Machine Learning Based Decision Support System for Atrial Fibrillation Detection using Electrocardiogram

Shrikanth Rao S K
Department of ECE,
VCET
Puttur, India
skanthrao@gmail.com

Roshan Joy Martis
Senior Member,IEEE
Department of CSE,Global Academy of Technology,
Bengaluru, India
roshaniitsmst@gmail.com

Abstract— **Atrial Fibrillation (AF) is a common sustained arrhythmia encountered in regular clinical practice. In order to diagnose AF, Electrocardiogram (ECG) is used in correlation with clinical symptoms. ECG is noninvasive and cost effective modality in order to diagnose cardiac abnormalities using AF. The complexity of ECG and its interrelationship with other physiological parameters make the AF detection a challenging task in the clinical practice. The traditional practice of diagnosing AF manually by the physician can cause intra physician variability leading to a need for automated algorithm based assisting system to detect AF. In the present methodology, the QRS complex is detected and each beat in the entire signal is segmented, the median beat is calculated for a given signal, the dimensionality is reduced using Principal Component Analysis (PCA) and the resultant components along with energy values are used for classification using decision tree. The methodology provided an improved average accuracy of 85.1 percent which is reasonably high. The system developed can be used in many practical applications and can provide acceptable results in clinical implementations. The developed methodology can be used as an adjunct tool by the physician in his clinical practice.**

Keywords—Atrial fibrillation, The cardiovascular disease, QRS complex, principal component analysis.

I. Introduction

The cardiovascular disease (CVD) is a group of disorders relating to the heart, blood vessels and circulation system. In the ancient days, the cause of mortality was mainly due to infectious diseases and malnutrition. However due to elevated fat rich food consumption, overweight, physical inactivity, diabetes mellitus, smoking, alcohol consumption etc. have contributed to epidemiological transition and many manmade diseases have come into existence. CVD is one of the manmade disease. American heart association and American stroke association [1], in its report in the says that, 2, 81,3503 resident deaths/year have taken place in the United States, which means average of 7708 death every day due to CVD. In 2017, ≈17.8 million deaths have taken place across the world due to CVD. This attributed to 21.1% increase in the mortality rate compared to that in 2007 [1].

CVDs can easily be diagnosed by measuring the electrical potential on the body surface by a method called electrocardiography (ECG). Electrocardiogram is the cardiac electrical activity of the heart shown by P, Q-R-S and T wave. The subtle duration and amplitude of ECG at various time positions provide diagnostic insight required to identify various CVDs. The cardiac abnormalities which involve rhythm disturbance are called as arrhythmia. The various arrhythmias having life-threatening implications include AF, ventricular fibrillation, atrial flutter, ventricular flutter etc.

AF is a common arrhythmia noticed in clinical practice which has life threatening implications. The physician generally diagnoses AF by evaluating the ECG. The subtle changes in amplitude and duration are used to diagnose AF by the physician. However it involves a laborious process to manually interpret the minute changes in the amplitude and duration. Automation of AF detection using computers and computational algorithms is an inevitable solution to this problem.

In this direction of automated AF detection using computational algorithms many studies are reported so far. Tateno et al. [2] proposed a method for automated detection of AF using coefficient of variation and density histogram of RR intervals and demonstrated sensitivity of 94.4% and specificity of 97.2%. Lian et al. [3] detected AF using RR intervals and demonstrated robust AF detection. Lee et al. [4] implemented AF detection using a smartphone and demonstrated 100% accuracy. Christov et al. [5] proposed a method to discriminate AF, normal sinus rhythm and other arrhythmia using heart rate variability analysis, beat morphology etc. and demonstrated an overall accuracy of 80%. Bin et al. [6] demonstrated AF detection with an overall accuracy of 82% using combination of morphology, interval and similarity index features. Ghiasi et al. [7] implemented AF detection using deep convolutional neural network to obtain an overall accuracy of 71% on test dataset. Garcis et al. [8] proposed a method to detect AF using various interval features and demonstrated an overall accuracy of 71%. Billeci et al. [9] implemented AF detection on a smartphone by using a set of 30 interval based features and reported an overall accuracy of 83%. Christov et al. [10] proposed a method for AF detection using various features such as heart rate variability, beat morphology, atrial activity etc. and demonstrated an overall accuracy of 80%. Kruger et al. [11] implemented a bimodal classification algorithm for AF detection using hand held device to achieve a sensitivity of 100% and specificity of 93%. Mishra et al [12] developed an algorithm to detect AF

978-1-7281-9885-9/20/$31.00 ©2020 IEEE

using Discrete Wavelet Transform (DWT) based on inverted T wave logic and ST-segment elevation to achieve accuracy of 57%.

Zabihi, et al [13] implemented AF detection algorithm using hand held ECG monitoring device to achieve overall score of 82.6%

There are many methods available in the literature to detect AF using various interval based features. During AF there will be some subtle change in the morphology, distribution of energy in the time domain samples, fibrillatory waves due to irregular beating of heart, etc. Due to this characterizing AF only by interval based features is not appropriate. The present study involves investigation of features derived from entire ECG beat. In view of this a methodology of classification of normal, AF and other rhythms on the basis of time samples and energy values is formulated.

Section II describes the methodology used, section III provides the results, section IV gives discussion and section V concludes the article.

II. Methodology

The proposed methodology is depicted in *Figure 1*. It consists of QRS detection and beat segmentation, Principal Component Analysis (PCA) and energy computation followed by pattern classification using decision tree.

A. Dataset Used

We have used the physionet computing cardiology challenge 2017 dataset [14]. AliveCor has used ECG recordings, obtained using the AliveCor system, for this challenge. The training set contains 8,528 ECG recordings of single lead of duration 9 s to just over 60s and the test set contains 3,658 ECG recordings of duration 9 s to just over 60s. The chosen data consists of 5,052 normal signals, 759 AF signals and 2,498 other rhythm signals. Each such signal is sampled at a rate of 300 Hz.

B. QRS Complex Detection and Segmentation

Each signal chosen from the database is subjected to discrete wavelet transform (DWT) denoising to remove unwanted portions in the signal such as baseline wander and other high frequency noise. The QRS complexes are detected using Pan Tompkins algorithm [15].

QRS complex detection involves passing the ECG signal through low pass filter followed by high pass filter, derivative filter to acquire information about the QRS slope. The filtered signal is squared to enhance the dominant peaks finally decision rules are applied to find R peak.

C. Data Reduction using Principal Component Analysis

Principal component analysis method is used for dimensionality reduction. The 200 samples of the median beats of all the signals are used in this stage. The dimensionality of the ECG beats is reduced using Principal the 99 samples before the QRS complex, the peak and the 100 samples after the QRS complex are considered as one ECG beat. So each beat is chosen as a 200 sample segment. All the beats present in a given signal are averaged by taking median and the resultant median beat is used for further processing.

Component Analysis (PCA) [16]. The PCA computes the data covariance matrix, the eigenvalue decomposition of data covariance matrix, the sorting of eigenvectors based on the descending order of eigenvalues and finally the projection of given data onto the sorted eigenvectors directions. Ten components of PCA are used such that most of the data variability is contained in the ten principal components.

The PCA consists of following steps.

a. Subtracting the mean, $\bar{X}$ from the given data, X_{org}

$$X = X_{org} - \bar{X} \quad (1)$$

where $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ and N is the dimension of the data, which is 200.

b. Computation of data covariance matrix, C as

$$C = \frac{1}{N} X X^T \quad (2)$$

c. Eigenvalue decomposition of C, into diagonal matrix, D and Eigenvector matrix, U as,

$$C = UDU^T \quad (3)$$

d. Sorting of Eigenvectors based on the descending order of Eigenvalues.

e. Projection of the given data onto the directions of sorted eigenvectors and consider the only first few components (10 in this case), which consists of most of the data variations.

D. Energy Computation

In addition to the principal components, the energy contained in each median heartbeat in the interval of beginning of the heart beat till the QRS midpoint (i.e, in the first 100 samples) is evaluated as,

$$E = \sum_{i=1}^{100} X_i \quad (4)$$

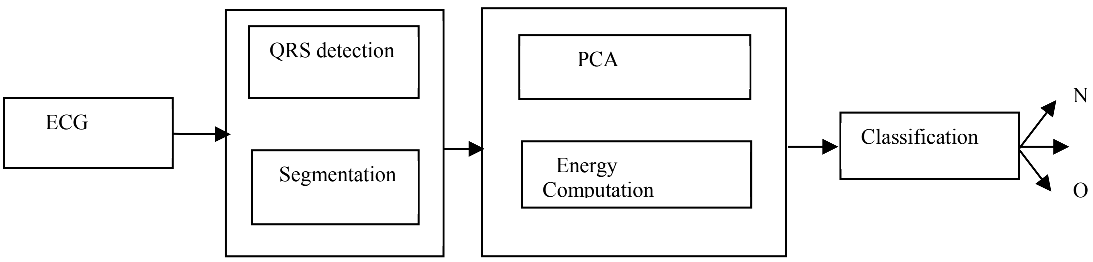

Figure1: Atrial fibrillation (AF) detection methodology using ECG

Where X_i is the ith sample in the median heartbeat. The ten principal components along with the energy value are used for the decision tree classifier for pattern identification.

E. Classification using Decision tree

Decision Tree is a supervised, nonparametric classifier. The purpose of using a Decision Tree is to build a model that can be used to classify the target variables class or value by learning basic rules of decision inferred from training set. The classifier has tree like structure, and has root node which is further branched into child nodes and leaf node based on if-else statements.

The decision tree (DT) is used for classification of normal, AF and other rhythm signals from the 11 features derived from the previous stages. The DT consists of different nodes and each node branches into child nodes till the leaf nodes.. The branching at each node is decided by the rule which is derived during training phase such that the discrimination is provided in separating the three classes. Based on the rules learnt during the training phase, the test data is classified into one of the three defined classes (N, AF and O).

F. Ten fold Cross Validation

The data is divided into 10 non overlapping subsets. In each fold there is training and testing phases. There are such ten folds. For each fold a subset is used for testing and remaining 9 subsets are combined to train the DT. In each fold the classification performance as class specific accuracy and overall accuracy is obtained.

III. RESULTS

The proposed methodology of classification of normal, AF and other rhythm signals is implemented in MATLAB. In total 5,052 normal signals, 759 AF signals and 2,498 other rhythm signals are studied. Figure 2 shows one of the original ECG signal from database. Each of ECG signal is subjected to DWT denoising followed by QRS complex detection. Figure 3 shows the detected QRS peaks using Pan Tompkins algorithm. The detected peaks are indicated by red asterisk.

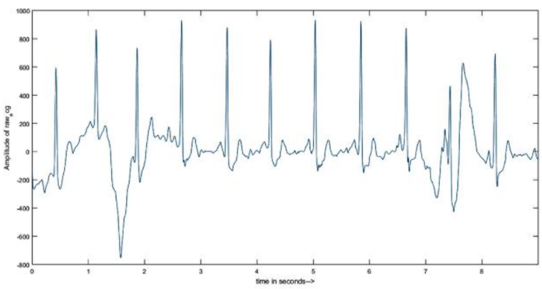

Figure 2: Original ECG signal of sample A00001

The energy of average beat is also computed. The reduced Based on the detected QRS peak ECG signal is segmented into different beats followed by median averaging of all the beats The resultant average beat is subjected to PCA to reduce the dimension. The energy of average beat is also computed. The reduced 10 principal component along with energy are combined to form feature vector. The features are classified using decision tree. The results of classification are shown in Table 1.

As seen from Table 1, the proposed methodology provides an overall accuracy of 85.1%, normal class specific accuracy of 83.1%, AF class specific accuracy of 87.5% and other rhythm class specific accuracy of 84.8%.

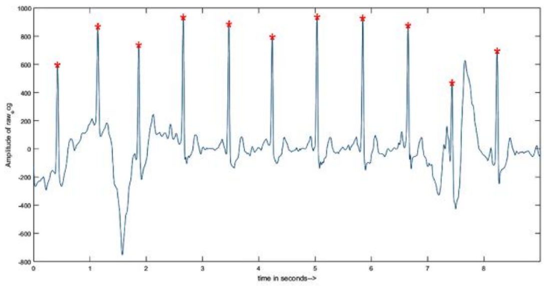

Figure 3: Detected QRS peaks using Pan Tompkins algorithm

Table 1: Average classification performance of DT

Class specific accuracy			Overall Accuracy
Normal	**AF**	**Other**	
83.1	87.5	84.8	85.1

IV. DISCUSSION

Summary of studies on physionet challenge 2017 is provided in Table 2. A method to discriminate AF, normal sinus rhythm and other arrhythmia using heart rate variability analysis, beat morphology etc. is proposed by Christov et al. and reported an overall accuracy of 80% [5]. By using combination of morphology, interval and similarity index features AF is detected with 82% of overall accuracy [6]. Ghiasi et al. [7] implemented AF detection using deep convolutional neural network to obtain an overall accuracy of 71% on test dataset. In [8], Garcis et al proposed a method to detect AF using various interval features and demonstrated an overall accuracy of 71%. Billeci et al. [9] implemented AF detection on a smartphone by using a set of 30 interval-based features and reported an overall accuracy of 83%. Christov et al. [10] proposed a method for AF detection using various features such as heart rate variability, beat morphology, atrial activity etc. and demonstrated an overall accuracy of 80%. Mishra et al [12] proposed filter diagonalization method to extract f-wave features to achieve overall accuracy of 57% Zabihi et al [13] proposed an algorithm to detect AF using random forest classifier to achieve overall score of 82.6%. The proposed methodology is applied on Physionet challenge 2017 database. A three class pattern analysis problem of classifying Normal, AF and other rhythm is addressed. An overall accuracy of 85.1% is obtained.

A prototype is implemented using MATLAB. The authors believe that the developed prototype model is of immense aid as an adjacent tool to the healthcare practitioners and primary care physicians. The developed methodology provides automated classification of AF, Normal and Other rhythm. It can be used for screening patients for atrial tachyarrhythmias, particularly geriatric patients.

Table 2: Summary of studies on classification of normal, AF and other rhythm using physionet challenge 2017 dataset

Literature	Methods	Classifier	Overall accuracy
Christov et al, 2017 [5]	HRV and beat morphology.	LDA	80%
Bin et al, 2017 [6]	Morphology, interval and similarity index features.	DT ensemble using Ada Boost M2	82%
Ghiasi et al, 2017 [7]	Geometric fractal dimension, correlation coefficient and variance of R-R intervals	Deep convolutional Neural Network	71%
Garcí et al, 2017 [8]	Ventricular activity variability, atrial activity variability.	SVM	71%
Zabihi et al 2017 [13]	HRV and beat morphology	Random forest	82.6%
Billeci et al, 2018 [9]	RR series derived features, frequency spectrum analysis features, P wave analysis features, etc.	LS SVM	83%
Christov et al, 2018 [10]	HRV, beat morphology, atrial activity etc.	LDA	80%
Mishra et al 2019 [12]	Filter Diagonalization Method	LDA	57%
Current study	Principal components, energy	DT	85.1%

Abbreviations: HRV = Heart Rate Variability; LDA = Linear Discriminant Analysis; DT = Decision Tree; SVM = Support Vector Machine; LS SVM = Least Square Support Vector Machine

V. Conclusion

The incidence of heart diseases is rising every year. Hence there is an increasing demand for development of computer assisted tools for heart disease screening. This will provide better outcome for the patients and also reduces healthcare cost. AF is a frequently encountered rhythm disorder of the heart in the clinical practice. If AF is unnoticed and not diagnosed in the early stage, it has life threatening implication such as occurrence of atrial flutter, incidence of other morbid condition of the heart. The proposed method provides a computational paradigm of AF detection with an overall accuracy of 85.1%, normal class specific accuracy of 87.5% and other rhythm class specific accuracy of 84.8%.

The authors believe that the proposed methodology provides valuable insights for further improvements. As a future direction the features used in this study be used along with RR interval derived features and further exploration is possible.

Acknowledgment

The authors thank the contributors of physionet challenge 2017 dataset [14] for making the dataset available.

References

[1] Virani, Salim S., Alvaro Alonso, Emelia J. Benjamin, Marcio S. Bittencourt, Clifton W. Callaway, April P. Carson, Alanna M. Chamberlain et al. "Heart disease and stroke statistics—2020 update: a report from the American Heart Association." *Circulation* (2020): E139-E596.

[2] Tateno, K., and L. Glass. "Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ΔRR intervals." in *Medical and Biological Engineering and Computing* 39, no. 6, 2001, pp. 664-671.

[3] Lian, Jie, Lian Wang, and Dirk Muessig. "A simple method to detect atrial fibrillation using RR intervals." *The American Journal of Cardiology.* vol.107, no. 10, pp. 1494-1497, 2011.

[4] Lee, Jinseok, Bersain A. Reyes, David D. McManus, Oscar Maitas, and Ki H. Chon. "Atrial fibrillation detection using an iPhone 4S." *IEEE Transactions on Biomedical Engineering.* V ol. 60, no. 1 203-206, 2013.

[5] Christov, Ivaylo, Vessela Krasteva, Iana Simova, Tatiana Neycheva, and Ramun Schmid. "Multi-parametric analysis for atrial fibrillation classification in ECG." In *2017 Computing in Cardiology (CinC)*, pp. 1-4. IEEE, 2017.

[6] Bin, Guangyu, Minggang Shao, Guanghong Bin, Jiao Huang, Dingchang Zheng, and Shuicai Wu. "Detection of atrial fibrillation using decision tree ensemble." In *2017 Computing in Cardiology (CinC)*, pp. 1-4. IEEE, 2017.

[7] Ghiasi, Shadi, Mostafa Abdollahpur, Nasimalsadat Madani, Kamran Kiani, and Ali Ghaffari. "Atrial fibrillation detection using feature based algorithm and deep convolutional neural network." In *2017 Computing in Cardiology (CinC)*, pp. 1-4. IEEE, 2017.

[8] García, Manuel, Juan Ródenas, Raúl Alcaraz, and José J. Rieta. "Atrial fibrillation screening through combined timing features of short single-lead electrocardiograms." In *2017 Computing in Cardiology (CinC)*, pp. 1-4. IEEE, 2017.

[9] Billeci, Lucia, Magda Costi, David Lombardi, Franco Chiarugi, and Maurizio Varanini. "Automatic Detection of Atrial Fibrillation and Other Arrhythmias in ECG Recordings Acquired by a Smartphone Device." *Electronics,* vol. 7, no. 9, pp. 199, 2018.

[10] Christov, Ivaylo, Vessela Krasteva, Iana Simova, Tatiana Neycheva, and Ramun Schmid. "Ranking of the most reliable beat morphology and heart rate variability features for the detection of atrial fibrillation in short single-lead ECG." *Physiological Measurement* vol. 39, no. 9 pp. 094005, 2018.

[11] Kruger, Grant H., Rakesh Latchamsetty, Nicholas B. Langhals, Miki Yokokawa, Aman Chugh, Fred Morady, Hakan Oral, and Omer Berenfeld. "Bimodal classification algorithm for atrial fibrillation detection from m-health ECG recordings." *Computers in Biology and Medicine* vol. 104, pp. 310-318, 2019.

[12] Mishra, Saumitra, Sreehari Rammohan, Khalid Z. Rajab, Gurpreet Dhillon, Pier Lambiase, Ross Hunter, and Elaine Chew. "Atrial Fibrillation Stratification via Fibrillatory Wave Characterization Using the Filter Diagonalization Method." *In Computing in Cardiology, pp. 1-4. IEEE, 2019.*

[13] Zabihi, Morteza, Ali Bahrami Rad, Aggelos K. Katsaggelos, Serkan Kiranyaz, Susanna Narkilahti, and Moncef Gabbouj. "Detection of atrial fibrillation in ECG hand-held devices using a random forest classifier.*" In 2017 Computing in Cardiology (CinC), pp. 1-4. IEEE, 2017.*

[14] Clifford, Gari D., Chengyu Liu, Benjamin Moody, H. Lehman Li-wei, Ikaro Silva, Qiao Li, A. E. Johnson, and Roger G. Mark. "AF Classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017." *In 2017 Computing in Cardiology (CinC), pp. 1-4. IEEE, 2017.*

[15] Pan, Jiapu, and Willis J. Tompkins. "A real-time QRS detection algorithm." *IEEE Trans. Biomed. En, vol. 32*, no. 3, pp. 230-236, 1985.

[16] Martis, Roshan Joy, U. Rajendra Acharya, K. M. Mandana, Ajoy Kumar Ray, and Chandan Chakraborty. "Application of principal component analysis to ECG signals for automated diagnosis of cardiac health.*" Expert Systems with Applications vol.39, no. 14, pp. 11792-11800, 2012.*