Classification of Electrocardiogram signals using Different Machine Learning Techniques: A Survey

Piyushkumar M. Prajapati,M.E student, Gujarat Technological University, Ahmedabad, INDIA
Mr. Amit Chavan, Project Engineer, C-DAC PUNE, INDIA
Mr. Aditya Kumar Sinha, Joint director, C-DAC PUNE, INDIA

ABSTRACT

Nowadays heart diseases are increasing with a remarkable rate. cardiovascular diseases are major reason for death of human. There for monitoring and detection of abnormal heart condition is important to recognize heart issue. By Analyze the ECG signal identified the heart problem. This presents paper gives a detailed survey of different pre-processing, feature extraction and classification techniques. Furthermore, our paper also presents comparison of performance parameter as sensitivity, specificity, and accuracy of the classifiers which help to evaluate the best classifiers techniques.

KEYWORDS

Electrocardiogram(ECG) Signal, MIT-BIT database, Cardiovascular Disease(CVD), Fuzzy logic, Neural Network, Support Vector Machine(SVM), Wavelet Transform(WT)

INTRODUCTION

An electrocardiogram (ECG) is a medical test which detects cardiac abnormality by measuring the electrical activity generated by the heart. A heart produces tiny electrical impulses which spread through the heart muscle. These impulses can be detected by an ECG machine. An ECG machine records the electrical activity of the heart and displays this data as a trace on a paper. This data is then interpreted by a medical practitioner. ECG helps to find the cause of symptoms helps to detect abnormal heart rhythm or cardiac (heart) abnormalities or that we identified or Cardiovascular Diseases. One of main Cardiovascular Diseases called Arrhythmia. Arrhythmia are critical heart diseases. Arrhythmia occurs due to the problem in the electrical conduction system of the heart. If these diseases are not detected in time, the activity and the structure of the heart gets deteriorated quickly [1]and that identified with electrocardiogram signal. The recognition of the beats in electrocardiography is a very important subject in the intensive care units, where the recognition and classification of the electrocardiographic (ECG) signals so important. Several algorithms have been developed in the literature for the detection and classification of ECG wave forms. For ECG signal analysis the main sequential steps are required a Pre-processing, feature extraction, normalization, and classification. Researchers have applied different pre-processing techniques for ECG classification. For noise removal, techniques such as low pass linear phase filter, linear phase high pass filter median filter, linear phase high pass filter, mean median filter etc. are used. Feature extraction techniques used by researchers are Wavelet Transform (WT), Continuous Wavelet Transform (CWT), Discrete Cosine Transform (DCT), S Transform (ST), Discrete Fourier transform (DFT), Principal Component Analysis (PCA), Daubechies wavelet (Db4), Pan- Tompkins algorithm, Independent Component Analysis (ICA) etc. Classification techniques used are Neural Network (NN), Fuzzy C-Means clustering (FCM), Feed forward neuro-fuzzy, Support Vector Machine (SVM), Quantum Neural Network (QNN), Radial Basis Function Neural Network (RBFNN), Type2 Fuzzy Clustering Neural Network (T2FCNN) and Probabilistic Neural Network (PNN) classifier etc. Most research commonly used MIT-BIH arrhythmias database. MIT-BIH database consist of several ECG signal that indicate different type of diseases and abnormalities of the heart rhythmse used.

The database contains 48 half-hour recordings each containing two ECG lead signals (denoted as lead A and lead B). In 45 recordings, lead A is a modified limb lead II (MLII), and lead B is commonly a modified lead V1. In the other recordings, lead A is V5 and lead B is V2 or MLII. The lead signals were band-pass filtered at 0.1-100 Hz and digitized at 360 Hz. The data collection usually taken randomly from patient in the hospital which commonly taken by the physician as their research studies [7].

ECG SIGNAL ANALYSIS

RR: interval between R wave and the next R wave normal duration is 0.6-1.2 s, P: first short upward movement of the ECG normal duration 80ms, PR: measured from the beginning of the P wave to the beginning of the QRS complex normal duration 120-200 ms, QRS normally begins with a downward deflection, Q: a larger upwards deflection R and ends with a downward S wave normal duration 80-120 ms, PR: connects the P wave and the QRS complex normal duration 50-120 ms.

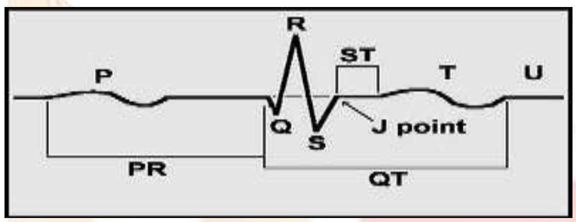

Figure 1: ECG SIGNAL [17]

J-point: The point at which the QRS complex finishes and the ST segment begins is called J-point. ST: connects the QRS complex and the T wave normal duration 80-120 ms,

T: normally a modest upward waveform normal duration 160 ms, ST: measured from the J point to the end of the T wave normal duration 320 ms, QT: measured from the beginning of the QRS complex to the end of the T wave normal duration 420 ms, U normally has low amplitude and often it is completely absent.

BACKGROUND SURVEY

1) Pre-processing: -Pre-processing of physiological signals is important part of this research. Since, the noise and artefact contamination of ECG signals will affect the efficient interpretation of clinical information and early diagnosis of CVDs diseases. For the ECG pre-processing signal, the previous research works have used such as discrete wavelet transform (DWT) and IIR filter for example Chebyshev Type 1, Elliptic filter and Butterworth filter [2] [4] [5]. These filters used to filter the unwanted signal and also decompose the signal to get the exact ECG signal that can be extracted to get the desire characteristics features to be classified. Conventional digital filtering technique such as Butterworth filter, Elliptic filters are used to remove the noises and artefacts from the ECG through windowing approach and filtered signals are smoothened using additional filtering such as median filter [6].

2) Features Extraction: -The purpose of the feature extraction process is to select and retain relevant information from the original signal. There are different feature extraction methods used in the previous studies in order to determine the early prediction of cardiovascular diseases [8] [9] [10]. Most of the studies used the P-QRS-T wave as the features. Different features represent different type of diseases and abnormalities. There is also study that used the QRS complex as the features because it has a sharp biphasic or diphasic wave of 1 mV amplitude and duration of approximately 80-100 ms, where the R-wave denoted the peak point of QRS complex as stated in [10]. while the QRS complex also have been analysing with the high and lower frequency QRS complex. In the features extraction stage, the previous researchers have used different methods feature extraction techniques such as Fast Fourier Transform (FFT), Discrete Wavelet Transform (DWT), Auto

Regressive approaches, etc. for extracting the statistical features from the raw ECG signal. ECG signals are usually varies from individual to individual, the specific waveform components, such as P-wave, the QRS-complex and the T-wave, will have different characteristics for different individuals. ECG signal of different cardiac disease generate distinct patterns in the time-frequency domain. In this research, DWT will be used as the features extraction tool. This proposed method will be applied before the QRS detection with db6 as the mother wavelet and decompose by 8 level of decompositions. The QRS detection is determined using Pan Tompkins's Algorithm likewise most of previous research used [8].

3) Classification: -In order to identify the abnormal CVDs due to the traditional risk factor such as tobacco smoking, there are several types of classifier can be used in the previous research works such as Artificial Neural Network (ANN) [11] [16], Fuzzy Logic system [12], Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM). Most of the researchers have used SVM and Fuzzy Logic system for CVDs classification using ECG signals [4] [5] [13]. The SVM technique was developed by Vapnik, in recent years it has proved to be the advanced tool in solving classification or pattern recognition scheme [8]. A fuzzy-logic system is supposed to mimic how a human would make decision but only much faster [9].

For performance measure of signal classification measures sensitivity, specificity, accuracy etc. Evaluation measures calculated from confusion matrix are Sensitivity, Specificity and Accuracy. Sensitivity is the ratio of true positive beats to total of true positive and false negative beats. Specificity is the ratio of true negative beats to total of true negative and false positive beats. The overall accuracy is the ratio of total number of true negative and true positive beats to total number of beats.

Specificity = 1 – [TN / (TN + FP)]

Sensitivity = TP/ (TP+ FN)

Accuracy = TN +TP/(TN+ FN +FP+ TP)

Precision =TP/ (TP + FP)

SURVEY OF ECG CLASSIFICATION METHODOLOGY

1.ECG Arrhythmia Classification with Support Vector Machines and Genetic Algorithm [18].

The proposed method combines both Support Vector Machine (SVM) and Genetic Algorithm approaches. In this twenty-two features from electrocardiogram signal are extracted. SVM classifier is optimized by searching for the best value of the parameters that tune its discriminate function, and looking for the best subset of features that optimizes the classification fitness function. Experimental results demonstrate that the approach adopted better classifies ECG signals. Four types of arrhythmias were distinguished with 93% accuracy. They used a MIT-BIH Arrhythmia Database to train and test the classifier.

2. Detection of Myocardial Infarction and Arrhythmia from Single-Lead ECG Data using Bagging Trees Classifier [22].

The robust detection of both MI and AR from ECG signal using a minimalistic (single channel) ECG data. In this study, a total of 440 records from 12-lead ECG signals (79 Healthy Person, 346 MI, and 15 AR) have been used from "PTB Diagnostic ECG Database" of Physio Bank The proposed algorithm automatically identifies P, Q, R, S, and T points of ECG signals, and then extracts 33 features: 15 interval type and 18 amplitude type. Bagging Tree, an ensemble method which is computationally efficient and at the same time can deal with the class imbalance problem within the data, is used for classification. The process flow are given below in figure2 [17].

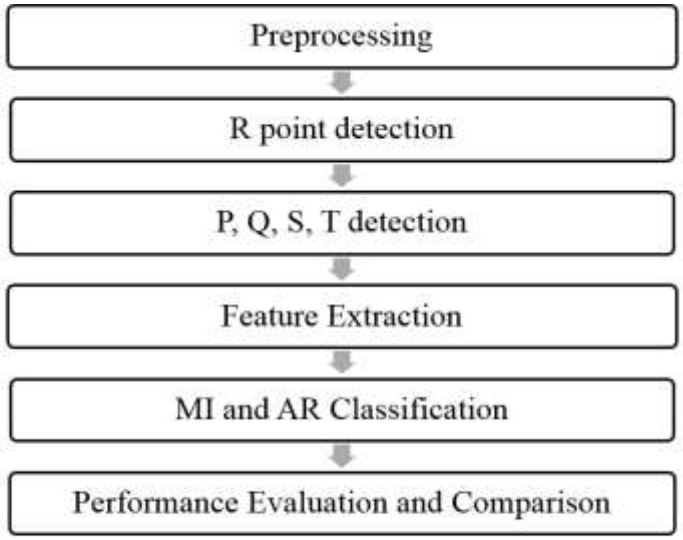

Figure 2: PROCESS FLOW

This work focuses on the detection of MI and AR from single-lead ECG data and compares different lead performances in order to find the most suitable lead so that any of these two diseases can be detected using a single classifier with minimum delay. To address the class imbalance issue and to reduce the computation, ensemble based bagging tree classifier is used. From the ECG signal of each lead, amplitude, amplitude difference, and time interval of the P, Q, R, S, T points are used to form the feature vector. Our results show that lead V4 performs the best to classify between MI, AR, and normal people among twelve leads in terms of cross validation accuracy, sensitivity, specificity, precision and demonstrates that Bagging Tree is able to identify MI, AR and normal patients with the cross-validation accuracy of 99.7%, sensitivity of 99.4%, specificity of above 99.5%, precision of 99.32%, and F1 score of 99.36% from a single lead ECG data (Lead V4). The proposed algorithm uses a minimalistic single channel ECG for robust detection of MI and AR that can beutilized in wearables for real-time patient monitoring.

3. Hybrid classification of Bayesian and Extreme Learning Machine for heartbeat classification of arrhythmia detection [21].

In this method Detection based on Extreme learning machine (ELM) has become a common technique in machine learning. However, it easily suffers from overfitting. This method hybrid classification technique using Bayesian and Extreme Learning Machine (B-ELM) technique for heartbeat recognition of arrhythmia detection AD. The proposed technique is capable of detecting arrhythmia classes with a maximum accuracy of (98.09%) and less computational time about 2.5s.

4. Arrhythmia Detection from Heartbeat Using Bayesian and Extreme Learning Machine

In this method provide Automatic interpretation of electrocardiography provides a non-invasive and inexpensive technique to analyze the heart activity for different cardiac conditions. The emergence of smartphones and wireless networks has made it possible to perform continuous Holter monitoring on patients or potential patients. Recently, much attention has been paid to the development of the monitoring methodologies of heart activity, which include both the detection of heartbeats in electrocardiography and the classification of types of heartbeats. However, many studies have focused on classifying limited types of heartbeats. We propose a system for classification into 17 types of heartbeats. This system consists of two parts, the detection and classification of heartbeats. The system detects heartbeats through repetitive features and classifies them using a *k*-nearest neighbour algorithm. Features such as the QRS complex and P wave were accurately extracted using the Pan- Tompkins algorithm. For the classifier, the distance metric is an adaptation of locally weighted regression. The system was validated with the MIT-BIH Arrhythmia Database. The system achieved a sensitivity of 97.22 % and a specificity of 97.4 % for heartbeat detection. The system also achieved a sensitivity of 97.1 % and a specificity of 96.9 % for classification.

5.Neuro-fuzzy-based Arrhythmia Classification Using Heart Rate Variability Features [23].

In this used a self-organizing neuro-fuzzy inference system called SONFIS [15] along with HRV-based feature extraction techniques to classify input ECG signals from several patients with the intention to determine whether they have or not a medical heart condition known as arrhythmia. The advantage of SONFIS over the use of ANN and SVM (besides the classification accuracy), is that the final model can be linguistically interpreted by means of the automatically generated fuzzy IF-THEN rules, thus providing physicians with a base of quantitative knowledge that can aid them for diagnosing this kind of medical condition. Although the aim of this work was not to obtain interpretable rules, the model itself is more interpretable than the SVM and FFANN models, so this gives an advantage over these approaches.

CONCLUSION

In this paper, survey of different technique of in ECG Classification using MIT-BIT data sate. The ECG signal processing remains the simplest non-invasive diagnosis method for determining various heart diseases. Hare different classification technique like Support Vector Machines and Genetic Algorithm, Bagging Trees Classifier, Bayesian and Extreme Learning Machine, Bayesian and Extreme Learning Machine, and a self-organizing neuro-fuzzy inference system(SONFIS) based classification in all of them the SONFIS classification technique give an advantage over these approaches for the ECG signal classifiers.

REFERENCES

[1] U.R. Acharya, et al., “Automated detection and localization of myocardial infarction using electrocardiogram: a comparative study of different leads,” Knowledge-Based Systemsn, vol. 99, 2016, pp. 146-15.

[2] Ghosh, D., Midya, B. L., Koley, C., & Purkait,P. (2005, 11-13 Dec. 2005). Wavelet Aided SVM Analysis of ECG Signals for Cardiac Abnormality Detection, Annual IEEE INDICON.

[3] Sufi, F., & Khalil, I. Diagnosis of Cardiovascular Abnormalities from Compressed ECG: A Data Mining-Base Approch. Information Technology in Biomedicine on.15(1),33-39.

[4] Er, K., Rajendra, A. U., Kannathal, N., & Choo Min, L. (2005, 2005). Data Fusion of Multimodal Cardiovascular Signals, 27th Annual International Conference of the Engineering in Medicine and Biology Society,2005."

[5] Pasupathi, P., Bakthavathsalam, G., Rao, Y. Y., & Farook, J. (2009). Cigarette Smoking-Effect of metabolic health risk: A review. Diabetes and Metabolic Syndrome: Clinical Research and Reviews, 3(2), 120-127.

[6] Alty, S. R., Millasseau, S. C., Chowienczyc, P. J., & Jakobsson, A. (2003, 30-30 Dec. 2003). Cardiovascular disease prediction using support vector machines, 2003 IEEE.

[7] Islam, M.r., Ahmad, s., Hirose, k., &moll m.
K.I. (May 30 2010-june 2 2010)

[8] A. Bakhshipour, M. Pooyan, H. Mohammadnejad, and A. Fallahi, "Myocardial ischemia detection with ECG analysis, using Wavelet Transform and Support Vector Machines," in Biomedical Engineering (ICBME), 2010 17th Iranian Conference of, pp. 1-4.

[9] T. Rocha, S. Paredes, P. Carvalho, J. Henriques, M. Harris, and J. Morais, "Ischemia detection in the context of a cardiovascular status assessment tool," in Engineering in Medicine and Biology Society, 2009. EMBC S2-1

[10] T. Rocha, S. Paredes, P. Carvalho, J. Henriques, M. Harris, J. Morais, and M. Antumes, "A lead dependent ischemicepisodes detection strategy using Hermite functions," Biomedical Signal Processing and Control, vol. 5, pp. 271-281.

[11] R. Silipo and C. Marchesi, "Neural for ST-T change detection,"Cardiology 1996, 1996, pp. 677-680.

[12] M. G. Tsipouras, C. Voglis, and D. I. Fotiadis, "A Framework for Fuzzy Expert System Creation— Application to Cardiovascular Diseases," Biomedical Engineering, IEEE Transactions on, vol. 54, pp. 2089-2105, 2007.

[13] D. Ghosh, B. L. Midya, C. Koley, and P. Purkait, "Wavelet Aided SVM Analysis of ECG Signals for Cardiac Abnormality Detection," in INDICON, 2005 Annual IEEE, 2005, pp. 9-13.

[14] S. Papadimitriou, S. Mavroudi, L. Vladutu, and A. Bezerianos, "Ischemia detection with a self-organizing map supplemented by supervised learning," Neural Networks, IEEE Transactions on, vol. 12, pp. 503-515, 2001

[15] H. Allende-Cid, A. Veloz, R. Salas, S. Chabert, and H. Allende, “Self-organizing neuro-fuzzy inference system,” in CIARP, 2008, pp. 429–436.

[16] A. De Gaetano, S. Panunzi, F. Rinaldi, A. Risi, and M. Sciandrone, "A patient adaptable ECG beat classifier based on neural networks," Applied Mathematics and Computation, vol. 213, pp. 243-249, 2009.

[17] Google.co.in. (2017). ecg signal with j point - Google Search. [online] Available at: https://www.google.co.in/search?dcr=0&biw=1366&bih=662&tbm=isch&sa=1&ei=wgcgWru0IcrTvgS3_KioCg&q=ecg+signal+with+j+point&oq=ecg+signal+with+j+point&gs_l=psyab.3...85058.93614.0.94793.13.10.0.3.3.0.492.1907.4j3j1j1j1.10.0....0...1c.1.64.psy-ab..0.6.534...0j0i30k1j0i24k1.0.s5Z3t8Hjcro#imgrc=F1Nt4lmGFFY2YM: [Accessed 30 Nov. 2017].

[18] Jalal A. Nasiri, Mahmoud Naghibzadeh, H. Sadoghi Yazdi , Bahram Naghibzadeh, “ECG Arrhythmia Classification with Support Vector Machines and GeneticAlgorithm” 2009 Third UKSim European Symposium on Computer Modeling and Simulation ,2009 IEEE.

[19] Juyoung Park, Kuyeon Lee, and Kyungtae Kang Arrhythmia Detection from Heartbeat Using-Nearest Neighbour classifier “International conformance of bioinformatics and iomadicien,2013 IEEE.

[20] Nurul Hikmah Kamaruddin**,** M.Murugappan**,** Mohammad Iqbal Oma, "Early Prediction of Cardiovascular Diseases Using ECG signal: Review” Student conformance on research and development 2012 IEEE.

[21] Fatin A. Elhaj, Naomie Salim, Taqwa Ahmed
Arief R. Harris, Tan Tian Swee, "Hybrid classification of Bayesian and Extreme Learning Machine for heartbeat classification of arrhythmia detection"2017 IEEE.

[22] Saleha Khatun; Bashir I. Morshed, **"**Detection of myocardial infarction and arrhythmia from single-lead ECGdata using Bagging Trees classifier**"** International Conference on Electro Information Technology (EIT) 2017 IEEE**.**

[23] Felipe Ramirez; Hector Allende-Cid; Alejandro
Veloz; HectorAllende"Neurofuzzybased Arrhythmia Classification Using Heart Rate Variability Features" International Conference of the Chilean Computer Science Society2010IEEE.