Classification and Identification of Arrhythmia using Machine Learning Technique

Haresh M. Nanarkar[1], Prof. Pramila M. Chawan[2]

[1]*M.Tech Student, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India*
[2]*Associate Professor, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India*

---***---

Abstract - *Heart arrhythmia is a state of heart causing irregularity in heartbeats which can be too fast, too slow or unstable. Electrocardiography (ECG) is the fundamental tool use for the detection of Heart arrhythmia. Generally, electrodes are attached to the skin of a patient and the electrical activities of the heart of a patient are recorded for a period of time. As ECG signals reflect the heart's physiological conditions, medical doctors prefer to use ECG signals for diagnosing heart arrhythmia. One of the important skills of medical professionals is the ability to identify the dangerous types of heart arrhythmia from ECG signals. However, it is a tedious and time consuming process, because interpretation of the ECG waveforms is performed by professional medical doctor manually. As a result to improve this process of manual interpretation, the development of automatic techniques for identifying abnormal conditions from daily recorded ECG data is of fundamental importance. Moreover, if such abnormal heart conditions can be detected automatically, timely first-aid measures can be effectively applied. Such automatic detection of abnormal heartbeat can be done using health monitoring equipment which internally uses machine learning algorithms. Thus, in this regard, machine learning plays an important role.*

Key Words: Electrocardiography, Neural Network, Pan Tompkins algorithm, Levenberg-Marquardt algorithm

1. INTRODUCTION

Heart arrhythmia is a typical symptom of heart diseases. Few types of heart arrhythmia may even cause strokes and cardiac arrest, such as ventricular escape, atrial fibrillation and ventricular fibrillation. An electrical impulse generated in the sinoatrial node regulates the rhythm of the heart. Due to disorders in the normal sinus rhythm, an arrhythmia beats occurs. Different ECG patterns are caused by different arrhythmias. The arrhythmias like ventricular, atrial fibrillations and flutters are life-threatening and may lead to stroke or unexpected cardiac death. There are high risks of dangerous heart arrhythmic beats for a patient who had previously suffered from a heart attack. In both urban and rural areas across the world, heart disease remains the leading cause of death. Annually 380,000 people die due to Coronary heart disease.

Visual interpretation of ECG is complex task which results in consuming major amount of time for detecting arrhythmia from large dataset of heartbeats. This may further lead to incorrect classifications of heartbeats in appropriate type of arrhythmia category. Simple time-domain features specific techniques for identification of arrhythmia itself cannot provide good segregation among normal and abnormal classes. These difficulties can be solved by using an intelligent diagnosis system having appropriate machine learning technique.

ECG Database

In current work, MIT–BIH arrhythmia database sampled at 360 Hz is used and is publicly available on PhysioNet. The entire dataset categorizes heartbeats into five arrhythmia classes as recommended by ANSI/AAMI EC57:1998 standard. There are 48 records in the MIT-BIH database. Every record has duration of 30 minutes and is sampled at frequency of 360 Hz. These records are chosen from 24 hours recordings of 47 different individuals. Our work is focused on the classification of heartbeats in the MIT-BIH arrhythmia database into four arrhythmia classes: Normal rhythm (N), Right bundle branch block (RBBB), Left bundle branch block (LBBB), Premature ventricular contraction (PVC).
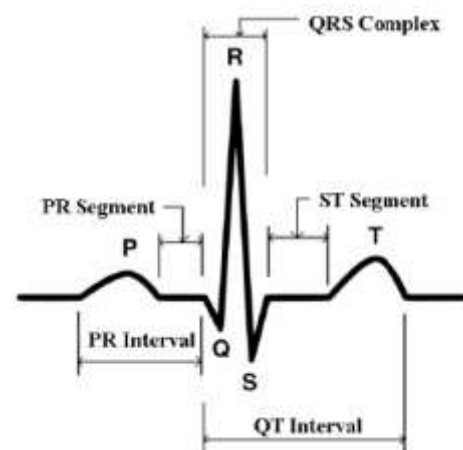

Fig -1: Components of ECG signal

In general, an ECG signal is composed of P wave, QRS complex followed by T wave as shown in Figure 1. Categorizing of heartbeats into different type of arrhythmia depends on investigating the shape, the relationship between these waves and the duration of each waves. However, analysis of the heart state for normal and abnormal ECG waves is a complicated task. In fact, the ECG signal is unsteady and thus, symptoms of a disease, if any, may not

occur continuously. Therefore, medical practitioners need to record and monitor the heartbeat for a long duration in order to classify the heart rhythm into normal or abnormal type. Thus, the size of the generated data can be huge for ECG signal analysis, which requires a lot of time and efforts, which arises a need for an automatic classification system.

2. LITERATURE REVIEW

In study conducted by [1], Support Vector Machine (SVM) and Multilayer Sensor (MLP) classifiers were used because the MLP and SVM classifiers gave the most successful results when working in this area. For classification and feature extraction operations, the calculation time is important. The performance of the classifiers to be used is compared to the time taken and other performance parameters. The contribution of this study was to apply different wave transformation techniques such as DCT, DWT, CWT on the ECG signals for improving the classification performance using these wave transformations.

In [2] this study, the major contribution was on the diagnosis of arrhythmia by introducing a new feature called amplitude difference to heartbeat classification based on two processes: 1) heartbeat detection and feature extraction; and 2) random forest classifier to classify heartbeats by their features. After extensive experiments investigating for adding a new feature in heartbeat classification using the MIT-BIH arrhythmia database showed that considering an amplitude difference feature can improve the performance of heartbeat classification by reducing false-positive and false negative rates.

The system proposed in [3], for the classification of arrhythmia, Deep Neural Network (DNN) is used as a classifier. It clearly showed that the patient is suffering from arrhythmia or sinus rhythm and the results obtained shows that the highest accuracy was achieved for the DNN. The specificity, accuracy, sensitivity and error rate of the classifier are delineated in the experimental result. 94% accuracy was achieved using the DNN classifier, making it more accurate than the previously existing system.

This [4] paper is concerned on the classification of ECG signal into different types of arrhythmia using a higher-order neural unit (HONU), i.e., a QNU, with error Back propagation in Batch optimization by Levenberg-Marquardt technique. The aim of the paper is to present a method that uses the classifier to aid the physicians in the recognition of ECG arrhythmias, and to evaluate the performance of the QNU for ECG arrhythmia classification.

The paper [5] has encouraged us to do research that consists of distinguishing between several arrhythmias by using deep neural network algorithms such as multi-layer perceptron (MLP) and convolution neural network (CNN).. The ECG database is accessible at PhysioBank.com and kaggle.com was used for training, testing, and validation of the MLP and CNN algorithms. The proposed algorithm in this paper consists of four hidden layers with weights, biases in MLP. It also uses a four-layer convolution neural network which has been used to map ECG samples to the different categories of arrhythmia.

In [6] this paper introduced a different technique for arrhythmia classification technique that can be used for efficient processing of large Electrocardiogram (ECG) records. By reducing size of the beats and quantizing the heartbeat numbers using Multi-Section Vector Quantization (MSVQ), this technique results in reduced time-complexity of arrhythmia classification without compromising on the accuracy of the classifier.

In [7], work is devoted to considering different neural networks in order to determine their accuracy in identification and separation of categories or classes of arrhythmia. Among all neural networks Feed Forward neural network using back propagation has been chosen for classification. A perfectly balanced input is used for training the network, taking same number of patterns from each class into consideration. This trained network is then used for classification of a completely different dataset.

This [8] paper proposes the system to predict eight cardiac arrhythmias using the radial basis function neural network (RBFN). In this study of neural network for heart rate time series, the prediction of Normal Sinus Rhythm (NSR) , Atrial flutter (AFL), Second degree block (BII), Left bundle branch block (LBBB), Sinus bradycardia (SBR), Atrial fibrillation (AFIB),), Right bundle branch block (RBBB) and Premature Ventricular Contraction (PVC) is done using proposed algorithm.
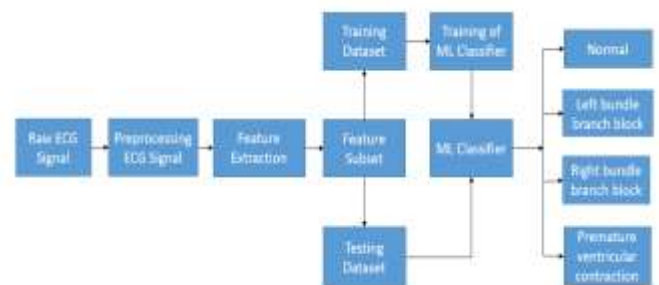
3. PROPOSED SYSTEM

Fig -2: Proposed system architecture

Figure 2 shows the system architecture of the proposed system. Input to the system is a raw ECG signals. This raw signal contains noise. In order to remove this noise from the ECG signal, preprocessing techniques is applied resulting in a filtered noise free ECG signal. After this features are extracted from the filtered ECG signal. The resulting feature dataset is then divided into training dataset and testing dataset. Training dataset is feed to the Machine Learning classifier. In the proposed system A Feed Forward Neural Network with Back Propagation classifier is used.

3.1 Preprocessing of raw ECG Signals

Preprocessing on raw ECG signal is important in order to remove noise from the signal. For this purpose, the proposed

system uses Pan-Tompkins preprocessing technique [13]. Firstly an integer coefficient band pass filter is composed using cascaded low-pass and high-pass filters. Its function is noise rejection. Next is a filter that approximates a derivative. After this, an amplitude squaring process takes place and the signal is passed through a moving-window integrator. Below figure 3 show the raw ECG signal and figure 4 shows Filtered ECG signal after preprocessing.
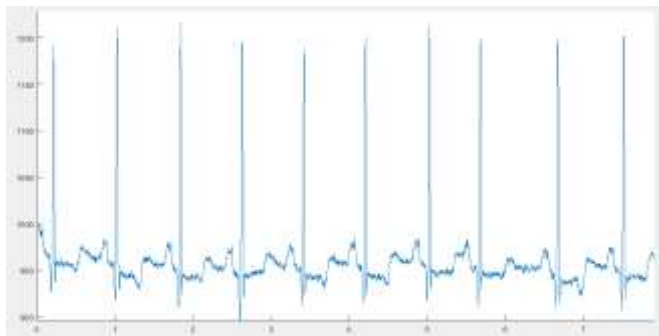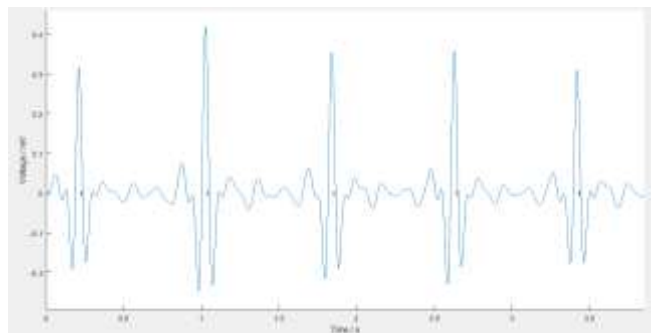

Fig -3: Raw ECG signal

Fig -4: Filtered ECG signal

3.2 Feature extraction and selection

For proper training of machine learning classifier, it is important to extract and select appropriate features from the dataset. From MIT BIH arrhythmia data, 11 features are extracted and selected for classifier. These are P-peak, P-wave duration, PR-segment, PR-interval, R-peak, QRS-wave duration, QT-interval, ST-interval, ST-segment, T-peak and T-wave duration. Details of each feature are given in the table 1.

Table -1: List of features selected from the ECG signal

Attributes	Description
P-peak	Peak of P-wave in mV
P-wave duration	Duration in sec.
PR-segment	Duration between end of P-wave and start of QRS complex in sec.
PR-interval	Duration between start of P-wave and start of QRS complex in sec.

R-peak	Peak of QRS wave in mV.
QRS-wave duration	Duration between start of Q-wave and end of S-wave in sec.
QT-interval	Duration between start of Q-wave and end of T-wave in sec.
ST-interval	Duration between end of S-wave and end of T-wave in sec.
ST-segment	Duration between end of S-wave and start of T-wave in sec.
T-peak	Peak of T-wave in mV.
T-wave duration	Duration in sec.

After filtering the ECG signal, R-peaks are located in signal. R-peaks are the point which having maximum amplitude in ECG signal. Once all R-peaks are detected, a window is created aroud R-peaks so that each window is indicating a beat i.e. one PQRST wave. In that window the highest point to the left of R-peak is P-peak and lowest point is Q-peak . Same hold on the right side, highest point on the right side is T-peak and lowest point is S-peak. This point is used for calculating the ECG features. Figure 5 show all the features detected and identified in the filtered ECG signal.
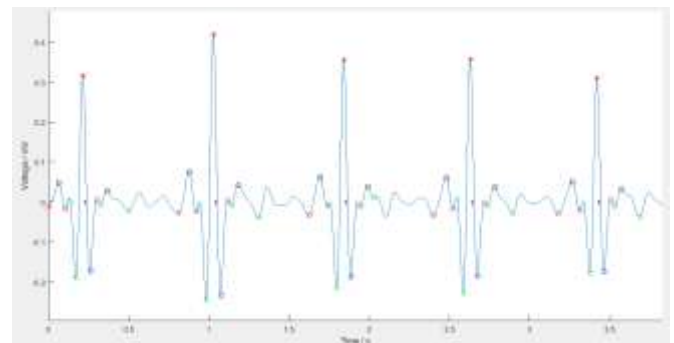

Fig -5: Plotting of the feature set in ECG signal

3.3 Neural Network Classifier training and testing:

In the proposed system, a feed forward neural network with back propagation is used as a classifier. A neural network has 11 input neurons representing the 11 features and 4 output neurons representing the 4 class to be predicted. There are 3 hidden layers in the neural network used in this proposed system. First hidden layer has 15 neurons, second hidden layer has 20 neurons and third hidden layer has 15 neurons. Figure 6 shows the structure of the neural network.
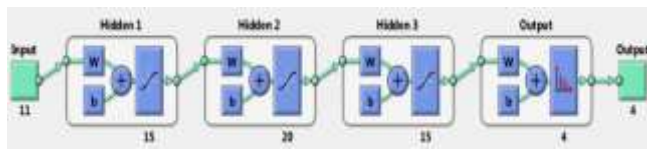
Fig -6: Structure of a neural network

Training algorithm used for this neural network is Levenberg-Marquardt algorithm [14]. Algorithm was designed to approach second-order training speed without computing the Hessian matrix. The Hessian matrix can be approximated as $H=J^TJ$ and the gradient can be computed as $g=J^Te$ when the performance function is in the form of a sum of squares as is typical in training feedforward networks. Here J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors. Also the Jacobian matrix can be computed using a standard backpropagation technique that is much less complex than computing the Hessian matrix.

Levenberg-Marquardt algorithm uses the approximation to the Hessian matrix in the following Newton-like update:

$$x_{k+1}=x_k-[J^TJ+\mu I]^{-1}J^Te$$

When the scalar μ is zero, this is Newton's method, using the approximate Hessian matrix. When μ is large, this is gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum; therefore, use of Newton's method would be benefical. Hence, μ is decreased after each successful step resulting in reduction in performance function and is increased only when a tentative step which would increase the performance function. Thus, the performance function is always reduced at each iteration of the algorithm.This algorithm is the fastest method for training moderate-sized feedforward neural networks up to several hundred weights. Results of neural network are given in table 2.

Table -2: Results of Neural network

Training algorithm	Levenberg-Marquardt algorithm
Epoch	97
Performance	0.0153
Gradient	0.117
Accuracy	96.1%

Fig -7: Confusion matrix

Confusion matrix for the trained neural network is shown in the figure 7. The rows in the matrix correspond to the predicted class i.e. Output Class and the columns correspond to the true class i.e. Target Class. The diagonal cells represent the observations that are correctly classified. The off-diagonal cells represent the incorrectly classified observations. Both the number of observations and the percentage of the total number of observations are shown in each cell.The column on the far right of the plot shows the percentages the precision or positive predictive value and false discovery rate, respectively. The row at the bottom of the plot shows the percentages of recall or true positive rate and false negative rate, respectively. The cell in the bottom right of the plot shows the overall accuracy which is 96.1%.

3. CONCLUSION

In the proposed system, a machine learning technique is used to identify the type of arrhythmia. In order to remove noise from raw ECG signal, prepossessing is done using Pan Tompkins algorithm. 11 features are extracted and selected from the filtered ECG signal dataset. A feed forward neural network with back propagation is used for training the classifier. The training algorithm used in neural network is a Levenberg-Marquardt algorithm for which the overall accuracy we got is 96.1%.

ACKNOWLEDGEMENT

REFERENCES

[1] Halil İbrahim BÜLBÜL and Neşe USTA, "CLASSIFICATION OF ECG ARRHYTHMIA WITH MACHINE LEARNING TECHNIQUES", 2017 16th IEEE International Conference on Machine Learning and Applications, DOI 10.1109/ICMLA.2017.0-104

[2] Juyoung Park, Seunghan Lee, and Kyungtae Kang, "Arrhythmia Detection using Amplitude Difference

Features Based on Random Forest", 978-1-4244-9270-1/15/$31.00 ©2015 IEEE

[3] Tanmay Paul, Arnab Chakraborty, and Subhrajit Kundu, "Hybrid Shallow and Deep Learned Feature Mixture Model for Arrhythmia Classification", 978-1-5386-5135-3/18/$31.00 2018 IEEE

[4] Ricardo Rodriguez, Jiri Bila, Osslan O. Vergara Villegas, Vianey G. Cruz Sánchez and Adriana Mexicano, "Arrhythmia disease classification using a higher-order neural unit", The Fourth International Conference on Future Generation Communication Technologies (FGCT 2015), 978-1-4799-8267-7/15/$31.00 ©2015 IEEE

[5] Shalin Savalia and Vahid Emamian, "Cardiac Arrhythmia Classification by Multi-Layer Perceptron and Convolution Neural Networks", Bioengineering 2018, 5, 35; doi:10.3390/bioengineering5020035

[6] Sandipan Chakroborty and Meru A. Patil, "Real-time Arrhythmia Classification for Large Databases", 978-1-4244-7929-0/14/$26.00 ©2014 IEEE

[7] Sanjit K.Dash and G.Sasibhushana Rao, "Robust Multiclass ECG Arrhythmia Detection Using Balanced Trained Neural Network", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016, 978-1-4673-9939-5/16/$31.00 ©2016 IEEE

[8] J. P. Kelwade and Dr. S. S. Salankar, "Radial basis function Neural Network for Prediction of Cardiac Arrhythmias based on Heart rate time series", 2016 IEEE First International Conference on Control, Measurement and Instrumentation (CMI), 978-1-4799-1769-3/16/$31.00 ©2016 IEEE

[9] Prof. Alka S. Barhatte, Dr. Rajesh Ghongade and Abhishek S. Thakare, "QRS Complex Detection and Arrhythmia Classification using SVM", 2015 International Conference on Communication, Control and Intelligent Systems (CCIS), 978-1-4673-7541-2/15/$31.00 it 2015 IEEE

[10] C. GURUDAS NAYAK, G. SESHIKALA, USHA DESAI , SAGAR G. NAYAK, "Identification of Arrhythmia Classes Using Machine-Learning Techniques", International Journal of Biology and Biomedicine, Volume 1, 2016, ISSN: 2367-9085

[11] Dr. Pritish Vardwaj, Abhinav Vishwa, Sharad Dixit and Mohit K. Lal, "Clasification Of ECG Arrhythmic Data Using Machine Learning", International Journal of Artificial Intelligence and Interactive Multimedia, Vol. 1, No 4., DOI: 10.9781/ijimai.2011.1411

[12] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: https://physionet.org/physiobank/database/mitdb/. Circulation 101(23): e215-e220 [2000 (June 13).

[13] JIAPU PAN AND WILLIS J. TOMPKINS," A Real-Time QRS Detection Algorithm", IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. BME-32, NO. 3, MARCH 1985, 00 18-9294/85/0300-0230$0 1.00 © 1985 IEEE

[14] Levenberg-Marquardt backpropagation - MATLAB trainlm

https://in.mathworks.com/help/deeplearning/ref/trainlm.html