A Survey on Classification and identification of Arrhythmia using Machine Learning techniques

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 heart state in which the heartbeat is irregular which can be too fast, too slow or unstable. Electrocardiography (ECG) is use for the detection of Heart arrhythmia. It records the electrical activities of the heart of a patient for a period using electrodes attached to the skin. Because ECG signals reflect the physiological conditions of the heart, medical doctors tend to use ECG signals to diagnose heart arrhythmia. Being able to identify the dangerous types of heart arrhythmia from ECG signals is an important skill of medical professionals. However, interpretation of the ECG waveforms performed by professional medical doctor manually is tedious and time-consuming. As a result, the development of automatic techniques for identifying abnormal conditions from daily recorded ECG data is of fundamental importance. Moreover, timely first-aid measures can be effectively applied if such abnormal heart conditions can be detected automatically using health monitoring equipment which internally uses machine learning algorithms. Thus, machine learning will play an important role in this regard.*

Key Words: *Electrocardiography*

1. INTRODUCTION

Heart arrhythmia is a common symptom of heart diseases. Some types of heart arrhythmia such as atrial fibrillation, ventricular escape and ventricular fibrillation may even cause strokes and cardiac arrest. The rhythm of a heartbeat is controlled by an electrical impulse generated in the sinoatrial node. An arrhythmia beats occurs when there is some disorders in the normal sinus rhythm. Different arrhythmias can cause different ECG patterns. The arrhythmias such as ventricular as well as atrial fibrillations and flutters are life-threatening and may lead to stroke or sudden cardiac death. There are more possibilities of arrhythmic beats for a patient who had previously suffered from a heart attack and also further the high risks of dangerous heart rhythms. Heart disease remains the leading cause of death across the world in both urban and rural areas. The most common type of heart disease is a Coronary heart disease which results in killing nearly 380,000 people annually.

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

ECG Database

In current study, publicly available PhysioNet, MIT–BIH arrhythmia database sampled at 360 Hz is used. Further, heartbeats from the entire dataset are categorized into five arrhythmia classes as recommended by ANSI/AAMI EC57:1998 standard. The MIT-BIH database contains 48 records. Each record has duration of 30 minutes with sampling frequency of360 Hz. These records are selected from 24 hours recordings of 47 different individuals. Our study is focused on the classification of four heartbeat classes in the MIT-BIH arrhythmia database: Normal rhythm (N), Left bundle branch block (LBBB), Right bundle branch block (RBBB), Premature ventricular contraction (PVC). Table 1 shows the distribution of these heartbeat types among the various ECG recordings present in the database.

Table -1: Distribution of heartbeats

Heartbeat	Type ECG Recording Containing Respective Type
N	100,101,105,112,115,000,000
LBBB	109,111 ,207,214
RBBB	124,212,231,232
PVC	105,109,116,119,214,000,000

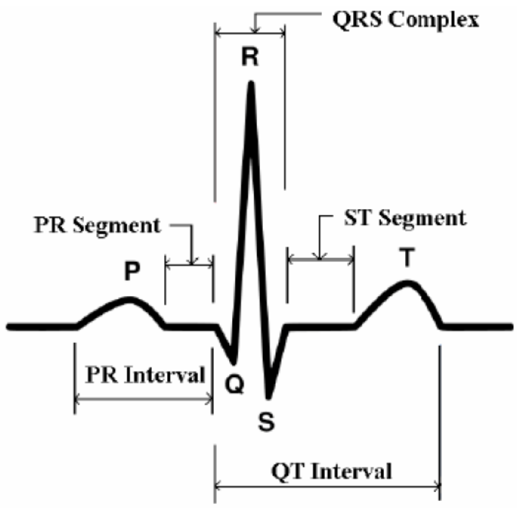

Figure 1. Components of ECG signal

The normal ECG signals are composed of P wave, QRS complex followed by T wave as shown in Figure 1. Diagnosing of the heartbeat is depends on investigating the shape, the relationship between these waves and the duration of each waves. However, analysis of the heart state or normal ECG waves is not an easy task. In fact, the ECG signal is non-stationary and thus, symptoms of a disease, if any, may not occur regularly. Therefore, physicians need to record and monitor the heartbeat for a long time to classify the rhythm into normal or abnormal type. For ECG signal analysis, the size of the generated data can be huge which requires a lot of time and effort, therefore need for an automatic classification system.

2.1. Machine learning

Being a field of science, machine learning deals with the ways in which machines learn using its past experience. From perspectives of many scientists, the term "machine learning" is interchangeably used along with the term "artificial intelligence", given that the possibility of learning is the main characteristic of an entity called intelligent agent. The main purpose of machine learning is the construction of computer program that can learn and adapt according using its past experience. A more detailed and formal definition of machine learning is given by Mitchel: A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, improves with experience E, as measured by P. With the advent of new techniques and approaches of Machine Learning, we currently possess an ability to find a solution to health related issues such as heart arrhythmia. We can develop a system using Machine learning techniques which has the ability to identify whether the patient has arrhythmia or not. Furthermore, detecting the arrhythmia in an early stage, leads to treating the patients before it becomes critical. Machine learning has the ability to extract hidden knowledge from a huge amount of heartbeat-related ECG data. Because of that, it has a significant role in heart arrhythmia research, now more than ever. The aim of this research is to develop a system which can predict the type of arrhythmia for a patient with a higher accuracy.

2.2. Supervised Learning

This survey is focused on study of a system based on classification methods which uses supervised learning .In supervised learning; the system must "learn" inductively using a function called target function. This target function is an expression of a model which describes the data. The objective function is used to predict the value of a variable, called dependent variable or output variable, from a set of variables, called independent variables or input variables or characteristics or features. The set of possible values as an input of the function, i.e. its domain, are called instances. Each case is described by a set of characteristics (attributes or features). A subset of all cases, for which the output variable value is labeled with the class it belongs, is called training data or examples. In order to infer the best target result, the learning system, given a training set, must take appropriate hypothesis for the function and denoted by h. In supervised learning, there are two kinds of learning tasks: classification and regression. Classification models predict distinct classes using its trained data, such as e.g. blood groups, while regression models predict numerical values. Some of the most common techniques are Support Vector Machines (SVM), Decision Trees (DT), Genetic Algorithms (GA), Artificial Neural Networks (ANN) and Instance Based Learning (IBL), such as k-Nearest Neighbors (k-NN).

2. LITERATURE REVIEW

In study conducted by [1], ECG signals 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. The calculation time is important for classification and feature extraction operations. The performance of the classifiers to be used is compared according to the time and other performance criteria. The contribution of this study was to apply some wave transformation techniques such as DWT, CWT, DCT to ECG signals in order to improve the classification performance by these wave transformations.

In [2] this study, the aim was to contribute to 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. Extensive experiments investigating the effects of adding a new feature in heartbeat classification using the MIT-BIH arrhythmia database show 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 Deep Neural Network (DNN) is implemented. It clearly visible that the patient is suffering from arrhythmia or sinus rhythm and the results obtained shows that the DNN has the highest accuracy. The specificity, accuracy, sensitivity and error rate of the classifier are delineated in the experimental result. An accuracy of 94% is acquired from the DNN classifier, making it more accurate than the existing system.

This [4] paper is concerned on the classification of electrocardiogram arrhythmia using a type of higher-order neural unit (HONU), i.e., a QNU, with error Backpropagation 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 databases accessible at PhysioBank.com and kaggle.com were 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 includes a four-layer convolution neural network which is used to map ECG samples to the different classes of arrhythmia.

In [6] this paper a different technique is introduced which is a coarse-to-fine arrhythmia classification technique that can be used for efficient processing of large Electrocardiogram (ECG) records. By reducing size of the beats as well as by quantizing the number of beats using Multi-Section Vector Quantization (MSVQ), this technique reduces time-complexity of arrhythmia classification without compromising on the accuracy of the classification.

In [7], a part of researches in this work is devoted to consideration of different neural networks in order to determine their accuracy in identification and separation of categories or classes. Among all neural networks Feed Forward 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. 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 Left bundle branch block (LBBB), Sinus bradycardia (SBR), Atrial fibrillation (AFIB),), Right bundle branch block (RBBB), Atrial flutter (AFL), Second degree block (BII) and Premature Ventricular Contraction (PVC) is done using proposed algorithm.

3. PROPOSED SYSTEM

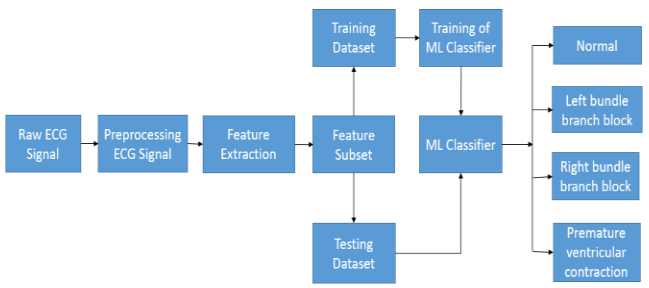

Figure 2. Proposed system architecture

The system architecture of the proposed system is as shown in Figure 2. Input to the system will be a raw ECG signals. This raw signal contains noise. Preprocessing of ECG signal removes this noise. Three different denoising techniques which can be used are median filter, moving average filter and notch filter. After this features are extracted from the filter ECG signal. Total 9 features are extracted for each beat using discrete wavelet transform, namely R point location, area under QRS complex, duration of QR, RS, RR points, R peak, R normal, area under autocorrelation and SVD of ECG. Different techniques i.e. FFT, CWT and DWT etc. will be used for the extraction of different features from the denoised ECG signal.

The resulting feature dataset will be then divided into training dataset and testing dataset. Training dataset will be feed to the different Machine Learning classifier. In the proposed system a SVM classifier and an ANN classifier will be used. Different hypothesis and weight will be taken into consideration in order to increase the accuracy of classification. At the end, the best combination of the preprocessing and classification techniques resulting on conducting this study will be used which can most accurately identify the type of heart arrhythmia.

For performance evaluation, we used three standard metrics: sensitivity, specificity, and accuracy [9]. These metrics are used to quantify the performance of the system. The sensitivity is measures of the capacity of test the positive samples.

Sensitivity (Sn) = (TP/ TP + FN) * 100

Where TP represents the true positive and FN represents the false negative. The specificity is measures of the capacity of test the negative samples.

Specificity (Sp) = (TN / TN + FP) * 100

Where TN represents the true negative and FP represents the false positive. The accuracy is defined as the ability of the test to correctly identify a classified type with and without positives. It reflects both sensitivity and specificity.

Accuracy (Ac) = (TP + TN) / (TP + TN + FP + FN) * 100

3. CONCLUSION

In the proposed survey we evaluate the various methods of classification techniques which can be used to identify the type of arrhythmia. The proposed system can classify the type of arrhythmia with high accuracy by using a proper combination of preprocessing and classification techniques. The system will be focused on using supervised learning technique i.e. SVM and ANN for classification. Sensitivity, Specificity and Accuracy will be used as a metric for performance evaluation.

ACKNOWLEDGEMENT

The author's would like to thank MIT-BIH for the dataset of heart arrhythmia ECG signal.

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).