A REVIEW ON ECG SIGNAL FEATURE EXTRACTION AND CLASSIFICATION TECHNIQUES

Dr. Naveen Kumar Dewangan[1*], Kiran Dewangan[2], Dr. Amar Kumar Dey[3], Sagar Singh Rathore[4]

Article History: Received: 26.04.2023 **Revised:** 21.06.2023 **Accepted:** 19.07.2023

Abstract

Development of automatic detection methods for identifying different heart irregularities or arrhythmias was required due to the rising number of heart patients in order to ease the burden on doctors. An electrocardiogram (ECG) records the heart's electrical activity and can be used to detect heart problems. ECG signals are widely utilized to categories the signals into several classes that aid medical professionals in identifying heart disorders. It is very challenging to classify an ECG accurately. In recent years, various techniques have been employed and investigated for the classification of cardiac signals and heart rhythm problems. The classification of ECG signals often uses temporal, morphological, fast Fourier transform, wavelet features, statistical features, correlation, and regression techniques. In this study, strategies for extracting and analyzing ECG signals are reviewed. It has been found that a hybrid feature extraction improves detection effectiveness. The majority of authors have worked on to classify ECG into 5 categories. Therefore, there is room to find the elements that work best together to deliver optimal effectiveness for a larger number of ECG classes. Future work is possible with a classifier that can deliver the highest accuracy with a larger number of classifications.

Keywords - ECG signal, preprocessing, feature extraction and selection, classification.

[1*,2,3] Department of E & TC, Bhilai Institute of Technology, Durg (C.G.), India
[4] G H Raisoni College of Engineering, Nagpur, India

[*]Corresponding Author:
Dr. Naveen Kumar Dewangan[1*]
Department of E & TC, Bhilai Institute of Technology, Durg (C.G.), India
Email: [1*]naveen.dewangan@bitdurg.ac.in

DOI: 10.31838/ecb/2023.12.6.210

1. INTRODUCTION

Cardiovascular diseases (CVDs), which account for 30% of all disease-related fatalities globally, are the leading cause of mortality as per the 2010 Global Status Report on Non-Communicable Diseases published by the World Health Organization (WHO) [1]. Before a heart attack occurs suddenly, symptoms including chest pain, faintness, shortness of breath, and palpitations may appear. It will be very beneficial if these peculiar symptoms can be timely and easily recognized and identified, allowing the individual to receive the appropriate care at right time. Consequently, it is essential to establish a quick, easy, and accurate way for identifying and classifying different heart disorders. The purpose of this study is to provide some relevant information from studies related to the improvement in heart disease classification with major contributions from ECG analysis from numerous researchers worldwide.

According to Chazal de Philip et al. (2004), the four basic phases in the ECG signal recognition system as shown in figure 1 are the signal preprocessing, P wave, T wave, QRS complex detection, suitable feature extraction, and categorization of ECG signal. [2]

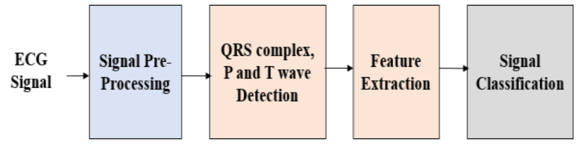

Figure. 1 ECG signal classification system.

2. REVIEW OF THE LITERATURE

The primary goal of ECG analysis is to increase degree of accuracy and incorporate more categorizable heart disorders. By examining the ECG signals, it is possible to identify a wide variety of heart problems. P, T waves, and QRS complexes are clearly delineated in ECG signals. Medical professionals or experts typically analyze manually the wave duration and amplitude [3]. Due to the enormous amount of data, manual analysis will be arduous and time-consuming, and there is a chance that the analyst may overlook crucial information [3]. That is why, developing a computer-based system for categorizing ECG signals and improving ECG analysis techniques are essential.

Based on the various extracted features that have been retrieved from the ECG signals, numerous methods have been developed to categorise ECG beats. Finding the smallest collection of features necessary to obtain respectable classification rates is the aim of the feature extraction stage. Without training and testing the classifier, the analyst is unable to estimate the performance of a set of features. As a result, a crucial and repetitive process is required for optimum feature selection which requires training various feature sets until acceptable classification performance is attained. [3]

Various studies and algorithms have been created recently to make it simple to analyse and categorise ECG signals.
Sarkaleh and Shahbahrami (2012) employed the Multi-Level Perceptron (MLP) Neural Network (NN) for the classification job while the Discrete Wavelet Transform (DWT) was used to extract some characteristic features. The suggested technique is able to identify two different forms of arrhythmias. The 24 statistics that make up the retrieved feature vector cover the whole range of wavelet coefficients, from level one to level eight. Maximum wavelet detail coefficients, minimum wavelet coefficients, and variance of wavelet detail coefficients at each level are included in the set of characteristics employed. 96.5% overall recognition rate was attained. [4]

To identify and categorize four different types of arrhythmias with a typical ECG signal, Saini and Saini (2012) employed Neural Network (NN) technology with error back propagation method. Twenty hidden nodes with sigmoid activation functions were used in an MLP network. 3 and 5 neurons are fixed at the input layer and output layer respectively. For the purpose of classifying arrhythmias, the three morphological features of the ECG signal - the RR interval, R peak amplitude, and QRS duration - were employed. More than 95% of sensitivity and accuracy were attained. [5]

Five different beat classifications of arrhythmia were examined by Martis R. J. et al. (2013) in their paper. For reducing the dimensionality, the DWT sub bands were subjected to individual linear discriminant analysis (LDA), principal component analysis (PCA), and independent component analysis (ICA) applications. For automated diagnosis, these dimensionality-reduced characteristics were input into the classifiers of the neural network (NN), probabilistic neural network (PNA) and Support Vector Machine (SVM). ICA features outperformed PCA and LDA when combined with a PNN with a spread value (σ) of 0.03. Using a ten-fold cross validation scheme, Average scores for different parameters is shown in Table I as according to [6].

Table I: Average Score Values of Parameters.

Sensitivity	Specificity	Positive Predictivity	Accuracy
99.97%	99.83%	99.21%	99.28%

A novel technique for automatic ECG beat classification for Holter monitoring was put forth by Park and Kang (2014). They precisely extracted features like the QRS complex and P wave using the Pan-Tompkins technique, and used a decision tree to categorize each beat according to these characteristics. Heartbeat classification accuracy is 94.6% and 99%, respectively, according to tests performed on the MIT-BIH arrhythmia database and on personalized patient's own ECG database. These demonstrate the effectiveness of our suggested strategy because they are comparable to outcomes from cutting-edge schemes. [7]

Dewangan Naveen Kumar et al. (2016) used 12 feature set; four morphological features: RR interval, QRS duration, PR interval, R peak amplitude and eight wavelet features. Wavelet features were obtained after decomposition of each ECG beat into eight levels and the variance of the detail coefficients d_1 to d_8 was calculated to form a feature set. First-degree AV block, paced beats, atrial premature beats, left bundle block, right bundle block were the five forms of arrhythmias that they classified apart from normal beats. Proposed 3-layer feedforward backpropagation neural network method achieved overall average accuracy of 87%. [8]

A novel approach for categorizing various types of arrhythmias using morphological and dynamic data was proposed by Bassiouni Mahmoud M. et al. in 2018. Each heartbeat is subjected to the discrete wavelet transform (DWT) to acquire the morphological information. As a dynamic feature, RR interval data is utilized. Teager energy operator is used to capture the RR interval's nonlinear dynamics, which enhances the categorization of arrhythmias. Twelve coefficients are selected as morphological features after redundant information is eliminated from DWT sub bands using independent component analysis. To classify arrhythmia, these hybrid traits are merged and given to a Neural network. Using three-fold cross validation, the proposed methodology improved the average accuracy for the class- and subject-oriented schemes, respectively, to 99.75% and 99.84%. [9]

According to Rangappa V. G. et al. (2018), A series of serious cardiac problems, including ischemic heart disease, heart failure, and myocardial infarction, can lead to coronary artery disease. In this work, a three-step process was used to distinguish between five types of ECG beats. The Pan-Tompkins method is used in the first step to locate the peaks in the ECG data. In the second phase, higher order ECG statistics are combined with the extracted three interval features. The third phase of the classification process for ECG beats uses the K-Nearest Neighbor (KNN) algorithm. This method accurately classified the ECG signals as abnormal or normal. The outcomes showed that the proposed method can separate signals with an accuracy of up to 98.40%. [10]

Because of the heart's erratic, dynamic, and nonlinear behaviour, Kiani Kourosh et al. (2019) claimed that the extracted properties of the ECG signal are a right depiction of the function of heart. The ECG signal is best represented by the fractal dimension because it can account for its hidden complexity. Back Propagation Neural Network (BPN) and the fractal dimension are used to analyze ECG signals. In this work, a novel method for accurately classifying seven arrhythmias from ECG signals utilizing the fractal dimension is presented. This technique can pinpoint the precise location of arrhythmias. The fractal dimension and BPN are utilized to categorize using a combination of 5 credible universal databases. Indices such as sensitivity and specificity are used to evaluate the effectiveness of the proposed strategy. Sensitivity 99.74% and specificity 96.84% was achieved. The data show that this strategy has a 98.83% accuracy rate. [11]

Using ECG signals, Dalal Sahil et al. (2021) established a novel, robust method for quickly and accurately identifying human being. Discrete wavelet transform (DWT) is used to remove noise and features are obtained with multi-cumulants. This method primarily relies on multi-cumulant features that are derived from the ECG data. Kernel Extreme Learning Machine (KELM) is used to classify the multi-cumulants feature based ECG data. Utilizing genetic algorithms (GA), the parameters of multi-cumulants and KELM are optimized. [12]

Halemirle A. D. et al. created a hybrid feature-based classification method in 2021 that makes use of the SVD-Entropy, Dual-Tree Complex Wavelet Transform, Autoregressive modelling, and feature extraction from multifractal analysis. The extracted features are subsequently categorized using the MIT-BIH database using the K-Nearest Neighbours (KNN), Random Forest Classifier, and Bayesian Optimized-KNN classifiers in order to achieve the best classification. The random forest classifier's highest accuracy is 98.29%. The findings support the methodology's validity and applicability as a tool to help detect cardiac disease in hospital settings. [13]

The main goal of Sehirli Eftal et al.'s (2021) work was to create an intelligent system based on the processing, analysis, and categorization of ECG signals using a hybrid machine learning model. This uses 837 fragments of ECG data from 7 classes from the MIT/BIH Arrhythmia database for a single lead. To smooth the signals and set the baseline correctly, the ECG signals are preprocessed. k-means clustering and local extrema points are used to segment ECG signals that contain the Q, R, and S waves (QRS complex). The measurement parameters for each QRS complex are independently calculated after feature extraction and selection to create a data set. Utilizing eight-fold cross validation, training sets and test sets are produced. Cardiovascular diseases (CVDs) were divided into seven categories in the suggested work. Several machine learning models, including decision trees, k-nearest neighbors, random forests, naive bays, linear discriminant analyses, support vector machines, and quadratic discriminant analyses are combined to create a hybrid model. The values for the CVD classes' sensitivity, specificity, accuracy, and MCC are 92.33%, 92.50%, 92.41%, and 0.85, respectively [14].

In order to categorize the five types of ECG beats in the MIT-BIH arrhythmia database, Wu Mengze et al. (2021) suggested a reliable and effective 12-layer deep one-dimensional convolutional neural network. The ECG signal is denoised using a wavelet self-adaptive threshold approach. The experimental results demonstrate that the method described in this research performs excellent in sensitivity, accuracy, anti-noise capabilities, and robustness than random forest, BP neural network, and other CNN networks. Its precise classification efficiently conserves medical resources, improving clinical practice. The suggested work classified ECG beats into five categories: normal, atrial premature beats, premature ventricular beats, eft bundle branch block, and right bundle branch block It is interesting that the proposed CNN network outperforms the BP neural network, random forests, and other CNN networks in terms of accuracy and robustness. The proposed CNN network performs exceptionally well with the classification parameter values shown in Table II. [15]

Table II: Average Score Values of Parameters.

Sensitivity	Specificity	Positive Predictivity	Accuracy
97.05%	99.35%	97.21%	97.41%

Using RCG signals from the MIT-BIH database, Sahoo Santanu et al. (2022) proposed a deep learning approach for automated identification of cardiac arrhythmia. For denoising the ECG signal several decomposition approaches, such as variational mode decomposition, empirical mode decomposition, and discrete wavelet transform, were used. Using denoised signals, the time-frequency based multi-domain characteristics were extracted from the various coefficients of the sub-bands. To choose the most informative features for higher classification accuracy, these derived features were prioritized using the Chi-squared test and particle swarm optimization (PSO)-based algorithms. Five different types of ECG beats were categorized using the hybrid characteristics by a deep neural network (DNN) using a ten-fold cross validation technique. The best outcomes were attained using the Chi squared selection technique, with a computational complexity of 0.14 second and an accuracy of 99.75%. In hospitals, the presented method can be used to automatically detect ECG beats that are abnormal [16].

Based on the ANSI-AAMI standard, a new bidirectional long-term short-term memory network (BLSTM) and deep convolutional neural network (CNN) model was developed by Bhatia, S. et al. (2022) to automatically classify ECG heartbeats into five separate groups. With this hybrid model, feature extraction and classification are combined to do end-to-end learning without the need for human feature extraction. The MIT-BIH arrhythmia database, which is accessible to the public, is used for the experiment. The results were contrasted with those of the two other hybrid deep learning models, which combine LSTM and CNN and Gated Recurrent Unit and CNN, respectively. Additionally, the model's performance was compared with earlier studies published in the literature. To solve the issue of the class imbalance, this database was artificially oversampled using the SMOTE method. This innovative hybrid model was developed using tenfold cross-validation on the real test dataset and trained on the oversampled ECG database. Experimental findings show that the proposed hybrid model performs better than existing techniques in terms of F-score performance, recall, accuracy and precision with scores of 91.67%, 94.36%, 98.36%, 89.4%, respectively. [17]

3. METHODOLOGY

Summary of different approaches for the classification of ECG signal is summarized in the Table III. It is observed that classification accuracy has been significantly increased by the selection of hybrid features [4]-[17] and maximum number of classes classified is 7 [11] and [14]. Neural network [6] and deep learning methods [16] are achieving highest accuracy of more than 99%. Based on these observations a hybrid approach will be proposed which will utilize neural network and deep learning method and hybrid set of features like temporal, morphological, fast Fourier transform, wavelet features, statistical features, entropy, correlation, regression, a non-fiducial, fiducial and a fusion approach are combined to classify more than 7 classes of ECG signal. The proposed system is expected to classify greater than 7 heart diseases based on ECG signal with multiclass accuracy of more than 99%.

4. CONCLUSION

The majority of researchers have created systems based on different methodologies and algorithms in the literature. The developed detection system's performance is quite encouraging, but it still has to be further assessed. Each method for classifying ECG signals that was previously developed has advantages and drawbacks. Table III lists different methods for classifying ECG signals. All of the methods mentioned above have the following drawbacks:

(1) Most approaches have only been tried on smaller data sets; thus, it is necessary to confirm their performance on bigger databases.

(2) All arrhythmia classes need to be tested; only certain classes have undergone evaluation.
(3) On classes of arrhythmia that occur infrequently, the classification accuracy is poor.

For the accurate diagnosis of heart disorders, automated ECG wave detection is essential. The QRS complex, as well as the T and P waves, must be reliably and accurately detected for an automatic ECG analyzing system to operate successfully. The majority of researchers solely focus on certain diseases. From the examined feature extraction and classification strategies for ECG analysis, it is observed that hybrid features and methods are one of the most recent ECG analysis techniques which is being carried out by the researchers. Expanding the types of cardiac disorders that may be accurately diagnosed by employing various algorithms for feature extraction, hybrid feature selection, and classification tasks should be the main objectives of research.

5. REFERENCES

1. Global status report on non-communicable diseases 2010. [Online]. Available: https://healthyoupromo.wordpress.com/
2. Chazal de Philip, O. Dwyer Maria and Reilly Richard B., 2004, "Automatic Classification of Heartbeats using ECG Morphology and Heartbeat Interval Features", IEEE Transactions on Biomedical Engineering, 51(7): 1196-1206.
3. Dewangan Naveen Kumar and Shukla, S.P., 2015, "A survey on ECG signal feature extraction and analysis techniques," International journal of innovative research in electrical, electronics, instrumentation and control engineering, 3(6): 12-19.
4. Sarkaleh Maedeh Kiani Sarkaleh and Shahbahrami Asadollah, 2012, "Classification of ECG arrhythmias using discrete wavelet transform and neural networks," International Journal of Computer Science, Engineering and Applications (IJCSEA), 2(1): 1-13
5. Saini Indu and Saini B. S., 2012, "Cardiac arrhythmia classification using error back propagation method," International Journal of Computer Theory and Engineering, 4(3): 462-464.
6. Martis R. J., Acharya U. R., and Min L. C., 2013, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform," Biomedical Signal Processing and Control, 8(5): 437–448.
7. Park Juyoung , Kang Kyungtae , 2014, "Personalized classification of heartbeat types using a decision tree", Computers in Biology and Medicine, 54: 79-88.
8. Dewangan N. K. and Shukla S. P., 2016, "ECG arrhythmia classification using discrete wavelet transform and artificial neural network," IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2016, pp. 1892-1896.
9. Bassiouni Mahmoud M., El-Dahshan El-Sayed A., Khalefa Wael and Salem Abdelbadeeh M., 2018, "Intelligent hybrid approaches for human ECG signals identification", Signal, Image and Video Processing, 12: 941–949.
10. Rangappa Vedavathi Gauribidanur, Varaprasad Sahani Venkata Appala and Agarwal Alok, 2018, "Classification of Cardiac Arrhythmia stages using Hybrid Features Extraction with K-Nearest Neighbour classifier of ECG Signals", International Journal of Intelligent Engineering & Systems, 11 (6): 21-32.
11. Kiani Kourosh and Maghsoudi Farzane, 2019, "Classification of 7 Arrhythmias from ECG Using Fractal Dimensions", Journal of Bioinformatics and Systems Biology, 2 (3): 053-065.
12. Dalal Sahil & Vishwakarma Virendra P., 2021, "Classification of ECG signals using multi-cumulants based evolutionary hybrid classifier", PubMed, 11(1): 15092.
13. Halemirle Appaji Deepak and Thippeswamy Vijayakumar, (2021), "ECG Signal Classification with Hybrid Features Using Bayesian Optimized K-Nearest Neighbors Classifier", International Journal of Intelligent Engineering & Systems, 14(6): 50-65.
14. Sehirli Eftal and Turan Mohammed Kamil, 2021, "A Novel Method for the Segmentation of QRS Complex on ECG Signals and Classify Cardiovascular Diseases via a Hybrid Model Based on Machine Learning", International Journal of Intelligent Systems and Applications in Engineering, 9(1): 1-10.

15. Wu Mengze, Lu Yongdi , Yang Wenli , Wong Shen Yuong, 2021, “A Study on Arrhythmia via ECG Signal Classification Using the Convolutional Neural Network”, Frontiers Computer Neuro Science, 14: 1-10.
16. Sahoo Santanu, Pratyusa Dash, Mishra B.S.P., Sabut Sukanta Kumar, 2022, “Deep learning-based system to predict cardiac arrhythmia using hybrid features of transform techniques”, Intelligent Systems with Applications, Science Direct, 16: 1-10.
17. Bhatia Surbhi, Pandey Saroj Kumar, Kumar Ankit and Alshuhail Asma, 2022, “Classification of Electrocardiogram Signals Based on Hybrid Deep Learning Models”, Sustainability 2022, 14, 16572: 2-15.

Table III: Summary of Approaches for ECG Signal Classification.

Authors	**Features Extracted**	**Classification Techniques Used**	**ECG Beats Classified**	**Performance of the Classifier**
Sarkaleh M. K. et al. (2012)	Used detail coefficients obtained by eight level decomposition of ECG signal	Multi-Level Perceptron (MLP) Neural Network (NN)	2 classes	Recognition rate 96.5%
Saini Indu and Saini B. S. (2012)	3 morphological features	Artificial neural network	4 classes	Sensitivity and accuracy 95%
Martis R. J. et al. (2013)	Dimensionalities reduced features obtained after DWT of the ECG signal	Support Vector Machine (SVM), neural network (NN) and probabilistic neural network (PNN)	5 classes	Accuracy 99.28%
Park Juyoung and Kang Kyungtae (2014)	QRS complex, P wave	Decision tree	2 classes	Accuracy 94.6
Dewangan Naveen Kumar et al. (2016)	4 morphological features and 8 wavelet features (R peak, QRS duration, PR interval and RR intervals as morphological features and variance of detail coefficients d_1-d_8 after 8 level decomposition as wavelet features)	Three-layer Feed forward Back Propagation Neural Network	6 classes	Average accuracy 87%
Bassiouni M. et al. (2018)	Features are obtained by using three different intelligent approaches: a non-fiducial, fiducial and a fusion approach between them.	Artificial neural network (ANN), K-nearest neighbor (KNN), support vector machine (SVM)	6 classes	Average accuracy of 99.75%
Rangappa V. G. et al. (2018)	Pan-Tompkins algorithm (PTA) is used for detecting the peaks in ECG signals. Extraction of three interval features combined with ECG higher order statistics.	K-Nearest Neighbour (KNN) technique	5 classes	Accuracy 98.40%
Kiani Kourosh et al. (2019)	Hybrid Features	Back Propagation neural network	7 classes	Accuracy 98.83%
Halemirle A. D. et al. (2021)	Hybrid features i.e., Dual-Tree Complex Wavelet Transform (DTCWT), SVD-Entropy, Autoregressive modeling, and Multifractal analysis-based feature extraction	Random Forest Classifier, K-Nearest Neighbors (KNN), and Bayesian Optimized-KNN classifiers	5 classes	Highest accuracy achieved in random forest classifier is 98.29 %
Sehirli Eftal et al. (2021)	Training sets and test sets are created using 8-fold cross validation	A hybrid model based on machine learning models such as decision tree (DT), k-nearest neighbor (KNN), random forest (RF), naive bays (NB), linear discriminant analysis (LDA), support vector machine (SVM), and quadratic discriminant analysis (QDA) is developed	7 classes	The values for the CVD classes' sensitivity, specificity, accuracy, and MCC are 92.33%, 92.50%, 92.41%, and 0.85, respectively
Santanu Sahoo et al. (2022)	DWT, EMD and VMD are used to de-noise the ECG signal. The time-frequency based multi-domain features are extracted from the various coefficients of the sub-bands from de-noised signals along with RR intervals.	A deep learning approach	5 classes	Accuracy of 99.75%