FEATURE EXTRACTION AND CLASSIFICATION OF SEGMENTED ECG SIGNALS BASED ON RADIAL BASIS FUNCTION AND RANDOM FOREST METHODOLOGY

RexyJ¹ and VelmaniP² Rajakumar T.C³

1Department of Computer Science, Manonmanium Sundaranar University

²Department of Computer Science, The M.D.T Hindu College

³Department of Computer Science, St. Xavier's College

ABSTRACT

Heart disease is the major cause of death ratio increase in this decade. Nowadays various people of different age sector undergo the high risk of heart problems and miss their precious life all of a sudden. Early detection of heart disease will save many people's life well in advance. Heart Diseases are predictable and they can be identified in earlier stage. First basic method to identify heart disease is ElectroCardioGram (ECG) which is the basic recording method of electrical activities of a functioning heart. ECG is the cheapest and painless method to detect the basic heart problems. This paper is an attempt to detect and classify heart beat signals which will serve as the basic step to predict basic and serious issues which may affect the functioning of the heart. The raw ECG signals are extracted and preprocessed to remove unwanted noises which will produce effective results. The preprocessed ECG signals are then are utilized to identify the heart beats which comprise of signals such as P,Q,R,S,T and U. After detecting the heart beats, they are segmented to extract the ECG Features. The temporal and spectral features are extracted from the segmented ECG signals for classification purpose. The extracted feature vectors are utilized to classify the signals. Radial Basis function and Random Forest method are commonly used classification methodologies; hence these two methodologies are applied to classify the ECG Signals into five basic classes. Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) Arrhythmia database and Noise Stress database are used for this implementation and the classes are identified based on the given dataset parameters. Performance metrics such as accuracy, specificity and sensitivity are computed to find out the best classification methodology among the applied two methodologies. This performance analysis provides a clear comparative view of both the existing methodologies and specifies that Radial Basis function well suits for the given segmented ECG signals and the extracted features. Hence this performance evaluation paves way for best classification algorithm selection or extension of the best methodology and it can be further optimized for better classification result. The implementation process has been carried out using Matlab software environment.

KEYWORDS

ECG, ECG Features, radial basis function, random forest method

1. INTRODUCTION

The human heart is a core organ that carries oxygen and other nutrients to all the tissues of the whole body. Electrocardiogram (ECG) analysis is the most common and basic cardiac examination, which is a useful to detect cardiac abnormalities. An electrocardiogram is a graphical record of bioelectrical signal generated by the human body during cardiac cycle [1]. Early diagnosis leads to prevention and creates way to save precious life. One complete ECG signal consists of P-QRS-T waves. The ECG signals may be recorded in different situations; there are possibilities for noisy signal. The noise in the ECG signals must be filtered prior heat beat detection, segmentation and classification. To improve the performance for better de-noising of ECG signal, the lower and upper level boundary values of Butterworth and Chebyshev Type I are concatenated [2]. The retrieved raw ECG signals are preprocessed using a novel hybrid methodology called ButterChev which is a combination of Butterworth and Chebyshev Type II filters. The heart beats must be detected and segmented before extracting the features from the filtered signals. The ECG signal must be segmented in terms of beats and its different waveforms (P wave, QRS complex and T wave) prior to the feature extraction stage [3]. The extracted features will serve as the basic source for classification. Hybrid methodology of Pan-Tompkins and Hamilton-Tompkins algorithm is used to identify the heart beats and segment the signals.[4]. Features are important properties and structural components of input patterns that adds more value to the given input pattern. Feature Extraction. Feature extraction is the method of gathering discriminative information from a set of samples [5]. Temporal, spectral and fiducial features are the main features to be extracted from the segmented ECG signals. This feature extraction process leads to better interpretation and classification process. The extracted features vector serves as the primary input for the signal classification process. In classification process the signals are classified according to the classes determined by the selected datasets. Classification is a supervised pattern of dividing the signals and variety of algorithms play a vital role in classifying the ECG signals based on the extracted features. As RDF and RF are commonly used classification methodologies both the algorithms are applied. The scope of this paper is to identify the best classifier methodology among RBF and RF. Performance metrics such as accuracy, sensitivity and specificity are calculated to evaluate the performance of both the methodologies. Fig.1 depicts the architectural diagram of the complete implementation.



Fig. 1 Architectural Diagram

2. RELATED WORK

ECG feature extraction and classification methodologies have been reviewed and analyzed. This section enlightens about various techniques proposed earlier and implemented for extracting ECG features and to classify the signals.

Prachi Lamba et al [6] reviewed various feature extraction and feature classification methods. Feature extraction methods such as DWT, PCA and LDA were reviewed. Classification methods such as SVM, RBF and KNN were reviewed and stated that DWT with combination of SM is best compared to other combinations.

Fatemeh Molaei Vaneghi at al [7] used Autoregressive (AR), Wavelet Transform (WT), Eigenvector, Fast Fourier Transform (FFT), Linear Prediction (LP), and Independent Component Analysis (ICA) methodologies and concluded that Eigenvector method gives better performance in frequency domain for the ECG feature extraction.

Wenliang Zhu et al [8] proposed to apply the SVM classification, after performing the combination of ECG morphology and segment feature extraction on the ECG signals. Morphological features and ECG segment features are extracted by using principal component analysis (PCA) and dynamic time warping (DTW). Support vector machine (SVM) is applied to the features and automatic diagnosis results are produced

Afseen Naaz et al [9] elucidated various techniques of extracting and selecting the vital features from the ECG signal in order to analyze the ECG signal automatically. It was found that the wavelet transform is one of the important tools for extracting out QRS complex and other features from the ECG signal.

Can Ye et al [10] proposed a new approach for heart-beat classification by combining morphological and dynamic features. Wavelet transform and independent component analysis (ICA)

are applied separately to each heartbeat to extract morphological features. After extracting the features, support vector machine classifier is utilized for the classification

K. Akilandeswari et al [11] used Walsh-Hadamard transformation (WHT) and Fast Fourier Transformation (FFT) to extract the required features from ECG signal. The extracted features are proposed for further classification.

Anoop Vylala et al [12] extracted spectral and statistical features from the wave components that yield the texture and the geometric nature of ECG. The classification is carried out using the Actor-Critic (AC) Neural Network that is trained using the Proposed Taylor-Sine Cosine Algorithm (Taylor-SCA).

Mohammed Belkheiri et al [13] implemented Radial Basis Function (RBF) artificial neural network to classify the preprocessed and ECG signals. ECG feature vectors composed of average amplitude values, DCT coefficients, DFT coefficients, and wavelet coefficients were used as input and it was found that an RBF which has one layer and the feature vector with 61 inputs, and 20 neurons produced the best performance

V.Mahesh et al [14] presents Random Forest for detecting cardiac arrhythmias. Heart Rate Variability (HRV) analysis and Discrete Wavelet Transform (DWT) are used to extract features from the ECG signals. The Random Forests classifier was provided with the combination of linear and nonlinear parameters derived as input. The proposed classifier showed satisfactory performances in discriminating the types of arrhythmia.

Emina Alickovic et al [15] proposed Random Forests (RF) classifier for ECG heartbeat signal classification. It is found that RF classifier achieves superior performances compared to other decision tree methods. It is elaborated that hybrid model with multi scale principal component analysis (MSPCA) de-noising, discrete wavelet transform(DWT) and RF classifier produces better performance with the area under the receiver operating characteristic(ROC) curve (AUC) and F-measure.

3. ECG SIGNAL PREPROCESSING

ECG Signals are the primary input for detecting the heart beats and the extract the features for classification purpose. A noisy ECG signal may not result in effective performance metrics; hence the noises are filtered from the given raw ECG signals. Several IIR filters such as Butterworth, Chebyshev Type I, Chebyshev Type II and Elliptic are commonly used for de-noising ECG signals to retrieve sharp ECG signal waves. To improve the performance for better de-noising and peak detection of ECG signal , the lower and upper level boundary values of Butterworth and Chebyshev Type I are concatenated.[2]. A hybrid method called ButterChev which is a combination version of Butterworth and Chebyshev Type I is used for effective filtering process. Fig.2 depicts the given raw ECG signal and filtered and peak identified signal.



Fig.2 Raw ECG Signal & Filtered Signal with Peak detection

4. HEARTBEAT DETECTION AND ECG SIGNAL SEGMENTATION

The preprocessed ECG signal must be segmented and the heart beat must be identified before the feature extraction process. To fully automate the heartbeat classification method, an automatic heartbeat detection module is required [16]. ECG Signal Segmentation is the basic step to be followed before extracting useful features from the ECG Signal. The ECG signal must be segmented in terms of beats and its different waveforms (P wave, QRS complex and T wave) prior to the feature extraction stage. The high level signals P,R,T,U and low level signal Q and S are segmented and the segmented signals makes way for efficient feature extraction and classification process. Fig. 3 presents the segmented ECG Signals view.



Fig. 3 Segmented ECG Signals View

5. ECG FEATURE EXTRACTION AND CLASSIFICATION

A. ECG Signal Features

Features of an ECG Signal are segments and intervals between fiducial points such as RR interval, amplitude of P,R and T waves [17]. The signal analysis provides adequate information about cardiac diseases and hence it serves as efficient input for extracting features from the ECG signals. ECG Feature Extraction

Feature Extraction is the process of retrieving the main properties of the given ECG signal which plays a vital role for classification. The main purpose of the feature extraction process is to select and retain relevant information from original signal.Cardiac diagnostic related attributes are extracted using Feature Extraction Methodologies. There are several Feature Extraction methodologies such as Independent Component Analysis (ICA), Eigen Vector, Principal Component Analysis, Wavelet Transform, Linear Prediction and Auto regression [18].Principal Component Analysis methodology is used to extract ECG features. Temporal, spectral and fiducial features are extracted which will be used for classification purpose. There are several Feature Extracting methodologies. Extracted features are divided into two subsets based on gender.

B. ECG Signal Classification

Classification is the supervised learning which helps to analyze the extracted features. MIT-BIH Arrhythmia Database and subset MIT-BIT Noise Stress Database from physionet are used which includes 48 ECG recordings [19].

The extracted features are classified into five standard ANSI/AAMI classes such as (a) normal (N), (b) supraventricular ectopic (S), (c) fusion (F), (d) ventricular ectopic (V), and (e) unknown (Q) [20].As the database contains both male and female ECG records ,the classification process is analyzed based on Gender category as a parameter. Neural network, machine learning and decision tree based classification algorithms are available. Machine Learning and decision tree based classification algorithms exists, among them Radial Basis Function and Random Forest Classifier Algorithm are used for classification.

1) **Radial Basis Function Classifier:** Radial Basis Function classifier is used in many applications such as classification, regression and signal processing. TheRBF classifier is composed of three layers called input layer, hidden layer, and output layer. The RBF networks are different from other types of neuralnetworks mainly in the hidden neurons. Each hiddenneuron has a radial basis function which is a center symmet-ric nonlinear function with local distribution. The radial basisfunction consists of a center position and a width parameter[21]

2) **Random Forest Classifier:** The RF algorithms form a family of classification methods that rely on the combination of several decision trees[22].Random forest classifier rely on ensemble method. Each tree in the RF splits into as class prediction and the class with more similarity will serve as the model for prediction and classification. The main idea of RF is to build many classification trees based on some randomly selected features from randomly selected samples with bagging strategy and then to use the trees to vote for a given input vector to get a class label. RF is constructed by many base learners and each base learner is an independent binary tree adopting recursive partitioning[23].Random Forest Classifier well suits for multi-class classification.

6. EXPERIMENTS AND RESULTS

Filtered ECG signals are received by applying ButerChev Filtering Method. Temporal, spectral and fiducial feature vector is generated with the help of PCA extracting methodology. Segmented signals are retrieved by applying hybrid version of Pan-Tompkins and Hamilton Tompkins algorithm. Having the retrieved feature vector as primary input the signals are classified using Radial basis function and Random Forest methodology. Performance metrics such as Accuracy, Specificity, and Sensitivity are calculated for measuring the performance of Radial Basis function and Random Forest methodology.

Where TP represents True Positive, TN represents True Negative, FP represents False Negative and FN represents False Negative,

TABLE I depicts the performance evaluation based on gender and overall dataset. The accuracy of the female ECG records based on RBF is 87.88 % and male ECG records is 86.04%. The overall accuracy rate of Radial Basis Function is 86.96%. The accuracy of the female ECG records based on Random Forest is 83.71 % and male ECG records is 81.52%. The overall accuracy rate of Random Forest methodology and Random Forest are 86.96% and 82.62% which reflects that RBF gives better accuracy than RF. The specificity of the female ECG records based on RBF is 94.49 % and male ECG records is 93.70%. The overall specificity rate of Radial Basis Function is 94.10%. The specificity of the female ECG records based on Random Forest are 94.10% and 94.36%. The overall specificity of RBF and RF differs slightly with .05% variation only. The sensitivity of the female ECG records based on Random Forest is 85.29 % and female records are 80.99%. The overall sensitivity of Random Forest methodology and Random Forest are 94.36% and 83.14%. Here the Radial Basis Function Performs well than Random Forest. So it clearly depicts that Radial Basis function well performs than Random Forest methodology.

Gender	Accuracy %		Specificity%		Sensitivity%	
	RBF	RF	RBF	RF	RBF	RF
Male	87.88	83.71	94.49	94.42	93.50	85.29
Female	86.04	81.52	93.70	94.31	94.49	80.99
Overall	86.96	82.62	94.10	94.36	94.00	83.14

T 11 '	1	•	C
Table 1:	gender	w1se	performance

Fig.4 represents the Accuracy performance metrics and clearly depicts that RBF performs more accurately than RF methodology.Fig.5 represents Specificity metrics and here RF methodology slightly out performs than RBF. Both of the methodologies reflect more or less same level of specificity.Fig.6 represents Sensitivity metrics and specifies that RBF well performs than RF.Fig.7 represents the overall performance metrics which clearly depicts that RBF well performs than RF.











Fig.6 Performance Metrics: Sensitivity



Fig.7 Overall Performance

7. CONCLUSION AND FUTURE WORK

This paper presents the ECG signal feature extraction and classification stage. This paper focuses on filtering the retrieved ECG signal using ButterChev Algorithm. After filtering the ECG signal heart beat is detected and the signals are segmented using hybrid approach of Pan& Tompkins and Hamilton & Tompkins methodology In this stage the high level signals P,R,T,U and low level signal Q and S are segmented. The segmented ECG signals are given as input for PCA Feature extraction methodology which extracts temporal, spectral and fiducial ECG features. This feature vector is applied to Radial Basis Function and Random Forest Methodology. The dataset is divided into two subsets for training and testing purpose. The signals are classified into five AAMI based classes. The accuracy, specificity and sensitivity of RBF and RF are 86.96%,94.1%,94%,82.62%,94.36% and 83.14%. This reveals that Radial Basis Function is more efficient that Random forest methodology. Further this signal classification can be extended by optimizing the classification process with the support of deep learning algorithms.

REFERENCES

- [1] N.Goldchlager,"Principles of clinical electrocardiography". Appleton & Lange, 13 th ed., Connecticut, USA, 1989.
- [2] J.Rexy, P.Velmani and T.C.Rajakumar,"Heart Beat Peak detection using signal filtering in ECG data ." International Journal of Advanced Technology and Engineering Exploration." (2019):12-24.
- [3] Rodolfo de Figueiredo Dalvi, Gabriel Tozatto Zago, Rodrigo Varejão Andreão, "Heartbeat classification system based on neural networks and dimensionality reduction.". (2016):318-326.
- [4] J.Rexy, P.Velmani and T.C.Rajakumar."A Novel Approach to Perform ECG Signal Identification and Segmentation Based On Pan- Tompkins And Hamilton-Tompkins Algorithm". International Journal of Pharmaceutical Research." (2021):Vol 13, Jssue 1.
- [5] P. Lamba and K. Rawal, "A survey of algorithms for feature extraction and feature classification methods," in Proc. Int. Conf. Autom., Comput. Technol. Manage. (ICACTM), Apr. 2019, pp. 338–341.
- [6] Prachi Lamba and Kirti Rawa,"A Survey of Algorithms for Feature Extraction and Feature Classification Methods", IEEE Xplore, 2019
- [7] Fatemeh Molaei Vaneghi, Maysam Oladazimi, F. Shiman, Afshan Kordi, M.J. Safari, F. Ibrahim," A Comparative Approach to ECG Feature Extraction Methods", 2 IEEE Computer Society,2012.
- [8] Wenliang Zhu, Xiaohe Chen, Yan Wang, Lirong Wang," Arrhythmia Recognition and Classification Using ECG Morphology and Segment Feature Analysis" ACM Transactions on Computational Biology and Bioinformatics, IEEE, 2018.

- [9] Afseen Naaz and Shikha Singh," Feature Extraction and Analysis of ECG signal for Cardiac Abnormalities-A Review", International Journal of Engineering Research & Technology (IJER, Vol. 3 Issue 11, November-2014.
- [10] Can Ye and Miguel Tavares Coimbra,"Heartbeat Classification Using Morphological and Dynamic Features of ECG Signals", IEEE Transactions On Biomedical Engineering, VOL. 59, NO. 10, October 2012.
- [11] K. Akilandeswari, R. Sathya, "Feature Extraction Of ECG Signals For Early Detection Of Heart Arrhythmia", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 12, December 2014.
- [12] Anoop Vylala ,Bipin Plakkottu Radhakrishnan," Spectral feature and optimization- based actor-critic neural network for arrhythmia classification using ECG signal" Journal Of Experimental & Theoretical Artificial Intelligence, 2019.
- [13] Mohammed Belkheiri, Zineb Douidi and Ahmed Belkheiri," ECG Beats Extraction and Classification Using Radial Basis Function Neural Networks", Proceedings of the Fourth International Conference on Signal and Image Processing 2012, Springer, 2013.
- [14] V.Mahesh, A.Kandaswamy, C.Vimal, B.Sathish," Random Forest Classifier Based ECG Arrhythmia Classification" International Journal of Healthcare Information Systems and Informatics, 5(2), 1-10, April-June 2010.
- [15] Emina Alickovic1 & Abdulhamit Subasi," Medical Decision Support System For Diagnosis Of Heartarrhythmia Using DWT And Random Forests Classifier" J Med System (2016), Springer.
- [16] Philip de Chazal, Richard B. Reilly," Automatic Classification of Heartbeats Using ECGMorphology and Heartbeat Interval Features". Ieee Transactions On Biomedical Engineering, Vol. 51, No. 7, July 2004.
- [17] Juie D. Peshave, Rajveer Shastr," Feature Extraction of ECG Signal", IEEE Advancing Technology for Humanity, 2014.
- [18] Sreelakshmi T G and Seethal Paul," An Accurate ECG Feature Extraction Method For Detecting Multiple Cardiovascular Diseases", International Journal of Innovative Science Engineering and Technology (IJISET), Vol 1,Issue 9.
- [19] Moody GB, Mark RG. "The impact of the MIT-BIH Arrhythmia Database". IEEE Eng in Med and Biol 20(3):45-50,2001.
- [20] N P. Joshi, Shweta A.Tambe1, P.S. Topannavar1,"Morphological & Dynamic Feature Based Heartbeat Classification" International Journal of Engineering Research and General Science Volume 2, Issue 3,April-May 2014
- [21] Jinna Lu, Hongping Hu, and Yanping Bai, "Radial Basis Function Neural Network Based on an Improved Exponential Decreasing Inertia Weight-Particle Swarm Optimization Algorithm for AQI Prediction" Hindawi Publishing Corporation Abstract and Applied Analysis Volume 2014, Article ID 178313
- [22] Nguyen C, Wang Y, Nguyen HN, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic". J Biomed Sci Eng [Internet]. 2013;6(5):551–60.
- [23] Taiyong Li and Min Zhou," ECG Classification Using Wavelet Packet Entropy and Random Forests", MDBI - Entropy 2016.

AUTHORS

RexyJ Department of Computer Science, Manonmanium Sundaranar University

VelmaniP Department of Computer Science, The M.D.T Hindu College

Rajakumar T.C Department of Computer Science, St. Xavier's College





