ECG Data Analysis using CNN and SVM

Mayur Jadhav[#1], Nikhil Nichit[#2], Swapnil Pujari[#3], Akash Kudale[#4]

[#1234](BE Student, Computer Engineering Department, D. Y. Patil College of Engineering, Akurdi, Savitribai Phule Pune University, Pune)

Abstract: ***Electrocardiogram (ECG) is a P, QRS and T wave signifying the electrical activity of the heart. Feature extraction and segmentation in ECG plays a major part in analysing most of the cardiac disease. According to a recent study by the Indian Council of Medical Research (ICMR) near about 25% of deaths between the ages of 25-69 years cause due to different heart-related problems.***

This paper provides a comprehensive review of the different techniques used by the researchers for diagnosing Arrhythmias and abstract view of the proposed system that we are going to implement with increasing accuracy using CNN. The system predicts the type of arrhythmia based on real-time ECG signals. The goal of this paper is to analyse, examine, evaluate, and make a summary of various approaches prediction of heart disease.

Keywords— Heart Disease, ECG Analysis, Feature Extraction, Signal Pre-processing, Live Patient Data

I. INTRODUCTION

Due to present unhealthy lifestyle diseases affecting heart increases. The healthcare industry has a huge amount of medical data, which is not mined. The medical data of the patients have hidden patterns that are necessary for data analysis in the detection of heart disease. Heart disease is a leading cause of death worldwide [16] from the past 15 years. As the clinical applications in remote health monitoring systems stored the patient's data for long-term recording, management and clinical access to patient's physiological information. As the body attributes of the human body like temperature, Heart Rate, Resting Electrocardiographic result, Blood Pressure [16][19] changes suddenly and abruptly in a cardiac patient, by providing the above parameters real-time, we can achieve the level of accuracy in prediction. Heart disease Diagnosis is basically based on patients Electrocardiogram (ECG) [9] [17] [18] test. The continuous recording of an electrocardiogram (ECG) by a wearable sensor provides a realistic view of a patient's heart condition by tracking such factors as high blood pressure, stress, anxiety, diabetes, and depression; during normal daily routines, we can predict the heart disease.

The ECG uses electrodes that have to be placed on the body of the patient from which the heart pulse is recorded [10] in the form of electrical activity. This signal depolarizes for every heartbeat reflected through the electrical activity on the skin. So by measuring the electrical pulse, we find any deep change in the ECG signal which is one of the main attributes in measuring heart disease. Diagnosing heart defects based on the relationship between ECG and clinical readings which can lead to high-performance heart diagnostics. Early detection of heart disease is essential because it can ease the treatment and also save people's lives.
We are developing a system for evaluating the real-time ECG values and the other parameter related to heart disease in a trained dataset and by applying the data mining technique i.e. Support vector machine prediction of the disease can be done.

II. LITERATURE SURVEY

The HDPS system predicts the probability of a patient getting a Heart disease. Miss. Chaitrali et al. [1] expanded the Heart Disease Prediction System (HDPS) system using a neural network. For prediction, the system uses sex, blood pressure, cholesterol-like 13 medical parameters. Here two more parameters are added i.e. obesity and smoking for better accuracy.

The use of the fuzzy measure and the relevant nonlinear integral gives a good performance in heart disease prediction. Sunita Soni, Jyoti Soni et al. [2] offers a frequent feature selection method for Heart Disease Prediction. None additively of the fuzzy measure imitates the importance of the feature characteristics as well as their interactions. Using features such as age, sex, blood pressure, and blood sugar can predict the likelihood of patients getting heart disease. And this improves accuracy and reduces the computational time.

D. Mendes, et al. [3] gives a decision tree model structure that uses a reduced set of six binary risk factors. The validation is performed using a recent dataset provided by the Portuguese Society of Cardiology of 11113 patients, which originally comprised 77 risk factors.
Several health monitoring systems use wearable sensors that produce continuous data and create many false alerts. Hence, these systems become inappropriate for use in clinical practice. To solve this problem some machine learning approaches are explained in [4] i.e. data generated by the wearable sensors are combined with clinical observations to provide early warning of serious physiological changes in the patients. Merging these data with manual observations the clinical staff makes important decisions about the patients.

Two data mining classification techniques like Artificial Neural Network (ANN) and Naive Bayes are used by Aishwarya B. Chavan Patil et al. in [5] 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 retain track of motion. The system is effective with low power consumption ability, easy setup, high performance and time to time response.

A system that can provide 24-hour continuous monitoring of the patient using a Discrete Wavelet Transform (DWT) for the ECG analysis and a Support Vector Machine (SVM) classifier is explained by Dimitra Azariadi et al. in [9]. It provides an algorithm for ECG analysis and classification for heartbeat diagnosis, and implement it on an IoT-based embedded platform.

The detection of cardiac disease using data mining classification techniques is explained in [11]. This application of classification technique gives the decision tree for the detection of heart disease. The classification tree uses the factors containing age, blood sugar, and blood pressure; it can detect the probability of patients fallen in CD by using fewer diagnostic tests which save time and money.

The fuzzy rule-based clinical decision support system (CDSS) for the diagnosis of heart disease, based on the knowledge from the patient's clinical data is explained in [12] by P.K. Anooj. The mining technique, attribute selection and attribute weighting methods are used to obtain the weighted fuzzy rules. Then, the fuzzy system is constructed by weighted fuzzy rules and chosen attributes. The experimentation is performed by the datasets obtained from the UCI repository and the performance of the system is compared with the neural network-based system utilizing accuracy, sensitivity and specificity.

Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients using a Cleveland Heart Disease dataset is given in [4]. The author examined if the accuracy could be enhanced by integrating voting with KNN. The results show that KNN achieved an accuracy of 97.4 %. The results also show that applying to vote could not enhance the KNN accuracy in the diagnosis of heart disease.

In the Year 2011, A.Q. Ansari et. al. [15] performed a work, "Automated Diagnosis of Coronary Heart Disease Using Neuro-Fuzzy Integrated System". Here, the author presented a Neuro-fuzzy integrated system for the analysis of heart diseases. To show the effectiveness of the projected system, Simulation for computerized diagnosis is performed utilizing the real causes of coronary heart disease. The author determined that this kind of system is right for the identification of patients with high/low cardiac risk.

Arrhythmia detection by using ECG data with the dataset (452 records) was retrieved from the UCI Machine Learning Repository. Information about Arrhythmia was extracted. Batra A. et al. [18] used machine learning algorithms and ECG diagnostic standards. The machine learning algorithms like neural networks, decision trees, random forest, gradient boosting and support vector machines were used for experimentations. However, more emphasis was laid upon SVM. With the combination of gradient boost and SVM, an accuracy of 83.04% was attained.

K. L. Clan et al. [22] have discovered the method of Hidden Markov Model (HMM) in classifying arrhythmia. They have developed a fast and reliable method of QRS detection algorithm based on a one-pole filter which is simple to implement and insensitive to low noise levels. The HMM method is not sufficient to represent one particular type of beat.

A. Ahmadian et al. [23] proposed a new piecewise modelling for approximation of ECG signal using Hermitian Basis. This method uses only the 5th order Hermitian basis functions. This method yields to weighing the approximation error of each segment base on its importance throughout the ECG complex. This method shows the total error obtained in this method is almost halved in comparison with similar non-segmented method.

The below figure shows the most common types of heart disease.

A. Types of Heart Disease [21][24]

Below are some type of heart attack and their exact condition in Arrhythmia:

Cardiac Arrest: The heartbeat is improper whether it may irregular, too slow or too fast.
High Blood Pressure: It has a condition that the force of the blood against the artery walls is too high.
Peripheral artery disease: The narrowed blood vessels which reduce the flow of blood in the limbs, is the circulatory condition.
Stroke: Interruption of blood supply occur damage to the brain.
Congestive heart failure: The heart does not pump blood as well as it should, it is the condition of chronic.
Congenital heart disease: The heart’s abnormality develops before birth.
Coronary artery disease: The heart does not pump blood as well as it should, it is the condition of chronic.

III. PROPOSED SYSTEM

Cardiovascular disease is one of the worldwide reasons for death. Normally the heart beats 60-100 beats/minute. However, heart rate higher than 76 beats per minute when resting may be linked to a higher risk of a heart attack. It is very difficult for a doctor to read an ECG report with bare eyes. At times, there is a high chance to miss out on any abnormality in the ECG report as the change in the ECG wave shape is hardly noticeable. Here we are developing a scheme that can analyse the ECG data of patients for predicting the type of arrhythmia. Figure 1 gives the architecture of the proposed system. Below is a detailed explanation of the proposed system.

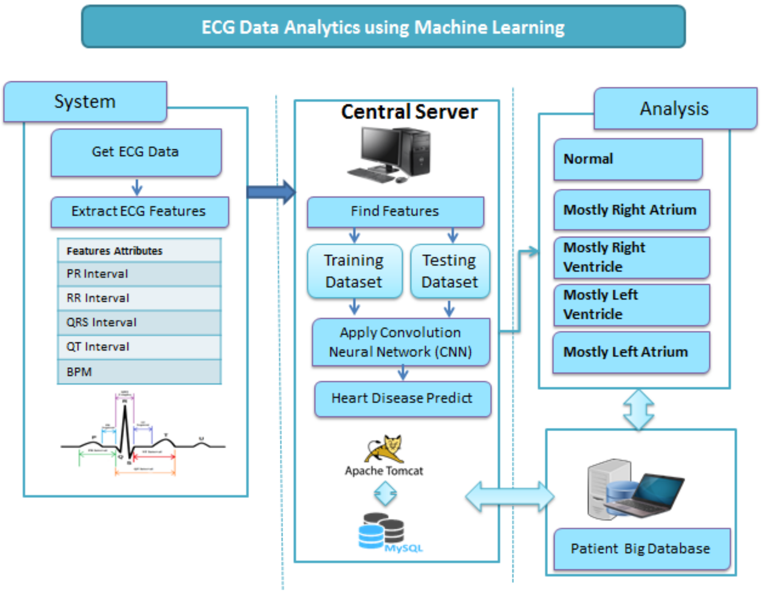

Figure 1: System Architecture

1. Load Patients Dataset: Get past data of the patient.

a. Pre-processing: The ECG signal getting from the sensor containing noise so pre-processing 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 pre-processed 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].

- Normal
  - Left/Right bundle branch block
  - Atrial escape
  - Nodal escape's
- S – Mostly Right Atrium
  - Atrial premature
  - Aberrant atrial premature
  - Nodal premature

- Supra-ventricular premature
- V – Mostly Right Ventricle
- Premature ventricular contraction
- Ventricular escape
- F – Mostly Left Ventricle
- Fusion of ventricular and normal
- Q – Mostly Left Atrium
- Paced
- Fusion of paced and normal
- Unclassifiable

3. Analyse 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.

a. **Atrial premature:** Right Atrium
b. **Premature ventricular contraction:** Right Ventricle
c. **Fusion of ventricular:** Left Ventricle
d. **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. CONCLUSION

Heart disease prediction is a standard research area in computer vision. The feature on which heart disease is mostly dependent is very susceptible and variant. So by receiving the historical information about the patient, we can forecast the heart disease
In this paper, we give a brief review of a different approach to the prediction of heart disease detection. A large collection of methods are identified for recognition of heart disease. But none of these can gives fine accuracy in the prediction. So there is a need to implement a system that can find heart disease with higher accuracy.

V. REFERENCES

[1] "A Data Mining approach for prediction of Heart Disease using Neural Networks", Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, International Journal of Computer Engineering and Technology, 2012.

[2]Sunita Soni, Jyoti Soni, Ujma Ansari, Dipesh Sharma, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer Application (IJCA, 0975 8887) Volume 17 No.8, March 2011.

[3] D. Mendes, S. Paredes, T. Rocha, P. Carvalho, J. Henriques, R. Cabiddu, J. Morais, "Assessment of Cardiovascular Risk based on a Data-driven Knowledge Discovery Approach", 2015 IEEE.

[4] Francis S. Collins, Mobile Technology and Healthcare, available at http://www.nlm.nih.gov/medlineplus/magazine/issues/winter11.

[5] Aieshwarya B. Chavan Patil, Prof. S. S. Sonawane, "To Predict Heart Disease Risk and Medication Using Data Mining Techniques With an IoT Based Monitoring System For Post Operative Heart Disease Patients", Sixth Post Graduate Conference for Computer Engineering (cPGCON 2017) Procedia International Journal on Emerging Trends in Technology (IJETT).

[6] Amir-Mohammad Rahmani, Nanda Kumar Thanigaivelan, Tuan Nguyen Gia, Jose Granados, Behailu Negash, "Smart e-Health Gateway: Bringing Intelligence to Internet-of-Things Based Ubiquitous Healthcare Systems", 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC).

[7] Fizar Ahmed, "An Internet of Things (IoT) Application for Predicting the Quantity of Future Heart Attack Patients", International Journal of Computer Applications (0975 8887) Volume 164 No 6, April 2017.

[8] Minas A. Karaolis, Member, IEEE, Joseph A. Moutiris, Demetra Hadjipanayi and Constantinos S. Pattichis, Senior Member, IEEE, "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees", IEEE Transactions On Information Technology In Biomedicine, Vol. 14, No. 3, May 2010.

[9] Dimitra Azariadi, Vasileios Tsoutsouras, Sotirios Xydis, Dimitrios Soudris, "ECG signal analysis and arrhythmia detection on IoT wearable medical devices", 2016 5th International Conference on Modern Circuits and Systems Technologies (MOCAST).

[10] 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).

[11] Abdul Aziz, Aziz Ur Rehman, "Detection of Cardiac Disease using Data Mining Classification Technique", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8 , No. 7, 2017

[12] P.K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules", Production and hosting by Elsevier.

[13] Deepali Chandna, "Diagnosis of Heart Disease Using Data Mining Algorithm", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1678-1680.

[14] Mai Shouman, Tim Turner and Rob Stocker, "Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients", International Journal of Information and Education Technology, Vol. 2, No. 3, June 2012.

[15] A.Q. Ansari et. al., "Automated Diagnosis of Coronary Heart Disease Using Neuro-Fuzzy Integrated System", 2011 World Congress on Information and Communication Technologies 978-1-4673-0125- 1@ 2011 IEEE (pp 1383-1388.

[16] Prachi W. Jambhulkar, Vaidehi Baporikar, "Wireless Sensor Network for Heart Disease Detection Using Data Mining Technique", IJIRCCE 2015.0306131

[17] Nourhan Bayasi, Member, IEEE, Temesghen Tekeste, Member, IEEE, Hani Saleh, Senior Member, IEEE, Baker Mohammad, Senior Member, IEEE, Ahsan Khandoker, Senior Member, IEEE, and Mohammed Ismail, Fellow, IEEE, "Low-Power ECG-Based Processor for Predicting Ventricular Arrhythmia", IEEE Transactions on Very Large Scale Integration (VLSI) SYSTEMS .

[18] Batra A. and Jawa V., "Classification of Arrhythmia using Conjunction of Machine Learning Algorithms and ECG Diagnostic Criteria", International Journal of Biology and Biomedicine, Vo1.1, 2016.

[19] 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).

[20] Jyoti Rohilla, Preeti Gulia, "Analysis of Data Mining Techniques for Diagnosing Heart Disease", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 7, ISSN:2277128X, 2015.

[21] V. Pavithra et al. "Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach", International Journal of Computer Applications (0975 –8887) Volume 181 –No. 18, September 2018.

[22] T. Cheng and K. L. Chan, "Classification of electrocardiogram using hidden Markov models", Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 20, No 1, 1998.

[23] A. Ahmadian, Karimifard S., Sadoughi H. and Abdoli M., "An Efficient Piecewise Modelling of ECG Signals Based on Hermitian Basis Functions", IEEE 2007.

[24] https://www.medicalnewstoday.com/articles/257484.php.