Machine Learning Based Cardiac Disease Recognition Approach

Anukunj Khandagle [1], Rushikesh Shinde[2], Gaurav Hendre[3], Aniruddha Kulkarni[4], Prof. Sarika P. Aundhakar [5]

[1234](Students, Computer Engineering Department, PVPIT College of Engineering, Pune)

[5](Professor, Computer Engineering Department, PVPIT College of Engineering, Pune)

Abstract— *Cardiac Arrhythmia is a disease that affects many people during middle or old age and can cause fatal complications. In Cardiac Arrhythmia the heartbeat beats for may be fast or slow than the normal ones. Nowadays, a consistent electrocardiogram (ECG) analysis and classification play an important role in the diagnosis of cardiac abnormalities. ECG provides useful information about the functional status of the heart. Analysis of ECG is of great importance in the detection of cardiac anomalies as it helps to detect abnormalities related to sinus rhythms.*

The proposed system helps to detect the type of arrhythmia i.e. blockage from the different areas like atrial premature, premature ventricular contraction, premature ventricular contraction, premature ventricular contraction with the help of CNN. Firstly, by evaluating the ECG feature values the heart disease dataset is trained, and by applying the machine learning technique i.e. convolution neural network (CNN) the prediction of the disease can be done.

Keywords: ECG Signal Analysis, Data Mining, Machine Learning Techniques, Cardiac Arrhythmia, Classification, ECG features, CNN. Right Atrium, Right Ventricle, Left Ventricle, Left Atrium.

I. INTRODUCTION

Early detection of Cardiac Arrhythmia has the potential to prevent the millions of deaths that cause worldwide every year. The traditional diagnosis approach for heart disease detection is not reliable in many aspects. So Accurate and precise prediction of the heart attack is important to save the patient's life.

Heart disease Diagnosis is basically based on patients Electrocardiogram (ECG) test. The ECG tool plays a vital role in diagnosing and treatment of several diseases related to cardiac. Normally hearts beat for 60 bits/minute. If the heartbeats may be fast or slow than the normal ones or they may be fluctuating then there are chances of getting heart abnormalities. Early diagnosis of heart-related problems can potentially reduce the mortality rate and help patients maintain a better quality of life. An electrocardiogram (ECG)[1][2] measures the electric activity of the heart and has been widely used for detecting heart diseases due to its simplicity and non-invasive nature. By analysing the electrical signal of each heartbeat, it is possible to detect some of its abnormalities. ECG is the electrical movement of the heart it produces electrical signals which are called PQRSTU waves. The most vital wave is the QRS complex [2][3].

A. ECG Interpretation

ECG Interpretation includes the appearance of the wave and interval on the ECG curve. ECG morphology simply is the waveform or perspective of the electrical activity of the cardiac muscle, depolarization, and repolarization, in a cardiac cycle. The heart produces electrical impulses that spread through the cardiac muscle to make the heart contract. A normal ECG morphology consists of PQRST waves and each of the PQRST waveforms represents a single heartbeat or a cardiac cycle as shown in Figure 1.

The main ECG features for normal waves with time intervals are 1) QRS duration 2) R-R interval 3) P-R interval 4) Q-T interval 5) R-wave amplitude 6) P-wave Duration 7) T wave duration. The range for normal P-R interval is 65-85BPM). P-R interval (Indicates the proper functioning of a Sino-atrial node), Q-T interval (Indicates the rate, the velocity of blood flow from atrial to ventricular chambers), Isoelectric line (indicates the resting time in the heart takes seconds in a single heartbeat, R wave amplitude (indicates the rate of blood flow from atrial to ventricular chambers), P wave amplitude (Indicates the extent of atrial excitation), T wave amplitude (indicates the extent of ventricular relaxation). The Figure as shown below describes the essential ECG analysis used to predict heart diseases.

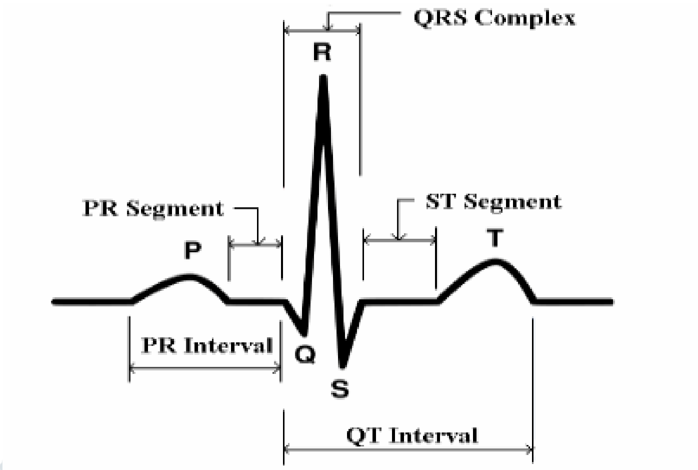


Figure 1: ECG Signal

II. LITERATURE SURVEY

A. Data Mining Based Approach

K.Gomathi1, Dr. Shanmugapriya et al.[8] present an analysis of the Heart disease for male patients using data mining techniques on data set consists of 210 records, which have all the available 8 fields from the database. The Naïve Bayes, Artificial neural network, and the J48 decision tree algorithms are used for prediction with final conclusion that Naïve Bayes predicts heart disease with higher Accuracy.

H. Benjamin Fredrick David et al.[9] used three data mining classification algorithms like Random Forest, Decision Tree and Naïve Bayes are addressed and used to develop a prediction system in order to analyse and predict the possibility of heart disease. , the UCI data repository is used for performing the comparative analysis of three algorithms. The paper shows that Random Forest algorithm performs best with 81% precision when compared to other algorithms for heart disease prediction.

T.Nagamani et al.[10] uses Mapreduce algorithm by comparing meta-heuristic approach along trained persistent fuzzy neural network on UCI machine learning repository dataset for predicting heart disease. The system gives 98.12 % accuracy for the 45 instances of testing set, when compared with meta-heuristic approach along with prepared neural network and conventional fuzzy artificial neural network.

KaanUyar et al. [11] proposed natural selection algorithm based trained neural networks for detecting the occurrence of heart diseases. The authors used totally 297 instances of patient data. Among these 252 were used for training and 45 of them were chosen for testing. The authors compared this RFNN approach with ANN-Fuzzy-AHP approach. By analysing the testing set 97.78% accuracy was attained as the outcome in the above said algorithm.

B. Machine Learning Based Approach

Alexander Schlemmer et al.[4] presents an machine learning algorithms for prediction of cardiac diseases. K Nearest Neighbors Classifiers (KNN), Support Vector Machines (SVM) and Random Forests (RF) are used for cardiological long-term study to improve classification results. Among which SVMs were found to yield the best results in Matthews Correlation Coefficient and are most stable with respect to a varying number of features.

Aditi Gavhane et al.[5] develop an application that can predict the vulnerability of a heart disease given basic symptoms like age, sex, pulse rate, etc. The machine learning algorithm neural networks are used to get the most accurate prediction results.

Jyoti Rohilla et al. [6] analyzed some of the data mining algorithms to predict heart disease. They have used a heart disease dataset from the UCI machine learning repository and analyzed using the WEKA tool, shown that decision tree algorithms performed well in predicting heart disease.

Aishwarya B. Chavan Patil et al. [7] use two data mining classification techniques like Artificial Neural Network(ANN) and Naive Bayes is used to assisting in the diagnosis of the heart disease to provide medication accordingly. The AVR-328 microcontroller is used as a gateway to communicate to the various sensors along with temperature sensor, heartbeat sensor, ECG sensor, the sensor for keeping a track of drip levels and a sensor to keep track of motion. The system is efficient with low power consumption capability, easy setup, high performance and time to time response.

III. PROPOSED SYSTEM

Often the doctors and medical staffs face problem in interpreting an ECG report. Very little change in any section of the ECG graph can result in different kind of diseases. During medical emergencies, like in ER or ICU, where time is of the essence, it would be more advantageous to find out what is ailing the patient for immediate treatment. Moreover, it is very difficult for a doctor to read an ECG report with bare eyes. At times, there is high chance to miss out any abnormality in the ECG report as the change in the ECG wave shape is hardly noticeable. So doctors often deny in concluding any disease from the ECG report of a patient with 100 percent accuracy until they conduct some more tests for the patient.\\
The proposed system can help out the doctors or practitioner to predict heart disease by evaluating the ECG feature values with convolution neural network (CNN). Figure shows the architecture of the proposed system.

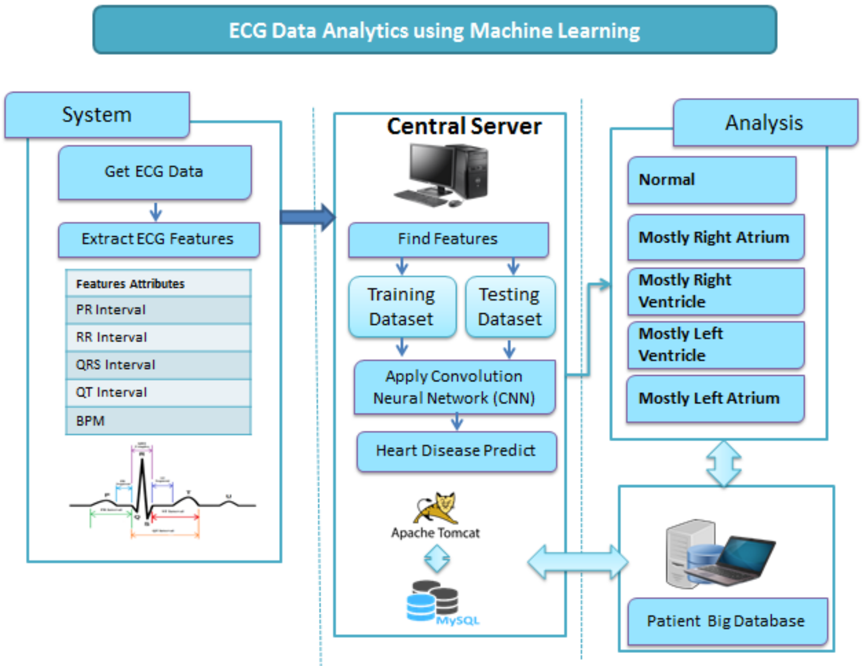


Figure 2: System Architecture

1. Load Patients Dataset: Get historical information/data of patient.

a. Preprocessing

The ECG signal getting from the sensor containing noise so preprocessing can be done to remove noise and get the final ECG value.

b. Detect Signal Parameters such P,Q ,R,S,T : Following Table shows the normal ECG Parameters.

Phase Duration Amplitude:

i) P Wave 0.06-0.11 <0.25 PR
ii) Interval 0.12-0.20 -
iii) PR Segment 0.08 -
iv) QRS Complex <0.12 0.8-1.2
v) ST Segment 0.12 -
vi) QT Interval 0.36-0.44 -
vii) T Wave 0.16 <0.5

ELSE

Abnormal heart

2. Heart Disease Prediction Using Machine Learning : System deals with existing arrhythmia data and performs analysis on that data. We are using Kaggle dataset for prediction of heart disease.

https://www.kaggle.com/shayanfazeli/heartbeat/data

This dataset is composed of two collections of heartbeat signals derived from two famous datasets in heartbeat classification, the MIT-BIH Arrhythmia Dataset and The PTB Diagnostic ECG Database. The number of samples in both collections is large enough for training a deep neural network.

This dataset has been used in exploring heartbeat classification using deep neural network architectures, and observing some of the capabilities of transfer learning on it. The signals correspond to electrocardiogram (ECG) shapes of heartbeats for the normal case and the cases affected by different arrhythmias and myocardial infarction. These signals are preprocessed and segmented, with each segment corresponding to a heartbeat.

Arrhythmia Dataset
Number of Samples: 109446
Number of Categories: 5
Sampling Frequency: 125Hz
Data Source: Physionet's MIT-BIH Arrhythmia Dataset
Classes: ['N': 0, 'S': 1, 'V': 2, 'F': 3, 'Q': 4].

3. Analyze Live Patient Data

Test samples are checked against the trained network and output label is predicted. Output label directly relates to type of arrhythmia the patient is having are given below. These are mainly depends on the values of the feature interval.

1. Atrial premature: Right Atrium
2. Premature ventricular contraction: Right Ventricle
3. Fusion of ventricular: Left Ventricle
4. Fusion of Paced and Normal: Left Atrium (Unclassifiable)

In Atrial premature causes as heart is occasionally beat irregularly. It may due to frightening or annoying, it's usually not dangerous unless you experience premature beats often or they impact the quality of your life.

Fusion of ventricles may occur early in the cycle R-on-T, after the T wave or late in the cycle - often fusing with the next QRS (fusion beat). R-on-T fusion may be especially dangerous in an acute ischemic situation, because the ventricles may be more vulnerable to ventricular tachycardia or fibrillation.

4. Compare Different algorithms

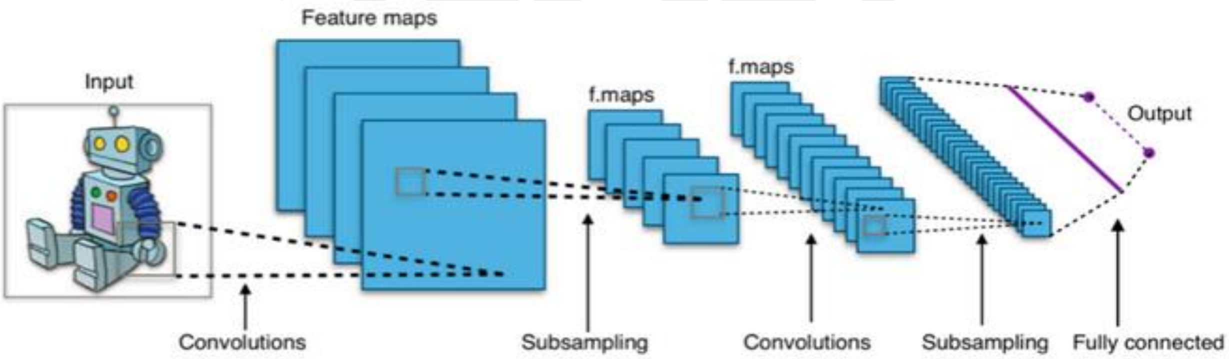
Different algorithms are applied on patient dataset to train the network. Network is trained using Neural Network. A clear confusion matrix needs to be plotted to show which algorithm provides good result for given dataset.

IV. ALGORITHM USED

Convolution Neural Network Traditional feature learning methods rely on semantic labels of images as supervision. They usually assume that the tags are evenly exclusive and thus do not pointing out towards the complication of labels. The learned features endow explicit semantic relations with words. We also develop a novel cross-modal feature that can both represent visual and textual contents. CNN is a method of categorizing the images as a part of deep learning. In which we apply a single neural network to the full image. The steps in CNN are as follows: convolution, subsampling, activation and full connectedness.

Step 1: Convolution it is the primary layers that accept an input signal are called convolution filters. Convolution is a procedure where the network tries to tag the input signal by referring to what it has learned in the past.

Step 2: Subsampling Inputs from the convolution layer can be smoothened to decrease the sensitivity of the filters to noise and variations. This smoothing procedure is labeled as sub- sampling, and can be attained by taking averages or considering the maximum over a sample of the signal.



Step 3: Activation the activation layer manages the signal flows from one layer to the subsequent Output signals which are strongly connected with past references would activate more neurons, enabling signals to be propagated more efficiently for identification.

Step 4: Fully connected the final layers in the network are fully connected, such that the neurons of preceding layers are connected to every neuron in subsequent layers. This imitates high Level reasoning where all feasible path ways from the input to output are measured.

V. EXPERIMENTAL RESULTS

In our proposed system we used a Kaggle dataset. Where accuracy and precision are calculated based on false positives items, i.e. which are items incorrectly labeled as belonging to the class and false negatives, which are items that were not labeled as belonging to the positive class but should have been.

The precision is the percentage of documents that are correctly classified as positive out of all the documents that are classified as positive. The confusion matrix is as shown below:

Table I: Confusion Matrix

Class	Precision	Recall	f1-Score
0	0.72	1	0.84
1	0.99	0.73	0.84
2	0.96	0.94	0.95
3	0.98	0.89	0.93
4	0.99	0.99	0.99

Here the results are classified in four classes like 0,1,2,3 and 4. Where,
0= Normal
1= Mostly Right Atrium
2= Mostly Right Atrium
3= Mostly Right Ventricle
4= Mostly Left Ventricle

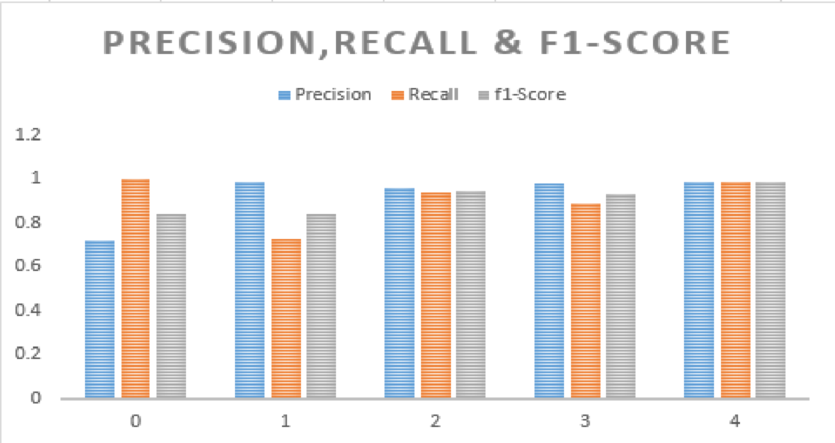


Figure 3: Graph for Precision & Recall

The overall accuracy obtained for the Classification is 98.76% using CNN.

A. CONCLUSION

Heart disease prediction is a popular research field in computer vision. The features on which heart disease is mostly dependent is intensely susceptible and variant. With the help of the proposed system, we can analyze the ECG signal data using the MIT-BIH Arrhythmia Dataset and The PTB Diagnostic ECG Database. The features like P Wave, PR interval, RR interval, QRS complex QT interval and T wave are taken into consideration for predicting the Arrhythmia. The dataset is trained and test samples are checked using CNN to predict the type of arrhythmia the patient is having. The system can predict four types of arrhythmia i.e. Atrial premature, Premature Ventricular Contraction, Fusion of ventricular and Fusion of Paced and Normal.

REFERENCES

[1] S. Shahbudin ; S. N. Shamsudin ; H. Mohamad ,"Discriminating ECG signals using Support Vector Machines", 2015 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE).

[2] Salma Banu N.K,Salma Banu N.K,"Prediction of Heart Disease at early stage using Data Mining and Big Data Analytics: A Survey",2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT).

[3] Aditi Gavhane ; Gouthami Kokkula ; Isha Pandya ; Prof. Kailas Devadkar ,"Prediction of Heart Disease Using Machine Learning", 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)

[4] Alexander Schlemmer ; Henning Zwirnmann ; Markus Zabel ; Ulrich Parlitz ; Stefan Luther,"Evaluation of machine learning methods for the long-term prediction of cardiac diseases",2014 8th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO).

[5] Aditi Gavhane ; Gouthami Kokkula ; Isha Pandya ; Prof. Kailas Devadkar ,"Prediction of Heart Disease Using Machine Learning", 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)

[6] Sarath Babu ; E M Vivek ; K P Famina ; K Fida ; P Aswathi ; M Shanid ; M Hena , "Heart disease diagnosis using data mining technique", 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA).

[7] Aieshwarya B. ChavanPatil, Prof. S. S. Sonawane, To Predict Heart Disease Risk and Medication Using Data Mining TechniquesWith an IoT Based Monitoring System For Post-Operative Heart Disease Patients Sixth Post Graduate Conference for Computer Engineering (cPGCON 2017) Procedi International Journal on Emerging Trends in Technology (IJETT).

[8] K.Gomathi1, Dr. Shanmugapriyaa,"Heart Disease Prediction Using Data Mining Classification",International Journal for Research in Applied Science & Engineering Technology (IJRASET).

[9] H. Benjamin Fredrick David and S. Antony Belcy, "HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES",ICTACT JOURNAL ON SOFT COMPUTING, OCTOBER 2018, VOLUME: 09, ISSUE: 01

[10] T.Nagamani, S.Logeswari, B.Gomathy,"Heart Disease Prediction using Data Mining with Mapreduce Algorithm",International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019 .

[11] KaanUyar ,Ahmetİlhan, “ Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks “, 9th International Conference on Theory and Application of Soft Computing, Procedia.