International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 1, Jan-Feb-2021, ISSN (Online): 2395-566X

Heart Disease Prediction using Data Mining Technique

Shivangi Agrawal, Asst.Prof. Ashish Tiwari(HOD)
Dept. of Computer Science & Engg.
Vindhya Institute of Technology & Science
ashishtiwari205@gmail.com, ashivangi15@gmail.com

***Abstract*- ECG is the most common and basic test performed on patients to detect anomalies in the heart. As a result of the ECG, 10 to 20 minutes of continuous patient heart data were collected and printed as a 1D plot. We have developed a program that extracts a continuous data set from the ECG machine and analyzes the data and extracts various ECG wave functions. First we divide the data by Wavelet decomposition. Then the data is reconstructed to 4 levels, removing noise from the signal. At the same time we detect important components of ECG wave, which are P wave, QRS complex and T wave. Electrocardiogram (ECG) is an important diagnostic tool for assessing cardiac arrhythmias in clinical practice. In this process, we present a deep learning-based convolutional neural network structure that was previously trained in a general data set of signals transmitted to perform automatic diagnosis of ECG arrhythmias by classifying ECG patients under similar cardiac conditions. The main goal of this process is to implement a simple, reliable and easy to use deep learning technique to classify two different selected conditions in the heart category. The results showed that cascading deep learning with conventional SVM was able to achieve very high performance. All this work is done by MATLAB simulation.**

Keywords- SVM, NN, RF,KNN,CNN,ECG

I. INTRODUCTION

The rhythmic pumping system of the heart needs electricity to contract which is regulated by a specialized conduction pathway [8]. The conduction pathway consists of five essentials components i.e. the sino-atrial (SA) node, the atrioventricular (AV) node, bound by His and its Purkinje fibers, shown in FIG. 1.4 . The cardiac action potential (AP) is generated due to the brief change in membrane potential across the cell membrane of the heart is shown in Fig. 1.3. Voltage is generated due to the movement of charged ions through ionic channels that connects the inside and outside of the cell. The action potentials also vary within the heart because of the presence of different ion channels in various cells. The resting membrane potential of ventricular cells is around -90 millivolts.

At rest state, the sodium (Na+) and chloride (Cl−) ions are found outside the cell, whereas the potassium (K+) ions found inside the cell [9]. The action potential starts with depolarization because of sodium channels opening that allow Na+ to flow into the cell. The depolarization begins after a brief delay, when K+ to leave the cell due to opening of potassium channels, creates a negative membrane potential. The calcium (Ca2+) ion found to be inside-outside of the cell to make sarcoplasmic reticulum (SR). bundle of Hans into the chambers. Its bundle is then divided into branches of the right and left bundles, which stimulates the right and left ventricles.

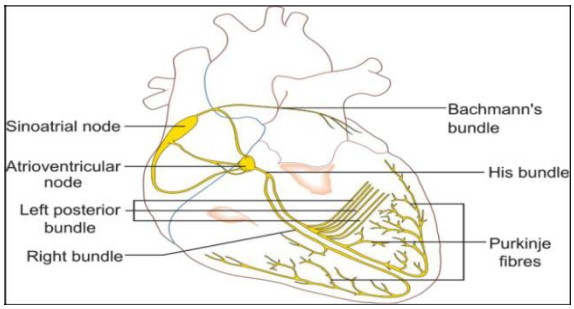

Fig. 1. Action potential of cardiac muscle ([Edited from [9]).

The peacemaking signal stimulates the right and left atrium to contract first, and then the right and left ventricles to allow the process of blood flow throughout the body The SA node also called natural pacemaker of the heart. It is a specialized tissue located in the atria and under normal condition of the heart generates an electrical stimulus of 60 to 100 times per minute at a regular interval of time. Each generated stimulus spreads rapidly through both atria in the form of a wave of contraction that passes through the myocardial cells. The electrical pulse travels from the SA node to the atrioventricular (AV) node, after which the pulses slow down for a very short time. The electrical stimulus moves through conductive pathways into the ventricles,which cause to contract and pumps out the blood. The two atrial chambers of the heart are stimulated first, then two ventricular chambers to contracts over a short period of time. The stimulus current travels in

conduction pathway via the The electrocardiogram (ECG) is a key diagnostic tool used to assess the health conditions of the heart. It records the electrical activity of heart during different phases of the cardiac cycle. The heart triggers tiny electrical impulses at SA node and spread through the conduction system of the heart to contract rhythmically.

These impulses can be recorded by the ECG machine by placing the surface electrodes over the skin of different part of the body. The tracings of the heart's electrical activity are called ECG waveform and the dips and spikes will show the conditions of the heart as shown in Fig.5. The ECG waveform is a series of positive and negative waves produced due to different deflection in each portion of heartbeat typical ECG tracing consists of P-wave, a QRS complex, and T-wave in each cardiac cycle. The ECG detects the transfer of ions through the myocardium, which changes in each heartbeat. The is electric line is the baseline voltage of ECG which is traced following the T-wave and preceding the next P-wave.

The upper chambers of heart make the first wave called P-wave. The P-wave is first to be generated due to contraction of the upper chamber of the heart followed by a flat line due to electrical impulse goes to the lower chambers. The contraction of ventricles makes the QRS complex and final T-wave produced for resting state of the ventricles [10]. The repetitive cycle of the electrical activity of heart is represented by the P-QRS-T sequences. The normal value of the different waveform of ECG is presented in the Table 1. [11].

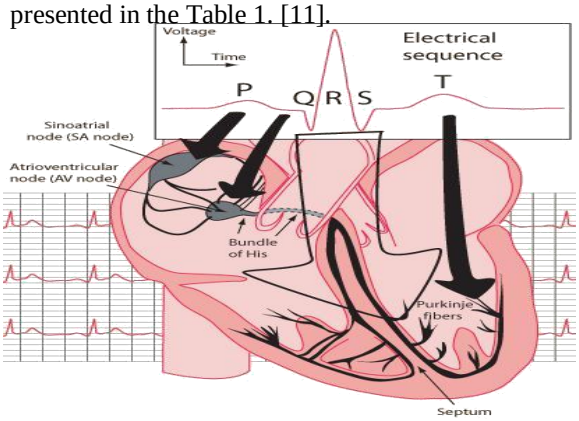

Fig.2 Generation of normal ECG signal ([Edited from[10]).

II. RELATED WORK

Several surveys have come up with the result that heart diseases are among the top five reasons of deaths worldwide. Conventional diagnosis of heart diseases relying on symptoms and standard tests like the ECG highly depends upon the experience of the physician and the fluctuations in the test data. Errors arising out of any of the above factors may have serious repercussions. In this, we present a technique in which consists of data pre-processing using the wavelet transform, classifying using the Euclidean Distance Classifier. Elimination of noise from ECG signals in pre-processing stage. Detection of precise R-peaks and QRS complex using different transform techniques such as wavelet, Hilbert and EMD in healthy and arrhythmia conditions. Extraction of both time and transform domain features (e.g. temporal features, heartbeat interval features and ECG morphology) from QRS complex and ECG waveforms. Selection and ranking of features to improve the classification accuracy in further tasks. Automated classification of arrhythmia beats using suitable machine learning techniques. Comparative performance analysis with published results in terms of sensitivity, specificity, positive predictive (Pp) and accuracy through arrhythmia beat classification. The main purpose of this work is to predict the ECG signal using an efficient classification method. To improve the accuracy of the classification and to reduce the miss class.

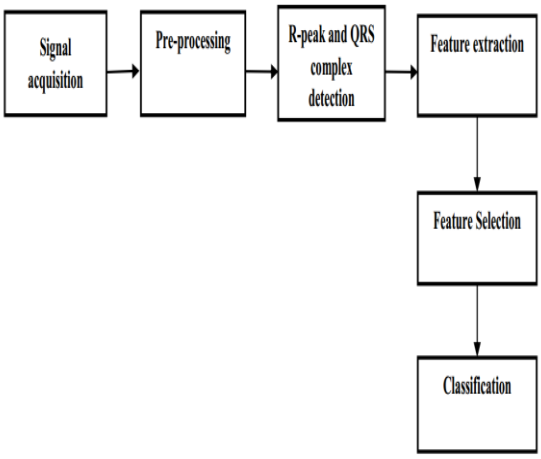

Fig. 3 proposed Block Diagram

III. PROPOSED APPROACH

In the process, we propose a deep arrhythmia diagnosis method, which is based on four classification models to automatically detect arrhythmia using ECG signals.

The classification model is mainly composed of four contiguous layers: two BLSTM layers and two fully connected layers. The data set of RR interval (referred to as set A) and heartbeat sequence (P-QRS-T wave, referred to as set B) is entered in the above model. Most importantly, our proposed method achieved 99.94% and 98.63% accuracy in training and validation sets in group A. respectively, in the test set (invisible dataset) we obtained an accuracy of 96.59%, a sensitivity of 99.93% and a specificity of 97.03%. Compared to other rating machines, the main advantage of multi SVM is that it reduces cross-validation and post-optimization features. Compared to other classifiers, SVM produces better classification results, mainly due to global optimization

functions included.SVM has shown a good concert in classification.

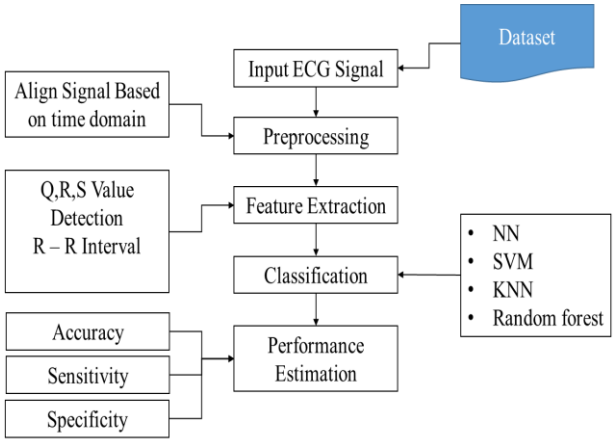

Fig. 4 Flow Diagram

Table 1 Different Attributes

S.No	Attribute
1	Age
2	Blood pressure (SI)
3	Urea
4	Sex
5	Blood pressure(DI)
6	Creatin
7	Bilirubin
8	PR (Per Rectum)
9	SGPT (Pyruvic)
10	Sugar
11	SGOT (Serum glutamic oxaloacetic transaminase
12	Sodium
13	HB(Hemoglobin
14	Family history
15	Stress level
16	Pottasium
17	Height
18	Smoker(yes/no)
19	Weight
20	Sedentary lifestyle

Preprocessing: Pre-processing: Pre-processing refers to "preparation" of samples / images to introduce them into the algorithm for a particular task: target tracking, recognition, feature extraction, etc. Data processing is a data mining technique that involves converting raw data into raw data. Understandable format. Real-world data is often incomplete, inconsistent, and / or lacking certain behaviors or trends and may contain many errors. Data processing is an effective way to solve such problems. We can convert data files (.xlsx) to .mat files and adjust the data. The purpose of optimization is to achieve a "best" design in relation to a set of priority criteria or constraints. Choosing the right features is very important for the pattern area, which can seriously affect the results. The quality and quantity of functions have a major influence on the goodness of the model. Various data processing techniques were used to improve the prediction model. Data integration is a combination of data from multiple sources that is continuously stored. With integration processing, there is always the problem of redundant data because attribute values from different sources have different names in the same real world, or an attribute can be an attribute derived from a different table. Therefore, researchers must use correlation analysis to carefully identify real-world entities from multiple data sources. Otherwise, careful integration of data from multiple sources can help reduce / prevent redundancy and discrepancies and improve mining speed and quality [9].

Data Extraction: Data extraction means analysis and review of data to retrieve relevant information from different data sources in a specific pattern. Here, the attributes for classification are identified

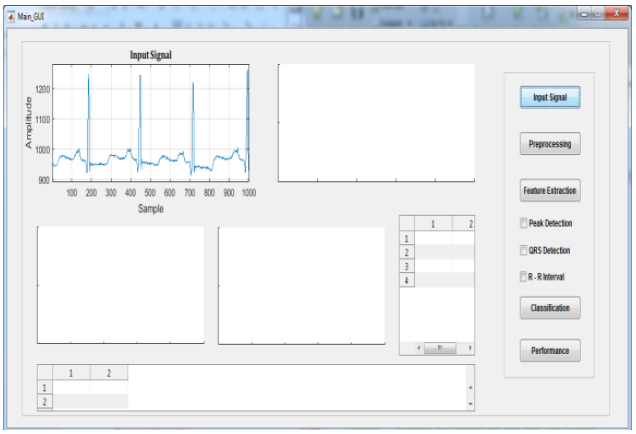

Fig. 5. Architecture Diagram.

Reconstruct Signal: After transforming the coefficients and drawing the coefficients at the same time, we observe that the frequency bands are separated and ca1, ca2, ca3 and ca4 are purer signals. However, due to down sampling, they will have fewer samples than the actual signal. We can see that the first signal is similar to the actual signal, but there is exactly one quarter of the sample because the signal is divided into 4 levels.

The number of samples in the second step is exactly half of the first step, and the number of samples in the third step is exactly half the second step. As the number of samples is reduced, this signal is also called a down-sampled signal. Obviously, second-level decomposition data is noise-free. Therefore, we consider this signal an ideal ECG signal from which to detect QRS. But the first R is in the third-level degradation signal for the 40th sample, and the first R is in the original signal in the 260th position. Therefore, when the R peak is detected in

International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 1, Jan-Feb-2021, ISSN (Online): 2395-566X

the third level reconstructed signal, it must be cross-validated in the current signal.

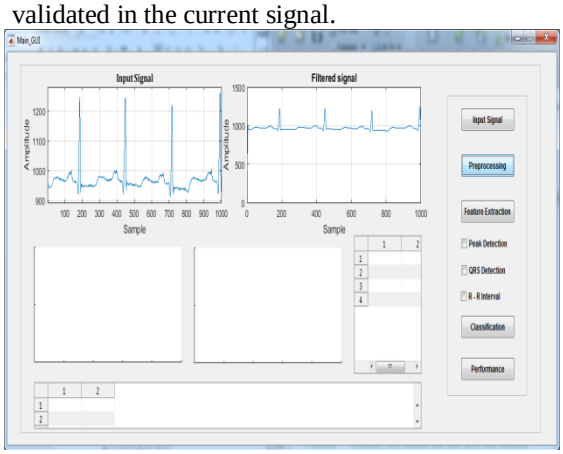

Fig. 6 Architecture Diagram.

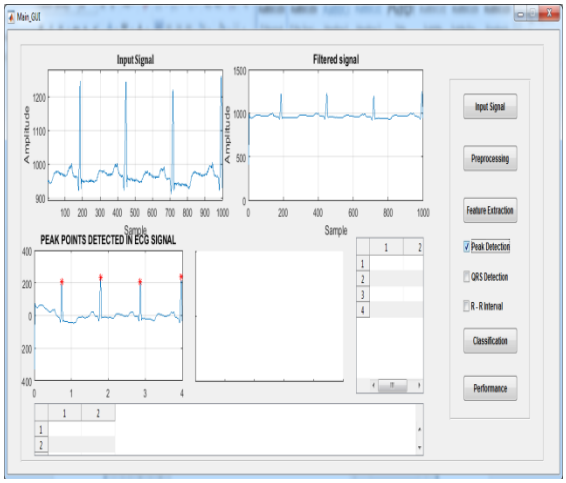

Fig. 7 Reconstructed Signal

```
[c,l]=wavedec(s,4,'db4');
ca1=appcoef(c,l,'db4',1);
ca2=appcoef(c,l,'db4',2);
ca3=appcoef(c,l,'db4',3);
ca4=appcoef(c,l,'db4',4);
```

QT Interval Measured from beginning of QRS to end of T Fluctuates in the frontal plane.
Normal: dependent on heart rate (corrected QT = QTc = measured QT, square root RR, in seconds; upper limit of QTc = 0.44 seconds).

Long-term QT syndrome: "LQTS" (based on the upper heart rate; QTc for men ≥ 0.47 seconds, QTc for women ≥ 0.48 seconds and hereditary LQTS can be diagnosed without other causes of increased QT) It has an important clinical meaning because it usually indicates an increased vulnerability to malignant ventricular arrhythmias, syncope and sudden death. The prototypical arrhythmia of the long QT interval syndrome (LQTS) is a polymorphic ventricular tachycardia, which is characterized by changes in QRS morphology and amplitude in electrical baselines. The reasons for LQTS include:

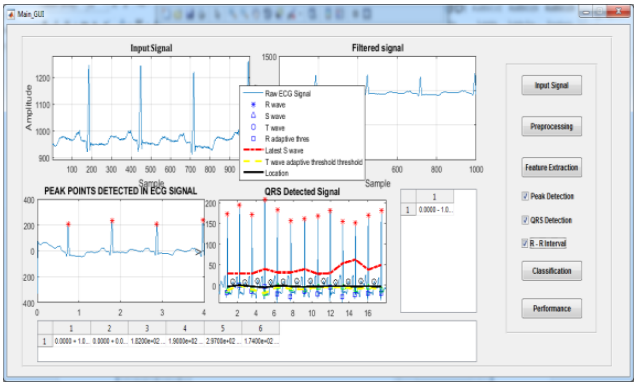

Fig. 8 QRS signal.

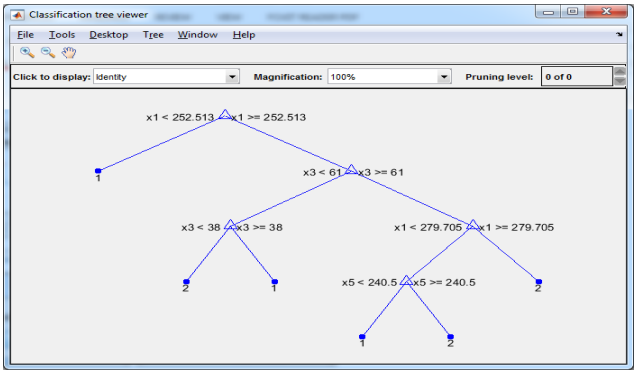

Fig. 9 RF Classifier tree

Table 2. Comparison Result

Classifier	Previous Work %	Proposed				
		Accuracy %	Sensitivity %	Specificity %	Precision %	Recall %
Svm	85.88	97.	99.9	95.8	95.0	99.8
Knn	83.2	97.8	95.8	99.2	99.65	95.5
Nn	92.21	93.0	88.9	99.1	99.09	88.0
Rf	85.88 %	89.7	99.6	90.7	89.11	99.9

IV. CONCLUSION

A new ECG heartbeat classification algorithm is proposed and signals are extracted from the database. Provide a brief ECG classification study. Electrocardiograms have been widely used to diagnose many heart conditions. Various techniques and conversion methods for ECG signals have been previously proposed in the literature.

This proposal provides an overview of noise cancellation, waveform detection and heart rate classification. This also shows a comparison table that evaluates the performance of various algorithms previously proposed for ECG signals. It also provides questions and guidance on existing work. We noticed that most of the noise elimination work was done using filter combinations. The QRS complex is most commonly used for heart rate classification. It can extract medical data with 20 attributes such as age, blood pressure and blood glucose to predict the potential of patients suffering from heart disease. These attributes are introduced into SVM, Random forest, KNN, and ANN classification algorithms, with SVM providing the best results with the highest accuracy. Using SVM algorithm can achieve effective performance in diagnosing heart disease and can further improve its performance by increasing the number of attributes.

REFERENCES

[1]. Sayali Ambekar Rashmi Phalnikar Disease Risk Prediction by Using Convolution Neural Network978-1-5386-5257-2/18/$31.00 ©2018 IEEE Pune, India
[2]. Shraddha Subhash Shirsath, Prof. Shubhangi Patil Disease Prediction Using Machine Learn.Over Big Data". I international Journal of Innovative Research in Science, Engineering and Technology, [2018]. ISSN (Online) : 2319-8753, ISSN (Print) : 2347-6710.
[3]. Vinitha S, Sweetlin S, Vinusha H, Sajini S. "Disease Prediction Using Machine Learning Over Big Data". Computer Science & Engineering: An International Journal (CSEIJ), Vol.8, No.1, [2018].DOI: 10.5121/cseij.2018.8101Sayali Ambekar and Dr.Rashmi Phalnikar. "Disease Prediction by using Machine Learning". International journal of computer engineering and applications, Volume XII, special issue, May 18. ISSN: 2321-3469.
[4]. Lohith S Y, Dr. Mohamed Rafi. "Prediction of Disease Using Learning over Big Data - Survey". International Journal on Future Revolution in Computer Science & Communication Engineering. ISSN: 2454-4248.
[5]. J. Senthil Kumar, S. Appavu. "The Personalized Disease Prediction Care from Harm using Big Data Analytics in Healthcare". Indian Journal of Science and Technology, vol 9(8), DOI: 10.17485/ijst/2016/v9i8/87846, [2016]. ISSN (Print): 0974-6846, ISSN (Online): 0974-5645
[6]. Gakwaya Nkundimana Joel, S. Manju Priya. "Improved Ant Colony on Feature Selection and Weighted Ensemble to Neural Network Based Multimodal Disease Risk Prediction (WENN-MDRP) Classifier for Disease Prediction Over Big Data". International Journal of Engineering & Technology, 7(3.27) (2018) 56-61.
[7]. Asadi Srinivasulu, S.Amrutha Valli, P.Hussainkhan, and P.Anitha. "A Survey on Disease Prediction in big data healthcare using extended convolutional neural network". National conference on Emerging Trends in information, management and Engineering Sciences, [2018].
[8]. Stephen J.Mooney and Vikas Pejaver. "Big data in public health: Terminology, Machine Learning, and Privacy", Annual Review of public Health [2018].
[9]. Smriti Mukesh Singh, Dr. Dinesh B. Hanchate."Improving Disease Prediction by Machine Learning". eISSN: 2395-0056, p-ISSN:2395-0072.
[10]. Joseph, Nisha, and B. Senthil Kumar. "Top-K Competitor Trust Mining and Customer Behavior Investigation Using Data Mining Technique."Journalof Network Communications and Emerging Technologies (JNCET) www. jncet. org 8.2 (2018).
[11]. Kumar, B. Senthil. "Adaptive Personalized Clinical Decision Support System Using Effective Data Mining Algorithms." Journal of Network Communications and Emerging Technologies JNCET) www. jncet. org 8.1 (2018).
[12]. Unnikrishnan, Asha, and B. Senthil Kumar. "Biosearch: A Domain Specific Energy Efficient Query Processing and Search Optimization in Healthcare Search Engine." Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org 8.1 (2017).
[13]. Kumar, B. Senthil. "Adaptive Personalized Clinical Decision Support System Using Effective Data Mining Algorithms." Journal of