchain-powered e-business concepts in process today. In [170], the authors propose a business model in which transactions between devices are handled using smart contracts operating within a decentralized system based on a blockchain network. Many businesses could benefit from blockchain-based solutions that

act as distributed Business Process Management systems. The blockchain could store each business process instance, and smart contracts could handle the workflow routing, automating and streamlining intra-organizational activities while lowering costs and increasing efficiency [116]. Supply chain management [3] and Energy Sector [117] are essential use cases. Blockchain is considered a decarbonization facilitator, allowing the energy sector to shift toward more decentralized energy sources [118].

The open-end issues in this regard are as follows: Despite some promising efforts, widespread implementation of blockchain technology in the energy and supply industries has yet to overcome significant obstacles. The primary concern in applying blockchain in supply chain management is determining the degree of usage of blockchain and offering higher data quality. Further, providing a secure access protocol for a legitimate user is another significant problem in using blockchain for supply chains. From the energy sector's perspective, blockchain adoption still couldn't solve problems related to power consumption and scalability.

K. Financial Human Relations (HR)

Human resources departments identify and hire the best candidates for the best jobs, train and educate them to help the company achieve its goals, and maintain a safe and pleasant workplace where employees can be creative and productive. While blockchain's immediate disruption in the capital markets is well-known, its role in transforming HR processes and the workforce is even more fascinating and significant. Companies worldwide have recognized the great promise of blockchain technology to improve human resources and recruitment operations [3][5]. Recruitment, background checks, employee data protection, and smart contracts are just a few critical use cases for blockchain in HR.

L. Other Applications and Use Cases

Blockchain technology can be used without a trusted third party or P2P transaction system for transparency, decentralization, immutability, integrity, privacy, and security. On the other hand, the technology has some restrictions, such as the significant delay produced by the consensus mechanism when many blocks are involved [3].

VI. Essential Findings of the Study

Some of the essential findings of the study are as follows:

- A more excellent way to build a highly transparent trust can be done via public blockchain, however, it lacks an efficient control system. Furthermore, its increased price volatility makes it unsuitable for financial products. Most existing schemes are in the nascent stage of development with narrowed implementation scope over the practical ground and less benchmarked models for a public blockchain.

- A better form of control can be established by adopting existing private blockchain-based schemes making them highly reliable and secure. However, its growth is limited owing to inferior incentives. In addition, the costs associated with its management and administration are pretty complex. Existing schemes of private blockchain offers better data security, but still, they suffer from higher coverage of security options.

- Although hybrid blockchain offers beneficial features for public and private blockchains, it's not affordable compared to a public blockchain. Conventional research work on hybrid blockchain is witnessed to provide a balance between security and data transmission; however, they were not testified over larger scale of network in existing scheme.

- Majority of the existing study towards blockchain technology was found to emphasize the consensus layer for improving its functionalities. However, it should be noted that the consensus layer plays a core role in authenticating transactions and incorporating decentralization and encryption. However, there is no report of a robust model where all the above three entities have been securely protected from intrusion over the consensus layer. Further, less emphasis towards the network layer and incentive layer in existing approaches is also witnessed, which could lead to the continuation of further intrusive activities.

- Although more significant research is being carried out towards improving consensus algorithms in blockchain, but these approaches suffer from low performance and weak scalability (PoW). They also have minimal fault tolerance (PBFT) and lower decentralization (DPoS) with a complex implementation process (PoS).

- Various existing review works are being carried out to exhibit the progressive work in blockchain technology. The work carried out by authors in [119] gave some essential information about the challenges and solutions of blockchain over IoT systems. However, the work is specific towards agriculture-based applications, whereas the presented review work encompasses all the internal approaches to make any application more resilient and secure. Similar environment-specific-based review work using blockchain was also carried out in [32] which draws potential conclusive remarks about the challenges of blockchain in IoT. In contrast, the presented review work focusses on discussing the strengths and weaknesses of different types of utilization and methods of blockchain over its different taxonomies. The notable review work carried out towards smart grid systems by authors in [120] and [121] furnishes some of the transforming techniques towards decentralization scheme; however, it lacks essential discussion of limitations and research gap, which is encompassed in the presented review work. The best notable review work was witnessed in manuscript framed by authors in [122] that contains almost all the potential strengths and weaknesses of a higher number of research-based schemes in blockchains. The proposed work further contributes towards upgrading this information by focusing more on the overall compact picture of blockchain, referring to all the significant works of literature published to date.

VII. OPEN CHALLENGES

Despite the rapid growth and research interest in blockchain technology, obstacles and concerns must be addressed before they can be widely adopted. Here, Fig. 13 illustrates a few blockchain concerns and challenges:

1) Privacy: It is necessary to safeguard the user's identity, location, and data from unauthorized users. The authentication of users entails using a private key to sign documents and the identification or verification of those documents utilizing the hash values compared to the public key. Based on all transactions shared between peers, anonymity cannot be guaranteed. The sharing of transactions makes it straightforward for other parties to investigate such activities and establish the genuine identities of the users involved in the transactions. While the decentralized consensus is built based on past transaction data that has been made publicly available, random users and their associated transaction data may also be made publicly available. In a similar vein, the location may be jeopardized. So, the question of confidentiality and privacy protection becomes a specific challenge.

2) Scalability: The distributed blockchain ledger includes several fields, such as events, users, and so forth. Users generate transactions, which are then appended to the blockchain as a block by a miner, forming a blockchain. This is a moderately challenging operation to do. As the application generates a significant number of transactions and related data, and as the communication infrastructure between nodes develops in size, the network's total performance will suffer. Though there are currently numerous blockchain protocols, research and development are still ongoing to create newer protocols that will address the difficulties in the technology. As a result, one of the most challenging tasks is ensuring that different protocols work together. More recent frameworks cannot scale after reaching a particular block and system size limit. As a result, the larger the blockchain, the longer it takes for the procedure to complete. Given this rapid growth in size and volume of data, scalability has emerged as a problem that must be addressed immediately.

3) Interoperability/Integration: Interoperability and integration - are important since blockchain technology is not a stand-alone application; instead, these applications are typically coordinated with diverse applications within and outside businesses. Blockchain will continue introducing new functionality to strengthen its capabilities and expand action plans. The integration of blockchain technology with current applications might be a difficult task. Specifically, interoperability and security issues are posing a significant barrier. Another concern is the numerous operating environments and platforms that may be required to interoperate to work on blockchain-based apps. Various developers may have produced old and new apps and hybrid applications utilizing multiple development processes,

programming languages, and environments. As a result, the integration technique turns out to be a significant amount of work that becomes increasingly difficult over time. Some work is underway to develop viable integration models that integrate blockchain technology into various industrial applications. Despite this, there is still work to find and establish increasingly practical and flawless solutions.

4) Standards: Because cryptocurrencies are involved, valid conclusions can impact blockchain applications immediately. As a result, it is a critical issue that cannot be ignored. Numerous nations have developed rules for regulating the use of cryptocurrencies. Still, these regulations are currently inaccessible to transactions involving these currencies in the real world due to black marketing. As a result, constructing a fully legal blockchain remains a pipe dream. Regulators like the General Data Protection Regulation (GDPR) safeguard users' data. However, the technology is not capable of complying with all applicable legislation. As a result, the information the user provides may raise concerns about security, accuracy, and privacy.

5) Software testing: Compared to standard software testing methodologies, applications produced using blockchain technology must be examined for special features. Verification and validation of smart contracts must be implemented, as well. Testing the individual components and the overall software engineering process is necessary. One of the unique issues is the testing of blockchain applications. Smart contracts provide the capability of trustworthy oracles that can be readily integrated into applications while verifying the validity of these external entities [3]. Smart contracts are becoming increasingly popular.

6) Consensus algorithms: A wide range of consensus algorithms is available, each with advantages and disadvantages [123]. Attacks such as the 51 per cent attack are feasible due to this. A new consensus approach must be adopted to meet the application's needs. Different algorithms have different types of loopholes based on the features that have been changed to distinguish them from one another. Working towards a consensus algorithm requires consideration of scalability, tolerance, performance, complexity, and energy usage. New organizations that provide innovative blockchain solutions for various enterprises and business domains are emerging. A few of these solutions have already been implemented in apps, whereas others are still in the development stage, with further research being conducted to improve them [3]. It is critical to improving the application's adaptability, usability, scalability, security, dependability, and effectiveness. This can be accomplished through a variety of methods. As already noted, it is crucial to consider the numerous challenges in developing solutions to make blockchain more helpful and easier to integrate into organizational applications.
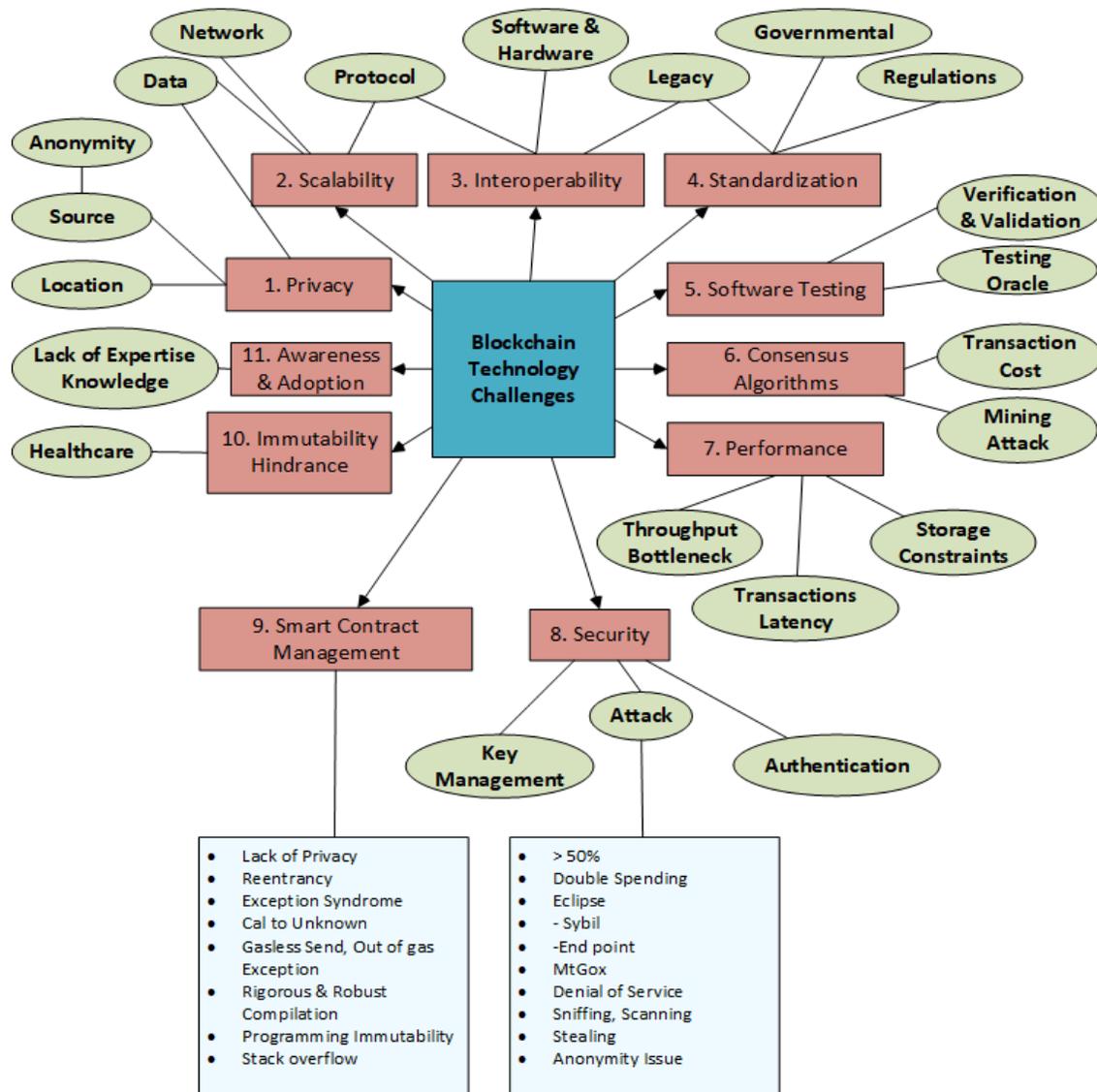
Fig. 13. Blockchain Challenges and Issues.

7) Performance issues: Blockchain systems have performance issues, like throughput bottleneck, storage constraints, and transaction latency authors in [124]. Bitcoin transactions are generally validated in an hour, which is acceptable but insufficient [9]. Blockchain-based solutions are efficient and effective when evaluated using empirical evaluation methods like experimental analysis and benchmarking [125].

8) Security: As blockchain apps become increasingly interconnected with the internet, they are subject to cyberattacks like scanning, sniffing, Sybil attacks, endpoint attacks, and DoS attacks. Blockchain miners and users control their entire system, unlike traditional methods where an external organization contains the end user. So, phishing can take the user's private keys. One example is the MtGox attack. In 2014, a bitcoin deal in Tokyo, Japan, lost $600 million. Another example is a $55 million loss versus Ether digital currencies. It is also known as a >50% attack [126]. Without adequate security measures against such attacks, all apps suffer. The idea of blockchain and its usage models is compelling. However, issues like many platforms, open access, impending developments, using open systems, and others may compromise the system's security. Existing security mechanisms help, but additional research is needed to develop more effective security models for blockchain. Attacks on key management and authentication of users and data are common.

a) Key management: A client's private key is an unrecoverable private key. However, since no central entity handles the blockchain and can track down and restore the changed data, stealing the attacker's private key is risky. Establishing reliable critical management systems in the blockchain is thus another considerable difficulty.

b) Authentication: Blockchain ensures the authenticity of a ledger and other sensitive data stored in blocks in a decentralized manner [210] by encrypting the data. To create, survey, or modify data in the blockchain, a user must first utilize a private key connected with a public key. These credentials are tied to an address in a wallet. Users can digitally sign transactions using such software keys and check their authenticity. The Elliptical Curve Digital Signature Algorithm (ECDSA) is the most often used Digital Signature Algorithm. There is still a need for a lot of studies to authenticate the information and the user. There is also the risk of a compromised wallet to be considered. Its ability to confirm identity and identity paperwork has made it a studied issue.

9) Smart contract management: Developing, deploying, and interacting with a blockchain among various members and stakeholders is a highly tough task [127]. Indeed, no matter what form of blockchain is used (public, private, or consortium), severe challenges such as management, administrative services, and smart contract debugging arise. A few of the smart contract-related concerns are as follows:

a) Privacy preservation: When building a smart contract, it is challenging to consider privacy concerns. Given the inclusion of bytecode analysis tools, the agreement faces the risk of being insecure [7][8].

b) Re-entrancy: The DAO assault extensively used re-entrance to achieve success. It impacts the execution of functions, says A and B; in most cases, A is the first to execute, followed by B. However, in this circumstance, B executes first, followed by A. Consequently, the situation deteriorates, resulting in a loss. A function is, for example, the "withdraw" function, which retrieves cryptocurrency from a record by giving cryptocurrency to a user and then "updating" the account balance. If the value of the currencies is less, the "update" should occur before the "withdraw".

c) Call to the unknown: Some Solidity functions can call the callee's fallback function when the callee's signature does not match any available users in a Solidity contract. When an attacker executes the malicious fallback function identified in their contract, they can use the call, transmit, or delegate call function to send the Ether to an address that does not exist [128]. Unlike other attacks on Ethereum, this one posed a high risk.

d) Exception syndrome: Incorrect handling of exceptions causes a syndrome that must be dealt with and a loss of currency. These occur due to an unintentional halt in the execution, and the gas quantity is deducted [128]. If an interruption happens, how to deal with it becomes a serious concern.

e) Gas exception: When a transaction is given, the "gas" is a measure of Ether used to cover the transaction's cost. If an exchange runs out of gas while being executed, it will be rescheduled; nevertheless, the sender will be obligated to pay the miner the total amount of available gas [128].

f) Programming smart contracts: Because smart contracts are immutable, security is one of the most challenging hurdles to overcome while programming them. If Ether is taken from users, it will be impossible to recover it. The other issue is data integrity and uniqueness, which are required for blockchain-based systems to be reliable and trustworthy. The economy is also a key source of concern. This is the cost of creating, executing, and deploying smart contracts, also known as the gas cost in Ethereum [128]. A single programming error will harm the economy as a whole.

10)Immutability hindrance: The immutability attribute of blockchain may make it challenging to employ the technology in some applications. For example, when used in healthcare area, the immutability feature may make it more challenging to comply with privacy rules, establishing an individual's right to have their health data erased and rendered inaccessible to third parties [129]. This is a delicate subject, and the use of technology in healthcare cannot proceed without first addressing this legal requirement [129].

11)Awareness and adoption: One of blockchain's most critical challenges is the lack of understanding and adoption. For example, many individuals are unfamiliar with how it operates. The technology's future development depends on how many parties use it, which is still an open subject.

a) Lack of expertise knowledge: A blockchain project's implementation and management are challenging tasks. The firm must have extensive expertise in the subject matter to complete the process. They must engage a large number of professionals in the blockchain sector, which creates difficulty, and as a result, it is considered one of the blockchain's downsides, according to some. Additionally, businesses must train their employees on using the technology by ensuring that the management team knows the intricacies and outcomes of running a blockchain-powered organization. They will be able to grasp client requirements and assist them in transforming their business operations to use blockchain. Not to mention that if you locate its developers and specialists, they will be more challenging to come by and more expensive than regular developers due to the demand and supply ratios. Table XXXI summarises key traits, literature references, and concerns.

TABLE XXXI.   CHARACTERISTICS OF A UNIFIED BLOCKCHAIN AND RELATED CONCERNS

	Mapped Literary Jargons	Related Concerns
Decentralization	Decentralization [33], [35], [128]	Complex security management Significant resource requirements Resource inefficiency Out-of-date and long chains are pruned out Poor performance and latency issues Scalability Inefficient use of energy
Transparency	Transparency [33], [35], [128] Openness [131] Auditability [32]	Privacy concerns Opposite of anonymity
Autonomy	Anonymity [130] Trust-Free [133] Trustlessness [131]	Few large firms rely on computing power to make decisions
Security	Security [32], [131], [18]	Various attacks
Immutability	Immutability [32], [35], [130][133] Unforgeable [128] Persistency [130] Untamperability [131]	Irreversible smart contract bugs Difficulty applying smart contracts patches Obstacles in some applications, like patient privacy in healthcare
Anonymity	Anonymity [22][131], [128] Pseudonymity [133] Transactional Privacy [32]	Lack of transparency about real transaction participants
Democratized	Democratized [33] Synchronized through Consensus [35] Collective Maintainability [131]	Low performance owing to an excessive number of decision nodes Few large firms use computational power to make decisions
Integrity	Integrity [21] Data Reliability & Integrity [32][133] Reliable Database [131]	Various types of attacks
Programmability	Programmability [23]-[132], [133] Open Source [130] Blockchain -Based Control [35] Openness [128]	Poor user interface
Fault Tolerance	Fault Tolerance [32]	Data duplication at several nodes necessitates substantial storage Synchronization Overhead
Automatic	Automatic [133] Independence [128]	Irreversible smart contract bugs Difficulty updating smart contracts Computer software is susceptible to hacking and attacks

VIII. FUTURE RESEARCH OPPORTUNITIES

Blockchain technology can completely transform the way businesses and payments are conducted worldwide without the need for trusted intermediaries or consideration of geographical limits. As a result, many parts of the blockchain will continue to be popular study areas, including consensus methods and managed services in terms of efficiency. The following section discusses some of the most important discoveries and future directions.

- Blockchain technology must address many existing challenges to become a cornerstone technology for various domains in the future. As a starting point, it should be scalable and overcome the restrictions of high latency, low throughput, and increasing storage requirements. The research community should work around the challenge of efficiently updating smart contracts with minimal overhead. Finally, the blockchain community must handle the enormous energy consumption of many nodes taking part in the consensus process, which could become a primary climate concern.

- To assist enterprise blockchain, Intel and Microsoft have already joined forces [134]. According to the alliance, the success of its enterprise will be determined by how well it addresses concerns such as performance, confidentiality, and governance [135]. Anomaly detection frameworks (ADFs) should fundamentally incorporate support for these issues if they are to be widely accepted across a wide range of business areas. The combined capabilities of blockchain and artificial intelligence offer enormous potential for developing applications in various fields. For example, the trust in blockchain technology and the decision-making capabilities of artificial intelligence in healthcare and driverless vehicles will be an excellent fit for special applications [136]. Therefore, it will be necessary for future ADFs to include intrinsic support for these types of functionalities.

- There will be two significant hurdles in the future for promoting blockchain security. The first is to strike a balance between an individual's privacy, security, and accountability, which is made feasible primarily

through Distributed Ledger Technology. On the other hand, it is necessary to cope with the privacy and security problems arising from the IoT. These issues include legal challenges, interoperability, developmental concerns, rights concerns, regulatory concerns, a lack of standards, and issues related to the emerging IoT economy, among other things.

- It is noted that blockchain is intended to operate as a decentralized system. Meanwhile, there is a growing trend toward miners being centralized within the mining pool. To date, 51 percent of the total hash power on the bitcoin blockchain is controlled by the top five mining pools combine [135][136]. Aside from that, the selfish mining method [137] shows that pools with more than 25 percent of the overall processing capacity might generate more money than their fair share of the revenue. The selfish pool's total power might fast approach 51 percent of the full power available. A solution should be given since the blockchain is not built for a few companies.

- Blockchain's economic benefits have been thoroughly researched [3][8]. Individual businesses must comprehend how the technology affects their organizational structure, method of operation, and management approach. The market must assess whether blockchain can address market failures caused by information asymmetry and improve market efficiency and societal welfare. However, more incredible research is required to understand how it affects business and market efficiency.

- Research on throughput, latency, size and bandwidth, hard forks, versioning, and many forks is also required. This is a significant research gap that will require further research. Understandably, given the tiny size of current blockchain applications, they are not the most exciting research areas for now. Currently, bitcoin, the most popular solution, has a much lower transaction volume than VISA. It is necessary to perform additional studies on scalability in the future when blockchain solutions are utilized by tens of millions of users and transaction volumes skyrocket.

- The other research gap is usability. This study found publications that explored usability from the user's perspective, but not the developer's, as the work [138] suggests. For example, the bitcoin API is still challenging to use. This has to be researched and improved as it could lead to more bitcoin applications and solutions. Future research will likely focus on bitcoin, other cryptocurrencies, and viable blockchain uses. This study identified some research on smart contracts, licensing, IoT, and smart properties in a blockchain setting. These studies will have a considerable impact in the future and may even be more exciting than cryptocurrencies. For example, using a decentralized environment to share a virtual property might alter how corporations sell their items. Considering this, it can be confidently posited that

further adoption of this technology by industry and academics will result in the significant new research.

- Improving blockchain's Interoperability and Compatibility with Existing Health Information Technology (HIT) infrastructure: Companies must know how to connect their HIT blockchain to other blockchains or non-blockchain platforms. Interoperability and data standards should be explored besides ascertaining effective integration governance mechanisms. Researchers should look at cross-authentication (for interoperable blockchains), oracles (which send external data to the blockchain for on-chain use), and application programming interfaces (APIs) (for incompatible blockchains). Many open standards exist for connecting proprietary systems to blockchain HIT, but the feasibility of using them to connect proprietary systems to blockchain HIT is still unknown [139]. Diverse stakeholders in blockchain silos cause complexity and interoperability issues that should be investigated. Security and privacy concerns require knowledge of blockchain interoperability's legal and regulatory consequences.

- HIT researchers are growing in awareness of the technological constraints of blockchain data storage and are exploring alternate techniques to comply with GDPR. Keeping addresses, hash values, and timestamps on the blockchain while storing Protected Health Information (PHI) off-chain in the cloud or on hospital servers is advised [140]. Data storage splitting can reduce system performance. On-chain and off-chain data storage should be optimized in future studies.

- The sharing economy is frequently defined as the P2P exchange of goods and services. In the future, the new sharing economy may include enterprise sharing. As a result, interconnecting blockchains may become a trend. These interfaces will let commercial operations connect identity authentication, Soft Computing and Measurements (SCM), and payment processes. They will also enable real-time data sharing and collaboration between businesses and industries.

- The vast amount of data stored in blockchain makes it possible to do Blockchain-based big data analytics. Other forms of ample data storage (for example, tensor computing) could also be improved to efficiently store and process blockchain data to save space, make it more accessible, and provide other benefits.

- Many blockchain consensus algorithms seek to replace PoW due to its massive energy waste. These new native protocols may introduce additional security risks or be impossible to deploy in reality [141]. The blockchain consensus mechanisms present a research opportunity. Future investigations are needed to seek more robust and more secure protocols than PoW while consuming less energy besides ensuring that the proper protocol is utilized in suitable applications [142].

- The threat of quantum computers to blockchain security necessitates efficient and well-proven post-quantum digital signature systems and other relevant investigations. The affordability of quantum computers is desirable, but systems like blockchain must be protected from quantum computers. Apart from ECDSA, RSA, and Digital Signature Algorithm (DSA), alternative cryptographic methods (post-quantum cryptos) are not impacted by quantum computers. Asymmetric key cryptography (Merkle signature system) and hash-based cryptography (McEliece public key system) both have such systems (e.g., AES). These include lattice-based (e.g., NTRU public key cryptosystem) and multivariate quadratic public key methods. Research is still needed to improve post-quantum cryptosystems' usability (key size), efficiency, and confidence against quantum computer threats [143]. More research is needed on quantum channels and effective post-quantum consensus techniques [144].

- Blockchain-based Decentralization Finance, sometimes known as 'DeFi,' is a new field based on blockchain technology. It is a unique, experimental type of finance, not relying on central authority or intermediaries such as exchanges or banks. DeFi is an umbrella term for many decentralized financial services and applications based on the technology. It has numerous potential study possibilities because it is new compared to the other blockchain research subjects stated above. A surge in interest in DeFi research in the coming years is anticipated.

- The interoperability of blockchain should also be investigated. Many blockchain platforms may interoperate to improve security, usability, and efficiency. Blockchain may also be used to supplement existing systems. Many businesses want to implement it but do not want to abandon their present systems. The optimal strategy to integrate the blockchain with an organization's existing systems must be researched. It is also essential to consider how different blockchain systems might work together for mutual gain.

IX. CONCLUSION

Blockchain technology is becoming increasingly popular for conducting decentralized transactions and data management without the involvement of an intermediary. Open and decentralized, the blockchain keeps track of all transactions between all parties involved in a form that can be independently verified. Since 2008, there has been a rise in interest in bitcoin and blockchain. However, it was just the beginning then, with no frequency of research papers. The data presented in the proposed paper (with respect to the percentage of increase in research publications in blockchain technology) now shows increasing attention towards evolving with more secure blockchain technology, witnessed only in the last three years. One of the prime reasons could be the adoption of blockchain as a new alternative for securing digital copyright data, supply chain management, the financial sector, etc. As a result of its inherent decentralization, persistence, anonymity, and auditability, the technology has drawn considerable interest from both the business and academic communities. A growing number of corporations and governments worldwide using blockchain technology to enhance their services' efficiency, scalability, and security. A complete description of blockchain technology was provided in this manuscript. At the outset, it introduced blockchain technology and its key features, including its architecture. After that, some of the most common consensus algorithms in the blockchain were discussed. Different aspects of these protocols were compared and evaluated. The study covers the basics of blockchain and some more interesting new applications and uses cases. For example, blockchain technology can have a revolutionary effect on the financial sector, e-Government, and Business Process Management (BPM) besides entertainment, healthcare, insurance, law firms, the Internet of Things, trading platforms, etc. However, there are just a few examples of the technology used with these systems. As a result, this technology is unlikely to replace existing systems or applications anytime soon wholly. Nonetheless, blockchain technology may undoubtedly be used in conjunction with existing systems and may even result in the construction of new techniques in the years ahead. For this reason, additional research into the technology is needed, as it is still in the experimental stage and numerous technical and legal challenges need to be overcome. This work serves as a helpful beginning point for future research issues related to the development of blockchain applications. It will be of use to practitioners and researchers alike in their efforts. Even though there are numerous ongoing active research projects in this field, this technology is still in its infancy; as a result, future study avenues have been identified, with an emphasis on in-depth research.

REFERENCES

[1] B. Thuraisingham, "Blockchain technologies and their applications in data science and cyber security," in *2020 3rd International Conference on Smart BlockChain (SmartBlock)*, Zhengzhou, China, 2020, pp. 1–4.

[2] C. Pike, "Blockchain technology and competition policy - issues paper by the Secretariat," *SSRN Electron. J*, 2018.

[3] F. Casino, T. Dasaklis, and C. Patsakis, "A systematic literature review of Blockchain-based applications: current status, classification and open issues," *Telematics and Inform*, vol. 36, pp. 55–81, 2019.

[4] H. Sheth and J. Dattani, "Overview of Blockchain technology," *Asian J. Convergence in Technol*, vol. 5, no. 1, pp. 1–4, 2019.

[5] M. Crosby, P. N. Pattanayak, S. Verma, and V. Kalyanaraman, *BlockChain Technology: Beyond Bitcoin. Appl. Innov*, no. 02, pp. 6–19, 2019.

[6] D. Vujicic, D. Jagodic, and S. Randic, "Blockchain technology, Bitcoin and Ethereum: A brief overview," in *17th International Symposium INFOTEH-JAHORINA (INFOTEH)*, 2018, pp. 1–6.

[7] R. Zhang, R. Xue, and L. Liu, "Security and privacy on Blockchain," *ACM Comput. Surveys*, vol. 52, no. 3, pp. 1–34, 2019.

[8] Chirag, C. Impact of Blockchain on the Economy | Appinventiv. Available online: https://appinventiv.com/blog/real-impact-of-blockchain-technology-on-economy/. (accessed on 12 August 2022).

[9]   S. Bragadeesh and A. Umamakeswari, "Role of Blockchain in the Internet-of-Things (IoT)," *Int. J. of Eng. & Technol*, vol. 7, no. 9, pp. 109–112, 2018.

[10]   W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Trans. Inf. Theory*, vol. 22, no. 6, pp. 644–654, 1976.

[11]   L. Lamport, "The part-time Parliament," *ACM Trans. Comput. Syst*, vol. 16, no. 2, pp. 133–169, 1998.

[12]   G. Pareek and B. Purushothama, "Blockchain-based decentralised access control scheme for dynamic hierarchies,". *Int. J. Inf. Comput. Secur.*, vol. 16, no. 3-4, pp. 324-354, 2021.

[13]   S. Zhai, Y. Yang, J. Li, C. Qiu, and J. Zhao, "Research on the application of cryptography on the blockchain," *J. Phys.: Conf. Ser.*, vol. 1168, no. 3, pp. 1-8, 2019.

[14]   K. Nelaturu, H. Du, and D. Le, "A review of blockchain in Fintech: Taxonomy, challenges, and future directions," *Cryptography*, vol. *6, no. 2,* p. 18, 2022.

[15]   X. Li, J. Xu, X. Fan, Y. Wang, and Z. Zhang, "Puncturable signatures and applications in proof-of-stake blockchain protocols," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 3872-3885, 2020.

[16]   F. Shahid, I. Ahmad, M. Imran, and M. Shoaib, "Novel One Time Signatures (NOTS): A compact post-quantum digital signature scheme," *IEEE Access*, vol. 8, pp. 15895-15906, 2020.

[17]   Z. Cai, J. Qu, P. Liu, and J. Yu, "A blockchain smart contract based on light-weighted quantum blind signature," *IEEE Access*, vol. 7, pp. 138657-138668, 2019.

[18]   S. Zhang and J. Lee, "A group signature and authentication scheme for blockchain-based mobile-edge computing," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4557-4565, 2020.

[19]   Y. Xiao, P. Zhang, and Y. Liu, "Secure and efficient multi-signature schemes for fabric: An enterprise blockchain platform," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 1782-1794, 2021.

[20]   F. Xiong, R. Xiao, W. Ren, R. Zheng, and J. Jiang, "A key protection scheme based on secret sharing for blockchain-based construction supply chain system," *IEEE Access*, vol. 7, pp. 126773-126786, 2019.

[21]   W. Zheng, K. Wang, and F. Wang, "GAN-based key secret-sharing scheme in blockchain," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 393-404, 2020.

[22]   Y. Kim, R. Raman, Y. Kim, L. Varshney, and N. Shanbhag, "Efficient local secret sharing for distributed blockchain systems," *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 282-285, 2019.

[23]   H. Yin *et al.*, "Attribute-based private data sharing with script-driven programmable ciphertext and decentralized key management in blockchain Internet of Things," *IEEE Internet of Things J.*, vol. 9, no. 13, pp. 10625-10639, 2022.

[24]   Q. Lyu *et al.*, "JRS: A joint regulating scheme for secretly shared content based on blockchain," *IEEE Trans. Netw. Serv. Manag.*, vol. 19, no. 3, pp. 2957-2971, 2022.

[25]   A. Mohsin *et al.*, "PSO–Blockchain-based image steganography: towards a new method to secure updating and sharing COVID-19 data in decentralised hospitals intelligence architecture," *Multimed. Tools. Appl.*, vol. 80, no. 9, pp. 14137-14161, 2021.

[26]   P. Sarkar, S. Ghoshal, and M. Sarkar, "Stego-chain: A framework to mine encoded stego-block in a decentralized network," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5349-5365, 2022.

[27]   A. Giron, J. Martina, and R. Custódio, "Steganographic analysis of blockchains," *Sensors*, vol. 21, no. 12, p. 4078, 2021.

[28]   J. Shi, X. Zeng, and R. Han, "A blockchain-based decentralized public key infrastructure for information-centric networks," *Information*, vol. 13, no. 5, p. 264, 2022.

[29]   F. Hashim, K. Shuaib, and F. Sallabi, "Connected blockchain federations for sharing electronic health records," *Cryptography*, vol. 6, no. 3, p. 47, 2022.

[30]   X. Boyen, U. Herath, M. McKague, and D. Stebila, "Associative blockchain for decentralized PKI transparency. *Cryptography*, vol. 5, no. 2, p. 14, 2021.

[31]   R. Longo, C. Mascia, A. Meneghetti, G. Santilli, and G. Tognolini, "Adaptable cryptographic primitives in blockchains via smart contracts," *Cryptography*, vol. 6, no. 3, p. 32, 2022.

[32]   B. Cambou, B. *et al.*, "Securing additive manufacturing with blockchains and distributed physically unclonable functions," *Cryptography*, vol. 4, no. 2, p. 17, 2020.

[33]   M. Ferrag, M. Derdour, M. Mukherjee, A. Derhab, L. Maglaras, and H. Janicke, "Blockchain technologies for the Internet Of Things: Research issues and challenges," *IEEE Internet of Things J*, vol. 6, no. 2, pp. 2188–2204, 2019.

[34]   Y. Xinyi, Z. Yi, and Y. He, "Technical characteristics and model of Blockchain," in *10th International Conference on Communication Software and Networks (ICCSN)*, Chengdu, China, 2018, pp. 562–566.

[35]   J. Xie, H. Tang, T. Huang, F. Yu, R. Xie, J. Liu, and Y. Liu, "A survey of Blockchain technology applied to smart cities: research issues and challenges," *IEEE Commun. Surveys Tuts*, vol. 21, no. 3, pp. 2794–2830, 2019.

[36]   R. Yang, F. Yu, P. Si, Z. Yang, and Y. Zhang, "Integrated Blockchain and edge computing systems: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts*, vol. 21, no. 2, pp. 1508–1532, 2019.

[37]   G. Irving and J. Holden, How Blockchain-Timestamped Protocols Could Improve The Trustworthiness Of Medical Science. *F1000Research* 2016, *5*, 222, 2016.

[38]   E. Politou, F. Casino, E. Alepis, and C. Patsakis, "Blockchain mutability: Challenges and proposed solutions," *IEEE Trans. on Emerging Topics in Comput*, vol. 9, no. 4, pp. 1972–1986, 2021.

[39]   "New kid on the Blockchain". *New Scientist*, vol. 225, no. 3009, pp. 7–7, 2015.

[40]   "How is blockchain verifiable by public and yet anonymous?", *Quora*, 2022. [Online]. Available: https://www.quora.com/How-is-Blockchain-verifiable-by-public-and-yet-anonymous. [Accessed: 22- Jul- 2022].

[41]   X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of Blockchain systems," *Future Gener. Comput. Syst*, vol. 107, pp. 841–853, 2020.

[42]   D. Puthal, N. Malik, S. Mohanty, E. Kougianos, and C. Yang, "The Blockchain as a decentralized security framework [future directions]," *IEEE Consum. Electron. Mag*, vol. 7, no. 2, pp. 18–21, 2018.

[43]   N. Y. Lee, J. Yang, M. M. H. Onik, C. S. Kim, and Modifiable, "Public Blockchains using truncated hashing and sidechains," *IEEE Access*, vol. 7, pp. 173571–173582, 2019.

[44]   Z. Guo, Z. Gao, H. Mei, M. Zhao, and J. Yang, "Design and optimization for storage mechanism of the public Blockchain based on redundant residual number system," *IEEE Access*, vol. 7, pp. 98546– 98554, 2019.

[45]   M. Baza, N. Lasla, M. M. E. A. Mahmoud, G. Srivastava, M. Abdallah, and B-Ride, "Ride sharing with privacy-preservation, trust and fair payment atop public Blockchain," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1214–1229, 2021.

[46]   A. Asheralieva and D. Niyato, "Learning-based mobile edge computing resource management to support public Blockchain networks," *IEEE Trans. Mob. Comput*, vol. 20, no. 3, pp. 1092–1109, 2021.

[47]   N. Mohammadzadeh, S. D. Nogoorani, and J. L. Muñoz-Tapia, "Invoice factoring registration based on a public Blockchain," *IEEE Access*, vol. 9, pp. 24221–24233, 2021.

[48]   C. Cai, Y. Zheng, Y. Du, Z. Qin, and C. Wang, "Towards private, robust, and verifiable crowdsensing systems via public Blockchains," *IEEE Trans. Dependable Secure Comput*, vol. 18, no. 4, pp. 1893–1907, 2021.

[49]   Y. Bai, Q. Hu, S. H Seo, K. Kang, and J. J. Lee, "Public participation consortium Blockchain for smart city governance," *IEEE Internet of Things J*, vol. 9, no. 3, pp. 2094–2108, 2022.

[50]   H. T. Wu and C. W. Tsai, "An intelligent agriculture network security system based on private blockchains," *J. Commun. Netw*, vol. 21, no. 5, pp. 503–508, 2019.

[51]   L. Hou, X. Xu, K. Zheng, and X. Wang, "An intelligent transaction migration scheme for RAFT-based private Blockchain in Internet of Things applications," *IEEE Commun. Letters*, vol. 25, no. 8, pp. 2753–2757, 2021.

[52] K. Toyoda, K. Machi, Y. Ohtake, and A. N. Zhang, "Function-level bottleneck analysis of private Proof-of-Authority Ethereum blockchain," *IEEE Access*, vol. 8, pp. 141611–141621, 2020.

[53] L. Xu, T. Bao, and L. Zhu, "Blockchain empowered differentially private and auditable data publishing in industrial IoT," *IEEE Trans. Industr. Inform*, vol. 17, no. 11, pp. 7659–7668, 2021.

[54] D. Huang, X. Ma, and S. Zhang, "Performance analysis of the raft consensus algorithm for private Blockchains," *IEEE Trans. Syst. Man Cybern. Syst*, vol. 50, no. 1, pp. 172–181, 2020.

[55] D. Chattaraj, B. Bera, A. K. Das, J. J. P. C. Rodrigues, and Y. Park, "Designing fine-grained access control for Software-Defined Networks using private Blockchain," *IEEE Internet of Things J*, vol. 9, no. 2, pp. 1542–1559, 2022.

[56] M. J. Baucas, S. A. Gadsden, and P. Spachos, "IoT-based smart home device monitor using private Blockchain technology and localization," *IEEE Netw. Letters*, vol. 3, no. 2, pp. 52–55, 2021.

[57] C. Shah, J. King, and R. W. Wies, "Distributed ADMM using private Blockchain for power flow optimization in distribution network with coupled and mixed-integer constraints," *IEEE Access*, vol. 9, pp. 46560–46572, 2021.

[58] H. M. Kim, H. Turesson, M. Laskowski, and A. F. Bahreini, "Permissionless and permissioned, technology-focused and business needs-driven: Understanding the hybrid opportunity in Blockchain through a case study of Insolar," *IEEE Trans. Eng. Manag*, vol. 69, no. 3, pp. 776–791, 2022.

[59] N. Liu, L. Tan, L. Zhou, and Q. Chen, "Multi-party energy management of energy hub: A hybrid approach with Stackelberg Game and Blockchain," *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 5, pp. 919–928, 2020.

[60] R. Akkaoui, X. Hei, and W. Cheng, "EdgeMediChain: A hybrid edge Blockchain-based framework for health data exchange," *IEEE Access*, vol. 8, pp. 113467–113486, 2020.

[61] J. Polge, S. Ghatpande, S. Kubler, J. Robert, and Y. L. Traon, "BlockPerf: A hybrid Blockchain emulator/simulator framework," *IEEE Access*, vol. 9, pp. 107858–107872, 2021.

[62] S. Zhu, Z. Cai, H. Hu, Y. Li, W. Li, and Zkcrowd, "A hybrid Blockchain-based crowdsourcing platform," *IEEE Trans. Industr. Inform*, vol. 16, no. 6, pp. 4196–4205, 2019.

[63] G. Subramanian, A. S. Thampy, N. V. Ugwuoke, and B. Ramnani, "Crypto Pharmacy - digital medicine: A mobile application integrated with hybrid Blockchain to tackle the issues in pharma supply chain," *IEEE Open J. Comput. Soc*, vol. 2, pp. 26–37, 2021.

[64] S. Fan, H. Zhang, Y. Zeng, and W. Cai, "Hybrid Blockchain-based resource trading system for federated learning in edge computing," *IEEE Internet of Things J*, vol. 8, no. 4, pp. 2252–2264, 2021.

[65] J. Zhang, R. Tian, Y. Cao, X. Yuan, Z. Yu, X. Yan, and X. Zhang, "A hybrid model for central bank digital currency based on Blockchain," *IEEE Access*, vol. 9, pp. 53589–53601, 2021.

[66] G. Subramanian and A. S. Thampy, "Implementation of hybrid Blockchain in a pre-owned electric vehicle supply chain," *IEEE Access*, vol. 9, pp. 82435–82454, 2021.

[67] Z. Cui, F. Xue, S. Zhang, X. Cai, Y. Cao, W. Zhang, and J. Chen, "A hybrid BlockChain-based identity authentication scheme for multi-WSN," *IEEE Trans. Serv. Comput*, vol. 13, no. 2, pp. 241–251, 2020.

[68] S. Kumari and S. Farheen, "Blockchain based data security for financial transaction system," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2020, pp. 829–833.

[69] B. Mohanta, D. Jena, S. Panda, and S. Sobhanayak, "Blockchain technology: A survey on applications and security privacy challenges," *Internet of Things*, vol. 8, pp. 1–30, 2019.

[70] M. Salimitari, M. Chatterjee, and Y. Fallah, "A survey on consensus methods in Blockchain for resource-constrained IoT networks," *Internet of Things*, vol. 11, pp. 1–23, 2020.

[71] D. Romano and G. Schmid, "Beyond Bitcoin: A critical look at Blockchain-based systems," *Cryptography*, vol. 1, no. 2, pp. 1–31, 2017.

[72] N. Shi, Architectures and Frameworks for Developing and Applying Blockchain Technology, United States of America: IGI Global, 2019.

[73] S. Bamakan, A. Motavali, and A. B. Bondarti, "A survey of Blockchain consensus algorithms performance evaluation criteria," *Expert Syst. with Appl*, vol. 154, pp. 1–39, 2020.

[74] A. Kiayias, A. Russell, B. David, and R. Oliynykov, "Ouroboros: A provably secure Proof-of-Stake Blockchain protocol," in *Annual International Cryptology Conference*, 2018, pp. 357–388.

[75] M. Castro and B. Liskov, "Practical Byzantine fault tolerance," in *Proceedings of the Third Symposium on Operating Systems Design and Implementation*, New Orleans, LA, USA, 1999.

[76] D. Larimer, *DPOS Consensus Algorithm-The Missing White Paper*, Steemit, New York, USA, White Paper, 2018.

[77] A. Baliga, "Understanding Blockchain consensus models," *Persistent*, vol. 2017, no. 4, pp. 1–16, 2017.

[78] "Neo Smart Economy", *Neo.org*. [Online]. Available: https://neo.org/. [Accessed: 25- Jun- 2022].

[79] C. Decker, J. Seidel, and R. Wattenhofer, "Bitcoin meets strong consistency," in *Proceedings of the 17th International Conference on Distributed Computing and Networking (ICDCN)*, Singapore, 2016, pp. 1–10.

[80] D. Kraft, "Difficulty control for Blockchain-based consensus systems," *Peer-to-Peer Netw. and Appl.*, vol. 9, no. 2, pp. 397–413, 2015.

[81] Y. Sompolinsky and A. Zohar, "Accelerating Bitcoin's transaction processing," *Fast Money Grows on Trees, Not Chains. IACR Cryptology ePrint Archive*, vol. 2013, pp. 1–31, 2013.

[82] A. Chepurnoy, M. Larangeira, and A. Ojiganov, 2016. [Online]. Available: https://arxiv.org/pdf/1603.07926.pdf.

[83] P. Raj, A. K. Dubey, A. Kumar, and P. S. Rathore, *Blockchain, Artificial Intelligence, and the Internet of Things*, 1ˢᵗ ed., Cham, Switzerland: Springer, 2021.

[84] "Blockchain breaks new ground on climate risk and performance," KPMG. [Online]. Available: https://home.kpmg/xx/en/home/insights/2021/03/blockchain-breaks-new-ground-on-climate-risk-and-performance.html. [Accessed: 19-Mar-2022].

[85] G. Hileman and M. Rauchs, "Global cryptocurrency benchmarking study," *SSRN Electronic J*, pp. 8–113, 2017.

[86] "11 Money Transfer Companies Using Blockchain Technology", [Online]. Available: https://gomedici.com/11-money-transfer-companies-using-blockchain-technology-2/. [Accessed: 10- Jun- 2022].

[87] H. Kim and M. Laskowski, "Towards an ontology-driven Blockchain design for supply chain provenance," *Intelligent Systems in Accounting, Finance and Management*, vol. 25, no. 1, pp. 18–27, 2018.

[88] W. Dai, Y. Lv, K. K. R. Choo, Z. Liu, D. Zou, and H. Jin, "CRSA: A cryptocurrency recovery scheme based on hidden assistance relationships," *IEEE Trans. Inf. Forensics Secur*, vol. 16, pp. 4291–4305, 2021.

[89] A. Johnson, "Everledger Is Using Blockchain To Combat Fraud, Starting With Diamonds", *Futurism*, 2015. [Online]. Available: https://futurism.com/everledger-is-using-blockchain-to-combat-fraud-starting-with-diamonds. [Accessed: 25- Jul- 2022].

[90] J. Lee, "Bidaas: Blockchain based ID as a service," *IEEE Access*, vol. 6, pp. 2274–2278, 2018.

[91] W. Reijers, F. O'brolcháin, and P. Haynes, "Governance in Blockchain technologies & social contract theories," *Ledger*, vol. 1, pp. 134–151, 2016.

[92] B. Leiding and A. Norta, "Mapping requirements specifications into a formalized Blockchain-enabled authentication protocol for secured personal identity assurance," in *International Conference on Future Data and Security Engineering*, Ho Chi Minh City, Vietnam, 2017, pp. 181–196.

[93] C. Sullivan and E. Burger, "E-residency and Blockchain," *Comput. Law Secur. Rep*, vol. 33, no. 4, pp. 470–481, 2017.

[94] "Cisco Visual Networking Index: Forecast and Trends, 2017–2022", 2018. [Online]. Available: https://cyrekdigital.com/uploads/content/files/white-paper-c11-741490.pdf. [Accessed: 25- Jul- 2022].

[95] T. Alladi, V. Chamola, N. Sahu, and M. Guizani, "Applications Of Blockchain In unmanned aerial vehicles: A review," *Veh. Commun*, vol. 23, pp. 1–28, 2020.

[96] P. Sharma, S. Moon, and J. Park, "Block-VN: A distributed Blockchain based vehicular network architecture in smart city," *J. Inf. Process. Syst*, vol. 13, no. 1, pp. 184–195, 2017.

[97] D. B. Rawat and C. Bajracharya, *Vehicular Cyber Physical Systems*, Cham, Switzerland: Springer, 2016.

[98] C. Akmeemana, Blockchain Takes Off: How Distributed Ledger Technology Will Transform Airlines. Blockchain Research Institute, 2017.

[99] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, "Blockchain for IoT security and privacy: The case study of a smart home," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Kona, HI, USA, 2017, pp. 618–623.

[100] D. K. Tosh, S. Shetty, P. Foytik, L. Njilla, and C. A. Kamhoua, "Blockchain-empowered secure Internet -Of- Battlefield Things (IoBT) architecture," in *MILCOM 2018-2018 IEEE Military Commu- nications Conference (MILCOM)*, Los Angeles, CA, USA, 2018, pp. 593–598.

[101] L. Serrano, O. S. Santos, and M, "Blockchain and the decentralisation of the cybersecurity industry," *DYNA*, vol. 96, no. 3, pp. 1–4, 2021.

[102] M. A. Uddin, A. Stranieri, I. Gondal, and V. Balasubramanian, "A survey on the adoption of Blockchain in IoT: Challenges and solutions," *Blockchain: Research and Applications*, vol. 2, no. 2, pp. 1–49, 2021.

[103] "Cryptocurrency Prices, Charts and Market Capitalizations", *Coinmarketcap.com*, 2022. [Online]. Available: https://coinmarketcap.com/. [Accessed: 14- Jul- 2022].

[104] X. Liang, S. Shetty, D. Tosh, C. Kamhoua, K. Kwiat, and L. Njilla, "ProvChain: A Blockchain based data provenance architecture in cloud environment with enhanced privacy and availability," in *17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, Madrid, Spain, 2017, pp. 468–477.

[105] "IBM Food Trust - Blockchain for the world's food supply", *Ibm.com*. [Online]. Available: https://www.ibm.com/in-en/blockchain/solutions/food-trust. [Accessed: 21- Jul- 2022].

[106] A. Tanzarian, "Understanding Smart Property", *Cointelegraph*, 2014. [Online]. Available: https://cointelegraph.com/news/understanding_smart_property. [Accessed: 22- Jul- 2022].

[107] D. Rawat, A. Alshaikhi, A. Alshammari, C. Bajracharya, and M. Song, "Payoff optimization through wireless network virtualization for IoT applications: A three layer game approach," *IEEE Internet of Things J*, vol. 6, no. 2, pp. 2797–2805, 2019.

[108] D. B. Rawat, M. S. Parwez, and A. Alshammari, "Edge computing enabled resilient wireless network virtualization for Internet of Things," in *2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC)*, San Jose, CA, USA, 2017, pp. 155–162.

[109] D. B. Rawat and A. Alshaikhi, "Leveraging distributed Blockchain based scheme for wireless network virtualization with security and QoS constraints," in *2018 International Conference on Computing, Networking and Communications (ICNC)*, Maui, HI, USA, 2018, pp. 332–336.

[110] M. Sharples and J. Domingue, "The Blockchain and Kudos: A distributed system for educational record, reputation and reward," in *European Conference on Technology Enhanced Learning*, Lyon France, 2016, pp. 490–496.

[111] A. Kamišalić, M. Turkanović, S. Mrdović, M. Heričko, "A preliminary review of blockchain-based solutions in higher education," in *International Workshop on Learning Technology for Education in Cloud*, Zamora, Spain, 2019, pp. 114-124.

[112] P. Franks, "Blockchain for Identity Management: Can a Case be made to Begin at Birth? - SJSU | School of Information", *SJSU | School of Information*, 2019. [Online]. Available: https://ischool.sjsu.edu/ciri-blog/Blockchain-identity-management-can-case-be-made-begin-birth. [Accessed: 30- Jul- 2022].

[113] C. Ellis, M. Last, G. Rana, J. Scottie and J. Cross, "World-Citizenship: Globally Orientated Citizenship With Private Passport Services Using Available Cryptographic Tools", *GitHub*, 2014. [Online]. Available: https://github.com/MrChrisJ/World-Citizenship/graphs/contributors. [Accessed: 15- Jul- 2022].

[114] D. B. Rawat, L. Njilla, K. Kwiat, C. Kamhoua, and Ishare, "Blockchain based privacy-aware multi-agent information sharing games for cybersecurity," in *2018 International Conference on Computing, Net-*

[115] W. Ying, S. Jia, and W. Du, "Digital enablement of Blockchain: Evidence from HNA group," *Int. J. Inf. Manage*, vol. 39, pp. 1–4, 2018.

[116] K. Bilal et al., "A taxonomy and survey on green data center networks," Future Gener. Comput. Syst, vol. 36, pp. 189–208, 2014.

[117] E. Mengelkamp, B. Notheisen, C. Beer, D. Dauer, and C. Weinhardt, "A Blockchain-based smart grid: towards sustainable local energy markets," *Comput. Sci. Res. Dev*, vol. 33, no. 1-2, pp. 207–214, 2017.

[118] N. Chaudhry and M. M. Yousaf, "Consensus algorithms in Blockchain: Comparative analysis, challenges and opportunities," in *12th International Conference on Open Source Systems and Technologies (ICOSST)*, Lahore, Pakistan, 2018, pp. 54–63.

[119] M. A. Ferrag, L. Shu, X. Yang, A. Derhab, and L. Maglaras, "Security and privacy for green IoT-based agriculture: Review, Blockchain solutions, and challenges," *IEEE Access*, vol. 8, pp. 32031–32053, 2020.

[120] M. B. Mollah *et al.*, "Blockchain for future smart grid: A comprehensive survey," *IEEE Internet of Things J*, vol. 8, no. 1, pp. 18–43, 2021.

[121] P. Zhuang, T. Zamir, and H. Liang, "Blockchain for cybersecurity in smart grid: A comprehensive survey," *IEEE Trans. Industr. Inform*, vol. 17, no. 1, pp. 3–19, 2021.

[122] M. A. Ferrag, M. Derdour, M. Mukherjee, A. Derhab, L. Maglaras and H. Janicke, "Blockchain Technologies for the Internet of Things: Research Issues and Challenges," in IEEE Internet of Things Journal, vol. 6, no. 2, pp. 2188-2204, April 2019.

[123] Y. Sompolinsky and A. Zohar, "Secure high-rate transaction process- ing in Bitcoin," in *International Conference on Financial Cryptography and Data Security*, San Juan, Puerto Rico, 2015, pp. 507–527.

[124] C. Fan, S. Ghaemi, H. Khazaei, and P. Musilek, "Performance evaluation of Blockchain systems: A systematic survey," *IEEE Access*, vol. 8, pp. 126927–126950, 2020.

[125] K. Salah, M. Rehman, N. Nizamuddin, and A. Al-Fuqaha, "Blockchain for AI: Review and open research challenges," *IEEE Access*, vol. 7, pp. 10127–10149, 2019.

[126] M. R. Islam, M. M. Rahman, M. Mahmud, M. A. Rahman, M. H. S. Mohamad, and A. H. Embong, "A review on Blockchain security issues and challenges," in *2021 IEEE 12th Control and System Graduate Research Colloquium (ICSGRC)*, Shah Alam, Malaysia, 2021, pp. 227–232.

[127] T. Mcghin, K. Choo, C. Liu, and D. He, "Blockchain in healthcare applications: Research challenges and opportunities," *J. Netw. Comput. Appl*, vol. 135, pp. 62–75, 2019.

[128] L. Luu, D. H. Chu, H. Olickel, P. Saxena, and A. Hobor, "Making smart contracts smarter," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 254–269.

[129] M. Kuperberg, "Blockchain-based identity management: A survey from the enterprise and ecosystem perspective," *IEEE Trans. Eng. Manag*, vol. 67, no. 4, pp. 1008–1027, 2020.

[130] I. Lin and T. Liao, "A survey of Blockchain security issues and challenges," *Int. J. Netw. Secur*, vol. 19, no. 5, pp. 653–659, 2017.

[131] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of Blockchain technology: Architecture, consensus, and future trends," in *2017 IEEE International Congress on Big Data*, Honolulu, HI, USA, 2017, pp. 557–564.

[132] Q. Xia, E. Sifah, K. Asamoah, J. Gao, X. Du, M. Guizani, and Medshare, "Trust-less medical data sharing among cloud service providers via Blockchain," *IEEE Access*, vol. 5, pp. 14757–14767, 2017.

[133] Y. Lu, "Blockchain: A survey on functions, applications and open issues," *J. Ind. Integr. Manage*, vol. 3, no. 4, pp. 1–23, 2018.

[134] "Microsoft and Intel Detail the Deep-Seated Problems with Blockchain," [Online]. Available: https://www.forbes.com/sites/davidblack/2019/05/13/microsoft-and-intel-detail-the-deep-seated-problems-with-Blockchain/?sh=75da7a256b06. [Accessed: 15- Jun- 2022].

[135] I. Eyal and E. G. Sirer, "Majority is not enough: Bitcoin mining is vulnerable," in *International Conference on Financial Cryptography and Data Security*, Christ Church, Barbados, 2014, pp. 436–454.

[136] "Biggest Bitcoin mining pools 2021 | Statista", *Statista*, 2022. [Online]. Available: https://www.statista.com/statistics/731416/market-share-of-mining-pools/. [Accessed: 25- Jul- 2022].

[137] A. Roehrs, C. D. Costa, and R. D. R. Righi, "Omniphr: A distributed architecture model to integrate personal health records," *J. Biomed. Inform*, vol. 71, pp. 70–81, 2017.

[138] M. Swan, *Blockchain: Blueprint for a New Economy*, 1st ed., Beijing: O'Reilly Media, 2015.

[139] Y. Park, E. Lee, W. Na, S. Park, Y. Lee, and J. Lee, "Is Blockchain technology suitable for managing personal health records? Mixed-methods study to test feasibility," *J. Med. Internet Res*, vol. 21, no. 2, 2019.

[140] A. Shoker, "Brief announcement: sustainable blockchains through Proof of Exercise," in *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing*, Egham, United Kingdom, 2018, pp. 269–271.

[141] M. Conti, E. S. Kumar, C. Lal, and S. Ruj, "A survey on security and privacy issues of Bitcoin," *IEEE Commun. Surveys & Tutorials*, vol. 20, no. 4, pp. 3416–3452, 2018.

[142] D. J. Bernstein, "Introduction to post-quantum cryptography," in *Post-Quantum Cryptography*, 2009, pp. 1–14.

[143] E. Kiktenko *et al.*, "Quantum-Secured Blockchain," *Quantum Sci. Technol*, vol. 3, no. 3, pp. 1–8, 2018.

[144] S. Abdulhakeem and Q. Hu, "Powered by Blockchain technology, DeFi (Decentralized Finance) strives to increase financial inclusion of the unbanked by reshaping the world financial system," *Mod. Econ*, vol. 12, no. 1, pp. 1–16, 2021.

Factors Influencing the Acceptance of Online Mobile Auctions using User-Centered Agile Software Development: An Early Technology Acceptance Model

Abdallah Namoun[1], Ahmed Alrehaili[2], Ali Tufail[3], Aseel Natour[4], Yaman Husari[5], Mohammed A. Al-Sharafi[6],
Albaraa M. Alsaadi[7], Hani Almoamari[8]

Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah, 42351, Saudi Arabia[1, 2, 4, 5, 7, 8]
School of Digital Science, Universiti Brunei Darussalam BE1410, Brunei Darussalam[3]
Department of Information Systems, Azman Hashim International Business School, Universiti Teknologi Malaysia, Malaysia[6]
Department of Business Analytics, Sunway University, 5, Jalan University, Petaling Jaya, Bandar Sunway, 47500, Selangor,
Malaysia[6]

Abstract—e-Commerce is booming everywhere, and Saudi Arabia is no exception. However, the adoption and prevalence of online mobile auctions (aka m-auction) remain unsatisfying in Saudi Arabia and the MENA region. This paper uncovers the enabling factors and hindering barriers against the use of mobile auctions by online consumers. To this end, a multiphase mixed methods design is applied to acquire an in-depth understanding of online mobile bidding or auctioning attitudes and practices of the Saudi auctioneers and bidders. Initially, an interactive mobile auction app was developed by applying the principles of user-centered agile software development (UCASD) methodology, which incorporated several design iterations based on feedback from 454 real users. The mobile auction requirements were collected using a mix of research methods, including a survey, focus groups, prototyping, and user testing. The UCASD methodology positively influenced the early evidence-based adoption and use of mobile auctions in the Saudi market. Subsequently, three consecutive focus groups were conducted with another 22 participants to induce further insights regarding the antecedents impacting the intention to embrace online auctions using mobile phones. A taxonomy of requirements coupled with thematic analysis of the discussions gave rise to 13 influential factors of mobile auctions, namely risk, quality of products, trust, ubiquity, usefulness, access to valuable products, ease of use, age, social influence, monetary costs, enjoyment, past experience, and facilitating conditions. Our inductive approach resulted in an early technology acceptance model of mobile auctions. We conclude by reflecting on the challenges observed to suggest some practical guidelines to pave the way for other researchers in this promising area to carry out experimental studies to ameliorate the proposed model.

Keywords—Online auction; mobile auction; technology acceptance model; eBay; human-centered design; agile software development; factors

I. INTRODUCTION

e-Commerce revenues in Saudi Arabia have approximated 10.44 billion USD in 2022, with projections that its value will surpass 23.46 billion USD by the end of 2027 [1]. Moreover, the number of Saudi consumers who indulged in e-commerce activities reached approximately 25.6 million users in 2020, and this number is prognosticated to surpass 34.5 million users with 92.5% user penetration by 2025 [2]. Although e-commerce websites in Saudi Arabia are gaining unprecedented popularity, their actual adoption and use by consumers are still below par compared to the world. e-Commerce takes various forms and includes many activities. In this paper, we shed light on one specific type of e-commerce, that is, online auctions performed using mobile devices (i.e., m-auction) for several motifs. Firstly, online shopping using mobile phones (i.e., mobile e-commerce) is constantly increasing. A Saudi Communications and Information Technology Commission report revealed that 93% of online shoppers use their smartphones to make online purchases in Saudi Arabia [3]. Secondly, the online auction market growth rate is anticipated to increase by 7.2% in 2022 [4], with global sales of art and antiques already exceeding 26 billion USD in 2021 [2]. However, the USA, China, and the UK hold 86% of the global auction market share leaving only 14% to the rest of the world.

Hence, our main intriguing research question is straightforward. Why does Saudi Arabia, despite its recent major digital transformations, still lag in the use of online mobile auctions? We believe that our findings about online mobile bidding attitudes and practices may be generalized to similar developing countries in the Middle East and North Africa (MENA) region. Software solutions often face critical challenges that lead to their failure. In fact, a weak understanding of user views and requirements and failure to engage users throughout the design process is deemed to be among the common reasons that lead to software failure [5][6]. However, developing easy to use software applications does not always guarantee successful adoption and use by the final users. Technology acceptance models have been proposed for decades as theoretical stipulations to understand the motivations and critical factors that must be present before users would start using a new technology [7]. This is the case for Saudi Arabia and the gulf region, where online auctions in general and mobile bidding in particular are still not being used

despite the high technological penetration levels experienced in these countries.

Studies investigating the adoption of mobile commerce have recently gained momentum worldwide [8][9][10][11] and in Saudi Arabia [12][13]. For example, Alkhunaizan and Love proposed a mobile commerce acceptance model where performance expectancy, cost, and effort expectancy were perceived as significant factors for user adoption [14]. However, to the best of our knowledge, there are no studies related to mobile auction adoption in Saudi Arabia. Moreover, the literature does not offer comprehensive studies exploring the critical success factors and challenges impacting the adoption of mobile bidding by Saudi consumers. In line with the above motivations, our research offers several crucial contributions.

1) The design of an interactive mobile auction app (which we call iBid) by applying an integrated user-centered design and agile development process [15][16]. We contribute a participative methodology that shows how to capture and apply user feedback in an ongoing process to improve software artifacts until an acceptable version is reached quickly. At the end of the design process, a real mobile application is produced and introduced to Saudi consumers to gauge their acceptance and capture genuine feedback.

2) The conceptualization of the factors that impact user attitudes, behavior, and adoption of live auctions using mobile devices. Such in-depth understanding was captured through a series of qualitative studies in the form of focus groups with real users. This phase induced a solid theoretical foundation elaborating the factors and potential influence on the intention to use online mobile auctions.

We organized the rest of the paper into five sections. Section II elaborates on the concept of mobile auctions and technology acceptance models that predict shoppers' intention to bid on products. Section III presents the UCASD research methodology that was applied to create our mobile auction app. Section IV presents the key results of the iterative development process along with the m-auction requirements. Section V reports on the theoretical observations concerning the factors influencing Saudi consumers' mobile bidding perceptions and behavior. Section VI discusses the practical implications and suggests a research roadmap for online mobile auctions in MENA countries.

II. RELATED WORKS

A. e-Commerce and Mobile Auction

This research revolves around online mobile auctions (aka m-auction), which can be conceived as a form of electronic commerce (e-commerce). e-Commerce is a type of business that incorporates the activities and operations of selling and/or buying services or goods/products online [17]. Payments may also take place using dedicated payment gateways, like PayPal. Various technologies and frameworks, such as transaction processing and inventory systems, were developed to achieve seamless e-commerce. e-Commerce prevails in various shapes. Nowadays, C2C and B2C e-commerce are offered through online auction channels. For instance, eBay reported a net revenue of 10.27 billion US dollars in the year 2020 [18]. Moreover, online C2C e-commerce usage in 2021 by the European Union reached 20% of its population [19].

Mobile commerce (m-commerce) was coined in the late 90s to refer to purchasing and selling goods or services using mobile devices [20]. The pervasiveness of mobile devices enabled quick access to various services, including, but not limited to, mobile banking, mobile marketing, mobile money transfer, and mobile auctions [21]. A Delphi study involving experts has revealed that the factors leading to m-commerce success included ease of use, convenience, trust, and ubiquity (real-time) [22].

Online auctions refer to online marketplaces where sellers display their services or products for sale through bidding [23]. The selling price is usually not determined at the start of the auction, and buyers compete for a particular item, which is won by the highest bidder [24]. This e-commerce model offers various advantages, such as access to services and products irrespective of the geographic location, accessibility to a large pool of customers, and lower prices compared to other traditional marketplaces. Online mobile bidding refers to the auction activities that take place within e-marketplaces using mobile devices. In mobile auctions, customers are empowered to watch live auctions directly from their phones and bid on the displayed services and goods [25].

B. Online Auction and Mobile Bidding Acceptance

Technology acceptance literature proposes and examines information systems theories and models to predict how people adopt and eventually use new technologies [26]. A variety of technology acceptance models were proposed for different electronic services, like e-commerce [27], e-government [28], e-payment [29], m-banking [30], and mobile learning [31]. These studies augmented the original technology acceptance model (TAM) with several key success factors, including perceived trust, perceived risk, cost, security, lifestyle, and compatibility. However, research efforts in online auctions / mobile bidding are limited despite their growing penetration among e-commerce customers [32]. Among the notable studies of online auction acceptance is that of Turel et al., [33]. Results of two surveys with eBay users showed that addiction toward online auctions impacts perceived enjoyment, usefulness, and usability, which consequently influence intentions to use online auction systems. Chang identified that the use of autonomous agents for online auctions is positively correlated with well-known technology acceptance factors such as usefulness, ease of use, and enjoyment [34]. This is an example of the unconventional factors that must be considered in modern technologies.

In Saudi Arabia, the acceptance, adoption, and use of e-commerce [35], online banking [36] and mobile commerce have been explored to a limited degree. Eid showed that e-commerce intention to use is influenced by the satisfaction of customers but not trust. The satisfaction, in turn, was influenced by the quality of the user interface and the quality of information [35]. Moreover, the use of e-shopping in Saudi Arabia was influenced by enjoyment, usefulness, and personal norms [37]. However, the level of education, resistance to change, and perceived trust were found to impact customers'

readiness to adopt online banking [36]. Alkhunaizan and Love suggested that successful customer adoption of mobile commerce in Saudi Arabia is strongly linked to perceived performance expectancy and incurred costs [14]. However, social influence and perceived trust were not found to impact mobile commerce usage intention.

In light of the above, we identify several challenges and unanswered questions that motivate our research agenda in this paper. Firstly, mobile bidding is a relatively new technology still yet to be explored in the Saudi context. Secondly, previous studies are confirmatory in nature (i.e., statistical), assuming links between constructs concerning online auctions and e-commerce theories. The aim here is to build a user-centered mobile app and explore the mobile features, characteristics, and context that truly influence consumers' intention to use mobile bidding and formulate those in a preliminary theoretical model of mobile bidding adoption, which can be confirmed in a subsequent confirmatory study.

C. Significance of Mobile Auction

Consumer behavior research has generated considerable attention, especially with the abundance of modern technologies in our daily life. Online auction is spreading rapidly amongst technology enthusiasts; for example, the number of eBay active users increased dramatically from 89 million users to 138 million active users in 2022 [38], with the annual bidding sales exceeding 10.42 billion US dollars in net revenue in 2021 [2]. These figures emphasize the importance of investigating mobile auctions for Saudi Arabia's economy and e-shopping customers within the Kingdom. There are other qualifying motivations for selecting this research topic, including the arguments below.

- e-Commerce is accelerating in Saudi Arabia, with more than 25 million users [1]; hence, it is worth investigating trendy e-commerce topics, such as mobile auctions, which is a promising C2C e-commerce model.

- There are conflicting findings in the literature regarding the specific factors that encourage or inhibit the use of e-commerce technologies among customers.

- Existing adoption models were introduced for non-Arab cultures (i.e., the developed countries), thus restricting these technology acceptance models to those environments and raising serious questions about their applicability in Saudi Arabia, a culture with unique characteristics and user needs. The literature has already demonstrated that different cultures perceive technologies differently (e.g., US vs. Korean) [39]. Such findings necessitate the development of m-auction models that fit the Saudi society.

- Existing technology acceptance models cannot continue to predict the correct use and adoption of modern technologies such as mobile bidding. For example, Röcker argues that as new technologies are introduced to support use in personal and specific contexts, old factors like perceived ease of use may not accurately predict technology use [40].

To the best of our knowledge, no explicit and systematic study explored the acceptance of online mobile auctions by Saudi e-shoppers. This is the first study in this direction, and its contributions will give real practical implications for e-commerce in the gulf region.

III. THE UCASD RESEARCH METHODOLOGY

Generally, a research methodology refers to the set of procedures to be followed to investigate a particular research problem scientifically, leading to reliable and valid results [41]. Creating theoretical models that accurately predict the acceptance and use of new technology has attracted significant attention [42] and usually follows a purely quantitative approach, which applies statistical tests to establish new constructs and relationships. However, such a deductive approach is practical only when a set of critical success factors have been identified a priori.

We argue that the topic of mobile auctions in Saudi Arabia and MENA is under-explored and Saudi e-shoppers are yet to adopt and use mobile bidding. In our view, the first step necessitates developing an in-depth understanding of the factors impacting mobile auction attitudes and use. Therefore, we opted for a qualitative approach to pave the way, through inductive reasoning, to suggest a tentative model that is ready for quantitative confirmation in future studies.

Fig. 1 depicts the integrated UCASD methodology that we applied to develop a mobile auction app and collect mobile auction requirements. Our methodology combined UCD and agile development processes, with focused system refinements emerging from the iterative testing of the 'm-auction' concept with real users [43]. Salah et al., argue that this complementary integration empowers the developers to 1) create a deep understanding of user needs, 2) consider user interface (UI) design and user experience (UX) during the development process and produce software quickly while reducing costs [44]. Although combining UCD activities with agile software development during software creation is no easy task [43], UCD, if implemented correctly, can strengthen the agile values [45], as presented in the next section. Table I summarizes the research methods used and the number of participants in each study. The details of each study are provided in the following sections.
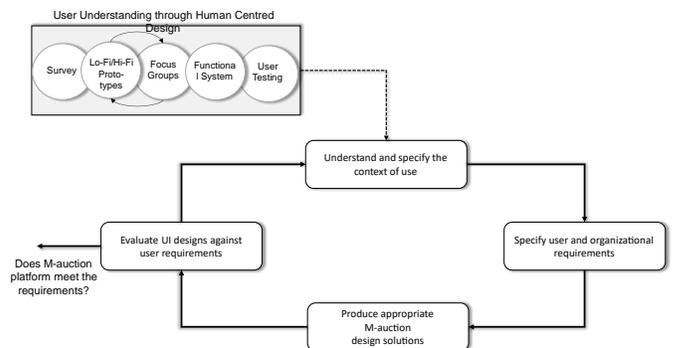

Fig. 1. Our Integrated User-Centered agile Software Development Methodology (UCASD).

TABLE I.   RESEARCH METHODS USED AND NUMBER OF USERS SOLICITED

Research Method	Type	No. of Users
Online Survey	Mixed	417
Focus Groups	Qualitative	9
Lo-Fi and Hi-Fi Prototyping	Qualitative	NA
User Testing	Mixed	28
Focus Groups	Qualitative	22

IV. THE USER-CENTERED DESIGN (UCD) OF THE IBID MOBILE APPLICATION

We argue that adopting a user-centered design philosophy is crucial to ensure various benefits ranging from rapid development, reduction in requirement changes, and high user satisfaction and acceptance. According to Wever et al., the User Centered Design (i.e., UCD) puts the final user at the center of the whole design process, where the user actively contributes his perceptions, requirements, and feedback so that a highly usable solution is created [41]. The added value of UCD in agile software development is further emphasized by Zorzetti et al., [46]. During our UCD, we captured the real requirements of bidders and introduced the mobile app for actual use. Subsequently, we held two focus groups and a user testing to discuss users' thoughts, perceptions, and reactions toward the designs of mobile auctions.

As a first contribution, this paper aims to showcase how it is possible to create a modern mobile app while applying the User-Centered Agile Software Development (UCASD) methodology. We closely followed and applied the principles suggested by Brhel et al., to realize the UCASD benefits [45]. The specific recommendations that were integrated during our development include:

- Principle One: the separation of 'product discovery' and 'product creation'.

- Principle Two: the design and development of systems through 'short, iterative, and incremental' software activities.

- Principle Three: parallel design and development tasks

- Principle Four: continuous stakeholder involvement from start to end.

- Principle Five: use tangible artifacts to facilitate communication of concepts to stakeholders.

A. Mobile Auction Discovery Through a Survey Study

Our research quest started by gathering and defining the requirements for mobile auctions. To this end, a survey was administered and distributed to university students via the university mailing lists. Surveys are cost-effective, enable access to a large pool of users quickly, and assist in capturing unique requirements [47][48][49]. In this first study, a total of 417 users responded to our online survey, which incorporated 59 quantitative and qualitative questions spanning across three different sections, namely (1) past e-shopping practices, (2) desired auction functions and features, and (3) respondents' demographic information, as shown in Fig. 2.
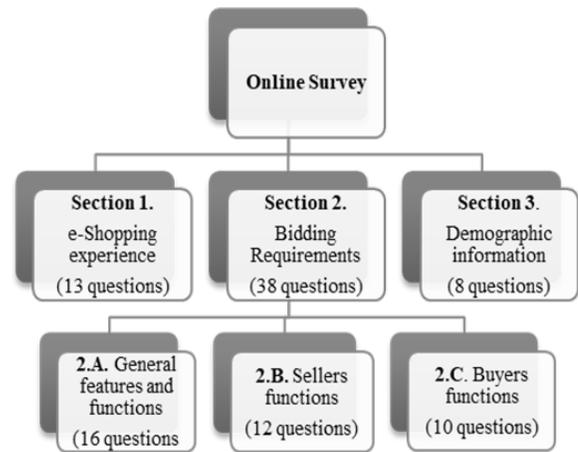

Fig. 2.   Distribution of Online Survey Questions.

The past e-shopping experiences section explored users' practices regarding e-commerce, e-shopping and online bidding that resulted in selling or buying products or services using the Internet or mobile applications. The bidding requirements section identified users' needs with respect to the general features of mobile bidding and functions supporting the seller and buyer. The requirements' questions were rated on a seven-point Likert scale (where 1=not necessary at all and 7= very important). The demographic section collected users' age, study level, mobile subscription package, smartphone brands, etc.

369 (88%) of our respondents were students, while the remaining respondents (i.e., 51 users) were professionals. 254 (60.9%) were males, and 163 (39.1%) were females. More than 75% of the respondents indicated they had an excellent experience using smartphones. Furthermore, 363 (87%) respondents in the past indulged in some sort of e-commerce activities using online websites or mobile apps. 51% of the respondents owned Apple smartphones, while the rest owned Android smartphones (e.g., Samsung and Huawei). Regarding mobile auctions, we succinctly summarize the key features and requirements desired by our respondents as follows.

- RQ1. Localization of platform: support of the local language.

- RQ2. Search functionality: availability of search functionality to look for products and services.

- RQ3. Information quality: adding photos and videos to the products or services.

- RQ4. Auction details: setting price limits for products and services on bidding.

- RQ5. Trustworthiness: displaying reviews about sellers, products, and bidding experience.

- RQ6. Live customer support via instant messaging.
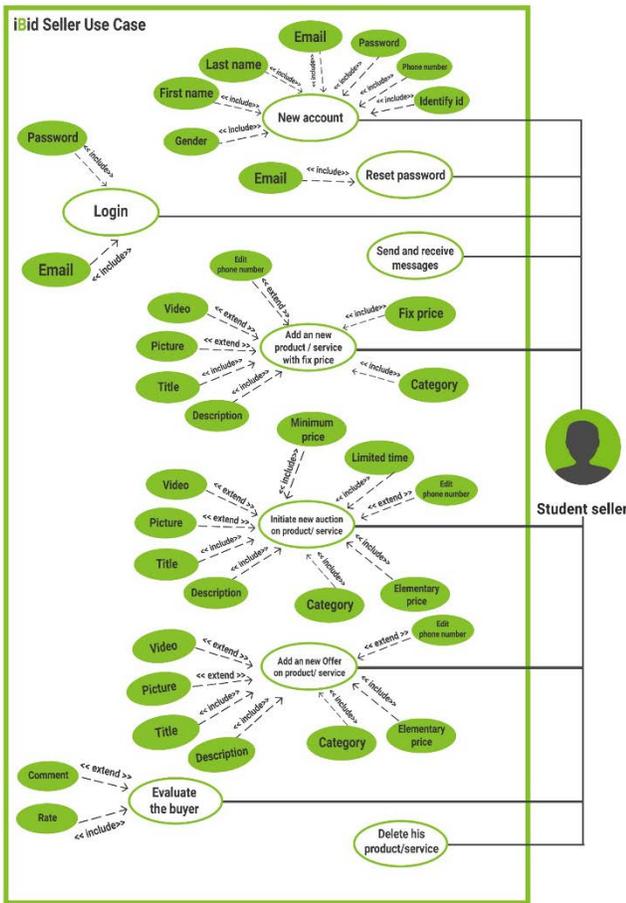
Fig. 3.   A Use Case Diagram for a Seller using the Mobile Auction System.
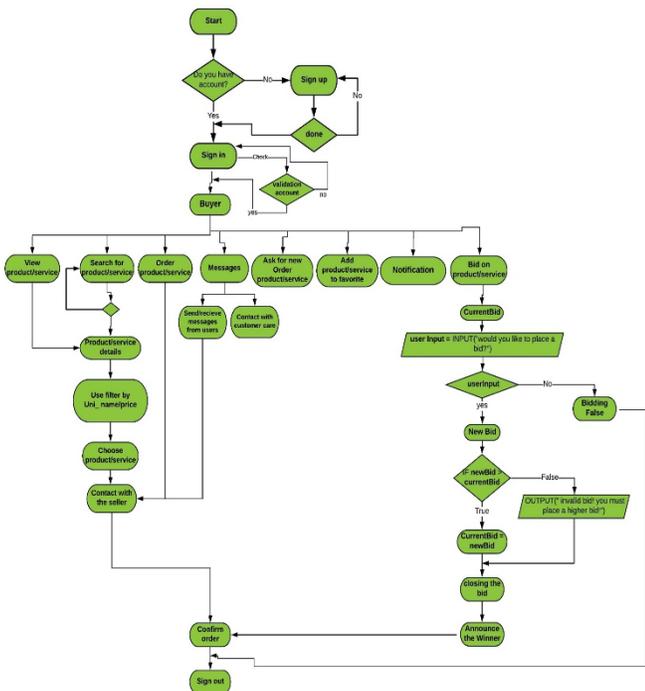

Fig. 4.   A Flow Chart of a Mobile Bidding Buyer.

We separated the concept discovery and product development tasks, as recommended by Brhel et al., [45], by carrying out an up-front analysis. The online survey helped us conceptualize the design models, such as, the use case and flow chart diagrams, of the main stakeholders of our mobile auction system. For the sake of brevity, we include only two exemplary UML diagrams derived from the user suggestions (as depicted in Fig. 3 and Fig. 4). Fig. 3 shows the use case diagram of the seller. Fig. 4 shows the flow chart of the bidder.

B. Low-Fidelity Prototypes of the Mobile Auction App

Next, we applied the second principle of effective UCASD, stipulating to design and develop the mobile app using iterative and incremental software activities. To this end, we created early prototypes of the intended mobile auction application. Low-fidelity prototyping is the process of converting product concepts and visions into simple visual representations using pen and paper [50]. The tangible benefits are well-known, ranging from cost-effectiveness, ability to inflict quick changes, and validation of requirements [51]. In this phase, we created eight low-fidelity screens to showcase the most prominent functionalities and requirements of mobile bidding, as depicted in Fig. 5. These functionalities and features were extracted from the results of an online survey study.
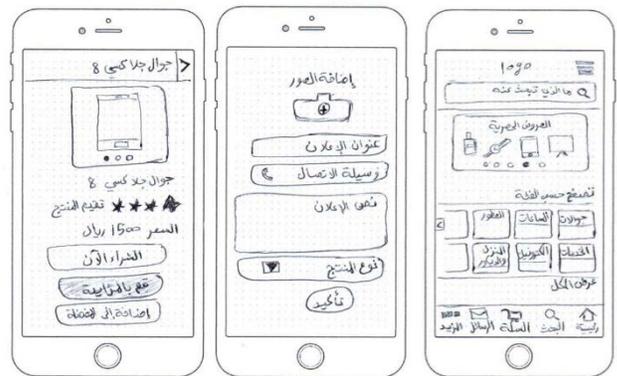

Fig. 5.   Early Prototypes of the Mobile Auction Concepts and App.

C. Mobile Auction Creation through Focus Group

Subsequently, we conducted three separate focus groups, including a total of 10 participants, to reflect on the mobile bidding requirements and early prototypes (depicted in Fig. 5). In each focus group, we mixed participants from different backgrounds and specialties to diversify the views and enrich the discussion. Focus groups are well-known methods for generating in-depth views and insights about the users' feelings and opinions about a particular technology [52].

However, to achieve the highest level of engagement and communication with our participants, we created high-fidelity interactive artifacts, as shown in Fig. 6. This empowered us to realize principles three, four, and five of Brhel et al. [45]. As we kept continuously collecting feedback from the participants, we augmented our mobile bidding platform with new requirements. High-fidelity prototypes are interactive visuals that exhibit various real functionalities, such as navigation and information architecture, of the software at hand [53].
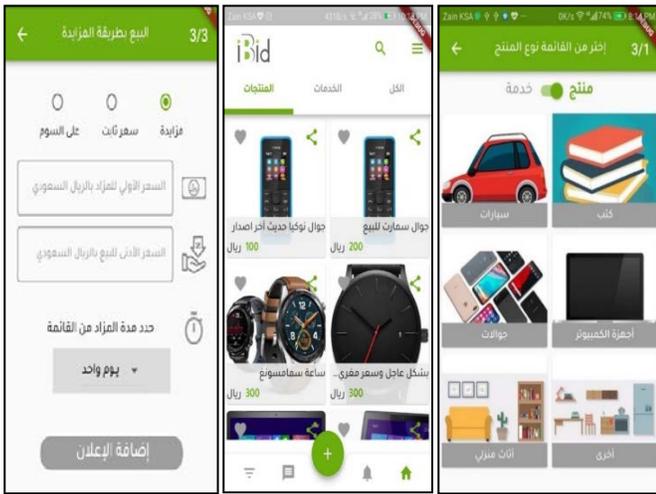
Fig. 6.   High-Fidelity Prototypes of the Mobile Bidding Concepts and App.

We applied consistent procedures in each focus group, where we first presented the idea of mobile auctions to the participants, followed by a discussion about their general impressions. Next, we presented the high-fidelity prototypes and tested various interactive functionalities to capture feedback from the participants. The results helped us hone the requirements and user interface designs of the iBid app. Overall, the focus groups gave rise to 33 new requirements, spanning the following concepts:

- RQ7. Social media sharing of products and services.

- RQ8. Consideration for culture-specific information, such as the gender.

- RQ9. Diverse payment methods, including E-payment systems.

- RQ10. Clarify of terms and conditions of auctions.

- RQ11. Measures to add trust factors, such as sellers' history and ratings.

- RQ 12. Preview products before committing to the auction.

- RQ13. Availability of a dispute process in case of auction frauds or complaints.

- RQ14. Logistic services to support the buying process (e.g., packaging and shipping).

D. Mobile Auction Implementation Decisions

Since we aimed to explore and promote online auctions using mobile devices, we first decided to create a mobile-friendly and cross-platform app. We selected robust and well-supported mobile development frameworks to expedite the implementation process. Moreover, our mobile bidding architecture implemented the three-tier client-server architecture [54], namely the presentation layer, business layer, and data management layer. E-bidders issue requests and events on the mobile interface, representing the presentation layer. The business logic layer is responsible for managing and executing the business processes and functionalities of the

bidding system. The data management layer handles product inventory, auctions, user profiles, and so on.

We used the Flutter framework [55] to develop the user interfaces of the aforementioned mobile bidding app (i.e., the presentation layer). Flutter is an open-source UI framework for creating high-performant native interfaces for Android and iOS applications. It offers plenty of ready-made and customizable widgets enabling fast development and production. Moreover, we used Google Firebase [56] to implement the business logic and database layers to achieve the architecture depicted in Fig. 7. There are several qualifying reasons for choosing Firebase. Firebase is a back-end-as-a-service web and mobile development framework, offering a wide range of functionalities, such as user authentication, real-time database, cloud storage, cloud functions and messaging, and ready-to-use API/HTTP requests, API/Authentication, among others.
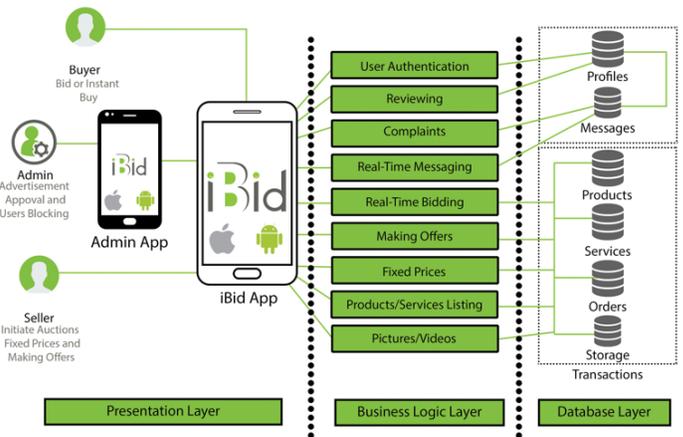

Fig. 7.   A Three Tier Architecture of our Mobile Auction App.

For the Database layer, NoSQL database [57] was preferred to store information about user profiles, messages, products and services, and orders and make them accessible through native mobile SDKs. The entities of these data were implemented in the Cloud Firestore Database, which empowers the creation of document databases that are stored as nested JSON objects. This choice was motivated by the fact that data are updated and synchronized in real-time to the devices of millions of connected clients, thus achieving the goal of real-time bidding. Moreover, Cloud Firestore Realtime Database enables scalability easily, which is critical to growing the concept of mobile bidding. Data are organized in the form of collections, where each collection contains a set of documents that can store simple or complex hierarchical data structures. Moreover, Cloud Firestore can protect and secure access to the data using Cloud Firestore security rules and firebase authentication.

E. Live Deployment of the Mobile Auction System

The primary API services and infrastructure for our live mobile auction system were implemented on Firebase, as shown in Fig. 8.

- Cloud Firestore: enables the real-time storage and access of auction collections (e.g., products, orders, etc.).

- Firebase Authentication: enables users (auctioneers and bidders) verification through a valid email and password.

- Cloud Storage: enables the storage of complex media such as videos and pictures.

- Real-time Database: enables the storage of real-time messages between the system's auctioneers and bidders.
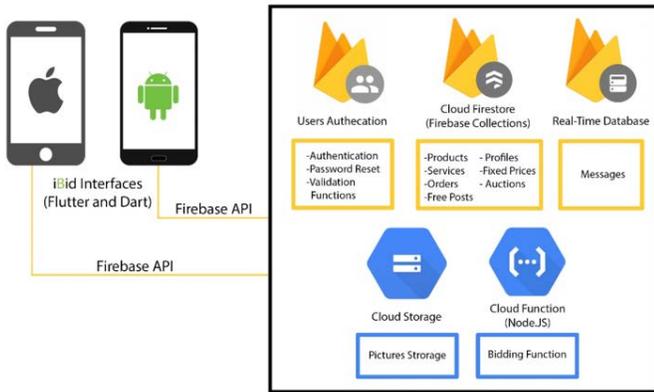

Fig. 8.   The Infrastructure of our iBid Mobile Auction App.

We deployed our bidding app on Google PlayStore to gauge users' real reactions and collect market-specific requirements. Within seven days, 103 real users downloaded our auction app (see Fig. 9). Next, we communicated directly to those registered users asking for feedback and suggestions upon their actual use of the iBid app, resulting in the below requirements:

- RQ15. Emphasis on allowing cash payment

- RQ16. Filter options to find specific products/services

- RQ17. Enforce user identities using phone numbers (e.g., OTP registration)

Fig. 9.   iBid App Installs during the Testing Period.

F. User Testing of the Mobile Auction App

In the subsequent phase of the UCASD methodology, we conducted an acceptance testing to confirm users' attitudes and perceptions toward the use of e-auctions using mobile phones. Therefore, we distributed an evaluation survey to the users who had installed and used our m-auction app. In total, 38 users responded to our survey. However, only 28 respondents

(approximately 73%) had actually used our app to bid for products and services. Therefore, we discarded the answers of the other ten users. The respondents rated the questions measuring their satisfaction with various aspects of mobile auctions. In doing so, they rated their agreement with the evaluation statements on a 5-point Likert scale, ranging from 1= strongly disagree to 5= strongly agree. In general, the rating questions assessed two main aspects; the first aspect is related to the usability of the mobile bidding app, and the second aspect is related to user acceptance and intention to continue using the app.

24 (85.71%) real users rated our mobile bidding app as 'easy to use' (4 or 5/5), while only two users disagreed (Fig. 10). 20 (71.42%) users agreed that 'learning to use iBid app is quick' (Fig. 11). On the other hand, 22 users (78.57%) disagreed that iBid app is cumbersome to use (Fig. 12).
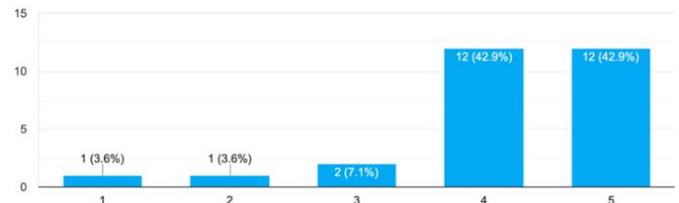

Fig. 10. Rating Distribution of: "iBid App is Easy to Use.".
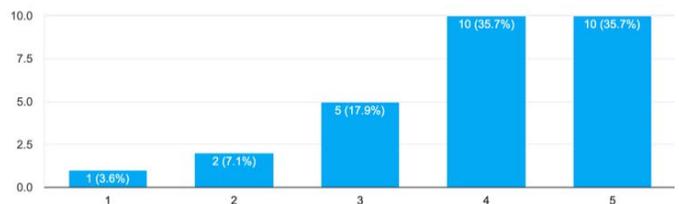
Fig. 11. Rating Distribution of: "Most Users will Learn to use iBid App Quickly.".
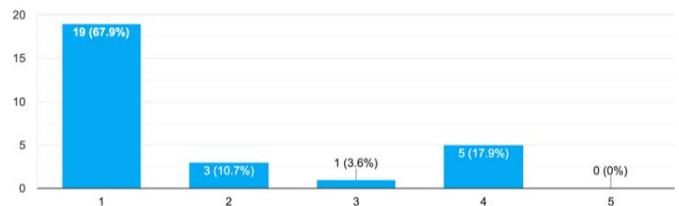
Fig. 12. Rating Distribution of: "iBid App is Cumbersome to Use.".

However, when it came to user intention to use mobile auctions, the ratings were quite encouraging. 19 (67.85%) users indicated their intention to use mobile bidding regularly, while only three users did not like the m-auction concept (Fig. 13). 18 (64.28%) users planned to reuse mobile auctions in the near future; however, five users did not show interest in auctioning again (Fig. 14). Approximately 20% of the respondents were neutral about the future use of m-auctions.
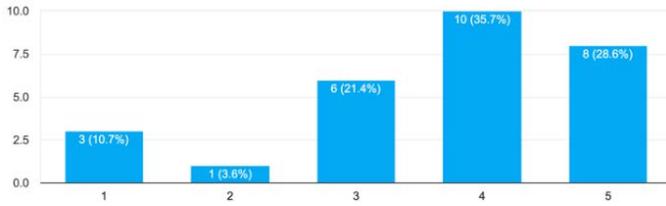
Fig. 13. Rating Distribution of: "I Intend to Use Mobile Auctions Regularly.".
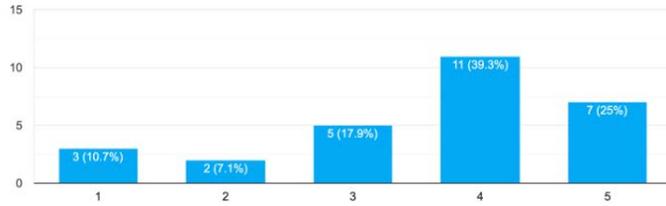
Fig. 14. Rating Distribution of: "I Plan to Reuse Mobile Auctions in the near Future.".

With respect to the qualitative feedback, we received a total of 46 positive comments and 20 negative comments from our users about the mobile bidding experience. We applied inductive thematic analysis [58] to identify potential m-auctions enabling drivers and deterrents. The emerging themes emphasized the following aspects:

- RQ24. Simplicity and aesthetics of bidding app
- RQ18. Time-constrained auctions
- RQ19. Transparency in showing sellers' details
- RQ20. Variety of products and services
- RQ21. Competitive pricing of offerings
- RQ22. Support of multi-languages (Internationalization of platform)
- RQ23. Platform quality and reliability (Free of glitches and errors)
- RQ24. Quality of products descriptions
- RQ25. Authenticity and seriousness of sellers
- RQ26. Consistency between product description and product received.

G. Card Sorting of m-Auction Requirements

In this phase, we applied card sorting to categorize the collected requirements into meaningful groups. Cart sorting is a well-tested technique used in user experience research to create a structured taxonomy of concepts [59]. We applied open card sorting (i.e., with no pre-determined groups) to establish a hierarchy of factors emerging from the UCASD activities. Fig. 15 depicts the final taxonomy of m-auctions requirements.
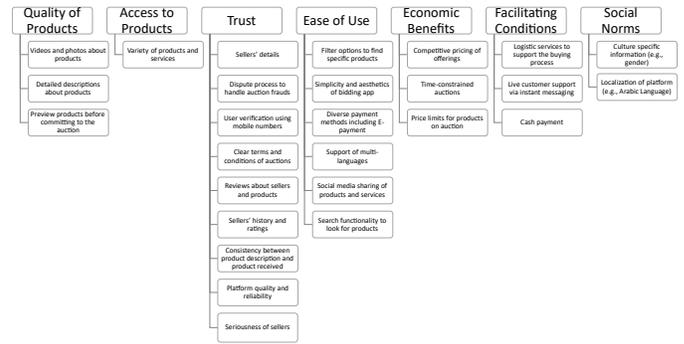

Fig. 15. Final Taxonomy of Mobile Auctions Requirements and Factors.

V. ENABLING AND INHIBITING FACTORS OF MOBILE AUCTIONS

Following the UCASD activities, we held three consecutive focus groups to establish an in-depth understanding of the factors that may influence consumers' acceptance of online mobile bidding in Saudi Arabia.

A. Focus Groups as a Means of Theory Formation

Qualitative research is adopted to develop theoretical explanations [60] that underpin users' attitudes toward mobile auctions. Focus groups are chosen as the method of exploration since they enable focused discussions with the selected participants on a particular topic [61], [62]. We organized three focus groups consecutively, where the outputs of one focus group were used to motivate the discussions of the next focus group. Overall, 22 undergraduate students participated in our focus groups (7, 8, and 7, respectively), with each focus group lasting approximately one hour and a half.

The focus groups' discussions were fully transcribed, and the data was anonymized and analyzed using thematic analysis. 156 high-level codes emerged, distributed across 28 major themes, as depicted in Table II. The comments are divided into enabling and inhibiting m-auction factors.

TABLE II. STATISTICS OF THE HIGH-LEVEL CODES AND THEMES EMERGING FROM THE FOCUS GROUPS

	Enabling	Inhibiting
High Level Codes	81	75
Themes	14	14

The below sub-sections summarize the classification of the key positive and negative factors that may impact the acceptance of mobile bidding by e-shoppers in Saudi Arabia.

B. Enablers of Mobile Auctions

The focus groups participants gave a total of 81 positive perceptions and 75 negative perceptions. These perceptions were classified into 13 critical factors: risk, quality of products, trust, ubiquity, usefulness, access to valuable products, ease of use, age, social influence, monetary costs, enjoyment, past experience, and facilitating conditions. Table III summarizes the key factors, themes, and occurrence percentage of each theme.

TABLE III.   THE PROMINENT FACTORS PERCEIVED TO INFLUENCE THE USE OF MOBILE AUCTIONS; PERCENTAGE OF POSITIVE (+) AND NEGATIVE (–) COMMENTS

Factor (%)	Themes	Positive (%)	Negative (%)
Risk (15.38%)	(-) Non-seriousness of bidders	-------	18.67%
	(-) Bid rigging and underpricing products and services	-------	5.33%
	(-) Non-eagerness of sellers	--------	6.67%
	(-) Monopolistic markets	-------	1.33%
Quality of Products (13.46%)	(+) Guarantees about products quality	6.17%	12%
	(-) Uncertainty about products quality		
	(+) Clear Information about products	8.64%	--------
Trust (12.82%)	(+)/(-) Trust of the platform and sellers	17.39%	5.33%
	(+) Sense of being secure using online transactional platforms	2.47%	--------
Ubiquity (12.18%)	(+) No geographic restrictions	12.36%	-------
	(+) Availability of auctions any time	3.70%	8%
	(-) Auctions are restricted to specified time windows		
Usefulness (9.62%)	(+) Functional benefits	11.07%	-------
	(+) Economic benefits	7.41%	-------
Access to Products (8.33%)	(+) Access to valuable products	12.28%	4%
	(-) Unavailability of products		
Ease of Use (7.69%)	(-) Technical issues	-------	8%
	(+) Easiness of communication	3.70%	-------
	(+) Easiness of bidding	3.70%	-------
Moderating factors (6.41%)	(-) Age	--------	10.67
	(+) Past experience using e-commerce platforms	2.47%	-------
Social Influence (5.13%)	(-) Community readiness and awareness	--------	8%
	(+) Marketing and advertisement	2.47%	--------
Monetary costs (5.13%)	(-) High prices	-------	8%
	(-) Platform commissions and charges	-------	2.67%
Enjoyment (3.21%)	(+) Experience of mobile auction	4.94%	-------
	(+) Excitement	1.23%	-------
Facilitating Conditions (0.64%)	(-) Slow delivery of products	-------	1.33%

Risk (15.38%): potential risk was perceived as the major concern during the use of mobile auctions by our participants. The fears were centered on the non-seriousness of auctioneers and bidders when committing to selling and buying decisions, bid rigging when merchandises are over-valued or under-valued by the platform users, and the monopolization of markets regarding specific products (see Table IV).

Quality of products (13.46%): participants indicated that the quality of products available in the auctions would highly influence their attitudes towards mobile bidding. The factors that determine their judgment include the availability of clear information about products (8.64%) and the guaranteed quality of the products (6.17%). However, concerns about the products' quality could be a major deterrent to indulging in bidding (12%), as shown in Table V.

Trust (12.82%): trust came out as the third important m-auction factor that our participants discussed. The high-level themes included trust of the auction platform and sellers

(22.61%) and the feeling of being secure using online systems because of the Corona pandemic (2.47%); see examples in Table VI.

TABLE IV.   EXAMPLES OF PARTICIPANTS' EXCERPTS ABOUT THE RISK FACTOR

Risk Themes (Frequency in Percent)	Evidence
(-) Non-seriousness of bidders (18.67%)	(-) Participant 2, Focus Group 3 " It is difficult to prevent interferences from bid/product owners." (-) Participant 3, Focus Group 2 " Many people don't participate on some mobile bidding websites because is impossible to enter a bid which goes with higher and higher prices so you might suspect that bidders are same people as sellers."
(-) Non-eagerness of sellers (6.67%)	(-) Participant 4, Focus Group 3 " The owner of the item cancels the sale." (-) Participant 2, Focus Group 2 " The seller's lack of seriousness by displaying pictures of the fake item, and when the sale is done, he does not send the item to you"
(-) Auction rigging and underpricing products and services (5.33%)	(-) Participant 3, Focus Group 3 " I guess bidder might don't give the product its fair price." (-) Participant 1, Focus Group 1 " Undervalued of goods that means you bid and give a lower price gives the commodity its value"
(-) Monopolistic markets (1.33%)	(-) Participant 6, Focus Group 1 " Actually, some applications provide a monopolized auctions on certain products." (-) Participant 2, Focus Group 3 " Some buyers buy all the quantity so once the bidding is over, they sell it at double its price"

TABLE V.   EXAMPLES OF PARTICIPANTS' EXCERPTS ABOUT THE PRODUCTS QUALITY FACTOR

Quality of Products Themes (Frequency in Percent)	Evidence
(+) Clear information about the products (8.64%)	(+) Participant 2, Focus Group 1 " A bidder can search on the Internet for the items that he is bidding on to learn more details about them on the web, where the person has a wide field to search for what is bidding on." (+) Participant 2, Focus Group 3 " Mobile bidding should provide a live show feature so that the seller can display his products clearly"
(-) Uncertainty about products quality (12%)	(-) Participant 5, Focus Group 3 " If I bid on a product, and once I got the product it appeared that it wasn't the same product, I mean it came out of with different specifications. How do I get my rights back? A solution to this point must be (product return guarantee)" (-) Participant 1, Focus Group 1 " If you are bidding on artifacts and goods, you don't know its exact specifications. It's best to inspect it personally so I feel there is less guarantee than the traditional bidding."
(+) Guarantees about products quality (6.17%)	(+) Participant 1, Focus Group 1 " Mobile bidding can provide a solution to this point (i.e., product return and guarantee); this will be the supporting factors to use the auction app." (+) Participant 4, Focus Group 3 " Sellers can also provide a warranty if the product, for example, has a defect. The whole process will be safe."

TABLE VI. EXAMPLES OF PARTICIPANTS' EXCERPTS ABOUT THE TRUST FACTOR

Trust Themes (Frequency)	Evidence
(+) Trust of the bidding platform and auctioneers (22.61%)	(+) Participant 6, Focus Group 3 " Through mobile auctions, you can register the seller's name so if you are confident in them, you can visit their shops and buy from them." (+) Participant 1, Focus Group 1 " I expect that a trusted party that can see the product and inspect what its defects are, and present it to us." (-) Participant 6, Focus Group 3 " Because what is happening now is that there are many biddings mobile applications that you feel it's kind of a fraud and manipulation." (-) Participant 1, Focus Group 1 "Online auction sites sometimes contain scam sellers"
(+) Feeling of being secure (2.47%)	(+) Participant 2, Focus Group 1 " If you want to spread the culture of electronic bidding, this is the best time. This is the right time, especially after the Corona pandemic people will feel secure right now." (+) Participant 6, Focus Group 1 " It will succeed especially after the current crisis in Corona, people feel more secure delaying with online transactions, there is no longer this freedom of movement"

Ubiquity (12.18%): participants' views were appreciative concerning the time and location freedom auctioneers, and bidders enjoy while using mobile auctions. The critical themes under ubiquity were no geographic barriers (12.36%) and availability of bidding all the time (3.7%). However, participants were worried about the timed auctions (8%), which could only make commodities available for certain time frames (see Table VII). In other words, mobile auctions must be available 'anytime from anywhere'.

TABLE VII. EXAMPLES OF PARTICIPANTS' EXCERPTS ABOUT THE UBIQUITY FACTOR

Ubiquity Themes (Frequency in Percent)	Evidence
(+) No geographic restrictions (12.36%)	(+) Participant 4, Focus Group 3: " It is great there is no compulsion about the geographic location; a person in one country can bid in an auction in another country, as it is considered as facilitating the bidding process. "
(-) Limited to specific times (8%)	(-) Participant 5, Focus Group 3 " I think mobile bidding is almost limited at a certain time, so it is not flexible like the traditional bidding." (-) Participant 1, Focus Group 1 " The time difference might be a problem imagine if the commodity is offered in faraway country, and you are bidding on it in your country."
(+) Availability of auctions any time (3.7%)	(+) Participant 2, Focus Group 1: "Facilitates meeting of buyers and sellers, it is possible to save time and effort, for seller and buyer there are no time obstacles." (+) Participant 1, Focus Group 1: " The auctions should be long and continuous. It does not stop, that means it should be available 24 hours."

Usefulness (9.62%): our participants emphasized that indulging in mobile auctions is motivated by two benefits: functional and economical. Examples of functional gains include convenience and saving time and effort. However, economic gains often refer to cost savings through bidding

activities. References were made mainly to the competitive pricing strategies that auctioneers set to win new bidders for their goods and services. See participants' comments in Table VIII.

TABLE VIII. EXAMPLES OF PARTICIPANTS' EXCERPTS ABOUT THE USEFULNESS FACTOR

Usefulness Themes (Frequency in Percent)	Evidence
(+) Functional benefits (11.07%)	(+) Participant 2, Focus Group 1 " You can do the bidding while you are relaxed at your home." (+) Participant 1, Focus Group 1 " Mobile bidding is a fascinating idea to bid on products and save your time and effort"
(+) Economic benefits (7.41%)	(+) Participant 7, Focus Group 3 " With mobile bidding you can buy the goods you want. For example, a specific device, you can get it used for a cheaper price." (+) Participant 6, Focus Group 3 " You also can find special prices I mean like special offers."

Access to products (9.88%): participants highlighted that m-auctions would enjoy a wider acceptance if they were to provide access to unique and rare merchandise, as demonstrated in Table IX. However, they should also act as a marketplace for other regular products (4%).

TABLE IX. EXAMPLES OF PARTICIPANTS' EXCERPTS ABOUT ACCESS TO PRODUCTS FACTOR

Access to Products Themes (Frequency in Percent)	Evidence
(+) Access to valuable products (9.88%)	(+) Participant 4, Focus Group 2 " If you want to buy from online mobile auctions, you must find a rare product or something special you need that is not available in the market." (+) Participant 5, Focus Group 1 " There is a possibility of finding a competition, but it will be for certain rare products."
(-) Unavailability of Products (4%)	(-) Participant 3, Focus Group 3 " There is no bidding mobile application that provides you with all the products that you may need it.

Ease of use (7.69%): 'ease of use' fared into the middle of the table with respect to frequency of occurrence, but it was still perceived to impact intentions to use m-auctions. The themes of ease of use included ease of communication (3.7%) and simplicity of the bidding process (3.7%). Moreover, the platform should have no technical failures and glitches (8%). In other words, the auction platform should be technically sound and reliable, as shown in Table X.

Control factors (6.41%): first, participants were quite doubtful that mobile auctions are suitable for the aging population (10.67%); second, there were indications that prior experience using e-commerce and transactional platforms would encourage users to place bids (2.47%), as listed in Table XI.

Social influence (5.13%): social influence, in the form of marketing, was expected to have a mass effect on the adoption of mobile bidding positively. As demonstrated in Table XII, there is an absolute necessity to improve awareness to ensure that society is ready to accept the concept (8%).

TABLE X.   EXAMPLES OF PARTICIPANTS' EXCERPTS ABOUT THE EASE OF USE FACTOR

Ease of Use Themes (Frequency in Percent)	Evidence
(-) Technical issues (8%)	(-) Participant 6, Focus Group 3 " Technical reasons, because the site might be out of service and the product bidding available only for a short time." (-) Participant 4, Focus Group 2 " There might be some problems with the mobile application which prevent some users from bidding in a timely manner"
(+) Easiness of communication (3.7%)	(+) Participant 1, Focus Group 3 " Of course, there will be a certain number of sellers and buyers. The buyer can browse hundreds of items at the same time and communicate with their sellers through the app." (+) Participant 3, Focus Group 1 " the platform will provide easy communication between the seller and buyer"
(+) Easiness of bidding (3.7%)	(+) Participant 6, Focus Group 2 " I tried an online biding auction, I found it ease without any difficulties." (+) Participant 2, Focus Group 13" The mobile biding application must have a button to go back and unbid if I have done a mistaken biding"

TABLE XI.   EXAMPLES OF PARTICIPANTS' EXCERPTS ABOUT THE AGE AND PAST EXPERIENCE FACTORS

Themes (Frequency in Percent)	Evidence
Age (10.67%)	(-) Participant 2, Focus Group 3 "Older people do not prefer to use mobile bidding. Older people prefer to go by themselves to the site and see the product for themselves." (-) Participant 5, Focus Group 2 " The elderly are very important part of our communities. How you can communicate with the elderly when they do not have a broad background in technology:"
(+) Past Experience (2.47%)	(+) Participant 2, Focus Group 1 "I can say that this is the right time, 85 percent of population have a background in electronic auctions." (+) Participant 3, Focus Group 1 "These days most people are using electronic ways of selling and buying"

TABLE XII.   EXAMPLES OF PARTICIPANTS' EXCERPTS ABOUT THE SOCIAL INFLUENCE FACTOR

Social Influence Themes (Frequency in Percent)	Evidence
(-) Community readiness and awareness (8%)	(-) Participant 6, Focus Group 3 "I guess people here in Saudi Arabia are not comfortable with mobile bidding, so we need to raise awareness about the concept of electronic bidding in general." (-) Participant 4, Focus Group 1 "Some communities don't have experience in electronic bidding. Individuals must be educated so they understand the electronic bidding process"
(+) Marketing and advertisement (2.47%)	(+) Participant 7, Focus Group 3 " It is necessary to intensify the advertisement about m-auctions in the targeted areas." (+) Participant 2, Focus Group 2 " The platform for mobile bidding must be free and offer gifts; Also, advertisement must be extensive about the concept."

Monetary costs (5.13%): concerns were raised about setting high prices in certain auctions and any platform charges, as exemplified in Table XIII.

TABLE XIII.   EXAMPLES OF PARTICIPANTS' EXCERPTS ABOUT THE MONETARY COSTS FACTOR

Monetary Costs Themes (Frequency in Percent)	Evidence
(-) High Prices (8%)	(-) Participant 6, Focus Group 2 " Sometimes in an auction the seller says this product, for example, starts from 100 riyals. Many buyers say we won't participate and withdraw from the auction." (-) Participant 2, Focus Group 1 " Celebrities can use these mobile bidding and the prices go high because of them and their fans."
(-) Platform commission (2.67%)	(-) Participant 4, Focus Group 3 " The commission must be at a symbolic price." (-) Participant 5, Focus Group 1 " The owner of the product might delay or even hesitate to display their goods because the platform raises the commission price"

Enjoyment (3.21%): one interesting motif, albeit infrequent, for e-shoppers to carry out m-auction was perceived enjoyment. The themes under enjoyment included the sense of excitement and auction experiences, as shown in Table XIV.

TABLE XIV.   EXAMPLES OF PARTICIPANTS' EXCERPTS ABOUT THE ENJOYMENT FACTOR

Enjoyment Themes (Frequency in Percent)	Evidence
(+) Experience of mobile auction (4.94%)	(+) Participant 2, Focus Group 1 " I want to try it after the current crisis of covid-19; there is no longer freedom of movement." (+) Participant 3, Focus Group 2 " We can go to the physical environment of such auctions; we can go to auctions and ask the bidders to see how they carry out the bidding"
(+) Excitement (1.23%)	(+) Participant 6, Focus Group 3 " I had seen this live bidding before which was on precious stones, and the situation was in which there was a competition and enthusiasm among bidders."

Facilitating conditions (0.64%): lastly the participants mentioned some comments about the logistic services supporting the after-sale activities (e.g., delivery) which could impact mobile auctions, as stated in Table XV.

TABLE XV.   EXAMPLES OF PARTICIPANTS' EXCERPTS ABOUT THE FACILITATING CONDITIONS FACTOR

Facilitating Conditions Themes (Frequency in Percent)	Evidence
(-) Slow delivery of products (1.33%)	(-) Participant 6, Focus Group 3 " I believe that a slow delivery of goods will be the main challenge in using the mobile bidding application."

VI.   DISCUSSION, PRACTICAL IMPLICATIONS, AND LIMITATIONS

To the best of our knowledge, our research is the first work to quest for the enablers and deterrents of using mobile auctions in Saudi Arabia. Two reasons inspired the selection of

this topic. First, despite the recent e-commerce boom and high user penetration, the adoption of online auctions is almost non-existent in the Saudi e-commerce market. Second, the online global auction industry is expanding year-on-year, with an estimated 7.2% yearly growth [4]. As such, the findings of our studies offer a foundational comprehension of the current attitudes and practices of Saudi auctioneers and bidders.

When coupled with user-centered design, agile development of software helps achieve numerous benefits [45] [63] In our case, the adoption of the User-Centered Agile Software Development (UCASD) empowered us to build a user-driven mobile application through a series of quick improvement cycles. The emerging design features were gathered from potential users (i.e., user-centered philosophy) and added to the m-auction platform in an incremental fashion (i.e., agile practices), thus reducing user frustration and improving customer traction. We demonstrate the procedures we followed and the lessons from applying UCASD. We were able to develop a user-validated m-auction platform since our design decisions were mainly based on user research. To this end, our focus was not to develop a feature-rich application but rather to create an excellent user experience by keeping our users at the center of the design and development activities. On another positive aspect, the short design iterations enabled us to tackle the tangible issues early, identify new relevant requirements, and quickly produce a functional mobile application to gauge users' attitudes toward mobile auctions. Another strong aspect of our research is the delivery and deployment of mobile bidding in the real world, where real bidders used the mobile platform and provided valuable feedback.

The use of UCASD, albeit beneficial in multiple ways [64], has caused us several challenges. We enumerate and suggest ways to remedy these challenges. The first challenge we faced was to fit usability studies, which typically take a long time, within quick development cycles. The second challenge was to accommodate some of the conflicting principles of UCD and agile development (e.g., up-front user research in UCD vs. reduction of up-front planning to produce working software in agile quickly). We found that building flexible team compositions, containing the necessary design, development, and testing skills, helped to minimize this challenge. We devised a cross-functional team to work on different tasks and processes, with the possibility of re-location if required. Using an online survey enabled us to reach a vast number of users within a short time, thus reducing up-front planning activities. We also found that considering user experience from the start of the project helps incorporate user views. Moreover, we had to adapt the complexity of our user-centered design activities to accommodate the agile sprints. In this sense, we shortened the usability studies and conducted them with fewer participants. We also employed low-fidelity prototyping to engage our users in a cost-effective manner. Finally, we found that planning our reasonably timed sprints (4/5 weeks) to embed design, implementation, and usability testing activities quite fruitful.

Now let us attempt to answer the main question of our research; why does Saudi Arabia, despite its recent major digital transformations, still lags regarding the use of online mobile auctions? The acceptance of new technologies is well-known to be triggered by the existence and amalgamation of several factors [65][66]. Our research is no exception. The qualitative inspection of a new phenomenon (e.g., the use of emerging technologies) is advocated by several studies [67][68]. Applying an inductive approach in our research, 13 distinct factors emerged as the key determinants of mobile bidding activities in Saudi Arabia (see Fig. 16).
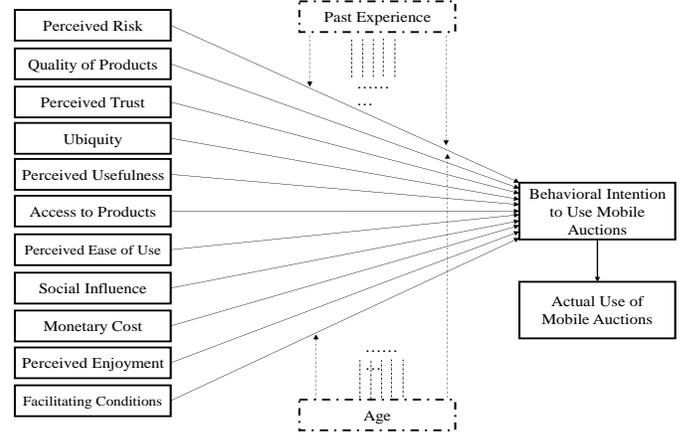

Fig. 16. Preliminary Online Mobile Auctions Technology Acceptance Model. Dashed Rectangles Represent the Moderating Factors, Which Influence all other Factors (Listed on Left Hand Side).

Perceived risk and trust were the dominant m-auction themes from our qualitative inspection, accounting for nearly a third (28.2%) of the concerns. Bidders were particularly worried about the genuineness of sellers, potential frauds, and auction rigging during the live mobile auctions, among other fears. Our findings are in line with the previous claims. For instance, Xu et al., demonstrated that perceived risk impacts the buyer's intentions to participate and engage in online consumer-to-consumer buy-it-now auctions [69]. Perceived risk, in return, was influenced by the buyer's risk attitude, the seller online reputation, and the product's price and type. Gull et al. showed that online shoppers are still concerned about their security in Saudi Arabia [70]. A recent laboratory study demonstrated that shill bidding remains a serious threat in mobile auction platforms [71].

The quality of products was the second most important factor to our participants in the bidding decision-making process. According to our bidders, products' quality is determined by clear information and specifications that reduce any uncertainties about the business deal. In a recent study, the authors investigated the behavior of bidders in mystery auctions where products' information was purposefully concealed [72]. The results showed that if information about the products' quality has not been provided, the seller's reputation will be greatly influenced. Moreover, when products' quality is unclear, reputation becomes more influential on the trustworthiness of the auctioneers.

Ubiquity refers to the property of being available anytime and anywhere. This is probably a natural factor to emerge as smartphones are characterized by their availability and mobility. Participants valued the possibility of making bids without geographic or time restrictions. Previous works back our findings [73], where ubiquity was found to be a significant

predictor of attitudes toward mobile payment services in Korea.

Perceived usefulness, represented by functional and economic gains, stood out as an influential factor in making online mobile bids. This is no surprise since usefulness has been a foundational construct in almost all types of technology acceptance models [74][75].

Similarly, when Vadovič explored Internet bidding behavior using dynamic auction settings, results showed that bidders with presumably higher search costs tend to offer a higher bid to secure the product [76]. In other words, bidders who are allowed to search for prices during the bidding process from other e-commerce websites are more likely to submit higher bids since they can assure of economic gains.

Interestingly, age and prior experience were discussed in our research to impact consumers' intention to conduct mobile bidding. Other e-commerce studies also found demographic characteristics (i.e., age) to have a moderating effect on technology acceptance [77][78]. However, Hernández et al., found that socioeconomic variables, such as age, gender, and income, do not alter e-commerce shopping intentions in developed countries [79]. Such observations are worth investigating in follow-up studies. Other constructs that arose in our studies include ease of using the mobile bidding platform, social influence, monetary costs, and enjoyment of m-auction activities. Once again, these factors are supported by several studies [80][81][82].

The implications of our results are both theoretical and practical. Theoretically, we developed a solid understanding of the factors that could encourage or inhibit the adoption of mobile auctions in Saudi Arabia. Our evidence-based factors, which were identified through several studies, represent the cornerstone for creating a technology acceptance model for live mobile auctions. The demonstration of applying the user-centered agile development will serve as a roadmap for practitioners and startup companies. Practically, we accumulated a set of m-auction requirements for e-commerce companies that might be interested in embedding bidding processes into their mobile platforms.

Our findings are indeed interesting and, we hope, will pave the way for further research about mobile auctions in developing countries. However, we would like to highlight a few limitations in this research that could restrict the validity of our observations, as follows.

The use of qualitative research empowered us to investigate consumers' behavior and probe their opinions and decision-making in detail; however, our interpretation of the themes and factors impacting mobile auctions might have been influenced by our subjective views. To reduce this bias, we checked the data analysis and themes' classification multiple times by four researchers to ensure the inter-rater reliability and accuracy of our analysis. However, there is also a likelihood that we might have missed some factors or misclassified some comments. Moreover, this type of research does not empower us to draw any causal relationships between the potential factors of m-auctions. More studies need to be conducted to verify any hypothesized relationships.

We used convenience sampling [83] to recruit participants in our user-centered design activities. This type of non-probabilistic sampling technique is subjective and, therefore, may not be representative of the intended population. It is, however, an effective approach when the research is restricted with respect to time and resources. Most of our participants were students, aged between 18 and 30 years, with a good technical background. Our sample lacked representative of elderly users. Moreover, our target group might have different characteristics, views, and behavior. We studied the e-shoppers' reactions to m-auctions in Saudi Arabia. Thus, our findings should be cautiously generalized to other user groups and MENA countries.

VII. CONCLUDING REMARKS AND FUTURE WORKS

We demonstrated the effectiveness of developing an interactive mobile auction platform through the application of a multi-phased User-Centered Agile Software Development (UCADS) methodology. 454 potential users participated in our mixed research studies, including an online survey, focus groups, prototyping, and user testing. Consequently, a taxonomy of the key mobile bidding requirements is presented. Integrating user-centered design and agile development principles showed that Saudi e-shoppers are ready to embrace mobile bidding activities. However, our follow-up qualitative studies identified 13 prominent factors that might impact consumers' intention to accept mobile auctions in Saudi Arabia. These enabling factors include trust, risk, usefulness, ubiquity, monetary costs, ease of use, merchandise information quality, access to valuable merchandise, enjoyment, social influence, facilitating conditions, age, and past experience. We recommend that these factors merit a profound investigation in upcoming confirmatory studies.

In the future, we plan to propose and test a mobile auction technology acceptance model for the MENA region. Our factors, coupled with evidence from current e-auction research works, will assist in establishing sound mobile auction hypotheses. Moreover, we intend to conduct a cross-sectional study to ensure the validity of the model in other developing countries that currently exhibit similar characteristics (e.g., Emirates, Malaysia, etc.).

REFERENCES

[1] Statista (2022) a, eCommerce - Saudi Arabia. https://www.statista.com/outlook/dmo/ecommerce/saudi-arabia, Accessed on 1 November 2022

[2] Statista (2022) b, Auction market worldwide - statistics & facts. https://www.statista.com/topics/8930/auction-market-worldwide/, Accessed on 1 November 2022

[3] Communications and Information Technology Commission (2017) ICT Report E-Commerce in Saudi Arabia. https://www.citc.gov.sa/en/reportsandstudies/Reports/Documents/CITC _ECOMMERCE_2017_ENGLISH.PDF, Accessed on 1 November 2022

[4] Research and Markets (2018), Global Online Auction Market 2018-2022. https://www.researchandmarkets.com/reports/4603459/global-online-auction-market-2018-2022, Accessed on 1 November 2022

[5] Charette, R. N. (2005). Why software fails [software failure]. IEEE spectrum, 42(9), 42-49.

[6] Issa, T. and Isaias, P., 2022. Usability and Human–Computer Interaction (HCI). In Sustainable design (pp. 23-40). Springer, London.

[7] Rondan-Cataluña, F. J., Arenas-Gaitán, J., & Ramírez-Correa, P. E. (2015). A comparison of the different versions of popular technology acceptance models: A non-linear perspective. Kybernetes.

[8] Jaradat, M. I. R. M., & Al Rababaa, M. S. (2013). Assessing key factor that influence on the acceptance of mobile commerce based on modified UTAUT. International Journal of Business and Management, 8(23), 102.

[9] Tarhini, A., Alalwan, A. A., Shammout, A. B., & Al-Badi, A. (2019). An analysis of the factors affecting mobile commerce adoption in developing countries: Towards an integrated model. Review of International Business and Strategy.

[10] Sari, E.N. and Subriadi, A.P., 2022, January. Consumer Behavior in Mobile Commerce Adoption: A Systematic Literature Review. In 2022 2nd International Conference on Information Technology and Education (ICIT&E) (pp. 287-292). IEEE.

[11] Blaise, R., Halloran, M., & Muchnick, M. (2018). Mobile commerce competitive advantage: A quantitative study of variables that predict m-commerce purchase intentions. Journal of Internet Commerce, 17(2), 96-114.

[12] Alfahl, H., Houghton, L., & Sanzogni, L. (2017). Mobile commerce adoption in Saudi organizations: a qualitative study. International Journal of Enterprise Information Systems (IJEIS), 13(4), 31-57.

[13] Turki, A. M. J., Pemberton, L., & Macpherson, H. (2017). Adoption and acceptance of mobile commerce in Saudi Arabia: the case of e-ticketing in the airline industry (Doctoral dissertation, University of Brighton).

[14] Alkhunaizan, A., & Love, S. (2012). What drives mobile commerce? An empirical evaluation of the revised UTAUT model. International Journal of Management and Marketing Academy, 2(1), 82-99.

[15] Salah, D., Paige, R. F., & Cairns, P. (2014, May). A systematic literature review for agile development processes and user centred design integration. In Proceedings of the 18th international conference on evaluation and assessment in software engineering (pp. 1-10).

[16] Cockton, G., Lárusdóttir, M., Gregory, P., & Cajander, Å. (2016). Integrating user-centred design in agile development. In Integrating User-Centred Design in Agile Development (pp. 1-46). Springer, Cham.

[17] Vulkan, N. (2020). The Economics of E-commerce. Princeton University Press.

[18] Statista (2022) c, eBay: annual net revenue 2013-2021, https://www.statista.com/statistics/507881/ebays-annual-net-revenue/, Accessed on 1 November 2022

[19] Statista (2022) d, Online C2C e-commerce penetration in the European Union 2007-2021, https://www.statista.com/statistics/381234/online-c2c-commerce-penetration-in-the-eu/, Accessed on 1 November 2022

[20] Niranjanamurthy, M., Kavyashree, N., Jagannath, S., & Chahar, D. (2013). Analysis of e-commerce and m-commerce: advantages, limitations and security issues. International Journal of Advanced Research in Computer and Communication Engineering, 2(6), 2360-2370.

[21] Pelletier, A., Khavul, S., & Estrin, S. (2020). Innovations in emerging markets: the case of mobile money. Industrial and Corporate Change, 29(2), 395-421.

[22] Xu, G., & Gutiérrez, J. A. (2006). An exploratory study of killer applications and critical success factors in M-commerce. Journal of Electronic Commerce in Organizations (JECO), 4(3), 63-79.

[23] Dass, M., Reddy, S. K., Newaz, M. T., & Reshadi, M. (2020). Discovering Market Structure of Ambiguously Appraised Products from Bid History in Online Auctions. In Continuing to Broaden the Marketing Concept. Emerald Publishing Limited.

[24] Ye, Q. C., Rhuggenaath, J., Zhang, Y., Verwer, S., & Hilgeman, M. J. (2021). Data driven design for online industrial auctions. Annals of Mathematics and Artificial Intelligence, 89(7), 675-691.

[25] Li, F., & Li, Y. (2011). Usability evaluation of e-commerce on B2C websites in China. Procedia Engineering, 15, 5299-5304.

[26] Marangunić, Nikola, and Andrina Granić. "Technology acceptance model: a literature review from 1986 to 2013." Universal access in the information society 14.1 (2015): 81-95.

[27] Mattia G, Di L A, Principato L (2021) On-line Consumer Behavior and Technology Acceptance Models. Springer Books:37-46

[28] Mohammad Ebrahimzadeh Sepasgozar, F., Ramzani, U., Ebrahimzadeh, S., Sargolzae, S., & Sepasgozar, S. (2020). Technology acceptance in e-governance: A case of a finance organization. Journal of Risk and Financial Management, 13(7), 138.

[29] Al-Dmour, A., Al-dmour, H. H., Brghuthi, R., & Al-Dmour, R. (2021). Technology Acceptance Dynamics and Adoption of E-Payment Systems: Empirical Evidence From Jordan. International Journal of E-Business Research (IJEBR), 17(2), 61-80.

[30] Tiwari, P., & Tiwari, S. K. (2020). Integration of technology acceptance model with perceived risk, perceived trust and perceived cost: Customers' adoption of m-banking. International Journal on Emerging Technologies, 11(2), 447-452.

[31] Chao, C. M. (2019). Factors determining the behavioral intention to use mobile learning: An application and extension of the UTAUT model. Frontiers in psychology, 10, 1652.

[32] Stern, B. B., Royne, M. B., Stafford, T. F., & Bienstock, C. C. (2008). Consumer acceptance of online auctions: An extension and revision of the TAM. Psychology & Marketing, 25(7), 619-636.

[33] Turel, O., Serenko, A., & Giles, P. (2011). Integrating technology addiction and use: An empirical investigation of online auction users. MIS quarterly, 1043-1061.

[34] Chang, H. H. (2010). Task-technology fit and user acceptance of online auction. International Journal of Human-Computer Studies, 68(1-2), 69-89.

[35] Eid, M. I. (2011). Determinants of e-commerce customer satisfaction, trust, and loyalty in Saudi Arabia. Journal of electronic commerce research, 12(1), 78.

[36] Al-Somali, S. A., Gholami, R., & Clegg, B. (2009). An investigation into the acceptance of online banking in Saudi Arabia. Technovation, 29(2), 130-141.

[37] Al-Maghrabi, A., Alharbi, B., & Tapper, N. (2011). Impact of the March 2009 dust event in Saudi Arabia on aerosol optical properties, meteorological parameters, sky temperature and emissivity. Atmospheric Environment, 45(13), 2164-2173.

[38] Statista (2022) e, Number of eBay's total active buyers from 1st quarter 2010 to 2nd quarter 2022 (in millions), viewed 01 November 2022, https://www.statista.com/statistics/242235/number-of-ebays-total-active-users/, Accessed on 1 November 2022

[39] Im, I., Hong, S., & Kang, M. S. (2011). An international comparison of technology adoption: Testing the UTAUT model. Information & management, 48(1), 1-8.

[40] Röcker, C. (2010). Why traditional technology acceptance models won't work for future information technologies. World Academy of Science, Engineering and Technology, 65, 237-243

[41] Wever R, Jasper V K, Casper B (2008) User-centred design for sustainable behaviour." International journal of sustainable engineering 1.1:9-20.

[42] Parvez, M.O., Arasli, H., Ozturen, A., Lodhi, R.N. and Ongsakul, V., 2022. Antecedents of human-robot collaboration: theoretical extension of the technology acceptance model. Journal of Hospitality and Tourism Technology.

[43] Cockton, G., Lárusdóttir, M., Gregory, P., & Cajander, Å. (2016). Integrating user-centred design in agile development. In Integrating User-Centred Design in Agile Development (pp. 1-46). Springer, Cham.

[44] Salah, D., Paige, R. F., & Cairns, P. (2014, May). A systematic literature review for agile development processes and user centred design integration. In Proceedings of the 18th international conference on evaluation and assessment in software engineering (pp. 1-10).

[45] Brhel, M., Meth, H., Maedche, A., & Werder, K. (2015). Exploring principles of user-centered agile software development: A literature review. Information and software technology, 61, 163-181.

[46] Zorzetti, M., Signoretti, I., Salerno, L., Marczak, S. and Bastos, R., 2022. Improving Agile Software Development using User-Centered

Design and Lean Startup. Information and Software Technology, 141, p.106718.

[47] Courage, C., & Baxter, K. (2005). Understanding your users: A practical guide to user requirements methods, tools, and techniques. Gulf Professional Publishing.

[48] Georgsson, M., & Staggers, N. (2016). An evaluation of patients' experienced usability of a diabetes mHealth system using a multi-method approach. Journal of biomedical informatics, 59, 115-129.

[49] Ehrlinger, L. and Wöß, W., 2022. A survey of data quality measurement and monitoring tools. Frontiers in Big Data, p.28.

[50] Christoforakos, L., & Diefenbach, S. (2018). Idealization Effects in UX Evaluation at Early Concept Stages: Challenges of Low-Fidelity Prototyping. In International Conference on Applied Human Factors and Ergonomics (pp. 3-14). Springer, Cham.

[51] Zaina, L.A., Fortes, R.P., Casadei, V., Nozaki, L.S. and Paiva, D.M.B., 2022. Preventing accessibility barriers: Guidelines for using user interface design patterns in mobile applications. Journal of Systems and Software, 186, p.111213.

[52] Hancerliogullari Koksalmis, G. and Damar, S., 2022. An empirical evaluation of a modified technology acceptance model for SAP ERP system. Engineering Management Journal, 34(2), pp.201-216.

[53] Li, J., W. Tigwell, G., & Shinohara, K. (2021, May). Accessibility of High-Fidelity Prototyping Tools. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-17).

[54] Ali, S., Alauldeen, R., & Ruaa, A. (2020). What is Client-Server System: Architecture, Issues and Challenge of Client-Server System. HBRP Publication, 1-6.

[55] Napoli, M. L. (2019). Beginning Flutter: A Hands On Guide To App Development. John Wiley & Sons.

[56] Khawas, C., & Shah, P. (2018). Application of firebase in android app development-a study. International Journal of Computer Applications, 179(46), 49-53.

[57] Gupta, A., Tyagi, S., Panwar, N., Sachdeva, S., & Saxena, U. (2017, October). NoSQL databases: Critical analysis and comparison. In 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN) (pp. 293-299). IEEE.

[58] Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. International journal of qualitative methods, 16(1), 1609406917733847.

[59] Conrad, L. Y., & Tucker, V. M. (2019). Making it tangible: hybrid card sorting within qualitative interviews. Journal of Documentation.

[60] Köhler, T., Smith, A. and Bhakoo, V., 2022. Templates in qualitative research methods: Origins, limitations, and new directions. Organizational Research Methods, 25(2), pp.183-210.

[61] Stewart, D. W., & Shamdasani, P. N. (2014). Focus groups: Theory and practice (Vol. 20). Sage publications.

[62] Labib, K., Evans, N., Roje, R., Kavouras, P., Reyes Elizondo, A., Kaltenbrunner, W., Buljan, I., Ravn, T., Widdershoven, G., Bouter, L. and Charitidis, C., 2022. Education and training policies for research integrity: Insights from a focus group study. Science and Public Policy, 49(2), pp.246-266.

[63] Tsangaris, E., Edelen, M., Means, J., Gregorowitsch, M., O'Gorman, J., Pattanaik, R., Dominici, L., Hassett, M., Witkowski, M.L., Schrieber, K. and Frank, E., 2022. User-centered design and agile development of a novel mobile health application and clinician dashboard to support the collection and reporting of patient-reported outcomes for breast cancer care. BMJ Surgery, Interventions, & Health Technologies, 4(1).

[64] Sy, D. (2007). Adapting usability investigations for agile user-centered design. Journal of usability Studies, 2(3), 112-132.

[65] Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. Information & management, 40(3), 191-204.

[66] Alsharhan, A., Salloum, S.A. and Aburayya, A., 2022. Technology acceptance drivers for AR smart glasses in the middle east: A quantitative study. International Journal of Data and Network Science, 6(1), pp.193-208.

[67] Kalu, F. A., & Bwalya, J. C. (2017). What makes qualitative research good research? An exploratory analysis of critical elements. International Journal of Social Science Research, 5(2), 43-56.

[68] Alase, A. (2017). The interpretative phenomenological analysis (IPA): A guide to a good qualitative research approach. International Journal of Education and Literacy Studies, 5(2), 9-19.

[69] Xu, B., Lin, Z., & Shao, B. (2010). Factors affecting consumer behaviors in online buy-it-now auctions. Internet Research.

[70] Gull, H., Saeed, S., Iqbal, S.Z., Bamarouf, Y.A., Alqahtani, M.A., Alabbad, D.A., Saqib, M., Al Qahtani, S.H. and Alamer, A., 2022. An empirical study of mobile commerce and customers security perception in Saudi Arabia. Electronics, 11(3), p.293.

[71] Carlson, J.I. and Wu, T., 2022. Shill bidding and information in eBay auctions: A Laboratory study. Journal of Economic Behavior & Organization, 202, pp.341-360.

[72] Chiou, L., & Pate, J. (2018). Trust, reputation and bidding behaviour in online mystery auctions. Applied Economics Letters, 25(10), 702-706.

[73] Lee, S. Y., & Park, J. (2016). A study on the intention of the use of mobile payment services: Application of the technology acceptance model. Korean Management Science Review, 33(2), 65-74.

[74] Surendran, P. (2012). Technology acceptance model: A survey of literature. International Journal of Business and Social Research, 2(4), 175-178.

[75] Tao, D., Fu, P., Wang, Y., Zhang, T. and Qu, X., 2022. Key characteristics in designing massive open online courses (MOOCs) for user acceptance: An application of the extended technology acceptance model. Interactive Learning Environments, 30(5), pp.882-895.

[76] Vadovič, R. (2017). Bidding behavior and price search in Internet auctions. International Journal of Industrial Organization, 54, 125-147.

[77] Al-Gahtani, S. S. (2011). Modeling the electronic transactions acceptance using an extended technology acceptance model. Applied computing and informatics, 9(1), 47-77.

[78] Katebi, A., Homami, P. and Najmeddin, M., 2022. Acceptance model of precast concrete components in building construction based on Technology Acceptance Model (TAM) and Technology, Organization, and Environment (TOE) framework. Journal of Building Engineering, 45, p.103518.

[79] Hernández, B., Jiménez, J., & Martín, M. J. (2011). Age, gender and income: do they really moderate online shopping behaviour?. Online information review.

[80] Chandra, C. P. (2015). The Adoption of E-Auction in Indonesia: The Extended Technology Acceptance Model Study. iBuss Management, 3(2).

[81] Tu, C. C., Fang, K., & Lin, C. Y. (2012). Perceived Ease of Use, Trust, and Satisfaction as Determinants of Loyalty in e-Auction Marketplace. J. Comput., 7(3), 645-652.

[82] Li, R., Chung, T. L. D., & Fiore, A. M. (2017). Factors affecting current users' attitude towards e-auctions in China: An extended TAM study. Journal of Retailing and Consumer Services, 34, 19-29.

[83] Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. American journal of theoretical and applied statistics, 5(1), 1-4.

Towards a Blockchain-based Medical Test Results Management System: A Case Study in Vietnam

Phuc Nguyen Trong, Hong Khanh Vo, Luong Hoang Huong, Khiem Huynh Gia, Khoa Tran Dang,
Hieu Le Van, Nghia Huynh Huu, Tran Nguyen Huyen, Loc Van Cao Phu, Duy Nguyen Truong Quoc,
Bang Le Khanh, Kiet Le Tuan
FPT University, Can Tho City, Viet Nam

Abstract— The role of the testing process cannot be denied in the diagnosis and treatment of patients' diseases in medical facilities today. The results from this process help doctors and nurses in medical centers make a preliminary and detailed assessment of symptoms and provide a specific course of treatment for their patients. In addition, these results are stored as a patient's medical record that serves as a reference for subsequent therapies. However, the storage of this information (i.e., paper-based, electronic-based) faces some difficulties for both approaches. Especially for developing countries (i.e., Vietnam), this process encounters some major obstacles at health centers in rural areas. Many centralized/decentralized storage methods have been proposed to solve the above problem. Besides, the current popular method is patient-centered (all information shared is decided by the patient) can solve the above problems and be applied by many research directions. However, these methods require the user (i.e., patient) to have a background in security and privacy as well as the cutting-edge technologies installed on their phones. This is extremely difficult to apply in rural areas in developing countries where people are not yet conscious of protecting their personal information. This paper proposes a mechanism for storing and managing test results of patients at medical centers based on blockchain technology - applicable to developing countries. We build a proof-of-concept based on the Hyperledger Fabric platform and exploit the Hyperledger Caliper to evaluate a variety of scenarios related to system performance (i.e., create, query, and update).

Keywords—*Blockchain-based system; hyperledger fabric; medical test results; medical institution at developing countries*

I. INTRODUCTION

All current treatments are based on the results of the clinical examination/test of the patient's symptoms and medical history. Indeed, doctors and nurses give their diagnosis about the patient's condition (i.e., consultation) and treatment. In other words, the testing process plays an extremely important role in the entire treatment of the patient [1]. In addition, these test results are recorded to build the patient's medical history. Nurses and doctors have a more general view of symptoms from early (i.e., first detected) to the current stage.

However, the storage of medical test results varies widely depending on the region (i.e., city or rural) and the country (i.e., developed or developing). Most hospitals in large cities and or in developed countries, adopt electronic records to store medical test results and patient medical records (aka electronic-based). While rural hospitals or some hospitals in developing countries still use paper records (aka paper-based). Paper records are extremely risky for the healthcare system because i) it is easy for patients to lose the medical test records

due to natural issues (e.g., flood, fire) or their failure (e.g., lost); ii) it is very difficult to back up those results because of technology and equipment limitations [2]. Due to these risks, some patients often take pictures of the medical test results to give to doctors for their next visit. For the recent model, almost all hospitals or medical centers in Vietnam applied manual input for the medical test record (see III-A for more details). Besides, determining timelines based on this approach is extremely difficult. For example, in addition to taking pictures, the patient must annotate all information related to those photos. Nevertheless, this method also carries risks, as loss of the device or memory card/memory limitations are the main barriers to this approach [3].

A study that collected patient responses to interviews about the role of using paper and electronic records was conducted by [4]. Their findings clearly indicate that electronic records are considered more reliable than paper records. Specifically, about 51% of interviewees indicated that they are less burdened in preserving paper records, and in case of loss of their records, they have to tell all their medical records to all. hospitals they visit if using paper records. An interesting finding of the author is that patients are willing to spend extra money each month to keep their health records in electronic form. In addition to the above reasons, the study by Muchangi and Nzuki [5] conducted a survey in medical centers in developing countries (i.e., India) that showed that patients did not trust the treatment. Using electronic records will reduce the risk of privacy invasion. They argue that the methods of building a user-centric health data sharing system are facing a lot of difficulties due to the limitations of the method of building a centralized data system (i.e., data) stored and processed centrally in cloud servers).

The risks that such systems may face come from unintended events such as natural disasters, fires or possible attacks by hackers to exploit sensitive patient information. For the first risk, studies[1] have shown that lost health information must be recreated, requiring in-depth time and resources. Because critical data such as patient health must require an electronic record using secure information systems to store and access patient health information and to ensure that the information contained updated and available when needed [2]. In addition, being attacked by hackers is inevitable. These hackers take advantage of existing security holes from the system to get personal information of patients and use them for purposes of violating privacy (e.g., selling personal information) [6]. It is impossible to recover stolen information. We can only find ways to overcome the current risks (i.e., centralized storage).

---

[1]https://www.hl7.org/FHIR/

The transition from centralized to decentralized medical data processing storage brings a lot of risks as well as benefits [7]. Nghia et al. [8] has argued that the challenges may outweigh the benefits, such as scalability, availability, data transparency (i.e., for the field). building systems on Blockchain), as well as giving more rights to users. Specifically, the authors argue that users can see who is working on their data and what benefits they have.

For the second approach (i.e., applying electronic record method to store patient information), in large cities in developing countries (e.g., Vietnam) archival methods are only centrally stored on a server of a center or a hospital in big cities (e.g., Ho Chi Minh City, Can Tho). Therefore, patients face a great challenge when they want to share medical data (i.e., test results, medical history) from previous facilities to the new facility [9]. Centralized data processing and storage systems exhibit at least one of the following disadvantages: instability due to a faulty central point, lack of security due to greater vulnerability, and greater potential for malicious attacks (i.e., unethical activities) due to the presence of central authority [10]. In contrast, the decentralized system has no central authority; instead, permissions are shared between each computer (node), each with equal permissions [11]. Instead of all processing and storage requirements being centralized in a central machine (i.e., cloud server), the distributed system divides computations into smaller computations to be performed by multiple nodes as well as storing the collected data in many different memory areas [12]. Thereby, increasing the processing and storage capacity for the whole system. This also increases interaction with users (i.e., patients, nurses, doctors) and reduces the risk of attacks from hackers because of the lack of a central point to attack [13]. This feature also makes them stable and fault tolerant since each node has the same role (i.e., regardless of Client-Server); thereby, privacy-invading operations are unlikely to be performed [14]. The peer-to-peer system uses these features and benefits to allow the network to remain fully operational even if one node fails.

In addition, because of the requirements for supporting equipment and supplies, the current testing procedures are all done in hospitals or medical centers (i.e., it is difficult to meet a large number of requests. patient testing at the same time). Patients have to wait a long time for testing and consultation results before their disease can be identified. Therefore, there is a need for a mechanism to support patients in storing and sharing medical records (e.g., test results, medical history) as well as limiting waiting time for results at centers. medical center. Besides, security and privacy issues depend greatly on the context in which the system is deployed. For example, in developing countries (i.e., Vietnam), exploitation and risk assessment and privacy have not been given due attention [15]. The developing countries' citizen do not have the concept of protecting personal privacy (especially countryside people). This argument is completely correct and can be applied to developed countries, where the education level is high and there are many supporting facilities and infrastructure.

To solve the problems mentioned above, this paper proposes a medical test result management model based on blockchain technology. Methods to prove the effectiveness of applying Blockchain technology to medical facilities to solve problems related to supply chain (e.g., blood and its produces [16], [17]) curative problem (e.g., emergency data assessment) that current health care systems have not fully addressed. Another example demonstrating the effectiveness of applying the strengths of Blockchain technology and patient care is introduced by Roehrs et al. [18]. Specifically, they emphasized non-functional requirements such as network usage, disk space, response time, CPU usage, and memory footprint, and their importance in implementing a pool management system. Health records based on blockchain technology. They evaluated the performance of both systems (i.e., traditional archiving system and blockchain-based storage system) by deploying two models on two hospital databases (i.e., 40,000 patients) adult person). Analyzing the results obtained, they concluded that the results achieved were much more effective than traditional practices in maintaining integrity, security, ownership, and decentralization. Various technologies can be used to create patient-centered healthcare systems where blockchain-based approaches provide a single solution [19]. Specifically, Blockchain uses features (security, stability, fault tolerance) to allow a network to remain fully functional even if a node fails [20]. However, these approaches (listed in the II section) suffer from many user-related (i.e., patient) limitations, which are highlighted in the Prior Work section.

Therefore, the research problem of this article is to introduce the blockchain-based approach for medical test result management used for developing countries (i.e., Vietnam) For the objective of this article, our contribution is threefold: i) proposing a model for managing test results for patients in developing countries (i.e., Vietnam); ii) building a proof-of-concept based on the proposed model via Hyperledger Fabric satisfying the specific properties of the regions/applicable areas; and iii) evaluate the system's capabilities based on how well it supports initialization, retrieval, and update requests (i.e., overall rating based on system performance - number of successful and failed requests; system-wide latency) based on exploiting Hyperledger Caliper.

The next section presents the state-of-the-art. Sections III and IV present our approach, processing model, and system implementation. Section V builds an environment for evaluating proposed models and makes comments on their strengths and weaknesses as well as future directions in Section VI. Finally, we summarize the study in Section VII.

II. RELATED WORK

There are many approaches that have proposed methods for remote diagnosis and treatment of diseases, which are data mining and other practical applications based on medical data by exploiting the strengths of the blockchain technology. For example, Chen et al. [21] proposes a model for storing and controlling personal data in a healthcare environment based on Blockchain technology. This system can collect information from IoT devices (i.e., medical devices in real time). To improve the security of the system, the authors build an anonymous data sharing environment and encrypt the patient's personal data before storing them on cloud servers. Similarly, Du et al. [22] and Son et al. [23] used medical centers (i.e., hospitals) to store data and manage access and those hospitals. Specifically, they categorize two types of medical data protection policies: global for all data shared outside

of the medical center, and local, which is accessed only by individuals at the medical center. medical (i.e., doctor, nurse).

However, one of the major limitations is that through this solution, patients do not have full control over their data as the data and policies are stored in the hospital. Patra et al. [24] proposes a cloud-based model to build an information system at the national level, providing a more convenient solution for patients in rural areas at the lowest cost. Specifically, instead of having to go to health care centers in large companies, they propose a solution to diagnose and treat diseases remotely. Specifically, citizens are encouraged to provide their personal healthcare information, which will be stored in the health cloud and accessed by health professionals and policymakers to provide more medical services. Similarly, Rolim et al. [25] proposes a framework that covers the process from data collection to cloud-based data delivery. Using sensors mounted on medical equipment, data can be collected and stored directly in the cloud, which can be accessed by authorized medical professionals.

Some other approaches build a user-centric (i.e., patient) model, who has full discretion to share their personal data with providers/health care facilities. economic (i.e., in a medical setting). For example, Makubalo et al. [26] has summarized the above approaches in their publication. They argue that the methods of building a user-centric health data sharing system are facing a lot of difficulties due to the limitations of the method of building centralized data system (i.e., data) stored and processed centrally in cloud servers). Yin et al. [27] introduced a patient-centric system built in the cloud with a data collection layer, data management layer, and medical service delivery layer based on medical records of the patient. To protect data privacy, many approaches have adopted attribute-based encryption (ABE), one of the most common encryption schemes used in cloud computing, to define patient data object. Depending on the context, the policy tells to lose (or not) grant the corresponding access rights. For example, Barua et al. [28] proposes an ABE-based access control model based on patience and privacy protection; Chen et al. [29] described a new framework with a cloud-based, privacy-aware Role-Based Access Control model that can be used for control, data traceability, and access allowed access to healthcare data resources. Methods for applying the Access Control model are also introduced for dynamic policies [30], [31] or protection policies for both security and privacy [32].

In addition, Madine et al. [33] has introduced a Smart Contract-based system that provides patients with reliable, traceable and secure control over their medical data (i.e., which is stored non-invasively). concentrate). To increase the security and privacy of medical data, they used the decentralized storage feature of the interplanetary file system (IPFS) to store and share patient medical data safely. For practical applications, HealthBank has proposed a healthcare system and surrounding ecosystems that allow users (i.e., patients) to manage and control their data.[2] This solution is recommended to be able to comply with strict security and privacy regulations (e.g., GDPR) and to assist users in using their services. In addition, the system also proposes solutions for storing personal data with complex data encryption algorithms, immutability and

accountability. Similarly, HealthNautica and Factom Announce Partnership have used blockchain technology to ensure the integrity of patient medical data while providing transparency based on blockchain technology and encryption of sensitive data (e.g., personal information, health status).[3]

With the same approach based on Blockchain technology and IPFS, Misbhauddin et al. [34] introduced the MedAccess platform, A Scalable Architecture for Blockchain-based Health Record Management. The platform supports on-chain storage and processing allowing doctors, lab technicians and patients to securely manage medical records. However, these systems face some problems in the processing and storage of personal data. Specifically, Le et al. [35] has argued that not all data collected must be processed on-chain. Instead, Son et al. [2] argues that personal data that is either not directly related to treatment or diagnosis may be stored off-chain (i.e., offchain). Similar to the above approach, to increase the processing capacity for the whole system, Zyskind et al. [36] presented an approach based on in-chain and out-of-chain processing. Onchain processes require all entities of a typical personnel management system, where patient and medical staff information is stored; in contrast, encrypted medical data is stored on a separate centralized storage server to enable faster access and low cost. However, the above methods have major limitations, including that any information that is validated must be executed on-chain instead of local processing. This only benefits storage but does not change data handling (i.e., since all information still executes on-chain) [17].

To solve the on-chain storage problem, Zhang et al. [37] have proposed FHIRChain (Fast Healthcare Interoperability Resources), a blockchain-based system that allows patients to securely share their clinical data in a medical setting. For this approach, users are allowed to share their personal data directly with hospitals and medical centers instead of having all their personal information stored directly on the banana. Another approach suggested by Patel et al. [38] has empowered hospitals to be the creators of medical records and patients to be owners of their records. In this approach, all medical data processing and updating are done off-chain (e.g., medical record sharing, and all query requests to patient data). In addition, the issue of threading while exporting medical data is also very important because data stored on the same system can be accessed by a malicious user on the same system [39], [14]. Therefore, Iryo is introduced as a healthcare ecosystem that uses blockchain technology to decentralize access to medical records.[4] Specifically, it uses the NuCypher KMS key management system (i.e., [40]) to address the limitations of adopting the peer2peer model for storing and executing on encrypted data. Also adopting an advanced cryptographic-based approach and blockchain technology, Chen et al. [41] proposed a system that only stores the searchable index of records on the blockchain. The patient information is organized as Key-Value. Where "key" contains records presented as hash (i.e., reduced index) while actual patient data (i.e., "value") is encrypted and stored on a public cloud server.

Tith et al. [42] proposed a system based on blockchain technology to ensure privacy, scalability, and availability of

---

[2]https://www.healthbank.coop/2018/10/30/healthbank-creates-the-first-patient-centric- healthcare-trust-ecosystem/

[3]https://www.factom.com/company-updates/healthnautica-factom-announce-partnership/

[4]$https://iryo.network/iryo_whitepaper.pdf$

patient data in the medical environment combined with data encryption methods. patient data using the public key. It uses a proxy re-encryption mechanism on a centralized server to transfer encrypted data from the patient to the doctor. In this solution, the patient-centered aspect is still lacking because the medical records are under the control of the hospitals. Another limitation is that the re-encryption is done on a single server. During the Covid-19 pandemic, there are several approaches to exploiting blockchain technology as a workaround in providing telemedicine support services (e.g., [43]). Specifically, they proposed a blockchain-based model to solve 5 problems at that time, including i) Managing patient consent for unwanted accesses from service providers health care service/center; ii) Traceability of remote diagnoses and treatments; iii) Traceability of home medical supplies and equipment; iv) Secure access to individual health records (i.e., combined with an IPFS-based approach); and v) Automated billing for all telehealth services (e.g., drug bills).

Instead of focusing only on the two main target groups in the medical environment, patients and staff at the healthcare facility (i.e., nurses, doctors), Kassab et al. [44] has expanded its blockchain-based medical data processing and storage system to include (insurance companies and regulatory agencies). In addition, the processed data is also extended to the supply chain of equipment and drugs from the suppliers/hosting agencies to the respective hospitals and pharmacies. However, the role of the patient is not mentioned as a major contribution of the paper. Similar to Xiao et al. [45] has proposed Healthcare Data Gateway (HGD) that allows patients to easily and securely own, control, and share their own data without violating privacy. This article presents a direction to combine Blockchain and Machine learning systems to achieve the system's ability to quickly handle "emergency" situations and ensure the privacy of medical data. [46] proposes a review, demonstrating how the inherent properties of the blockchain (e.g., synchronous processing, decentralized storage) can enhance or hinder current healthcare systems in improving healthcare services in healthcare facilities.

However, the above approaches (i.e., state-of-the-art) have brought many solutions to today's traditional healthcare systems. But those approaches only consider the general problem rather than consider a specific area (i.e., country). Thereby, in developing countries (e.g., Vietnam) where medical equipment and supplies are one of the barriers that directly affect people's healthcare process. In addition, the above approaches require a certain knowledge of information technology as well as the risks related to security and privacy. It is for the above reasons that a few case studies (i.e., applied to a specific geographical area - country, region) address the upper limits of [47]. In this article, we provide Blockchain-based support for the management of test results in medical centers.

III. The Blockchain-based Medical Test Results Management System

A. Traditional Model

To build the traditional model, we conducted a survey of 10 hospitals and medical centers in Can Tho city. We act as the medical check-up patient and the interviewer taking information directly from the test patients. We also surveyed
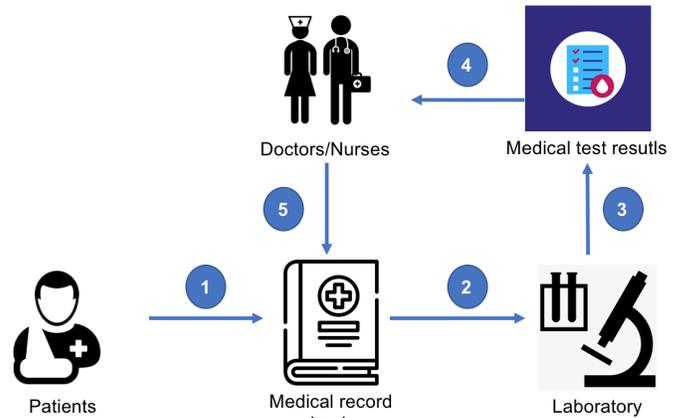

Fig. 1. The Traditional Process for Medical Test Results Management System.

medical opinions on the diagnosis process of a doctor who used to work at Children's Hospital - Can Tho. The traditional model surveys were conducted from June to August 2022. All the collected procedures are evaluated and graded step-by-step by the team before building the medical test results management model as shown in Fig. 1.

Fig. 1 shows the basic steps of the traditional medical test results management process. This model describes the five main steps, excluding the risks of losing medical test results. In other words, this process will be repeated for each patient whenever they have a routine checkup or a health-related reason. Specifically, in the first step, the patient registers for a medical test result, which includes basic information about the patient, such as full name, address, phone number, or medical condition. The medical test result number is also the patient number at that hospital. In the second step, the patient brings the medical test result to a specialist at the hospital, called a laboratory (e.g., eye, blood, urine) for sampling. This procedure requires a very long wait time from the patient. The patient then receives information about this form in the third step before forwarding this information to the doctors and nurses for consultation in the fourth step. Finally, the consultation results are updated in the medical test result of the patient in the fifth one.

For the process of storing patient information, the storage of their information is completely manual. Only a few major medical centers in major cities support the storage of medical results on their centralized database. This demonstrates that it is not feasible to share a patient's medical result between different healthcare facilities. It is easy to see that there are many inconveniences for both patients and hospital staff when using the current testing/receiving process, respectively. The first limit comes from the patient, all information stored on the medical test result must be ensured carefully, and the medical test result must not be lost otherwise, all procedures will have to be repeated from the beginning with a new medical test result. Changing the place of treatment/examination is extremely difficult because the patient has to bring the medical test results issued at the previous medical facility to a new one. In addition, the loss of medical test results is extremely risky, besides the reason for having to repeat the entire sampling

process, since they relate to the diagnosis process. Regarding the responsibility of physicians (i.e., doctors/nurses), they must reread a patient's entire medical history each time their patient has a follow-up visit. This is similar to the process of examining a new patient.

B. Proposed Model

To solve the above problems, we introduce a model based on Blockchain technology, where all information related to the testing process and the storage of patient's medical test results are updated and shared freely in the healthcare environment. Fig. 2 shows our proposal system based on Blockchain technology and distributed ledger (i.e., Distributed Ledger). As a first step, the patient initializes a global ID for not only a certain healthcare facility but also for others ones (e.g., the hospital in the same city). Unlike the traditional process, in another word, this ID will identify the user globally, which means that the patient can be examined at another medical facility without affecting the diagnosis process. Specifically, doctors/nurses can retrieve information about a patient's medical history based on their global ID (this will be covered in more detail in the next steps). From the initial global ID, users can generate more than 1 medical test result (i.e., per medical facility or healthcare service). These records store all test results and related patient information (i.e., similar to a medical test result in the paper). The data stored on the medical test result is always updated to Distributed Ledger (step 3). Users will then go to the respective Laboratories to take samples (step 4) before seeing a doctor in person to receive advice on their health status (step 6). This is the biggest difference between our model compared to the traditional model. Patients do not need to wait a long time at the facility; instead, an appointment is delivered to their device (e.g., smartphone) whenever their result is available. Meanwhile, the remaining steps will be executed independently at the system under the confirmation of the relevant parties. Specifically, after testing, the results are updated to the Distributed Ledger, and this information includes the user's corresponding medical test results and metadata about the time and location of the test as well as the doctors participating in the consultation. In case the patient goes to another medical facility, the patient's permission (or the patient's family member's/relatives in some special cases) must be obtained before accessing the patient's medical data (i.e., over-privileged permission). After receiving the request from the system, the doctors will enter the diagnostic results into the system (i.e., Distributed Ledger). The whole process will be confirmed by the stakeholders during the execution. The data will be encrypted when there is no request for access or update from the relevant parties (e.g., patient, nurse, doctor). The next section presents our approach based on Hyperledger Fabric.

IV. IMPLEMENTATION

A. Permission Diagram

Fig. 3 presents the working mechanism of the request authentication process in this paper. Specifically, we built two organizations with corresponding encrypted material certificates, each organization includes two users and two peers. Each peer is responsible for maintaining the version of the

ledger so that the network and data can be maintained even if other peers are shut down.

When the user initiates a request and sends it to the service. The back-end service processes the data and sends it to the smart contract API. When receiving the request and the data, the smart contract sends this to the peers in the network for authentication and data interaction purposes. During the creation, querying or updating data processes, peers check the identity of the request to decide whether to allow access to the data at the distributed ledger. If the identified user of the request is not defined in the data collection, the system denies access and sends a message to the back-end API to notify the user; the system allows access and proceeds with further processing steps.[5]

B. Hyperledger Component

The model in this paper is implemented on the Hyperledger Fabric platform. Fabric is a permissionless blockchain platform that integrates smart contracts, the storage of data to the distributed ledger is controlled through the smart contract APIs, from which the data is simplified and easily traced. Each request that goes through the smart contract is verified with public and private key pairs. In other words, if the user does not exist in the system, the system is better protected from malicious requests outside the system.

The Fabric system in this paper includes two organizations. Each organization consists of 2 peers to store smart contracts, where each peer registers two users and is authenticated with public and private key pairs. The components of the model are shown in Fig. 4

When user devices access the system to initiate/query or update data for a particular transaction, requests are sent from the client to the services of the existing system. Then, these services send access information to the peers belonging to the organization located in the blockchain network. At this step, the peers conduct verification of that user's key pairs, and if the successful peer authentication process proceeds to send information to the smart contract with the transaction type declared in a smart contract requested by the user, the smart contract will go through the designed features function to access the distributed ledger to initiate/query or update specific data.

C. Our Proposed Model's Diagram

One of the most important parts of the model lies in the validation and interaction with the patient's global ID and their medical data described in Fig. 5 and 6. In particular, the main functions include initializing and querying the patient's global ID and their medical data.

Fig. 5 depicts the process of storing new record data (e.g., patients' global ID and their medical data). In step 1, when the user initializes information about a certain ID, the data is sent to the back-end service of the health center's information management system. In the next step, the back-end APIs (i.e., backend) check, authenticate, and initialize the

---

[5]For more detail of the basic Blockchain following https://ethereum.org/en/whitepaper/ and https://www.hyperledger.org/learn/white-papers for Ethereum and Hyperledger Fabric, respectively

Fig. 2. The Proposed Model for Blockchain-Based Medical Test Results Management System.
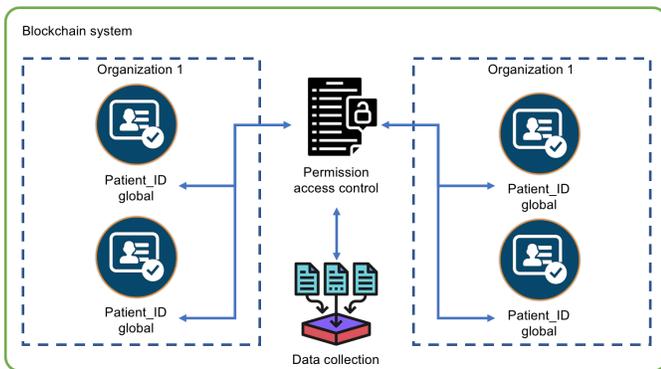

Fig. 3. Permission Diagram.


Fig. 5. Initializing and Storing the New Data.

Fig. 4. Hyperledger Fabric Component.

null field data.

Fig. 6 presents the process of retrieving data of a particular (e.g., patients' global ID and their medical data). When the user sends a query request to the system, the service query data checks and confirms whether the parameter ID of their medical data exists or not. Then, the smart contract's APIs are called and passed into the corresponding parameter. Next, the smart contract's APIs check for the existence of data in the request before querying. In the case that the ID does not exist, the smart contract sends an error notification to the user's device; otherwise, it returns the relevant data/record of the patient corresponding to the requested ID.

V. EVALUATION SCENARIOS

A. Environment Setting

Our paradigm is deployed on the Hyperledger Fabric network maintained inside docker containers. In this section, we measure the performance of chaincode in the two scenarios: initializing (i.e., creating data) and accessing data. The experiments are deployed on Ubuntu 20.01 configuration, core i5 2.7Ghz, and 8GB RAM.

default values, then pass the parameters to the API inside the smart contract. At this point, a smart contract transfers data and stores transactions to the distributed ledger of the blockchain network. The default values for parameters sent from the request are intended to minimize errors caused by

Fig. 6. Retrieving/Querying Data Process.

#Request/second	Success	Fail	Percentage of the fail requests
1000	38506	10524	21.4644%
2000	37289	9918	21.0096%
3000	38754	8128	17.3371%
4000	35208	14705	29.4613%
5000	36699	12872	25.9668%
6000	37098	11833	24.1830%
7000	38405	10847	22.0235%
8000	37769	11535	23.3957%
9000	37392	12456	24.9880%
10000	37852	11516	23.3269%

Fig. 8. Data Access (i.e., Retrieving/Querying) for the Medical Test Results of the Patient in Terms of Latency.


Fig. 7. Data Creation/Initialization for the Medical Test Results of the Patient in Terms of Latency.

To prove the effectiveness of our model, we also define several experiments by exploiting the Hyperledger Caliper[6] that is used to design the test scenarios and collect all the information regarding the performance.

B. Evaluation Results

As introduced in the whole article, we consider the three main execution tasks for the blockchain-based system, namely data creation, data access, and data update. The three following sub-section will target our evaluation of the ten scenarios in terms of the supported performance of the system.

1) Data Creation: In this scenario, the study measures the performance of the data initialization function/data created (e.g., medical record book) performed through smart contracts. The number of requests sent simultaneously from two users[7]. Table I shows the execution results of the data initialization/creation function in terms of the success and fail requests. The data initialization/creation script is conducted with two users concurrently making 1000 - 10000 requests to the system. Based on the execution results in Table I, it can be seen that the number of successful and failed requests is stable (except in the case of 10000 requests/second). Specifically, the number of failed requests is limited to less than 15,000 requests (i.e., on average 11.4K requests - 23.32%). Meanwhile, the

lowest case was with only 17.34% (8,128 requests at 3K requests/second). The maximum of failed request rate is in the fourth row 4,000/s request with 14,705 requests - 29.46%. In contrast, the system is more stable in terms of data creation with only an average of approximately 37.5K success requests for each scenario (from 1K requests to 10K).

We also measure the parameters of the system latency for each request (i.e., max, min, avg) in Fig. 7. In particular, for system-wide latency, we recorded the number of requests with response delays per 1,000 requests/second to 10,000 requests per second. The data in Fig. 7 demonstrate that the highest latency ranges from 1,461.58 to 1,642.91 seconds. The minimum is less than 10 seconds. The average delay when creating new data is less than 940 seconds. This is acceptable because creating thousands of new records at the same time is very unlikely in medical centers. The results observed in this scenario also demonstrate that the system supports very well with the continuous generation of new profiles.

2) Data Access (Retrieving/Querying): In the second experiment, we consider the data access (e.g., medical record book). We also set up 6 scenarios from 1000 to 10000 requests which access the medical record book from two users. Table II shows the execution results of the data access function (e.g., medical record book). Compared with the creation tasks, the results of 10 scenarios to evaluate the data accessibility of our proposed blockchain-based system are more balanced. Given the number of successful and failed requests, we also collect the number of requests at every 1000 to 6000 requests per second. The number of successful and failed requests is fairly balanced, with only 21 fail request for 10 scenarios (especially without the fail request at the seventh and ninth scenarios) around 99.99% of the requests are successful in all 6 scenarios.

---

[6]https://www.hyperledger.org/use/caliper

[7]We set up one organization with two users and two peers

TABLE II. Data Access (Retrieving/Querying) for the Medical Test Results of the Patient in Terms of the Success and Fail Requests

#Request/second	Success	Fail	Percentage of the fail requests
1000	104435	2	0.0019%
2000	105102	3	0.0029%
3000	105103	1	0.0010%
4000	105294	4	0.0038%
5000	105348	4	0.0038%
6000	105373	1	0.0009%
7000	105119	0	0.0000%
8000	105222	1	0.0010%
9000	105104	5	0.0048%
10000	105122	0	0.0000%

TABLE III. Data Edit/Update for the Medical Test Results of the Patient in Terms of the Success and Fail Requests

#Request/second	Success	Fail	Percentage of the fail requests
1000	23018	23509	50.5277%
2000	21503	25012	53.7719%
3000	22642	25497	52.9654%
4000	22042	24782	52.9258%
5000	24547	23376	48.7782%
6000	22499	24091	51.7085%
7000	21775	23820	52.2426%
8000	22690	22972	50.3088%
9000	22496	23656	51.2567%
10000	25557	20470	44.4739%

Retrieval of stored data is extremely important. Indeed, considering health data retrieval time directly affects the patient's health care. To solve this problem, we consider the latency of the system (i.e., the maximum/average/minimum time it takes to process the request of data accessed from the system) which presented in Fig. 8. Specifically, the maximum time to wait for a data retrieval request is 0.6 seconds at the third scenario with on average 0.513 seconds. [8] The minimum wait time is almost instant response (i.e., with only 0.01 seconds for the whole scenarios - from 1K to 10K requests/second). The average time for each data retrieval request is less than 0.1 seconds - between 0.08 and 0.1 seconds.

3) Data Edit/Update: Finally, we look at the user's ability to update the medical test result's data. This parameter reflects whether a doctor or nurse updates information about a patient's medical record (e.g., new symptoms, diagnoses). In this scenario, we also conduct a review of 10 different scenarios, each of which will require processing from 1000 to 10000 requests per second. We also measure two parameters, similar to the two scenarios above, the number of successful and failed requests in Table III and the overall latency, which is shown Fig. 9.

For the first aspect (i.e., #request for success and fail), The number of failed requests was also higher than the success requests in all 10 scenarios (with an average of about 51%). On average, there are 22,876.9 request is success and this amount for fail requests is 23,718.5. Generally, the number of fail requests in data update task is higher than that in the two above tasks (see section VI for the details of reason).

For the latency aspect of the data update requirement, we
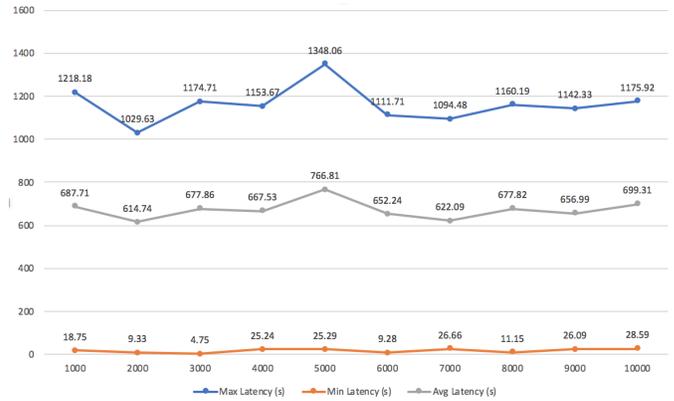
Fig. 9. Data Edit/Update for the Medical Test Results of the Patient in Terms of Latency.

present the collected performance in Fig. 9. In terms of time for execution, updating data is more complex than the previous two previous tasks (i.e., creation and access data). Specifically, we must determine if user information exists in Distributed ledger, then we determine which information needs updating (e.g., symptoms, disease diagnoses). Because of the above requirement, the execution time for the update task is longer. Specifically, the latency of all scenarios ranges from 1029.63 to 1348.06 seconds in the maximum case. The minimum latency ranges from 4.75 seconds to 28.59 seconds, while the average latency required by an application to process ranges from 614.74 seconds to 766.81 seconds.

VI. Discussion

Comparing all three evaluation scenarios, we find that real-time is acceptable. We also describe why there is a difference between the time lag in the execution of requests from the system depending on the complexity of the query. Specifically, the most prolonged time lag was recorded in data initialization due to updating to Hyperledger. This is different from the traditional way of storing data, where the information is only stored in tables and is done by the system administrator. On the contrary, initiating a medical test result requires confirmation from all relevant parties. In addition, defining constraints in an update request is more complex than in a retrieval request. The update time clearly defines the information the requester wishes to add/update to the existing medical test results. Finally, the fastest execution time is the data retrieval request which offers more promise for a Blockchain-based system than traditional storage systems.

However, Section V provides a marked change in all three data creation, retrieval, and update scenarios regarding the number of success and failed requests. Specifically, in the update scenario, the failure rate of requests is much higher than in the data initialization scenario (with more than 50% compared to less than 20%). A similar method occurs when comparing initialization/data creation and access/data retrieval with more than 20% and nearly 0% of failed requests, respectively. This happens because we build a system that simulates the interactions between the parties (e.g., patient, nurse, doctor). In particular, the update and retrieval request must require the data to be initialized before. Otherwise, the request is

considered a failure. For the update scenario, the system also requires that the updated information be initialized before being replaced with new data (e.g., patient information and medical history). Initializing a dummy data system according to the above requirements is extremely difficult because we do it on two separate user groups.

For the system specification, we have not included encryption and decryption times for the data stored on Hyperledger. We assume that a trusted third party will take care of this. In terms of execution time, including the user critical generation time, as well as encryption and decryption, will increase the execution time for the whole system. This is hard to meet on our simulation system. In addition, this proposed model is also the first attempt to build a blockchain-based system that aims to offer a test management model in medical centers in developing countries.

For future work, we intend many potential research directions to follow after this work. One of the mandatory requirements for health systems is confidentiality (i.e., authentication and authorization). We apply the proposed models based on the dynamic data support the environment of IoT devices [11], [48]. For authorization, a model based on ABAC [31], [30] and supporting dynamic policy [49], [50] is an appropriate choice in the context of the current health system. For encryption requirements, we use a trusted authority that provides a solution to store and protect patient data on Hyperledger [51]. Moreover, we plan to combine the NFT and IPFS to define the medical test result for the patients.

VII. Conclusion

We have built a blockchain-based system for healthcare facilities in developing countries (i.e., Vietnam) to manage and store patient medical test results. Our proposed model is based on a balance between many unique criteria including limitations of medical facilities in developing countries (i.e., in terms of equipment), patient effort (i.e., waiting time, requirement background about security and privacy as well as technique), sharing information about test results and patient medical records between different healthcare centers; and system transparency. In addition, we build a proof-of-concept based on the Hyperledger Fabric platform. To demonstrate the feasibility of the proposed model, we evaluate the actual w.r.t performance of the system (i.e., create, access and update) based on 10 scenarios that change the number of requests per second (i.e., 1000 requests increaing for each scenario) by exploiting Hyperledger Caliper. The review highlighted our findings based on a review of the system as well as of the proof-of-concept; thereby suggesting possible future development directions.

References

[1] R. A. McPherson and M. R. Pincus, *Henry's clinical diagnosis and management by laboratory methods E-book.* Elsevier Health Sciences, 2021.

[2] H. T. Le, L. N. T. Thanh, H. K. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, K. H. N. Vuong, H. X. Son *et al.*, "Patient-chain: Patient-centered healthcare system a blockchain-based technology in dealing with emergencies," in *International Conference on Parallel and Distributed Computing: Applications and Technologies.* Springer, 2022, pp. 576–583.

[3] L. N. T. Thanh *et al.*, "Toward a unique iot network via single sign-on protocol and message queue," in *International Conference on Computer Information Systems and Industrial Management.* Springer, 2021.

[4] N. Mostert-Phipps, D. Pottas, and M. Korpela, "Improving continuity of care through the use of electronic records: a south african perspective," *South African Family Practice*, vol. 54, no. 4, pp. 326–331, 2012.

[5] D. M. Mugo and D. Nzuki, "Determinants of electronic health in developing countries," 2014.

[6] H. X. Son and E. Chen, "Towards a fine-grained access control mechanism for privacy protection and policy conflict resolution," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, 2019.

[7] N. Duong-Trung, H. X. Son, H. T. Le, and T. T. Phan, "Smart care: Integrating blockchain technology into the design of patient-centered healthcare systems," in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*, ser. ICCSP 2020, 2020, p. 105–109.

[8] ——, "On components of a patient-centered healthcare system using smart contract," in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*, 2020, p. 31–35.

[9] H. X. Son, T. H. Le, N. T. T. Quynh, H. N. D. Huy, N. Duong-Trung, and H. H. Luong, "Toward a blockchain-based technology in dealing with emergencies in patient-centered healthcare systems," in *International Conference on Mobile, Secure, and Programmable Networking.* Springer, 2020, pp. 44–56.

[10] L. N. T. Thanh, N. N. Phien, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, H. X. Son *et al.*, "Sip-mba: A secure iot platform with brokerless and micro-service architecture," 2021.

[11] N. T. T. Lam, H. X. Son, T. H. Le, T. A. Nguyen, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, and H. V. K. Nguyen, "Bmdd: A novel approach for iot platform (broker-less and microservice architecture, decentralized identity, and dynamic transmission messages)," *International Journal of Advanced Computer Science and Applications*, 2022.

[12] T. T. L. Nguyen, H. K. Vo, H. H. Luong, H. T. K. Nguyen, A. T. Dao, X. S. Ha *et al.*, "Toward a unique iot network via single sign-on protocol and message queue," in *International Conference on Computer Information Systems and Industrial Management.* Springer, 2021, pp. 270–284.

[13] L. N. T. Thanh, N. N. Phien, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, H. X. Son *et al.*, "Uip2sop: a unique iot network applying single sign-on and message queue protocol," 2021.

[14] N. Duong-Trung, X. S. Ha, T. T. Phan, P. N. Trieu, Q. N. Nguyen, D. Pham, T. T. Huynh, and H. T. Le, "Multi-sessions mechanism for decentralized cash on delivery system," *Int. J. Adv. Comput. Sci. Appl*, vol. 10, no. 9, 2019.

[15] C. Frazzoli, O. E. Orisakwe, R. Dragone, and A. Mantovani, "Diagnostic health risk assessment of electronic waste on the general population in developing countries' scenarios," *Environmental Impact Assessment Review*, vol. 30, no. 6, pp. 388–399, 2010.

[16] H. T. Le, T. T. L. Nguyen, T. A. Nguyen, X. S. Ha, and N. Duong-Trung, "Bloodchain: A blood donation network managed by blockchain technologies," *Network*, vol. 2, no. 1, pp. 21–35, 2022.

[17] N. T. T. Quynh, H. X. Son, T. H. Le, H. N. D. Huy, K. H. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, N. Duong-Trung *et al.*, "Toward a design of blood donation management by blockchain technologies," in *International Conference on Computational Science and Its Applications.* Springer, 2021, pp. 78–90.

[18] A. Roehrs, C. A. da Costa, R. da Rosa Righi, V. F. da Silva, J. R. Goldim, and D. C. Schmidt, "Analyzing the performance of a blockchain-based personal health record implementation," *Journal of biomedical informatics*, vol. 92, p. 103140, 2019.

[19] X. S. Ha, T. H. Le, T. T. Phan, H. H. D. Nguyen, H. K. Vo, and N. Duong-Trung, "Scrutinizing trust and transparency in cash on delivery systems," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage.* Springer, 2020, pp. 214–227.

[20] X. S. Ha, H. T. Le, N. Metoui, and N. Duong-Trung, "Dem-cod: Novel access-control-based cash on delivery mechanism for decentralized marketplace," in *2020 IEEE 19th International Conference on Trust,*

Security and Privacy in Computing and Communications (TrustCom). IEEE, 2020, pp. 71–78.

[21] Z. Chen, W. Xu, B. Wang, and H. Yu, "A blockchain-based preserving and sharing system for medical data privacy," *Future Generation Computer Systems*, vol. 124, pp. 338–350, 2021.

[22] M. Du, Q. Chen, J. Xiao, H. Yang, and X. Ma, "Supply chain finance innovation using blockchain," *IEEE Transactions on Engineering Management*, vol. 67, no. 4, pp. 1045–1058, 2020.

[23] H. X. Son, M. H. Nguyen, H. K. Vo *et al.*, "Toward an privacy protection based on access control model in hybrid cloud for healthcare systems," in *International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on EUropean Transnational Education (ICEUTE 2019).* Springer, 2019, pp. 77–86.

[24] M. R. Patra, R. K. Das, and R. P. Padhy, "Crhis: cloud based rural healthcare information system," in *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance*, 2012, pp. 402–405.

[25] C. O. Rolim, F. L. Koch, C. B. Westphall, J. Werner, A. Fracalossi, and G. S. Salvador, "A cloud computing solution for patient's data collection in health care institutions," in *2010 Second International Conference on eHealth, Telemedicine, and Social Medicine.* IEEE, 2010, pp. 95–99.

[26] T. Makubalo, B. Scholtz, and T. O. Tokosi, "Blockchain technology for empowering patient-centred healthcare: A pilot study," in *Conference on e-Business, e-Services and e-Society.* Springer, 2020, pp. 15–26.

[27] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-cps: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Systems Journal*, vol. 11, no. 1, pp. 88–95, 2015.

[28] M. Barua, X. Liang, R. Lu, and X. Shen, "Espac: Enabling security and patient-centric access control for ehealth in cloud computing," *International Journal of Security and Networks*, vol. 6, no. 2-3, pp. 67–76, 2011.

[29] L. Chen and D. B. Hoang, "Novel data protection model in healthcare cloud," in *2011 IEEE International Conference on High Performance Computing and Communications.* IEEE, 2011, pp. 550–555.

[30] N. M. Hoang and H. X. Son, "A dynamic solution for fine-grained policy conflict resolution," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 116–120.

[31] H. X. Son and N. M. Hoang, "A novel attribute-based access control system for fine-grained privacy protection," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 76–80.

[32] Q. N. T. Thi, T. K. Dang, H. L. Van, and H. X. Son, "Using json to specify privacy preserving-enabled attribute-based access control policies," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage.* Springer, 2017, pp. 561–570.

[33] M. M. Madine, A. A. Battah, I. Yaqoob, K. Salah, R. Jayaraman, Y. Al-Hammadi, S. Pesic, and S. Ellahham, "Blockchain for giving patients control over their medical records," *IEEE Access*, vol. 8, pp. 193 102–193 115, 2020.

[34] M. Misbhauddin, A. AlAbdulatheam, M. Aloufi, H. Al-Hajji, and A. Al-Ghuwainem, "Medaccess: A scalable architecture for blockchain-based health record management," in *2020 2nd International Conference on Computer and Information Sciences (ICCIS).* IEEE, 2020, pp. 1–5.

[35] N. T. T. Le, Q. N. Nguyen, N. N. Phien, N. Duong-Trung, T. T. Huynh, T. P. Nguyen, and H. X. Son, "Assuring non-fraudulent transactions in cash on delivery by introducing double smart contracts," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, pp. 677–684, 2019.

[36] G. Zyskind, O. Nathan *et al.*, "Decentralizing privacy: Using blockchain to protect personal data," in *2015 IEEE Security and Privacy Workshops.* IEEE, 2015, pp. 180–184.

[37] P. Zhang, J. White, D. C. Schmidt, G. Lenz, and S. T. Rosenbloom, "Fhirchain: applying blockchain to securely and scalably share clinical data," *Computational and structural biotechnology journal*, vol. 16, pp. 267–278, 2018.

[38] V. Patel, "A framework for secure and decentralized sharing of medical imaging data via blockchain consensus," *Health informatics journal*, vol. 25, no. 4, pp. 1398–1411, 2019.

[39] H. X. Son, M. H. Nguyen, N. N. Phien, H. T. Le, Q. N. Nguyen, V. Dinh, P. Tru, and P. Nguyen, "Towards a mechanism for protecting seller's interest of cash on delivery by using smart contract in hyperledger," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 4, pp. 45–50, 2019.

[40] M. Egorov, M. Wilkison, and D. Nuñez, "Nucypher kms: decentralized key management system," *arXiv preprint arXiv:1707.06140*, 2017.

[41] L. Chen, W.-K. Lee, C.-C. Chang, K.-K. R. Choo, and N. Zhang, "Blockchain based searchable encryption for electronic health record sharing," *Future generation computer systems*, vol. 95, pp. 420–429, 2019.

[42] D. Tith, J.-S. Lee, H. Suzuki, W. Wijesundara, N. Taira, T. Obi, and N. Ohyama, "Application of blockchain to maintaining patient records in electronic health record for enhanced privacy, scalability, and availability," *Healthcare informatics research*, vol. 26, no. 1, pp. 3–12, 2020.

[43] R. W. Ahmad, K. Salah, R. Jayaraman, I. Yaqoob, S. Ellahham, and M. Omar, "The role of blockchain technology in telehealth and telemedicine," *International journal of medical informatics*, vol. 148, p. 104399, 2021.

[44] M. Kassab, J. DeFranco, T. Malas, P. Laplante, G. Destefanis, and V. V. G. Neto, "Exploring research in blockchain for healthcare and a roadmap for the future," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 4, pp. 1835–1852, 2019.

[45] X. Yue, H. Wang, D. Jin, M. Li, and W. Jiang, "Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control," *Journal of medical systems*, vol. 40, no. 10, pp. 1–8, 2016.

[46] M. Kassab, J. DeFranco, T. Malas, G. Destefanis, and V. V. G. Neto, "Investigating quality requirements for blockchain-based healthcare systems," in *2019 IEEE/ACM 2nd International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB).* IEEE, 2019, pp. 52–55.

[47] K. L. Quoc, H. K. Vo, L. H. Huong, K. H. Gia, K. T. Dang, H. L. Van, N. H. Huu, T. N. Huyen, L. Van Cao Phu, D. N. T. Quoc *et al.*, "Sssb: An approach to insurance for cross-border exchange by using smart contracts," in *International Conference on Mobile Web and Intelligent Information Systems.* Springer, 2022, pp. 179–192.

[48] H. H. Luong, T. D. Anh, K. N. H. Tuan, and H. X. Son, "Ioht-mba: An internet of healthcare things (ioht) platform based on microservice and brokerless architecture," 2021.

[49] S. H. Xuan, L. K. Tran, T. K. Dang, and Y. N. Pham, "Rew-xac: an approach to rewriting request for elastic abac enforcement with dynamic policies," in *2016 International Conference on Advanced Computing and Applications (ACOMP).* IEEE, 2016, pp. 25–31.

[50] H. X. Son, T. K. Dang, and F. Massacci, "Rew-smt: a new approach for rewriting xacml request with dynamic big data security policies," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage.* Springer, 2017, pp. 501–515.

[51] M. Uddin, "Blockchain medledger: Hyperledger fabric enabled drug traceability system for counterfeit drugs in pharmaceutical industry," *International Journal of Pharmaceutics*, vol. 597, p. 120235, 2021.