




HRV Signal Feature Estimation and Classification for Healthcare System Based on Machine Learning

Ranjeet Kumar¹ , Agya Ram Verma², Manoj Kumar Panda²,
and Papendra Kumar²

¹ Madanapalle Institute of Technology and Science,
Madanapalle 517325, Andhra Pradesh, India
ranjeet281@gmail.com

² G.B. Pant Engineering College, Pauri Garhwal 246194, Uttarakhand, India
agyaram06ei03@gmail.com, pandgbpec@gmail.com,
papendral@yahoo.co.in

Abstract. A system for heartbeat classification is provided a solution against the mortality due to cardiac arrest. Here, an automated real-time system proposed to identify and classify the abnormalities in the electrocardiogram (ECG) signal or its variability. In healthcare services, multiple monitoring systems are available to monitor cardiac health conditions. Although, efficient monitoring and the alert system needed for prevention from any health loss or mortality. In this work, a public database is utilized named MIT-BIH Arrhythmia (MITA) ECG Database and Normal Sinus Rhythm Database (NSRDB). In the classification process features play a key role to identify the class of the signal that may indicate the particular health condition. Here, a statistical technique is used for the analysis of HRV data of ECG signal and decision tree for classification of parameters extracted from HRV signals. In this paper, a method developed for arrhythmia detection using time and time-frequency domain statistical features. Therefore, five statistical parameters of HRV signals were computed and considered as features of normal and arrhythmia HRV signals, used in training and test data in the decision tree classifiers. The use of multiple sizes of training dataset gives accuracy variation as well as the classification rate. As per the comparison of both feature sets and training dataset, Time-frequency feature is efficiently employable for identification of signal class that represents the cardiac health condition, its accuracy in classification reached up to 99.2% with 20:80 data distribution as a training and testing dataset.

Keywords: Arrhythmia · Sinus rhythm · ECG signal · Feature extraction · Classification

1 Introduction

In field of ECG signal processing, classification of ECG signals and its accuracy is most important and challenging task due to complex signal nature and associated health risk. Here, HRV signal itself a characteristic/feature; its exploited for the classification

of ECG characteristics [1]. It is very important in study of cardiac health condition. In the classification process features play a key role to identify the class of the signal that may indicate the particular health condition. These features basically consider as signature or characteristics of large data size samples.

In this study, the HRV signal that is a time-series data of *R-R intervals* in ECG signal is analyzed with its time domain and wavelet-based time-frequency domain statistical parameters to classify the signal classes; these classes directly explain the cardiac health condition. In the continuation of the introductory section, a short review of machine learning techniques in ECG signal processing. In Sect. 2, the time domain and wavelet-based time-frequency domain statistical parameters/features of the HRV signal are discussed, and its classification analysis is briefly explained in Sect. 3. Further, results, discussion, and a conclusion are elaborated in Sects. 4 and 5 respectively.

1.1 Machine Learning in ECG Signal Processing

Nowadays, automated, artificial intelligence or machine with intelligence system helps a lot in many services similar to healthcare; its intelligent machines work as assistive tool or system that help to medical experts in decision or healthcare planning. In the past two decades, several techniques are reported in the literature as systems developed for health monitoring and disease detection based on the vital features from signals and images [2–8]. Here, Machine learning (ML) techniques make the rules based on the labelled data in a system training; it can work as substitute of expert or physician involvement in case of accurate label data [9–13]. In the ECG signal processing, classification problems such as diagnosis of myocardial infarction (MI), arrhythmia detection, supraventricular condition, etc. causes cardiac arrest are identified and classified and 85% to 99% accuracy achieved with different machine learning methods using several ECG signal features and transform techniques [14]. However, ECG signal or variability feature are not standardized due to several conditions like age, stress, activity, lifestyle. Therefore, an ECG signal having a similar waveform pattern like *P*, *QRS*, *T*, and *U* in humans with normal health conditions; although it's different in individuals with abnormalities [1, 2]. Such that very difficult task for cardiologist or data scientist to recommend a common subjective feature for different health conditions to train the ML classifier.

The primary observation with ECG is heart rate and its variability that reflect the several cardiac conditions similar to waveforms. Although, heart rate variability (HRV) is itself a feature for the health interpretation by the physician. Here, HRV signal further explored for the classification of cardiac health condition using ML techniques.

2 HRV Signal and Feature Estimation

An automated real-time system helps to identify and classify the abnormalities in electrocardiogram (ECG) signal or its variability such as heart rate (HR).

The HRV signal are represent the variation of heart rate, it can be measured in term of R-R intervals. The variance of R-R intervals clearly represents the variability in cardiac rhythm; it may be normal/abnormal or specific cardiac decease. Here, different feature is calculated for two different domains such as time and wavelet-based time-frequency features. In Fig. 1, *R-R intervals* shown for the 1-min duration of arrhythmia and normal sinus rhythm ECG signal [15].

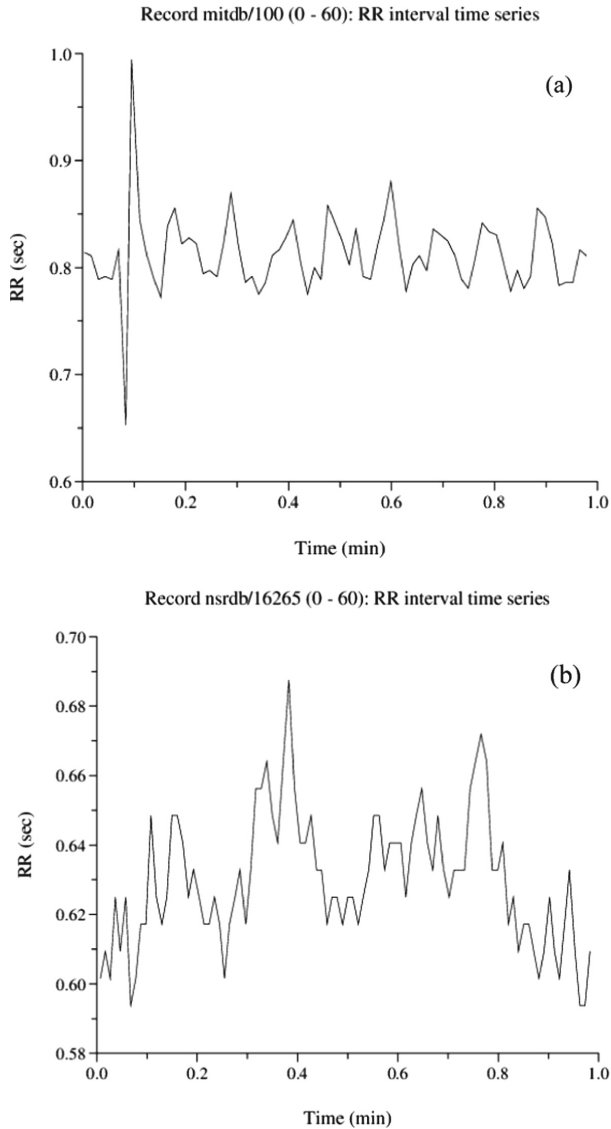


Fig. 1. *R-R interval* variability for the two different ECG signals: (a) MIT-BIH arrhythmia Rec. 100, and (b) NSR Rec. 16265 [15].

As per the pattern of *R-R interval* and HR for any ECG signal doctors or physician can identify the abnormality or normal condition with very short duration of time-series data. It is very difficult to identify or classify the cardiac health condition or pathological class with long duration time-series data or during real-time/ambulatory monitoring. The main feature for the any time-series data are mean, standard deviation, skewness and kurtosis that represent the center tendency, group deviation from mean value, symmetric behavior around the mean, and the tails of a distribution differ from the tails of normal distributed data respectively [15].

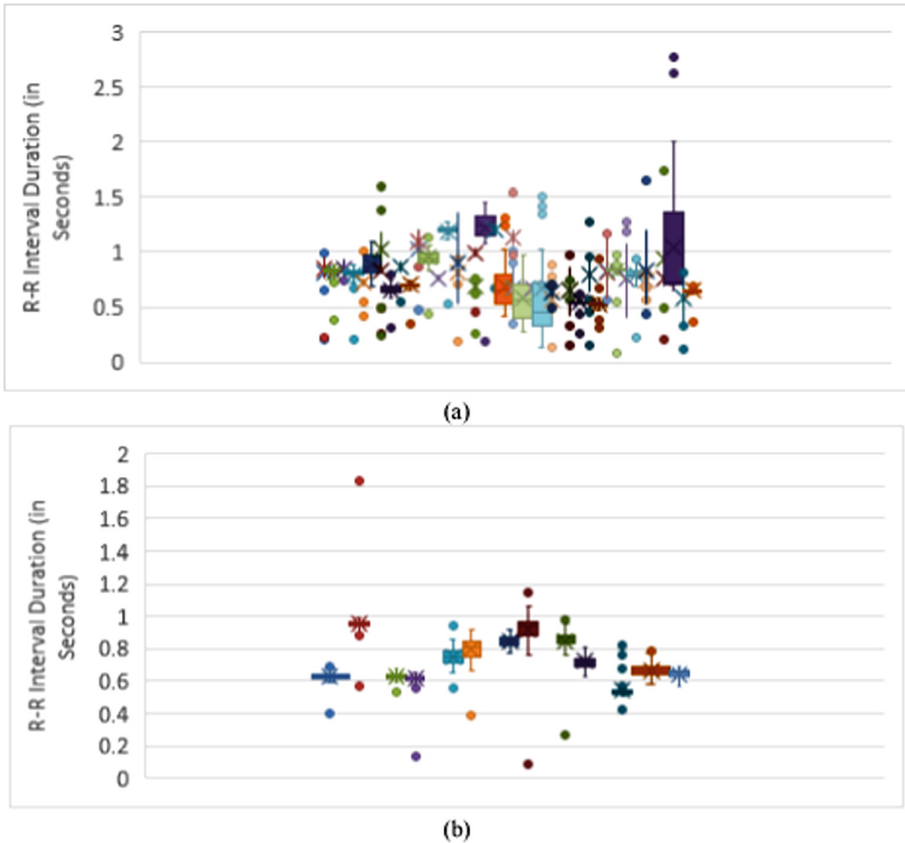


Fig. 2. R-R Interval distribution of different signals for 1-minute duration data from (a) MIT-BIH Arrhythmia (MITA), (b) Normal Sinus Rhythm Database (NSRDB). It represents the values from lower to upper quartile and the central line is representative of the centre data tendency.

In the Fig. 1 and 2, its clearly shown the variability of *R-R interval* for different signals from the both datasets. Although, range of variability is clearly visible for both groups of signals such as arrhythmia signal HRV spread between 0.5–1.2 s and NSR HRV spread

between the 0.6–0.9 s. However, these range of variability are overlapped with some signals of both the records, in such cases difficult to separate both case from each other.

Further, statistical measure evaluated for the different analysis like time-domain and time-frequency based on the wavelet coefficients. These statistical measures considered as features for the classifier as shown in Fig. 3 and summarized as follow:

- Step-1: HRV Signal Extracted for ECG Signal/database
- Step-2: Feature estimation in time/time-frequency domain
- Step-3: feature index and data label for the classification
- Step-4: Classification.

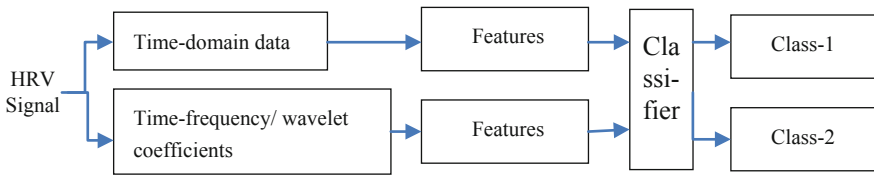


Fig. 3. Feature estimation for HRV signals in time-domain and time-frequency domain, and classification.

As discussed above, five features such as heart rate (HR), and mean, standard deviation (S_d), skewness (S_k) and kurtosis (K_t) estimated for the time domain data and wavelet domain time-frequency data [16]. Here, HRV data is decomposed at two-level with the coiflet (coif-5) wavelet function that produces one-approximation and three-detailed coefficients bands as reported in many literatures of wavelet and filter bank theory [17–22]. Therefore, features computed in both the domain such as time and time-frequency.

2.1 Time Domain Feature Analysis of HRV Signal

The detailed illustration is presented with Fig. 4 that represent the feature variability for both datasets. A comparison of both the dataset is illustrated in Fig. 4 with Box-plots and whisker diagram that represent the data/value distribution with its mean/median value of heart rate (HR), Mean, S_d , S_k , K_t features. The median HR is 78 and 81, Mean is 0.9 and 0.7, S_d is 0.09 and 0.05, S_k is -1.8 and -0.8 , K_t is 10 and 5 for MITA and NSRDB database respectively. Here, some features are clearly separable like S_d , S_k , and K_t and some are not clearly separable in both the dataset of $R-R$ interval time series data. It may affect the performance of classification.

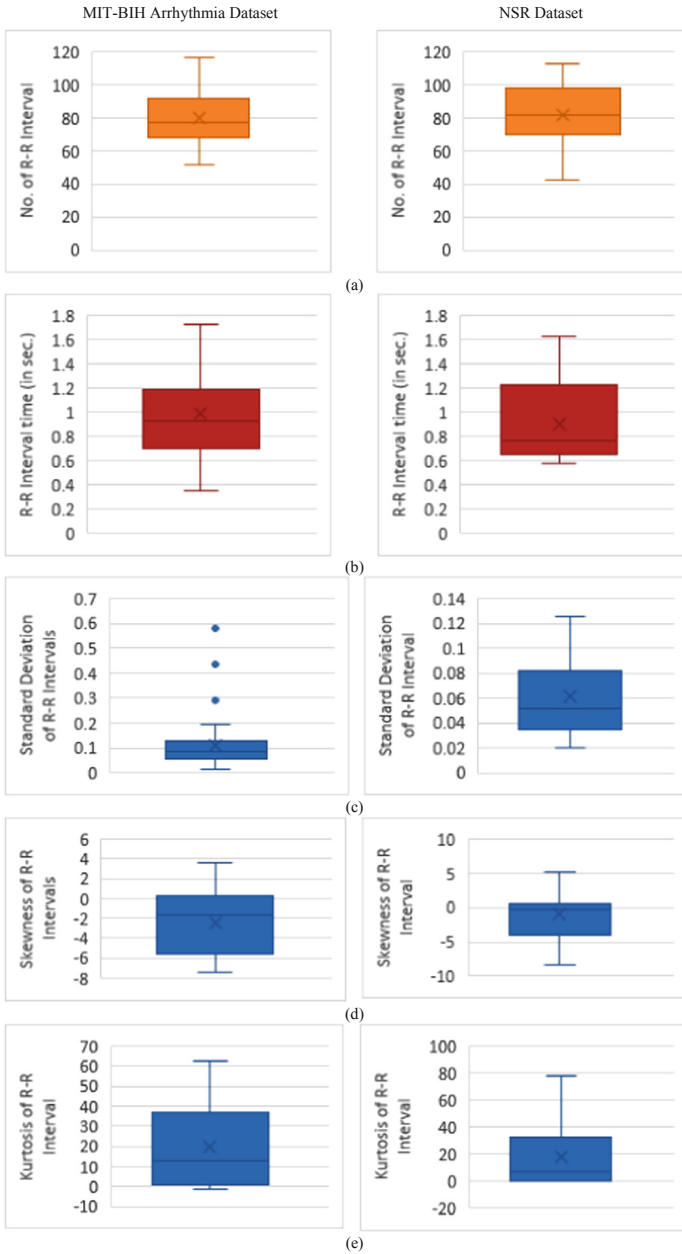


Fig. 4. Distribution of different time features from center tendency of different signals for MIT-BIH Arrhythmia (MITA), Normal Sinus Rhythm Database (NSRDB). The box represents the values from lower to upper quartile and the central line is representative of the median. The whiskers are expanded from lower to upper values. (a) HR, (b) mean, (c) standard deviation, (d) skewness, and (e) kurtosis

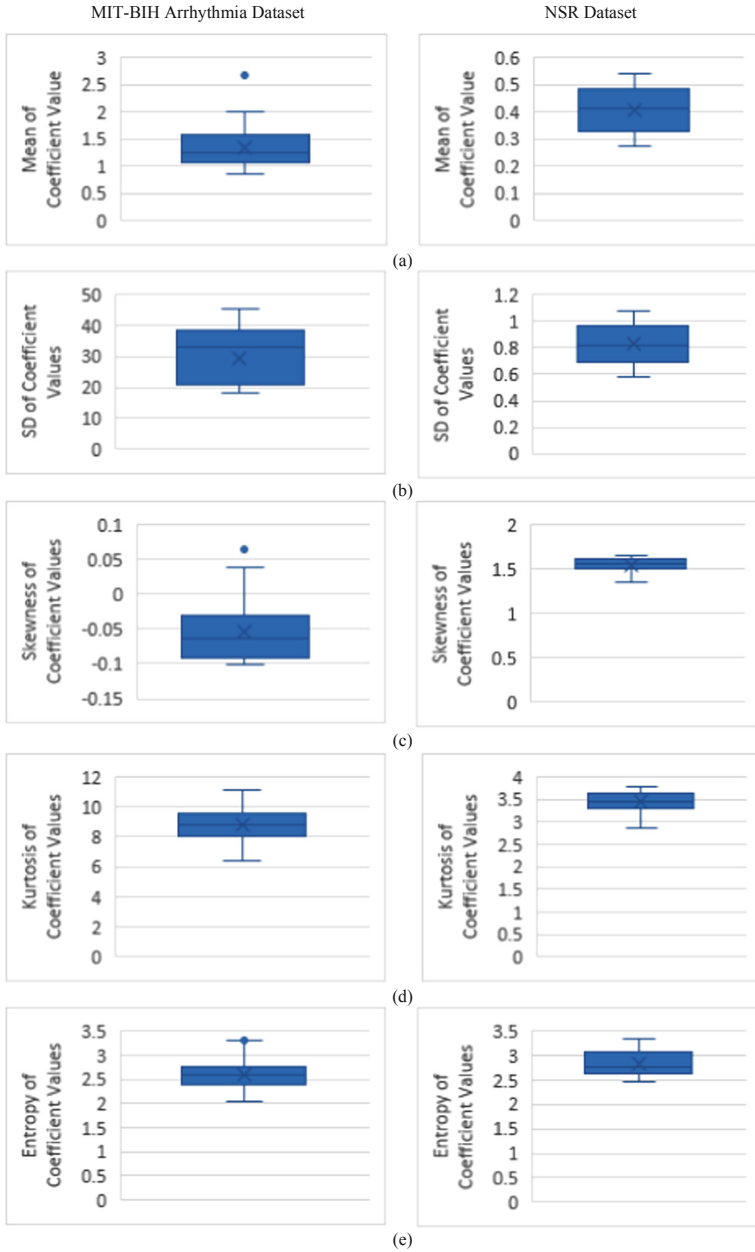


Fig. 5. Distribution of different wavelet-based time-frequency features of different signals for MIT-BIH Arrhythmia (MITA), Normal Sinus Rhythm Database (NSRDB). The box represents the values from lower to upper quartile and the central line is representative of the median. The whiskers are expanded from lower to upper values of wavelet coefficients. (a) mean, (b) standard deviation, (c) skewness, (d) kurtosis, and (e) entropy.

2.2 Time-Frequency Domain Feature Analysis of HRV Signal

The wavelet-based time-frequency domain features like Mean, standard deviation (S_d), Skewness (S_k), Kurtosis (K_t) and Entropy are calculated for all the records of both datasets. A comparison of both the dataset is illustrated in Fig. 5 with Box-plots and whisker diagram that represent the data/value distribution with its mean/median value of heart rate (HR), Mean, S_d , S_k , K_t features. The median of Mean is 1.2 and 0.4, S_d , is 35 and 0.82, S_k is 1.6 and -0.06 , K_t is 9 and 3.5, and entropy is 2.6 and 2.8 for MITA and NSRDB database respectively.

Here, some features are clearly separable like Mean, S_d , S_k , and K_t ; where, entropy are not clearly separable in both the dataset of $R-R$ interval time series data. It may affect the performance of classification.

As discussed, HRV signal itself a feature for the cardiac rhythm and may considered for the primary examination. However, statistical analysis as shown in the Fig. 4 and 5 time based features such as HR and $R-R$ interval not clearly separable for health condition. Further, other features are exploited for the classification of signal.

3 Classification of HRV Signal: Proposed Framework

In this paper, MIT-BIH Arrhythmia and Normal sinus rhythm (NSR) ECG databased is studied for the different classes of disease related to the heart [15]. Here, total 61 HRV signals divided into 48 records that containing Arrhythmia signals (MITA) beats and 13 Normal sinus rhythm (NSRDB) records. As the duration of MITA and NSRDB records is 30 min and 24 h respectively, the training and tests data are composed as 1 min of data for features extraction in classification application. The dataset processed with two different type of features such as time domain feature and wavelet-based time-frequency features as discussed. Here, Decision-Tree (DT) classification explore the for the classification of two-class data set based on the feature's characteristics; it makes the rules for classify the binary class problem with minimum computation cost and fast processing [14, 23].

The performance of the system is evaluated from the statistical rules. The terms used in evaluating are defined as: TP: true positive, TN: true negative, FP: false positive, FN: false negative. In order to estimate the classifiers performances another index namely Predictability (Pr), Sensibility (Se) and Accuracy (Acc) are calculated as follow [5, 14]:

$$Se = \frac{TP}{TP + FN}, Pr = \frac{TP}{TP + FP}, Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

Here, two different domain features are evaluated for the classification of HRV signal such as Arrhythmia and Normal Sinus classes. As per the Fig. 4 and 5, it is clearly shown the time domain features are not clearly separated for the different class and wavelet-based time-frequency features are clearly distinguishing to each other. It is clearly illustrated through statistical data in Table 1 and 2. Here, both the feature dataset

is divided as training and test dataset as ratio of 10:90, 20:80, and 40:60 respectively, further dataset exploited for classification based on the decision-tree method.

4 Results and Discussion

In Table 1, signals are classified 80%, 79.6%, and 78.4% as a TP for 10%, 20% and 40% dataset as training set from time-domain features of total 61 signals respectively. Similarly, signals are classified 78.2%, 99%, and 99% as a TP for 10%, 20% and 40% dataset as training set from wavelet-based time-frequency domain features of total 61 signals respectively as listed in Table 2.

Table 1. Statistical data result for time domain features

Parameters	Size of training dataset (in %)		
	10	20	40
TP	0.8	0.796	0.784
TN	0	0	0
FP	0.2	0.204	0.216
FN	0	0	0

Table 2. Statistical data result for wavelet-based time-frequency domain features

Parameters	Size of training dataset (in %)		
	10	20	40
TP	0.782	0.99	0.99
TN	0	0	0
FP	0.218	0	0
FN	0	0	0

As discussed, previous section, training and test feature datasets are classified using the decision tree algorithm. From Tables 3 and 4, it clearly shown that wavelet-based feature is more suitable for the classification task. The analysis of features as illustrated in Figs. 4 and 5 is also support the achieved results of classifier, because of time based feature are not distinguish in between two classes.

Table 3. Performance of classifier for time domain features

Performance	Size of training dataset (in %)		
	10	20	40
Sensitivity (%)	100	100	100
Predictivity (%)	80	79.6	78.4
Accuracy (%)	80	79.6	78.4

Table 4. Performance of classifier wavelet-based time-frequency domain features

Performance	Size of training dataset (in %)		
	10	20	40
Sensitivity (%)	100	100	100
Predictivity (%)	78.2	99.2	99.2
Accuracy (%)	78.2	99.2	99.2

The performance of classifier based on two features dataset is summarized in Table 3 and 4 as sensitivity, predictivity and accuracy parameters. The accuracy of wavelet-based time-frequency features is 99.2% for 20% and 40% of training dataset. Where, maximum accuracy 80% for 10% of training dataset time domain feature.

These results achieved based on the separable and non-separable features of time-series and time-frequency data for the binary class problem using the decision tree technique.

5 Conclusion

A system for heart rate variability (HRV) classification is provided assistance to the cardiologist in the planning of healthcare against the mortality due to cardiac arrest. Here, a statistical technique is used for analysis of HRV data and decision tree for classification of parameters extracted from HRV signals. Different separable and non-separable statistical features/parameters of HRV signals were computed and considered as features of normal and arrhythmia HRV signals, used in training and test data in the Decision Tree classifiers. Further, division in feature dataset in multiple size as training dataset give variation in accuracy as well as classification rate. As per comparison of both feature sets and training dataset, Time-frequency feature are efficiently employable for identification of signal class that represent the cardiac health condition. It gives the 99.2% accuracy and classification rate with 20% and 40% size of training dataset due to clear separation of both domain features.

Acknowledgement. This work is supported by research grant under Competitive Research of Technical Education Quality Improvement Programme (TEQIP-III) of Uttarakhand Technical University, Dehradun.

References

1. Acharya, U.R., Joseph, K.P., Kannathal, N., Lim, C., Suri, J.: Heart rate variability: a review. *Med. Biol. Eng. Comput.* **44**, 1031–1051 (2006)
2. Navin, O., Kumar, G., Kumar, N., Baderia, K., Kumar, R., Kumar, A.: R-peaks detection using shannon energy for HRV analysis. In: Rawat, B.S., Trivedi, A., Manhas, S., Karwal, V. (eds.) *Advances in Signal Processing and Communication*. LNEE, vol. 526, pp. 401–409. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-2553-3_39

3. Budoff, M.J., Shinbane, J.S.: Cardiac CT Imaging: Diagnosis of Cardiovascular Disease. Springer, Berlin (2016). <https://doi.org/10.1007/978-1-84882-650-2>
4. Ceylan, R., Özbay, Y.: Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network. *Expert Syst. Appl.* **33**, 286–295 (2007)
5. Fujita, H., et al.: Sudden cardiac death (SCD) prediction based on nonlinear heart rate variability features and SCD index. *Appl. Soft Comput.* **43**, 510–519 (2016). <https://doi.org/10.1016/J.ASOC.2016.02.049>
6. Bahadure, N.B., Ray, A.K., Thethi, H.P.: Comparative approach of MRI-based brain tumor segmentation and classification using genetic algorithm. *J. Digit. Imaging* **31**(4), 477–489 (2018). <https://doi.org/10.1007/s10278-018-0050-6>
7. Avola, D., Cinque, L., Placidi, G.: Customized first and second order statistics based operators to support advanced texture analysis of MRI images. *Comput. Math. Methods Med.* **2013**, 213901 (2013). <https://doi.org/10.1155/2013/213901>
8. Banday, S.A., Mir, A.H.: Statistical textural feature and deformable model based brain tumor segmentation and volume estimation. *Multimed. Tools Appl.* **76**(3), 3809–3828 (2016). <https://doi.org/10.1007/s11042-016-3979-9>
9. Salem, A.M., Revett, K., El-Dahshan, E.A.: Machine learning in electrocardiogram diagnosis. In: 2009 International Multiconference on Computer Science and Information Technology, pp. 429–433 (2009)
10. Jambukia, S.H., Dabhi, V.K., Prajapati, H.B.: Classification of ECG signals using machine learning techniques: a survey. In: 2015 International Conference on Advances in Computer Engineering and Applications, pp. 714–721 (2015)
11. Zubair, M., Kim, J., Yoon, C.: An automated ECG beat classification system using convolutional neural networks. In: 2016 6th International Conference on IT Convergence and Security (ICITCS). pp. 1–5 (2016)
12. Hegde, V.N., Deekshit, R., Satyanarayana, P.S.: Comparison of characterizing and data analysis methods for detecting abnormalities in ECG. In: 2011 Second Asian Himalayas International Conference on Internet (AH-ICI), pp. 1–5 (2011)
13. Yu, S.-N., Chen, Y.-H.: Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network. *Pattern Recognit. Lett.* **28**, 1142–1150 (2007). <https://doi.org/10.1016/J.PATREC.2007.01.017>
14. Alarsan, F.I., Younes, M.: Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. *J. Big Data* **6**(1), 1–15 (2019). <https://doi.org/10.1186/s40537-019-0244-x>
15. Mark, R., Moody, G.: The impact of the MIT-BIH arrhythmia database history, lessons learned, and its influence on current and future databases (2001)
16. Alekseev, M.A.A.: Statistical methods of ECG signal processing in diagnostics of coronary artery disease. *Int. J. Cardiovasc. Res.* **7** (2018). <https://doi.org/10.4172/2324-8602.1000339>
17. Kumar, A., Singh, G.K., Kumar, R.: Electrocardiogram signal compression using singular coefficient truncation and wavelet coefficient coding. *IET Sci. Meas. Technol.* **10**, 266–274 (2016). <https://doi.org/10.1049/iet-smt.2015.0150>
18. Kumar, R., Kumar, A., Singh, G.K.: Electrocardiogram signal compression based on singular value decomposition (SVD) and adaptive scanning wavelet difference reduction (ASWDR) technique. *AEU - Int. J. Electron. Commun.* **69**, 1810–1822 (2015). <https://doi.org/10.1016/j.aeue.2015.09.011>
19. Kumar, R., Kumar, A., Singh, G.K.: Electrocardiogram signal compression using singular coefficient truncation and wavelet coefficient coding. *IET Sci. Meas. Technol.* **10** (2016). <https://doi.org/10.1049/iet-smt.2015.0150>

20. Ranjeet, K., Kumar, A., Pandey, R.K.: ECG signal compression using different techniques. In: Unnikrishnan, S., Surve, S., Bhoir, D. (eds.) ICAC3 2011. CCIS, vol. 125, pp. 231–241. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-18440-6_29
21. Ranjeet, K., Farida, J.: Retained signal energy based optimal wavelet selection for denoising of ECG signal using modified thresholding. In: 2011 International Conference on Multimedia, Signal Processing and Communication Technologies, IMPACT 2011 (2011)
22. Kumar, A., Ranjeet: Wavelet based electrocardiogram compression at different quantization levels. In: Das, V.V., Thomas, G., Lumban, G.F. (eds.) Information Technology and Mobile Communication. AIM 2011. Communications in Computer and Information Science, vol. 147. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-20573-6_69
23. Sannino, G., De Pietro, G.: A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. *Futur. Gener. Comput. Syst.* **86**, 446–455 (2018). <https://doi.org/10.1016/J.FUTURE.2018.03.057>