Cardiac Arrythmia Detection Using Naïve Bayes And Svm Models

Rashmi Priyadarshini D K[1], Aman Sarathi B[2], Arun Kumar[3], Akshay Nayak[4], Akshay Kumar[5]

[1]School of Electronics and Communications Engineering, REVA University, India
[2]School of Electronics and Communications Engineering, REVA University, India
[3]School of Electronics and Communications Engineering, REVA University, India
[4]School of Electronics and Communications Engineering, REVA University, India
[5]School of Electronics and Communications Engineering, REVA University, India

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 23 May 2021

Abstract: Arrhythmia in any case called cardiovascular arrythmia, is a social event of conditions where the heartbeat is inconsistent, unnecessarily fast, or exorbitantly drowsy. Arrhythmia tends to a critical by and large broad ailment, addressing 15–20 % of all passing's. Early acknowledgment and investigation remains the best approach to perseverance, which can be refined by using novel procedures and shrewd development. As of now, arrhythmia ID is refined through ECG signal examination. In ECG signals, QRS structures which address the depolarization of ventricles are analyzed for arrhythmia revelation. The examination of time period of these PQRST waves essentially implies the presence of arrhythmia or commonness. In this assignment, a motorized structure to channel and bit ECG signals using different estimations is created using Python and MATLAB. For division computations, for instance, Two Moving Average are used. Unmistakable Machine Learning models, for instance, Naïve Bayes Model is used for the following examination of the divided ECG signals. Finally, the precision has using diverse division estimations and Machine Learning Models are researched and considered. Through this the most exact and beneficial system is settled. The eventual outcomes of this undertaking can appropriately help as a manual for clinicians for the distinguishing proof of arrhythmia.

Keywords: Hamilton detector segmentation, Navie-Bayes algorithm, Python, Support Vector Machine algorithm, Two moving averages segmentation.

1. Introduction

Electrocardiogram or all the more generally, ECG is the understanding of the electrical movement of the heart throughout some stretch of time, estimated utilizing anodes set on the skin. The examination of ECG signal gives data in regards to the state of the heart-either routineness or arrhythmia, if present. Arrhythmia depicts an unpredictable heartbeat - the heart may pulsate quick, moderate, early or sporadically. Consequently, discovery of arrhythmia infers cardiovascular turmoil in the subject. By and by, ECG signals are observed and recorded in gadgets, for example, holter screen and are therefore perused and dissected by specialists. Likelihood of misdiagnosis is higher for this situation as the undeveloped eye may neglect the slight abnormalities. Mechanization in this circle is essentialto beaten human blunders and reduction the general time taken for determination.

The inspiration of this venture is to help in the early discovery of heart-related problems by examining and arranging ECG signals through mechanization to accomplish more exact analysis, indeed, even without gifted experts. As of now, there is a need for research in the cycles of separating, division and distinguishing proof of suitable preparing calculations for the identification of ECG signals. When refined, this novel interaction with its higher precision will be liked absurd techniques for conclusion. Zerina Maseticet al.,(2016) portray utilization of autoregressive(AR) Burg strategy for extraction of highlights on the drawn out ECG time arrangement. Grouping worked on five heart-related disorders by analyzing and classifying ECG signals through automation in order to achieve more accurate diagnosis, even in the absence of skilled professionals. Currently, there is a need for research in the processes of filtering, segmentation and identification of appropriate training algorithms for the detection of ECG signals. Once accomplished, this novel process with its higher accuracy will be preferred over the current methods of diagnosis

Problem Statement

In the context of Cardiac Arrythmia Detection Using Naïve Bayes and SVM Models, we aim to prevent any heart related issues in the patient with the early detection .

2. Related work

Researchers across the world have conducted various studies in the field of ECG signal analysis and classification through different machine learning algorithms.

The work of S Celin et al.,(2018) focuses on comparative study of various learning algorithms for classification of ECG signal. The input signal is pre-processed using low pass, high pass and Butterworth filter to eliminate high frequency noise. After pre-processing, peak points are detected by using peak detection algorithm and features for the signal are extracted using statistical parameters. Extracted features are classified using SVM, Adaboost, ANN and Naïve Bayes classifier to classify the ECG signal database into normal or abnormal ECG signal. Results yielded an accuracy of 87.5%, 93%, 94 and 99.7% respectively.

Javier Aspuru et al., (2019) appraise the segmentation of ECG signals using a Linear Regression algorithm. A novel method for detecting fiducial points of ECG waves, using linear regression to identify maximums and minimums from an acquired ECG signal was proposed. Low pass Butterworth filter was operated for pre-processing the signal. The detection of R point was achieved using the proposed algorithm. The ECG signal was then segmented into periods, using the average distance among all R points as window size. Testing of the detection algorithm with signals acquired in real time by an ECG sensor was conducted. Results generated sensitivity of 97.5% for Q-point and 100% for the rest of ECG peaks.

The study by Renato Cuocolo et al., (2019) describes the current applications of machine learning in Cardiology. It elaborates on the different findings in this area of research such as that of Isin, which used machine learning for arrhythmia detection on ECG using an online dataset which showed a correct recognition rate of 98.5% and accuracy of 92%. This study includes a comparison of the different algorithms implemented and the corresponding accuracy, sensitivity obtained.

Zerina Masetic et al.,(2016) describe usage of autoregressive(AR) Burg method for extraction of features on the long term ECG time series. Classification operated on five different classifiers namely,C4.5 decision tree, k-nearest neighbour, support vector machine, artificial neural networks and random forest classifier. Results produced outperform standard metrics.The efficiency of path planning reduces as the amount of UAVs increases in the air space with leads to traffic congestion and delay in its operations while on flight.

3. Proposed system

As shown in Fig 3.1, the signals need to be first segregated into training database for training the model and testing database for testing the model. These signals then need pre-processing and denoising by passing through filters. The QRS complexes which are then obtained form the segments. The time periods for these segments need to be calculated, from which the segments will be classified as Normal or Arrhythmic. The machine learning algorithms will then be trained using the training data. Once trained, the model is used to predict the output of the test data. The obtained output will be compared with the expected annotations from the annotation files and the accuracy of the model will be calculated

3.1 System Architecture

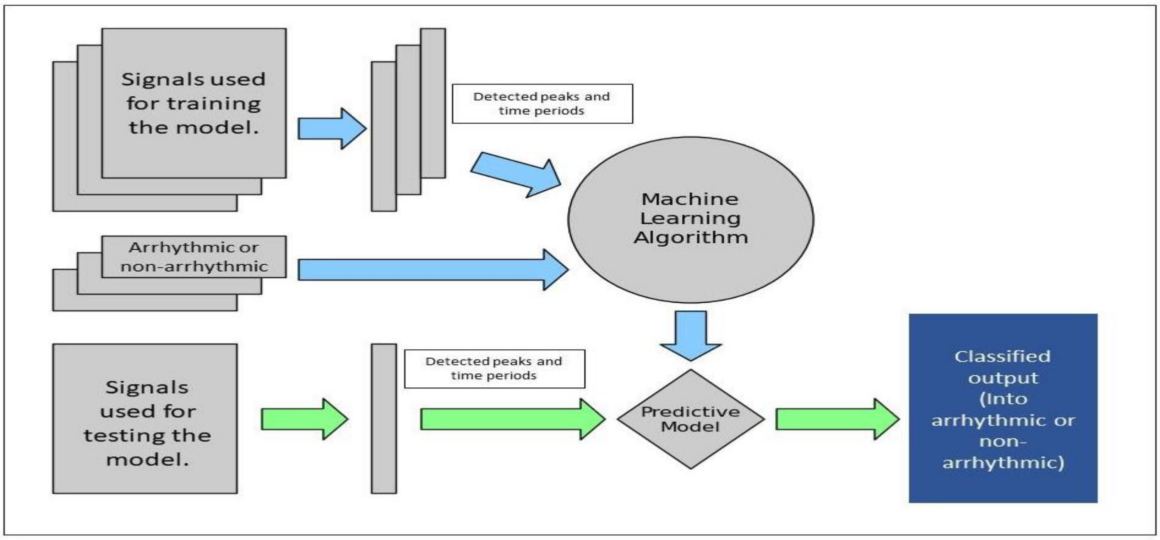


Fig:3.1 Architecture of proposed model

As shown in Fig 3.1, the signals need to be first segregated into training database for training the model and testing database for testing the model. These signals then need pre-processing and denoising by passing through filters. The QRS complexes which are then obtained form the segments. The time periods for these segments need to be calculated, from which the segments will be classified as Normal or Arrhythmic. The machine learning algorithms will then be trained using the training data. Once trained, the model is used to predict the output of the test data. The obtained output will be compared with the expected annotations from the annotation files and the accuracy of the model will be calculated

3.2. Implementation strategy:

Steps 1-7 are followed as an implementation strategy for arrhythmia detection

1. Obtain the signals from MIT-BIH Arrhythmia database.
2. Pre-process the signals to remove existing noise.
3. Detect QRS complexes using different detection algorithms and segment the signals R-R.
4. Calculate the time periods of the segments and classify it as Normal and Arrhythmic.
5. Calculate the accuracies of the detection algorithms by comparing it with the QRS complexes mentioned in the annotation files of the signals.
6. From the segmentation data, train and test the different machine learning models.

Calculate the accuracies of the machine learning models by comparing it with the data from the annotation files of the signals.

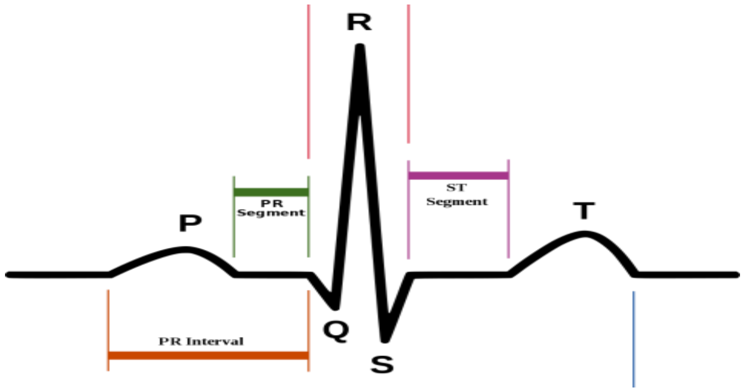


Fig:3.2 Diagram of PQRST wave

Above fig :3.2 consists of a **PQRST** complex. The sinoatrial node (SA) is the pacemaker of the heart and produces the P **wave**. The P **wave** in an **ECG** complex indicates atrial depolarization. The QRS is responsible for ventricular depolarization and the T **wave** is ventricular repolarization.

3.3. Obtaining the Database:

1. The MIT-BIH Arrhythmia database consists of 48 half-hour excerpts of two-channel ambulatory ECG recordings corresponding to 47 patients.
2. Each file is a uniquely named .csv file and it consists of 650,000 samples per channel, which at a sampling rate of 360 samples per second gives 30:055 minutes of two-channel ECG readings.
3. Each of these files are complemented with two more files

An annotations file in .txt format containing information about time of R-peaks, samples denoting R-peaks, and the type of arrhythmia.

An information file stored as .info file containing information about source, the sampling frequency, the sampling interval, duration of the signals, and name of the channels in the file.

3.4. Pre-processing/Decomposition of the signals:

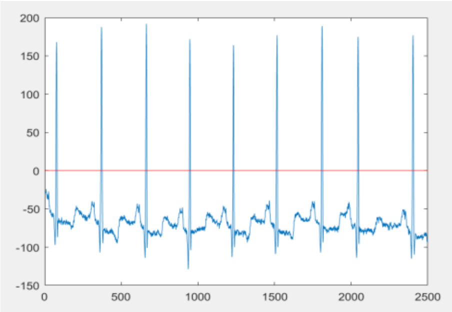
For processing the signals and removing the noises involved, MATLAB version R2018a is used in reference to the paper "Signal Processing Techniques for Removing Noise from ECG Signals"[2019].

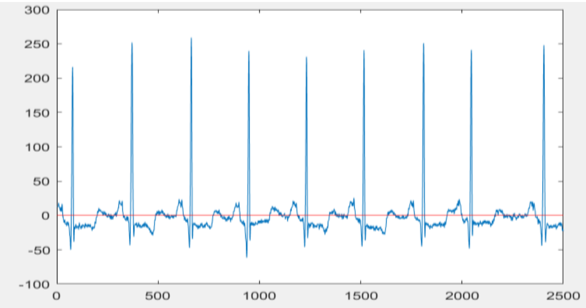
The ECG signals in the database primarily consist of 3 types of artifacts:

1. Baseline Wander
2. Powerline Interference
3. Electrode Motion Artifacts

3.4.1. Baseline Wander Effect:

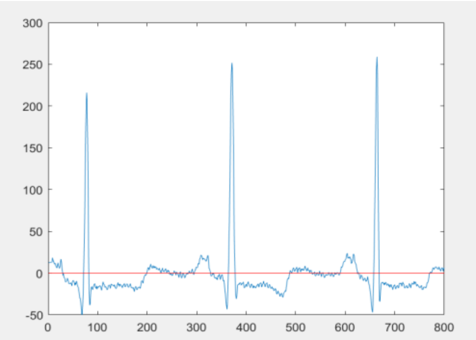
Baseline Wander is the effect where the base axis of the signal shifts from its normal base (Fig. 3.4.1). Baseline error is caused due to improper electrodes, patient's movement and breathing. To remove Baseline Wander from the ECG signal, Discrete Wavelet Transform is being used. In this method, the signal is decomposed using subsequent LPF and HPF. The cut-off frequency for both these filters will be half the sampling frequency. Hence nine-level decomposition is required to remove the baseline wander of approximately 0.5 Hz frequency

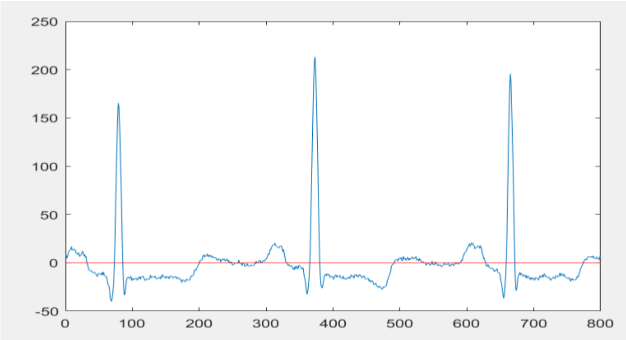




3.4.2. Powerline Interference:

Powerline interference is caused by the electromagnetic fields in the powerline. This error makes analysis and interpretation of ECG signals difficult as it completely superimposes the low frequency P, T waves. Powerline interference is reduced in the signal by using a filter defined by a complex conjugate pair of zeros that lie on a unit circle at the interfering frequency. To make sure the filter doesn't attenuate the ECG wave along with the powerline interference, it is modified such that the notch becomes more selective and removes only the powerline interference.

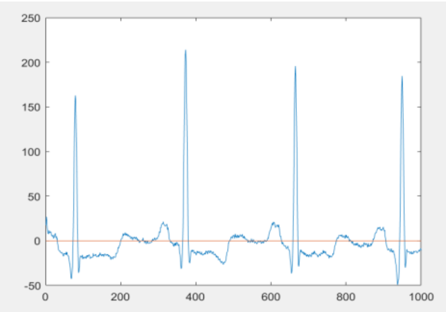


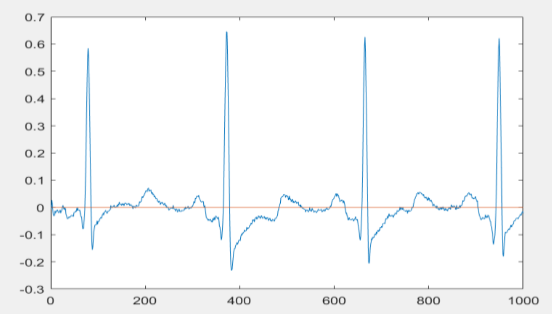


3.4.3. Electrode Motion Artifacts

Electrode Motion Artifacts are mainly caused by variation of impedance of the skin around the electrode mainly caused by skin stretching. They resemble errors caused by Baseline Wander but are more problematic because they are manifested in large amplitudes and sometimes can be mistaken for QRS complex.

Widely used method to remove Electrode Motion Artifacts is by using adaptive filters. Adaptive filters require two inputs, the primary input and a reference signal. Here we are using Least Mean Square Adaptive Filter for Electrode Motion Artifacts removal.





4.Results and Discussions

Segmentation Results:

The segmentation of the signals from the MIT BIH Arrhythmia Database was tried on two software, MATLAB and Python.

For segmentation using Python, the following methods were implemented :

Two moving averages algorithm: Moving average smoothing is a naive and effective technique in time series forecasting. It can be used for data preparation, feature engineering and even directly for making predictions.

Hamilton Detector: Hamilton Detector is an easy and effective algorithm which can be easily modified for different sample rates.

The accuracies were determined by comparing the number of R-peaks in the segmented signals with the number in the annotation files.

Two Moving Averages Algorithm:

This is a calculation executed regarding "Recurrence Bands Effects On QRS Detection",Elgendi, Mohamed &Jonkman, Mirjam& De Boer, Friso. (2010). This calculation comprises of three primary stages: Bandpass separating, Generating possible squares and Thresholding. This calculation is required to give 99.92% exactness as referenced in the paper.

Hamilton Detector:

This algorithm implemented with reference to “Open Source ECG Analysis Software Documentation”, E.P. Limited, 2002. This algorithm consists of the following stages: Low-pass filtering, High pass filtering, Differentiation, Moving average, Peak detection, and Checking detection rules. This algorithm is expected to give 96.48% accuracy as mentioned in the paper.

Implementation of Algorithms

For easier and more efficient implementation of the algorithms, the py-ecg-detectors library is imported and used in the program implemented in Python 3.5. This package includes all the detection algorithms mentioned above as easy to implement functions which were designed with reference to the above-mentioned papers. The complexes are detected and the time periods for each segment is found and stored in .csv files

Naïve–Bayes Classifier:

It is a classification technique based on Bayes Theorem with an assumption of independence among predictors.We implemented our own Naive Bayes binomial and multinomial classifiers in Python. We use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction. In the first case, the training-testing data was split 80% - 20% and in the second case, the training-testing data was split 80% - 20%. The results are summarized below–

Table 4.1: Naïve-Bayes with PCA

After running the test dataset obtained from different types of segmentation in Naïve Bayes Model with PCA the following data was obtained.

Training Testing Size	Training Accuracy	Testing Accuracy
80%-20%	79%	62%
70%-30%	78%	65%

Table 4.2:Navie Bayes with RF

After running the test dataset obtained from different types of segmentation in Naïve Bayes Model with RF the following data was obtained.

Training Testing Size	Training Accuracy	Testing Accuracy
80%-20%	75%	63%
70%-30%	74%	62%

SVM(Support Vector Machines):

In SVM, a hyperplane is chosen to best separate the focuses in the info variable space by their group, either class 0 or class 1. In two-measurements you can picture this as a line. You can make characterizations utilizing this line. By connecting input esteems into the line condition, we as certain whether another point is above or underneath the line. We attempted both the polynomial and the direct parts for the SVM and discovered that the straight piece beat the polynomial.

Table 4.3: SVM with PCA

After running the test dataset obtained from different types of segmentation in SVM Model with PCA the following data was obtained.

Training Testing Size	Training Accuracy	Testing Accuracy
80%-20%	99%	75%
70%-30%	99%	71%

Table 4.4 :SVM with RF

After running the test dataset obtained from different types of segmentation in SVM Model with RF the following data was obtained.

Training Testing Size	Training Accuracy	Testing Accuracy
80%-20%	100%	71%
70%-30%	99%	70%

5.Conclusion

In this project, a comparative study was performed on the different segmentation and Machine Learning algorithms to determine the best, novel method to detect arrhythmia from ECG signals. Segmentation of ECG signals using Two Moving Averages algorithm on Python proved to be have the highest degree of accuracy over other segmentation algorithms. Analysis of various Machine Learning models yielded SVM (Linear Kernel) as the most accurate models for the detection of arrhythmia. A milestone towards the aim of creating a high accuracy, minimum error system in the analysis of ECG signals and detection of arrhythmia was reached. This project can prove to be the stepping stone to the future of Medicine by providing clinical aid to experts and as a patient self-monitoring system in areas where access to an expert is difficult. Through this, improved and faster diagnosis and treatment can help decrease the occurrences of a large category of cardiovascular diseases, paving the way for a healthier world

References

1) Javier Aspuru, Alberto Ochoa-Brust, Ramon A Felix, Walter Mata-Lopez, Luiz J. Mena, Rodolfo Ostos and Rafael Martinez-Palaez. (2019) “Segmentation of the ECG Signal by Means of a Linear Regression Algorithm”. In : US National Library of Medicine, Feb 2019.
2) C. Celin and K. Vasanth. “ECG Signal Classification Using Various Machine Learning Techniques”. In : L Med Syst 42, 241 (2018).
3) Zerina Masetic, Abdulhamit Subasi. “Congestive heart failure detection using random forest classifier”. In : Computer Methods and Programs in Biomedicine (March 2018).
4) Renato Cuocolo, Teresa Perillo, Mario Petretta. (2019) “Current Applications of Big Data and Machine Learning in Cardiology”. In : Institute of Geriatric Cardiology, Chinese PLA General Hospital.
5) Miquel Alfaras, Miguel C. Soriano, Silvia Ortin. (2019) “A Fast Machine Learning Model for ECG-Based Heartbeat Classification and Arrhythmia Detection”. In : Institute for Cross-Disciplinary Physics and Complex Systems IFISC (UIB-CSIC).

6) Jens- Rainer Ohm. “Signal Decomposition”. In : RWTH Aachen University, Aachen, Germany (2004). From the book Signals and Communication Technology (pp.295-332)
7) Shailesh Kumar, P. K. Sahu . “Denoising of Electrocardiogram (ECG) Signals by using empirical mode decomposition (EMD) with non-local mean (NLM) technique”. In : Dept. of Electrical Engineering, NIT Rourkela, India (2018).
8) Rashid Ghorbani Afkhami, Ghanbar Azarnia, Mohammad Ali Tinati. “Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals”. In : Pattern Recognition Letters, Volume 70, 15 January 2016, Pages 45-51, (November 2015).
9) Ivaylo I Christov. “Real time electrocardiogram QRS detection using combined adaptive threshold”. In : BioMedical Engineering Online, (27th August 2004).
10) Rahul Kher. “Signal Processing Techniques for Removing Noise from ECG Signals”.In : J Niomed Eng 1: 1-9 (2019).
11) Mohamed Elgandi, Mirjam Jonkman, Frisco DeBoer. “ Frequency Bands Effects on QRS Detection”. In : DataBase Systems and Logic Programming (2004).
12) Jiapu Pans , Willis J Tompkins. “A real-time QRS Detection Algorithm”. In : IEEE Transactions on Biomedical Engineering, Vol. BMW -32. (March 1985)
13) Jiapu Pans , Willis J Tompkins. “A real-time QRS Detection Algorithm”. In : IEEE Transactions on Biomedical Engineering, Vol. BMW -32. (March 1985)
14) National Heart, Lung, and Blood Institute. July 1, 2011. from the original on 19 February 2015. Retrieved 7 March 2015
15) GBD 2013 Mortality and Causes of Death, Collaborators (17 December 2014).
16)]Meigas, K; Kaik, J; Anier, A (2008). "Device and methods for performing transesophageal stimulation at reduced pacing current threshold". Estonian Journal of Engineering.
17) Software-Defined Perimeter (SDP): The Next-Generation Secure Vpn Solution Built For Future Networks, A.Shaji George, Bashiru Aremu, International Journal Of Advance Research In Science And Engineering http://www.ijarse.com IJARSE, Volume No. 09, Issue No. 10, October 2020 ISSN-2319-8354(E).
18) Pehrson, Steen M.; Blomströ-LUNDQVIST, Carina; Ljungströ, Erik; Blomströ, Per (1994)
19) Mandel, William J., ed. (1995). Cardiac Arrhythmias: Their Mechanisms, Diagnosis, and Management (3 ed.). Lippincott Williams & Wilkins
20) Wiener, Norbert; Rosenblueth, Arturo (1946). "The mathematical formulation of the problem of conduction of impulses in a network of connected excitable elements, specifically in cardiac muscle".
21) Peressutti, Caroline; Martín-González, Juan M.; m.García-Manso, Juan; Mesa, Denkô (2010). "Heart rate dynamics in different levels of Zen meditation". International Journal of Cardiology.