An IoT-based Framework for Detecting Heart Conditions using Machine Learning

Mona Alnaggar[1], Mohamed Handosa[2], T. Medhat[3], M. Z. Rashad[4]
Robotics and Intelligent Machines Department-Faculty of Artificial Intelligence, Kafrelsheikh University, Egypt[1]
Computer Science Department-Faculty of Computers and Information, Mansoura University, Egypt[2, 4]
Electrical Engineering Department-Faculty of Engineering, Kafrelsheikh University, Egypt[3]

***Abstract*—A lot of diseases may be preventable if they can be analyzed or predicted from patient historical and family data. Predicting diagnosis depends on the gathered clinical and physiological data of patients. The more collected clinical and medical healthcare data, the more knowledge the medical support system may support. Hence, real monitoring clinical and healthcare data for patients is the trend of this decade based on Internet of Things technologies (IoT). IoT models facilitate human life by easily collecting clinical data remotely for recognizing diseases that are easily treatable if it is diagnosed early. This paper proposes a framework consisting of two models: (i) heart attack detection model (HADM); (ii) Electrocardiosignal ECG heartbeat multiclass-classification model (ECG-HMCM). Gridsearch is used to the hyperparameters optimization for different machine learning (ML) techniques. The used dataset in HADM consists of 1190 patients and 14 features. As the foundation of diagnosing cardiovascular disease is arrhythmia detection hence, we propose an ECG heartbeat multi-class classification model using MIT-BIH Arrhythmia and PTB Diagnostic ECG signals dataset which contains five categories with 109446 samples. K Nearest Neighbor (KNN) technique is applied to build ECG-HMCM in addition to the using of Gridsearch algorithm for hyperparameter optimization aiming to improve the accuracy of classification which achieved 97.5%. The proposed framework aims to facilitate human life by easily collecting clinical data remotely. The outcomes of the experiments show that the suggested framework works well in a practical setting.**

Keywords—Medical healthcare; medical support system; real monitoring; Internet of Things (IoT); ECG signals, Gridsearch; hyperparameter optimization

I. INTRODUCTION

Heart disease and other cardiovascular diseases (CVDs) are among the prominent reasons of mortality in the globe[1]. Corresponding to the World Health Organization (WHO), around 31% of fatalities globally are caused by cardiac disorders[2]. Furthermore, almost a quarter of a million Americans lose their lives each year to heart attacks. Heart disease is difficult for medical professionals and specialists to diagnose since it takes knowledge and experience to do so. Cardiac disorders, which are roughly categorized as various sorts of faulty heart problems, are responsible for about 1 in every 4 fatalities. However, diagnosing CVDs requires human analysis of data from several clinical tests, which takes time.

To give medical professionals faster analysis by lowering the time it takes to get a diagnosis and improving outcomes, new methods for automating the identification of such anomalies in human heart diseases should be created. Electronic health records (EHRs) [3] are frequently used to find insightful data patterns that enhanced Machine Learning(ML) algorithms' ability to anticipate. ML makes a substantial contribution to the resolution of problems like forecasts in numerous fields, such as healthcare. Doctors and healthcare workers always face different epidemics. They suffer from their infection by these epidemics during their work. Hence, researchers and scientists must support them with new innovative ideas, tools, smart devices, and medical systems aiming to eliminate the increased infection and protect medical staff and patients from these epidemics. Applying ML and DL algorithms for prediction and diagnosing supports many medical branches such as oncology[4], surgery[5], Dentistry [6], Cardiology [7] Fetal Heart Diagnosis [8], diseases classification [9], and remote respiration rate monitoring [10, 11]. To aid in everyday medical inspections, a multifunctional, portable health monitoring device was developed and put into use[12].

Even though most feature-based ML approaches employ Heart rate variability (HRV) analysis to diagnose Electrocardiosignal (ECG), their robustness cannot be assured. Most crucial HRV characteristics are always influenced by unrehearsed variations, breathing, medication interactions, age, and gender. As an outcome, analyzing HRV should not serve as the major pillar for evaluating heart disorders. Most current ECG classification techniques offer accurate classification outcomes that discriminate between CHF and NSR patients. Building an effective automated framework that can correctly differentiate CHF, ARR, and NSR scenarios and operate in real-time with minimal hardware complexity is still quite challenging.

To the extent of our knowledge, all previous works focus on either heart attack detection or ECG classification. This work combines the two objectives in an IoT based framework with a notification system. Additionally, the highest accuracy achieved for heart attack detection is 92.28%. This accuracy is still to be enhanced. This work tries to enhance the accuracy of heart attack detection based on the combination of two different models for building the proposed framework. Therefore, improving heart disease detection models using ML algorithms is essential for spotting heart illness early on. However, using ML technique requires figuring out how the heart failure dataset's characteristics relate to one another. The main contribution of this paper is to:

- Propose a multi-class classification model for heart attack detection.
- Enhance the accuracy of heart attack detection model using ML techniques.
- Propose a classification model for multi-class classification of ECG readings that represent the heartbeat.
- Optimize the finding results of ECG heartbeat signals classification.
- Propose an IoT framework with a notification system to provide suitable alerts or assist in the medical decision support systems and real monitoring for patients.

The rest of the paper is designed to contain the related work in Section II. The material and methods of the proposed models will be presented in Section III. Section IV will discuss implementation and methods. Section V will show the achieved results with discussions. Finally, the conclusion of our proposed paper and the future work will be discussed in Section VI.

II. Related Work

Currently, transforming traditional healthcare services depends on applying models of ML on the clinical and healthcare data. Applying ML techniques in the medicine and healthcare area have proven great results in different specializations such as using medical images for body organs recognition, tumor detection, reconstruction of medical images, lung nodules detection and diseases classification, to name a few[13]. The doctor may monitor the patient's health, wellness, and health using an IoT medical gadget. The history of physiological medical tools and procedures is simple to find. The patient's physiological indicators should be timely monitored to prevent any health issues before they arise[14]. Sensors are adopted with real monitoring systems. This adaptation concerned the user's daily life, which is very simple to enable the collected data to be simply recorded over long hours. Through this means, collected data can be recorded over long hours to produce a huge amount of data. These data need processing and transformations to be more useful when applying Artificial intelligence algorithms. This is for increasing accuracy, decreasing the error rate, or improving computational efficiency to support suitable decisions. The IoT Monitoring system consists of three sections, the network of used medical sensor devices, IoT cloud, and graphical user interface (GUI). Hence, this section will mention some related work in (i) heart attack detection, (ii) ECG heartbeat signal classification.

A. Heart Attack Detection

One of the important techniques of DL is the Convolutional neural network (CNN) which is employed for building applications of image processing. There are many other techniques of DL have common uses[15]. Diagnosing diseases is a very attractive research point. It is the hope for patients to live a more easily life, especially, if it is prevented from being infected with diseases according to analyzing and predicting of patient's data. Using UCI repository and CNN for possible heart disease prediction[16, 17] supported in the medical decision for patients. In this study [18], a distance-based ensemble for the KNN (k Nearest Neighbor) approach is introduced, and the results of its use in the diagnosis of heart disease are shown. There are two ensemble formations in use. One employs three distances, the previous five. For all the formations and versions, the evaluation was with the Cleveland dataset of UCI heart disease, this ensemble had an accuracy of close to 85%. The dataset for Cleveland heart disease was employed in the suggested study, and data mining algorithms, including regression and classification were applied. Random Forest (RF) and Decision Tree (DT) algorithms are used. Three ML methods are utilized in the proposed implementation: RF, DT, and Hybrid Model (Hybrid of random forest and decision tree). According to experimental findings, the achieved accuracy based on hybrid model was 88.7%. A hybrid model of DT and RF was utilized to forecast heart disease using the user's input response in the interface[19]. One of the most widely utilized ML algorithms, KNN, is frequently used for data categorization. Analyzing each patient's health metrics can also help forecast heart disease. The accuracy of the parameter comparison approach can be improved by combining KNN. The accuracy of the 13 definite parameters obtained by this study, which employs UCI ML dataset, was 86%[20]. Verma et al., for instance, introduced a hybrid model for the prediction of heart illness using particle swarm optimization (PSO) and ML classifiers, namely K-nearest neighbor and multi-layer perceptron (MLP), which reached a 90.28% accuracy [21]. Five ML algorithms were used by Alotaibi et al.[22] (LR, RF, Naive Bayes, DT, SVM). Rapid Miner is utilized for implementation, and it produces good results when compared to prior studies. 1013 patient records from the Cleveland dataset (UCI) were employed. Shah et al.[23]'s efficient model applies KNN, DT, RF, and Naive Bayes while utilizing just 14 key features from the Cleveland dataset from UC Irvine. The best outcome is produced by KNN among these algorithms. Another cardiac disease detection model was created by Jindal et al.[24]. To build a model, RF, LR, and KNN algorithms are applied. Of RD and LR, KNN provides the best accuracy.

B. ECG Heartbeat Signal Classification

The gold specification for noninvasive diagnosis of several forms of cardiac problems is the ECG. When identifying CVDs, the ECG signal is quite important. We can learn about the heartbeat from the ECG readings. Cardiac arrhythmia can be found using ECGs. ECG is a non-invasive diagnostic that measures the electrical activity of the heart muscles to evaluate heart function. Since it gives cardiologists all the information they need regarding heart problems, ECG serves as an effective tool for detecting a variety of cardiac illnesses. Yet, a critical factor in increasing treatment options and slowing the course of CHF is early diagnosed of the condition. The arrhythmia is a significant cardiac condition that contributes to several instances of sudden death (ARR). ARR stands for aberrant heart rhythm, which is a result of erratic heart rate. The provision of quality care necessitates accurate patient diagnosis as well as the identification of appropriate treatments while avoiding incorrect diagnoses. Early CVD identification also lowers costs and decreases CVD mortality. A classification method, which is crucial in clinical research, may be used to do the task effectively and inexpensively with data mining approaches.

ECG data is made up of a collection of electrical impulses from the heart that are collected, and it may be used to spot several different cardiac disorders including arrhythmias, coronary artery disease, and cardiomyopathies. One of the most crucial techniques for assessing the general condition of the heart is heart rate variability (HRV). It shows how flexible the cardiac system is to changes in internal and external stimuli. The fluctuation in the interval between successive heartbeats is called a HRV time series (RR intervals) [25]. Congestive heart failure (CHF) is a dangerous cardiac condition that significantly raises death rates across the world. With CHF, the heart can no longer pump blood across the body efficiently enough to meet tissue demands for oxygen and metabolism. CHF affects more than 26 million individuals worldwide, with 3.6 million new cases added each year[26].

Robustness, efficiency and accuracy of applying ML/ DL techniques in healthcare encourage researchers to keep going to improve and enhance the used methodologies in medical systems, such as the proposed DL Modified Neural Network (DLMNN) to support in medication, particularly in heart diseases diagnosing by using a wearable IoT sensor device, then saving these normal and abnormal signals on cloud [27]. The information gained on this job can be used to the myocardial infarction (MI) classification challenge, according to a recommended strategy. On the MIT-BIH and PTB diagnostics datasets from PhysionNet, the proposed methodology is assessed. The findings show that the proposed technique can predict arrhythmia classification with average accuracies of 93.4% and 95.9%, respectively[28]. This paper [29] suggested an enhanced ResNet-18 model for ECG heartbeat signals classification of based on a convolutional neural network (CNN) method through suitable model training and parameter tuning. The model's distinct residual structure allows for the deepening of CNN layered structure to improve classification performance. The MIT-BIH arrhythmia database application findings show that the suggested model attains greater accuracy (96.50%). The suggested technique[30] is tested on the benchmark MIT-BIH arrhythmia database before being prototyped on a commercially accessible ARM-based embedded device. Moreover, the prototype is assessed using two schemes, namely class and customized schemes, and both schemes indicated greater overall accuracy for state-of-the-art CVDs diagnosis than the current works, at 96.29% and 96.08%, respectively.

III. Material and Methods

In this section, the steps of the proposed framework will be discussed as a preview architecture to facilitate decision making for doctors and remotely support patients with medical decisions. Our proposed architecture aims to achieve a more compatible framework to follow up the patient using IoT components. The framework phases can be listed as the following:

Collecting Data phase: depending on the selected IoT sensor for gathering the patient's medical, then start the phase of gathering data.

Data Preparation and cleaning: data processing, this step is responsible for raw data changing to be more useful.

Model selection: according to the issue and collected data, the algorithm should be implemented.

Train the model: using Artificial Intelligence Algorithms to enhance the accuracy of the prediction or decision making.

Model Evaluation: evaluating the proposed model and comparing it with others.

Notification system: when using the proposed system, and abnormality of signals appear then the family and doctors receive alerts about the case to take a right decision.

Even though most feature-based ML approaches employ HRV analysis to diagnose ECGs, their robustness cannot be guaranteed. Most crucial HRV characteristics are always influenced by unrehearsed variations, breathing, medication interactions, age, and gender. As a result, analyzing HRV should not serve as the major pillar for evaluating heart disorders. By creating novel diagnostic procedures for heart disorders, these difficulties can be eliminated.

The proposed framework consists of several phases such as cleaning data, storage of data, analysis data, implementing models, and notification system. Fig. 1 shows the basis of the proposed framework which oversees gathering physical data from the human body's surface and transmitting it wirelessly to the IoT cloud. Then, the signals are put through a variety of steps to increase their quality and make them suitable for wireless transmission, such as amplification, filtering, etc. Typically, a smart terminal (for example: a smartphone or computer) is required to receive the data before transmitting it wirelessly to the IoT cloud. After making data processing, artificial intelligence algorithms applied to predict if there will be any issue. If it exists, the monitoring system sends alarm to the connected computer or smart phone. As the foundation of diagnosing cardiovascular illness is arrhythmia detection, we propose this framework which consists of two different models: (i) Heart attack detection and (ii) ECG heartbeat signals classification.

A. *The Heart Attack Detection Model (HADM)*

A heart attack, also identified as a myocardial infarction, happens when a part of the heart muscle is not receiving enough blood, according to the center of disease control, and prevention (CDC). Hence, with the help of the dataset's variables, we will create a model to forecast the chance of a heart attack. ML techniques can be highly helpful in the early detection and treatment of patients with cardiovascular disease or who are at high cardiovascular risk beneficial (due to the presence of one or more risk factors, including hypertension [31], diabetes, hyperlipidemia [32], or previously existing illness).

Some of the preparation steps were executed such as dropping duplicates, Outlier Detection, Features encoding such as Binary features encoding and One-hot encoding were applied on the used dataset. Dividing dataset into training and testing datasets is executed for building and evaluating the HADM.

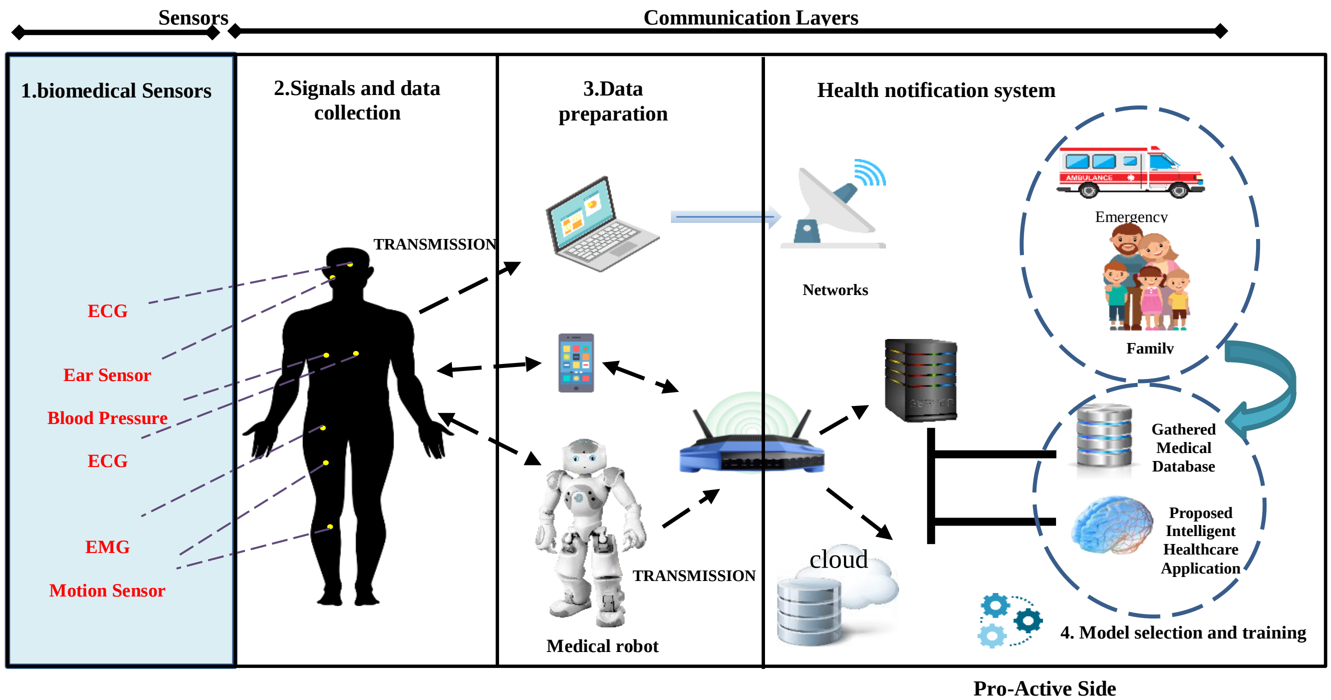


Fig. 1. Proposed framework phases to enhance the performance of the medical system based on the IoT sensors.

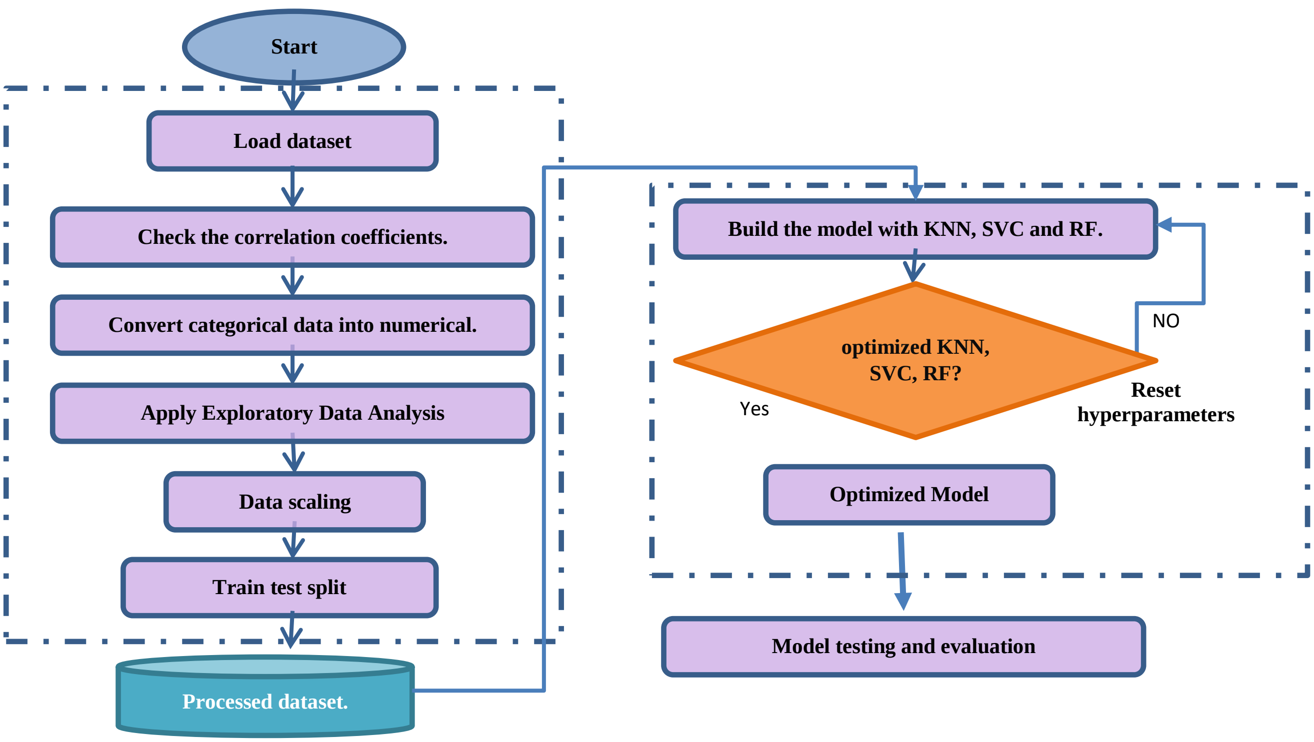


Fig. 2. Flowchart of the proposed Heart Attack Detection Model (HADM).

The proposed model depends on using the K Nearest Neighbors algorithm, support vector machine classifier (SVM) and random forest (RF)to determine whether cardiac disease is present or not. In Fig. 2, the HADM flowchart will be shown. Gridsearch technique is used for hyperparameters optimization. These parameters might be a continuous variable, a categorical variable, or an integer with values spanning from the lower to upper bounds. As hyperparameters remain constant throughout training, the model's accuracy is increased while training time and memory requirements are concurrently decreased. Different models employ a variety of hyperparameters depending on the issue description. There are no hyperparameters that are optimal for all models.

B. ECG Heartbeat Signals Multi-class Classification Model (ECG- HMCM)

A rigorous and consistent examination by cardiologists is required, which is difficult and time-consuming, to give an effective and precise diagnosis of ARR and CHF. For the proper detection of cardiac disorders, a completely automated diagnosis system is urgently required. The creation of diagnostic systems can help cardiologists diagnose ECG recordings quickly and accurately while also cutting down on the time and money needed for clinical interpretation. The heartbeat classification dataset from the MITBIH is used in this suggested model. Up sampling is done such that there is the same number of instances in each of the five classes, each of which has a different number of samples. The KNN model is trained on the data once it has been divided into training and testing set. The test data is run through the trained model. Individual successfully and erroneously categorized signals are examined in terms of correlation to the class mean as well as plotted to examine explainability. Here, we looked at whether such affordable and straightforward algorithms may be useful enough for clinical usage and pave the way to better services. The model architecture will be shown in Fig. 3. The steps of applying the proposed model will be summarized in the next algorithm.

The proposed algorithm for ECG heartbeat multi-class classification model
Step 1) *load the data set,*
Step 2) *define classes,*
Step 3) *assign a data frame to each class,*
Step 4) *extract shape of the training and test sets, Upsample the minority classes.*
Step 5) *split original training data into a training and test subset (test_size =0.2)*
Step 6) *apply KNN algorithm on the dataset.*
Step 7) *attributes are ranked according to their value.*
Step 8) *apply* ***(Grid search)*** *for hyperparameters optimization.*
Step 9) *estimate the classifier's accuracy, which measures the capability of the classifier to classify heartbeat ECG signals correctly.*

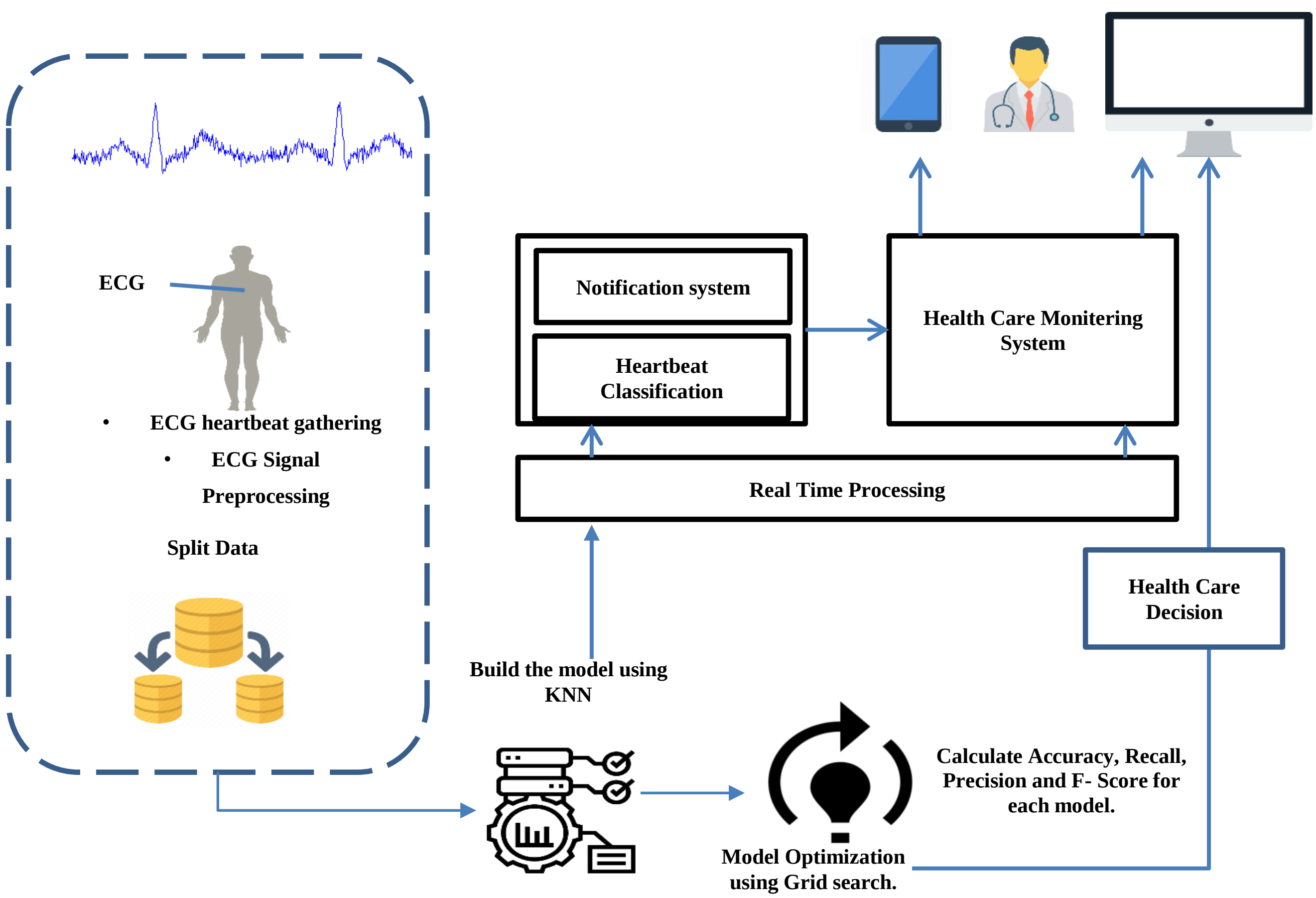


Fig. 3. ECG heartbeat multi-class classification model architecture.

IV. Implementation and Methods

In this section, the experiments of the proposed framework will be discussed which are divided into (i) heart attack detection, and (ii) ECG heartbeat signals classification, in addition to the description of the used datasets for both models.

To support the implementation of our suggested modal, our experiments were implemented in Python notebook version (6.4.8) and several open-source libraries. This system makes use of the packages Matplotlib, NumPy, sklearn, and Keras. It was carried out a laptop with a 2.30GHz, 11th-generation Intel Core (TM) i7-11800H processor, and 16.0 GB (15.8 GB usable).

A. Heart Attack Detection Model (HADM)

The heart is a vital organ in humans. It sends blood to all the organs in our body. If it malfunctions, the body's other organs will stop working and the person will pass away in a matter of seconds. Several different things can lead to heart failure. These components have been split up by scientists into two categories: risk variables that cannot be changed and risk factors that can be. Historical Family data [33], gender, and age are risk aspects that cannot be changed. Excessive cholesterol, smoking [34], inactivity, and blood pleasure are risk factors.

The heart attack detection model depends on ML algorithms, namely, KNN, SVM and RF. The KNN algorithm is distinguished by its straightforward design and practical use. KNN was selected because it can compete with the most accurate models and offers incredibly exact predictions. How accurate the forecasts are determined by the distance. KNN technique can be used in applications that demand great accuracy as a result. The SVM algorithm was also selected to participate in our classification model because of its mathematical base in statistical learning theory which provides a systematic solution to ML challenges. A portion of the training input is used to create the solution via SVM. Classification, regression, novelty detection, and feature reduction problems have all seen widespread usage of SVM.

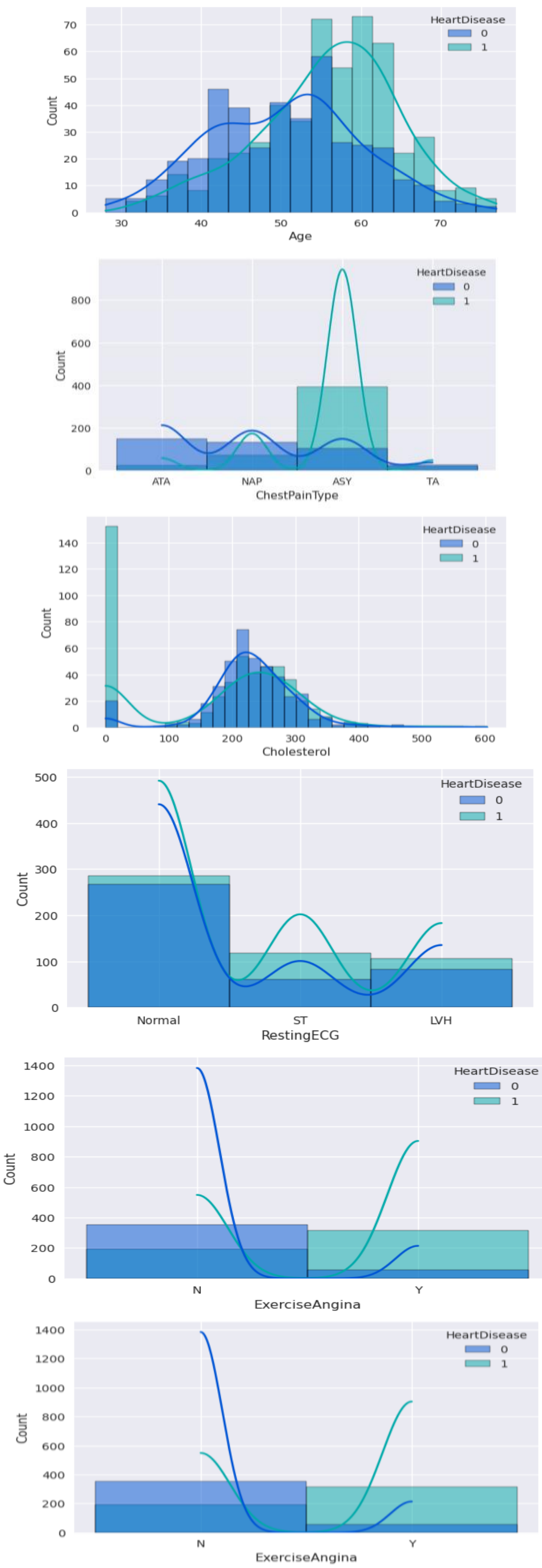
Random forest classifier is also used to build our model. A method for lowering the variance of an estimated prediction function is RF. Regression and classification are also possible uses for RF. A random forest collects class votes from each tree before classifying using a majority vote when used for classification. The predictions from each tree at the target point x are simply summed when used for regression. When used for classification tasks, RF gets a class vote for each tree and then classifies the data using the results of the majority vote. The predictions from all trees at a target point x are merely averaged when RF is used to regression problems. The values of the parameters that are most appropriate for a job rely on those parameters, which are understood and employed as tuning parameters. Then, Gridsearch is separately applied to each technique to get optimized hyperparameters. To assess the intensity of the key elements that influence the prognosis of heart disease, a generalized algorithm was developed and optimized. Cleveland, Hungary, Switzerland, Long Beach and Stalog (CHSLBS) datasets were used to assess the models. The selected ML algorithms were used to build the proposed models for classification. Also, we will use grid search for tuning models' hyperparameters, additionally we will evaluate their performance.

1) The used dataset: The original source for this dataset was the [35]. The dataset variables are mentioned in [35]. By combining multiple datasets that were previously available separately, one dataset was created. This dataset, which includes five heart datasets including eleven features in public, is the greatest heart disease dataset currently available for research objectives. The five datasets utilized for its curation are: Cleveland: 303 observations; Hungarian: 294 observations; Switzerland: 123 observations; Long Beach VA: 200 observations; and 270 observations in the Stalog (Heart) (CHSLBS) datasets. The total number of observations is 1190, but we found 272 duplicated rows. Hence, the preprocessing step is to drop this duplication to finally get 918 rows. Analyzing CHSLBS shows that there are 508 with heart disease and 410 normal cases. This leads us to have a balanced dataset. It was found that the presence of HeartDisease is substantially linked with all numerical characteristics. Hence, we will first get dummies for categorical characteristics. In our experiment, some dataset analysis has been extracted such as maximum, minimum, average, and median to determine whether the data indicate a skewness distribution. Each of these traits is looked at individually because they all have different units and meanings.

The used dataset [35] contains different attributes. Sex, FastingBs and Exercise angina are binary data. Chest Pain type, Resting ECG and ST_ slope are categorical data, and the continuous set consists of Age, Resting blood pressure, cholesterol, MaxHR and old peak. Some of preparations steps were executed such as dropping duplicates, Outlier Detection, Features encoding such as Binary features encoding and One-hot encoding were applied on the used dataset. 4.58% of instances that are outliers in resting blood pressure (restingBP) were removed. Since characteristics like cholesterol and resting blood pressure are exceptional situations that do not accurately reflect the state of the general population and are not suitable for training our model of predictions, preprocessing is necessary.

Exploratory Data Analysis EDA [36] seeks to condense descriptions to make them easier to process with the cognitive capacity at hand; and tries to make the description more effective by looking behind the previously mentioned surfaces. The used dataset needs some steps of preprocessing and exploratory data analysis. First, we investigate the correlations between the patient's numerical characteristics and the target column. It was found that the presence of cardiac disease is closely associated with all numerical characteristics. Then, the categorical data is handled to be numerical.

The distribution analysis is a graphical representation of a dataset's distribution that displays the frequency of data points within various intervals. The histogram for distribution analysis of the features is a common graphing technique used to include both continuous and discrete data that was collected on an interval scale. It is widely used to conveniently illustrate the main ideas of data distribution. The conclusion after applying EDA on the used dataset shows the following information: the risk of suffering CVDs increases with age. Fig. 4 shows the histogram for distribution analysis of some features such as age, oldpeak, resting Blood Pressure, and cholesterol.



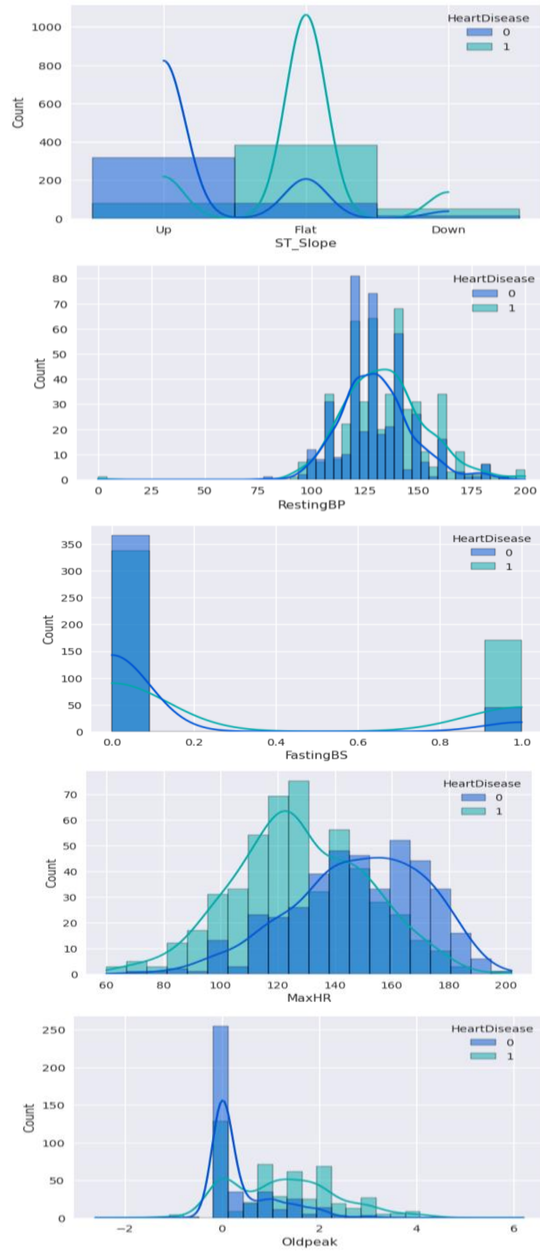
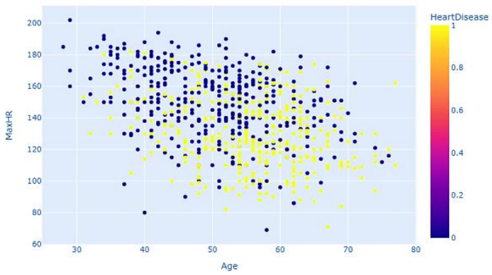
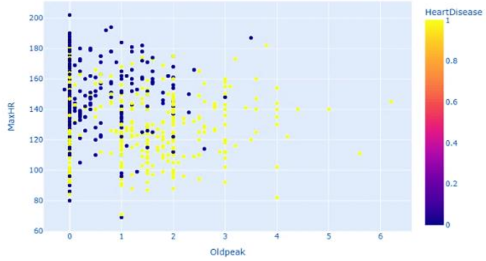


Fig. 4. Distribution analysis of some dataset features.

Apparently, the men have a major risk than women of suffering CVDs. Most cases of CVDs present absence of chest pain or the usual anginal equivalents followed by the cases that present chest pain non-anginal. We also see in this histogram a slight increase in CVDs cases when increasing the resting blood pressure. Fig. 5 will show the histogram for distribution analysis according to Age and oldpeak. Fig. 6 shows the correlation between dataset features.



(a) Heart disease dataset distribution according to age.



(b) Heart disease dataset distribution according to oldpeak.

Fig. 5. The histogram for distribution analysis according to Age, oldpeak.

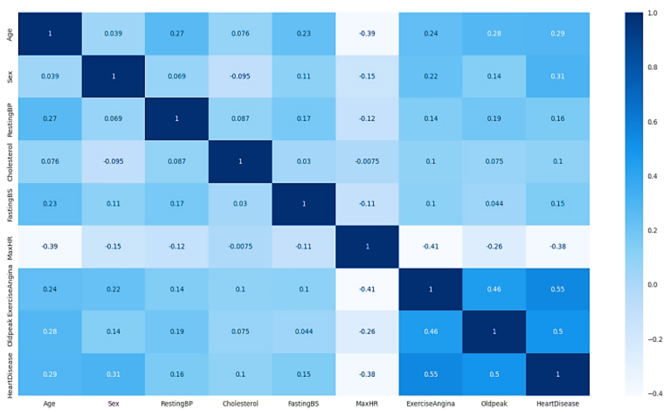


Fig. 6. The correlation between dataset variables.

B. ECG Heartbeat Signal Multi-class Classification

A default KNN model is trained on the data once it has been divided into a training and testing set. The test data is run through the trained model. Individual successfully and erroneously categorized signals are examined in terms of correlation to the class mean as well as plotted to examine explainability. Gridsearch algorithm is used to get the optimized k-parameter, then the model is retrained on the original test set by achieving an enhancement of accuracy.

1) The used datasets: Two collections of the signals of heartbeat that add up to this dataset are derived from two well-known datasets in classification of heartbeat, the MIT-BIH Arrhythmia Dataset [37] and The PTB diagnostic ECG database [38]. The MITBIH heartbeat dataset categorization is the subject of this experiment. The dataset contains five classes with these labels (i) normal beat ('N'): 0, (ii) supraventricular ectopic beats ('S'): 1, (iii) ventricular ectopic beats ('V'): 2, (iv) fusion Beats ('F'): 3, (v) unknown Beats ('Q'): 4. Upsampling is done such that there are the same number of instances in each of the five classes, each of which has a various number of samples. Both sets provide a large enough number of samples to support deep neural network training. This dataset has been used to research the classification of heartbeats and to investigate some of the transfer learning possibilities using deep neural network architecture. The signals correspond to the ECG types of heartbeats for both the normal case and cases affected by different arrhythmias and myocardial infarction. These signals are preprocessed and divided into segments, each of which represents a heartbeat. Arrhythmia Dataset consists of five categories which are labeled to be ['N': 0, 'S': 1, 'V': 2, 'F': 3, 'Q': 4] with 109446 samples, the sampling frequency is 125HZ and the source is from Data Source: Physionet's MIT-BIH Arrhythmia Dataset. The PTB Diagnostic ECG dataset contains two categories with 14552 samples and the source of data is from Physionet's PTB Diagnostic dataset.

C. Evaluation Matrix

This section presents the results of the proposed models and evaluates their performance using AUC. The AUC has been chosen according to the providing equations[39]. Eq. (1) - (5) described the evaluation way of the proposed models. Precision, recall or sensitivity, f1-score, and other model performance indicators were also assessed, in addition to accuracy. To evaluate the model's performance, the true positive, true negative, false positive, and false negative values must first be established. When the model correctly identified positive and negative predictions, they are referred to as true positive and negative values. False positives and negatives, on the other hand, occur when the model incorrectly divides positive and negative predictions into positive and negative categories.

Accuracy:

$$Accuracy = \frac{True\ Negative + True\ Positive}{Total\ Data} \tag{1}$$

Precision:

$$Precision = \frac{True\ Positive}{True\ \ Positive + False\ Positive} \tag{2}$$

Recall:

$$Recall\ (Sensitivity) = \frac{True\ Positive}{True\ \ Positive + False\ Negative} \tag{3}$$

Specificity:

$$Specificity = \frac{True\ Negative}{True\ \ Negative + False\ Positive} \tag{4}$$

F1 score:

$$F1 = \frac{2x\ Precision\ x\ recall}{Precision + \ recall} \tag{5}$$

D. Model Optimization

Hyperparameters are peripheral parameters[40] that are not a model component and cannot be predicted from the dataset. For classification challenges, a user must manage the hyperparameters setting procedure, which varies for various algorithms and datasets, to obtain high accuracy. Finding a tuple that suggests an ultimate model and minimizes the loss function is the goal of optimization [41]. A Gridsearch algorithm was used to optimize the hyperparameters for the experimental classification model when developing HADM using KNN, SVC and RF classification techniques. Table I lists the top hyperparameters for each used technique.

TABLE I. THE OPTIMIZED HYPERPARAMETERS FOR THE USED TECHNIQUES

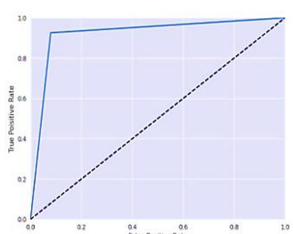
Technique	Best hyperparameters
KNN	'n_neighbors': 9, 'p': 2, 'weights': 'distance'
SVC	'C': 1.0, 'gamma': 'auto', 'kernel': 'sigmoid'
RF	'criterion': 'gini', 'max_features': 30, 'n_estimators': 700,

V. RESULTS AND DISCUSSIONS

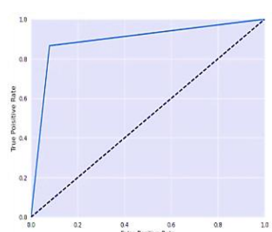
This evaluation section consists of (i) heart attack detection model, (ii) ECG heartbeat signal multi-class classification to be proposed models that can early support with an appropriate decision for saving patients.

A. *Heart Attack Detection Results*

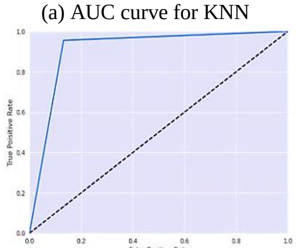
The heart attack detection model depends on the use of KNN, SVC and RF classifiers achieved the best accuracy compared with another previous research when applied on the greatest heart disease dataset which consists of five datasets with 1190 observations formed (CHSLBS) datasets. During the execution of the experiments, Gridsearch is applied to optimize the hyperparameters for achieving higher accuracy. Area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity were among the results that were reported. AUC curves will be shown in Fig. 7 for KNN, SVC and RF. Table II shows the summarized results for building the model of Heart Attack Detection.



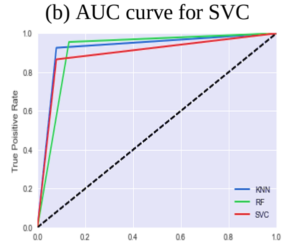
(a) AUC curve for KNN



(b) AUC curve for SVC



(c) AUC curve for Random forest



(d) AUC curve for KNN, RF and SVC

Fig. 7. The dataset's AUC curve for test data using the proposed models.

TABLE II. THE EVALUATION OF HEART ATTACK DETECTION MODEL USING THE EVALUATION MATRIX

Technique	K-Nearest Neighbors	C-Support Vector	Random Forest
Accuracy	92.307692	89.510490	90.909091
Precision	91.176471	90.625000	86.486486
Recall	92.537313	86.567164	95.522388
f1- score	92.310711	89.491892	90.915316

We must find a balance between recall score and accuracy score; in the case of our focus, the Random Forest classifier had the greatest recall score, but also the poorest precision score. Hence, the KNN classifier is the best algorithm in this experiment. This study demonstrates that the RF classifier outperforms other classification algorithms when employing Gridsearch for hyperparameters optimization [40] and evaluating depending on the recall evaluation matrix parameter. KNN achieved higher accuracy compared with random forest and support vector classifier algorithms. Table III shows comparison between the result of our proposed result and other used techniques.

TABLE III. COMPARISON OF ACHIEVED RESULTS AND PREVIOUS RESEARCH

References	Year	The used technique	Accuracy
Saqlain, et al. [42]	2019	SVM and MCC	81.91%
Beunza, et al. [43]	2019	NB, BN, RF, and MP	85.48%.
Kavitha, et al.[19]	2020	Hybrid DT and RF	88%
Kavitha, et al.[44]	2021	KNN with SBS feature selection	90%
Pawlovsky, et al. [18]	2022	KNN	85%
Anderies, et al. [20]	2022	KNN	86%
Chauhan, et al. [21]	2023	PSO, KNN	90.28%
The proposed model	**2023**	**KNN, SVC and RF**	**92.31%, 89.51%, 90.91%**

B. *ECG Heartbeat Monitoring Results*

In this model, KNN is used for classifying the ECG heartbeat for patient monitoring as a proposed solution to assist in the decision-making support health system. The achieved results were optimized according to using the grid search optimizer. The model is Instantiated and performed depending on the use of a grid search for setting the number of neighbors (the k parameter value) fitting five folds for each of five candidates, for a total of 25 fits, for the first 40,000 shuffled training data. In Fig. 8, the preview of 5 ECG beats from the five categories will be shown when using the training dataset.

The finding results have been enhanced 0.5 % after using the grid search hyperparameter optimization and 97.5% Overall classification accuracy. Fig. 9 will preview the signals for the correct prediction according to our proposed model when applied on the test set.

This shows how the suggested approach might help cardiologists improve the precision of ECG diagnosis in real-time clinical situations.

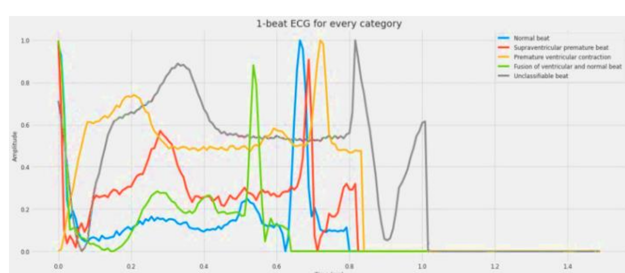


Fig. 8. Previewing 1 beat ECG from each category.

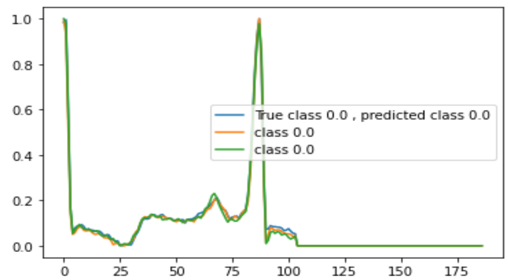


Fig. 9. Sample of correct predicted signal.

The proposed model has some advantages over previous works. The model combines heart rate detection and ECG classification in one IoT based framework with a notification system. The model also enhanced the accuracy of heart attack detection over most previous works. It is anticipated that the created method would make it much easier for doctors to diagnose heart issues in a more practical way. Above all, the study has significantly improved the computation of strength ratings, which are excellent indicators of the prognosis of heart conditions depending on the proposed framework.

VI. Conclusions and Future Work

The leading cause of death is heart disease. Heart disease mortality will decrease because of the identification of key risk factors, the creation of decision support systems, effective control methods, and health education initiatives. If CVDs are detected and treated early, many lives might be saved. Cardiologists cannot manually examine the vast quantity of collected data for a patient to develop a timely medication plan. One of the most widely utilized ML algorithms, KNN, SVM and RF are frequently used for data categorization in our proposed model. The five datasets utilized for its curation are: Cleveland: 303 observations; Hungarian: 294 observations; Switzerland: 123 observations; Long Beach VA: 200 observations; and 270 observations in the Stalog (Heart) (CHSLBS) datasets. The total number of observations is 1190 and the achieved accuracy were 92.3%, 89.51% and 90.91% when applying KNN, SVM classifier and RF recursively with hyperparameters optimization using Gridsearch technique. Analyzing each patient's health metrics can also help to forecast heart disease. Hence, KNN is used for heart attack detection using UCI dataset, in addition to using KNN to build an ECG heartbeat multi class classification model. Two heartbeat signals were examined and evaluated in this study. The Arrhythmia and PTB diagnostic ECG datasets, both of which have 125 Hz sampling rates, each comprise 109446 samples in five categories and 14552 samples in two categories. Using two ECG datasets, the KNN methodology with the hyperparameters optimizer employing Gridsearch algorithm was applied, and higher accuracy has been achieved when compared with previous recent research in this field was attained. The results of this model have been optimized using grid search for hyperparameters optimization. The major goals of the proposed framework are to provide a method that can effectively identify heart attacks and help professionals to choose the best decision for saving human life. In future work, other different ML techniques can be used. By adding more data and testing with more important or statistically produced data, such as numeric data augmentation, the applicable procedure may be enhanced. The writers see this as a future work that may be improved.

References

[1] Centers for Disease Control and Prevention (CDC). Deaths: leading causes. Available: https://www.cdc.gov/nchs/fastats/leading-causes-of-death.

[2] S. Kaptoge, L. Pennells, D. De Bacquer, M. T. Cooney, M. Kavousi, G. Stevens, et al., "World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions," The Lancet Global Health, vol. 7, pp. e1332-e1345, 2019.

[3] M. R. Cowie, J. I. Blomster, L. H. Curtis, S. Duclaux, I. Ford, F. Fritz, et al., "Electronic health records to facilitate clinical research," Clinical Research in Cardiology, vol. 106, pp. 1-9, 2017.

[4] H. Shimizu and K. I. Nakayama, "Artificial intelligence in oncology," Cancer science, vol. 111, pp. 1452-1460, 2020.

[5] P. Mascagni, A. Vardazaryan, D. Alapatt, T. Urade, T. Emre, C. Fiorillo, et al., "Artificial intelligence for surgical safety: automatic assessment of the critical view of safety in laparoscopic cholecystectomy using deep learning," Annals of surgery, vol. 275, pp. 955-961, 2022.

[6] T. Shan, F. Tay, and L. Gu, "Application of artificial intelligence in dentistry," Journal of dental research, vol. 100, pp. 232-244, 2021.

[7] D. Itchhaporia, "Artificial intelligence in cardiology," Trends in cardiovascular medicine, 2020.

[8] S. Liu, Y. Sun, and N. Luo, "Doppler Ultrasound Imaging Combined with Fetal Heart Detection in Predicting Fetal Distress in Pregnancy-Induced Hypertension under the Guidance of Artificial Intelligence Algorithm," Journal of Healthcare Engineering, vol. 2021, 2021.

[9] M. Alnaggar, M. Handosa, T. Medhat, and M. Z Rashad, "Thyroid Disease Multi-class Classification based on Optimized Gradient Boosting Model," Egyptian Journal of Artificial Intelligence, 2023.

[10] A. I. Siam, N. A. El-Bahnasawy, G. M. El Banby, A. Abou Elazm, and F. E. Abd El-Samie, "Efficient video-based breathing pattern and respiration rate monitoring for remote health monitoring," JOSA A, vol. 37, pp. C118-C124, 2020.

[11] M. Alnaggar, A. I. Siam, M. Handosa, T. Medhat, and M. Rashad, "Video-based Real-Time Monitoring for Heart Rate and Respiration Rate," Expert Systems with Applications, p. 120135, 2023.

[12] A. I. Siam, M. A. El-Affendi, A. Abou Elazm, G. M. El-Banby, N. A. El-Bahnasawy, F. E. Abd El-Samie, et al., "Portable and Real-Time IoT-Based Healthcare Monitoring System for Daily Medical Applications," IEEE Transactions on Computational Social Systems, 2022.

[13] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and robust machine learning for healthcare: A survey," IEEE Reviews in Biomedical Engineering, vol. 14, pp. 156-180, 2020.

[14] R. Karthick, R. Ramkumar, M. Akram, and M. V. Kumar, "Overcome the challenges in bio-medical instruments using IOT–A review," Materials Today: Proceedings, vol. 45, pp. 1614-1619, 2021.

[15] S. Serte, A. Serener, and F. Al-Turjman, "Deep learning in medical imaging: A brief review," Transactions on Emerging Telecommunications Technologies, vol. 33, p. e4080, 2022.

[16] A. Mehmood, M. Iqbal, Z. Mehmood, A. Irtaza, M. Nawaz, T. Nazir, et al., "Prediction of heart disease using deep convolutional neural networks," Arabian Journal for Science and Engineering, vol. 46, pp. 3409-3422, 2021.

[17] M. M. Eisa and M. H. Alnaggar, "Hybrid Rough-Genetic Classification Model for IoT Heart Disease Monitoring System," in Digital Transformation Technology, ed: Springer, 2022, pp. 437-451.

[18] A. P. Pawlovsky, "An ensemble based on distances for a kNN method for heart disease diagnosis," in 2018 international conference on electronics, information, and communication (ICEIC), 2018, pp. 1-4.

[19] R. Indrakumari, T. Poongodi, and S. R. Jena, "Heart disease prediction using exploratory data analysis," Procedia Computer Science, vol. 173, pp. 130-139, 2020.

[20] A. Anderies, J. A. R. W. Tchin, P. H. Putro, Y. P. Darmawan, and A. A. S. Gunawan, "Prediction of Heart Disease UCI Dataset Using Machine

Learning Algorithms," Engineering, MAthematics and Computer Science (EMACS) Journal, vol. 4, pp. 87-93, 2022.

[21] A. Chauhan, A. Jain, P. Sharma, and V. Deep, "Heart disease prediction using evolutionary rule learning," in 2018 4th International conference on computational intelligence & communication technology (CICT), 2018, pp. 1-4.

[22] F. S. Alotaibi, "Implementation of machine learning model to predict heart failure disease," International Journal of Advanced Computer Science and Applications, vol. 10, 2019.

[23] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," SN Computer Science, vol. 1, pp. 1-6, 2020.

[24] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," in IOP conference series: materials science and engineering, 2021, p. 012072.

[25] H. ChuDuc, K. NguyenPhan, and D. NguyenViet, "A review of heart rate variability and its applications," APCBEE procedia, vol. 7, pp. 80-85, 2013.

[26] P. Ponikowski, A. A. Voors, S. D. Anker, H. Bueno, J. G. Cleland, A. J. Coats, et al., "2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure," Kardiologia Polska (Polish Heart Journal), vol. 74, pp. 1037-1147, 2016.

[27] S. S. Sarmah, "An efficient IoT-based patient monitoring and heart disease prediction system using deep learning modified neural network," Ieee access, vol. 8, pp. 135784-135797, 2020.

[28] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "Ecg heartbeat classification: A deep transferable representation," in 2018 IEEE international conference on healthcare informatics (ICHI), 2018, pp. 443-444.

[29] E. Jing, H. Zhang, Z. Li, Y. Liu, Z. Ji, and I. Ganchev, "ECG heartbeat classification based on an improved ResNet-18 model," Computational and Mathematical Methods in Medicine, vol. 2021, 2021.

[30] S. Raj and K. C. Ray, "A personalized arrhythmia monitoring platform," Scientific reports, vol. 8, pp. 1-11, 2018.

[31] I. Shiue, "Are urinary polyaromatic hydrocarbons associated with adult hypertension, heart attack, and cancer? USA NHANES, 2011–2012," Environmental Science and Pollution Research, vol. 22, pp. 16962-16968, 2015.

[32] J. Stewart, T. McCallin, J. Martinez, S. Chacko, and S. Yusuf, "Hyperlipidemia," Pediatrics in review, vol. 41, pp. 393-402, 2020.

[33] K.-T. Khaw and E. Barrett-Connor, "Family history of heart attack: a modifiable risk factor?," Circulation, vol. 74, pp. 239-244, 1986.

[34] D. Cook, S. Pocock, A. Shaper, and S. Kussick, "Giving up smoking and the risk of heart attacks: a report from the British Regional Heart Study," The Lancet, vol. 328, pp. 1376-1380, 1986.

[35] Heart Failure Prediction Dataset. Retrieved [Date Retrieved] from [Online]. Available: https://www.kaggle.com/fedesoriano/heart-failure-prediction.

[36] G. Wang, B. Zhao, B. Wu, C. Zhang, and W. Liu, "Intelligent prediction of slope stability based on visual exploratory data analysis of 77 in situ cases," International Journal of Mining Science and Technology, vol. 33, pp. 47-59, 2023.

[37] MIT-BIH Arrhythmia Database [Online]. Available: https://www.physionet.org/content/mitdb/1.0.0/.

[38] PTB Diagnostic ECG Database [Online]. Available: https://www.physionet.org/content/ptbdb/1.0.0/.

[39] M. J. Zaki, W. Meira Jr, and W. Meira, Data mining and analysis: fundamental concepts and algorithms: Cambridge University Press, 2014.

[40] H. Alibrahim and S. A. Ludwig, "Hyperparameter optimization: comparing genetic algorithm against grid search and bayesian optimization," in 2021 IEEE Congress on Evolutionary Computation (CEC), 2021, pp. 1551-1559.

[41] I. Priyadarshini and C. Cotton, "A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis," The Journal of Supercomputing, vol. 77, pp. 13911-13932, 2021.

[42] C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," Informatics in Medicine Unlocked, vol. 16, p. 100203, 2019.

[43] J.-J. Beunza, E. Puertas, E. García-Ovejero, G. Villalba, E. Condes, G. Koleva, et al., "Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease)," Journal of biomedical informatics, vol. 97, p. 103257, 2019.

[44] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart disease prediction using hybrid machine learning model," in 2021 6th international conference on inventive computation technologies (ICICT), 2021, pp. 1329-1333.