Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020)
IEEE Xplore Part Number: CFP20P17-ART; ISBN: 978-1-7281-5821-1

Machine Learning based Cardiac Arrhythmia detection from ECG signal

Kavya Subramanian ,
Department of Electrical and Electronics Engineering
Amrita School of Engineering, Coimbatore,
Amrita Vishwa Vidyapeetham, India
kavyasubramanian20@gmail.com

N Krishna Prakash
Department of Electrical and Electronics Engineering
Amrita School of Engineering, Coimbatore,
Amrita Vishwa Vidyapeetham, India
n_krishnaprakash@cb.amrita.edu

Abstract—**This paper presents the analysis of heart diseases that are categorized as arrhythmia based on Electrocardiogram (ECG). ECG database of different disease conditions was analyzed. The ECG signals are filtered to remove noise which is caused due to powerline interface or Electromyogram. This filtered signal is segmented to smaller pieces of ECG so that feature extraction is accurate. The features extracted are Peak to peak Interval (R-R Interval), BPM (Beats per minute), P wave to QRS peak. The data set is classified using an SVM classifier algorithm. This algorithm classifies the input ECG signal with varying feature parameters to two different types of arrhythmia. This approach has achieved an accuracy of 91% and the performance regarding other criteria such as precision, recall and F1 score were significantly better indicating the success of the proposed method.**

Keywords— ***Electrocardiogram (ECG), Peak to peak Interval (R-R Interval), BPM (Beats per minute), P wave to QRS peak, Support Vector Machine (SVM)***.

I. INTRODUCTION

Cardiac arrhythmia is a collective term for a group of conditions such as bradycardia, tachycardia, irregular heartbeat, fibrillation, premature contraction. Cardiac arrhythmia can be considered as a leading cause of death, about 25,000 people per year. If detected at earlier stages, treatment can be done. A regular heartbeat is considered as 60-80 bpm, where bpm stands for beats per minute. Out of the types, defined bradycardia and tachycardia can be identified from Electrocardiogram [1]. Bradyarrhythmia is the medical term for commonly known bradycardia where patients have slower heart rates than a regular heartbeat. In some cases, even healthy people may have a heart rate less than 50 bpm. Tachycardia is a medical condition where the heart beats faster than 100 bpm. This usually happens because the electrical pulses that are sent to the heart misfires. The beats are so fast that the upper chamber of the heart doesn't get filled before its contracted. Some other heart irregularities include premature atrial contractions, Paroxysmal supraventricular tachycardia (PSVT), Premature Ventricular contractions (PVC), Atrial flutter, long QT syndrome, atrial fibrillation. Each type of arrhythmia has a different pattern, these patterns can be detected using automation [2]. The Electrocardiogram is widely used for diagnosing heart diseases. The electrocardiogram is a painless, non-invasive technique used to record electrical signals in the heart. It consists of 12 electrodes placed around the heart and limbs which records the electric flow. The electrical signal produced by the heart is very feeble and can go up to a few microvolts which are collected from the surface of the skin. This signal is a one-dimensional representation of potential against time series. The various features that are extracted from ECG including morphological features are used to detect arrhythmia [3]. The chances of getting cardiac arrhythmia arise due to the following risk factors the coronary arteries that bring blood to the heart muscles might get damaged if arrhythmia persists. Deposits in the ventricular portions of the heart make it difficult to regulate heart rate.

Studies suggest that classification based on a patient's symptoms as symptoms for arrhythmia can be confused with other diseases. The chances of clinical error can be reduced if machine learning algorithms are trained to recognize a pattern in the ECG. Machine learning uses algorithms to generalize scientific findings. Instead of just using it on IT systems can be expanded to biomedical sectors too where analysis and diagnosis can be made simpler. The features for patients with and without heart disabilities differ in minute values which can be differentiated by pre-programmed algorithms. If an arrhythmia is detected from features extracted, then the classification will be more accurate. The accuracy of such algorithms has proven to be close to 98%. This paper puts forward a method to classify ECG based on extracting features. After extracting features, an SVM classifier is utilized to categorize the features.

The organization of the paper is in this manner. Section II discusses the previous researches that have taken place on arrhythmia detection using ECG signals. Section III explores a method proposed for detecting arrhythmia using the ECG signal. Section IV provides with the results of the system. Section V concludes the paper.

II. RELATED WORK

The ECG signal detects abnormalities by recording variation in potential of the human heart. The ECG signal generally has noise and baseline drift which can be removed by using smoothing and baseline elimination [4]. Features such as RR interval, QRS interval, Beats per minutes are used to classify ECG into different types of arrhythmia [5]. There are various methods for the extraction [6] of these features such as Discrete Wavelet Transform (DWT) [7], Principal Component Analysis (PCA) [8][9]. With the help of DWT, a signal can be represented as a set of wavelets which can be scaled according to the requirement. DWT is used to compute the spectral energy of any signal. DWT, when performed on an ECG signal, can be used as a feature for classification. Using PCA a set of co-related variables can be converted to a set of uncorrelated variables. PCA is used to decrease the number of features that are taken into consideration based on their relevance. After extracting features, machine learning algorithms were implemented to classify these features into appropriate categories. Classifiers such as Convolutional Neural Network (CNN), Support Vector Machine (SVM), Genetic Algorithm (GA) [10], Artificial Neural Network (ANN), Recurring Neural Network (RNN) [11] were used to classify the database. While the genetic algorithm uses process like natural selection to classify features, Neural networks use training data to understand the characteristics of signals which can then be used in classification. Weights can be assigned to each feature to obtain a fast optimization algorithm [12]. SVM classifier classifies data based on morphological features such as R-R interval [13], QRS peak [14][15]. These features help to accurately determine whether a person has any heart arrhythmia-related issues.

In this paper, a simple method which uses features extracted from ECG and uses an SVM classifier is discussed. The method has higher accuracy (>0.91) in discrimination of other complex methods.

III. PROPOSED WORK

The system for detection of arrhythmia consists of the modules shown in Fig 1.

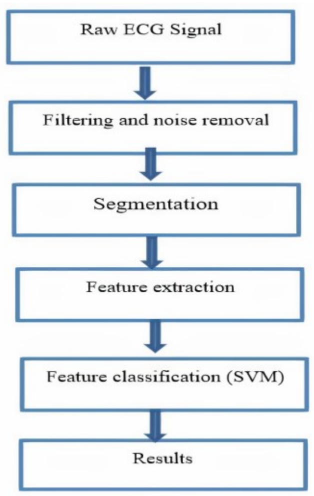

Fig 1. Flowchart for Arrhythmia detection from ECG

The data was collected from the MIT-BIH database which had a collection of arrhythmia-based samples. There are 3 components in the file namely .dat, ars and. xsws which are different formats in which the ECG data is available. These files can be read as an array and plotted with time to get ECG signals. ECG signals have power line interferences at 50Hz, to remove that a notch filter of 50Hz cut-off frequency is used. The bandstop filter has a small bandwidth through which the signal will be attenuated. The ECG signals were filtered using a Butterworth high pass filter to remove noise. The files were plotted in the time domain and the visual differences were noted. The signals are normalized within the range of -1 and 1. The filtered signals are split into four seconds. The R-R peaks are detected by calculating the moving average of the signal and then maxima at each interval specified in the sampling rate are determined. The features that are required for detecting arrhythmia are peak to peak intervals and beats per minute.

For detecting peak to peak interval (QRS-interval), Pan Tompkins algorithm is used. Pan Tompkins algorithm consists of a series which highlights the frequency content of rapid depolarization on the heart. Then the derivative filter is applied to detect the slope of QRS. The filtered signal is squared to enhance QRS peaks and reduce the chances of it being recognized as a T wave or P wave. For each ECG data given in the database after filtering and pre-processing these features are extracted. These features are used by SVM classifier to train itself to recognize whether the incoming data is Normal, Tachycardia or bradycardia.

A. Algorithm for arrhythmia classification

Once features are detected, SVM classifier is used for classification. SVM generally creates a hyperplane between the classes which can be visualized as data points in a plane. The data set consists of more than hundreds of data which have ECG with tachycardia, bradycardia and normal ECG data out of which 15 data sets 5 of each category are used as training set to train the SVM classifier.

TABLE 1. Training data provided to SVM classifier

ECG data	Heart rate bpm	Peak to peak Interval (s)	PR Interval (s)	Class
1	60	0.04938	0.18320	NL
2	68	0.04160	0.17557	NL
3	72	0.04870	0.18320	NL
4	100	0.06916	0.41984	TC
5	126	0.06138	0.18320	TC
6	145	0.05913	0.18320	TC
7	56	0.91603	0.17557	BD
8	42	0.068702	0.18320	BD

There are two types of comparison beat to beat and beat to bat. The beat to beat variation consists of variation in QRS axis and beat to bat variation depends on P axis. The features used for reference here are heart rate, PR interval and QRS interval which are given to an SVM classifier. Thus, each data point is classified as normal (NL) Tachycardia (TC) and bradycardia (BD). The dimensional spaces use kernel functions in decision making. Using kernel functions the input ECG data are classified into one of the categories. This kernel type is defined for non-linear classifiers, where simple SVM classifiers cannot be used.

The kernel SVM displays data that is non-linear with fewer dimensions as linearly segregable data with more dimensions. This way data points belonging to different classes are allocated to different dimensions. But while classification algorithms are being used, the data can either be classified into the correct class or incorrectly. In the case of medical problems, the impact of incorrectly classified data is higher.

The performance of SVM classifier can be quantified by measures such as accuracy and F1 score

B. Performance analysis of algorithm

For the performance analysis of the classifier, accuracy and F1 score can be calculated

$$\text{Accuracy}= \frac{(TP+TN)}{(P+N)} \quad \text{------(1)}$$

In the equation (1), True Positive (TP) which represents data correctly classified and True Negative (TN) which represents data incorrectly classified. P and N are the classes true or false to which classification of data points are done. Accuracy having higher value can be considered that the model has predicted larger members to a class correctly. But the members that are incorrectly classified also come into the picture that is why F1 score is calculated.

$$\text{F1 score}=\frac{2*precision*recall}{precision+recall} \quad \text{------(2)}$$

The harmonic mean of precision and recall gives a better measure for incorrectly classified data. It is based on False negatives and false positives. F1 score of a perfect model is 1 and that of a failed model is considered as 0. In the equation (2) the value of precision is the fraction of True positive values to the sum of True Positives and False Positive values whereas recall is the ratio of True Positive values and the sum of True Positives and False Negatives.

IV. RESULTS

Data from arrhythmia database were used to extract features such as R peak, QRS peak and heartbeat. These features were used to categorise the data values into different arrhythmia types.

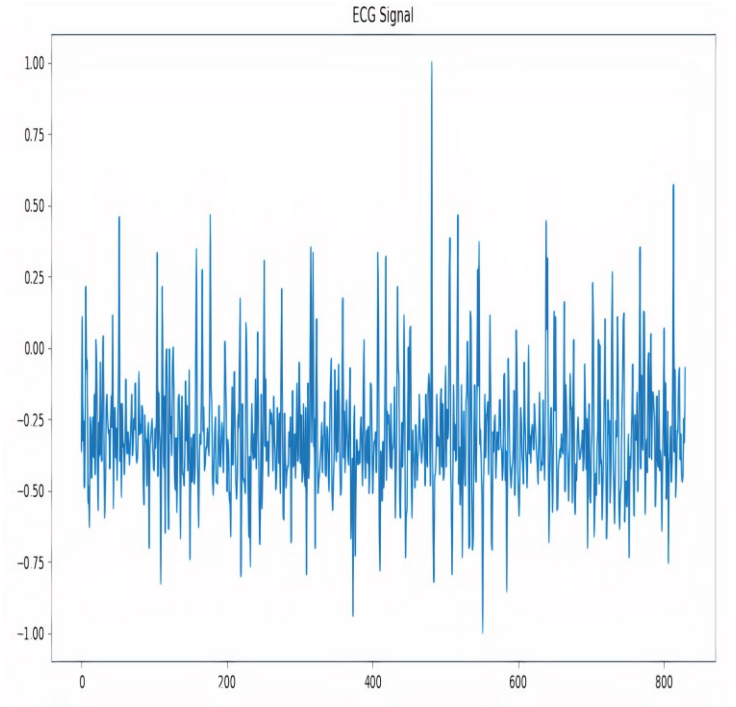

Fig 2. Representation of raw ECG signal in the time domain before processing

The input ECG signal is displayed in Fig.2 and the signal after removing noise and baseline wandering is shown in Fig.3.

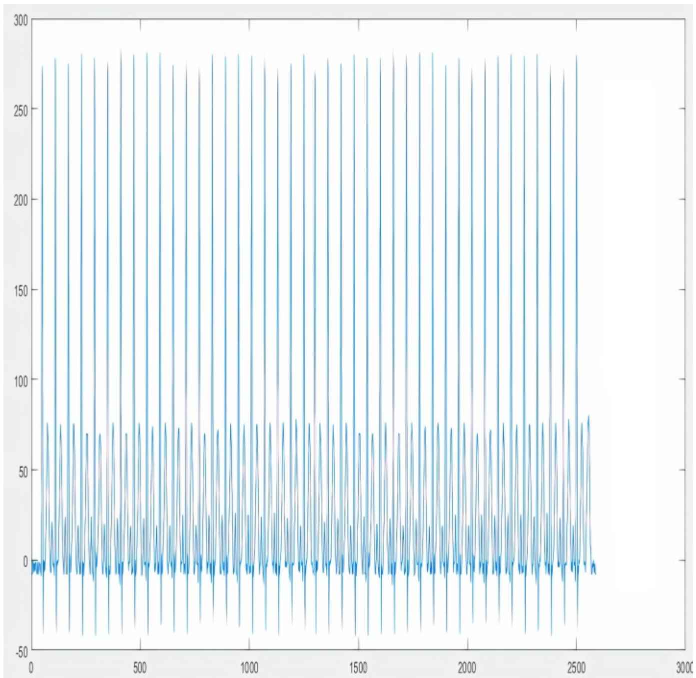

Fig 3. Representation of filtered tachycardiac ECG signal in the time domain.

The input signal shown in Fig.2 is sampled and then features are extracted. After training the SVM classifier with the training set shown in Table 1, another set of input data containing ECG from the MIT-BIH database are given to the SVM classifier to check its performance.

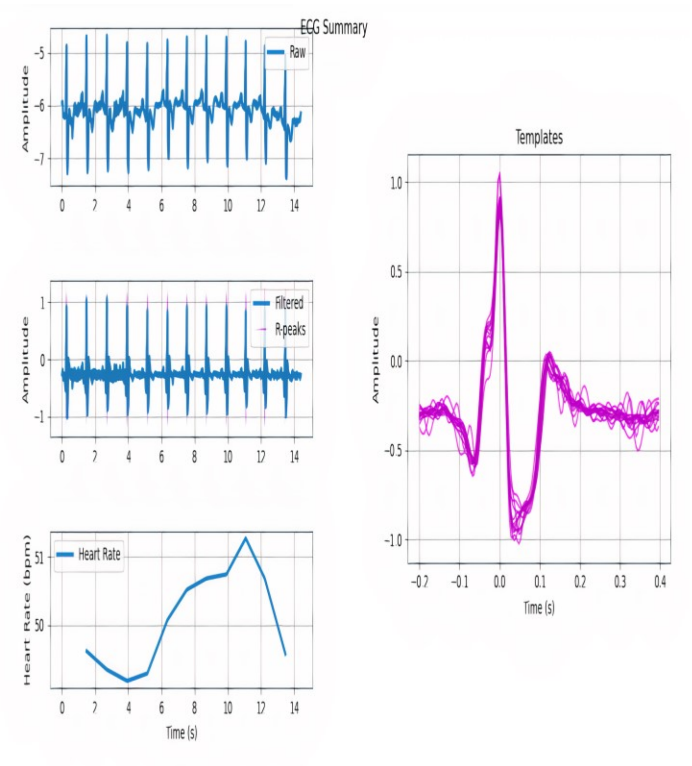

Fig 4. Time-domain representation of bradycardic ECG signal after filtering along with R-peaks detected.

TABLE 2. Comparison of Standard values and algorithm prediction

No.	Heart rate bpm	Peak to peak Interval (s)	PR Interval (s)	Actual	Predicted
1	62	0.0393	0.1192	NL	NL
2	66	0.0456	0.1237	NL	NL
3	71	0.0487	0.1232	NL	NL
4	108	0.0791	0.1594	TC	TC
5	120	0.0593	0.1601	TC	TC
6	150	0.0491	0.1683	TC	TC
7	55	0.0816	0.1755	BD	BD
8	74	0.0597	0.1783	BD	TC

Table 2 shows a comparison of actual and the predicted class of the input ECG data. The features extracted are represented in the table which was used to categorize ECG data as bradycardia or tachycardia. The prediction of such diseases on an early onset basis will make the cure easier. The use of pacemakers, regulation in diet or other solutions can be prescribed to people who opt for this mode of disease prediction. The medical field is emerging and the use of machine learning in places where more precision and accuracy are required is remarkable.

TABLE 3. Confusion matrix for the predicted and actual condition

No. of ECGs (23)	Predicted Bradycardia	Predicted Tachycardia
Actual Bradycardia	11	1
Actual Tachycardia	10	1

Out of 23 ECG that are considered from the database, 11 ECGs were correctly classified to bradycardia, 1 ECG which was supposed to be bradycardia was classified as tachycardia, 10 ECGs were correctly classified as tachycardia and 1 ECGs were misclassified as bradycardia. The data that was classified correctly and incorrectly was noted using a confusion matrix shown in Table 3.

Thus, accuracy is calculated as 0.91 or close to 91%. After calculating precision and recall F1 score is calculated as 0.906593 which is closer to one and can be considered as a good score. An algorithm for successfully classifying different ECG signals with the help of features extracted was tested. With pre-processing and SVM classifier it is easier to diagnose diseases like tachycardia or bradycardia which can save many lives.

V. CONCLUSION

This paper proposes an SVM based solution that classifies ECG data into types of arrhythmia. Python was used to implement SVM classifier and for the extraction of features. To improve the robustness of the system, training is performed using ECG data having varying feature parameters. This system can classify input ECG data into types of arrhythmia by training the SVM classifier with previously determined data. For this system to be successfully implemented, the training set must be precise and diverse as a slight variation in the peak to peak values can alter the condition of the heart. This paper explores simple extraction and classification techniques to classify ECG data without compromising accuracy.

REFERENCES

[1] Tanvi Sharma, Sahil Verma, Kavita, "Intelligent heart disease prediction system using Machine Learning: A review", International Journal of Recent Research Aspects, Vol. 4, Issue 2, June 2017.

[2] D.S. Medhekar, M.P. Bote, and S.D. Deshmukh, "Heart disease prediction system using Naive Bayes," International Journal of Enhanced Research in Science Technology and Engineering, vol. 2, no. 3, Elsevier 2013.

[3] B.Kohler, C. Hennig and R. Orglmeister, "The principles of software QRS detection," *in* IEEE Engineering in Medicine and Biology Magazine, vol. 21, no. 1, pp. 42-57, Jan.-Feb. 2002.

[4] Caroubalos, C. Perche, C. Metaxaki-Kossionides, Sangriotis, and D. Maroulis, "Method for automatic analysis of the ECG," Journal of Biomedical Engineering Volume 10, Issue 4, July 1988, Pages 343-347

[5] Siva A., Hari Sundar M., Siddharth S., Nithin M. and Rajesh C. B., "Classification of arrhythmia using Wavelet Transform and Neural Network model", Journal of Bioengineering & Biomedical Science, Volume 8, Issue 1, p.244 (2018).

[6] H. H. So and K. L. Chan, "Development of QRS detection method for real-time ambulatory cardiac monitor," Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering' (Cat. No.97CH36136), Chicago, IL, USA, 1997, pp. 289-292

[7] Jeppesen J, Beniczky S, Fuglsang Frederiksen A, Sidenius P, Johansen P., "Modified automatic R-peak detection algorithm for patients with epilepsy using a portable electrocardiogram recorder. "Conf Proc IEEE Eng Med Biol Soc. 2017.

[8] Hongquang Li, Danyang Yuan, Youxi wang, Dianyin Cui, Lu Cao, "Arrhythmia classification based on multi-Domain Feature extraction for an ECG recognition system", *Sensors* 2016, *16*, 1744.

[9] Portet F, Hernández AI, Carrault G,"Evaluation of real-time QRS detection algorithms in variable contexts. ", Med Biol Eng Comput. 2005;43(3):379-385.

[10] Kuo-Kun Tseng, Dachao Lee and Charles Chen, "ECG identification system using neural network with global and local features", International Association for Development of the Information Society,2016.

[11] Swapna G, Dr. Soman K. P.,Vinayakumar R, "Automated detection of cardiac arrhythmia using deep learning techniques", Procedia Computer Science, Volume 132, p.1192 - 1201 (2018).

[12] Sandeep Gutta, Qi Cheng, "Joint feature extraction and classifier design for ECG-based biometric recognition", IEEE J Biomed HealthInform. 2016;20(2):460-468.

[13] N. K. Prakash, S. M. R. Banu and S. M. H. Banu, "Denoising of ECG by statistical adaptative thresholding and detection of T-Wave Alternans using Principal Component Analysis," Proceedings of International Conference on Advances in Computing, Control, and Telecommunication Technologies, Kerala, pp. 778-780 (2009)

[14] Adnane M, Jiang Z, Choi S.," Development of QRS detection algorithm designed for wearable cardiorespiratory system." Computer Methods Programs Biomed. 2009;93(1):20-31.

[15] Paoletti M, Marchesi C. "Discovering dangerous patterns in long-term ambulatory ECG recordings using a fast QRS detection algorithm and explorative data analysis", Computer methods and programs in biomedicine 20-30, Volume-82. 2006.