Implementation of IoT-Based Healthcare Kit

Tanya Chanchalani
Electronics and Communication Engineering,
PES University
Bengaluru, India
tanyachanchalani@gmail.com

Gaurav R
Electronics and Communication Engineering,
PES University
Bengaluru, India
gaurav.gaurav.r53@gmail.com

Bhushan Kiran Munoli
Electronics and Communication Engineering,
PES University
Bengaluru, India
bhushanmunoli01@gmail.com

Sinchitha H V
Electronics and Communication Engineering,
PES University
Bengaluru, India
sinchitha18@gmail.com

Purushotham U
Project Guide,
PES University
Bengaluru, India
purushothamu@pes.edu

2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS) | 979-8-3503-2097-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ECBIOS57802.2023.10218615

***Abstract*—Cardiovascular diseases and Cardiac Arrhythmia are the most familiar reasons for death throughout the world over the last few decades across the world. However, it is difficult to examine patients in all cases accurately, and consultation with a patient for 24 hours by a doctor is not possible as it needs extra patience, expertise, and time. Thus, with ECG sensors, Arduino, and Raspberry Pi, we implemented machine learning models based on K-Nearest Neighbour, Logistic Regression, Support Vector Machine, and Random Forest for heart disease prediction based on the parameters and attributes related to cardiovascular disease. The datasets in this research are available publicly on the UCI website. The early diagnosis of cardiovascular diseases assists in making decisions on lifestyle changes in patients prone to high risk of heart diseases and minimizing the complications. The result of this research canbe a milestone in medicine.**

Keywords—cardiovascular diseases, cardiac arrhythmia, ECG sensors, Arduino Uno, Raspberry Pi, K-Nearest neighbor, logistic regression, support vector machine, random forest classifier, decision trees, data augmentation.

I. Introduction

The heart is a prevalent organ of the human body. We can see a steady increase in heart diseases all around the world. It is known as a silent killer leading and a major cause of death. If not identified in the primary stage, the person's condition might worsen leading to death due to reluctance. In this research, we recommend a solution based on two machine learning models, each of which is for cardiovascular disease and cardiac arrhythmia. Along with the development, three hardware components are used to obtain ECG-related data from the patient. The models aim to predict the presence of the diseases and provide the best solution for the classification and prediction of the two types of heart diseases. Various Machine Learning (ML) models are compared in terms of accuracy and applying the best model for the requirement. For the prediction of cardiovascular diseases (CVD), electrocardiogram (ECG)-related attributes are recorded from the hardware and inputted to the graphic user interface (GUI). Other important features are input directly into the GUI which is built using HTML, CSS, and Flask modules on Python. The GUI for the prediction of cardiac arrhythmia is built using Tkinter on Python and all attribute values are directly inputted into the GUI.

II. Literature Review

We reviewed research papers for a literature survey. In Real-Time Multi-Class Arrhythmia classification in IoT-cloud platform [1] showed Data procurement is a mechanism for collecting raw data. AD8232 sensor is used in the above process. Signal processing is applied to denoise and refactor. Classification performance is enhanced by building a machine learning model with Deep Neural Network (DNN) [1] with three hidden layers. This algorithm is a subset of the Artificial Neural Network (ANN). IoT-Based Health Monitoring Systems are developed using Raspberry Pi and ECG signals [2] in a monitoring system that tracks the heart pulse of the patient. It results in fast and noninvasive monitoring of the heart.

The Wearable Monitoring Structure is for detecting Abnormal Heart [3]. A portable ECG device is made to find abnormal heart conditions and to analyze the data prior to diagnosis. The architecture is a full vertical system with the advantages of the latest technology and includes devices such as ECG electrodes, Bluetooth modules, and IoT modules. This deviceis connected to the cloud. This system informs the family and doctor in charge of any abnormalities that are detected. Reference [4] gives a view on removing the noise from ECG signals by Signal Processing Techniques. It focuses on removing noises from the signal. An Analog filter (Butterworth filter) is considered to do the above filtering task. This type of filter produces sequences with no ripple count.

III. Methodology

A. Dataset

We used a dataset from the Universityof Irvine (UCI) to classify and predict two types of heart diseases [6]. The arrhythmia UCI dataset differentiates between the presence and absence of cardiac arrhythmia andis classified into 16 groups. The cardiovascular UCI dataset consists of 14 important attributes. There are no missing values in the Cardiovascular UCI dataset but the dataset for Cardiac Arrhythmia has 376 missing values. We use Arduino Uno, Rpi4 and ECG AD8232 (Fig. 1).

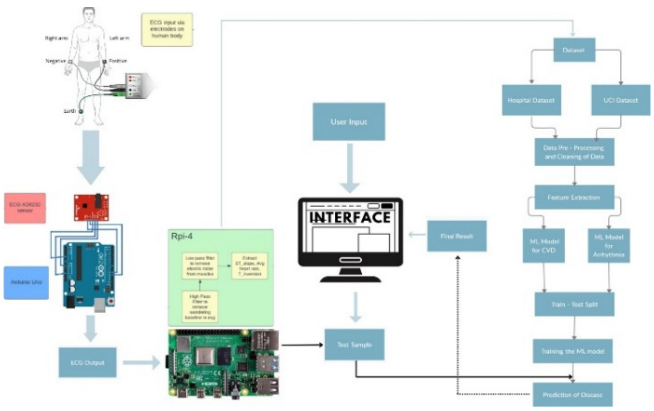

Fig. 1. Block diagram of this research.

B. Hardware Implementation

1) ECG signals

The ECG electrodes are placed in the form of Einthoven's triangle. The red electrode is placed on the right arm, the yellow electrode is on the left arm, and the green electrode is on the right leg. Lead II is extracted from the ECG sensor. A change in the arrangement of the electrodes can leadto a different lead ECG signal according to Einthoven's triangle.

2) Arduino

The signal collected in the Arduino is passed throughthe inbuilt filters in the sensors and a final single lead ECG is passed to the Arduino [7].

3) Communication between Arduino and Rpi

With Universal Asynchronous Receiver-Transmitter (UART) protocol, the data collected by the Arduino is sent to the Rpi. A USB port of Rpi is configured to receive UART packets with a set baud rate of 9600 bps.

4) Noise in ECG signals

When ECG readings are taken, there are afew basic noise signals which are inevitable. Noise includes baseline wandering caused by the breathing of a patient at a frequency of 0.5 to 2 Hz. The power line interference exists at a frequency of 50 to 60 Hz. Noise is caused by the electric activities in the muscles which are usually above 100 Hz. Noise caused by the movement of the electrodes during feature extraction is in the frequency range of1 to 10 Hz. These are the basic noise present in ECG.

5) Removal of noise

The baseline wandering is removed by using a high pass Butterworth filter of cutoff frequency 2 Hz [6]. The power line interference is removed by using a low pass Butterworth filter of cutoff frequency 50 Hz. The y axis represents the magnitude of the ECG signal and the x axis represents the count of the signal in Fig. 2.

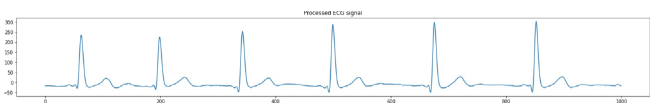

Fig. 2. ECG filtered signal.

6) Peak detection

The P, Q, R, S, and T peaks are detected using the Python library neurokit2. This returns an array of x coordinates of the peaks. This can be further used to detect different parameters.

7) ST slope detection

The array of x and y coordinates of the complete ST considered, first order difference of both the x and y-axes are calculated in the NumPy arrays. They are divided, and the mean of the array is considered as the slope of the ST slope. The ECG signal has 32 ST segments. This results in finding 32 separate slopes. To narrow it down to only one value, the mean of all the slopes is considered. The normal range of the ST slope is 0 to 0.5. Further, efficiency can be increased by detecting the J point in ECG and considering theST slope from the J point to the T point.

8) Maximum heart rate during sample collection

The count of the total number of R peaks in the sample is obtained. The sample is collected for 30 s. To find BPM, the count of R peaks is doubled.

9) T inversion

The baseline in the ECG signal has a variation of 0.1 mV. If the amplitude of the T peak is less than -0.1 mV, then there is a possible T inversion in the ECG signal.

10) Feature Extraction from ECG

There are 3 parameters extracted from ECG: the slope of ST segment, heart rate and presence or absence of Tinversion. The ST segment is the connecting segment s between S and T peaks in ECG. The slope of this line is obtained in two ways. When using the slope, the segment is not a straight line which is inefficient. To reduce the error, the first discrete difference of coordinate values of the ST segment are considered. To find the T inversion, the magnitude of the T peaks is checked if they are less than -0.1 mV. If so, 1 is returned, else 0 is returned. To find the heart rate, the number of R peaks are calculated and doubled as the signal is collected for a time duration of 30 s. These 3 parameters are stored in the .csv file with the age and patient ID. These three parameters obtained from the hardware are the maximum heart rate at the time of collection, ST peak, and T inversion. They are inputs to the interface for CVD along with other important parameters.

C. Implementation of Machine Learning Models

1) Feature selection

The UCI Dataset for CVD prediction consists of multivariate data where attributes are categorical, integer, and real. The considered dataset contains 303 instances and 75 attributes. The dataset is trained and tested with a train-test split ratio of 75:25 (75% training dataset and 25% testing dataset) which makes the dimensions 222 × 75 records. The instance and attributes size is 297 × 14. Hence, the number of attributes considered is 14, amongst the entire number of attributes which is 75. The number of positive cases present in the dataset is 137 and that of negative cases found in the datasetis 160. The results are displayed as two values: 0 and 1. If the result obtained is 0, it indicates no disease. If the result obtained is 1, it indicates the disease is present. Among these, a subset of fourteen attributes is mostly included in published experiments for CVD, which are: age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, num, or the predicted attribute.

The UCI Dataset for the prediction of Cardiac Arrhythmia consists of multivariate data where attributes are categorical, integer, and real. The dataset has 452 instances and 279 attributes out of which 206 attributes are valued linearly and the remaining are nominal. Hence, the total number of positive cases observed is 207, and that of negative cases is 245. The train test split of this dataset is selected to be 70:30 dividing it into 316 and 136 records. Therefore, the output includes 16 classes based on the value of attributes. Class 1 indicates normal cases or no presence of this disease, and classes 2 to 15 of them indicate various types of Arrhythmia. Lastly, class 16 indicates the unknowns due to a lot of missing values in the dataset. Hence, the fifteen attributes selected to predict this disease are Age, Sex, Height, QRS duration, P-R interval, T interval, P interval, Vector angles in degrees on the front plane of QRS, T, P, QRST and J, and Heart rate.

Further, two separate machine learning models are developed and trained using the UCI datasets repository for CVD and Cardiac Arrhythmia. The data required is cumulated, cleaned, and processed. This is done to extract features according to the parameters necessary for the two heart diseases selected. In this approach, multiple ML algorithms are implemented to increase efficiency and get better accuracy.

D. Algorithms

Algorithms used for the prediction of CVD are Logistic Regression, K-Nearest Neighbour, Supportvector machine, Random Forest, and Decision Tree Classifier. Logistic regression works better on classification problems along with the classification results, and it produces probabilities that are accurately calibrated. KNN is simple and robust to noisy data. It is highly effective in cases of exceptionally large datasets. The SVM classifier is used as the UCI dataset's two classes are separated by a large margin. A decision tree classifier is employed as data is divided into chunks for better interpretability and work faster with categorical/numerical values thus reducing the computation power and minimizing the implementation time required to compute the final results. To enhancethe results Random Forest Classifier was applied to have more randomness to the model while growing trees.

Two algorithms are used for the prediction of Cardiac Arrhythmia: SVM with linear and radial basis function (RBF) and polynomial kernels, and Linear SVM [8]. For feature selection, Principal Component Analysis (PCA) is applied. This is followed by SVM [5] to predict the data in high dimensions which are later combined with a Random Forest Classifier to enhance the accuracy, generalization, and comprehensibility. Kernels are employed to enhance the SVM classifier output. Data preprocessing is done first by deleting columns that have more than 40% missing values. These missing values are later imputed by substituting the attribute median and then the attributes are normalized. For attributes having higher deviation, it is recommended to use the median instead of the mean of the missing values. For feature selection using PCA, according to Kaiser's rule components with eigenvalues, greater than or equal to 1 are selected.

After training the models, in order to test them, data is manually entered into the interface according to the parameters specified for each disease. The inputs are processed in the ML models connected to the two interfaces. And the final outputs are displayed through the GUI built for both diseases.

E. Building GUI

The interface for CVD is built using Flask which is a microframework written in Python, along with HTML to provide page structure, and CSS to provide page layout. A few of the other required modules are Jinja2 as it enables the clean addition of loops and variables into the Htmldocument. Other than that. Gunicorn and MarkupSafe have to be installed. Then, the machine learning model is used to obtain pickle (.pkl) files of the required algorithm, indicating that the model has finished being trained and is ready for testing. The command to start the server has to be executed for the web page to run on the web browser with all input fields. At last, the user provides the necessary details into the input fields, and the final results are calculated using the machine learning model by clicking the predict button, output is displayed for the user on the second page (Figs. 3 and 4).

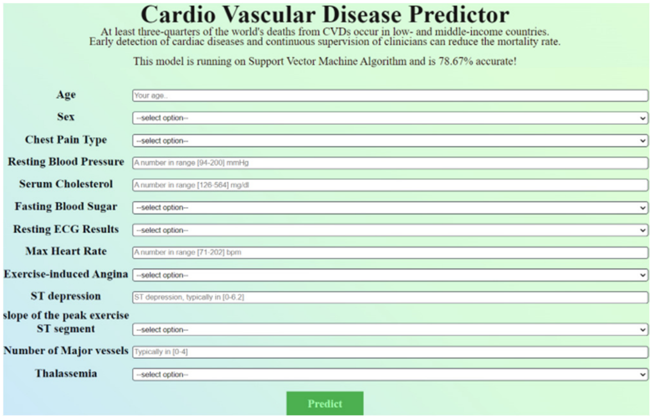

Fig. 3. GUI to detect cardiovascular disease.

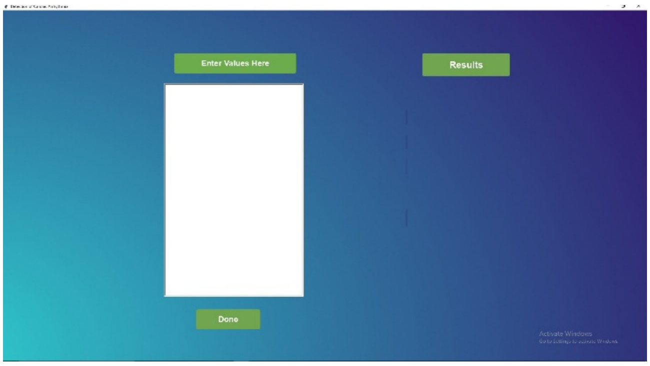

Fig. 4. GUI to detect cardiac arrhythmia.

The interface for the prediction of Cardiac Arrhythmia is built using a Python module Tkinter. GUI application is developed to be quick, simple, and dependable with Tkinter. Custom-made buttons are incorporated into the Tkinter module to facilitate the functioning of the GUI. On the website, the initial page has a title and a button to move to another page. The user enters data from the UCI dataset in a text box on the second page for the prediction of Cardiac Arrhythmia. The input data is tested with SVM, Decision Tree, and KNN which output the class of Arrhythmia. The final part is displaying the most probable class of arrhythmia from the obtained results.

IV. RESULT

The main operation is carried out with an ECG sensor to measure electrical activity and record the electricalsignals in the form of graphs. The main discussion of the above project is to Classify the two types of heart disease (CVD and Cardiac Arrhythmia). The secondary operation is to detect the risk of disease by creating an application interface. The output value is obtained in binary format in which 0 indicates no disease and 1 indicates the person has a chance ofhaving

CVD. Among multiple machine learning algorithms, the Support Vector Classifier shows the highest accuracy of 78.67%. For cardiac Arrhythmia, the output includes 16 classes based on the value of attributes. Initially, PCA is applied for feature extraction, followed by SVM with various kernel functions. The results of the random forest classifier show that the SVM with Linear Kernel and regularization parameter of 0.01 is 74% accurate amongst the other kernels used for predicting CardiacArrhythmia (Tables I and II).

TABLE I. RESULTS OF CVD PREDICTION

Sl.No.	Model	Accuracy
1	Random Forest Classifier	77.33%
2	Decision Tree	70.67%
3	Linear SVM	78.67%
4	Logistic Regression	76.0%
5	K-nearest Neighbors	74.67%

TABLE II. RESULTS OF CARDIAC AARRHYTHMIA PREDICTION

Sl.No.	Model	Accuracy
1	SVM with Random Forest(Linear, RBF, Polynomial Kernel)	73%, 66%, 68%
2	Linear SVM	74%

V. CONCLUSION AND FUTURE WORK

We implement machine learning algorithms to classify CVD and Cardiac Arrhythmia. Data acquisition for heart parameters is implemented using the hardware, and signal processing is used to detect peaks in ECG signals/ waveforms with inputs to the interface along with other important parameters. Using UCI datasets available for CVD and Cardiac Arrhythmia, we train and test the accuracy of the algorithms used in the machine-learning models. The SVM algorithm shows the best results for both CVD and Arrhythmia among other algorithms with an accuracy of 78.67 and 74% respectively. Most countries have seen an increase in death rates due to various heart diseases. Hence, early detection of disease reduces death rates. For the best accuracy in the proposed system, we incorporate recent machine learning algorithms. However, in addition to the approach in this research, it can be further enhanced by incorporating various deep learning techniques, augmenting the data set, using alternative methods for attribute selection, and using it to forecast several other diseases. Especially in rural areas, healthcare monitoring, and forecasting frameworks have proven indispensable in saving lives. By implementing the Internet of Things, we can improve patient care, administrative efficiency, and operational effectiveness.

ACKNOWLEDGMENT

We would like to express our gratitude to Dr. Purushotham U, Department of Electronics and Communication Engineering, PES University, for his continuous guidance, assistance, and encouragement throughout the development of this Capstone Project. We take this opportunity to thank Dr. Anuradha M, Chairperson, Department of Electronics and Communication Engineering, PES University, for all the knowledge and support we have received from the department. We are deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro-Chancellor, PES University, Dr Suryaprasad J, Vice-Chancellor, PES University for providing various opportunities and enlightenment to every step of the way. Finally, this project could not have been completed without the continual support and encouragement we have received from our family and friends.

REFERENCES

[1] K. Abdulnazar, A. Naz, R S Jeena and P. Niyas, “Deep Neural Network Based Real Time Multi-Class Arrhythmia Classification in IoT-Cloud Platform,” research square, v1, 2021, in press.

[2] PP. Kora, A. Rajani, M. C. Chinnaiah, K. Swaraja and K. Meenakshi, "IoT Based Wearable Monitoring structure for detecting Abnormal Heart," International Conference on Sustainable Energy and Future Electric Transportation (SEFET), 2021, pp, in press.

[3] S. Nashif, R. Raihan, M. R. Islam, M. H. Imam, “Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System,” World Journal of Engineering and Technology, 06. 854-873, 2018, in press.

[4] R. Kher, “Signal Processing Techniques for Removing Noise from ECG Signals,” J Biomed Eng 1: 1-9, 2019, in press.

[5] M. Gudadhe, K. Wankhade and S. Dongre, "Decision support system for heart disease based on support vector machine and Artificial Neural Network," International Conference on Computer and Communication Technology (ICCCT), 2010, pp. 741-745, in press.

[6] H. Yan, J. Zheng, Y. Jiang, C. Peng and Q. Li, "Development of a decision support system for heart disease diagnosis using multilayer perceptron," IEEE International Symposium on Circuits and Systems (ISCAS), 2003, pp. V-V, in press.

[7] G Sowmya, “IOT Based Health Monitoring System,” International Journal for Research in Applied Science and Engineering Technology, 2021, 9. 1176-1178, in press.

[8] S. K. Jain and B. Bhaumik, "An Energy Efficient ECG Signal Processor Detecting Cardiovascular Diseases on Smartphone," in IEEE Transactions on Biomedical Circuits and Systems, vol. 11, no. 2, pp. 314-323, 2017, in press.