Feature Extraction and Classification for ECG Signal Processing based on Artificial Neural Network and Machine Learning Approach

Stalin Subbiah[1], Rajkumar Patro[2] , Subbuthai P[3]

[1]Research Scholar, Research and Development Center, Bharathiar University, Coimbatore,

[2]Associate Professor, Department of Electronics, Guru Nanak Engineering College, Andhrapradesh,

[3]Research Scholar, Department of Electronics and Instrumentation, Bharathiar University, Coimbatore.

Abstract– *In present day, several types of developments are carried toward the medical application. There has been increased improvement in the processing of ECG signals. Get the accurate detection of ECG signals with the help of detection of P, Q, R and S waveform. However these waveforms are suffered from some disturbances like noise. Initially denoising the ECG signal using filters and detect the PQRS waveforms. ECG signal is analyzed or classify using Extreme Learning Machine (ELM) and it compared with Support Vector Machine (SVM) and Back Propagation Neural Network (BPN). The paper classifies the ECG signal into two classes, Normal and Abnormal. ECG waveform is detected and analyzed using the 48 records of the MIT-BIH arrhythmia database. The classifier performance is measured in terms of Sensitivity (Se), Positive Predictivity (PP) and Specificity (SP).*

Keywords: Electrocardiogram, Extreme Learning Machine, Support Vector Machine, Back Prorogation Neural Network, MIT-BIH arrhythmia database.

I   INTRODUCTION

ECG is abbreviated as Electrocardiogram and it is used to represent the electrical activity of the heart. It means that it showing the heart muscle contraction and relaxation. Electrocardiography is another term is used to give or provide the heart condition. Analysis of the ECG waveform is used to identify the heart normal and abnormalities. Basically ECG waveform has some basic waves they are P, Q, R, S, T and U. These waveforms are used to analyses the heart condition. The main important one in ECG is depolarization and repolarization. Atrial depolarization is represented by P waveform and ventricular depolarization is represented by QRS complex waveform and then repolarization of the ventricle is represented by T waveform [1].

There are several existing methods for ECG waveform analysis and provides the system with accuracy and sensitivity. These methods are based on various techniques that are wavelet, RBF Neural network, fuzzy with clustering technique, machine learning of SOM and autoregressive modeling [2, 3, 4, 5 and 6]. There are several techniques are analyzed the ECG signal and it is described in [7-11].

Combination of recurrent neural networks and eigenvectors are used to analyze the ECG signal [12]. Another combination of novel method for analyzed the ECG signal using particle swarm optimization and radial basis function neural network [13]. Apart from MIT-BIH database, ECG signal is collected from European ST-T database and it is described in [14] for normal and abnormal ECG, ECG beat segmentation technique is introduced in [15]. ECG noise removal by sparse derivatives is explained in [16].

The paper is organized as follows. In Section 2, describes about some related work. In Section 3, the proposed ECG algorithm, In Section 4, we show how the classification method helps increase the overall detection accuracy. The system is applied to the whole MIT-BIH Arrhythmia database and the performance is compared to some other state-of-art methods.

II   RELATED WORK

In [18] ECG signal is analyzed using cluster analysis method, this technique consists of three major steps they are extracting the QRS waveform, second stage is selecting qualitative features and the third stage is determining heartbeat case. This method analyses and classified the normal and abnormal heartbeat. In [17] ECG signal noise reductions are carried through the novel approach of combination of discrete wavelet transform and artificial neural network. The work of the wavelet transform is decomposes the ECG signal and remove the noise then the second stage of the artificial neural network is implemented the inverse transform and adaptive filtering for remove the remaining noise. In [19], author presented a KNN algorithm as a classifier for detection of QRS complex. In [20], author has presented an algorithm for detection of R-peak. In [21], authors have used three type of classification that are (i) Back Propagation Network (BPN), (ii) Feed Forward Network (FFN) and (iii) Multilayered Perceptron (MLP)

III   METHODOLOGY

Block diagram of proposed framework is shown in figure 1. The figure provides that the whole framework is divided into three stages that are preprocessing, peak detection, heart beat classification. Input signal are picked from MIT-BIH arrhythmia database s given to preprocessing. Preprocessing is used to denoised the ECG signal and this process is given as an input to the next stage.
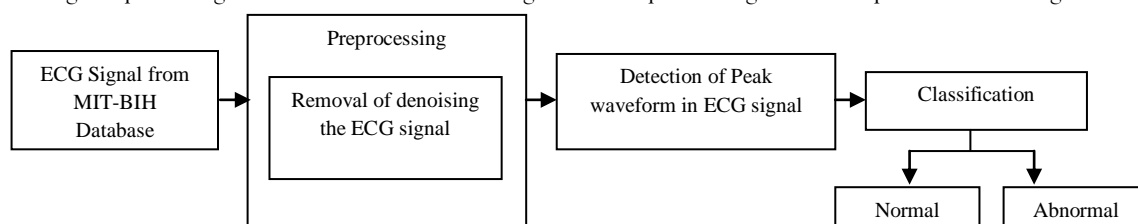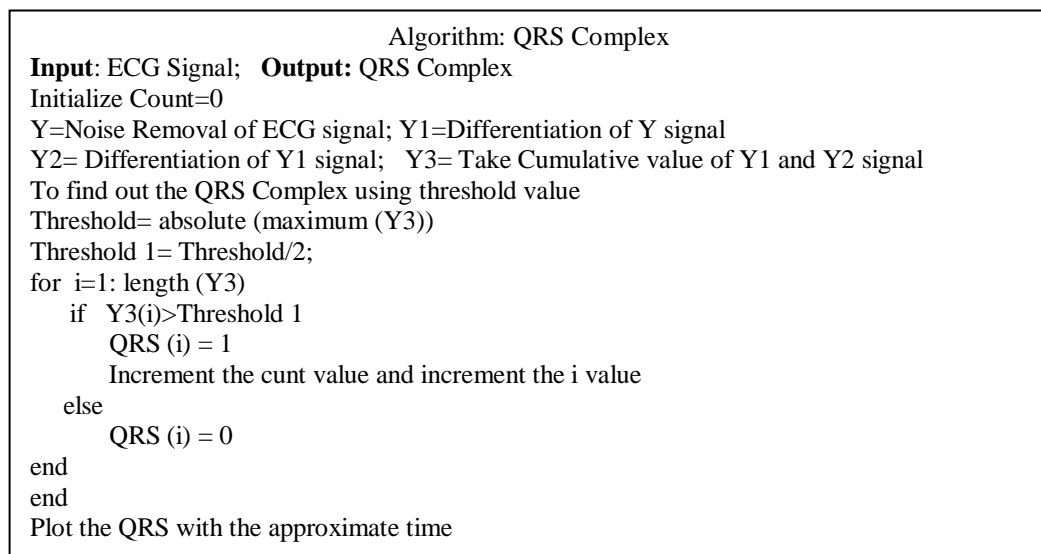

Figure. 1 Block Diagram of proposed Framework

```
                          Algorithm: QRS Complex
Input: ECG Signal;   Output: QRS Complex
Initialize Count=0
Y=Noise Removal of ECG signal; Y1=Differentiation of Y signal
Y2= Differentiation of Y1 signal;   Y3= Take Cumulative value of Y1 and Y2 signal
To find out the QRS Complex using threshold value
Threshold= absolute (maximum (Y3))
Threshold 1= Threshold/2;
for  i=1: length (Y3)
    if   Y3(i)>Threshold 1
        QRS (i) = 1
        Increment the cunt value and increment the i value
    else
        QRS (i) = 0
end
end
Plot the QRS with the approximate time
```

The next stage of the proposed framework is peak detection and QRS complex, before QRS complex ECG signal is decomposed used Wavelet transform. It is one of the powerful techniques in biomedical signal processing. QRS complex features are given as input to classifier and identified the heartbeat as a normal and abnormal. ECG signal are classified by BPN, SVM and ELM. Compared to BPN and SVM, ELM provides the simpler implementation, learning speed is fast and provide better performance

A.   Preprocessing

Preprocessing is one of the important tasks in signal and image processing. This is the first step for proposed ECG classification. The work of the preprocessing is to eliminate the noise in the input ECG signal using various filters approaches. Proposed work handled median filter, FIR filter, Gaussian filter and Butterworth filter for noise removal in ECG signal [23]. Preprocess result is used to get the better efficiency of the ECG signal. Peak detection efficiency is increased due to preprocessing work. Various filters approaches used to remove the noise of baseline wander, power interference, and instrumental error. Elimination of baseline wander is therefore needed in the ECG signal analysis to diminish the irregularities in beat morphology.

B.   Peak detection

ECG signal is extracted using wavelet decomposition; this is done by Daubechies6 (DB6) multiresolution wavelet. First denotes the wavelet decomposition vector and their wavelet bookkeeping vector. Then reconstruct the signal based on wavelet decomposition. ECG signal is decomposed up to eight levels. This type of decomposition is used to extract the features and helpful for the selection of P, Q, R and S wave. Then all the peaks are identified by minimum and maximum value. Figure 2 illustrates the algorithm for QRS complex. Using QRS complex, in this paper find out the heart beat and these beats are given to the classification to deliver the classify output as normal and abnormal.

C. Classification

Feature extraction output is given as an input to classifier. In this paper three types of classification are carried over to classify as normal and abnormal heart beat that are BPN, SVM, and ELM.

1) Back Propagation Neural Network (BPN)

Given a finite length input patterns $X_1(k), X_2(k), \ldots, X_n(k) \in R, (1 \leq k \leq k)$ and the desired patterns $X_1(k), X_2(k), \ldots, X_m(k) \in R$.

Step 1: Select the total number of layers M, the number $n_i (i = 1,2, \ldots, M-1)$ of the neurons in each hidden layer, and an error tolerance parameter $\varepsilon > 0$.

Step 2: Randomly select the initial values of the weight vectors $W_{aj}^{(i)}$ for $i = 1,2, \ldots, n_i$.

Step 3: Initialization:

$$W_{aj}^{(i)} \leftarrow W_{aj}^{(i)}(0), E \leftarrow 0, k \leftarrow 1 \tag{1}$$

Step 4: Calculate the neural outputs

$$\begin{cases} s_j^{(i)} = \left(W_{aj}^{(i)}\right)^T X_a^{(i-1)} \\ X_j^{(i)} = \sigma\left(S_j^{(i)}\right) \end{cases} \tag{2}$$

For $i = 1,2, \ldots, M$ and $j = 1,2, \ldots, n_i$

Step 5: Calculate the output error

$$e_j = d_j - X_j^{(M)} \tag{3}$$

For $j = 1,2, \ldots, m$

Step 6: Calculate the output deltas

$$\delta_j^{(M)} = e_j \sigma'\left(s_j^{(M)}\right) \tag{4}$$

Step 7: Recursively calculate the propagation errors of the hidden neurons

$$e_j^{(i)} = \sum_{l=1}^{n_i+1} \delta_l^{(i+1)} W_{lj}^{(i+1)} \tag{5}$$

From the layer $M-1, M-2, \ldots,$ to layer 1.

Step 8: Recursively calculate the hidden neural delta values.

$$\delta_j^{(i)} = e_j \sigma'\left(s_j^{(i)}\right) \tag{6}$$

Step 9: Update weight vectors

$$W_{aj}^{(i)} = W_{aj}^{(i)} + \eta \delta_j^{(i)} X_a^{(i-1)} \tag{7}$$

Step 10: Calculate the error function

$$E = E + \frac{1}{k}\sum_{j=1}^m e_j^2 \tag{8}$$

Step 11: if k=K then go to step 12; otherwise, $k \leftarrow k+1$ and go to step 4.
Step 12: if $E \leq \varepsilon$ then go to step 13; otherwise go to step 3.
Step 13: learning is completed. Output the weights

2) Support Vector Machine (SVM)

Support Vector Machine (SVM) is one of the learning system and it is used mainly in classification. It was developed in the year of 1998 by Vapnik and it is one of the most techniques to solve the supervised classification problem. In essence, SVM classifiers maximize the margin between training data and the decision boundary (optimal separating hyperplane), which can be formulated as a quadratic optimization problem in a feature space. The subset of patterns those are closest to the decision boundary are called support vectors.

Consider a set of training examples $(x_1, y_1), \ldots, (x_l, y_l)$, here input $x_i \in R^N$ and class labels $y_i \in \{R^N\}$ and class labels $y_i \in \{-1, +1\}$. For a linearly separable classification problem, the construction of a hyperplane is $w^T x + b = 0$ so that the margin between the hyperplane and the nearest point is maximized and can be posed as the following quadratic optimization problem:

$$\min_w \frac{1}{2}(w^T w) \tag{9}$$

Subject to $y_i\left((w^T x_i) + b \geq 1\right),$ where $i = 1, \ldots, l$

Based on below equation, forces a rescaling on (w, b) so that the point nearest to the hyperplane has a distance of $(1/\|w\|)$ [24].

In many practical situations, a separating hyperplane does not exist. To allow the possibilities of violating (2), slack variables ξ_i are introduced like

$$\xi_i \geq 0, i = 1, \dots, l \tag{10}$$

To get

$$y_i\big((w^T x_i) + b\big) \geq 1 - \xi_i, i = 1, \dots, l \tag{11}$$

The optimization problem now becomes as follows:

$$\min_{w,\xi} \frac{1}{2}(w^T w) + C \sum_{i=1}^{l} \xi_i \tag{12}$$

subject to constraints (10) and (11). The C is a user defined constant. It is called regularizing parameter and determines the balance between the maximization of the margin and the minimization of the classification error.

By introducing Lagrange multipliers $\propto_i$ and using Karush–Kuhn–Tucker theorem of optimization theory, the solution is given by;

$$w = \sum_{i=1}^{l} y_i \propto_i x_i \tag{13}$$

Only a small fraction of the $\propto_i$ coefficients are nonzero. The corresponding pairs of x_i entries are known as support vectors and they fully define the decision boundary. All other training examples with corresponding zero $\propto_i$ values are now rendered irrelevant and automatically satisfy constraint (4) with $\xi_i = 0$. The hyperplane decision function for the vector x can be written as follows

$$f(x) = \text{sgn}\Big[\sum_{i=1}^{l} y_i \propto_i (x^T x_i) + b\Big] \tag{14}$$

By replacing the inner product $(x^T x_i)$ with kernel function $K(x, x_i)$; the input data are mapped to a higher dimensional space [24]. It is then in this higher dimensional space that a separating hyperplane is constructed to maximize the margin.

3) Extreme Learning Machine

Extreme Learning algorithm use a finite number of input and outputs for training in supervised batch learning system. In this system consider N arbitrary distinct samples $(X_i, t_i) \in R^n \times R^m$, in this X_i is an $n \times 1$ input vectors and t_i is an $m \times 1$ target vector. If an SLFN with $\widetilde{N}$ hidden nodes can approximate these N samples with zero error, it then implies that there exist β_i, a_i and b_i such that

$$f_{\widetilde{N}}(X_j) = \sum_{i=1}^{\widetilde{N}} \beta_i G(a_i, b_i, X_j) = t_j, j = 1, \dots, N \tag{15}$$

The above equation can be written as

$$H\beta = T \tag{16}$$

Where

$$
\begin{aligned}
&H(a_1, \dots, a_{\widetilde{N}}, b_1, \dots, b_{\widetilde{N}}, X_1, \dots, X_N) \\
&= \begin{bmatrix} G(a_1, b_1, X_1) & \dots & G(a_{\widetilde{N}}, b_{\widetilde{N}}, X_1) \\ \vdots & \dots & \vdots \\ G(a_1, b_1, X_N) & \dots & G(a_{\widetilde{N}}, b_{\widetilde{N}}, X_N) \end{bmatrix}_{N \times \widetilde{N}}
\end{aligned} \tag{17}
$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\widetilde{N}}^T \end{bmatrix}_{\widetilde{N} \times m} \text{ and } T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \tag{18}$$

H is called the hidden layer output matrix of the network [14]; the ith column of H is the ith hidden node's output vector with respect to inputs $X_1, X_2, \dots, X_N$ and the jth row of H is the output vector of the hidden layer with respect to input X_j.

In real applications, the number of hidden nodes, $\widetilde{N}$, wil always be less than the number of training samples, N, and, hence, the training error cannot be made exactly zero but can approach a nonzero training error $\in$. The hidden node parameters a_i and b_i (input weights and biases or centers and impact factors) of SLFNs need not be tuned during training and may simply be assigned with random values according to any continuous sampling distribution [9], [10], [11]. Linear system and the output weights β are estimated as

$$\hat{\beta} = H^+ T \tag{19}$$

Where H^+ the Moore-Penrose is generalized inverse [15] of the hidden layer output matrix H. The ELM algorithm which consists of only three steps, can then be summarized as

ELM Algorithm: Given a training set

$\aleph = \{(X_i, t_i) | X_i \in R^n, t_i \in R^m, i = 1, \dots, N\}$, activation function $g(x)$, and hidden node number $\widetilde{N}$.

STEP 1: Assign random hidden nodes by randomly generating parameters (a_i, b_i) according to any continuous sampling distribution, $i = 1, \dots, \widetilde{N}$.

STEP 2: Calculate the hidden layer output matrix H.

STEP 3: Calculate the output weight β such as $\beta = H^+ T$

IV   EXPERIMENTAL RESULT

The proposed framework results are carried out in MATLAB. The input for proposed work is collected from the MIT-BIH arrhythmia database, and it gives the classification output as Normal and Abnormal. Totally 48 patient records are collected from that database. In this 50% are used for training and another 50% is used for training. The classification is done by three methods that are Back Propagation Neural Network (BPNN), Support Vector Machine (SVM) and Extreme Learning Machine (ELM). In this paper

classification performance are evaluated using some metrics that are Sensitivity, Positive Predictivity and Specificity. The following equations are used to calculated these metrics

$$Se(\%) = \frac{TP}{TP+FN} \times 100 \qquad (20)$$

$$PP(\%) = \frac{TP}{TP+FP} \times 100 \qquad (21)$$

$$Sp(\%) = \frac{TN}{TN+FP} \times 100 \qquad (22)$$

In the above three equation TP denotes the number of true positive samples, FN indicates the number of false negative samples, TN denotes the number of true negative samples and FP indicates the number of false positive samples. These TP, TN, FP and FN are used for classification and it is defined as
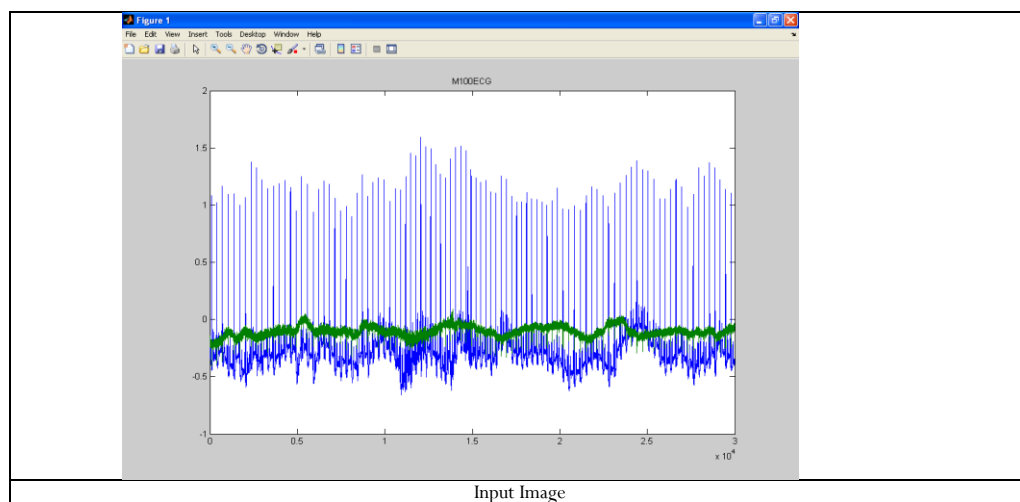
FP: Normal class classifies as abnormal; TP: Abnormal class classifies as abnormal; FN: Abnormal classifies as normal.

TN: Normal class classifies as normal. Then Overall classification accuracy is evaluated using below equation

$$Overall\ Accuracy\ (\%) = \frac{Correctly\ classified\ samples}{Total\ Number\ of\ samples} \qquad (23)$$

TABLE 1

OVERALL PERFORMANCE OF ELM, SVM AND BPN

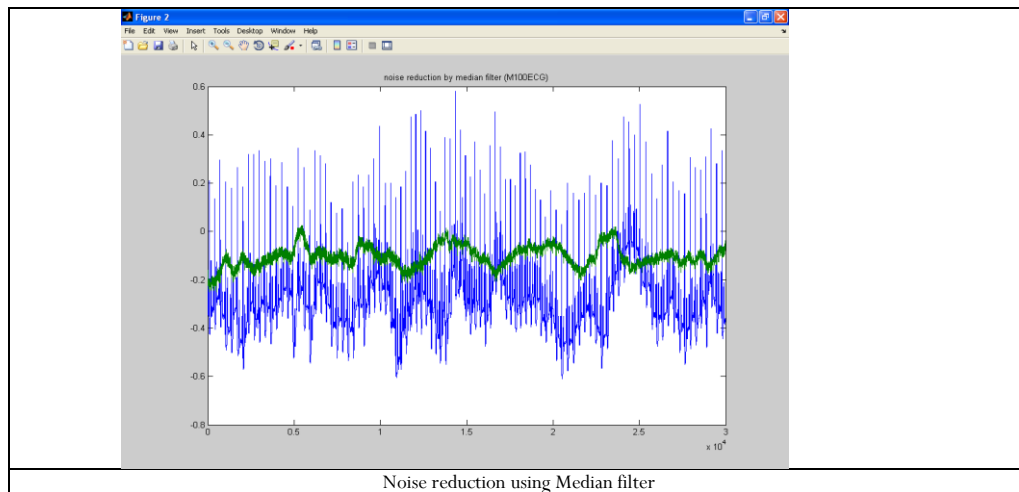| Methods	Proposed Targets	Normal	Abnormal	Se (%)	PP(%)	Sp(%)	Accuracy
ELM	Normal class	24	1				
	Abnormal class	0	20	100	95	96	97
	Total	24	21				
SVM	Normal class	22	3				
	Abnormal class	2	18	90	86	88	73%
	Total	24	21				
BPN	Normal class	20	5				
	Abnormal class	2	18	90	78	72	64%
	Total	22	23				

Input Image

Noise reduction using Median filter

Figure 3: (a) Input image (b) Noise Reduction using Median filter
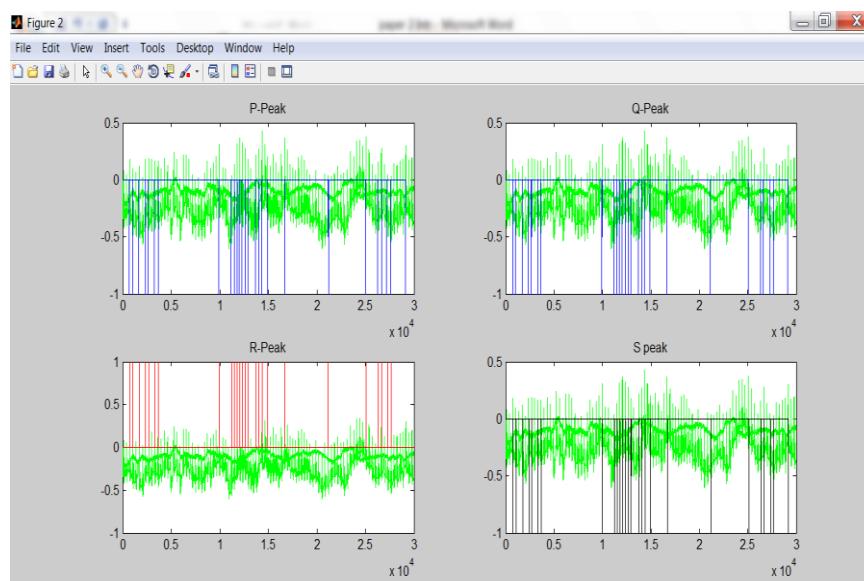
Figure 4: Feature Extraction results of PQRS waveform peak detection

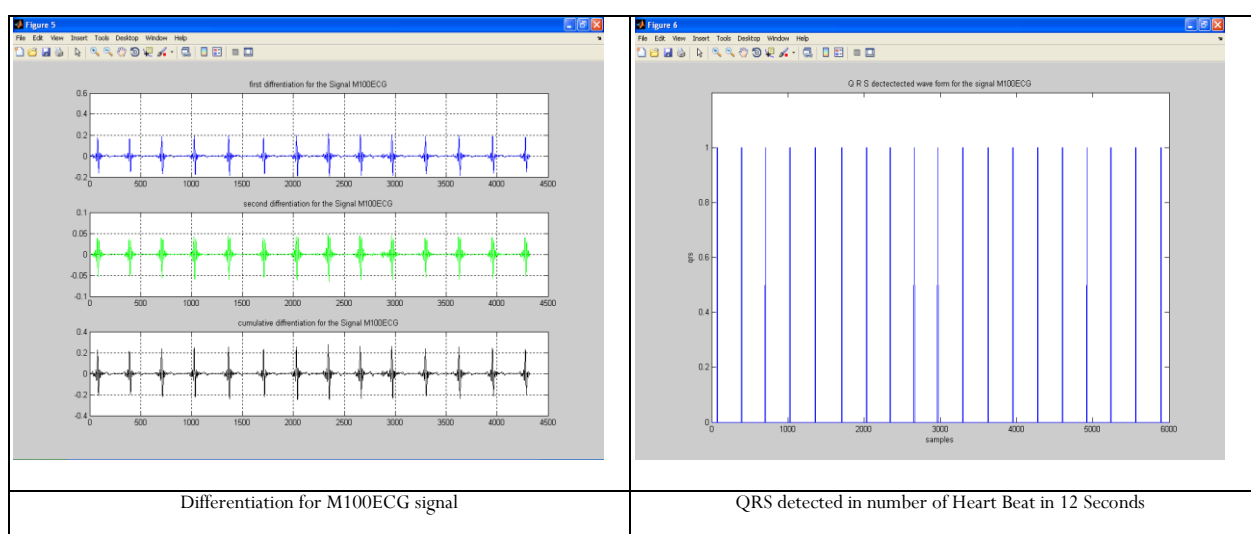| Differentiation for M100ECG signal | QRS detected in number of Heart Beat in 12 Seconds |

Figure 5. QRS Complex

Figure 3 illustrates the preprocessing output, figure 3(a) shows the original image and 3(b) illustrates the noise reduction using median filter. The original image is collected from the MIT-BIH arrhythmia database and each signal is digitized at 360 samples per second per

channel with 11-bit resolution over a 10 mV range. The preprocessing output of median filter is given as the input to the feature extraction. Feature extraction is used to detect the PQRS peak waveform and it is shown in figure 4. Peak detection is done by maximum and minimum value of the signal. Figure 5 illustrates the QRS complex and their heart beats. Table 1 provides the results such as overall performance metrics for ELM, SVM and BPN. Classification results are shown and given by a confusion matrix, through this confusion matrix accuracy, specificity, sensitivity and Positive Predictivity is produced. Totally 48 ECG signals are picked from the MIT BIH arrhythmia database, in this 50 signals are used for training and 50 signals are used testing. In testing 25 and 20 signals are divided as normal and abnormal. From the confusion matrix of TP, TN, FP and FN, accuracy is calculated and given the output of 97% for ELM, 89% for SVM and 84% for BPN it is given in table 1 and also provides the comparative result of individual class sensitivities, Positive Predictivity, and specificity of various classifiers such as ELM, SVM and BPN.

V CONCLUSION

The proposed paper analysis ECG signals using Extreme learning machine. Classified the ECG signal as normal and abnormal classes and it is collected from the MIT-BIH arrhythmia database, 48 records are collected from this database, and split 50 for training and 50 for testing. PQRS features and QRS complex are extracted in this paper. This extraction is useful for the classification of normal and abnormal beats. The method achieves are shown in experimental results that ELM gives the best result that is 97% accuracy, a sensitivity of 100%, specificity of 96% and a positive predictively of 95%. Different kinds of noise and artifacts contained in the ECG signals of the database are reduced using median filter.

REFERENCES

[1] N. Maglaveras, T. Stamkapoulos, K. Diamantaras, C. Pappas, M. Strintzis, ECG pattern recognition and classification using non-linear transformations and neural networks: a review, Int. J. Med. Inform. 52 (1998) 191–208.

[2] P. de Chazal, B.G. Celler, R.B. Rei, Using wavelet coefficients for the classification of the electrocardiogram, in: Proceedings of the 22nd Annual EMBS International Conference, Chicago IL, July 23–28, 2000.

[3] N. Srinivasan, D.F. Ge, S.M. Krishnan, Autoregressive modeling and classification of cardiac arrhythmias, in: Proceedings of the Second Joint Conference Houston, TX, USA, October 23–26, 2002.

[4] Hafizah Hussain, Lai Len Fatt, Efficient ECG signal classification using sparsely connected radial basis function neural network, in: Proceeding of the 6th WSEAS International Conference on Circuits, Systems, Electronics, Control and Signal Processing, December 2007, pp. 412–416.

[5] Marcel0 R. Risk, Jamil F. Sobh, J. Philip Saul, Beat Detection and Classification of ECG using self organizing maps, in: Proceedings –19th International Conference – IEEEIEMBS, Chicago, IL. USA, October 30–November 2, 1997.

[6] Yuksel. Ozbay, Rahime. Ceylan, Bekir. Karlik, Integration of type-2 fuzzy clustering and wavelet transform in a neural network based ECG classifier, Expert Syst. Appl. 38 (2011) 1004–1010.

[7] S.C. Saxena, V. Kumar, S.T. Hamde, Feature extraction from ECG signals using wavelet transforms for disease diagnostics, Int. J. Syst. Sci. 33 (13) (2002) 1073–1085.

[8] S.Y. Foo, G. Stuart, B. Harvey, A. Meyer-Baese, Neural network-based EKG pattern recognition, Eng. Appl. Artif. Intell. 15 (2002) 253–260.

[9] N. Maglaveras, T. Stamkopoulos, K. Diamantaras, C. Pappas, M. Strintzis, ECG pattern recognition and classification using non-linear transformations and neural networks: A review, Int. J. Med. Inform. 52 (1998) 191–208.

[10] K. Sternickel, Automatic pattern recognition in ECG time series, Comput. Methods Programs Biomed. 68 (2002) 109–115.

[11] I. Guler, E.D. Ubeyli, ECG beat classifier designed by combined neural network model, Pattern Recogn. 38 (2) (2005) 199–208

[12] Elif Derya Ubeyli, "Combining recurrent neural networks with eigenvector methods for classification of ECG beats", Digital Signal Processing 19 (2009) 320–329.

[13] Mehmet Korürek and Berat Dogan, "ECG beat classification using particle swarm optimization and radial basis function neural network", Expert Systems with Applications 37 (2010) 7563–7569.

[14] A. De Gaetano, S. Panunzi, F. Rinaldi, A. Risi, M. Sciandrone, "A patient adaptable ECG beat classifier based on neural networks", Applied Mathematics and Computation 213 (2009) 243–249.

[15] Samjin Choi, Mourad Adnane, Gi-Ja Lee, Hoyoung Jang, Zhongwei Jiang, Hun-Kuk Park, "Development of ECG beat segmentation method by combining lowpass filter and irregular R–R interval checkup strategy", Expert Systems with Applications 37 (2010) 5208–5218.

[16] Xiaoran Ning and Ivan W. Selesnihack, "ECG Enhancement and QRS Detection Based on Sparse Derivatives", Biomedical Signal Processing and Control 8 (2013) 713– 723.

[17] Suranai Poungponsria and Xiao-Hua Yu, "An adaptive filtering approach for electrocardiogram (ECG) signal noise reduction using neural networks", Neurocomputing 117(2013)206–213.

[18] Yun-Chi Yeh, Che Wun Chiou and Hong-Jhih Lin, "Analyzing ECG for cardiac arrhythmia using cluster analysis", Expert Systems with Applications 39 (2012) 1000–1010.

[19] Indu Saini, Dilbag Singh, Arun Khosla, "QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases", Journal of Advanced Research (2013) 4, 331–344.

[20]   Deboleena Sadhukhan and Madhuchhanda Mitra, "R-peak detection algorithm for ECG using double difference and RR interval processing", Procedia Technology 4 (2012) 873 − 877.

[21]   Hari Mohan Rai, Anurag Trivedi and Shailja Shukla, "ECG signal processing for abnormalities detection using multi-resolution wavelet transform and Artificial Neural Network classifier", Measurement 46 (2013) 3238−3246.

[22]   R. Mark, G. Moody, MIT−BIH arrhythmia database directory <http:// ecg.mit.edu/dbinfo.html.>

[23]   Stallin Subbiah, Rajkumar Patro and K. Rajendran, "Reduction of Noises In ECG Signal by Various Filters", International Journal of Engineering Research & Technology, Vol. 3 - Issue 1, 2014.

[24]   Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery, 2, 955−971.