IMAGE & SIGNAL PROCESSING

CrossMark

ECG Signal Classification Using Various Machine Learning Techniques

S Celin[1] (iD) · K. Vasanth[2]

Abstract
Electrocardiogram (ECG) signal is a process that records the heart rate by using electrodes and detects small electrical changes for each heat rate. It is used to investigate some types of abnormal heart function including arrhythmias and conduction disturbance. In this paper the proposed method is used to classify the ECG signal by using classification technique. First the Input signal is preprocessed by using filtering method such as low pass, high pass and butter worth filter to remove the high frequency noise. Butter worth filter is to remove the excess noise in the signal. After preprocessing peak points are detected by using peak detection algorithm and extract the features for the signal are extracted using statistical parameters. Finally, extracted features are classified by using SVM, Adaboost, ANN and Naïve Bayes classifier to classify the ECG signal database into normal or abnormal ECG signal. Experimental result shows that the accuracy of the SVM, Adaboost, ANN and Naïve Bayes classifier is 87.5%, 93%, 94 and 99.7%. Compared to other classifier naïve bayes classifier accuracy is high.

Keywords ECG signal · Butter worth filter · SVM · Adaboost · ANN · Naïve bayes

Introduction

Electrocardiogram (ECG) is a periodic signal which reflects the activity of the heart. From ECG a lot information is obtained for normal and pathological physiology of heart. The ECG signal is non-stationary in nature and very difficult to analyze. Clinical observation takes long time and the signal is non-stationary. So, computer-based technique is used in ECG analysis. The principle of ECG signal is electrical activity. This is transmitted through the body and picked up on the skin. Finally, the ECG machine records the activity using electrodes and display graphically. The ECG signal consists of repetitive complex waveform with a frequency value of 1 Hz. Each cardiac cycle consists of three waves such as P wave QRS wave and T wave. These waves are produced due to the function of atria and ventricular parts of the heart. Cardiovascular disease is one of the major causes of death in the western world where more than 16 million die annually [1].

An Electro cardio graphic signal is recorded on a long timescale for the purpose of identifying intermittently occurring disturbances in the heart rhythm. To reduce the cardiovascular diseases the life style changes such as reduced cholesterol intake and regular exercise are mandatory. Early detection is a crucial step in the cardiovascular disease. To measure the electrical signal from different parts of heart two types of electrodes are used. They are limb electrodes and chest electrodes. The P wave is produced by atrial depolarization and the duration is less than 120 ms and this can be considered as low frequency wave. ORS complex wave has duration of 70-110 ms and T wave is produced by the repolarization of ventricles having duration greater than 300 ms [2]. Feed forward multilayer neural network with error back propagation used to classify the ECG signal as normal or abnormal signal and two data set are tested for different NN structures.

Neural network [3–7] is to classify the ECG signal and it consists of two-dimensional array to implement the digital hardware [8]. Fuzzy logic neural network [9–12] is to classify and detect the ECG signal. The amplitude and intervals of the

✉ S Celin
  celin10sam@gmail.com

  K. Vasanth
  vasanthecek@gmail.com

[1] Satyabama Institute of Science and Technology, Chennai, India

[2] Vidya Jyothi Institute of Technology, Hyderabad, Telangana 500075, India
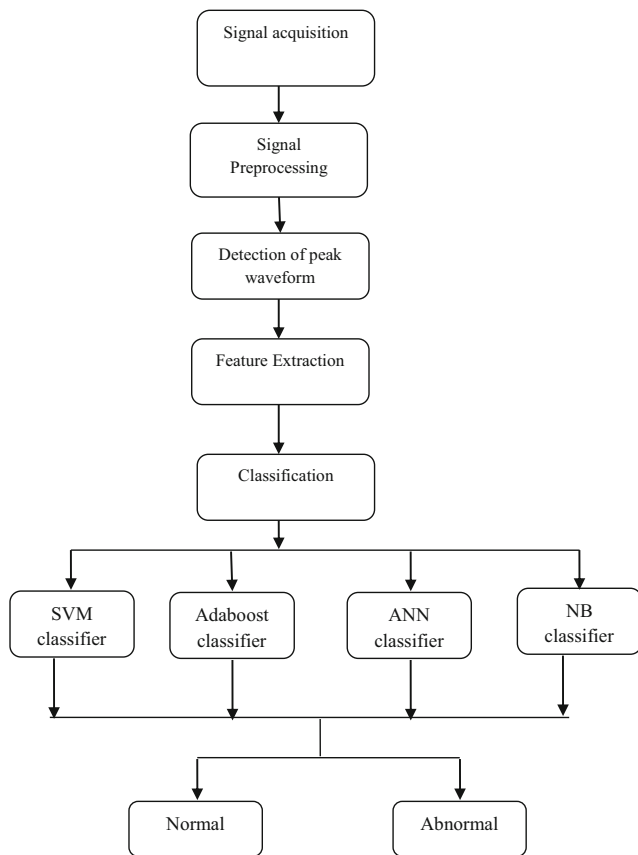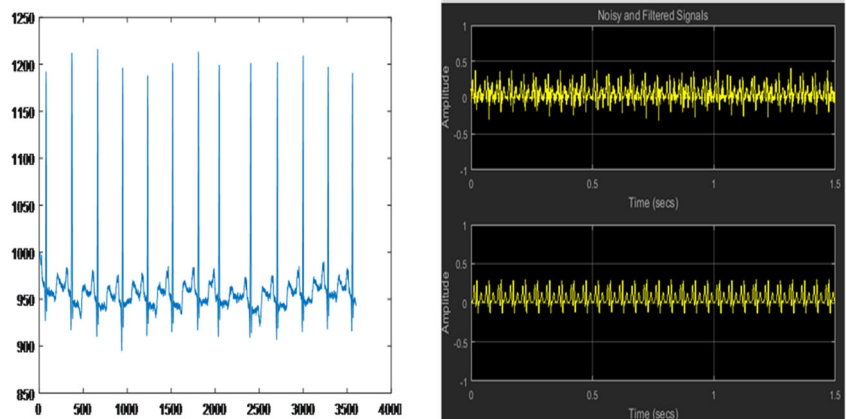
Fig. 1 Block diagram of proposed system

Table 1 Symbols

Symbol	Notation
ρ	Blood density
r	Inner radius of vessel
E	Elastic modules
h	Vessel wall thickness

cardiac waves are obtained. The amplitude and duration of the wave is more. To remove the unwanted noise in the ECG signal

median filter [13–18] is used. For extracting the features, classification preprocessing is the important process. For the analysis of low frequency components median filter play an important role. Time and frequency features based random forest [19–21] is to classify ECG signal. Different classification techniques [22–25] are used to classify the ECG signal.

The contribution of the paper is as follows:

- Remove high frequency noise by low, high pass filter and butter worth filter.
- Detect the R peak by peak detection algorithm.
- Classify the ECG signal by using SVM, ANN, Adaboost and Naïve Bayes classifier.

The main aim of the proposed method is to detect the R peak from QRS waveform and classify the ECG signal as normal or abnormal signal. To avoid over fitting problem in the signal cross validation method is used in the classifier. Section "Proposed methodology" describes in detail about feature extraction and classification of the proposed method. Section "Result and discussions" discusses about experimental results of the proposed method followed by conclusion in section "Conclusion".

Proposed methodology

The aim of the proposed method is to classify the normal and abnormal cell in the ECG signal. Figure 1 shows the block diagram of the proposed method. It mainly consists of three stages such as preprocessing, peak detection and classification. Preprocessing is done by using low, high and butter worth filter to remove the high frequency noise in the signal. Butter worth filter is used to remove excess noise in the signal. After preprocessing detect the QRS [26] peak for Heart Rate Variability (HRV) signal using peak detection algorithm. Statistical features are extracted from ECG signal and finally the signal is classified using

Fig. 2 Original ECG, noise signal for normal signal

Fig. 3 Raw ECG data plotting, filtered ECG signal for normal signal
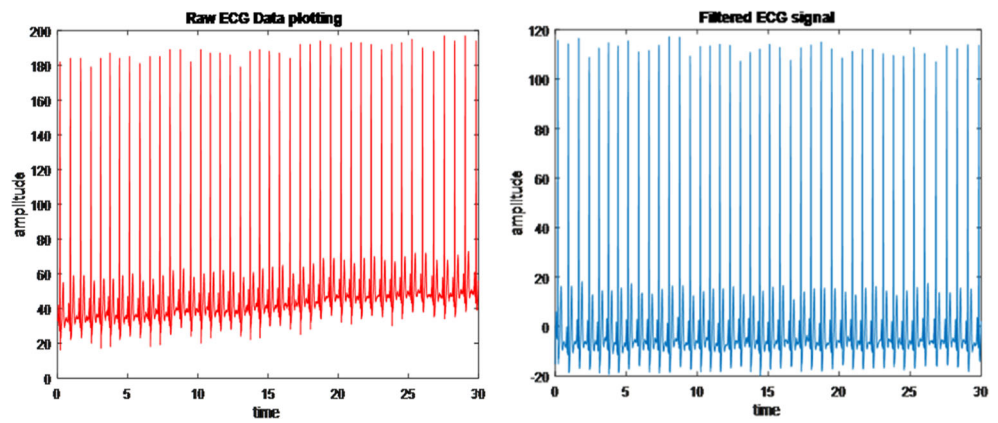
classification algorithm such as Feed Forward Neural Network (FFNN), Naïve Bayes and Cascade algorithm to detect if the signal is normal or abnormal. Cascade algorithm is the combination of Support Vector Machine (SVM) and Adaboost algorithm.

Signal acquisition

The database for ECG signal is taken from MIT-BIH arrhythmia database. The recording is digitized by 360 samples per second per channel with a resolution of 10mv range. Other database is ECG –ID database and it also used to classify the ECG signal.

Preprocessing

Preprocessing is one of the important tasks in signal and image processing. To eliminate the high frequency noise in the ECG signal filtering techniques such as low pass filter, high pass filter and butter worth filter are used [27]. Filtering techniques avoids the overlap between ECG and muscle noise. Low pass and high pass filter removes the

high frequency noise. After removal of noise in the ECG signal high frequency noise sometimes appears [28–30]. So, to remove the excess noises in the signal apply Butter worth filter with a sampling frequency of 1000 Hz is used and cut off frequency of 265khz. Hence it is a simple type of filter and in pass band it shows flat response. At zero herz it has 2 N-1 derivative of the magnitude response for order N. The reliability of the proposed work can be incorporated with the protection schemes [27, 31–33].

Peak detection

Detect the QRS complex of ECG signal using peak detection algorithm's complex wave is important for HRV analysis of ECG signal. In QRS wave detection algorithm R wave has the highest or lowest value in the QRS complex.

Feature extraction

Different signal features are extracted using statistical parameters. Based on the peak detection features are extracted in order to classify the ECG signal. Features such as mean,

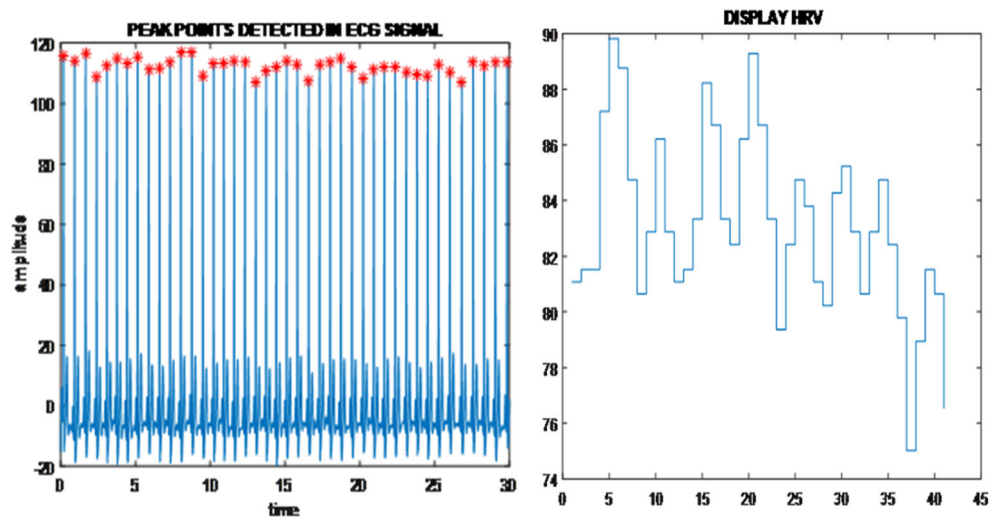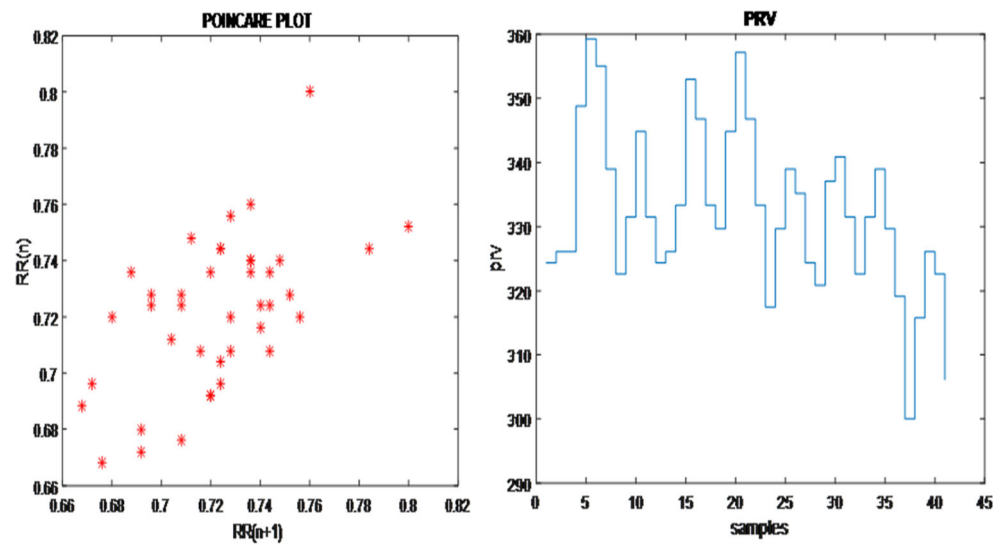**Fig. 4** Peak points detected, HRV display for normal signal

Fig. 5 Point care plot and PRV for the normal signal

standard deviation, root mean square, pulse transit time, pulse rate variability are extracted.

Pulse transit time (PTT) - It is the time interval for arterial pulse pressure wave travel from aortic valve to peripheral site. Pulse transit time is calculated by following equation. Table 1 gives the symbol for the parameters.

$$PWV = \frac{L}{PTT} = \sqrt{\frac{E.h}{2r\rho}} \qquad (1)$$

Where PWV is the pulse wave velocity vessel length, ρ is the blood density, r is the inner radius of vessel, h is the

vessel wall thickness and E is the elastic modulus.
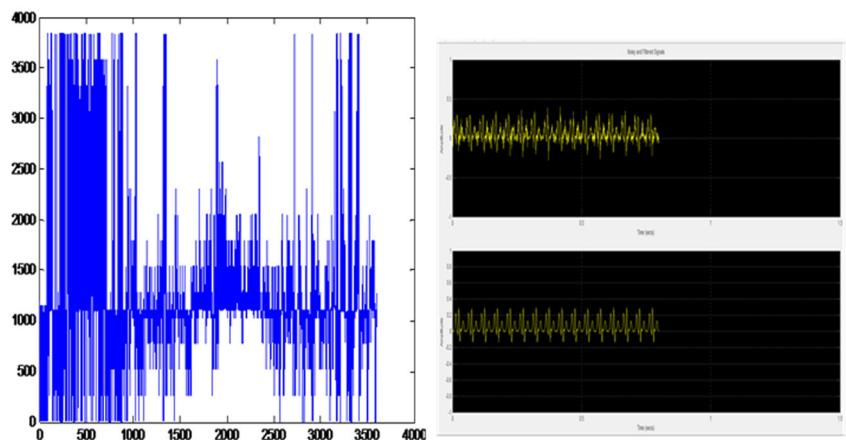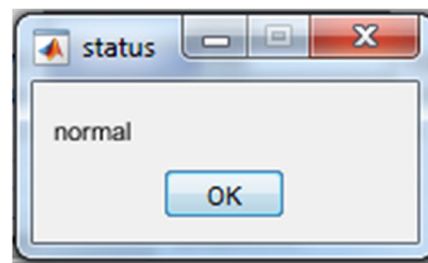
Pulse rate variability (PRV) – For pulse rate variability the fiducial point for each pulse in PPG is the medium amplitude point η_{M_i}.
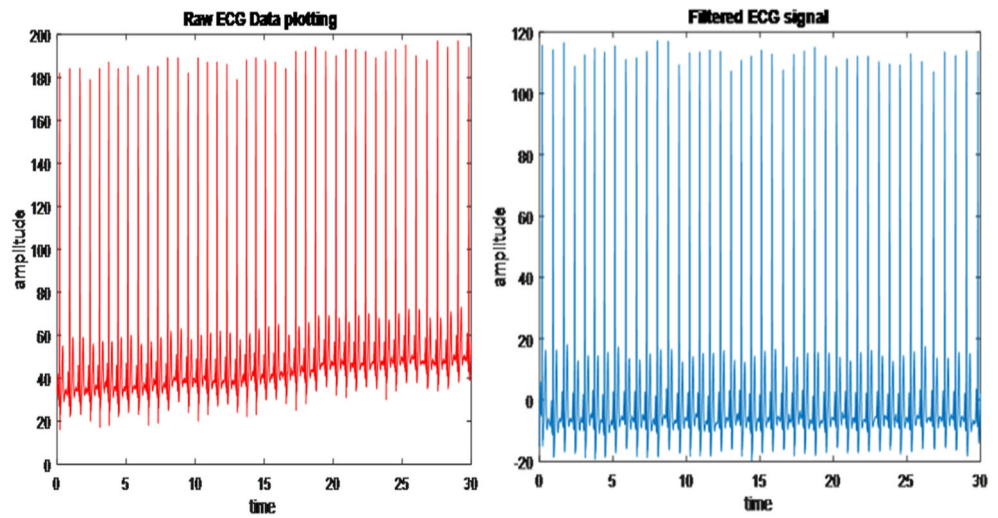
$$THB = N \qquad (2)$$

$$MRR = \bar{I} = \frac{1}{N-1}\sum_{n-2}^{N} I(n) \qquad (3)$$

$$SDNN = \sqrt{\frac{1}{N-1}\sum_{n=2}^{N}\left(I(n)-\bar{I}\right)^2} \qquad (4)$$

Fig. 6 Original ECG, noise signal for abnormal signal

Fig. 7 Raw ECG data plotting, filtered ECG signal for abnormal signal

$$RMSSD = \sqrt{\frac{1}{N-2}\sum_{n=3}^{N}(I(n)-I(n-1))^2} \qquad (5)$$

THB   Total heart beats
MRR   Mean of RR intervals
SDNN   Standard deviation of normal to normal RR intervals
RMSSD Root mean square of successive NN intervals

Classification

Support vector machine

Support Vector Machine (SVM) (16) is a supervised learning method used for classification and regression prediction tool to maximize the accuracy. It uses hypothesis space of a linear function and it can be used for different applications such as face analysis, handwriting analysis, mainly for pattern classification and regression. It achieves high generalization by maximizing the margin and supports an efficient learning of nonlinear functions is the main properties of SVM. The initial form of SVM classifier is binary classifier, output is either negative or positive. Two class SVM is used to train the data and classify the test signals between normal and abnormal ECG signals. Different combination of features is combined and the output is given to the Adaboost classifier.

Adaboost classifier

Adaboost is the short form of Adaptive Boosting algorithm and is a machine learning meta algorithm to improve the performance. The aim of adaboost classifier is to reduce the error rate of the weak classifier by introducing a new

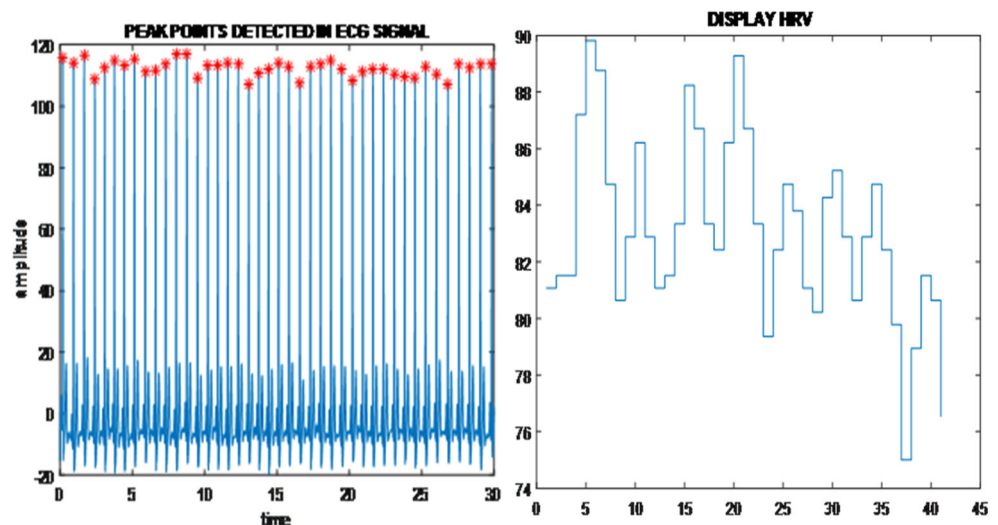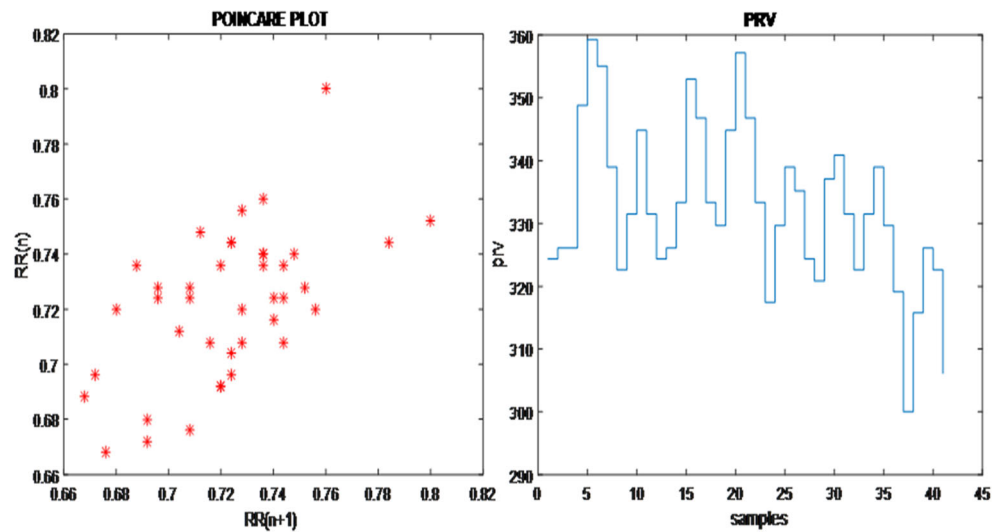**Fig. 8** Peak points detected, HRV signal for abnormal signal

Fig. 9 Point care plot and PRV for the abnormal signal

classifier. Here the output of SVM classifier is combined with adaboost classifier represents the final output of the adaboost classifier. Adaboost is sensitive to noise data and the adaboost algorithm is shown below.

$$H(x) = \sin\left(\sum_{t=1}^{T} \alpha_i h_t(x)\right) \tag{6}$$

Where $h_t(x)$ is the weak classifier, α_i weight is chosen such that error rate is miminum.

Artificial neural network

Artificial neural network (ANN) is a biologically inspired network useful in different areas such as classification, pattern recognition etc. Based on the features of input pattern, the decision making of ANN is holistic and is suitable for the classification of biomedical data. ANN is done by using back propagation algorithm and its aim is to reduce the error to minimum. The weights are randomly assigned and for each epoch the weights are updated, the error between desired output and actual output is minimum. The activation function used is sigmoid function. ANN classifier classifies the ECG database into normal or abnormal signal. Sixteen signal dataset is used for classification and ANN classifier plots five various charactertics. They are train state, error histogram, performance, confusion and receiver operating charactertics.

Naive bayes classifier

Naive Bayes (NB) classifier is the form of data analysis that extracts the models describing important data classes. The naïve bayes classifier has various applications such as fraud detection, performance, target marketing and medical

analysis. Naive Bayes classifieris used to classify the ECG signal as normal or abnormal implemented in rapid miner tool and it has exhibit high accuracy, applied to large database. In this method naive bayes classifier is to classify the ECG signal into normal or abnormal signal. It is the statistical classifier that predicts the class membership and has high accuracy compared to another classifier. Assume, attribute value on a given class is independent of the values of the other attributes and this is called class conditional independence; simplify the computation and this state is called naïve.
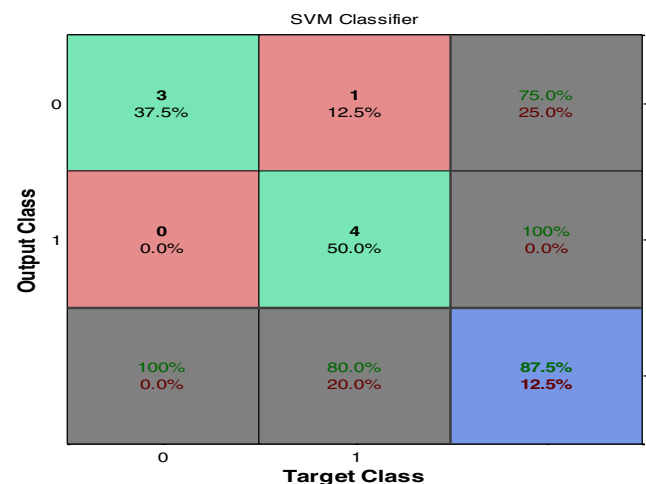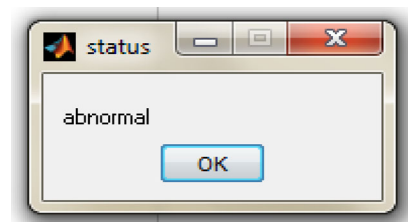

Fig. 10 Confusion matrix of SVM classifier
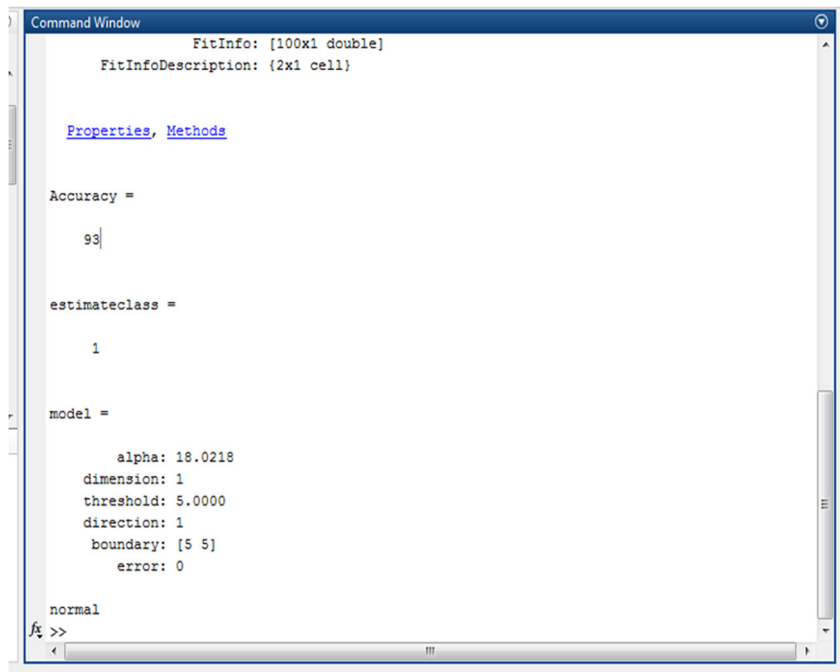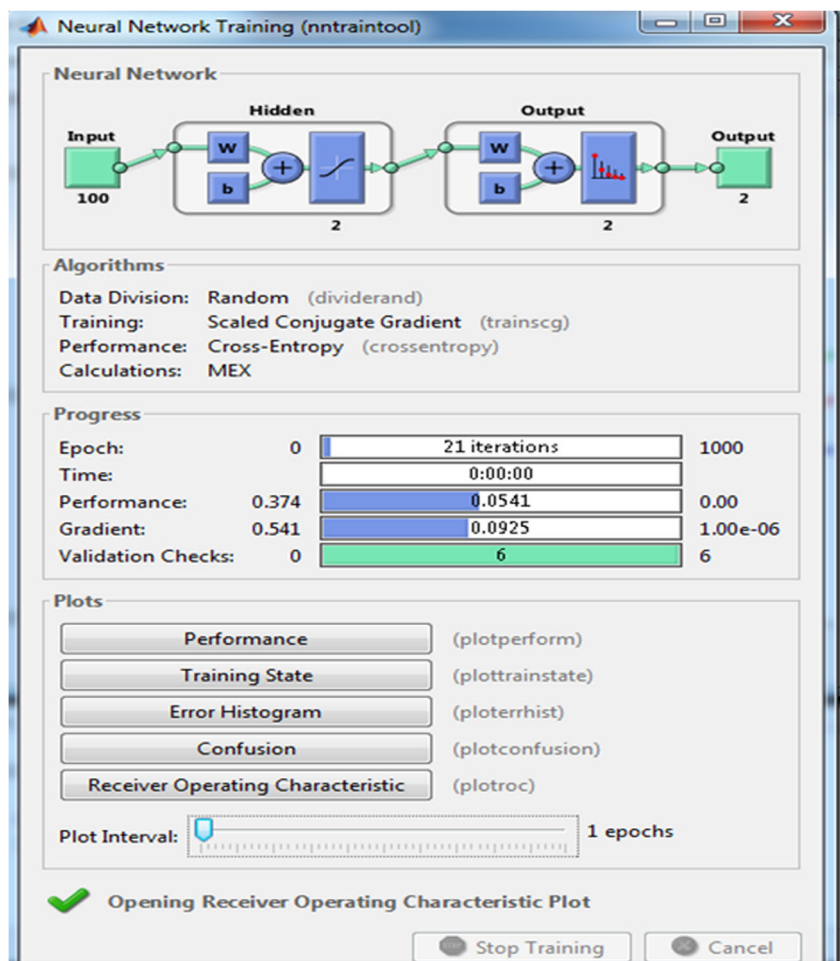
Fig. 11 Adaboost classifier

```
Command Window
                    FitInfo: [100x1 double]
           FitInfoDescription: {2x1 cell}


    Properties, Methods


Accuracy =

    93


estimateclass =

    1


model =

        alpha: 18.0218
    dimension: 1
    threshold: 5.0000
    direction: 1
     boundary: [5 5]
        error: 0

normal
fx >>
```

Fig. 12 Neural network classifier
tool box

```
Neural Network Training (nntraintool)

Neural Network
              Hidden              Output
    Input                                        Output
                 W                  W
                 b                  b
     100                                            2
                 2                  2

Algorithms
Data Division:  Random    (dividerand)
Training:       Scaled Conjugate Gradient  (trainscg)
Performance:    Cross-Entropy   (crossentropy)
Calculations:   MEX

Progress
Epoch:              0          21 iterations          1000
Time:                         0:00:00
Performance:    0.374          0.0541                 0.00
Gradient:       0.541          0.0925                 1.00e-06
Validation Checks:  0             6                   6

Plots
        Performance               (plotperform)
        Training State            (plottrainstate)
        Error Histogram           (ploterrhist)
        Confusion                 (plotconfusion)
        Receiver Operating Characteristic  (plotroc)

Plot Interval:                              1 epochs

    Opening Receiver Operating Characteristic Plot

                            Stop Training      Cancel
```
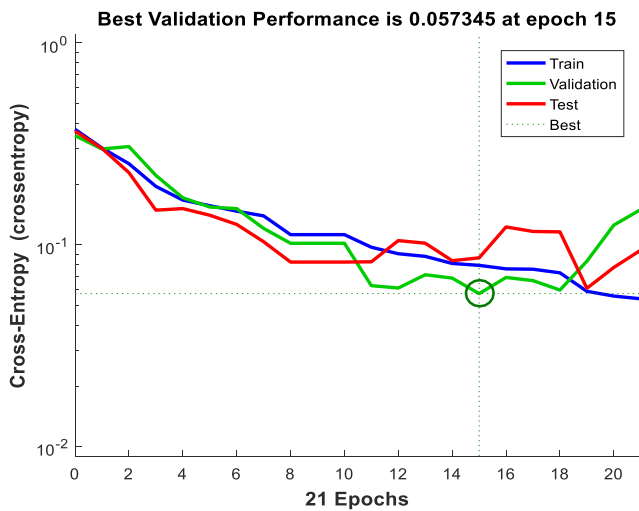
Fig. 13 performance plot

Error measures the discrepancy between the two borders and it is calculated by

$$\text{Border error} = \frac{FP + FN}{TP + FN} \times 100\% \qquad (7)$$

Result and discussions

The proposed method uses ECG signal and it is taken from MIT-BIH database and classify the output as normal or abnormal signal. The input signal is preprocessed by using filter methods such as low pass, high pass and butter worth filter. By using low pass and high pass filter low frequency noise is removed and in order to remove the excess noise butter worth filter is used with a sampling frequency of 1000 Hz. Detect the peak for the QRS waveform and extract the features for the HRV signal. Classify the signal
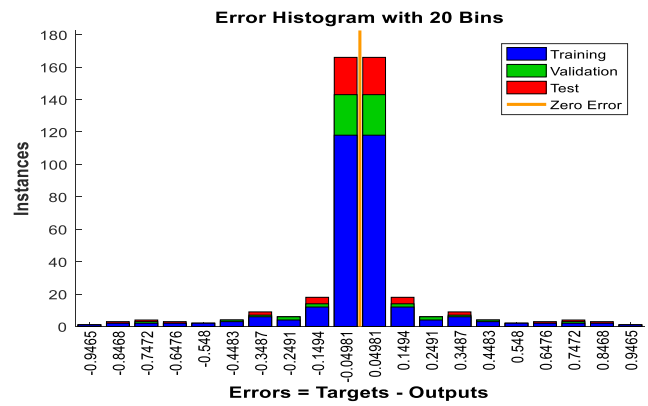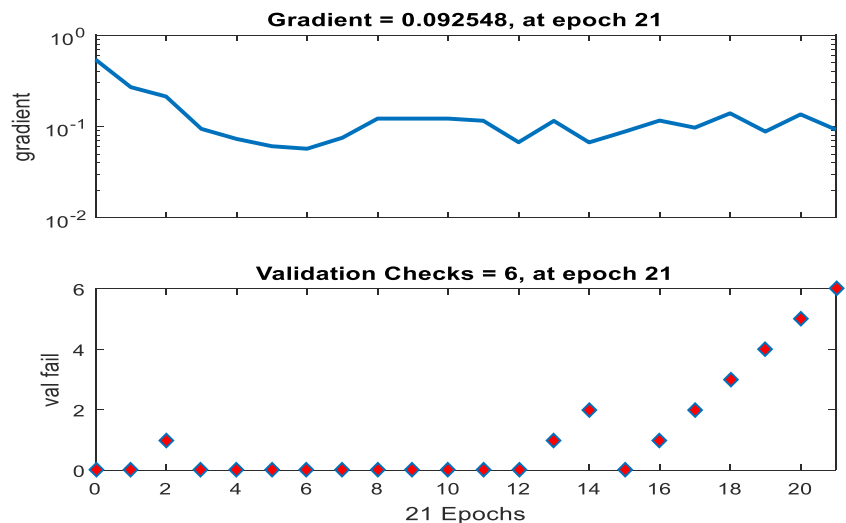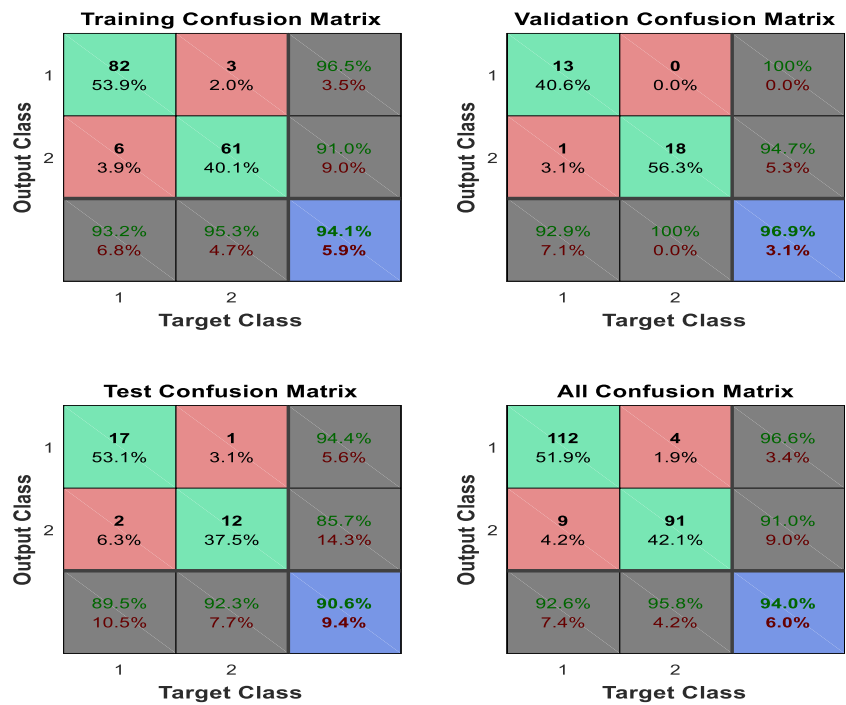

Fig. 15 Error histogram plot

using four techniques such as ANN, SVM, Naïve Bayes, Adaboost classifier. The experimental result is simulated by using Matlab 2016a.

Figure 2 to 5 shows the ECG signal preprocessing and peak detection of the normal signal. The ECG signal is filtered by using low, high, and butter worth filter to remove the unwanted noise in the signal. Raw ECG is plotted for the signal and peak points are detected for the QRS waveform using peak detection algorithm. In this the peaks are detected for the signal and HRV (Heart Rate Variable) signal is displayed. After HRV signal is displayed, point care is detected for the signal and finally PRV (Pulse Rate Variable) is detected for the ECG signal. The same process is repeated for abnormal signal as shown in Figs. 6, 7, 8 and 9.

Figure 10 shows the confusion matrix output of SVM classifier and here the accuracy of the classifier is 87.5%, sensitivity is 75%,specificity is 100%.Figure 11 shows the Adaboost classifier output. To classify the ECG signal into normal or abnormal signal totally 16 ECG signal are taken from the database. The accuracy of the SVM classifier is 87.5%and the accuracy of the Adaboost

Fig. 14 Training state plot

Fig. 16 Confusion matrix for ANN

Training Confusion Matrix

	1	2	
1	82 / 53.9%	3 / 2.0%	96.5% / 3.5%
2	6 / 3.9%	61 / 40.1%	91.0% / 9.0%
	93.2% / 6.8%	95.3% / 4.7%	94.1% / 5.9%

Validation Confusion Matrix

	1	2	
1	13 / 40.6%	0 / 0.0%	100% / 0.0%
2	1 / 3.1%	18 / 56.3%	94.7% / 5.3%
	92.9% / 7.1%	100% / 0.0%	96.9% / 3.1%

Test Confusion Matrix

	1	2	
1	17 / 53.1%	1 / 3.1%	94.4% / 5.6%
2	2 / 6.3%	12 / 37.5%	85.7% / 14.3%
	89.5% / 10.5%	92.3% / 7.7%	90.6% / 9.4%

All Confusion Matrix

	1	2	
1	112 / 51.9%	4 / 1.9%	96.6% / 3.4%
2	9 / 4.2%	91 / 42.1%	91.0% / 9.0%
	92.6% / 7.4%	95.8% / 4.2%	94.0% / 6.0%

classifier is 93%. An SVM classifier is added with adaboost classifier in the final output of the adaboost classifier. Figure 12 shows the neural network classifier tool box and Figures 13 to 17 show the various characteristics of the ANN classifier. The overall accuracy of the ANN classifier is 94%. The total number of iteration used for the process is 21. Figure 18 shows the naïve bayes classifier output and the accuracy of the classifier is 99.7%.Naïve bayes classifier shows the better accuracy compared to another classifier.
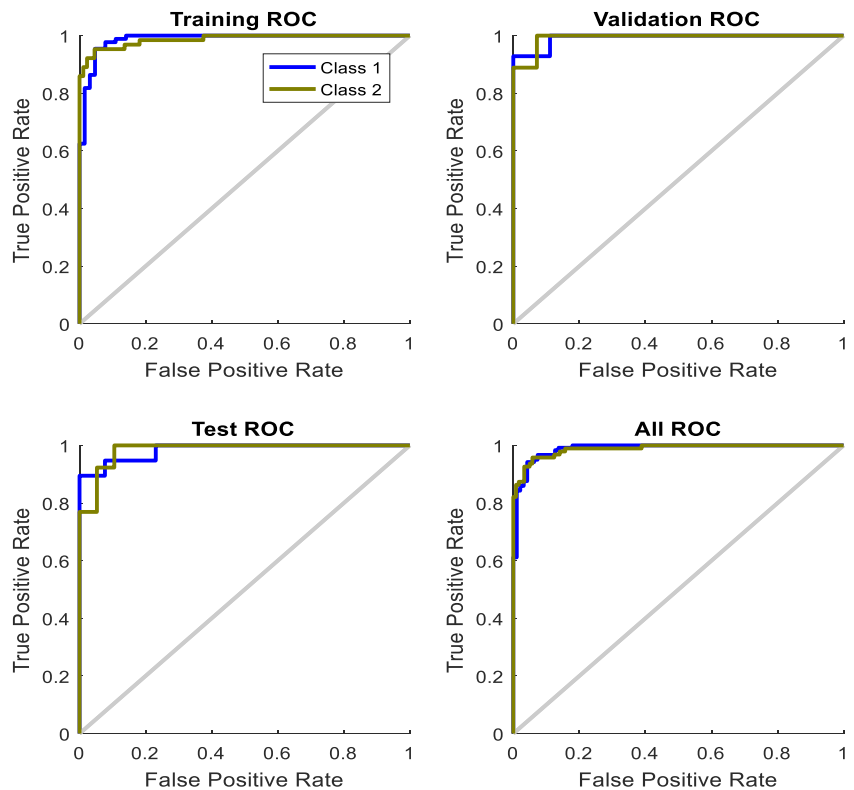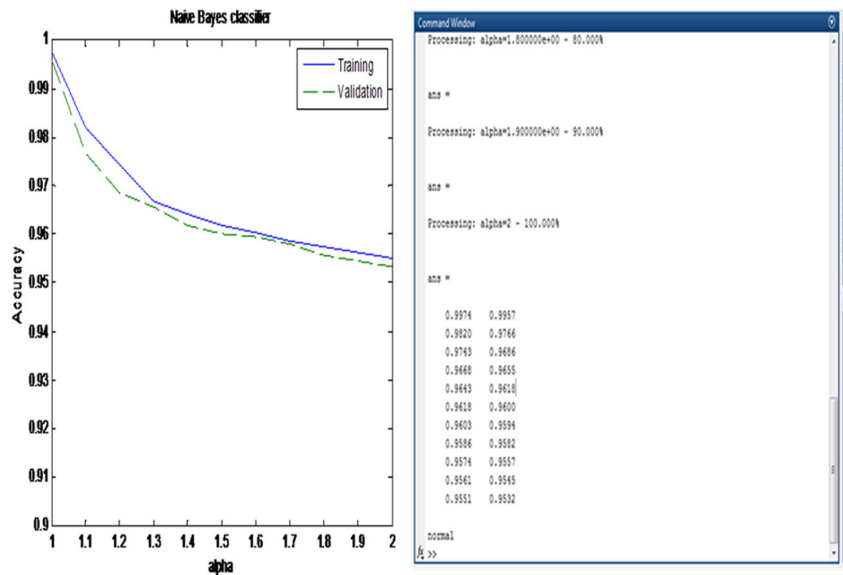
Fig. 17 Receiver operating charactertics plot

Fig. 18 Naïve bayes classifier

Totally four classifiers are used to classify the ECG signal database such as ANN, SVM, Adaboost and Naïve bayes classifier. These classifiers classify the ECG signal into normal or abnormal ECG signal. The accuracy of SVM classifier is 87.5%, Adaboost is 93%, ANN classifier is 94% and for Naïve Bayes 99.7%. The accuracy of Naïve Bayes classifier is high compared to SVM, ANN and Adaboost classifier. The classifier accuracy and error values of all the classifier are compared. Table 2 shows the comparison table for classifier accuracy and error values.

Figure 19 shows the classifier accuracy and error comparison, the accuracy of the Naïve bayes classifier is high compared to other classifier. If the error value is low the accuracy is high.

Conclusion

In this paper the various types of classifiers are used to classify the ECG signal into normal or abnormal ECG signal. The input image is preprocessed by various filtering technique to remove high frequency noise. Peak value is detected by peak detection algorithm and HRV signal is detected. Features are extracted for signal and the signals are classified by ANN, SVM, Adaboost and Naive bayes classifier. From experimental results the accuracy of the SVM, Adaboost, ANN and Naïve Bayes classifier is 87.5%,93%,94 and 99.7%.Naive Bayes classifier has the highest accuracy compared to other classifier.

Future scope

In future detect the peak for other waveform such as P and T waveform and classify the signal.

Table 2   Comparison of classifier accuracy and error values

Classifier	SVM	Adaboost	ANN	Naïve bayes
Accuracy (%)	87.5	93	94	99.7
Error (%)	12.5	7	6	0.3

Compliance with ethical standards

Fig. 19 Comparison of classifier accuracy and error

References

1. Zhao, Q., and Zhan, L., ECG Feature Extraction and Classification Using Wavelet Transform and Support Vector Machines. International Conference on Neural Networks and Brain, ICNN&B 2:1089–1092, 2005.

2. El-Khafif, S. H., and El-Brawany, M. A., Artificial Neural Network-Based Automated ECG Signal Classifier. ISRN Biomedical Engineering 2013:1–6, 2013.

3. Jiang, W., Kong, S. G., and Peterson, G. D., ECG Signal Classification using Block-based Neural Networks. Proceedings of International Joint Conference on Neural Networks, pp. 326–331, 2005.

4. Jiang, W., Member, S., and Kong, S. G., Block-Based Neural Networks for Personalized ECG Signal Classification. IEEE Trans. Neural Netw. 18(6):1750–1761, 2007.

5. Subbiah, S., Patro, R., and Subbuthai, Feature extraction and classification for ECG signal processing based on artificial neural network and machine learning approach. International Conference on Inter Disciplinary Research in Engineering and Technology, pp. 50–57, 2015.

6. Djahida, M., Talha, K., and Slimane, S.., Neural networks and SVM for heartbeat classification. 11th International Conference on Information Sciences, Signal Processing and their Applications, pp. 830–835, 2012.

7. Güler, I., and Übeyli, E. D., ECG beat classifier designed by combined neural network model. Pattern Recogn. 38(2):199–208, 2005.

8. Tayel, M. B., Eltrass, A. S., and Ammar, A. I., A new multi-stage combined kernel filtering approach for ECG noise removal. J. Electrocardiol. 51:265–275, 2018.

9. Mahapatraa, S., Mohantab, D., Mohantyc, P., Nayakd, S., and Beharie, P., A Neuro-fuzzy based model for analysis of an ECG signal using Wavelet Packet Tree. International Conference on Intelligent Computing, Communication & Convergence 92:175–180, 2016.

10. Mahapatra, S., Nayak, S. K., and Sabat, S. L., Neuro -Fuzzy model for adaptive filtering of oscillatory Signals. Elsevier Science, Measurement 30:231–239, 2001.

11. Chen, Y. C., and Teng, C. C., Fuzzy neural network in model reference control systems. Fuzzy Logic and Expert Systems Applications, San Diego-Toronto: Academic Press, pp. 285–313, 1998.

12. Engin, M., ECG beat classification using neuro-fuzzy network. Elsevier pattern Recognisation Letter 25:1715–1722, 2004.

13. Jaykumar S. Karnewar, Milind V. Sarode," The Combined Effect of Median and FIR Filter in Pre-processing of ECG Signal using Matlab", International Journal of Computer Applications,pp.30-33March 2013.

14. Upganlawar, I. V., and Chowhan, H., Pre-processing of ECG Signals Using Filters. International Journal of Computer Trends and Technology (IJCTT) 11(4):166–168, 2014.

15. Lyons, R., Interpolated narrowband lowpass FIR filters. IEEE Signal Process. Mag. 20(1):50–57, 2003.

16. Saramaki, T., Neuvo, Y., and Mitra, S. K., Design of computationally efficient interpolated FIR filters. IEEE Transactions on Circuits and Systems 5(1):70–88, 1988.

17. Daskalov, I. K., and Christov, I. I., Electrocardiogram signal pre-processing for automatic detection of QRS boundaries. Journal of Elsevier Medical Engineering & Physics 21:37–44, 1999.

18. Arias-Castro, E., and Donoho, D. L., Does median filtering truly preserve edges better than linear filtering? Ann. Stat. 37(3):1172–1206, 2009.

19. Kropf, M., Hayn, D., and Schreier, G., ECG Classification Based on Time and Frequency Domain Features Using Random Forests. Comput. Cardiol. 44:1–4, 2017.

20. Li, T., and Zhou, M., ECG Classification Using Wavelet Packet Entropy and Random Forests. Entropy 18(8):1–16, 2016.

21. Emanet, N., ECG beat classification by using discrete wavelet transform and random forest algorithm. International Conference on Soft Computing, 2009.

22. Padmavathia, S., and Ramanujamb, E., Naïve Bayes Classifier for ECG abnormalities using Multivariate Maximal Time Series Motif. Procedia Computer Science 47:222–228, 2015.

23. Kandala, N. V. P. S., and Rajesh, R. D., Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier. Biomedical Signal Processing and Control 41:242–254, 2018.

24. Maciejewski, M., Dzida, G., ECG parameter extraction and classification in noisy signals, Signal Processing, pp. 243–248, 2017.

25. Kaistha, T., Mahajan, A., and Ahuja, K., A novel approach for Extraction and Classification of ECG signal using SVM, International Journal of Computer Technology and Application. 9(41):177–182, 2016.

26. Xiong, Z., Stiles, M. K., Zhao, J., Robust ECG signal classification for detection of atrial fibrillation using a novel neural network. Cardiology, 44, 2017.

27. Ahilan, A., & Deepa, P., Improving lifetime of memory devices using evolutionary computing based error correction coding. Computational Intelligence, Cyber Security and Computational Models. pp. 237–245, 2016.

28. Harikumar, R., Shivappriya, S. N., Analysis of QRS Detection Algorithm for Cardiac Abnormalities – A Review. International Journal of Soft Computing and Engineering (IJSCE). 1 (5), 2011.

29. Golrizkhatami, Z., Taheri, S., and Acan, A., Multi-scale features for heartbeat classification using directed acyclic graph CNN. An International Journal of Applied Artificial Intelligence, 2018.

30. Nayak, S., Soni, M. K., and Bansal, D., Filtering Techniques For ECG Signal Processing", International Journal of Research in Engineering & Applied Sciences 2(2), 2012.

31. Ahilan, A., and Deepa, P., Radiation Induced Multiple Bit Upset Mitigation and Correction in Memories using Cost Efficient CMC. Microelectronics, Electronic Components and Materials 46(4):257–266, 2016.

32. Ahilan, A., and Deepa, P., Design for built-in FPGA reliability via fine-grained 2-D error correction codes. Microelectron. Reliab. 55(9–10):2108–2112, 2015.

33. Ahilan, A., & Deepa, P., Design for Reliability: A Novel Counter Matrix Code for FPGA based Quality Applications. 6th Asia Symposium on Quality Electronic Design Electronic Design (ASQED 2015), Kuala Lumpur, Malaysia Vol. 6, pp. 56–61, 2015.

34. Gaikwad, K. M., Chavan, M. S., Removal of high frequency noise from ECG signal using digital IIR butterworth filter. IEEE Global Conference on Wireless Computing and Networking, pp. 121–124, 2014.

35. Sarma, P., Nirmala, S. R., Sarma, K. K., Classification of ECG using some Novel Features. IEEE International Conference on Emerging Trends and Applications in Computer Science, pp. 187–191, 2013.