Received: June 16, 2018

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Classification of Cardiac Arrhythmia stages using Hybrid Features Extraction with K-Nearest Neighbour classifier of ECG Signals

Vedavathi Gauribidanur Rangappa [1]* **Sahani Venkata Appala Varaprasad Prasad [1]**
Alok Agarwal [1]

[1]Lingaya's Vidyapeeth, India
* Corresponding author's Email: vedavathi.gr@gmail.com

Abstract: The catastrophic heart diseases such as Myocardial Infarction (MI), Heart Failure (HF) and Ischemic Heart Disease (IHD) is a chain process leads to Coronary Artery Disease (CAD). The analysis of CAD from Electrocardiogram (ECG) signals by manual techniques are quite difficult. Therefore, there is need of techniques without human interaction for classifying the CAD should be improved. This work presented the recognition of five types of ECG beats by using a three-step system. In the first step, Pan-Tompkins algorithm (PTA) is used for detecting the peaks in ECG signals. The second step includes extraction of three interval features combined with ECG higher order statistics. In the third step for classifying ECG beats K-Nearest Neighbour (KNN) technique is employed. This approach analysed the heartbeats as normal or abnormal as accurately, and the experiments were conducted on MIT/BIH arrhythmia database for classifying the ECG signals. The results stated that the accuracy of the proposed approach is up to 98.40% for segregating the signals.

Keywords: Coronary artery disease, Myocardial infarction, Electrocardiogram, Pan-Tompkins algorithm, K-nearest neighbour, MIT/BIH arrhythmia database.

1. Introduction

The life draining cardiac abnormalities causing sudden cardiac arrest are heart muscle disease (cardiomyopathy and hypertrophy), Bundle Branch Block (BBB) and MI [1]. The abnormalities of these cardiac diseases are detected by an automated process that can lead to a challenging problem in ECG. BBB is the delay in the conduction process of the heart, whereas, MI is a CAD and the pathological characteristics are the appearance of abnormal Q wave, the inversion in T-wave and the ST-segment elevation [2]. Due to the formation of cholesterol plaque on the artery wall or contraction of the whole wall for other reasons, decreasing the diameter of the arteries that leads to some pathological conditions called CAD [3]. The ECG is a non-invasive method of measuring the electrical properties of the heart. ECG become an important method for diagnosing the heart disease which provides more information about the physiological state of the heart [4]. The cardiovascular diseases can be detected by analysing computer-aided ECG results by assisting physicians. The most of the existing arrhythmia detection system considered linear data of ECG signal. The experiments on non-linear data represent the ECG signals as more sensitive, achieve good performance under noisy conditions and extracting the hidden information from ECG signals [5].

In the field of biomedical signal processing, the different abnormalities of ECG are investigated by researchers by using detection and classification methods. The process of transformation of original features into novel features which are a weighted combination of original features is defined as Feature Extraction (FE) [6]. A typical FE method is Higher Order Statistical/ Spectral (HOS), spectral features and temporal features. The filter and wrappers are the two categories of feature selection algorithms. In wrapper methods, the subset selection procedure depends upon the learning algorithm that is used to train the model itself, whereas, in filter method, the learning process is independent [7]. The different parameters and insight are offered by spectral domain provide a distinctive representation of signals that can be used for diagnosis [8]. The conventional lower (first and second) order statistics are well-known in the field of bio-signal processing. However, for non-linear signals, the lower order statistics are not sufficient for a proper representation [9]. The temporal features describe the next higher order statistics which is known as skewness and kurtosis. The extracted features are classified with the help of Support Vector Machine (SVM) classifier. According to the structural risk minimization principle, SVM can construct an Optimal Separating Hyperplane (OSH) in the feature space [10]. The classification of both the unseen and training sample with minimum risk in the test set by OSH. The non-linear correlation method is more effective which is proved by the results by classifying the ECG signals.

In this work, the identification of irregular shape of the signal is carried by FE and classify it with the help of KNN. The features that are going to extract are higher order statistics and temporal features which are taken from the ECG signal. These features increased the classification accuracy by providing more value of the ECG signal. These are classified by KNN classifier and compared with other existing techniques. The accuracy, sensitivity, specificity, and f-measure of the classifier are shown in the experimental result. The KNN classifier accuracy was achieved up to 98.40% which is higher than the existing systems.

This paper is composed as follows. Section 2 presents a broad survey of several recent papers on arrhythmia identification and classification. In section 3, classification methodology (KNN) is presented with an effective linear and non-linear features. In Section 4, comparative study of proposed and existing methodology is presented. The conclusion is made in Section 5.

2. Literature review

The arrhythmia identification and classification reviewed by several techniques suggested by many researchers. In this scenario, brief evaluations of some important contributions to the existing techniques are presented below.

S. S. Qurraie and R. G. Afkhami, [11] proposed an algorithm based on time-frequency representation to extract features for cardiac arrhythmia classification. The time-frequency plane was split into the frequency bandwidth and the time duration of ECG segments and peaks into nine windows. The windows were selected as pseudo-energy features such as RR-interval and HOS and the subject-oriented scheme with decision tree was also used in classification, the test data included samples from different subjects. The subject-oriented method guaranteed the classification authenticity and avoided the issues of overfitting. The method showed extremely high accuracy in the heartbeat classification. Average over any finite time duration was absent, so the Wigner-Ville distribution had an infinite resolution in frequency and time domains.

F. A. Elhaj, N. Salim, A.R. Harris, T.T. Swee, and T. Ahmed [12] investigated the representation of the ability of features such as non-linear and linear features and proposed a combination of features by improving the ECG data classification. The method analysed the different classes of the heartbeat of arrhythmia namely supra-ventricular ectopic beats, fusion beats, unclassifiable and paced beats, ventricular ectopic beats, and non-ectopic beats. The non-linear features such as cumulants and HOS, whereas non-linear feature reduction methods were combined with linear features were analysed. The features were tested for their ability using different classifiers such as SVM and neural network methods with tenfold cross-validation. The experiments on the method showed that they provided equal average accuracy, sensitivity, and specificity. The FE method did not include the symmetry and reflection properties and didn't follow the superposition principle.

U. R. Acharya, V.K. Sudarshan, J.E. Koh, R.J. Martis, J.H. Tan, S.L. Oh, and C.K. Chua [13] presented a methodology for the automated identification of CAD by applying non-linear HOS FE methods by using ECG signals. The FE such as cumulant and HOS bispectrum were extracted from each ECG beats. The extracted features were applied to Principal Component Analysis (PCA) dimension reduction technique and then PCA coefficients were ranked by using several methods. The ranked features subjects to different classifiers such as Decision Tree (DT) and KNN classifiers for obtaining the highest classification performance. An integrated index using CAD Index (CADI) was formulated and developed by the method for characterization of ECG signals as normal with CAD conditions using a single number. The non-linear method was unable to identify the vast amount of blockage in artery that leads to MI.

U. Desai, C. Gurudas Nayak, and G. Seshikala [14] presented an approach for diagnosing MI conditions using ECG beats as automatic. The method addressed the importance of ECG in higher-dimensional visualization plots symbolize the Normal Sinus Rhythm (NSR), management of MI risk factors, and MI classes. The non-linearity in ECG was finely captured using HOS and then, dimensionality was reduced. The performance of extracted features was measured by using different classifiers through ADaBoost (ADB), BaGGing (BGG), Rotation Forest (ROF), and Random Forest (RAF) in which DT was used as the major classifier. The method revealed that ROF achieved higher accuracy among them to identify the MI and NSR conditions. The disadvantage of the method was the abnormalities and the risk stages of MI were diagnosed by using a small dataset.

S. S. Kumar and H. H. Inbarani, [15] presented a Multi-Granulation Rough Set (MGRS) based classification approaches that were applied to explore the process of making the decision. The method used the process of FE such as Pan Tomkins (PT) and Wavelet Transform (WT) extraction method. The PT method extracted the morphological features of peak intervals that were used for determining the heart rate. The algorithm searched for the local maximum or minimum peak near the heart beat label to establish local points. The WT FE approach used for extraction of wavelet coefficients were applied to ECG signal classification. The experiments were conducted and the results proved that the method performed higher accuracies compared to other well-known techniques. The computational time was larger in the method while comparing with the existing methods.

Raghu Nanjundegowda, and Vaibhav Aniruddha Meshram, [16] implemented the Hybrid features of T-wave in ECG for arrhythmia classification. The system had three major phases such as windowing technique, extracting the features and classification. Several features were extracted namely Peak-Magnitude RMS ratio, Auto Regressive feature based Yule Walker, Burgs method and Differential Entropy (DE). The classifier called Deep Neural Network (DNN) was used in classification phase. By using this method, signals were classified as normal or abnormal. The experimental results proved that the hybrid features Arrhythmia classification provided 98.3% accuracy. The classification method consumes more time for classifying the signals, because this method presented the accuracy in 100th iteration only.

An effective combination of non-linear and linear features is implemented with an appropriate binary classifier: KNN for enhancing the performance of arrhythmia identification and classification to overcome the above-mentioned issues.

3. Proposed methodology

The KNN method is designed to classify the ECG signals as normal or abnormal signals. The novelty of this method is taking into account the arrhythmia classification data. The proposed automated arrhythmia prediction system contains the following procedures: segmentation, FE, and classification. Fig.1 shows the basic structure of the proposed system. Each section of the proposed work is briefly explained in the following sections.

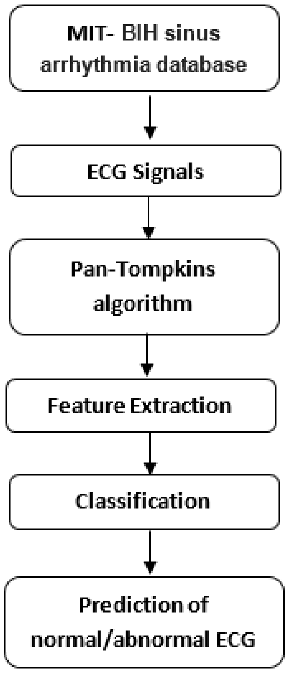

Figure. 1 The general architecture of the proposed method

3.1 Materials and methods

In the initial stage of the arrhythmia prediction system, an acquisition of an ECG signal helps to detect heart rate, disease, and other essential information of heart. There are huge amount of datasets for providing the information regarding heart such as NSR dataset, Personal Health Record (PHR) dataset, MIT-BIH sinus arrhythmia etc. In the proposed method, the identification of normal and abnormal signals is taken as input from the MIT-BIH sinus arrhythmia database. Twenty-three records were randomly taken from a set of 4,000 twenty-four hours ambulatory ECG data gathered from individuals including the two inpatients (around 60%) and outpatients (around 40%) at Boston's Beth Israel Hospital. Remaining 25 records were chosen from a similar set in order to incorporate clinically significant arrhythmias. The ECG records are examined at 360 Hz. Per channel includes 11-bit resolution more than 10 mV extend. A brief description of the various kinds of ECG beats is explained below.

3.1.1. Normal beat

The normal ECG beat comprises peaks: P, QRS complex and T wave. The PR peaks range between 110 ms and 200 ms and the heart rate interval range between 55 and 105 beats per minute.

3.1.2. RBBB (right bundle branch block)

The QRS complex of the ECG signal demonstrates an additional diversion, which indicates the slower depolarization of the right ventricle that follows the quick depolarisation of the left ventricle.

3.1.3. LBBB (left bundle branch block)

The compression of the left ventricle is due to the delay activation of left ventricle rather than the right ventricles. The length of the QRS complex exceeds 120 ms.

3.1.4. APC (atrail premature contraction beat)

It is described by early heartbeats starting in the atria. The sinoatrial node directs the pulse amid normal sinus rhythm, APCs happen the sinoatrial node is slower than the other region of the atria for depolarization and it triggers a premature beat.

3.1.5. PVC (premature ventricular contraction beat)

Issue exists outside the SA node, a QRS complex is extended, not related to the following P wave and T wave is inverted. The acquired data set is utilized for non-linear and linear FE.

3.2 Pan-Tompkins algorithm

The MIT-BIH arrhythmia database provides the raw ECG signals that are processed and this step includes the elimination of noises present in the signals. The specific bands are obtained by moderating the low spectral frequencies of signals which are high and checked that have improved quality of frequency ranges. The signals are also influenced by some of the noises like muscle noise, artifacts due to movement, power-line interference, baseline wandering, and T waves with higher frequency characteristics similar to QRS complexes. Here, PTA is used for findings the peaks by using filtering and pre-processing techniques. Moreover, the architecture of the proposed algorithm is given below Fig. 2. Generally, the process of overall detection process can be segregated into four stages such as Filtering, Derivative, Squaring and R-Peak.

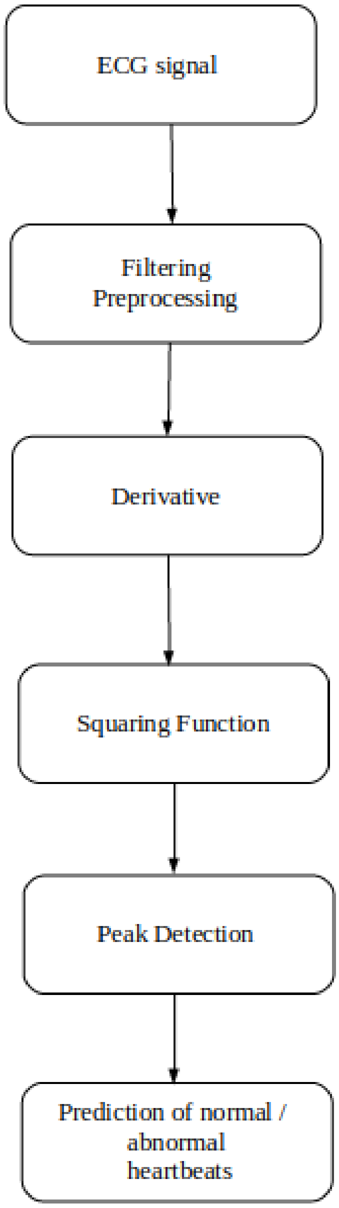

Figure. 2 Basic Model of Pan-Tompkins architecture for QRS detection

3.2.1. Filtering

When the raw ECG signal enters the system, the first step is pre-process the bandpass filter for reducing the influence of baseline wander, muscle noise, 60Hz interference, and interference in T- wave. The desirable range of pass-band 4-15 Hz is used for maximizing the QRS complex energy. The low-pass and high-pass filters are cascaded to form Band-pass filter where the low-frequency interference and higher frequency artifacts removed automatically. Thus bandpass allows particular frequencies to analyse the nature of a QRS complex while both frequency signals such as high and low frequencies get attenuated.

3.2.2. Derivative

To acquire the slope of QRS complex the noise-free ECG signal is given to derivative block. The amplitude threshold is applied to the signal in order to cut horizontally the ECG signal to reduce influence of the P and T waves compared to the R wave. The random signal distortion is present in the real-world signal will cause many false peak detections with higher peak distortion, and also cause maximum peak width and height of the wave. This can be avoided by, smoothing the derivative output of the signal. Smoothing technique determines the position of the wave appearance, amplitude of the peak, and width of each peak with the particular interval of time.

3.2.3. Squaring function

After derivative, the output signal is squared sequentially. The data points are positive, then the derivative output is amplified non-linearly, and the QRS complexes are emphasized as eq. 1.

$$y(nT) = [x(nT)]^2 \tag{1}$$

where, *x(nT)* = Input ECG signal
y(nT) = Squared input signal

The higher frequencies in the signal are shown by the squaring function, which is mainly because of the QRS complex.

3.2.4. Peak detection

A common purpose for processing the signal is used for identifying the peaks in a signal and to determine their positions, heights, and width. The process detects the peak in the signal and comes after extracting the ECG features that contain QRS complex that is more emphasized compared to the noise. A peak is determined when the signal changes direction within the certain time interval. It must able to detect a large number of different QRS morphologies to be clinically useful and must be able to follow sudden changes and gradual changes. The rising edge point of the integrated ECG Squared waveform is marked by the temporal location of QRS complex. In the final step, two thresholds are varied according to the detection reliability. The value with a higher threshold is marked as the peak of the signal. Through the peak prediction algorithm, the peak is detected which is more effective than the amplitude thresholding technique.

Three steps to detect a QRS peak are

Step 1: Select the value and store it to temp local maximum

Step 2: If newer point value is greater than temp local maximum, then store it to temp local maximum.

Step 3: If newer point value is smaller than half of temp local max, a peak is identified and detected. Otherwise, go to Step 2.

Then, peak is distinguished as QRS complex or noise, and the signal is saved for later diagnosis. PTA Neglect all larger peaks by a refractory blanking of less than 200ms to ignore T waves and multiple detections of QRS complex waves. If the peak is greater than the adaptive detection threshold then it is identified as a QRS complex instead of noise.

$$DT = NPL + TC \times (QRSPL - NPL) \tag{2}$$

Where,
DT represents the threshold of detection,
TC is the thresholding coefficient
NPL is the level of noise peak and
QRSPL is the QRS complex peak level.

To detect the true peak in ECG signal and to neglect the peaks which are too wider, too smaller or too narrow, the parameter such as slope width and threshold amplitude helps the process. This technique is used for measuring positions of peak widths and heights accurately.

The peak detection by applying the PTA for both stages of signals (i.e. normal and abnormal) are represented in the following Figs. 3 and 4.

3.3 Feature extraction

The extraction and strengthening the characteristic features of the signal is the main aim of this stage and therefore, the efficiency of recognition of these features increased in the arrhythmia classification. Higher Order Statistics (HOS), Temporal features and Min-Max features are extracted due to periodic nature of ECG signals. The features extractions are briefly explained in the below sections.

3.3.1. Higher order statistics

The HOS or cumulant is one of the robust methods applied for the non-linear signal analysis. The first two order statistics have achieved great importance in bio-signal processing field, which is not sufficient for representing non-linear signals. Therefore, next two order statistics are used in this analysis. The expectations over the process multiplied by lagged versions as the calculation of the n^{th} order moment as:

Received: June 16, 2018

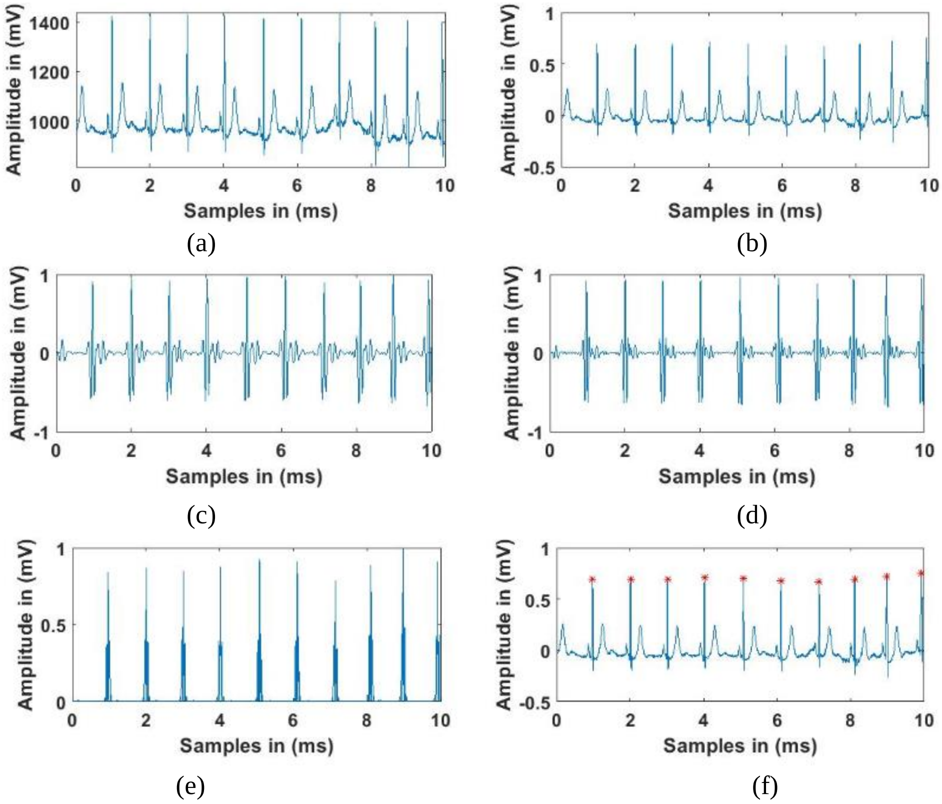

Figure. 3 (a) Input signal, (b) Normalization, (c) Band pass filtered, (d) Filtered with derivative filter, (e) Squared, and (f) R-Peak detection.

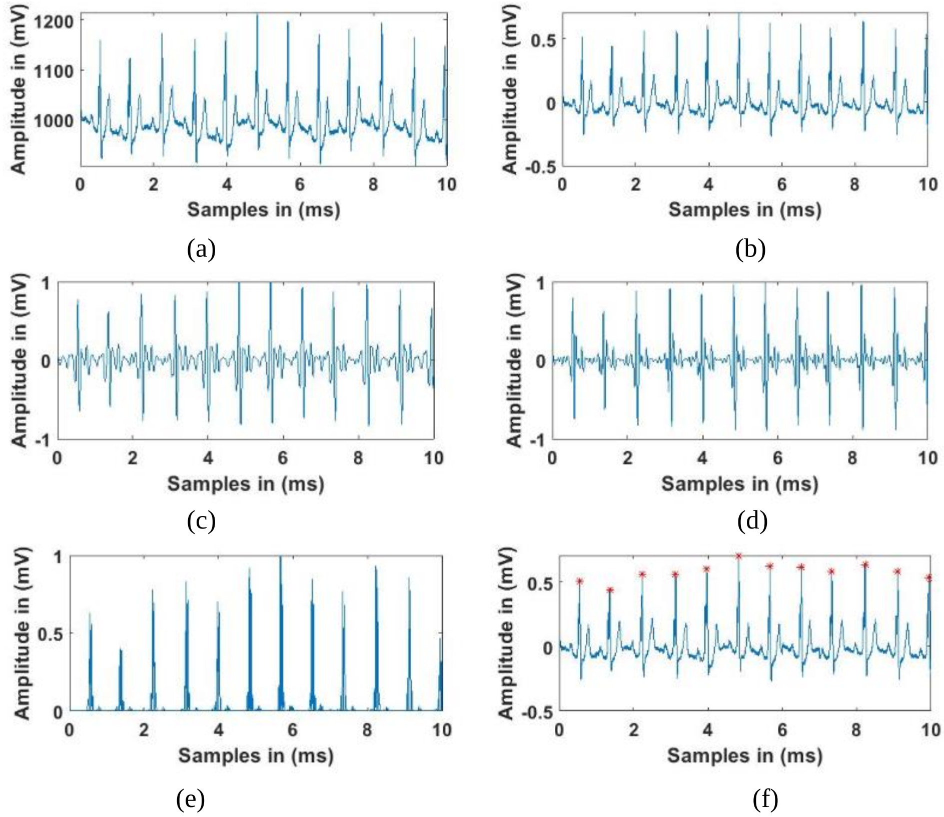

Figure. 4 (a) Input signal, (b) Normalization, (c) Band pass filtered, (d) Filtered with derivative filter, (e) Squared, (f) R-Peak detection

 DOI: 10.22266/ijies2018.1231.03

$$m_1^X = E[X(n)] \tag{3}$$

$$m_2^X(i) = E[X(n)X(n+i)] \tag{4}$$

$$m_3^X(i,j) = E[X(n)X(n+i)X(n+j)] \tag{5}$$

$$m_4^X(i,j,k) = E[X(n)X(n+i)X(n+j)X(n+k)] \tag{6}$$

Where, $m_1^X, m_2^X, m_3^X, m_4^X$ are the first four moments. The cumulants can be examined as non-linear combinations as:

$$C_1^X = m_1^X \tag{7}$$

$$C_2^X = m_2^X(i) \tag{8}$$

$$C_3^X = m_3^X(i,j) \tag{9}$$

$$C_4^X = m_4^X(i,j,k) - m_2^X(i)m_2^X(j-k) - m_2^X(k-i) - m_2^X(k)m_2^X(i-j) \tag{10}$$

where $C_1^X, C_2^X, C_3^X, C_4^X$ are termed as cumulants orders. Here, *X(n)* is considered as a zero mean process. The cumulant features and ECG waveform are extracted by using the R wave as found in the annotation of the database which is chosen from a window size nearly −*300 ms to +400 ms.*

3.3.2. Temporal features

FE has a significant role in classification. In this paper, the sequence of both statistical and temporal features has been used. The relative spread of the ECG characteristics is reduced by the cumulant characterization of complexes in QRS features which belongs to the same type of heart rhythm and this method makes the classification easier.

The difference in time between the two consecutive peaks is evaluated as temporal parameters such as the QRS intervals. Temporal features are the QRS interval between the current beat and the previous beat (QRS1), between the previous beat and the one before it (QRS0), between the current beat and the next beat (QRS2), the ratio of QRS1 and QRS2, the ratio of QRS1 and QRS0 (which is an indication of an abnormal timing sequence and helps in identifying an abnormal beat), T-wave maximum value and PQRST-minimum value. These features are extracted for each individual beat in the database.

3.3.3. Min max features

In FE phase, statistical features are used for extracting the features from ECG signal, a signal is decomposed to determine its features. The properties of both non-statistical and statistical components such as energy, mean, median, entropy, standard deviation, skewness, kurtosis, covariance are calculated from the coefficients which are generated after decomposition to create the feature set. Here, minimum and maximum voltage features can be obtained from the ECG signal. In different normal and abnormal signals, the peak voltages are different. So, this features able to recognize the classification of signals as abnormal or normal.

3.4 Classification

After the extracting the features of pre-processed data, the ECG data are used for the prediction of the abnormal and normal signal. The classification uses KNN, which minimize the noise between the actual and predicted outcomes. This usually involves neural layers learning on huge dataset.

3.4.1. K-nearest neighbour (KNN) classifier

The classifier KNN is a supervised method having a certain computational speed along with classification accuracy. According to mathematics and simple theory, the classifier is developed, whereas the training stage does not require for the KNN-based classifier. Compared to the Artificial Neural Networks (ANN) and Support Vector Machine (SVM) classifiers, the KNN classification structure contains lower computational burden. Consider the pair $(x_i, \delta(x_i))$ contains the feature vector x_i and its corresponding label $\delta(x_i)$ is needed for formulating the KNN classification algorithm, where $\delta \epsilon \{1,2,\ldots,n\}andi = 1,2,\ldots.,N$ (*n and N* are the number of classes and the number of train feature vectors, respectively). For an arbitrary feature vector x_i, calculation of a defined distance between this feature and the vector $x_j$is possible as follows,

$$d(i,j) = f(x_i, x_j) \tag{11}$$

Where $f(x_i, x_j)$is a scalar distance function. For instance, $f(x_i x_j)$can be defined as

$$f(x_i, x_j) = (x_i, x_j)^T \Sigma, (x_i, x_j) \tag{12}$$

$$f(x_i, x_j) = \left(\sum_{k=1}^{p} \left(x_i(k) - x_j(k)\right)^r\right)^{1/r} \tag{13}$$

$$f(x_i, x_j) = \frac{1}{p}\sum_{k=1}^{p} abso\left(x_i(k) - x_j(k)\right) \quad (14)$$

Where the Eq. (12) called generalized distance, the famous Euclidean norm is achieved by using weight matrix. While the Eq. (13) describes the Minkovski distance of degree *r* and again the Euclidean distance appears for *r =2*. The Eq. (14), is called the City Block distance and is used in many pattern recognition cases. The following equation describes the distance vector *D(i)*,

$$D(i) = \{d(i,j) \vee i = 1,2,\ldots.N_{test}, j = 1,2,\ldots.,N_{train}\} \quad (15)$$

The D(i) vector is sorted in an ascending order, and choose the first K elements (which is called K nearest neighbors) as follows

$$D_N(i) = \underset{Ascending}{sort}\left(D(i)\right) \quad (16)$$

$$V = \left\{\delta\left(D_N(i)(1)\right), \ldots., \left(\delta D_N(i)(K)\right)\right\} \quad (17)$$

According to the KNN algorithm, the test feature x_i belongs to the class with the major votes in the K-nearest vote vector *V*. In order to determine the optimum *K* corresponding to the best accuracy, a simple way is to alter the *K* from *1* to a large enough value and choose the *K* for which the best accuracy is obtained for all test features.

4 Experimental result

The Classification was done in the MATLAB (version 8.0) with a system requirement of 2 GB of RAM and 2 GHz of Intel dual core processor. The MIT-BIH arrhythmia database is used for evaluating the performance of proposed method which contains the ECG signals. The label file consists of both non-linear and linear data which contains the 18 normal and 18 abnormal signals. In this paper, the abnormal signals consist of 5 windows for each signal depends upon the above five stages in which 20 signals are used for training and 5 signals for testing for the analysis of the method.

4.1 Performance evaluation

The ECG signals were extracted from proposed combined features of HOS and temporal features. The KNN classifier decides whether the signals are normal (SNR) or abnormal (Arrhythmia). In this experimental analysis, the proposed KNN classifier performance was compared with the other existing classifiers such as Navie Bayes (Navy), Random Forest (RF), Neural Network (NN), Support Vector Machine (SVM), in terms of Accuracy, Specificity, Sensitivity, and F-measure. The estimation has been done for these parameters using True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The calculation of parameters is described below.

4.1.1. Accuracy

Accuracy is the most important performance measures and is able to predict the correct observations from the total observations. The accuracy is directly proportional to true results by considering both TNs and TPs from the total number of cases analyzed. The parameter of accuracy is calculated in Eq. (18),

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (18)$$

4.1.2. Specificity and sensitivity

The measurement of the proportion of negatives that are correctly identified is described by Specificity which is described in Eq. (19).

$$Specificity = \frac{TN}{TN+FP} \times 100 \quad (19)$$

The sensitivity calculates the ratio of positives that appropriately recognize signals and a mathematical equation of sensitivity is described in Eq. (20).

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \quad (20)$$

4.1.3. F-measures

In addition, the f-measure is the suitable evaluation metrics for finding the effectiveness of abnormality and normality of ECG signals. These measures are statistical variability and a representation of random errors. The general formula of f-measure for determining the normality and abnormality is given in the Eq. (21).

$$F - measure = \frac{2TP}{(2TP+FP+FN)} \times 100 \quad (21)$$

In the experimental analysis, the signals of ECG are classified based on the prediction of normal and abnormal signals. In below section, the Arrhythmia disease signal and Normal signals are shown below in Fig. 5:

Received: June 16, 2018

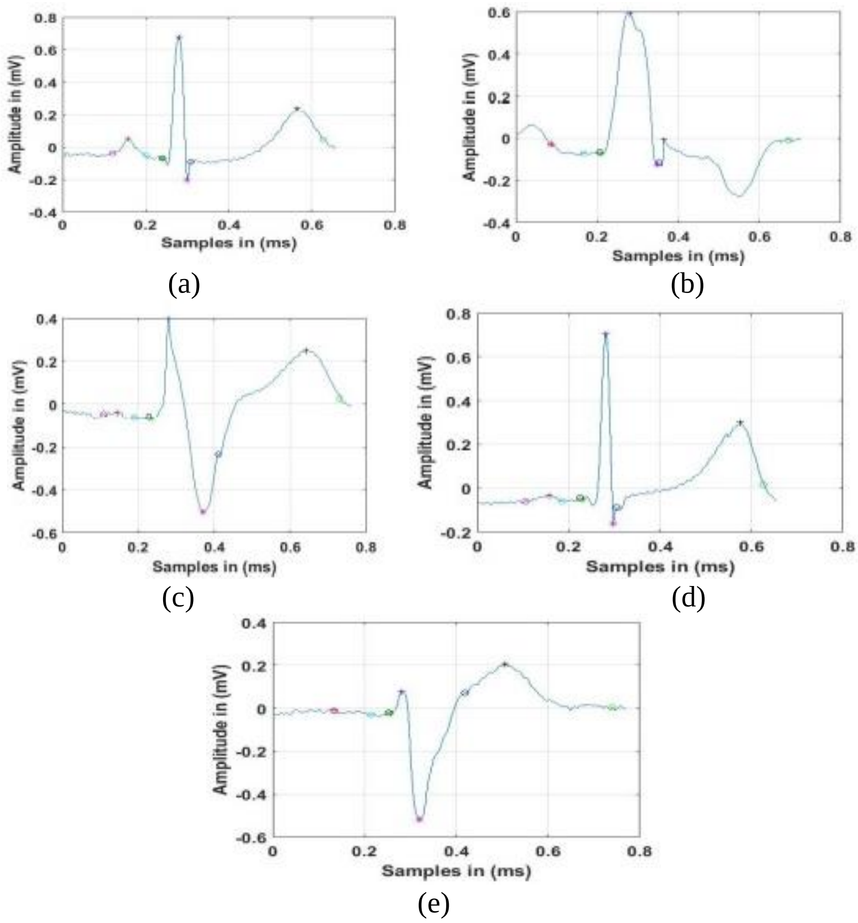

Figure. 5 ECG signals of Normal and Arrhythmia disease signal: (a) Normal signal, (b) LBBB signal, (c) RBBB signal, (d) APC signal, and (e) PVC signal

Table 1. Performance of the proposed classifier with existing classifier

Classifiers	Features	Specificity(%)	Sensitivity(%)	F-Measure(%)	Accuracy(%)
NN[17]	Min	2.00	95.00	2.00	19.20
	HOS	12.00	93.00	6.00	21.60
	T-Peak	4.00	99.00	4.00	20.00
	QRS Time	4.00	95.00	2.00	20.80
	Hybrid	8.00	91.00	5.33	18.40
SVM[17]	Min	8.00	100.00	50.67	53.60
	HOS	100.00	61.00	34.00	53.60
	T-Peak	100.00	85.00	100.00	47.20
	QRS Time	16.00	90.00	42.00	54.40
	Hybrid	100.00	100.00	97.33	77.60
RF[17]	Min	56.00	93.00	68.00	77.60
	HOS	44.00	76.00	44.67	50.40
	T-Peak	100.00	100.00	100.00	63.20
	QRS Time	48.00	91.00	46.00	69.60
	Hybrid	100.00	98.00	97.33	92.80
	Min	72.00	93.00	75.20	75.20

Navy[17]	HOS	52.00	80.00	53.60	53.60
	T-Peak	100.00	100.00	65.60	65.60
	QRS Time	64.00	77.00	60.00	60.00
	Hybrid	100.00	98.00	92.80	92.80
Proposed KNN	Min	56.00	88.00	45.33	74.40
	HOS	28.00	75.00	20.67	47.20
	T-Peak	100.00	100.00	100.00	56.80
	QRS Time	48.00	94.00	44.00	72.00
	Hybrid	100.00	100.00	100.00	98.40

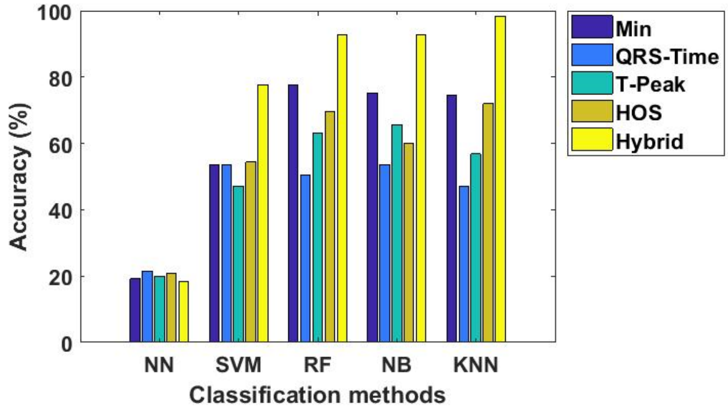

Figure. 6 Accuracy performance measure

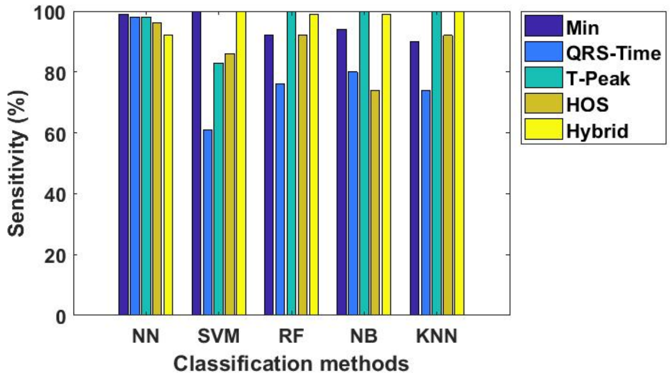

Figure. 7 describes the efficiency of Sensitivity

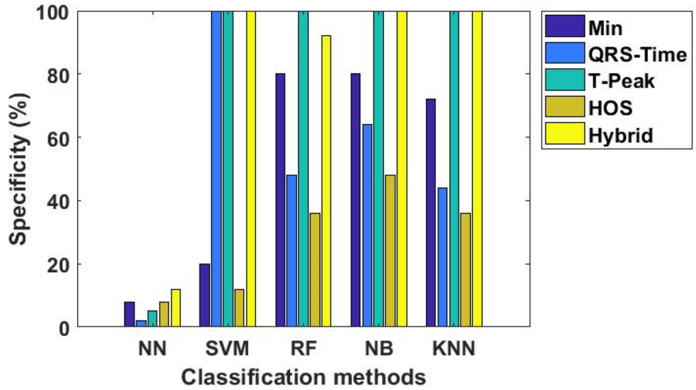

Figure. 8 represents the specificity performance

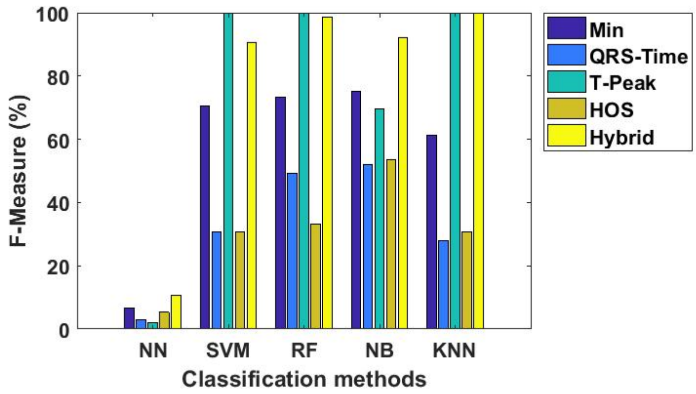

Figure. 9 F-measure performance

Table 2. The performance analysis of proposed method with existing methodologies.

References	Database	Classification method	Accuracy
A.E. Zadeh, et al. [12]	MIT-BIH arrhythmia dataset	RBF_BA	95.18%
U. Rajendra Acharya, et al. [13]	MIT-BIH arrhythmia dataset	KNN-DT	98.17%
Proposed work	MIT-BIH arrhythmia dataset	KNN	98.40%

Table 1 represents the window method of the ECG signal classification with the four classifiers such as NAVY, RF, NN, SVM and KNN classifier and also represents the parameters of four classifications namely accuracy, sensitivity, specificity and f-measures are shown in the following Figs. 6, 7, 8, and 9.

4.4. Comparative Analysis

Table 2 presents the comparative study of existing work and the proposed work performance. A.E. Zadeh, et al. [12] presented a new methodology, that contained HOS as FE with other some interval features. The classification process was able to classify these FE by using RBF_BA classifier. This experiment was also performed on MIT-BIH arrhythmia dataset and achieved 95.18% of classification accuracy. In addition, U. Rajendra Acharya, et al. [13] developed a prediction algorithm for arrhythmia by combining cumulant features with HOS bispectrum were extracted. The classification of the ECG signals was provided by a binary classifier (KNN-DT) after the extraction of feature information. This experiment was carried out on a publicly available database (i.e., MIT-BIH arrhythmia dataset) to validate its result by means of accuracy and achieved 98.17% of accuracy. Whereas, the proposed hybrid method achieved 98.40% of accuracy that was higher than the existing work.

5 Conclusion

ECG signal is highly used in the medical field to analyse the health of a patient. An automated classification of a CAD in ECG signals with good accuracy is a great challenging task. The proposed method includes steps such as data pre-processing, extracting the modules and classifying the extracted features. In this method, hybrid features are obtained from the ECG signal to classify the signal. The ECG signal was collected from the real-world database such as MIT-BHE arrhythmia database which contains a vast amount of ECG signals for both sinus and arrhythmia rhythm. The database also contains the label values and it shows the normal and sinus rhythm, which was used in the testing and training process. The peak can be detected by using PTA from the ECG signals which is the first step. In the FE module, the efficient features such as QRS time, HOS, temporal are extracted to differentiate various types of ECG signal, then KNN is used for classifying these signals. The classification accuracy of 98.40% for the dataset are obtained and the results also proved that the KNN classifiers provide higher accuracy and performed well compared to the existing methods. In future work, the stages of diseases can be increased and also the peak detection can be improved using other algorithms.

References

[1] L. Duan, H. Zhang, and M. Zhang, “Wavelet De-Noising and Genetic Algorithm-Based Least Squares Twin SVM for Classification of

Arrhythmias", *Circuits, Systems, and Signal Processing*, Vol.36, No.7, pp.2828-2846, 2017.

[2] H. Li, H. Liang, C. Miao, L. Cao, X. Feng, C. Tang, and E. Li, "Novel ECG signal classification based on KICA nonlinear feature extraction", *Circuits, Systems, and Signal Processing*, Vol.35, No.4, pp.1187-1197, 2016.

[3] R.K. Tripathy and S. Dandapat, "Detection of cardiac abnormalities from multi lead ECG using multiscale phase alternation features", *Journal of medical systems*, Vol.40, No.6, pp.143, 2016.

[4] A. Emina and A. Subasi, "Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier", *Journal of medical systems*, Vol.40, No.4, pp.108, 2016.

[5] W.F. Wang, C.Y. Yang, and Y.F. Wu, "SVM-based classification method to identify alcohol consumption using ECG and PPG monitoring", *Personal and Ubiquitous Computing*, Vol.22, No.2, pp.275-287, 2017.

[6] G. Pan, Z. Xin, S. Shi, and D. Jin, "Arrhythmia classification based on wavelet transformation and random forests", *Multimedia Tools and Applications*, pp. 1-18, 2017.

[7] I. Kaur, R. Rajni, and A. Marwaha, "ECG Signal Analysis and Arrhythmia Detection using Wavelet Transform", *Journal of the Institution of Engineers (India): Series B*, Vol.97, No.4, pp.499-507, 2016.

[8] O. Singh and R.K. Sunkaria, "ECG signal denoising via empirical wavelet transform", *Australasian Physical & Engineering Sciences in Medicine*, Vol.40, No.1, pp.219-229, 2017.

[9] U.R. Acharya, O. Faust, V. Sree, G. Swapna, R. J. Martis, N. A. Kadri, and J. S. Suri, "Linear and nonlinear analysis of normal and CAD-affected heart rate signals", *Computer Methods and Programs in Biomedicine*, Vol.113, No.1, pp.55-68, 2015.

[10] Y.H. Chen and S.N. Yu, "Selection of effective features for ECG beat recognition based on nonlinear correlations", *Artificial Intelligence in Medicine*, Vol.54, No.1, pp.43-52, 2014.

[11] S.S. Qurraie and R.G. Afkhami, "ECG arrhythmia classification using time frequency distribution techniques", *Biomedical Engineering Letters*, Vol.7, No.4, pp.325-332, 2017.

[12] F.A. Elhaj, N. Salim, A.R. Harris, T.T. Swee, and T. Ahmed, "Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals", *Computer Methods and Programs in Biomedicine*, Vol.127, pp.52-63, 2016.

[13] U.R. Acharya, V.K. Sudarshan, J.E. Koh, R.J. Martis, J.H. Tan, S.L. Oh, and C.K. Chua, "Application of higher-order spectra for the characterization of Coronary artery disease using electrocardiogram signals", *Biomedical Signal Processing and Control*, Vol.31, pp.31-43, 2017.

[14] U. Desai, C. Gurudas Nayak, and G. Seshikala, "Application of ensemble classifiers in accurate diagnosis of myocardial ischemia conditions", *Progress in Artificial Intelligence*, Vol.6, No.3, pp.245-253., 2017.

[15] S.S. Kumar, and H.H. Inbarani, "Cardiac arrhythmia classification using multi-granulation rough set approaches", *International Journal of Machine Learning and Cybernetics*, Vol.9, No.4, pp.651-666, 2016.

[16] R. Nanjundegowda and V. Aniruddha Meshram, "Arrhythmia Detection Based on Hybrid Features of T-wave in Electrocardiogram", *International Journal of Intelligent Engineering and Systems*, Vol.11, No.1, pp.153-162, 2017.

[17] G.R. Vedavathi and S.V.A.V. Prasad, "Robust Method for The Classification of the Arrhythmia with DNN Classifier in the Diverse ECG Signal", *International Journal of Computer Engineering & Technology*, Vol.8, No.5, pp.107-119, 2017.