Probabilistic Feature Extraction Techniques for Electrocardiogram Signal-A Review

M. Ramkumar, C. Ganesh Babu, S. Karthikeyani, G.S. Priyanka, R. Sarath Kumar

Abstract. Extraction of features for an ECG signal plays a vital role in making the diagnosis of most of the diseases associated with cardiac muscle as well as the different states of arrhythmias. In this study, a literature review over the comprehensive manner has been done for the probabilistic feature extraction technique of an electrocardiogram signal in making the analysis of various classification methods of ECG arrhythmia signals that has been proposed over the past years of research. The ECG arrhythmia classification methods includes few digital signal processing techniques, Fuzzy Logic techniques, Hidden Markov Model, Support Vector Machines, Genetic Algorithm, Particle Swarm Optimization, Artificial Neural Networks, Transductive Transfer Learning etc. with which each individual approach exhibits its own disadvantages and advantages. For diagnosing the heart’s clinical condition, Echocardiogram is considered to be an essential tool but it consumes more time for processing and analyzing the ECG rhythms which possess huge number of heart beats. Therefore, it is in the urge that the system of diagnosis has to be made automated for classifying the clinical states of heart rhythms and heart beats in making the diagnosis very accurately and in the précised manner. For analyzing the electrocardiogram signal subsequently, the fundamental characteristics in terms of features like peak amplitudes, time intervals are very much required in determining the heart functions.

Keywords: Artificial Neural Networks, ECG, Support Vector Machine, Fuzzy Logic, Discrete Wavelet Transform.

1. INTRODUCTION

Electrocardiogram (ECG) signal is denoted as the time periodic signal which makes reflection on the cardiac activity. A huge information pathological and normal physiology of the cardiac muscle could be easily acquired from ECG signal. Inspite of ECG signal by nature itself being considered as the non-stationary signal its, very hard for visually analyzing those signals. Hence it is highly necessary for developing the computerized techniques in analyzing the ECG signal over its clinical states. Observing the ECG signals clinically might take huge hours and it could be very hard process because of its non-regularities in its time period and additionally analyzing in visual manner could also be not relied over it.
This type of factors creates the necessities for the development of techniques on the basis of computer technology which could be very precise and accurate. At each and every instant of heart beat, the cardiac muscle is allowed to depolarize for triggering its state of contraction. This transmission of electrical pulses all throughout the human body could be acquired through the skin which acts as the basic theory or the principle behind the Electrocardiogram. An Electrocardiogram instrument makes the record of these activities through the electrodes that has been placed over the surface of skin and makes its graphical representation of display. An ECG makes the involvement over the attachment of about ten electrical wires to the segment of body with which one to each individual limb and 6 numbers across the segment of chest. ECG is considered to be as the signal which makes the representation of the wave over the heart’s electrical activity such as depolarization of atrium and the ventricles, repolarization of the atrium and the ventricles. The ECG signal characterizes with the repeated complex cycles of waveforms along with certain approximations of frequency parameter with 1 Hz. One heart beat cycle of ECG wave component is consisting of the PQRST waveforms which denotes its peak. The major characterization of ECG signals are its amplitude peaks and its time intervals which will be considered as features in determining the clinical state of heart.

Cardiovascular disease holds the major cause of death mortality in all parts of the world and it is responsible for more than 17 million of death all over the world for a year [1]. Variations in the life-style such as reduction in the intake of cholesterol foods and periodic cycles of doing regular exercises can produce the chances for the reduction of dangerous events that has been related with cardiovascular diseases. Hence, detection in the early stage is a crucial step in preventing the death which is related with cardiovascular disease. A periodic visit by the patient to the clinical doctors inclusive of an ECG is an essential step towards the earlier detection which would result in maximum quantum of subject clinical information for which the scrutinization has to be made more carefully. Older methods of diagnosing and monitoring the variations in the ECG signal lies on making the detection of specific features over the signal by the peculiar observation of human [2]. QRS Complex is one of the most challenging features in ECG because of its external envelope and hence it is considered as the base process in performing the process of extracting the features.

![Figure 1. ECG Wave representation](image)

The development of the system for medical diagnosis on the basis of automated computerized process is made for assisting the clinicians in analyzing the huge volume of subject information. Utilization of different signal processing methodologies has been made by the process of feature extraction process from the clinical signals and determines the analysis of these features which possess their own demerits and merits. These methodologies functions by making the transformation for the maximal chances for determining the estimation for diagnosis in order to select the desired features that has to establish the classification process [3]. These techniques have been utilized for addressing these problems such as ECG signal analysis for detecting the variations in echocardiogram by utilizing the function of autocorrelation, feature analysis in frequency domain, wavelet transform and analysis over frequency-time. Few techniques that possess a line of band pass filters consisting of the frequency range which is approximately equal to the QRS complex though these techniques have very lower value of accuracy for making the analysis of features in ECG respect to the presence of noise and baseline drift.
that largely influence the ECG signal. In the recent works it could be observed that most of the classification studies of ECG signal arrhythmias is being carried out by dynamic systems in performing the experimentation and performing the analysis in non-linear manner [4]. The main aim of these type of analysis is to determine the estimated features which has been considered for classification of ECG signals under non-linear fashion will be used for clinicians in preceding their treatment based on the selected features under dynamic environment [5]. The proposal for many different methodologies has been made at the beginning of this study for analyzing the ECG classification by establishing the efficient feature extraction process. In this study, various techniques in performing the ECG feature extraction and classification has been proposed by considering certain criteria such as less complex and efficient system, high accuracy in predicting the clinical state by the proper extraction and selection of features.

2. ECG FEATURE EXTRACTION

Initially the method of feature extraction by utilizing the Wavelet Transform process and implying the classification by utilizing Support Vector Machines (SVM) has been made with acceptable value of accuracy in detecting the arrhythmia [1]. A novelty has been implied in presenting the new method of feature extraction for recognizing the reliable heart clinical state. The three essential stages were followed and they are Preprocessing of data, Extraction of features and ECG signal classification. Applications of two methods were mailed simultaneously for extracting the ECG signal features which provides the vector of features for the ECG signal. For extracting the features in terms of transform coefficients, the wavelet transform of ECG coefficients is being utilized. Subsequently, the application of autoregressive modelling is also made for the accumulation of temporal structures in ECG signal. Then as a terminal part, the Support Vector Machine (SVM) possessed with gaussian kernel is being utilized for classifying various rhythms of ECG heart beats. The simulated results have been acquired with the overall average accuracy of about 99%. The depolarization of cardiac root has been depicted in figure 2. The signal of depolarization then continues firstly to bundle of his with which it gets partitioned into two directions and moves through right and left branches of bundle [5]. The impulse moves through the bundle length although the septum of intra ventricles to the base of cardiac muscle with which the bundle gets partitioned into the system of purkinje [6]. The distribution of depolarization state is being made to the wall of ventricles and it gets initiated with the contraction of ventricles. The initial step in the feature extraction of ECG signal begins from the accurate R peak detection from the section of QRS Complex.

![Figure 2. Cardiac anatomy and physiological state representation](image)

Table 1. Electrophysiological State

| S No | Potential Action | Effect |
|------|------------------|--------|
| 1    | Action of Depolarization | Electrolyte shifting across the membrane of cell that produces variations in electric charge |
| 2    | Action of Repolarization | Restoration of internal negative charge and the return of cells to the state at rest |

Table 2. System of conduction and its associated functions

| S No | Name of the Structure | Function and Location |
|------|-----------------------|-----------------------|
| 1    | Sinoatrial Node (SA)  | Pacemaker with the dominant characteristics of cardiac |
A robust detector of R peak waveform utilizing the wavelets is being formulated [2]. The wavelets utilized for detection are symmetric and Daubechies. The acquisition of database has been obtained from MIT-BIH arrhythmia database and the analysis of signals from lead II has been made. The coefficient of detail (d4) is being selected on the basis of following essential parameters. They are frequency, energy and the analysis over cross-correlation for the ECG signal structure of decomposition. Totally, testing of 42 records of ECG signal is made for the R wave peak. The cumulative average accuracy as on whole utilizing sym11 and db6 are about 84% and 96% respectively. The essentialities of utilizing the wavelet transform has been enhanced with which the filtration of noise is being done at each decomposition level thus thereby providing the elimination for the process subjected o preprocessing. This quality ensures that this methodology is very robust. Herewith it is confirmed by utilizing the various database records in the presence of noise. The stability has been determined with the results obtained with db6 wavelets by creating the variation of threshold when compared to sym11 that picks up the false amplitudes [5]. The elimination of the zero-padding effect is also being done at the time of analyzing the energy in determining the algorithm to be simpler and more operated with minimum consumption of time.

The proposal for the new approach for extracting the ECG signal features has been made [4]. In that proposed study, the presentation of the algorithm has been made on the basis of wavelet transform for extracting the ECG signal features and recognizing the heartbeats which is abnormal in state. Since wavelet transform determines the localization over time-frequency domain, the development for the method has been made for selecting the mother wavelet which is optimal in its characteristic from the desired set of bi-orthogonal and orthogonal bank of wavelet filters. Selecting the function of wavelet is purely dependent on the capability of reconstructing the signal from the decomposition of wavelet for inhibiting the preservation of energy of transformation. The subsequent step of the approach is for the removal of noise from the ECG waveform by the hard or soft threshold provided with the 99% of limitation in enhancing the reconstruction ability and decomposition of PQRST into the vectors of coefficients by the function of wavelet with optimal characteristics [3]. The approximations and the coefficients of the terminal level scaling function, all the levels of details are being utilized for the analysis of ECG. Each cycle’s coefficients are being segmented into three different partitions with which it is being related to P-peak, T peak and QRS peak. The cumulative addition of these partitions determines the vectors for the features of one cycle. The testing of the algorithm is being made for two different ECG signals with which the initial is being acquired from biomedical MIT database and its decomposition is made into 4 levels and the denoising is induced by wavelet with optimal characteristics “sym4” possessed with 1.30 as threshold value globally. The resultant ECG waveform is acquired with abnormal state of heart beat [5]. The recording of later ECG signal is being obtained from the subject at the time of an epileptic seizure. The function of optimal wavelet is denoted as “coif5” with the threshold
value of 23.2. The three signals which is being possessed with the QRS complex makes the representation of depolarization of ventricles [4].

a. The correspondence of minute Q waves are subjected to the intraventricular depolarization of septum. The relationship of Q waves can also be made with the effect of breathing and are commonly very thin and small. They could also extract the characteristics of the waveform of an ancient myocardial infraction with which it is in such a way that it must be wide and maximum.

b. The reflection of R wave is being subjected to the essential mass of ventricle depolarization and hence it is represented as the biggest wave.

c. The significance of S wave denotes the terminal ventricle depolarization at the base of cardiac muscle.

Yet another proposal for extracting the features of ECG waveform by utilizing the Daubechies wavelet transform is being made [5]. The system for extracting the features of ECG signal is on the basis of wavelet transform in the multiresolution domain with which its evaluation and its development has been made. The signals of ECG that has been acquired from modified lead II were selected for further processing. For performing the detection better, the wavelet filter possessed with the function of scaling similar to the envelope of ECG waveform is being selected. The initial step of their technique is for denoising the ECG wave component by the removal of equivalent coefficients of wavelet at the higher order scaling region. Then the detection of QRS complex is made and each complex is utilized for tracing the peaks of each waves inclusive of offsets and onsets of the T peak and P peak which indicates its presence in single cardiac cycle. The results over the experimentation has proved that the approach over the proposed methodology for extracting the features of ECG signal has achieved with the positive predictivity and sensitivity of 98% and 99.1% respectively.

For detecting the complexities in the QRS waveform a presentation of the algorithm has been made. The QRS complex recognizes the determination of the origin for the maximal automated algorithm for analyzing the ECG signal. The consideration of feature in this algorithm for making the detection of QRS complex is the ECG signal slope. The maximal probability of transformation the baseline drift and the filtered component of the ECG wave component is utilized to mine the new variable feature of the slope. In this proposed algorithm, the procedure of filtering on the basis of sliding means determines the ECG signal with the characteristics of smooth and nil spikes that makes the approximation for extraction of slope features. The initial step is for extorting the feature of slope from the drift corrected and filtered ECG wave component. The feature extracted wave component is significantly made with the enhancement in the region of QRS complex and its suppression is created over the non-QRS complex portion. This presented study has proved to be the positive prediction and the rate of detection with 99.1% and 98.5% respectively.

The image classification technique over the classification of ECG images by the extraction of wavelet transform features and the neural network features. The extraction of features is being made from the decomposition of wavelets of the intensity for the images of ECG. The acquired features of ECG are being processed further by the utilization of artificial neural networks. The features are represented as follows.

- Median and Mean
- Minimum value and the Maximum value
- Mean absolute Deviation, variance and standard deviation

The training over the Artificial Neural Network has been made sixty-three essential ECG image features of various diseases. The result over the testing has been shown that the accuracy of classification over ANN has been resulted with 91%. The proposal over another algorithm for the for the extraction of features in the ECG wave component has been made [7]. This study has been focused on the evaluation of the performance over the classification automatically for detecting the abnonmal beats of ECG. The concept over the extraction of features is the novel and the robust technique. The acquisition for the sets of features has been made on the basis of morphological characteristics of ECG and the R-R interval. The adaptation for this configuration has been made by the self-organizing maps (SOM) for examining the clustering and the features of ECG signal. The development of classifier is made with learning vector quantization (LVQ) and SOM by utilizing the datasets acquired from the ECG records with which its recommendation is made by the standard of ANSI/AAMI EC-57. However, the proposed study enumerates the comparison of two classification strategies with which its annotation is made by the QRS complexes.

- Based on the original morphological features of ECG
- The new proposed approach on the basis of morphological features of preprocessed ECG signal.
The morphological filtering which has been carried out mathematically is being utilized for the ECG signal pre-processing [9]. The proposal for the alternate feature extraction technique is being made. Three models of ANN have been made and they are Multi-Layer Perceptron (MLP), Radial Basis Function (RBF) Neural Network and SOFM for which its consideration is made for classifying the ECG signals. 4 different sorts of ECG beats were recognized to be selected and its inclusive of Normal Sinus Rhythm (NSR), Atrial Premature Beat, Premature Ventricular Contraction and Left & Right bundle branch block.

Six various features such as average value of R peak, average power spectral density, QRS Complex area, signal energy Q-S interval and the value of autocorrelation is being acquired in terms of feature extraction for the characterization of four various sorts of heart beats. MLP provides the overall best performance in determining the accuracy.

The technique of feature extraction module by the technique of Discrete Wavelet Transform (DWT) is formulated for addressing the non-stationary ECG signal problem. Its derivation is being made from the single function of generation called as the parent wavelet i.e. the mother wavelet by the operation of dilation and translation. In the process of feature extraction, by utilizing the Discrete Wavelet Transform might lead to the resolution of optimal frequency in all the range of frequencies as it possesses the different size of windows which is broad at minimum frequencies and narrow at higher frequencies. The characterization of DWT will determine the features with high stability to the variations in the ECG morphology. The consideration over the non-linear characteristics of ECG signals were made and this theory has been moved forward with the behavioral study over the dynamic system from the series of experiments acquired from ECG signal [12]. The testing over this consideration is being made by utilizing the tools of non-linear dynamics similar to that of Lyapunov exponent computation. The classification of ECG beats is being made from the database which has been acquired from MIT-BIH arrhythmia Physionet recorded files. The computation over the ECG waveform Lyapunov exponent is being utilized as the inputs for the trained MLPNNs with the back-propagation network, delta-bar-delta in the extended form and the Levenberg-Marquardt algorithms.

Another algorithm which is represented as the slope vector waveform for establishing the QRS complex detection and the evaluation of RR interval is being discussed [13]. In this proposed study, the variable stage differentiation is being utilized for achieving the desired slope vectors for the extraction of features and the amplification in non-linear manner is utilized for acquiring the better SNR. This technique permits the accurate and speed search of R peak position, duration of QRS complex and the interval of RR that yields the best feature extraction results of ECG. For acquiring the duration of QRS complex the rules over the feature extraction is much needed.

3. CONCLUSION AND FUTURE ENHANCEMENT

ECG signal component plays an essential role for the diagnosis of different disorders of cardiac muscle and it is utilized for extracting the ECG signal features [15]. Many methodologies over the frequency domain and the time domain [16] is being utilized for the extraction of features which is consisting of its own limitations and advantages. The focus on the future work is made in such a way that the various techniques which is being utilized for the feature extraction process must lend high accuracy and it must possess the fast-responsive characteristics and easy implementation. This study lends the overview of different algorithms and techniques for the ECG wave feature extraction which has been proposed in the previous surveys of literature. The discussions over the advantages and disadvantages has been made [17][18]. The description over the requirements of techniques over the frequency domain has been made [19] and its not necessary to determine the analysis over entire features [20]. The development over the feature extraction algorithm and the technique [21] has been made for ECG which must be very high in its accuracy and must ensure the speedy extraction of ECG signal features.

References
[1] Xiaomin Xu, and Ying Liu,“ECG QRS Complex Detection Using Slope Vector Waveform (SVW) Algorithm,” Proceedings of the 26th Annual International Conference of the IEEE EMBS, pp. 3597-3600, 2004.
[2] Qibin Zhao, and Liqing Zhan,“ECG Feature Extraction and Classification Using Wavelet Transforim and Support Vector Machines,” International Conference on Neural Networks and Brain, ICNN&B, vol. 2, pp. 1089-1092, 2005.
[3] Ashley EA, Niebauer J., “Conquering the ECG,” London: Remedica; 2004.
[4] M.B. Tayel, and Mohamed E. El-Bouridy, “ECG Images Classification Using Feature Extraction Based On Wavelet Transformation And Neural Network,” ICGST, International Conference on AIML, June 2006.
[5] S. C. Saxena, A. Sharma, and S. C. Chaudhary, “Data compression and feature extraction of ECG signals,” International Journal of Systems Science, vol. 28, no. 5, pp. 483-498, 1997.

[6] E.D.Übeyli, “Detecting variabilities of ECG Signals by Lyapunov Exponents,” Neural computing and applications, vol.18, no. 7, pp. 653-662, 2009.

[7] F. Sufi, S. Mahmoud, I. Khalil, “A new ECG obfuscation method: A joint feature extraction & corruption approach,” International Conference on Information Technology and Applications in Biomedicine, pp. 334-337, May 2008.

[8] Chen, J.; Xie, Y.; Wang, K.; Wang, Z.H.; Lahoti, G.; Zhang, C.; Vannan, M.A.; Wang, B.; Qian, Z. Generative Invertible Networks (GIN): Pathophysiology-Interpretable Feature Mapping and Virtual Patient Generation. In Medical Image Computing and Computer Assisted Intervention—MICCAI 2018; Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 537–545.

[9] Dong, S.; Luo, G.; Wang, K.; Cao, S.; Mercado, A.; Shmuelovich, O.; Zhang, H.; Li, S. VoxelAtlasGAN: 3D Left Ventricle Segmentation on Echocardiography with Atlas Guided Generation and Voxel-to-Voxel Discrimination. In Medical Image Computing and Computer Assisted Intervention—MICCAI 2018; Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 622–629.

[10] Quan, T.M.; Nguyen-Duc, T.; Jeong, W. Compressed Sensing MRI Reconstruction Using a Generative Adversarial Network With a Cyclic Loss. IEEE Trans. Med. Imaging 2018, 37, 1488–1497, doi:10.1109/TMI.2018.2820120.

[11] Han, L.; Yin, Z. A Cascaded Refinement GAN for Phase Contrast Microscopy Image Super Resolution. In Medical Image Computing and Computer Assisted Intervention—MICCAI 2018; Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 347–355.

[12] Zhu, X.; Zheng, W.L.; Lu, B.L.; Chen, X.; Chen, S.; Wang, C. EOG-based drowsiness detection using convolutional neural networks. In Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China, 6–11 July 2014; pp. 128–134, doi:10.1109/IJCNN.2014.6889642.

[13] M.B. Tayel, and Mohamed E. El-Bouridy, “ECG Images Classification Using Feature Extraction Based On Wavelet Transformation And Neural Network,” ICOST, International Conference on AI&M, June 2006.

[14] Kar A, Das L, (2011),“A Technical Review on Statistical Feature Extraction of ECG signal”, IJCA Special Issue on “2nd National Conference- Computing, Communication, and Sensor Network” CCSN, 2011, pp.35-40, 2011.

[15] Kelwade J P, Salankar S S,(2015),“Prediction of Cardiac Arrhythmia using Artificial Neural Network”, International Journal of Computer Applications, ISSN:0975-8887, Vol.115, Issue No.20, pp.30-35, April 2015.

[16] Kohler B, Hennig C, Orglmeister R, (2002),” The Principles of Software QRS Detection Reviewing and Comparing Algorithms for Detecting this Important ECG Waveform”, IEEE Engineering in Medicine And Biology, pp: 42-57, January 2002.

[17] Kutlu Y, KuntalpD,(2012),“Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients”, Computer Methods and Programs in Biomedicine, Vol.105, Issue No.3, pp:257-267, 2012.

[18] Li Qiao, Rajagopalan C, Clifford G D,(2014),"Ventricular Fibrillation and Tachycardia Classification Using a Machine Learning Approach" IEEE Transactions On Biomedical Engineering, Vol. 61, No. 6, pp. 1607-1613, June 2014.

[19] Luz E J D S, Nunes T M, Albuquerque V H C D, Papa J P, Menotti D,(2013),“ECG arrhythmia classification based on optimum-path forest”, Expert Systems with Applications, Vol.40, Issue No.9, pp:3561–3573, 2013

[20] Malviya N, Rao T V K H,(2013),” De-Noiseing ECG Signals Using Adaptive Filtering Algorithms”, International Journal for Technological Research in Engineering, ISSN: 2347–4718, Vol.1, Issue No.1, pp. 75-79, September 2013.

[21] Ebrahimi A, Adedeji,(2015),”Classification of ECG Arrhythmias Using Adaptive Neuro-Fuzzy Inference System and Cuckoo Optimization Algorithm”, CRPASE, ISSN 2423-4591, Vol. 01, Issue No.04,pp.134-140, November 2015.