



## ECG Signal Classification with Hybrid Features Using Bayesian Optimized K-Nearest Neighbors Classifier

Halemirle Appaji Deepak<sup>1\*</sup>

Thippeswamy Vijayakumar<sup>1</sup>

<sup>1</sup>*Department of Electronics and Communication Engineering, SJB Institute of Technology, BGS Health & Education City, Bengaluru-60, Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India*

\* Corresponding author's Email: [deepakgowdal@gmail.com](mailto:deepakgowdal@gmail.com)

---

**Abstract:** In recent years, several methods have been established for the automatic detection of the characteristics of electrocardiogram (ECG) signals based on mathematical tools, among which the Fourier Transform, the Continuous Wavelet Transform (CWT), the Discrete Wavelet Transform (DWT) stand out among others. This type of procedure is important because it can be more efficiently detected if a certain patient has heart disease, such as an arrhythmia or ischemia. The primary goal of this paper is the development and implementation of a classification approach based on hybrid features i. e., Dual-Tree Complex Wavelet Transform (DTCWT), SVD-Entropy, Autoregressive modeling, and Multifractal analysis based feature extraction. To accomplish the better classification, the extracted features are further classified by Random Forest Classifier, K-Nearest Neighbors (KNN), and Bayesian Optimized-KNN classifiers utilizing the MIT-BIH database. Highest accuracy achieved in random forest classifier is 98.29 %. The results obtained show the feasibility and practical efficiency of the methodology as a tool to aid in the diagnosis of heart disease in hospital environments.

**Keywords:** AR, Bayesian Optimization, DT-CWT, ECG, KNN, MFF, SVD-Entropy, WL.

---

### 1. Introduction

The purpose of medical signal processing is to process raw biological signals and obtain information with a diagnostic value from this mark. These signs can be interpreted or classified using different methods. These classification processes can be used as Decision Support System in medicine.

This study proposes a combination of hybrid traits. Coefficients of the autoregressive (AR) model of the 4<sup>th</sup> order. The SVD entropy value for the MODPWT (maximal overlap discrete wavelet packet transform) at level 4 of the reference exponent or singularity spectrum.

In addition, an estimate of the multiscale wavelet variant based on DTCWT was extracted for each signal along the entire data length. An unbiased estimate of the wavelet variant is used. This requires that when estimating variance, only levels with at least one wavelet coefficient are used, which are not

affected by the boundary conditions. Important properties extracted from the wavelet coefficient are the maximum, minimum, mean, standard deviation and median of the corresponding wavelet coefficients in registers 100, 106, 109, 118 and 209, which represent normal ventricular premature signals, contraction of the left bundle branch, right bundle branch, or premature atrial contraction.

In addition, these features are trained and tested with Random Forest and the KNN Classifier. In addition, the random forest hyperparameter and the KNN classifier are combined with Bayesian optimization to achieve greater accuracy. Experimental results are presented for five segmented heart rate signals obtained from the MIT-BIH arrhythmia database. The accuracy, precision, recoverability, and F-score of the performance evaluation parameters are calculated for the proposed system analysis.

The present work is concluded to the detection of

hybrid features (DTCWT, SVD-Entropy, Autoregressive modeling, and Multifractal analysis) followed by the classification by using random forest, k-nearest neighbors, and Bayesian optimized-KNN classifiers.

With the advancement of computational processing, several techniques for digital analysis of electrocardiogram (ECG) signals have been developed. The ECG is a medical examination that can identify different types of heart disease, which is non-invasive and economical. The time and amplitude of the waves of ECG signals are parameters of fundamental importance for the diagnosis of these diseases. Among the diseases that the ECG exam can identify are several cardiac arrhythmias, such as tachycardia, bradycardia and sinus arrest, myocardial infarction, ventricular blocks, ventricular contractions, among others.

The ECG signal is composed of the repetitive succession of 5 waves, each of which represents a part of the cardiac cycle. These are the P wave, the Q wave, the R wave, the S wave, and the T wave, each of which has a predetermined range of amplitude and duration, and whose values outside these ranges can identify some type of heart disease.

Direct classification is ineffective in terms of performance, so the automatic identification of the segmented pulse signal with the extracted characteristics plays an important role as input for the classification. Wavelet transform is one of the methods commonly used to obtain the properties of ECG signals [1-9]. Using various statistical methods such as principal component analysis (PCA) and linear discriminant analysis (LDA) for these coefficients, low-dimensional characteristic vectors are generated [3, 4, 6] by obtaining the wavelet coefficients from different signal levels. Li and Ming [7] used the WPA (Wavelet Packet Decomposition) method to obtain characteristic properties for detecting five different types of beats. They calculated the entropy of the coefficients decomposed using WPA. Authors of [8] calculated the detail and approximate wave factor of the ECG signal to create a vector of characteristics. Authors of [9] used various feature extraction techniques to automatically detect arrhythmias in an ECG sample to achieve the best results using DWT properties, RR interval, and signal energy. Other methods such as high-order statistics (HOS), independent component analysis (ICA), PCA and cumulative have also been used to extract properties of the ECG signal [5, 6, 10-14]. Authors of [6] extracted the linear and non-linear properties of the signal to automatically determine the ECG pulse. Authors of [15] have successfully used various entropy properties to diagnose

computerized ECG signals. Authors of [16] used sophisticated PQRST feature extraction to characterize heart rate data into five grades. This complex PQRST function includes position, duration, amplitude, and waveform. Alickovic and Subasi [17] proposed an autoregressive (AR) model in feature extraction to classify five classes of arrhythmia data. Plaviak [18] proposed a new method for classifying heart diseases based on ECG signals. Smelt uses Power Spectral Density with Welch and Discrete Fourier Transforms to extract various properties of ECG signal fragments.

Finally, various classification methods are used with the AdaBoost method [19], radial basis function [20], neuro-diffuse adaptive function [21], convolutional neural networks (CNN) [22-24] and the Extreme Learning Machine. (ELM) [25, 26]. IT assistance based on the ELM system was carried out in the classification of ECG and EEG, etc. due to the ability to quickly learn and generalize [24-28]. However, there are some difficulties, such as a local minimum of availability, an indeterminate learning rate, the choice of the number of hidden neurons, and overfitting [29]. To overcome the weakness of ELM, there are several nature-inspired population methods with global search functions such as differential evolution (DE) [30-33], Algorithm for Artificial Bee Colony (ABC) [34], Ant Colony Optimization (ACO) [35], Genetic algorithm (GA) [36], and particle swarm optimization (PSO) [37]. Following are the pros and cons of above-mentioned conventional techniques:

- The ABC algorithm has the advantages of high reliability, fast convergence, and high flexibility. However, it has a major drawback - premature convergence in subsequent IPs.
- The ACO optimization is well received, resulting in good solutions being found quickly, and the distributed processing prevents premature convergence. With ACO, convergence is guaranteed, but the convergence time is undefined, which is a big drawback.
- Differential evolution is easy to apply to a variety of problems, apart from noisy, multimodal, and multidimensional spaces, which often make it difficult to optimize the problem. The overall minimum search efficiency is very sensitive to the configuration of the control parameters, and this is a weak point for DE for some global optimization problems.
- Genetic algorithms require less information about the problem, but it can be difficult to

design objective functions and get correct representations and operators.

- PSOs have the advantage of fast convergence, but they can also converge prematurely and, especially in the case of complex problems, stagnate at local lows.

ECG diagnostics usually use features derived from the P wave, QRS complex, and T wave, which are an important part of the cardiac recording.

Rest of the paper is organised as follows; section two presents pre-processing and post-processing based proposed methodology. Results and discussion are provided in section three followed by the conclusion and future perspective in section four.

### 1.1 Materials and methods

#### 1.1.1. Electrocardiogram:

An electrocardiogram is a medical test that records all of the electrical activity of the heart. Each beat is represented by a series of waves, marked by peaks and valleys. The ECG signal frequency is typically 0.05 to 100 Hz with a signal amplitude range of 1 to 10 mV. The ECG signal is characterized by 5 waves, which are called P, Q, R, S, and T waves. In some cases, in the wave shown in Fig. 1, one can see a sixth wave, called a U wave. The P wave corresponds to the atrial depolarization wave; The QRS complex, which is the junction of the Q, R and S waves, is related with ventricular depolarization, while the T wave is related with ventricular repolarization [38].

For ECG analysis, you need to analyze the intervals between these waves, such as: P-R interval, P-Q interval, S-T interval, Q-S interval. This range has a value as defined by the Brazilian Society of Cardiology (SBC), where values outside this range may indicate structural and functional changes in the heart.

At a normal heart rate known as sinus rhythm, the PR interval is 0.12 to 0.2 seconds. QRS interval from 0.04 to 0.12 seconds; The R-R interval is 0.6 to 1.2 seconds, and the heart rate (HR) is 60 to 100 beats per minute (beats per minute) [39].

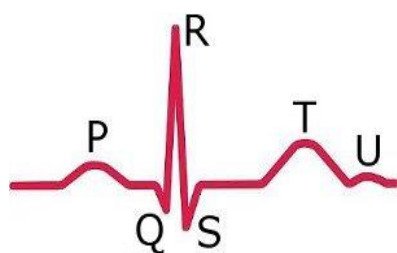


Figure. 1 Characteristic waves of the ECG signal [38]

The effectiveness of an ECG signal analysis system usually depends on the accuracy and reliability of complex QRS detection [40, 41].

### 1.2 MIT-BIH arrhythmia database signals

An ECG signal database called the "MIT-BIH Arrhythmia Database" is widely used in science for ECG signal analysis research. This database consists of 48 ECG samples divided into two phases. The first includes 23 lists containing the most common diseases such as tachycardia and bradycardia. The second phase, on the other hand, includes signs indicating rare but clinically important events such as conduction disturbances, morphological changes, and other types of arrhythmias. Each of these signals has a duration of 30 minutes, collected at a sampling rate of 360 Hz [42, 43]. In this work, we used signals from this database [43]. Most heart rate values are recorded in this database. The AAMI standard defines how annotations should be generated. He recommended dividing heartbeats into five classes, including normal heartbeats (N), supraventricular ectopic heartbeats (S), ventricular ectopic heartbeats (V), combined V and N heartbeats (F), and unknown pulse types (Q).

## 2. Proposed methodology

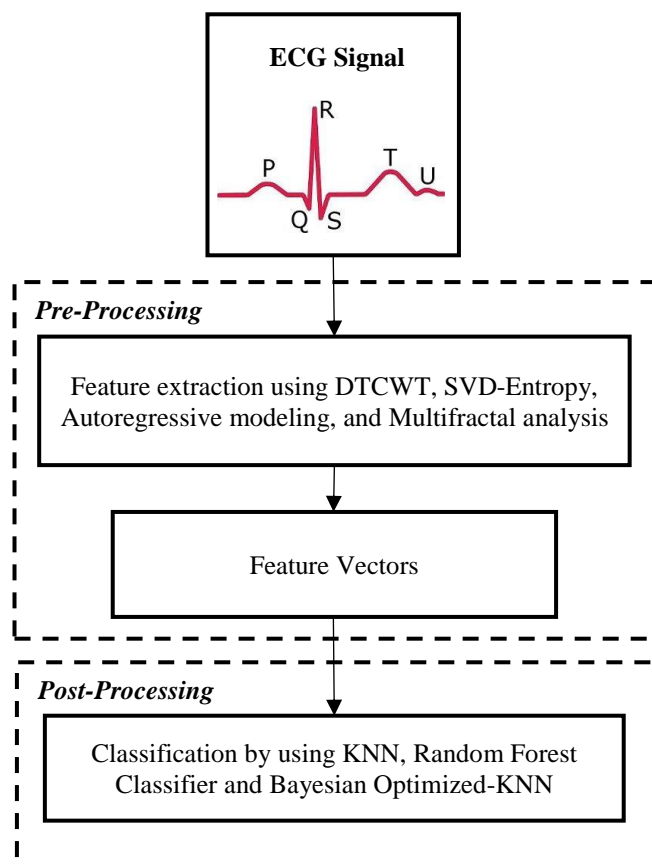


Figure. 2 Flow diagram for proposed research work

## 2.1 Methods

A block diagram of the proposed method for classifying ECG signals is shown in Fig. 2. This method can be divided into three parts, as shown in Fig. 2. First, the ECG signal from the MIT database is provided for preprocessing. The preprocessing process is carried out in two stages, namely baseline wandering, noise suppression and QRS detection. Feature extraction is applied to the previously processed signal and hybrid feature set, which are also provided as input to ANN and random forest classifiers to classify them as normal signal, ventricular premature contraction, right bundle branch, left bundle branch, contraction, and premature atria respectively.

## 2.2 Preprocessing

Interference and noise can be unwanted, which means they cause jitter in signal processing applications. Interference and noise can be unwanted, which means they cause jitter in signal processing applications. The ECG contains too many unsweetened sounds, so this step plays an important role in the application of ECG classification [44-45]. In addition, it is used as a routine analytical step in most biomedical procedures to obtain a clear and stable signal. ECG signals cover many types of disturbances such as: system interaction and baseline wandering. Symmetric scale filters and denoising operations were used to correct baseline deviations. In addition, linear and systemic interactions were reduced at this stage. These distractions in the recording of the heart were then eliminated using the technique suggested by Deepak and Vijayakumar [46]. The Pan-Tompkins (PT) [42] algorithm was then used to detect QRS.

## 2.3 Feature extraction

There will different hybrid features are proposed for ECG beat classification represented in Figure 3. Following features are extracted on after QRS detection and additionally DTCWT features are extracted on entire ECG signal set on a given time frame. This hybridization ensures the system model more generalize on different conditions.

### 2.3.1. Dual-tree complex wavelet transform (DT-CWT) based feature extraction

The Hilbert transformation of a signal resembles the Fourier plane to a complex gain filtering  $-i \text{sign}(v)$  [47]. According to distribution theory, this corresponds to an impulse response  $vp \left( \frac{1}{\pi t} \right)$  where  $vp$  denotes the Cauchy principal value.

We then build the analytical signal [47]:

$$z(t) = x(t) + i\mathcal{H}\{x(t)\} = x(t) + \frac{i}{\pi} vp \int_{-\infty}^{\infty} \frac{x(s)}{t-s} ds \quad (1)$$

Where,  $z(t)$  is output signal in Fourier plane.  $x(t)$  is finite energy analog signal.  $\mathcal{H}$  represents the Hilbert transform.

It is clear that the Eq. (1) has only positive frequencies. The Hilbert transform of a real signal is real and is quadrature of phase with respect to the signal.

Rather than taking the Hilbert transform of the wavelet (which is defined by the intermediary of the related filters), we can take the Hilbert transform of the signal, and analyze it with the initial wavelet, because  $\langle f, \mathcal{H}\psi_{a,t} \rangle = \langle \mathcal{H}f, \psi_{a,t} \rangle$  since the Hilbert transformation is a linear filtering.

The scheme is therefore the following: we have a signal  $X(n)$ , we analyze it with a real wavelet by Mallat's algorithm to obtain the wavelet coefficients  $d_1(j, k)$ . Then, we analyze  $\mathcal{H}X(n)$  by the same wavelet and we obtain the coefficients  $d_2(j, k)$ . We construct the complex coefficients:  $d_{complex}(j, k) =$

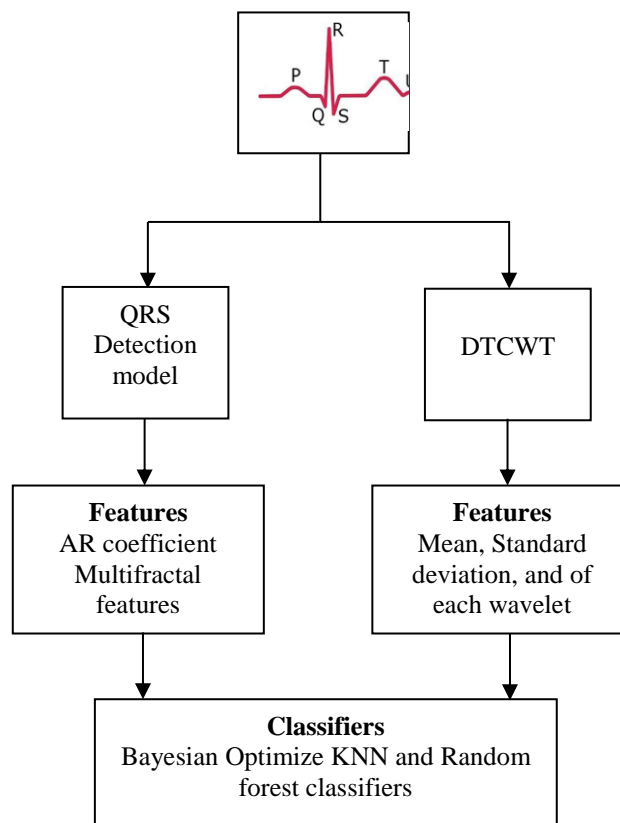


Figure. 3 Features for training and testing in RF and KNN classifiers

$d_1(j, k) + id_2(j, k)$ . In the following, the modulus of these coefficients is called Hilbert modulus or modulus by Hilbert transformation.

The disadvantages of this technique are twofold:

- The sustenance of the Hilbert transform of a compact support wavelet is infinite. We therefore lose in locality;
- There is a computational drawback since the cost is two wavelet transform plus one Hilbert transform (i.e. two FFT).

It is theoretically possible to limit the inconvenience linked to the non-compact sustenance of the Hilbert transform of the wavelet by being satisfied with an estimated Hilbert transform. However, the approximation cannot then be optimized for all scales because we only do the Hilbert transformation once at the start.

The DT-CWT consists of analyzing a signal using two different DWT trees and choosing a filter that is finally decomposed by an analytical wavelet. Fig. 4 shows how the algorithm works using two different sets of filters:  $h_1$  and  $g_1$  are high pass filters for the first and second trees, and  $h_0$  and  $g_0$  are low pass filters for the first and second trees.

The first tree gives the coefficient of the real part  $d_r(j, k)$ , and the second tree gives the coefficient of the imaginary part  $d_i(j, k)$ . Then we build a complex coefficient  $d_{complex}(j, k) = d_r(j, k) + id_i(j, k)$ . This modulus of the coefficient is called the dual-tree modulus [47, 48].

Finally, the use of this structure requires a pre-filtering operation, i.e. the filters used in the first step are not the same as those used in the following steps. The first scale will therefore never be taken into account in our analysis problem.

The advantages of this method over simple Hilbert signal transform (Eq. (1)):

- Lower calculation costs (only two DWT),
- Estimates of the Hilbert transform, optimized for each scale.
- The possibility of adequate reconstruction remains.

The major drawback is the inability to use the well-known DWT wavelets (Daubechies wavelets, spline), and therefore also to select the number of zero moments (all  $qshift$  filters give wavelets to at two zero moments).

### 2.3.2. Autoregressive (AR) modeling

A time series from a stochastic process can be

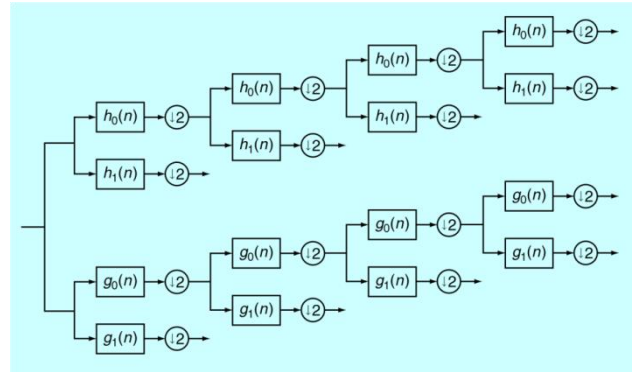


Figure. 4 Tree representing the functioning of Dual Tree [16]

analyzed through an autoregressive parametric model. Such a model states that the sample  $x_t$  of a time series at a given moment can be expressed as a weighted linear summation of  $p$  previous values of the time series  $x_{t-1}, x_{t-2}, \dots, x_{t-p}$ , plus a component uncertainty or error (Eq. 2). The  $p$ -value is known as the order of the model and is normally selected much smaller than the amount of data in the time series [49].

$$x_t = \sum_{i=1}^p a_i x_{t-i} + \varepsilon_t \quad (2)$$

Where  $x_t$  is the time series,  $\alpha_1, \alpha_2, \dots, \alpha_p$  are the parameters of the autoregressive model of order  $p$  and the factor  $\varepsilon_t$  is known as the prediction error of the process, which is white noise. This type of parametric modeling has been widely described for the analysis of EEG signals, and its use extends to different topics, such as the classification and segmentation of EEG signals, the identification and cancellation of artifacts, the description of the different brain rhythms, the simulation and generation of EEG signals, among others [50].

One of the most important assumptions that this parametric modeling technique has, and at the same time a limitation, is the stationary condition of the stochastic process, which from the statistical point of view corresponds to when its first and second moment, mean and variance respectively, do not vary with respect to time [49, 51]. In EEG signals, stationarity is highly dependent on signal length, and it has been found that EEG signals can be assumed locally stationary when analyzed in short time intervals [52].

To establish the proper order of the model, the use of functions called "information criteria" is common. The most appropriate order for the model is the one that minimizes the information criterion when it is evaluated over a range of given orders. One of these information criteria is the Hannan-Quinn criterion [53], denoted by  $HQ(p)$  (Equation 3). This criterion

is defined as a consistent estimator, which converges to its true value when the number of data tends to infinity.

$$HQ(p) = \log(V) + \frac{2p \log(\log(N))}{N} \quad (3)$$

Where  $V$  is the variance of the input white noise,  $N$  is the number of samples of the signal and  $p$  is the order of the evaluated model. The previous function is composed of two terms, the first one seeks to minimize the variance of the process, and the second characterizes the number of parameters to be estimated in the model. The interpretation of the  $HQ$  criterion is that the proper order corresponds to the lowest value of  $HQ$  when it is evaluated in a range of orders.

In addition, the adjustment index or percentage of the change in output has been estimated. This index provides a measure of the fit of the model to the real data (Eq. (4)).

$$FIT = \left( 1 - \frac{\|x_t - \hat{x}_t\|}{\|x_t - \text{mean}(x_t)\|} \right) \times 100, \quad [\%] \quad (4)$$

Where  $x_t$  is the time series and  $\hat{x}_t$  is the time series simulated from the coefficients of the AR model. FIT values close to 100% indicate a good fit of the model, while values close to 0% denote an inappropriate model [54].

The stability of the AR model is evaluated through the reflection coefficients calculated by the Burg method and verifies that the magnitude of these coefficients is less than or equal to unity [55, 56]. Finally, the power spectral density (PSD) has been estimated, using the Welch periogram [57], both for the original EEG signals and for the simulated signals, above in order to verify that the characteristics in the domain of the frequency of the simulated signals correspond to those of the original signals. For each spectrum, the value of the average energy has been obtained in each of the representative frequency bands of the EEG, described previously, in addition, two parameters have been calculated that provide information about the displacement of the power spectrum of the signals. These parameters are the mean frequency ( $F_m$ ): corresponding to the mean value of the frequency in the PSD, and the central frequency ( $F_c$ ): defined as the median of the PSD.

### 2.3.3. Multifractal analysis

**Hölder Exponent and Spectrum of Singularities:** Let  $x_0 \in \mathbb{R}$  and  $\alpha \geq 0$ . A bounded function  $f: \mathbb{R} \rightarrow \mathbb{R}$  is  $C^\alpha(x_0)$  if there is a constant  $C > 0$  and a polynomial  $P$  that satisfies degree  $(P) < \alpha$  such that, in a neighborhood of  $x_0$ , the relation  $|f(x) - P(x - x_0)| \leq C|x - x_0|^\alpha$ . The Hölder exponent of  $f$  at  $x_0$  is  $h_f(x_0) = \sup\{\alpha: f \in C^\alpha(x_0)\}$  [58]. It measures the local regularity of  $f$  at the point  $x_0$ .

Usually, the variations of the Hölder exponent along a signal are described by the spectrum of singularities (or multifractal spectrum - MFS)  $D(h)$ . This is defined as the Hausdorff dimension of the set of points at which the Hölder exponent proceeds the value  $h$ . In other words, the value of  $D(h)$  indicates the proportion of points in the signal that have local regularity  $h$ .

**Multifractal Formalism (MFF):** Determining the spectrum of singularities of a signal is essential to analyze its properties. On the other hand, it is not possible to do it from its definition. This is due to the fact that, in general, in multifractal signals, the local regularity varies abruptly between one instant and the next, and the limitations of finite resolution and sampling period do not make their discrimination possible [58]. To solve this problem, we use the introduction of MFF, which offers an alternative way to obtain a spectrum of singularities using an easily computable element: structure-function (SF). As mentioned in [58], a wavelet leader (WL) based MFF was proposed. This approach overcomes many of the disadvantages of the previous method. Its main attraction is that it is based on DWT, which allows fast implementation of the filterbank algorithm and its numerical stability.

**Wavelet Leaders:** A review of the definition of WLs and the MFF based on them is presented below [58]:

Let  $\psi_0$  be a mother wave with compact support and a number  $N \geq 1$  of null moments. Let  $\{\psi_{j,k}(t) = 2^{-j}\psi_0(2^{-j}t - k), j \in \mathbb{N}, k \in \mathbb{N}\}$  be the orthonormal basis of  $L^2(\mathbb{R})$  formed by the versions of  $\psi_0$  dilated to the scales  $2^j$  and translated to the positions  $2^j k$ . The coefficients of the discrete waveform transform of a signal  $f$  are  $d_f(j, k) = \int_{\mathbb{R}} f(t) 2^{-j} \psi_0(2^{-j}t - k) dt$ .

A special notation is defined for dyadic intervals. Let  $\lambda = \lambda_{j,k} = [k2^j, (k+1)2^j]$  such that  $d_\lambda = d_f(j, k)$ . Lastly, let  $3\lambda$  be the union of  $\lambda$  and its two adjacent dyadic intervals:  $3\lambda_{j,k} = \lambda_{j,k-1} \cup \lambda_{j,k} \cup \lambda_{j,k+1}$ .

WLs are defined as [59]:

$$L_f(j, k) \equiv L_\lambda = \sup_{\lambda' \in 3\lambda} |d_{\lambda'}| \quad (5)$$

That is, the calculation of the WL for a particular time and scale involves searching for the value of the supreme in a temporal neighborhood of the point in question, for all the finer scales.

**Multifractal Formalism based on Wavelet Leaders:** From the WL, the structure functions (SF) are calculated:

$$S_L(q, j) = \frac{1}{n_j} \sum_{k=1}^{n_j} |L_f(j, k)|^q \quad (6)$$

Using the SF, the scaling exponent (SE) is calculated:

$$\zeta_L(q) = \lim_{j \rightarrow 0} \inf \left( \frac{\log_2 S_L(q, j)}{j} \right) \quad (7)$$

Finally, the MFS is obtained by:

$$D(h) = \inf_{q \neq 0} (1 + qh - \zeta_L(q)) \quad (8)$$

### 2.3.4. SVD- entropy based feature extraction approach

Here, SVD-based entropy was originally proposed to extract features of the ECG signal. The formula for calculating the singular value decomposition of each feature vector is presented below.

$$V_j = \frac{S_j^2}{\sum_w S_w^2} \quad (9)$$

Where  $S_j$  represents the singular value of the feature vector,  $S_w^2$  represents the eigenvalues of the feature vector,  $w$  represents the window size. The entropy of SVD is as follows [60]:

$$E = -\frac{1}{\log(w)} \sum_{j=1}^w V_j \log(V_j) \quad (10)$$

Features are divided into three categories [60]. Let  $m$  be the mean of all entropy in SVD and the standard deviation equal to  $n$ .

1.  $E < m + n$ , features with high contribution.
2.  $m + n > E > m - n$ , features with regular influence.
3.  $E < m - n$ , features with a destructive influence.

The acquired features in the first phase were considered significant for the extracted features from the ECG. Features of the second phase are considered neutral, while those of the third phase reduces the overall structure of the SVD.

**Algorithm:** SVD-Entropy Based Feature Extraction

**Input:** ECG Signal

**Output:** Extracted Features.

**Process:**

**Step-1:** Entropy Calculation

For  $i = 1$  to Number of features

SVD calculation for each feature vector

Let  $K =$  Number of non-zero SVD entries

along window size  $w$

For  $j=1, \dots, K$

Calculate SVD-Entropy using equation (10)

End For

End For

**Step-2:** Extraction of Features from ECG Signal

If entropy of each feature vector is less than threshold value then select those features for classification process.

Else

Eliminate the features from classification process

End If

## 2.4 Classification

Further after feature extraction by above mentioned methods it is concatenated as a set of hybrid feature for classifier input. All the hybrid feature is divided as a set for testing and training samples for classifiers according to 10 k fold verification method.

The goal is to automatically classify the ECG beats. The choice of the classifier is very important, it constitutes the decision element in a pattern recognition system.

### 2.4.1. Classification by using random forest classifier

It is a classifier containing a set of base classifiers such as a decision tree shown:

$$\{h(x, \Theta_k), k = 1, \dots, L\} \quad (11)$$

Random forests are composed of a set of binary decision trees in which randomness has been introduced.

Random forests were introduced by Breiman (2001) by the following very general definition [61]: Let  $(\hat{h}(\theta_1), \dots, \hat{h}(\theta_q))$  a collection of tree predictors, with  $\theta_1, \dots, \theta_q$  random variables independent of  $\mathcal{L}_n$ . The predictor of random forests  $\hat{h}_{RF}$  is obtained is aggregating this collection of random trees as follows:

- $\hat{h}_{RF}(x) = \frac{1}{q} \sum_{l=1}^q \hat{h}(x, \theta_l)$  Average of individual tree predictions in regression.
- $\hat{h}_{RF}(x) = \arg \max_{1 \leq k \leq K} \sum_{l=1}^q 1_{\hat{h}(x, \theta_l)=k}$  Majority vote among individual predictions trees in classification.

The term random forest comes from the fact that individual predictors are, here, explicitly predictors per tree, and that each tree depends on an additional random variable (that is, in addition to  $\mathcal{L}_n$ ).

#### 2.4.2. Classification by k-nearest neighbors (KNN)

The K-nearest neighbors algorithm is simple, but who can give interesting results if the data range is large enough. It's about a classification method widely used in many fields and is also found among the top 10 data mining algorithms [62]. Typically, houses that are close to each other have similar characteristics. We can group them and give them a classification. The algorithm uses this same logic to try to group the elements that are close to each other. KNN is an example of instance-based learning. It works in situations where each instance can be defined by an n-dimensional vector, where n is the number of attributes used to describe each instance and the ranges are discrete values. The training data is written and when new samples are found, it is compared with the training data to find the nearest neighbors.

The nearest neighbors are those who are closest following the Euclidean distance. The distance between two elements  $A = \langle a_1, \dots, a_n \rangle$  and  $B = \langle b_1, \dots, b_n \rangle$  is calculated as follows [63]:

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \tag{12}$$

Sorted by the  $k$  neighbors closest to the new instance, the classification assigned to it will be the class with the highest occurrence among them.

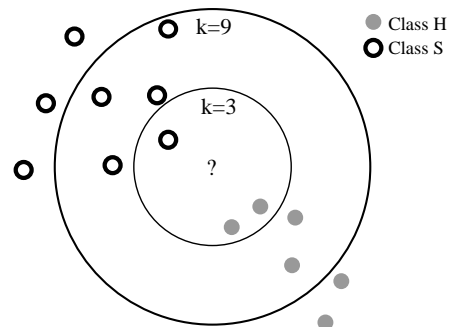


Figure. 5 Operation of a simple KNN with  $k = 3$  and  $k = 9$  [62]

Here is the pseudo-code representing the algorithm:

#### Pseudo-Code for KNN

Requires 3 parameters: a set of examples  $X$ , a given  $x$  and  $k \in \{1, \dots, k\}$

For each example  $x_i \in X$

Calculate the distance between  $x_i$  and  $x$ :  $\delta(x_i, x)$

End for

For  $j \in \{1, \dots, k\}$  do

$KNN(j) \leftarrow \arg \min \delta(x_i, x) \ i \in 1, \dots, n$   
 $\delta(x_i, x) \leftarrow +\infty$

End for

Determine the class of  $x$  from the class of examples whose number is stored in the KNN.

#### 2.4.3. Classification by using bayesian optimized KNN

First of all, three parameters are to be taken into consideration: the sample data, the number of nearest neighbors to select ( $k$ ), and the point we want to evaluate ( $x$ ). Subsequently, for each element of the sample, we evaluate the distance between reference point  $X$  and point  $x$ ; of the set of learning and we check if the distance between them is less than one of the distances contained in the list of nearest neighbours. If so, the point is added to the list. If the number of items in the list is more significant than  $k$ , the last value is simply removed from the list. The algorithm itself is not very complicated and can give a good result with brute force if sampling is not too big. However, since we are talking about data mining, the number of individuals to be evaluated is often very big, that's why an optimization algorithm is needed.

The main idea of Bayesian Optimization (BO) is to construct a surrogate probabilistic model sequentially to try to infer the objective function. Iteratively, new observations are made, and the model is updated, reducing its uncertainty allows working with a known and cheaper model, which is used to construct a utility function that determines the



next point to evaluate. The different steps of the BO methodology are described below.

First, the apriori model must be chosen over the possible space of functions. For this, different parametric approaches can be used, such as Beta-Bernoulli Bandit or Linear Models (Generalized), or non-parametric models such as t-Student Processes or Gaussian processes [64].

Then repeatedly until a particular stopping criterion [65]:

The prior and the likelihood of the observations so far are joined to get a posterior distribution. This is done using Bayes' theorem, hence the origin of the name.

Recall Bayes' theorem. Let  $A$  and  $B$  be two events such that the conditional probability  $P(B|A)$  is known, then the probability  $P(A|B)$  is given by:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (13)$$

Where  $P(A)$  is the a priori probability,  $P(B|A)$  is the probability of event  $B$  conditional on the occurrence of event  $A$ , and  $P(A|B)$  is the posterior probability.

Then, a particular utility function is maximized on the a posteriori model to determine the next point to evaluate and the new observation is collected to repeat until the stop criterion.

Since the KNN approach uses a discretization technique for the continuous parameter, therefore it results in less accurate results with the data loss. This proposed work discusses the algorithm that can tune KNN parameter  $k$ .

### 3. Results and discussions

#### 3.1 Evaluation parameters

Table 1. Evaluation parameters

<b>Accuracy</b>	$\frac{TP + TN}{TP + TN + FP + FN}$
<b>Recall (Sensitivity)</b>	$\frac{TP}{TP + FN}$
<b>Precision</b>	$\frac{TP}{TP + FP}$
<b>Specificity</b>	$\frac{TN}{TN + FP}$
<b>F-Score</b>	$\frac{2TP}{2TP + FP + FN}$
<b>Error Rate</b>	$\frac{FP + FN}{TP + TN + FP + FN}$

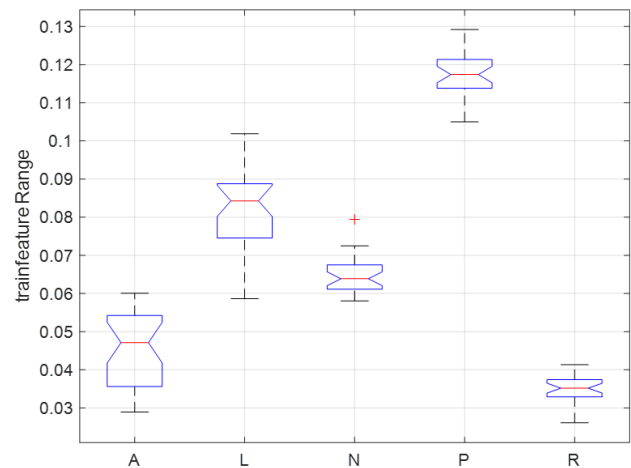


Figure. 6 Range of training features extracted by proposed hybrid approach

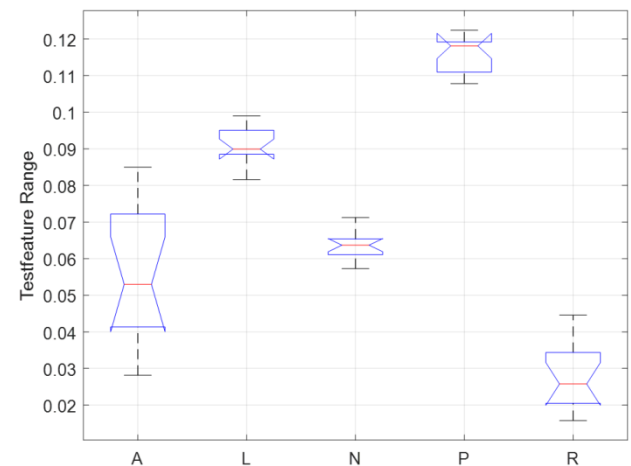


Figure. 7 Range of testing features extracted by proposed hybrid approach

#### 3.2 Simulation results

The hybrid feature distribution across datasets in the boxplot from Fig. 6 and Fig. 7 show the performance of the system with -training and testing. There are two primary observations from this plot:

- The hybrid feature in the proposed system is less skewed than other features. Skewness indicates that the data may not be normally distributed. Hence, the extracted hybrid feature has a stable distribution of data for the classifier as a training sample.
- The L, N, P and R classes are less skewed in the hybrid-feature-based plot as compared to class A from testing features. The median range is in the same range for hybrid features ranging from 0.056 to 0.07. The variation in all the class, the median weights of the notch plots are nearly similar.

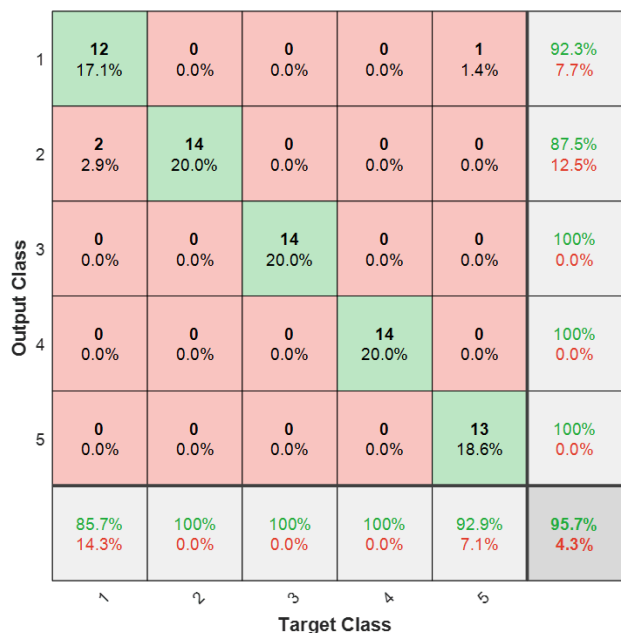


Figure. 8 Confusion matrix plot for proposed approach by using random forest classifier

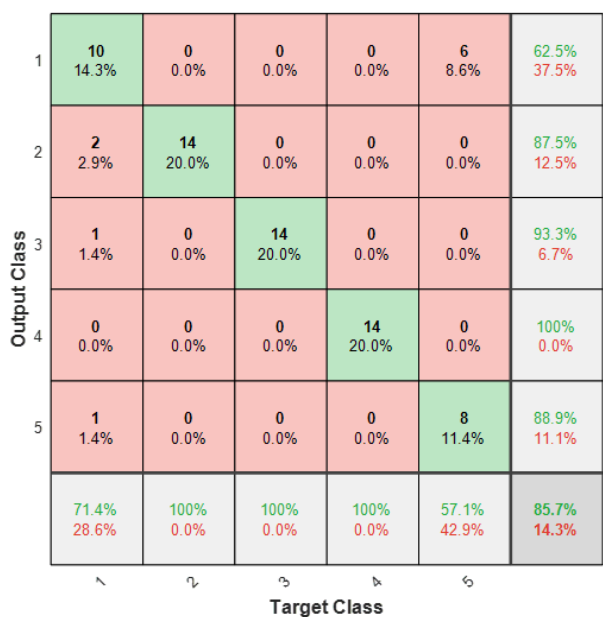


Figure. 9 Confusion matrix plot for proposed approach by using k-nearest neighbors classifier

Fig. 8 shows a confusion matrix plot for random forest classifier with five different classes. Classes 1, 2, 3, 4 and 5 represent the A, L, N, P and R waves respectively. The specificity achieved by this method is 95.7%.

Fig. 9 shows a confusion matrix plot for KNN classifier with five different classes. Classes 1, 2, 3, 4 and 5 represent the A, L, N, P and R waves respectively. The specificity achieved by this method is 85.7%.

Fig. 10 shows a confusion matrix plot for Bayesian Optimized-KNN classifier with five

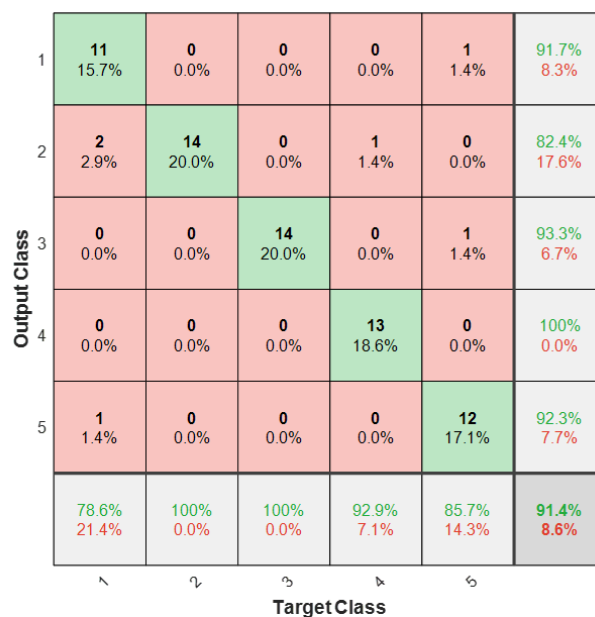


Figure. 10 Confusion matrix plot for proposed approach by using Bayesian optimized-KNN

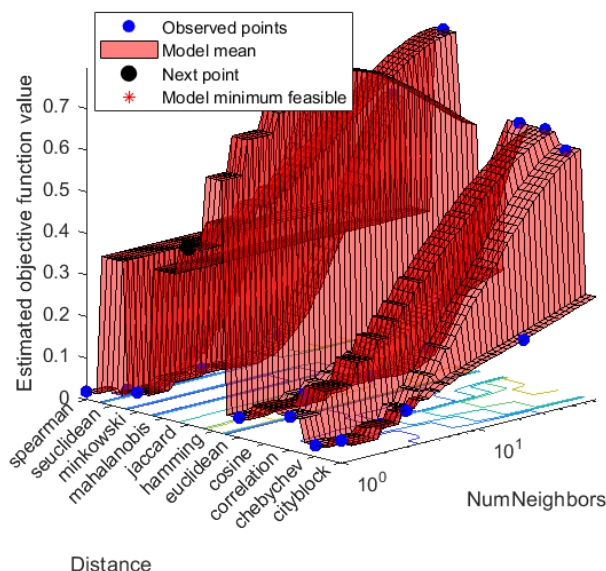


Figure. 11 Objective function model

different classes. Classes 1, 2, 3, 4 and 5 represent the A, L, N, P and R waves respectively. The specificity achieved by this method is 91.4%.

Fig. 11 shows the objective functional model of KNN, where various distance model is sampled for quantile error function evaluations in Fig. 12. It can be seen from the above graph convergence is achieved at 15 iteration itself.

Tables 2 present the classification results with respect to magnification for MIT-BIH dataset, trained with different sets of feature vectors. The hybrid feature-based system model produced 98.29 % for five class classification in Random forest classifier, while SVD-entropy Multifractal and AR coefficient-based

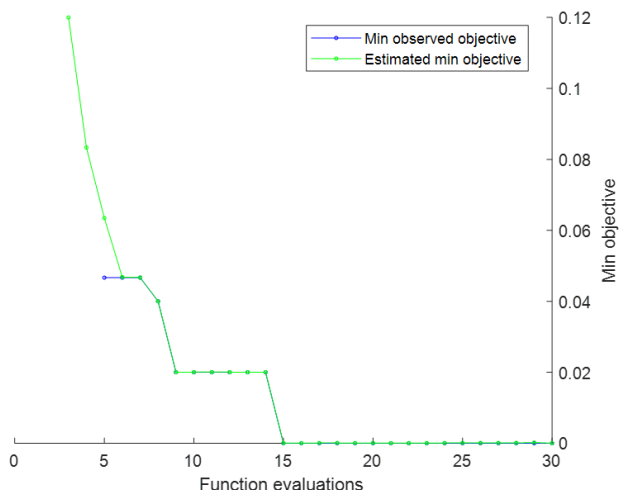


Figure. 12 Min. objective vs. number of function evaluations

Table 2. Comparative results with different features RF classifiers

Parameter	Features Extracted			Hybrid Features
	AR coefficient	Multifrac tal features	SVD Entropy	
Accuracy	0.8889	0.8611	0.9083	98.29%
Error	0.9167	0.8889	0.9083	01.71%
Sensitivity	0.9722	0.963	0.9083	98.93%
Specificity	0.9168	0.8865	0.9083	95.71%
Precision	0.8903	0.8607	0.9083	98.93%
False Positive Rate	0.8889	0.8611	0.9083	4.29%
F-Score	0.9167	0.8889	0.9083	98.93%
Matthews Correlation Coefficient	0.9722	0.963	0.9083	94.64%
Kappa	0.9168	0.8865	0.9083	94.64%

features produces 90.83%,86.11% and 88.89 % accuracy respectively. For the hybrid feature set system model, the average F-score, MCC, and kappa statistics were higher (98.93%, 94.64 %, and 94.64%).Higher F-Score indicates better classification in classifiers.

The proposed framework was also evaluated with different classifiers along with hybrid features for heartbeat classification. For training and testing, the respective beat information was considered to analyze the performance of the proposed model concerning each class of heartbeats. Table 4

Table 3. Comparative results with different features KNN classifiers

Parameter	Features Extracted			Hybrid Features
	AR coefficient	Multifrac tal features	SVD Entropy	
Accuracy	70%	75%	87.5%	94.29%
Error	30%	25%	12.5%	5.71%
Sensitivity	45%	75%	100%	96.43%
Specificity	95%	75%	75%	85.71%
Precision	90%	75%	80%	96.43%
False Positive Rate	50%	25%	25%	14.29%
F-Score	60%	75%	88.9%	96.43%
Matthews Correlation Coefficient	46%	50%	77.5%	82.14%
Kappa	40%	50%	75%	82.14%

Table 4. Comparative results of different classifiers

Parameter	Proposed Method		
	KNN	Bayesian Optimized-KNN	Random Forest Classifier
Accuracy	94.29%	96.57%	98.29%
Error	5.71%	3.43%	01.71%
Sensitivity	96.43%	97.86%	98.93%
Specificity	85.71%	91.43%	95.71%
Precision	96.43%	97.86%	98.93%
False Positive Rate	14.29%	8.57%	4.29%
F-Score	96.43%	97.86%	98.93%
Matthews Correlation Coefficient	82.14%	89.29%	94.64%
Kappa	82.14%	89.29%	94.64%

illustrates the accuracy calculated i.e., 94.29%, 96.57%, and 98.29% for the MIT dataset with five classes in KNN, Bayesian Optimized KNN, and RF classifier respectively. For the RF classifier proposed features provides higher accuracy than at KNN. F-Score of 96.43%, 97.86%, and 98.93% are observed on different classifiers. Overall, by analyzing the statistical measures, we can conclude that the proposed heartbeat framework provides considerably accurate outcomes, even when each is considered.

Authors of [66] has presented a technique on

Table 5. Comparison with previous research works

Research works	Features	Classifier	Accuracy
Yang and Wei, 2020 [66]	Combined parameter and visual pattern features of ECG morphology KNN	KNN	97.70 %
Mondéjar-Guerra et al., 2019 [67]	Wavelets+LBP+HOS +several amplitude values RF	Random forest	94.5%
Mougoufan, et al., 2020 [68]	Permutation entropy (PE) and the conditional entropy of ordinal patterns (CEOP)	LR	93.67 %
Proposed Hybrid approach using KNN	DTCWT, SVD-Entropy, Autoregressive modeling, and Multifractal analysis based features	K-Nearest Neighbors	94.29 %
Proposed Hybrid approach using Bayesian Optimized-KNN	DTCWT, SVD-Entropy, Autoregressive modeling, and Multifractal analysis based features	Bayesian Optimized-K-Nearest Neighbors	96.57 %
Proposed Hybrid approach Random Forest Classifier	DTCWT, SVD-Entropy, Autoregressive modeling, and Multifractal analysis based features	Random Forest Classifier	98.29 %

specific dataset with morphological features tuned for that dataset only using KNN classifier. The approach is carried out in [66] is not generalized and deviation of features may lead to improper results. Authors of paper [67] has proposed the automatic classification of electrocardiograms (ECG) based on the combination of multiple Support Vector Machines (SVMs). The method relies on the time intervals between consequent beats and their morphology for the ECG characterisation. Different descriptors based on wavelets, local binary patterns (LBP), higher order statistics (HOS) and several amplitude values were employed. Instead of concatenating all these features to feed a single SVM model, in this approach different SVM models are defined according to their specific features and decisions of the different models are combined with the product, sum, and majority rules, which creates more complex system. Authors

of [68] proposes permutation entropy (PE) and the conditional entropy of ordinal patterns (CEOP) to Electrocardiogram (ECG) data analysis. It is dependent on threshold formulation which is sensitive to the database. The proposed method proposes a combination of hybrid traits. The present work is concluded to the detection of hybrid features (DTCWT, SVD-Entropy, Autoregressive modeling, and Multifractal analysis) followed by the classification by using random forest, k-nearest neighbors, and Bayesian optimized-KNN classifiers. An estimate of the multiscale wavelet variant based on DTCWT was extracted for each signal along the entire data length and further incorporation with specific features after QRS detection ensures the proposed algorithm is less sensitive to the parameter setting and more dataset independent.

Comparative analysis of the proposed framework with the existing techniques on the benchmark datasets i.e. MIT-BIH Arrhythmia is illustrated in Table 5. The proposed model is a five-class for beat classification framework that was evaluated on the dataset with various feature extraction, where the proposed framework yields an accuracy of 98.29%, 96.57 %, and 94.29% respectively. It shows that in previous studies [66], [67] and [68], models were evaluated classification was performed for the 5-class beats.

#### 4. Conclusion

The main advantage of using computational detection of ECG arrhythmias is that this analysis makes the diagnosis more efficient than conventional analysis. This form of detection is very useful when analyzing long-term electrocardiogram signals, because, in conventional analysis, errors caused by the professional carrying out the analysis may occur. This work aimed to implement an ECG signal classification framework using hybrid features (DTCWT, SVD-Entropy, Autoregressive modeling, and Multifractal analysis), which was applied to the MIT-BIH ECG signal database in order to assess the performance of the detection. The results obtained by the Random forest classifier were quite satisfactory and outperforms the KNN and Bayesian Optimized-KNN.

It was found that the KNN based approach achieves an accuracy of 94.29% while the Bayesian optimization of KNN provides 2.28% of improvement in accuracy i.e. 96.57%. Moreover the Random Forest classifier attains 4% improvement in accuracy compared to the KNN approach and 1.72% of improvement than the Bayesian optimized KNN technique.

Future work may be dedicated to the detection of other types of cardiac arrhythmias, the use of improved versions of classification, as well as the implementation of a prototype for applications in hospital environments.

### Conflicts of interest

The authors declare no conflict of interest.

### Author contributions

The paper conceptualizations, methodology, software, validation, formal analysis, investigation, resources, data collection, writing—original draft preparation, writing—review and editing, visualization, have been done by 1<sup>st</sup> author. The supervision and project administration have been done by 2<sup>nd</sup> author.

### Acknowledgments

The authors would like to express sincere thanks to Research Centre, Department of Electronics and Communication Engineering, SJB Institute of Technology, for providing facilities and support during this research work.

### References

- [1] O. T. Inan, L. Giovangrandi, and G. T. Kovacs, "Robust Neural-Network-Based Classification of Premature Ventricular Contractions Using Wavelet Transform and Timing Interval Features", *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 12, pp. 2507-2515, 2006.
- [2] S. Sahoo, B. Kanungo, S. Behera and, S. Sabut, "Multiresolution Wavelet Transform Based Feature Extraction and ECG Classification to Detect Cardiac Abnormalities", *Measurement*, Vol. 108, pp. 55-66, 2017.
- [3] M. Thomas, M. K. Das, and S. Ari, "Automatic ECG Arrhythmia Classification Using Dual Tree Complex Wavelet Based Features", *AEU-International Journal of Electronics and Communications*, Vol. 69 No. 4, pp. 715-721, 2015.
- [4] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG Beat Classification Using PCA, LDA, ICA and Discrete Wavelet Transform", *Biomedical Signal Processing and Control*, Vol. 8, No. 5, pp. 437-448, 2013.
- [5] Y. Kutlu and D. Kuntalp, "Feature Extraction for ECG Heartbeats Using Higher Order Statistics of WPD Coefficients", *Computer Methods and Programs in Biomedicine*, Vol. 105, No. 3, pp. 257-267, 2012.
- [6] 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.
- [7] T. Li and M. Zhou, "ECG Classification Using Wavelet Packet Entropy and Random Forests", *Entropy*, Vol. 18, No. 8, p. 285, 2016.
- [8] Y. Özbay, R. Ceylan, and B. Karlik, "Integration of Type-2 Fuzzy Clustering And Wavelet Transform in A Neural Network Based ECG Classifier", *Expert Systems with Applications*, Vol. 38, No. 1, pp. 1004-1010, 2011.
- [9] V. H. C. D. Albuquerque, T. M. Nunes, D. R. Pereira, E. J. D. S. Luz, D. Menotti, J. P. Papa, and J. M. R. Tavares, "Robust Automated Cardiac Arrhythmia Detection in ECG Beat Signals", *Neural Computing and Applications*, Vol. 29, No. 3, pp. 679-693, 2018.
- [10] S. N. Yu and K. T. Chou, "Integration of Independent Component Analysis and Neural Networks for ECG Beat Classification", *Expert Systems with Applications*, Vol. 34, No. 4, pp. 2841-2846, 2008.
- [11] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support Vector Machine-Based Expert System For Reliable Heartbeat Recognition", *IEEE Transactions on Biomedical Engineering*, Vol. 51, No. 4, pp. 582-589, 2004.
- [12] R. J. Martis, U. R. Acharya, C. M. Lim, K. M. Mandana, A. K. Ray, and C. Chakraborty, 2013. Application of Higher Order Cumulant Features for Cardiac Health Diagnosis Using ECG Signals. *International Journal of Neural Systems*, 23 (04), p. 1350014.
- [13] H. F. Huang, G. S. Hu, and L. Zhu, "Sparse Representation-Based Heartbeat Classification using Independent Component Analysis", *Journal of Medical Systems*, Vol. 36, No. 3, pp. 1235-1247, 2012.
- [14] R. J. Martis, U. R. Acharya, K. M. Mandana, A. K. Ray, and C. Chakraborty, "Application of Principal Component Analysis to ECG Signals for Automated Diagnosis of Cardiac Health", *Expert Systems with Applications*, Vol. 39, No. 14, pp. 11792-11800, 2012.
- [15] U. R. Acharya, Y. Hagiwara, J. E. W. Koh, S. L. Oh, J. H. Tan, M. Adam, and R. San Tan, "Entropies for Automated Detection of Coronary Artery Disease using ECG Signals: A Review", *Biocybernetics and Biomedical Engineering*, Vol. 38, No. 2, pp. 373-384, 2018.

- [16] Y. C. Yeh, C. W. Chiou, and H. J. Lin, "Analyzing ECG for Cardiac Arrhythmia using Cluster Analysis", *Expert Systems with Applications*, Vol. 39, No. 1, pp. 1000-1010, 2012.
- [17] E. Alickovic and A. Subasi, "Effect of Multiscale PCA De-noising in ECG Beat Classification for Diagnosis of Cardiovascular Diseases", *Circuits, Systems, and Signal Processing*, Vol. 34, No. 2, pp. 513-533, 2015.
- [18] P. Pławiak, "Novel Methodology of Cardiac Health Recognition Based on ECG Signals and Evolutionary-Neural System", *Expert Systems with Applications*, Vol. 92, pp. 334-349, 2018.
- [19] Y. Zhang and Z. Zhao, "Fetal State Assessment based on Cardiotocography Parameters using PCA and Adaboost", In: *Proc. of 10<sup>th</sup> International Congress on Image and Signal Processing, Biomedical Engineering and Informatics (CISP-BMEI)*, pp. 1-6, 2017.
- [20] M. Korürek and B. Doğan, "ECG Beat Classification using Particle Swarm Optimization and Radial Basis Function Neural Network", *Expert Systems with Applications*, Vol. 37, No. 12, pp. 7563-7569, 2010.
- [21] I. Bogdanova, F. Rincón, and D. Atienza, "A Multi-Lead ECG Classification based on Random Projection Features", In: *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 625-628, 2012.
- [22] T. J. Jun, H. M. Nguyen, D. Kang, D. Kim, D. Kim, and Y. H. Kim, "ECG arrhythmia classification using a 2-D convolutional neural network", *Computer Vision and Pattern Recognition, arXiv preprint arXiv:1804.06812*, 2018.
- [23] H. Kutlu and E. Avci, "A Novel Method for Classifying Liver and Brain Tumors Using Convolutional Neural Networks, Discrete Wavelet Transform and Long Short-Term Memory Networks", *Sensors*, Vol. 19, No. 9, p. 1992, 2019.
- [24] H. Kutlu, E. Avci, and F. Özyurt, "White Blood Cells Detection and Classification Based on Regional Convolutional Neural Networks", *Medical Hypotheses*, Vo. 135, p. 109472, 2020.
- [25] J. Tang, C. Deng, and G. B. Huang, "Extreme Learning Machine for Multilayer Perceptron", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 27, No. 4, pp. 809-821, 2015.
- [26] Y. Sönmez, H. Kutlu, and E. Avci, "A Novel Approach in Analyzing Traffic Flow by Extreme Learning Machine Method", *Tehničkivjesnik*, Vol. 26, No. 1, pp. 107-113, 2019.
- [27] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme Learning Machine: Theory and Applications", *Neurocomputing*, Vol. 70, No. (1-3), pp. 489-501, 2006.
- [28] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme Learning Machine for Regression and Multiclass Classification", *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 42, No. 2, pp. 513-529, 2011.
- [29] R. Singh and S. Balasundaram, "Application of Extreme Learning Machine Method for Time Series Analysis", *International Journal of Intelligent Technology*, Vol. 2, No. 4, pp. 256-262, 2007.
- [30] L. D. Tavares, R. R. Saldanha, and D. A. Vieira, "Extreme Learning Machine with Parallel Layer Perceptrons", *Neurocomputing*, Vol. 166, pp. 164-171, 2015.
- [31] E. Avci and R. Coteli, "A New Automatic Target Recognition System Based on Wavelet Extreme Learning Machine", *Expert Systems with Applications*, Vol. 39, No. 16, pp. 12340-12348, 2012.
- [32] M. G. V. Cervantes, A. R. Molina, C. V. G. Mendoza, O. P. Mejia, and G. S. Cervantes, "Multi-Objective On-Line Optimization Approach for the DC Motor Controller Tuning Using Differential Evolution", *IEEE Access*, Vol. 5, pp. 20393-20407, 2017.
- [33] J. Hu, C. Wang, C. Liu, and Z. Ye, "Improved K-means Algorithm Based on Hybrid Fruit Fly Optimization and Differential Evolution", In: *Proc. of 12<sup>th</sup> International Conference on Computer Science and Education (ICCSE)*, pp. 464-467, 2017.
- [34] H. F. Farahani and F. Rashidi, "Optimal Allocation of Plug-In Electric Vehicle Capacity to Produce Active, Reactive and Distorted Powers Using Differential Evolution Based Artificial Bee Colony Algorithm", *IET Science, Measurement & Technology*, Vol. 11, No. 8, pp. 1058-1070, 2017.
- [35] X. Chen, P. Zhang, G. Du, and F. Li, "Ant Colony Optimization Based Memetic Algorithm to Solve Bi-Objective Multiple Traveling Salesmen Problem for Multi-Robot Systems", *IEEE Access*, Vol. 6, pp. 21745-21757, 2018.
- [36] E. Avci, "A New Method for Expert Target Recognition System: Genetic Wavelet Extreme Learning Machine (GAWELM)", *Expert Systems with Applications*, Vol. 40, No. 10, pp. 3984-3993, 2013.

- [37] F. Melgani and Y. Bazi, "Classification of Electrocardiogram Signals with Support Vector Machines and Particle Swarm Optimization", *IEEE Transactions on Information Technology in Biomedicine*, Vol. 12, No. 5, pp. 667-677, 2008.
- [38] V. Gupta and M. Mittal, "KNN and PCA Classifier with Autoregressive Modelling during Different ECG Signal Interpretation", *Procedia Computer Science*, Vol. 125, pp. 18-24, 2018.
- [39] E. S. Williams, D. S. Owens, J. A. Drezner, and J. M. Prutkin, "Electrocardiogram Interpretation in the Athlete", *Herzschrittmachertherapie+Elektrophysiologie*, Vol. 23, No. 2, pp. 65-71, 2012.
- [40] C. C. Lin, H. Y. Chang, Y. H. Huang, and C. Y. Yeh, "A Novel Wavelet-Based Algorithm for Detection of QRS Complex", *Applied Sciences*, Vol. 9, No. 10, p. 2142, 2019.
- [41] A. H. Matamoros, H. Fujita, E. E. Hernandez, H. P. Meana, and M. N. Miyatake, "Recognition of ECG Signals Using Wavelet Based on Atomic Functions", *Biocybernetics and Biomedical Engineering*, Vol. 40, No. 2, pp. 803-814, 2020.
- [42] Y. N. Soe and K. K. K. Oo, "ECG Signal Classification using Discrete Wavelet Transform and Pan Tompkins Algorithm", *International Journal of All Research Writings*, Vol. 2, No. 11, pp. 14-19, 2020.
- [43] S. Sumathi and M. Y. Sanavullah, "Comparative Study of QRS Complex Detection in ECG based on Discrete Wavelet Transform", *International Journal of Recent Trends in Engineering*, Vol. 2, No. 5, p. 273, 2009.
- [44] S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A Survey on ECG Analysis", *Biomedical Signal Processing and Control*, Vol. 43, pp. 216-235, 2018.
- [45] M. Vozda, T. Peterek, and M. Cerny, "Novel Method for Deriving Vectorcardiographic Leads Based on Artificial Neural Networks", In: *Proc. of the 41<sup>st</sup> International Congress on Electromyography and Clinical Neurophysiology, Bratislava, Slovakia*, pp. 4-7, 2014.
- [46] H. A. Deepak and Dr. T. Vijayakumar., "Review of ECG Signal Classification using Deep Learning and Traditional Methods", *International Journal of Scientific & Technology Research*, Vol. 9, No. 3, pp. 5683-5690, 2020.
- [47] O. El B'charri, R. Latif, K. Elmansouri, A. Abenaou, and W. Jenkal, "ECG Signal Performance De-Noising Assessment Based on Threshold Tuning of Dual-Tree Wavelet Transform", *Biomedical Engineering Online*, Vol. 16, No. 1, pp. 1-18, 2017.
- [48] N. Prashar, M. Sood, and S. Jain, "Dual-tree Complex Wavelet Transform Technique-Based Optimal Threshold Tuning System to Deliver Denoised ECG Signal", *Transactions of the Institute of Measurement and Control*, Vol. 42, No. 4, pp. 854-869, 2020.
- [49] K. Blinowska and J. Zygierevicz., "Series in Medical Physics and Biomedical Engineering. Practical Biomedical Signal Analysis Using MATLAB", 1<sup>st</sup> ed. Florida: CRC Press. Taylor & Francis Group, 2021.
- [50] V. Lawhern, W. D. Hairston, K. McDowell, M. Westerfield, and K. Robbins, "Detection and Classification of Subject-Generated Artifacts in EEG Signals Using Autoregressive Models", *Journal of Neuroscience Methods*, Vol. 208, No. 2, pp. 181-189, 2012.
- [51] J. Muthuswamy and N. V. Thakor, "Spectral Analysis Methods for Neurological Signals", *Journal of Neuroscience Methods*, Vol. 83, No. 1, pp. 1-14, 1998.
- [52] L. Sörnmo and P. Laguna, "Bioelectrical Signal Processing in Cardiac and Neurological Applications", *Academic Press*, Vol. 8, 2005.
- [53] E. J. Hannan and B. G. Quinn, "The Determination of the Order of An Autoregression", *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 41, No. 2, pp. 190-195, 1979.
- [54] S. Privara, Z. Vána, E. Žáčková, and J. Cigler, "Building Modeling: Selection of The Most Appropriate Model for Predictive Control", *Energy and Buildings*, Vol. 55, pp. 341-350, 2012.
- [55] J. Pardey, S. Roberts, and L. Tarassenko, "A Review of Parametric Modelling Techniques for EEG Analysis", *Medical Engineering & Physics*, Vol. 18, No. 1, pp. 2-11, 1996.
- [56] B. H. Jansen, J. R. Bourne, and J. W. Ward, "Autoregressive Estimation of Short Segment Spectra for Computerized EEG Analysis", *IEEE Transactions on Biomedical Engineering*, Vol. 9, pp. 630-638, 1981.
- [57] J. L. Semmlow and B. Griffel, "Biosignal and Medical Image Processing", *CRC press*, 2014.
- [58] B. Lashermes, S. Jaffard, and P. Abry, "Wavelet Leader Based Multifractional Analysis", In: *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 4, pp. iv-161, 2005.
- [59] S. M. Shekatkar, Y. Kotriwar, K. P. Harikrishnan, and G. Ambika, "Detecting Abnormality in Heart Dynamics From

- Multifractal Analysis of ECG Signals”, *Scientific Reports*, Vol. 7, No. 1, pp. 1-11, 2017.
- [60] H. Li, T. Liu, X. Wu, and Q. Chen, “Research on Bearing Fault Feature Extraction Based on Singular Value Decomposition and Optimized Frequency Band Entropy”, *Mechanical Systems and Signal Processing*, Vol. 118, pp. 477-502, 2019.
- [61] L. Breiman, “Random forests”, *Machine Learning*, Vol. 45, No. 1, pp. 5-32, 2001.
- [62] I. Saini, D. Singh, and A. Khosla, “QRS Detection Using K-Nearest Neighbor Algorithm (KNN) and Evaluation on Standard ECG Databases”, *Journal of Advanced Research*, Vol. 4, No. 4, pp. 331-344, 2013.
- [63] R. He, K. Wang, Q. Li, Y. Yuan, N. Zhao, Y. Liu, and H. Zhang, “A Novel Method For The Detection of R-Peaks in ECG Based on K-Nearest Neighbors And Particle Swarm Optimization”, *EURASIP Journal on Advances in Signal Processing*, No. 1, pp. 1-14, 2017.
- [64] D. Acuna and P. Schrater, “Bayesian Modeling of Human Sequential Decision-Making on The Multi-Armed Bandit Problem”, In: *Proc. of the 30<sup>th</sup> Annual Conference of The Cognitive Science Society, Washington, DC*, Vol. 100, pp. 200-300, 2008.
- [65] M. Pelikan, D. E. Goldberg, and E. C. Paz, “BOA: The Bayesian Optimization Algorithm”, In: *Proc. of The Genetic and Evolutionary Computation Conference (GECCO-99)*, Vol. 1, pp. 525-532, 1999.
- [66] H. Yang and Z. Wei, “Arrhythmia Recognition and Classification Using Combined Parametric and Visual Pattern Features of ECG Morphology”, *IEEE Access*, Vol. 8, pp. 47103-47117, 2020.
- [67] V. M. Guerra, J. Novo, J. Rouco, M. G. Penedo, and M. Ortega, “Heartbeat Classification Fusing Temporal and Morphological Information of ECGs Via Ensemble of Classifiers”, *Biomedical Signal Processing and Control*, Vol. 47, pp. 41-48, 2019.
- [68] J. B. B. àMougoufan, J. A. E. Fouda, M. Tchunte, and W. Koepf, “Adaptive ECG Beat Classification by Ordinal Pattern Based Entropies”, *Communications in Nonlinear Science and Numerical Simulation*, Vol. 84, p. 105156, 2020.