International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 4 Issue 3, April 2020 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

# An Automated ECG Signal Diagnosing Methodology using Random Forest Classification with Quality Aware Techniques

**Akshara Jayanthan M B[1], Prof. K. Kalai Selvi[2] M.E., MISTE**

[1]PG Student, Department of Communication Systems,
[2]Assistant Professor(RD), Department of Electronics and Communication Engineering,
[1,2]Government College of Engineering, Tirunelveli, Tamil Nadu, India

**ABSTRACT**

In this project, we put forward a new automated quality-aware ECG beat classification method for effectual diagnosis of ECG arrhythmias under unsubstantiated health concern environments. The suggested method contains three foremost junctures: (i) ECG signal quality assessment (ECG-SQA) based whether it is "acceptable" or "unacceptable" based on our preceding adapted complete ensemble empirical mode decomposition (CEEMD) and temporal features, (ii) reconstruction of ECG signal and R-peak detection (iii) the ECG beat classification as well as the ECG beat extraction, beat alignment and Random forest (RF) based beat classification. The accuracy and robustness of the anticipated method is evaluated by means of different normal and abnormal ECG signals taken from the standard MIT-BIH arrhythmia database. The suggested ECG beat extraction approach can recover the categorization accuracy by protecting the QRS complex portion and background noises is suppressed under an acceptable level of noise . The quality-aware ECG beat classification techniques attains higher kappa values for the classification accuracies which can be reliable as evaluated to the heartbeat classification methods without the ECG quality assessment process.

***KEYWORDS:*** *ECG beat classification, ECG arrhythmia recognition, ECG signal quality assessment, Random forest classifier*

***How to cite this paper:*** Akshara Jayanthan M B | Prof. K. Kalai Selvi "An Automated ECG Signal Diagnosing Methodology using Random Forest Classification with Quality Aware Techniques" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-3, April 2020, pp.1174-1179, URL: www.ijtsrd.com/papers/ijtsrd30750.pdf

IJTSRD30750



## I. INTRODUCTION

Accurate and reliable classification of electrocardiogram (ECG) beats is most significant in automatic ECG analysis applications beneath resting, exercise, and ambulatory ECG recording circumstances. Several methods were introduced using various signal processing techniques and classificators. The ECG beat classification system generally consists of three foremost junctures: (i) preprocessing, (ii) feature extraction, and (iii) classification.

The preprocessing stage is commonly designed to suppress background noises using the denoising techniques such as the two median filters, highpass filter (HPF) with cut-off frequency of 1 Hz, bandpass filter through 0.1-100 Hz, morphological filtering, multiscale principal component analysis (MSPCA), wavelet transform, band pass filtering with 5-12 Hz for removal of baseline wander; second-order Butterworth low pass filter (LPF) with 30-Hz cutoff frequency, band pass filtering, 12-tap LPF, MSPCA filtering, morphological filtering, and notch filter. In the past methods, different signal processing techniques were proposed for extracting the features from ECG signals. The features are: temporal morphological features, frequency domain features, wavelet morphological features, Stock well transform (ST) features, Hermite coefficient features, statistical features (time-domain, frequency-domain and time-frequency domain), RR interval features, wavelet cross-spectrum (WCS) and wavelet coherence (WCOH) features and independent component analysis (ICA).

Based on the extracted features, the beat classification was performed using the linear discriminant analysis (LDA), neural network, neuro-fuzzy network, rule-based rough sets, geometric template matching, block-based neural networks (BbNNs), support vector machine (SVM), particle swarm optimization (PSO), multidimensional PSO (MD PSO) based multilayer perceptrons (MLPs), hidden Markov models, mixture of experts with self-organizing maps (SOM) and learning vector quantization (LVQ) algorithms, random forests (RF) classifier, extreme learning machine (ELM) and 1-D convolutional neural networks (CNNs). Patient-specific ECG beat classification approach based on the beat detection, the raw ECG morphology waveform, beat timing information and adaptive 1-D convolutional neural networks (CNNs). The authors observed that there is a significant variation in the system's accuracy and reliability for the larger databases and noisy ECG signals with physiological artifacts and external noises.

Most aforementioned methods include two major steps: heartbeat feature extraction and signal quality grading. For computing the signal quality indexes (SQIs), different time-domain and spectral features, RR-interval and QRS complex-based features, higher-order statistical features are extracted from the processed ECG signal. Some of the methods used a set of decision rules and machine learning approaches to classify the recorded ECG signals into two-four quality groups such as acceptable and unacceptable; acceptable, intermediate and unacceptable; and excellent, very good, good and bad based on the measured SQI values. The limitation of most methods is the accurate and reliable extraction of the ECG morphological features that can be very difficult under time-varying ECG morphological patterns and heart rates.

## II. RELATED WORK

**Zaunsederet, al., (2011)** propose the ECG classification problem make use of a methodology, which can augment classification performance while concurrently reducing the computational resources, making it exceptionally adequate for its application in the progressment of ambulatory settings. For this rationale, the sequential forward floating search (SFFS) algorithm was applied with a new standard function index based on linear discriminants.

**Coimbraet, al., (2012)** establish a new approach for heartbeat classification based on a mixture of morphological and dynamic features. Wavelet transform and independent component analysis (ICA) are applied individually to each heartbeat to extort morphological features. Besides, RR interval information is computed to provide dynamic features. These two dissimilar types of features are concatenated and a support vector machine classifier is make use of for the classification of heartbeats keen on one of 16 classes. The procedure is self-regulatingly applied to the data from two ECG leads and the two decisions are combined for the final classification decision.

**Banerjee et, al., (2014)** put forward a cross wavelet transform (XWT) for the analysis and classification of electrocardiogram (ECG) signals. The cross-correlation flanked by two time-domain signals gives a measure of alike between two waveforms. **The application of** the continuous wavelet transform to two-time series and the cross-examination of the two decompositions expose confined similarities in time and frequency. Relevance of the XWT to a pair of data acquiesces wavelet cross-spectrum (WCS) and wavelet coherence (WCOH). The proposed algorithm examines ECG data utilizing XWT and surveys the resulting spectral differences.

Kiranyazet, al., (2016) presents a simple and reliable classification and monitoring system for patient-specific electrocardiogram (ECG). Methods: An adaptive accomplishment of 1-D convolutional neural networks (CNNs) where feature extraction and classification are obtained by combining the two foremost blocks of the ECG classification into a distinct learning body. Therefore, for each patient, using relatively small common and patient-specific training data, an individual and simple CNN will be trained and thus, such patient-specific feature extraction ability can additionally improve the classification performance. Since this also contradicts the necessity to extort hand-crafted manual features, once a devoted CNN is trained for a exacting patient, it can exclusively be used to classify probably long ECG statistics stream in a fast and accurate manner or alternatively, such a resolution can suitablely use for real-time ECG monitoring and premature alert organization on a light-weight wearable device.

## III. SYSTEM IMPLEMENTATION

In this project, to present a quality-aware ECG beat classification method for unsupervised ECG monitoring applications. It consists of three major stages:

A. The ECG signal quality assessment (ECG-SQA) based on whether it is "acceptable" or "unacceptable" and preceding adapted complete ensemble empirical mode decomposition (CEEMD) and temporal features,
B. The ECG signal reconstruction and R-peak detection and
C. The ECG beat classification including the ECG beat extraction beat alignment and Random Forest (RF) based beat classification. The ECG signal quality assessment was implemented based on the modified CEEMD algorithm and temporal features such as the number of zero crossings (NZC), maximum absolute amplitude (MAA), and short-term NZC envelope as described in our previous work. In the second stage, the acceptable ECG signals are further processed for classifying the ECG beats present in the ECG signal. In the third stage, the heartbeat classification is performed using the RF-based classification similarity metric score which is computed between a test heartbeat template and the reference templates that are stored in the heartbeat database.

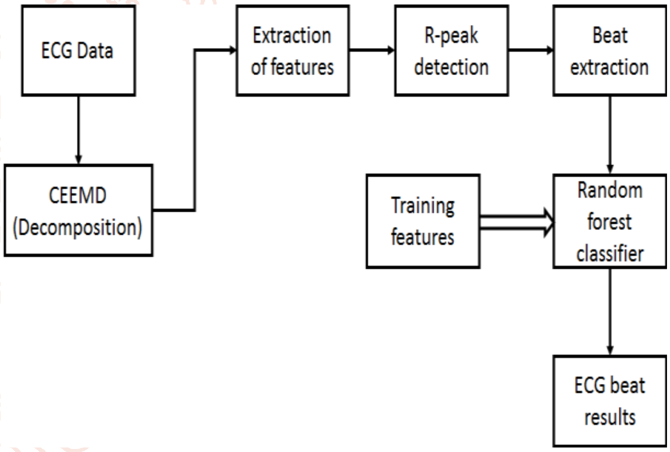



**Fig.1 Proposed system**

A simplified block diagram of the proposed quality-aware ECG beat classification method is illustrated in Fig.3.1 which consists of five steps: modified CEEMD based ECG decomposition, the CEEMD based ECG signal quality assessment, the combined R-peak detection and ECG enhancement, R-peak alignment and the ECG beat extraction and the beat similarity matching by random forest classifier.

### 1. COMPLETE ENSEMBLE EMPIRICAL MODE DECOMPOSITION (CEEMD)

The CEEMD is a data-dependent method of decomposing a signal into some oscillatory components, known as intrinsic mode functions (IMFs). EMD does not make any assumptions about the stationarity or linearity of the data. The aim of EMD is to decompose a signal into a number of IMFs, each one of them satisfying the two basic conditions: 1) the number of extrema or zero-crossings must be the same or differ by at most one; 2) at any point, the average value of the envelope defined by local maxima and the envelope defined by the local minima is zero. Given that we have a signal, the calculation of its IMFs involves the following steps:

1. Identify all extrema (maxima and minima) in $x(t)$.
2. Interpolate between minima and maxima, generating the envelopes $e_l(t)$and $e_m(t)$
3. Determine the local mean as $a(t)=e_m(t)+e_l(t)/2$.
4. Extract the detail i.e., $h(t)=x(t)-a(t)$.
5. Decide whether $h_l(t)$ is an IMF or not based on two basic conditions for IMFs mentioned above.
6. Repeat steps 1 to 4 until an IMF is obtained.

Once the first IMF is obtained, define $c_l(t)=h_l(t)$, which is the smallest temporal scale in $x(t)$. A residual signal is obtained as $r_l(t)=x(t)-c_l(t)$. The residue is treated as the next signal and the above-mentioned process is repeated until the final residue is a constant (having no more IMFs). At the end of the decomposition, the original signal can be represented as follows:

$$x(t) = \sum_{m=1}^{M} c_m(t) + r_M(t)$$

where M is the number of IMFs, $c_m(t)$is the th IMF and $r_M(t)$is the final residue.

### Analytic Representation of IMFs

After IMFs have been extracted from EEG signals their analytical representation is obtained. This representation eliminates the DC offset from the signal spectral portion, which is an essential part of compensating for the non-stationary nature of the signals. Given that we have an IMF $c_m(t)$, its analytic representation is given as,

$$y(t)=c_m(t)+iH\{c_m(t)\}$$

where$H\{c_m(t)\}$is the Hilbert transform of $c_m(t)$, which is the $m$th IMF extracted from the signal $x(t)$. After performing EMD of the signal, the IMFs are used for feature extraction purposes.

## 2. FEATURE EXTRACTION

The rationale of the feature extraction process is to choose and retain appropriate information from the original signal. The Feature Extraction stage extracts analytical information from the ECG signal. In order to discover the peaks, specific details of the signal are elected. In feature extraction, detection of the R peak is the first step . The R peak in the Modified Lead II (MLII) lead signal has the highest amplitude of all waves compared to other leads. The QRS complex recognition consists of the influential R point of the heartbeat, which is, in general, the point where the heartbeat has the highest amplitude. A normal QRS complex designates that the electrical impulse has progressed usually from the bundle of His to the Purkinje network through the right and left bundle branches and that the right and left ventricles normal depolarization occurs. Most of the energy of the QRS complex lies among 3 Hz and 40 Hz. The 3-dB frequencies of the Fourier Transform of the wavelets designate that most of the energy of the QRS complex lies among scales of 23 and 24, with the largest at 25. The energy decreases if the scale is larger than 25. The energy of motion objects and baseline wander (i.e., noise) enlarges for scales superior than 25. Therefore, we decide to use distinctive scales of 21 to 25 for the wavelet. In the anticipated algorithm ECG signal is squared after eradicating noise (e. g .baseline wander) and decomposed up to level 5 using Db 4 wavelet thus extrication approximate and detail coefficients. Then inverse Discrete Wavelet transform is applied to recreate the signal inexactly. Then number of QRS complex wavelet transform features was extorted by selecting a window of -300ms to +400ms about the R wave as found in the database annotation. The 252-illustration vectors were downsampled to 21, 25, 31, 42 or 63 samples (corresponding to 12$\boldsymbol{x}$, 10$\boldsymbol{x}$, 8$\boldsymbol{x}$, 6$\boldsymbol{x}$, 4$\boldsymbol{x}$ decimation, respectively), and normalized to a mean of zero and standard deviation of unity. This reduced the DC offset and eradicated the amplitude variance since file to file. QRS width is computed from the onset and the offset of the QRS complex. The onset is the inauguration of the Q wave and the offset is the finale of the S wave. Normally, the onset of the QRS complex consists the high-frequency components, which are recognised at finer scales.

### Temporal Statistic features

Researchers have shown that IMF's statistical features are useful in distinguishing between normal and abnormal EEG signals. Its use is driven by the fact that the sample distribution in the data is characterized by its asymmetry, dispersion and concentration around the mean. A visual examination of the IMFs collected from healthy patients and patients with epilepsy during interictal and ictal cycles after Hilbert transforms shows that they are very different. Ironically, using the IMF data, certain variations are correctly recorded. For an IMF, these statistics can be obtained by the following quantities:

$$\mu_t = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$\sigma_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \mu_t)^2}$$

$$\beta_t = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i - \mu_t}{\sigma_t}\right)^3$$

Where $N$ is the number of samples in the IMF$\mu t$is the mean, $\sigma_t$ is the variance and $\beta_t$ is skewness of the corresponding IMF.

## 3. R-PEAK DETECTION

A simple and robust automated algorithm for the detection of R-peaks of a long-term ECG signal. Figure 3.2 shows a block diagram of our R-peak detection algorithm that consists of the following steps:

- ➢ Bandpass Filtering and Differentiation
- ➢ New Nonlinear Transformation
- ➢ New Peak-Finding Technique
- ➢ Finding Location of True R-Peaks.

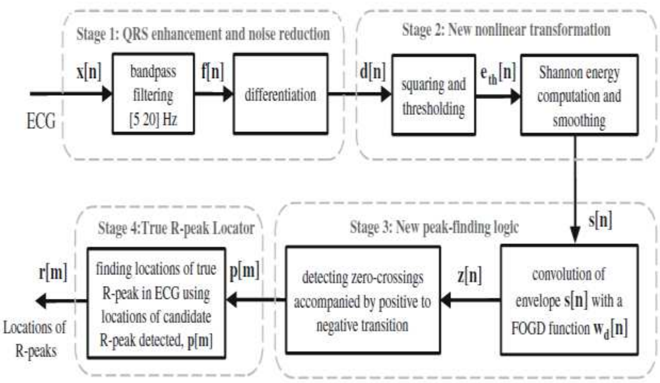
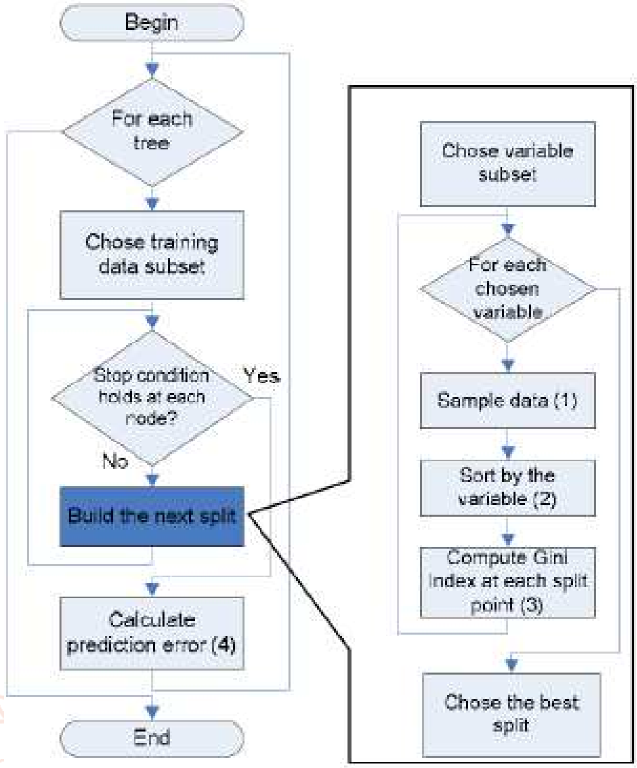



**Fig.2 R-Peak detection algorithm**

The detection algorithm contains of four stages. In the first point, band pass filtering and differentiation is used to boost QRS complexes and reduce out - of-band noise. In the second stage to obtain a positive-valued feature signal which comprises large candidate peaks corresponding to the QRS complex regions a new nonlinear transformation basis on energy thresholding, Shannon energy computation, and smoothing processes was introduced. The energy thresholding minimises the effect of spurious noise spikes as of muscle artifacts. The Shannon energy transformation amplifies average amplitudes and outcomes in small deviations among successive peaks. Therefore, the anticipated nonlinear transformation is capable of minimizing the number of false positives and false-negatives under small-QRS and wide-QRS complexes and noisy ECG signals. A simple peak-finding strategy based on the first-order Gaussian differentiator (FOGD) is proposed in the third stage that accurately identifies locations of candidate R-peaks in a feature signal. This juncture computes the convolution of the smooth feature signal and FOGD operator. The resultant convolution output has the candidate peaks of feature signal , negative zero-crossings (ZCs) suitable to the anti-symmetric nature of the FOGD operator. Thus, these negative ZCS are perceived and used as channels to find locations of real R-peaks in an original signal at the fourth stage.

**Flowchart:**




## 4. RANDOM FOREST CLASSIFICATION ALGORITHM

Random Forest is a popular machine learning algorithm used for several types of classification tasks. A Random Forest is a tree-structured classificator ensemble. That forest tree gives a unit vote which assigns that input to the most likely class label. It is a fast method, robust to noise and it is a successful ensemble that can identify non-linear patterns in the data. It can handle numeric as well as categorical data easily. One of the major advantages of Random Forest is that it does not suffer from over fitting, even if more trees are appended to the forest.

**Each tree is constructed using the following algorithm:**

1. Let N and M the number of training cases and the number of variables in the classifier
2. m the number of input variables to be used to determine the decision at a node of the tree; m should be much less than *M*.
3. Prefer a training set for this tree by choosing n times with replacement from all N offered training cases (i.e. take a bootstrap sample). Use the rest of the cases to approximation the error of the tree, by envisaging their classes.
4. For each node of the tree, at random prefer m variables on which to base the decision at that node. Calculate the best split anchored in these m variables in the training set.
5. Each tree is entirely developed and not shortend (as may be done in constructing a normal tree classifier).

For prophecy, a new sample is short of down the tree. The label of the training sample is assigned in the terminal node it ends up in. This procedure is iterated over all trees in the collection, and the average vote of all trees is statemented as random forest prediction.

### A. Improved-RFC approach

Improved-RFC approach uses a Random Forest algorithm, an evaluator attribute method and a process-Resample instance filter. The method aims to increase the classification accuracy for multi-class classification problems of the Random Forest algorithm.

### B. Algorithm of improved-RFC approach

The pseudo-code of the improved-RFC approach is given below.

**Algorithm1.** Improved-Random Forest classifier

- **Input:** DTrain = {x1,x2 . . .xn} // Training dataset which consists of a It runs efficiently on large databases.
- Thousands of input variables can be managed without variable deletion..
- This gives estimates of the essential variables in the classification.
- This produces an internal objective generalization error calculation as forest development progresses.
- Where a significant proportion of the data is incomplete, it has an efficient method for estimating incomplete data and preserves accuracy.

set of training examples and their linked class labels.

**Output:** classification-accuracy A.

**Method:**

Step 1 : Select an attribute evaluator method and apply it on training dataset-Dtrain to obtain a subset of attributes Am.

Step 2 : Apply instance filter-Resample for Am of Dtrain and obtain Dtrain-resample.

Step 3 : Select a Random Forest classification algorithm on Dtrain-resample and obtain classification accuracy A

Step 4 : Output classification-accuracy A.

**The advantages of the random forest are:**

- It is one of the most accurate learning algorithms available. For several data sets, it generates a highly accurate classifier.

## IV. SIMULATION RESULTS& DISCUSSION

The following figure represents the sampled ECG signal data tested with this proposed work.

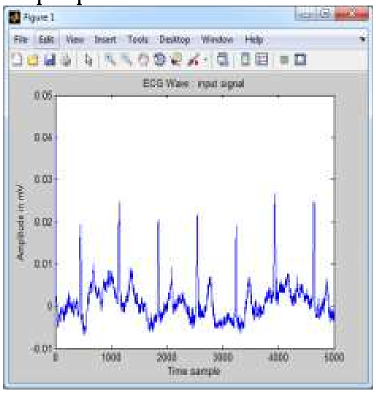



**Fig.3 ECG wave – Input signal**

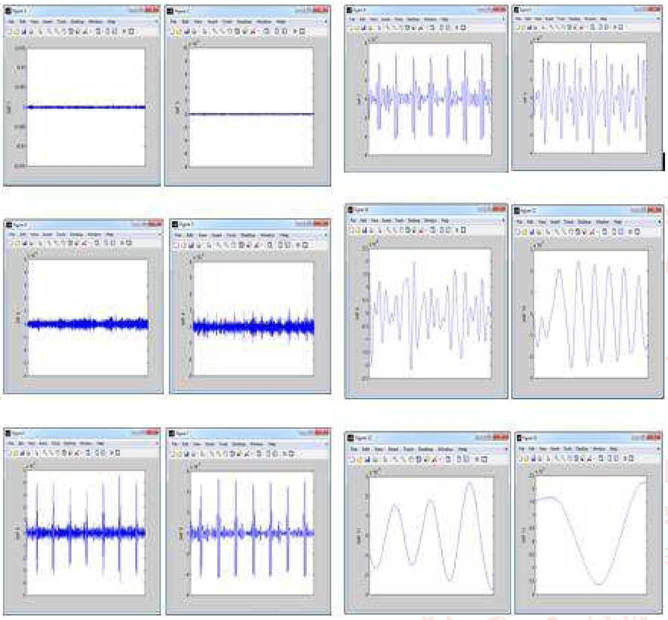


**Fig.4 ECG at 1 to 12thlevel decomposition**

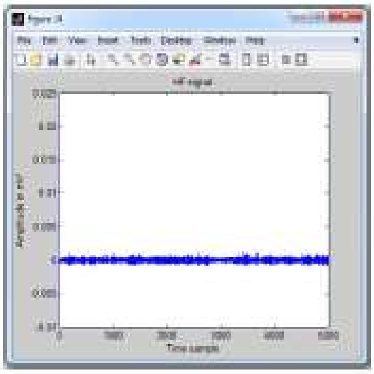



**Fig.5 HF- High-frequency signal**

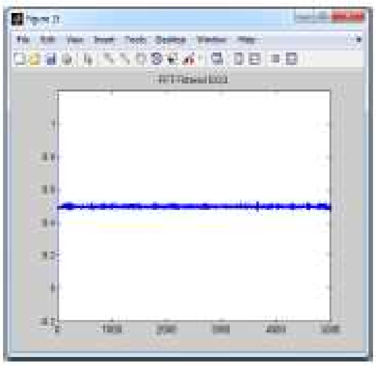


**Fig.6 Filtered ECG signal**

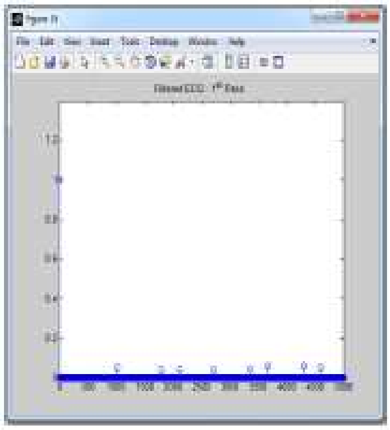


**Fig.7 Filtered ECG signal -1st pass**

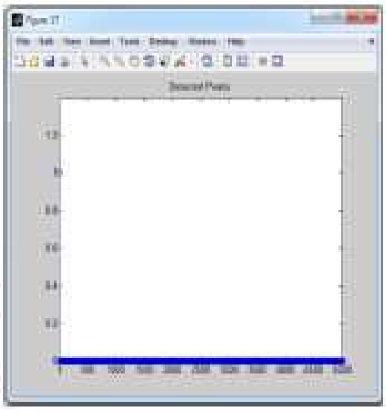



**Fig.8 Detected peak signal**


**Fig.9 Classifier result**

## CONCLUSION

In this project, we present a new quality-aware ECG beat classification method that can be capable of reducing the false alarms and ensuring the consistency of class-specific accuracies for the four classes of heartbeats under noisy ECG recordings. Evaluation results on the standard MIT-BIH arrhythmia database demonstrate that the preservation of QRS complexes is most essential for improving the beat classification when the denoising process is applied for suppression of background noises. Classification results show that the proposed random forest heartbeat classification method improves the consistency with improved classification accuracy and F1-score. For each of the heartbeat classes, the proposed and existing heartbeat classification methods had significant improvement in the false alarm reduction (FAR). Results further demonstrate that a quality-aware ECG analysis system is most essential to ensure the accuracy and reliability of the diagnosis of different types of arrhythmias under noisy ECG recording environments.

## REFERENCES


[1] J. Wannenburg, R. Malekian and G. P. Hancke, “Wireless capacitive based ECG sensing for feature extraction and mobile health monitoring,” IEEE Sensors J., vol. 18, no. 14, pp. 6023-6032, July 2018.

[2] E. Span, S. Di Pascoli and G. Iannaccone, “Low-power wearable ECG monitoring system for multiple-patient remote monitoring,” IEEE Sensors J., vol. 16, no. 13, pp. 5452-5462, July 2016.

[3] D. Labate, F. L. Foresta, G. Occhiuto, F. C. Morabito, A. Lay-Ekuakille and P. Vergallo, “Empirical mode decomposition vs. wavelet decomposition for the extraction of respiratory signal from single-channel ECG: A Comparison,” IEEE Sensors J., vol. 13, no. 7, pp. 2666-2674, July 2013.

[4] P. de Chazal, M. O’Dwyer, and R. B. Reilly, “Automatic classification of heartbeats using ECG morphology and heartbeat interval features,” IEEE Trans. Biomed. Eng., vol. 51, no. 7, pp. 1196-1206, July 2004.

[5] P. de Chazal, R. B. Reilly, “A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval feature,” IEEE Trans. Biomed. Eng., vol. 53, pp. 2535-2543, 2006.

[6] T. Mar, S. Zaunseder, J. P. Martnez, M. Llamedo and R. Poll, “Optimization of ECG classification by means of feature selection,” IEEE Trans. Biomed. Eng., vol. 58, no. 8, pp. 2168-2177, Aug. 2011.

[7] C. C. Lin, and C. M. Yang, “Heartbeat classification using normalized RR intervals and morphological features,” Mathematical Problems in Engineering, 2014.

[8] Q. Li, C. Rajagopalan, and G. D. Clifford, “Ventricular fibrillation and tachycardia classification using a machine learning approach,” IEEE Trans. Biomed. Eng., vol. 61, no. 6, pp. 1607-1613, June 2014.

[9] M. K. Das, and S. Ari, “Patient-specific ECG beat classification technique,” Healthcare Technol. Lett., vol. 1, no. 3, 2014.

[10] J. Kim, S. D. Min, and M. Lee, “An arrhythmia classification algorithm using a dedicated wavelet adapted to different subjects,” Biomed. Eng. Online, vol. 10, no. 1, 2011.

[11] E. Alickovic, and A. Subasi, “Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier,” J. Medical Systems, vol. 40, no. 4, pp. 1-12, 2016.

[12] C. Ye, B.V.K. Vijaya Kumar, and M.T. Coimbra, “Heartbeat classification using morphological and dynamic features of ECG signals,” IEEE Trans. Biomed. Eng., vol. 59, no. 10, pp.2930-2941, Oct. 2012.