

Arabic Dialogue Processing and Act Classification using Support Vector Machine

Abraheem Mohammed Sulayman Alsubayhay, Md Sah Hj Salam, Farhan Bin Mohamed
School of Computing-Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Johor, Malaysia

Abstract—Text classification is the technique of grouping documents according to their content into classes and groups. As a result of the vast amount of textual material available online, this procedure is becoming increasingly crucial. The primary challenge in text categorization is enhancing classification accuracy. This role is receiving more attention due to its importance in the development of these systems and the categorization of Arabic dialogue processing. In the research, attempts were made to define dialogue processing. It concentrates on classifying words that are used in dialogue. There are various types of dialogue processing, including hello, farewell, thank you, confirm, and apologies. The words are used in the study without context. The proposed approach recovers the properties of function words by replacing collocations with standard number tokens and each substantive keyword with a numerical approximation token. With the use of the linear support vector machine (SVM) technique, the classification method for this study was obtained. The act is classified using the linear SVM technique, and the anticipated accuracy is evaluated against that of alternative algorithms. This study encompasses Arabic dialogue acts corpora, annotation schema, and classification problems. It describes the outcomes of contemporary approaches to classifying Arabic dialogue acts. A custom database in the domains of banks, chat, and airline tickets is used in the research to assess the effectiveness of the suggested solutions. The linear SVM approach produced the best results.

Keywords—Dialogue processing; Act; Arabic language; linear support vector machine; without cue

I. INTRODUCTION

Arabic is recognized as a challenging dialect by non-native speakers of the language in the field of automated language processing. The recognized dialogue acts provide accurate information when the user asks questions and remain silent are typically utilized as an input to the Dialogue Manager component to assist the system decide what to do next. Depending on the dialogue system domain, the Dialogue Acts taxonomy varies [1]. To identify speech acts in dialogues, a variety of techniques have been employed, ranging from rule-based approaches to deep learning and machine learning techniques. To overcome the lack of appropriate training data, novel language interpretation methods have been created [2].

The area of artificial intelligence is employed rarely. Several computer researchers and scientists are engaged in the humanities in order to create computer programmes that accurately scan Arabic writings and convert them into digital formulas. This is because Arabic and artificial intelligence are closely related fields of study. With the advent of new

approaches like Intersent, Elmo, wordszvec and Use, AI was impacted by the introduction of some sophisticated methods for presenting the semantic meaning of texts [3]. In text extraction and processing, conventional ML techniques are typically employed. These techniques include Naïve Bayes, Stochastic Gradient Descent (SGD), and SVM with a linear kernel. The STS problem is addressed using these techniques. Utilizing statistical-based, string and character-based and distance-based similarity metrics, several features are extracted and utilized [4]. The several perspectives for a specific speech sound have typically been represented by statistical methods like Gaussian mixture frameworks [5]. These methods can be employed in a variety of tasks since they are in fact rich in mathematical techniques.

The simplest language communication unit, or dialogue acts (DA) are those that convey the intention of the speakers. Automated dialogue act identification is a crucial indicator for many activities, including subject detection, machine translation, summarization, dialogue systems and human conversation interpretation. Annotation and segmentation are the two key subtasks in dialogue act detection [6]. These two phases can be completed separately—segmentation comes first, then annotation is joined in a single stage. In the context for recognizing dialogue acts, the annotation task is crucial. A label describing the user's intention for every segmented statement during the dialogue is assigned. ML techniques have been tested in studies to detect new DAs. Linear methods and vector-based methods are examples of supervised modeling techniques that are often employed [7]. Limited vocabulary limits are utilized to achieve high speech recognition rates. In reality, it is thought that this outcome is adequate for the majority of voice activation devices to be implemented. As a result, both the learning time and error rate are gradually reduced. Additionally, this rate is plainly erratic and is influenced by lexicon and language. Nevertheless, the creation of advanced ASR systems with increased speech recognition rates has emerged as an intriguing topic of interest for all academics working in the field of speech recognition. In fact, a number of categorization and parametric techniques have surfaced to complete the task. Various categorization techniques, including Gaussian mixture model, Decision tree, machine learning via ANN, K-Nearest Neighbor, Dynamic Time Warping (DTW) and HMM, have been utilized in a number of publications for isolated-words identification systems [8]. But these are inefficient and consume time while attaining the results.

In the Speech Recognition area, computer scientists provided particular consideration for creating effective

techniques to recognize spoken words. Because of the considerable advancements made in this area, Automatic Speech Recognition (ASR) technique is widely utilized. Voice recognition technology called ASR has been designed to assist individuals to communicate better with one another. By utilizing these ASR technologies, it is possible to overcome the challenges imposed on the world's various dialects. As a result, a number of strategies have been developed in the ASR field. An effective method for dialogue processing is the Support Vector Machine [9]. Support Vector Machine (SVM) is used as an estimator of posterior probabilities to improve the performance of identification systems as it has a strong predictive capacity and discrimination. They also have a structural risk minimization (SRM) foundation, where the goal is to create a classifier that minimizes a limit on the predicted risk instead of the empirical risk [5]. Segregating classifier of SVM is considered to have a few practical applications. An algorithm of machine learning for binary categorization is SVM categorization [8]. For multiclass assignments in the real life, it is necessary to change the system's decision-making portion.

SVM is initially intended to function in situations when data has only two classes. In other terms, binary classifier issues can be solved using it. The goal of the multiclass SVM issue is to label cases by selecting labels from a limited number of various elements. Reducing the single multiclass issue into many multiple binary classification issues is the conventional method for utilizing SVM to solve this issue. Building one-versus-all classifiers and selecting the class that correctly categorizes the test examples with the largest margin of error is the most widely used strategy in practice [10]. The presented method first transforms the reader's sound waves into Mel-Frequency Cepstrum Coefficients (MFCC) features before generating a features vector matrix. To train the modified SVM-based technique, portions of the retrieved features are utilized. The trained SVM performed really well when tested twice with the other portion of the collected information. The suggested SVM-based technique was examined employing real-world data. When assessing and contrasting the findings, the tests had very good outcomes. Utilizing the identical datasets of gathered waveforms, the outcomes of the presented SVM-based technique were evaluated with those of other methods for comparison study. The developed SVM-based method surpasses competing methods in terms of accuracy.

Background: The situation is complicated, nevertheless, by the degrading language used in speeches delivered by professionals. As a result, it becomes vital to create new tools to aid in the comprehension of flawed Arabic discourses. Speech that is illogical will then be automatically corrected [11]. In Arabic, there are four divisions: Model Standard Arabic (MSA), which is used in newspapers, books, broadcasts, documentaries, formal situations, or when the reader or audience is from a different nation of Arabic, and Dialectal Arabic (DA), which is a dialect of everyday experiences, are all still in use today. Ancient Arabic is no longer in use [12]. In formal situations like journalism, presentations, and courtrooms, the MSA is widely trained in schools and colleges. All countries that speak Arabic

acknowledge MSA as their primary language. Additionally, people frequently use their own languages in everyday conversation. These languages are taught in schools, but there is no organized written record of them. Arabic dialects have extremely distinct morphology, vocabulary, phonology, and syntax from MSA. Dialectal Arabic is the language that is organically spoken in daily life (DA). It varies from one country to the next and can be seen within a country [13]. The user's inquiries and silence are often used as input to the Dialogue Manager component to help the system decide what to do next. Recognized dialogue acts give accurate information when the user asks questions and remains silent.

The remainder of this strategy is distributed into the succeeding divisions: Section II presents the pertinent works and provides a comprehensive analysis of them. Section III describes the problem statement. A thorough review of the proposed method and analysis for the prediction of act classification is provided in Section IV. The results of the experiment are given, examined, and in-depth assessed in Section V, along with a comparison. Finally, the paper is concluded.

II. RELATED WORKS

The paper [14] utilized DL-based method to show how the meaning of the original and suspect reports was identical. By projecting each potential investment onto its neighbors, the word2vec method did in fact identify the pertinent information. Sentence relating to various could then be produced by combining the acquired dimensions. CNN was then employed to gather more environmental data and determine the level of measuring the similarity. Due to the dearth of accessible to the public material, an edited collection was created using the skip gramme approach. It performed better when substituting an original term with one from language skills that was most comparable and belonged to the same grammatical class. In terms of precision (85%) and recall (86.8%), the work improved efficient environmental connection recognition between Arabic materials compared to earlier studies. for measuring semantic similarity and document modelling, CNN model with various statistical regularities was utilized. A word2vec-based Arabic paraphrased corpus was created for the trials due to the dearth of Arabic paraphrased materials that were accessible to the general audience. Every word from the OSAC source corpus was replaced with its vocabulary counterpart that was the most similar and shared the same grammatical class.

The automatic comprehension of Arabic dialogues for the Egyptian dialect at the utterance level is a challenge that addresses in the paper [15] utilized a machine learning approach, called YOSR. The extractor has been tested on a dataset of impromptu conversations and instant messages for the Egyptian dialect. The results from the YOSR classification are really encouraging. These are the first outcomes for comprehension of the Egyptian dialect that have been recorded, as far as aware. As a viewpoint, we intend to enhance YOSR by including standard call-center regarding the service, context-based characteristics, regional phrase modifications, and morphological information such the initial verb form and Lemma. In order to enhance the categorization

outcomes, also plan to expand the "JANA" dataset for training. However, the efficiency of the method is comparatively slow down the process.

Modern Arabic natural language processing, or ANLP solutions are being developed using machine learning techniques. ML algorithms are commonly employed in NLP because to their high level of accuracy regardless of how strong the input signal is and how simple they are to utilize. On the other hand, the method used in ML-based ANLP implementations entails a number of phases. This evaluation clarifies the concept in detail, illustrates how ML techniques were used to develop these tools, and identifies well-liked ANLP techniques. The paper [16] covers the importance and specifications of ANLP as well as the characteristics and challenges of the Arabic language. Arabic sentiment expressions in tweets can be difficult to distinguish in sentiments categorization software. The complexity of the Arabic script and the unorganized nature of Twitter usage may both contribute to this problem; however, Textual data usability evaluation still needs improvement.

The paper [17] utilized DNN framework was used for sentiment classification, text translation, and text analysis. To text categorization are assessed using a sequence-to-sequence encoder-decoder paradigm, which would consist of NN trained concurrently on outputs and inputs. DNN make use of huge datasets to enhance their accuracy. These linkages are supported by the method, which by identifying textual focal areas, may also be able to manage extensive texts more successfully. This work has re-implementing the basic text useful when analyzing and adapting the sequence-to-sequence structure to Arabic because Arabs have never seen the usage of this approach for text summarizing. About 300,000 elements make up the data set, and each one comprises the title that corresponds with the corresponding article information. After applying standard summarizing techniques on the prior data set, the results are contrasted using the ROUGE index. However, the volume of the AHS information created and raised to the Gig word collection, but it is not much larger than the CNN/Daily Mail database. Additionally, it might make use of other strategies that would be advantageous in the sequential design, such as See et AI demonstrations of a comprehensive approach that uses a various performance vector. The perfect place to concentrate on would be trying to infer new designs that are useful with this vocabulary knowledge in multiple languages, considering that Arabic is a particular textual phrase read from right to left. The research does not take the possibility of increasing the data set into account.

In [18], used an Arabic Alphabet Sign Language Recognition System that is vision-based. Four distinct stages make up the process. The technique can be used with three possible database types: data with hands bound and an alongside a dark tiled wood, information with bare hands and a white base, and data with face buried in a darker-collared glove. By using one of the offered methods in AArSLRS, a hand is first removed from the image and separated from the background before the hand features are eliminated using the search technique that was employed to eliminate them. For the classification of the 28-letter Arabic alphabet using 9240

pictures in this work used methods of supervised learning. To focused on categorizing the 14 special characters and using Quranic sign language to sign the first verses of the Qur'an. It has been proven that picking a resource allocation problem is the optimum choice and that the neighborhood k value of the K-mean clustering algorithm affects rating dependability. However, the method needs more time for the processing phase when compared with other methods.

For an experimental tests of the connection between traditional in the paper [19] employed SR corpus compensation approaches (feature vectors, data selection, sexual identity acoustic frameworks, and dialect-dependent/register-dependent variance across Arabic ASR samples). The first interaction examined in the paper was that between intermittent syntax variance and auditory tracking performance. By removing speakers with inadequate teaching data and switching to grapheme-based acoustic models instead of phone-based ones, discrete specific language difference can be rewarded for. The latter method also helps to make up for poor capture performance, which is further made up for by removing delta-delta acoustic features. Together, the three methods decrease Word Error Rate by 3.24 percentage points to 5.35 percentage. Alteration in the perfectly alright acoustic speech sounds from each phoneme in the phrase is the second feature of regional and registration diversity to be taken into account. Building way of predicting and dialect-specific algorithms contributes to significant reductions in WER since empirical findings show that sexual identity and language are the primary contributors to diversity in speech. Cross-dialect investigations are carried out to gauge how far different Arabic dialects are in terms of the acoustic variations between phone models needed for each of them.

Natural Language Understanding has been vastly enhanced by DL methodologies such as word representations and DNN architectures. This [20] work provides a method for Arabic home automation using DL approaches for text categorization and text categorization recognition. In order to achieve this, to provide an NLU component that can be additionally connected with ASR, a dialogue administrator, and a natural language generator component to develop a fully functional making the change. The procedure of gathering and categorizing the data, constructing the purpose classification and concept extraction frameworks, and eventually the evaluation of these techniques against benchmark datasets are all included in the study. The benchmark results showed that the LSTM effectiveness, with an F-Score of 92.0, was marginally superior to the CNN achievement for both intent classification techniques. The paper employed a hybrid representation of word representations and personality language models, which is then fed to a Bidirectional LSTM network, to retrieve the user objectives and purposes from the information. A high F-Score of 94.0 for the BiLSTM with the Char Embedding's test suggests that its efficiency is extremely comparable to the most recent English Named Entity Recognition benchmarks. However, to create an undertaking dialogue system, the Natural Language Understanding module can be combined with automatic speech recognition and natural language generation modules.

III. PROBLEM STATEMENT

This issue may be exacerbated by the intricacy of the Arabic script and the disorganised Twitter usage. The evaluation of the usability of textual data still needs work, though. The potential for expanding the data set is not considered in the current research. When compared to alternative ways, the processing step of the current method takes more time. To determine how diverse Arabic dialects are from one another in terms of the acoustic differences between phone models required for each of them, cross-dialect examinations are conducted. However, the Natural Language Understanding module can be integrated with the automatic speech recognition and natural language creation modules to form a complex dialogue system. To develop a classification model for sentential act prediction, the existing technique used a selected group of features extracted from annotated utterances and applied in "YOSR", an SVM technique based on ML. SVM designs ought to be used as a result for improved processing.

IV. METHODOLOGY

The suggested approach is based on Arabic conversation processing and employs a machine learning algorithm to classify Acts. The contact among the operator and clients in various industries is the subject of the study. The data was gathered from the Jana Corpus in three distinct industries, including banking, travel, and chat. The research's major objective is to establish a connection with the speaker and operator and help them comprehend the dialogue, whether it's a request, an obligation, or something else entirely. Previous

studies have used a variety of algorithms, including deep learning, gradient boosting, convolution neural networks, and natural language processing, but they have not been able to achieve higher accuracy rates. As a result, the research focused on interpreting Arabic dialogue without cues and makes use of chatbot technology and machine learning algorithms to increase productivity. The research is conducted in the stages listed below, and Fig. 1 shows the flow diagram based on the suggested method.

A. Dataset Collection

The data collection used in this study was compiled from a variety of sources, including chat, banks, and flights. The vast amount of material on the website is useful for any chatbot application. The CAMEL tool module for Python was used to parse Arabic content and retrieve data. To address a variety of issues, researchers present CAMEL Tools, a user-friendly Python toolkit for Arabic and Arabic dialect pre-processing, morphological modelling, dialect recognition, text categorization recognition, and sentiment analysis. The application programming interfaces (APIs) and command-line interfaces (CLIs) offered by CAMEL Tools cover these functions. This also processes Arabic dialogue using data from chat, planes, and banks. The bot can practise on a dataset, which is a collection of different input expressions and outputs. This effort will allow Arabic programmers to use more dialects of Arabic. The gathered data were split into training and testing samples, with training data making up 70% of the samples and testing data making up 30% of the samples. The sample's division for the testing and training processes is shown in Table I and Fig. 2.

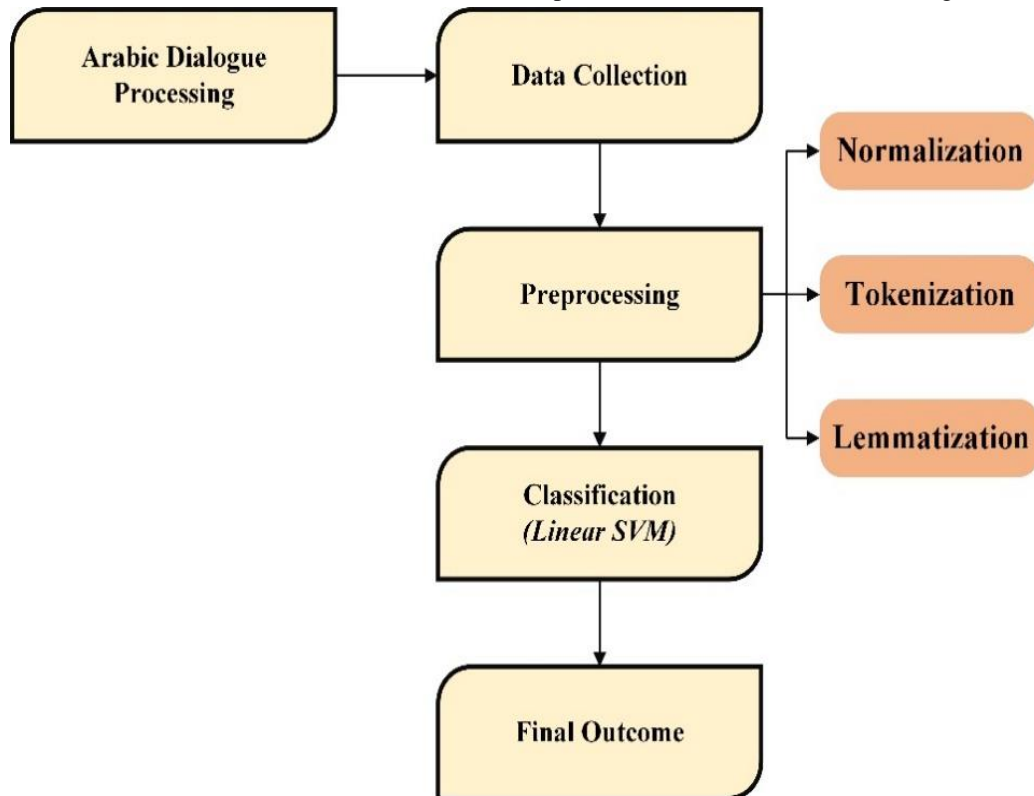


Fig. 1. Proposed method.

Out[31]:

	TID	UID	Text	Speaker	PrimaryClass	SecondaryClass	Act
0	D01T01	D01T01U01	مساء الخير	operator	response	socialobligation	greeting
1	D01T01	D01T01U02	الاهلي فون	operator	other	socialobligation	selfintroduce
2	D01T01	D01T01U03	احمد سامي	operator	other	socialobligation	selfintroduce
3	D01T02	D01T02U01	مساء النور	customer	response	socialobligation	greeting
4	D01T03	D01T03U01	أتشرف بالاسم	operator	request	question	otherquestion
...
1133	D56T18	D56T18U01	شكرا ليك اخي سعود	customer	response	argumentation	agree
1134	D56T19	D56T19U01	...أي إستفسارات أو خدمات أخرى اقدر اساعدك فيها ال	operator	request	question	confirmquestion
1135	D56T20	D56T20U01	شكرا	customer	response	argumentation	disagree
1136	D56T21	D56T21U01	- شكرا لتواصلك بخدمه الدعم المياشر	operator	response	socialobligation	thanking
1137	D56T21	D56T21U02	زين عالم جميل	operator	other	socialobligation	selfintroduce

4725 rows × 7 columns

Fig. 2. Input data.

TABLE I. DATA SAMPLE

Data from the JANA Corpus involved in the study	
Transcribed Dialogues	
Telephone Calls	52
Instant Messaging Conversations	30
Total number of words	20311

B. Preprocessing

Preprocessing's function is to add morphological and other additional context to the lemmatizes (and other following elements') input in order to simplify the lemmatization process. The preprocessing performed by the lemmatization technique is quick and light in comparison to cutting-edge tools. Part-of-speech tagging, performed as a quick machine-learning-based sequence labeler, is used to first enhance the tokenized input text. A very basic word segmented component is used to progressively segment it, reducing the complexity and ambiguity of the words. The procedure's simplicity, which is described in full, can be attributed to the POS tagger's output of a wealth of morphological features, which makes word segmentation essentially easy [21].

In order to make the lemmatization process simpler, preprocessing's role is to add morphology and other contextual info to the lemmatizes (and other following sections') input. The algorithm performs tokenization, function word elimination, and lemmatization for each article in the database. There are numerous Arabic dialects. The analysis of dialect and vernacular terminology follows. Assessments are often written by researchers in their original tongues. Depending on the dialect, they use different terms to convey the same meaning. The solution to this problem was to create and use a dictionary of dialectical terms and their equivalents in standard Arabic [22].

C. Morphology Analyzer

Morphological analysis is the first stage of text preparation. Breaking down morphologically complicated words into their individual morphemes is known as morphological analysis, also known as analysis of structures (word meaning parts). The morphological analysis of these phrases is then carried out using the Alkhalil analyzer. As a consequence, researchers are provided with all potential lemmas and associated morpho-syntactic information for each individual word in the text that has been analysed out of contextual [23].

D. Normalization

Experts in Arabic dialogue processing typically employ orthographic normalization as a fundamental technique with the goal of reducing data noise. This is true whether the intention is to produce parallel texts for machine learning, information retrieval data, or computational linguistics text. Examples of normalizing include Tatweel elimination (an effort to eliminate the Tatweel sign), Diacritic removal, and alphabet normalization (variant forms to one form conversion) [24]. The tokenization procedure should become more confusing at this point even if these normalizations would help us in the search or discovering phase. As an illustration, if Ta-Marbuta (ة) and Ha (ه) are normalized, the latter could be tokenized as a pronoun. Because of this, normalizations are treated as a stepping stone for querying, finding, and other operations in the research.

E. Tokenization

Reducing the raw text into smaller, more manageable chunks is tokenization. It is not an easy problem to solve since tokenization is "closely related to the morphology". This is especially true for dialects like Arabic, which have rich yet complex morphologies. Flowing text is divided into tokens by tokenizers so that they can be processed further by

morphological transducers or POS taggers. The tokenizer recognizes numbers, word boundaries, clitics, multi-word phrases, and abbreviations. By doing this, the original text is divided into tokens, or words and sentences. These tokens support simulation techniques for various processes or contextual comprehension. Tokenization makes it easier to understand the text's significance by looking at the word order. Stop words are typically eliminated in the initial phases of tokenization even though there isn't however a defined list of Arabic stop words [25]. There could be several degrees of Arabic tokenization built, depending on how complicated the linguistic study involved is. Researchers worked with the Arabic grammar to develop three distinct strategies, or frameworks, for Arabic tokenization. In terms of durability, adherence to the modular concept, and ability to remove unnecessary ambiguity, these systems differ greatly from one another. The tokenizer uses white space and punctuation to distinguish among major tokens. However, when defining sub-tokens, the tokenizer needs more morphological information. This information can be provided either deterministically or in-deterministically by a morphology converter or a token forecaster. In the end, both main tokens and sub tokens are identified by the same token borders, which are denoted by the symbol "@" throughout the entire article. It is believed that separating the text into primary tokens using capitalization and white spaces is a straightforward strategy [26].

1) *Utterance feature*: The smallest speech unit in spoken language analysis is termed an utterance. It is a continuous chunk of speech with a distinct pause at the start and the finish. Oral languages typically, but not always, have silence as their boundary. Only their depictions can be found in written language; utterances are absent. The following Table II shows the extraction of utterance.

The utterances' meta data could be used to help with dialogue act classification. Additionally, the classification procedure could benefit from knowing what happened before the current utterance. Researchers made use of:

Utterance Speaking Style: The speakers or listeners of the current utterance may be able to discern the act of speech based on their speaking style. The act of "Service-Question," for instance, relates to the customer because they contacted customer service to ask about just one service offered, whereas the acts of "Other-Question" and "Choice-Question" relate to the control system since the operator chooses one of the services offered or asks the client's identity.

Previous Utterance Act: The classification may anticipate the act of the present utterance by understanding the sequencing of previous statements in the dialogue. For example, the act "Confirm Question" almost immediately follows the acts "Agree" and "Disagree".

F. Lemmatization

The main element of our lemmatization strategy is a classifier that uses machine learning. It accepts m -word segments as input along with their matching POS tag, taking context (utterances and labels) into consideration. The inherent ambiguity of Arabic words without diacritics, whose meanings are often inferred by both individuals and computers from context, justifies the learning-based method. The introduction of new cases necessitates retraining the analyzers. Another disadvantage is that classifications, like the one from OpenNLP that the researcher utilized, frequently settle on an output variable that may or may not be accurate. In this case, additional NLP processing methods from the entire collection of probable lemmas may be able to yield reliable results. To solve these situations, investigators supplement the learning-based lemmatize with a dictionary-based lemmatize.

Fig. 3 illustrates the results of the lemmatization. The analyzer is first used by researchers to look at the words in the database's vowelized version. Once this has been done, researchers only save lemmas whose given lexical tags (clitics, root, and stem) coincide with the word's in-document clitics, root, and stem. The analyzer has located and examined the corpus's words. One potential lemma of the phrases was also discovered during this initial step's study into prospective lemmas. Following the identification of the likely lemmas for each word in the corpus, researchers asked two linguists to select the appropriate lemma from the available options. If the correct lemma is not given among the potential lemmas, the speaker provides the proper lemma to the phrase.

After being lemmatized, the text is again processed to provide the class feature. The class feature obtained in this process is split into final features and final labels, as shown in Fig. 4 and 5. Out from clean text, the first word is acquired, the first verb is filtered and the class feature is extracted, and eventually the class feature is again filtered to obtain the last feature and final labels, as shown in Fig. 4 and 5. As part of this procedure, the provided data were divided into training and testing phases, which were used to classify the data using machine learning.

TABLE II. UTTERANCE FEATURE

Data	
Conversation Turns	3001
Average Conversation/turn	6.7
Utterances	4725
Average words/Utterance	4.3

Out[38]:

	TID	UID	Text	Speaker	PrimaryClass	SecondaryClass	Act	Clean Text	First Verb	Utterance Length
0	D01T01	D01T01U01	مساء الخير	operator	response	socialobligation	greeting	مساء خير	الخير	2
1	D01T01	D01T01U02	الاهلي فون	operator	other	socialobligation	selfintroduce	اهلي فون	فون	2
2	D01T01	D01T01U03	احمد سامي	operator	other	socialobligation	selfintroduce	احمد سام	سامي	2
3	D01T02	D01T02U01	مساء النور	customer	response	socialobligation	greeting	مساء نور	النور	2
4	D01T03	D01T03U01	أتشرف بالاسم	operator	request	question	otherquestion	تشرف اسم	بالاسم	2
...
95	D03T08	D03T08U01	انا لي شهادات استثمار مجموعة ب في فرع شريف	customer	response	argumentation	inform	ان لي شهادة استثمار مجموعة ب في فرع شريف	شريف	9
96	D03T09	D03T09U01	تمام	operator	response	argumentation	agree	تمام	تمام	1
97	D03T10	D03T10U01	البنك الاهلي	customer	response	argumentation	inform	البنك اهلي	الاهلي	2
98	D03T10	D03T10U02	ممکن احولها الي الشهادة البلاطينية	customer	request	question	servicequestion	ممکن احول الي شهادة البلاطينيه	البلاطينيه	5
99	D03T11	D03T11U01	تمام	operator	response	argumentation	agree	تمام	تمام	1

100 rows × 10 columns

Fig. 3. Text obtained after lemmatization process.

TID	UID	Text	Speaker	PrimaryClass	SecondaryClass	Act	Clean Text	First Verb	Utterance Length	Class Features
0	D01T01	D01T01U01	مساء الخير	operator	response	socialobligation	greeting	الخير	2	Start Start Start
1	D01T01	D01T01U02	الاهلي فون	operator	other	socialobligation	selfintroduce	فون	2	greeting
2	D01T01	D01T01U03	احمد سامي	operator	other	socialobligation	selfintroduce	سامي	2	selfintroduce
3	D01T02	D01T02U01	مساء النور	customer	response	socialobligation	greeting	النور	2	selfintroduce
4	D01T03	D01T03U01	أتشرف بالاسم	operator	request	question	otherquestion	بالاسم	2	greeting
...
1133	D56T18	D56T18U01	شكرا ليك اخي سعود	customer	response	argumentation	agree	سعود	4	confirmquestion
1134	D56T19	D56T19U01	أي إستفسارات أو خدمات أخرى ...أقدر اساعدك فيها ال	operator	request	question	confirmquestion	؟	11	agree
1135	D56T20	D56T20U01	شكرا	customer	response	argumentation	disagree	شكرا	1	confirmquestion
1136	D56T21	D56T21U01	شكرا لتواصلك بخدمه الدعم - المباشر	operator	response	socialobligation	thanking	-	6	disagree
1137	D56T21	D56T21U02	زين عالم جميل	operator	other	socialobligation	selfintroduce	جميل	3	thanking

Fig. 4. Class feature.

TID	UID	Text	Speaker	PrimaryClass	SecondaryClass	Act	Clean Text	First Verb	Utterance Length	Class Features	Final Features	Final Labels	
0	D01T01	D01T01U01	مساء الخير	operator	response	socialobligation	greeting	مساء خير	الخير	2	Start Start Start	2 مساء خير الخير operator Start Start Start	[response, socialobligation, greeting]
1	D01T01	D01T01U02	الاهلي فون	operator	other	socialobligation	selfintroduce	اهلي فون	فون	2	greeting	2 اهلي فون فون operator greeting	[other, socialobligation, selfintroduce]
2	D01T01	D01T01U03	احمد سامي	operator	other	socialobligation	selfintroduce	احمد سام	سامي	2	selfintroduce	2 احمد سام سامي operator selfintroduce	[other, socialobligation, selfintroduce]
3	D01T02	D01T02U01	مساء النور	customer	response	socialobligation	greeting	مساء نور	النور	2	selfintroduce	2 مساء نور النور customer selfintroduce	[response, socialobligation, greeting]
4	D01T03	D01T03U01	أتشرف بالاسم	operator	request	question	otherquestion	تشرف اسم	بالاسم	2	greeting	2 تشرف اسم بالاسم operator greeting	[request, question, otherquestion]
...	
1133	D56T18	D56T18U01	شكرا ليك اخي سعود	customer	response	argumentation	agree	شكر لي اخ سعود	سعود	4	confirmquestion	4 شكر لي اخ سعود سعود customer confirmquestion	[response, argumentation, agree]
1134	D56T19	D56T19U01	أي استفسارات أو خدمات أخرى اقدر اساعدك فيها ال	operator	request	question	confirmquestion	اي استفسار او خدمه اخر اقدر ساعد في الله عافي ؟	؟	11	agree	11 اي استفسار او خدمه اخر اقدر ساعد في الله عا customer confirmquestion	[request, question, confirmquestion]
1135	D56T20	D56T20U01	شكرا	customer	response	argumentation	disagree	شكر	شكرا	1	confirmquestion	1 شكر شكرا customer confirmquestion	[response, argumentation, disagree]
1136	D56T21	D56T21U01	شكرا لتواصلك بخدمه الدعم - المياشر	operator	response	socialobligation	thanking	شكر تواصل خدمه دعمه -	-	6	disagree	6 شكر تواصل خدمه دعمه مياشر operator disagree	[response, socialobligation, thanking]
1137	D56T21	D56T21U02	زين عالم جميل	operator	other	socialobligation	selfintroduce	زين عالم جميل	جميل	3	thanking	3 زين عالم جميل جميل operator thanking	[other, socialobligation, selfintroduce]

4725 rows × 13 columns

Activate Windows

Fig. 5. Final feature and label text.

G. Multilayer outcome and Three Layer Hierarchical Classifications

A kind of machine learning called multi-output classification predicts many outcomes at once. When generating any prediction in multi-output classification, the system would generate multiple or more outcomes. In other classes, the system usually forecasts just one result. Three models—one for each layer—are primarily used in the construction of the suggested technique. As a result, researchers believe that proposed method of classification is quicker than existing classification and may be a more effective dialogue act classification model in real-time systems. The second distinction is that, while "this approach" does not change with the number of dialogue acts or classes, the variety of models does change when binary categorization is utilized, as in the researchers' strategy [27].

1) *Three layer hierarchical*: Traditional classification requires the construction of 26 categories, one for each dialogue act, as opposed to hierarchical structure categorization, which is separate. Every word stated in a discussion is communicated to succeeding tiers to determine the next course of action, which is the fundamental tenet of a hierarchical organization. The recommended method classifies dialogue utterances using a three-layer hierarchical framework. In the first layer, the utterance had already been assigned to one of the three fundamental categories—request, answer, or other. The speech was categorized into various discourse actions in the second layer based on its primary

category, which was established or categorized in the first layer. Question, turn management, social responsibility, social courtesy, argumentation, response, and discourse structure are some of them. Fig. 6 illustrates the hierarchical classification that may be made using the multi-class classification technique.

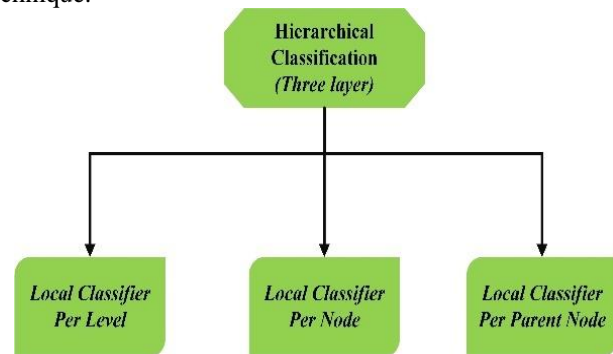


Fig. 6. Hierarchical classification.

Local Classifier per Level:

- For each level of the classification, there is a single multi-class classifier in this system.
- It is a trim, efficient categorization device.

Local Classifier per Node:

- Each node should have a binary classifier.

- Multi-labeling comes easy to this strategy.

Local Classifier per Parent Node:

- Every parent node receiving a single multi-class classifier.
- This is the most logical method.
- This strategy is far more frugal than the prior one.

H. Classification

The categorization process is conducted out following the lemmatization technique. In earlier studies, the majority of researchers classify the text using NLP, deep learning, SVM classifiers, gradient boosting, etc. Linear Support Vector Machine is utilized in this study to categorize the provided data. The given dataset was divided into test and training data, and the classification method was applied to the data.

1) *Linear support vector machine*: One of the most popular and well-known machine learning methods is called a support vector machine (SVM). SVMs are supervised learning models that are further used to evaluate and classify text data. SVM is also frequently used for regression evaluation and text classification. Research introduced and applied SVM in text classification and categorization in 1998. The basic purpose of the SVM training method is to create a model that classifies fresh documents into a number of predetermined groups. SVMs could also be employed as non-linear and linear classifiers. A relatively recent class of machine learning algorithms called Support Vector Machines (SVM) was first developed. Based on the data mining and machine learning theory's structure risk minimization concept, SVM seeks a decision surface to divide the training data points into two classifications and bases its judgments on the training examples that are chosen as the only useful components in the training dataset. The SVM Classifier is used for the classification of the input. The Stages convoluted in the procedure are:

- Stage 1: The necessary libraries are added.
- Stage 2: Including the necessary dataset.
- Stage 3: Data pre-processing is carried out.
- Stage 4: Preparing the training and test datasets
- Stage 5: The dataset is encoded to distinguish between labels, and values of 0 or 1 are assigned.
- Stage 6: involves converting text data to polygons. (The user may choose to utilize Term Frequency-Inverse Document Frequency TF-IDF)
- Step 7 involves utilizing SVM to accomplish machine learning.

According to the definition of natural language processing, this describes the conversion of any raw input into highly comprehending. Fig. 7 depicts the pre-processing stages that are involved.

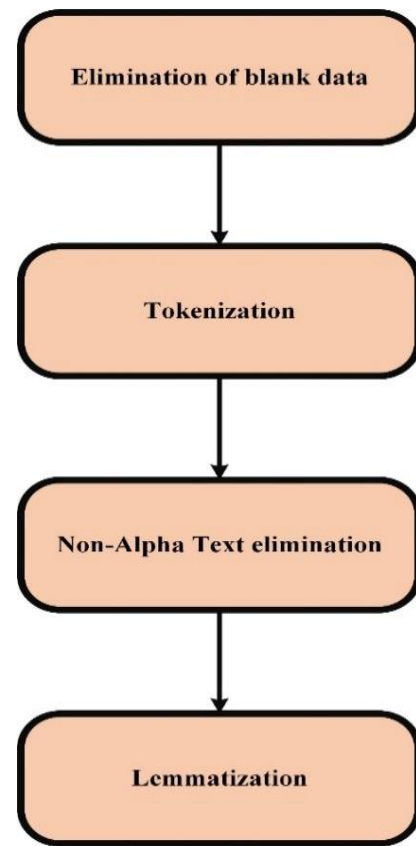


Fig. 7. Preprocessing stages.

V. RESULTS AND DISCUSSIONS

Three datasets—the Banks dataset, the Flights dataset, the chat dataset is presented in the experimental findings. SVM, Nave Bayesian, and logistic regression approaches are used independently for each dataset. Moreover, the performance evaluation of different dataset for both multioutput classification and 3-layer hierarchical structure is also presented. Multioutput and local hierarchical classification apply to classifier 3-layer or 3 level structure (primary Categories, secondary Categories and dialogue act) Primary Categories is (Request, Response, and Other) secondary Categories is (Question, Turn Management, Social Obligation, Social Courtesy, Argumentation, Answer, and Dialogue Structure. First classifier primary categorise next classifier second categories depends on the first classification and last act classification depends on first and second classification.

A. Performance Evaluation

Accuracy: Accuracy processes how accurately the system model operates. Usually, it is the proportion of correctly predicted observations to all observations. Accuracy is uttered in eq. (1),

$$A = \frac{\text{Total number of Dialogue act}}{\text{No of dialogue act classified} + \text{No of correct dialogue act}} \quad (1)$$

Precision: Precision is estimated as the number of correct positive estimates of the correct text. It is the fraction of precise identification of the text that is computed utilizing eq. (2),

$$P = \frac{\text{No of correctly classified dialogue act}}{\text{number of dialogue act classified}} \quad (2)$$

Recall: The recall is defined as the ratio of the entire true positives and false negatives to the right positive forecasting accuracy as shown in eqn. (3).

$$R = \frac{\text{No of correctly classified dialogue act}}{\text{number of correct dialogue act}} \quad (3)$$

F1-Score: The F1-score measurement combines precision and recall. Precision and recall are used to calculate the F1-score measure that is symbolized in eqn. (4).

$$F1 - score = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

In the following research Arabic dialogue processing is carried out in machine learning algorithm and the accuracy is obtained and compared with the existing method.

B. Logistic Regression

The following Table III displays the logistic regression's performance measure for the three parameters of recall, precision, and F1-score for various interactions, including Confirm inquiry, Correct, Disagree, Greet, Inform, misunderstanding sign, Pause, Self-introduce, Service response, Service question, Suggest, accepting request, Thanking, Other answer, other question Warning: Turn assignment.

TABLE III. ACT CLASSIFICATION AND PERFORMANCE MEASURE OF LOGISTIC REGRESSION

Act	Precision	Recall	F1-Score
Agree	0.93	0.80	0.86
apology	0.00	0.00	0.00
Closing	0.82	1.00	0.90
Confirm question	0.34	0.54	0.42
Correct	0.00	0.00	0.00
Disagree	0.63	1.00	0.77
Greeting	0.95	0.90	0.93
Inform	0.15	0.57	0.24
Mis-understanding sign	0.25	1.00	0.40
Offer	0.00	0.00	0.00
Other answer	0.59	0.62	0.60
Other question	0.61	0.86	0.71
pausing	0.38	1.00	0.55
Promise	0.00	0.00	0.00
Repeat	0.00	0.00	0.00
Self-introduce	0.76	0.94	0.84
Service answer	0.92	0.62	0.74
Service question	0.57	0.58	0.58
Suggest	0.00	0.00	0.00
Taking request	1.00	1.00	1.00
Thanking	1.00	1.00	1.00
Turn assign	0.67	0.85	0.75
Warning	0.10	0.75	0.18

C. Performance Evaluation using SVM Multi-output Classification without Cue

Data are displayed without a cue in Table IV using local support vector machines. Without cues utilizing a support vector machine is explained in Fig. 8. For all datasets, data without cue using support vector machine classification produced better results in terms of recall, precision, accuracy, and f1-score for data without cue. Without employing cue data, a support vector machine was used to achieve an accuracy value of 0.83 and a recall value of 0.85.

A baseline system employing the suggested feature set even without hierarchical characteristics (the major category of the present utterance and the major category of the prior utterance) has been developed in order to assess the efficacy of a Local hierarchical classification in categorizing dialogue acts when used in discussions, flights, and banks. In terms of recall, F measure, precision, and accuracy macro averages, the proposed systems' performances are shown in the table and figure, with multi-output three-layer classification producing the largest result and highest result.

Tables V and VI depict the multioutput classification and local hierarchical structure to interpret data without cue. Performance assessment accuracy findings for bank datasets are higher utilizing Multioutput classification and the Hierarchical structure for supplemented data without a cue. Fig. 9 displays multioutput classification with enhanced data without a prompt. Fig. 10 illustrates data without a cue using a hierarchical structure.

TABLE IV. PERFORMANCE EVALUATION USING SVM

Performance Metrics	SVM Method
Accuracy	0.83
Precision	0.84
Recall	0.85
F1-Score	0.84

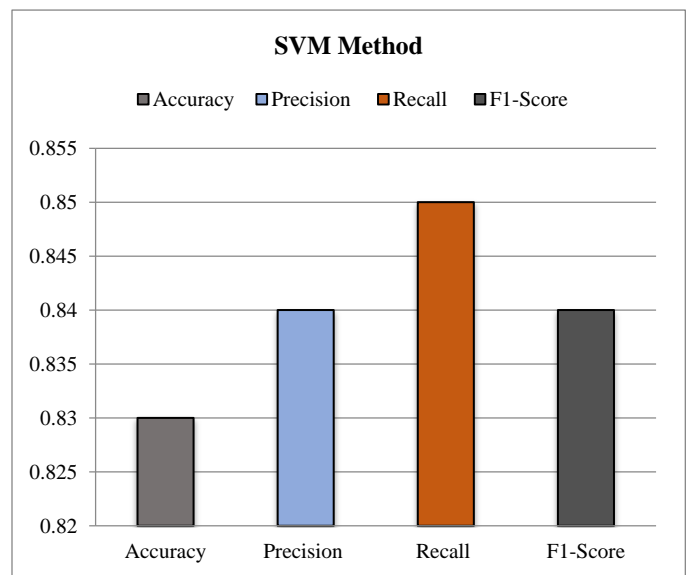


Fig. 8. Performance evaluation using SVM.

TABLE V. MULTI-OUTPUT THREE-LAYER CLASSIFICATION

Multioutput Classification				
	Accuracy	Precision	Recall	F1-score
Banks	0.96	0.95	0.95	0.95
Chats	0.96	0.95	0.94	0.94
Flight	0.95	0.94	0.95	0.94

TABLE VI. THREE-LAYER HIERARCHICAL STRUCTURE

Local hierarchical classification				
	Accuracy	Precision	Recall	F1-score
Banks	0.96	0.95	0.95	0.95
Chats	0.95	0.94	0.94	0.94
Flight	0.95	0.94	0.95	0.94

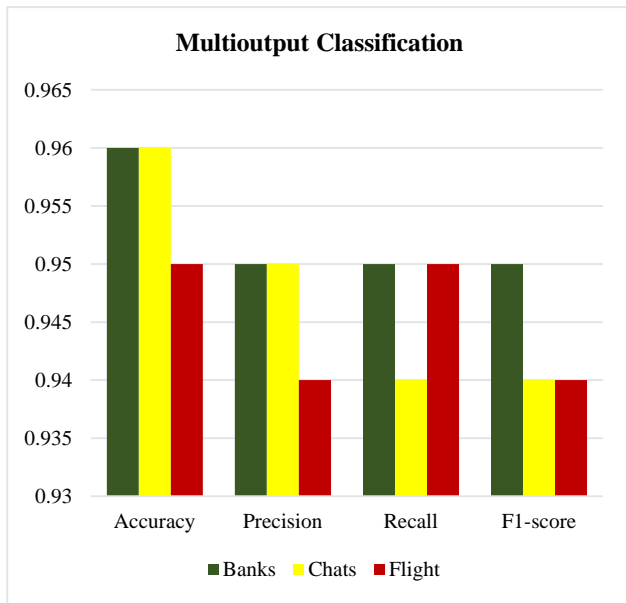


Fig. 9. Multioutput classification.

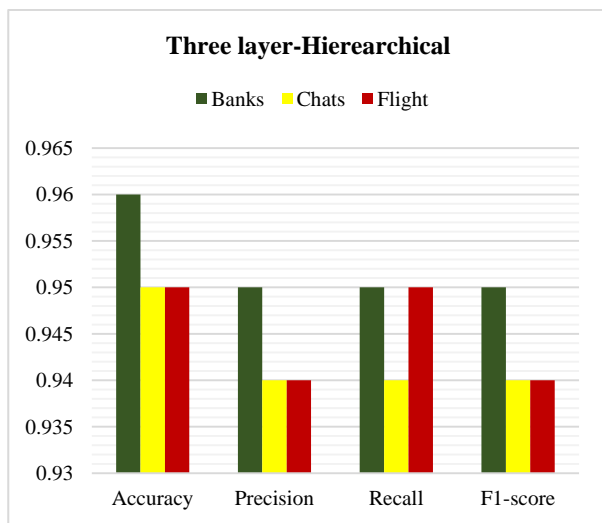


Fig. 10. Local hierarchical classifications.

D. Performance Comparison

Compared to the logistic regression approach shown in Table VII, the projected methodology support vector machine achieves a greater level of efficiency. Compared to performance assessed using the logistic regression method, the novel support vector machine produces greater accuracy which shown in Fig. 11. Using the support vector machine model, an accuracy level of 0.83 was achieved in this case. This suggests that a support vector machine is more effective at classifying acts than the logistic regression approach.

TABLE VII. ACCURACY COMPARISON

Method	Accuracy
Logistic Regression	0.72
Linear SV	0.83

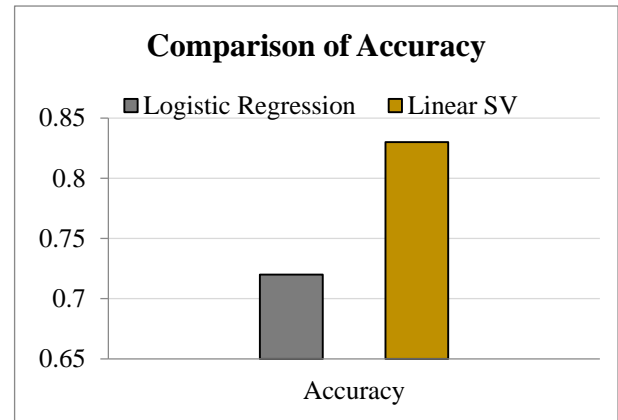


Fig. 11. Accuracy comparison of logistic regression and linear SVM.

VI. CONCLUSION

The structure that supports Arabic dialogue processing has been modified for better machine comprehension. The proposed classifier was tested using the Jana corpus of real-world spoken dialogues, and the results are quite positive. Researchers made progress in this work by developing an Arabic Dialogue Processing that will allow users to interact via chat, banking, and flights using textual documents. Additionally, researchers provide the JANA corpus, a multi-genre collection of spontaneous Arabic dialogues that has been tagged at the utterance level in Arabic Dialogues Act Understanding. A technique for classifying utterances according to whether or not they contain function words was created in this work. Among the attributes of this methodology are a database and a set of function words. On the selected dataset, a number of classifiers were tried to see which one would be most useful over the long term. The support vector machine technique is employed in the study to find the customer and operator responses. Better accuracy is attained when comparing the classification system's performance metrics between the logistic regression approach and other methods in the results section. When compared to the other method, the accuracy of the support vector machine method is 0.83. Researchers want to eventually enable clients to make reservations over the phone by connecting our Dialogue Processing with an active booking system, banking, and chat.

To serve as a standard annotated corpus for testing and assessing several Arabic software applications in the future. Finding the appropriate Support Vector Machine parameters can increase accuracy, therefore future research will examine the impact of each parameter. Future study may focus on determining the best approach to integrate various dialogue act classification development techniques in order to enhance the utilization of any technique, even the most basic.

ACKNOWLEDGMENT

I'd like to express my heartfelt gratitude to Dr. Md Sah Hj Salam for his insightful and constructive suggestions throughout the planning and development of this research project. His generosity in donating his time has been greatly appreciated. The authors would like to thank Universiti Teknologi Malaysia (UTM) under UTM Encouragement Research grant PY/2020/04246 for the support in the research.

REFERENCES

- [1] B. Algotiml, A. Elmadany, and W. Magdy, "Arabic tweet-act: Speech act recognition for Arabic asynchronous conversations," in Proceedings of the Fourth Arabic Natural Language Processing Workshop, 2019, vol. Proceedings of the Fourth Arabic Natural Language Processing Workshop, pp. 183–191. doi: 10.18653/v1/W19-4620.
- [2] Joukhadar, N. Ghneim, and G. Rebdawi, "Impact of Using Bidirectional Encoder Representations from Transformers (BERT) Models for Arabic Dialogue Acts Identification," *Ingénierie Systèmes Inf*, vol. 26, no. 5, pp. 469–475, 2021, doi: <https://doi.org/10.18280/isi.260506>.
- [3] M. M. Khalatia and T. A. H. Al-Romanyb, "Artificial Intelligence Development and Challenges (Arabic Language as a Model)," *Artif. Intell.*, vol. 13, no. 5, 2020.
- [4] M. O. Alhawarat, H. Abdeljaber, and A. Hilal, "Effect of stemming on text similarity for Arabic language at sentence level," *PeerJ Comput. Sci.*, vol. 7, p. e530, 2021, doi: <https://doi.org/10.7717/peerj-cs.530>.
- [5] Zarrouk, Y. Benayed, and F. Gargouri, "Hybrid SVM/HMM model for the recognition of Arabic triphones-based continuous speech," in 10th International Multi-Conferences on Systems, Signals & Devices 2013 (SSD13), Hammamet, Tunisia, Mar. 2013, pp. 1–7. doi: 10.1109/SSD.2013.6564036.
- [6] M. Bashir, A. Hassan, B. Rosman, D. Duma, and M. Ahmed, "Implementation of A Neural Natural Language Understanding Component for Arabic Dialogue Systems," *Procedia Comput. Sci.*, vol. 142, pp. 222–229, 2018, doi: 10.1016/j.procs.2018.10.479.
- [7] S. B. Dbabis, H. Ghorbel, and L. H. Belguith, "Sequential dialogue act recognition for Arabic argumentative debates," 2018, doi: <https://doi.org/10.26342/2018-60-6>.
- [8] W. Mohamed, B. Souha, and C. Adnen, "Speech recognition system based on discrete wave atoms transform partial noisy environment," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, 2019.
- [9] K. M. Nahar, M. Ra'ed, A. Moy'awiah, and M. Malek, "An efficient holy Quran recitation recognizer based on SVM learning model," *Jordanian J. Comput. Inf. Technol. JJCIT*, vol. 6, no. 04, 2020.
- [10] M. Ahed, B. H. Hammo, and M. A. Abushariah, "An Enhanced Twitter Corpus for the Classification of Arabic Speech Acts," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 3, 2020.
- [11] N. Terbeh, M. Maraoui, and M. Zrigui, "Arabic discourse analysis: a naïve algorithm for defective pronunciation correction," *Comput. Syst.*, vol. 23, no. 1, pp. 153–168, 2019, doi: <https://doi.org/10.13053/cys-23-1-2820>.
- [12] K. M. Nahar, O. M. Al-Hazaimh, A. Abu-Ein, and M. A. Al-Betar, "Arabic Dialect Identification Using Different Machine Learning Methods," 2022, doi: <https://doi.org/10.21203/rs.3.rs-1726491/v1>.
- [13] Alsharhan and A. Ramsay, "Investigating the effects of gender, dialect, and training size on the performance of Arabic speech recognition," *Lang. Resour. Eval.*, vol. 54, no. 4, pp. 975–998, Dec. 2020, doi: 10.1007/s10579-020-09505-5.
- [14] Mahmoud and M. Zrigui, "Sentence Embedding and Convolutional Neural Network for Semantic Textual Similarity Detection in Arabic Language," *Arab. J. Sci. Eng.*, vol. 44, no. 11, pp. 9263–9274, Nov. 2019, doi: 10.1007/s13369-019-04039-7.
- [15] A.Elmadany, S. M. Abdou, and M. Gheith, "Towards Understanding Egyptian Arabic Dialogues," *Int. J. Comput. Appl.*, vol. 120, no. 22, pp. 7–12, Jun. 2015, doi: 10.5120/21390-4427.
- [16] S. Larabi Marie-Sainte, N. Alalyani, S. Alotaibi, S. Ghouzali, and I. Abunadi, "Arabic Natural Language Processing and Machine Learning-Based Systems," *IEEE Access*, vol. 7, pp. 7011–7020, 2019, doi: 10.1109/ACCESS.2018.2890076.
- [17] M. Al-Maleh and S. Desouki, "Arabic text summarization using deep learning approach," *J. Big Data*, vol. 7, no. 1, p. 109, Dec. 2020, doi: 10.1186/s40537-020-00386-7.
- [18] Tharwat, A. M. Ahmed, and B. Bouallegue, "Arabic Sign Language Recognition System for Alphabets Using Machine Learning Techniques," *J. Electr. Comput. Eng.*, vol. 2021, pp. 1–17, Oct. 2021, doi: 10.1155/2021/2995851.
- [19] Alsharhan and A. Ramsay, "Investigating the effects of gender, dialect, and training size on the performance of Arabic speech recognition," *Lang. Resour. Eval.*, vol. 54, no. 4, pp. 975–998, Dec. 2020, doi: 10.1007/s10579-020-09505-5.
- [20] M. Bashir, A. Hassan, B. Rosman, D. Duma, and M. Ahmed, "Implementation of A Neural Natural Language Understanding Component for Arabic Dialogue Systems," *Procedia Comput. Sci.*, vol. 142, pp. 222–229, 2018, doi: 10.1016/j.procs.2018.10.479.
- [21] Elmadany, S. M. Abdou, and M. Gheith, "Recent Approaches to Arabic Dialogue Acts Classifications," in *Computer Science & Information Technology (CS & IT)*, Feb. 2015, pp. 117–129. doi: 10.5121/csit.2015.50412.
- [22] Hmeidi, R. F. Al-Shalabi, A. T. Al-Taani, H. Najadat, and S. A. Al-Hazaimh, "A novel approach to the extraction of roots from Arabic words using bigrams," *J. Am. Soc. Inf. Sci. Technol.*, p. n/a-n/a, 2009, doi: 10.1002/asi.21247.
- [23] S. Larabi Marie-Sainte, N. Alalyani, S. Alotaibi, S. Ghouzali, and I. Abunadi, "Arabic Natural Language Processing and Machine Learning-Based Systems," *IEEE Access*, vol. 7, pp. 7011–7020, 2019, doi: 10.1109/ACCESS.2018.2890076.
- [24] M. M. Kamruzzaman, "Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network," *Wirel. Commun. Mob. Comput.*, vol. 2020, pp. 1–9, May 2020, doi: 10.1155/2020/3685614.
- [25] S. Elons, M. Ahmed, H. Shedid, and M. F. Tolba, "Arabic sign language recognition using leap motion sensor," in 2014 9th International Conference on Computer Engineering & Systems (ICCES), Cairo, Egypt, Dec. 2014, pp. 368–373. doi: 10.1109/ICCES.2014.7030987.
- [26] Joukhadar, N. Ghneim, and G. Rebdawi, "Impact of Using Bidirectional Encoder Representations from Transformers (BERT) Models for Arabic Dialogue Acts Identification," *Ingénierie Systèmes Inf.*, vol. 26, no. 5, pp. 469–475, Oct. 2021, doi: 10.18280/isi.260506.
- [27] Elmadany, S. Abdou, and M. Gheith, "Improving Dialogue Act Classification for Spontaneous Arabic Speech and Instant Messages at Utterance Level," 2018, doi: 10.48550/ARXIV.1806.00522.