ISSN: 2581-3404 (Online) IF: 5.68 (SJIF)
IJIRTM, Volume-8, Issue-5, October-2024.

A Survey on ECG Signal Preprocessing Techniques for Heart Disease Diagnosis

Tanuja Goswami[1], Dr. Amit Saxena[2]
[1]Research Scholar, Department of Computer Science & Engineering, T.I.E.I.T, Bhopal, India
[2]Principal, T.I.E.I.T, Bhopal, India
[1]tanujagoswami1993@gmail.com, [2]amitsaxena@trubainstitute.ac.in

Abstract. *ECG signal preprocessing and feature extraction are critical components in improving the accuracy and reliability of heart disease diagnosis, especially in the detection of arrhythmias and other cardiac abnormalities. Raw ECG signals are often contaminated with noise and artifacts from various sources, making it essential to apply effective preprocessing techniques to enhance signal quality. This paper explores several key preprocessing methods, such as denoising, baseline correction, and normalization. Denoising techniques help remove unwanted noise without distorting the important features of the signal, while baseline correction addresses drift and shifts in the signal baseline that can affect analysis. Normalization ensures that the signal amplitude remains consistent, making it suitable for further processing. In addition to preprocessing, feature extraction plays a crucial role in identifying distinctive characteristics of the ECG signal that are necessary for accurate classification. Techniques such as wavelet transform, time-domain analysis, and frequency-domain analysis are explored for their ability to capture both time- and frequency-related features. The study concludes that combining optimal preprocessing and feature extraction methods significantly enhances ECG-based diagnostic systems, making them a valuable tool for early detection and clinical decision support, ultimately contributing to better patient care and outcomes.*

Keywords: *ECG Signal Preprocessing, Feature Extraction, Heart Disease Diagnosis, Arrhythmia Detection.*

I. Introduction

An electrocardiogram (ECG) is simply a recording of the electrical activity generated by the heart. The heart produces the electrical activity that measures by a medical test called an ECG, which identifies the cardiac abnormality [1]. A heart produces tiny electrical impulses that spread through the heart muscle. Then, a medical practitioner interprets this data; ECG leads to find the cause of symptoms of chest pain and also leads to detect abnormal heart rhythm. An ECG signal has a total of five primary turns, counting P, Q, R, S, and T waves, plus the depolarization of the atria causes a small turn before atria contraction as the activation (depolarization) wave-front propagates from the Sino atria node through the atria [6]. The Q wave is a downward deflection after the P wave [2]. The R wave follows as an upward deflection, and the S wave is a downward deflection following the R wave. Q, R, and S waves together indicate a single event. Hence, they are usually considered to be QRS complex, as shown in Figure 1. The QRS complex is a key feature for ECG analysis, caused by currents generated during ventricular depolarization before

contraction. Although atrial depolarization occurs first, its waveform is not visible due to the higher amplitude of the QRS complex. The T wave follows the S wave, representing ventricular repolarization, and the U wave follows the T wave. Arrhythmias can be classified into two types: morphological arrhythmias, caused by a single irregular ECG signal, and rhythmic arrhythmias, caused by irregular heartbeats. [3] A challenge in arrhythmia detection is that ECG signals can vary between individuals, and similar ECG patterns may occur in different diseases, complicating heart disease diagnosis.

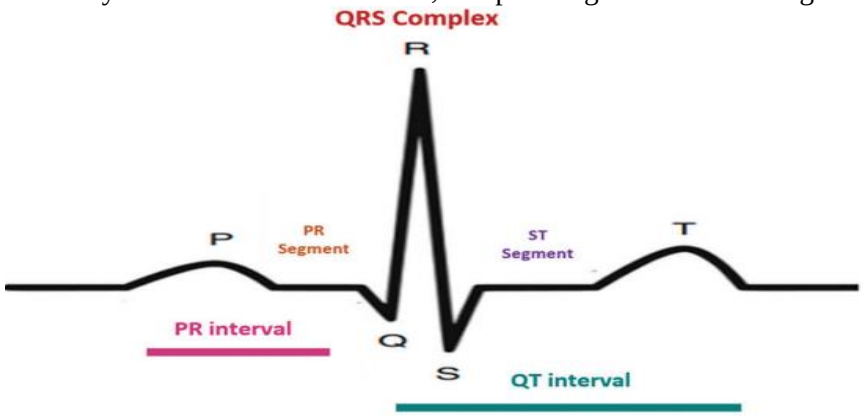

Figure 1: A typical electrocardiogram signal

II. ECG Signal Processing Stage

Generally, ECG signal processing stages are data acquisition, pre-processing [4], feature extraction and feature selection, and finally, classification.

1) Pre-processing Stage- The pre-processing stage plays a critical role in ECG signal analysis by preparing the raw data for subsequent processing and analysis. This stage focuses primarily on eliminating noise and other artifacts that can obscure the underlying signal..

2) Feature Extraction Stage- Feature extraction from ECG signals is a fundamental step in the analysis and classification process, as it transforms raw ECG data into meaningful and interpretable metrics. This step involves decomposing the ECG signal into its constituent components, such as Q-, R-, T-, and U-waves, which are critical for understanding the physiological and pathological characteristics of the heart.

3) Training and Testing Stage- During the training and testing stages of ECG signal analysis, the process focuses on selecting the most important features from the signal to achieve effective classification. Feature selection is a critical step, as the choice of features directly influences the performance and accuracy of the classification model.

4) Classification Stage- At this stage, the ECG signals are grouped or categorized based on the results produced by the classification method. This step is the culmination of the analysis process, where the model applies the learned patterns and insights from the training and testing phases to assign each ECG signal to a specific category or class.

5) Validation Stage- The final stage in the process is validation, a critical step that evaluates the performance and reliability of the proposed model. Validation aims to determine whether the model has achieved an acceptable level of accuracy using the given training dataset.

III. Literature Review

This section reviews various research efforts aimed at improving ECG classification accuracy. V. C. et al. [1] used SVM and KNN classifiers to detect arrhythmias and coronary heart disease, achieving 97.5% accuracy. Mohammad et al. [2] proposed a Random Feature Extractor (RFE) for feature selection, reaching 99.79% accuracy in arrhythmia classification. S. T. Sanamdikar et al. [3] combined SVM with kernel PCA for ECG signal classification, focusing on noise removal and feature extraction. Y. Kaya et al. [4] used feature extraction, PCA, ICA, and genetic algorithms for arrhythmia detection, employing various classifiers such as SVM and decision trees. Hemant Amhia et al. [5] proposed a method for QRS peak identification and ECG classification using fuzzy classifiers, achieving 76.67% accuracy. H. Kaur et al. [6] introduced continuous wavelet transformation (CWT) and Kalman filtering to reduce noise and identify R-peaks. M. Ramkumar et al. [7] used DWT for preprocessing, ICA for feature extraction, and MLP neural networks for classification. P. Malleswari et al. [8] employed DWT and various classifiers to achieve good precision in ECG classification on MIT-BIH data. Rizal et al. [9] focused on LSTM for atrial identification and compared it with other methods. Subramanian et al. [10] used SVM for ECG classification by extracting features such as the R-R interval and beats per minute (BPM) from segmented signals. Ansari et al. [11] explored deep learning approaches for ECG anomaly detection, highlighting the need for large datasets. Li et al. [12] proposed a heart rhythm recognition algorithm using wavelet transforms and SVM, optimized by genetic algorithms. Vedavathi et al. [13] applied the Pan-Tompkins method for peak identification and KNN for classifying five types of ECG pulses. Barhatte et al. [14] used wavelet frequency distribution alongside SVM for ECG classification, focusing on QRS complex features. Lastly, Kaya et al. [15] combined feature extraction, dimensionality reduction, and machine learning techniques to detect arrhythmias effectively. Overall, SVM and KNN are widely used for ECG classification, and ongoing research continues to improve classification accuracy.

IV. P-QRS-T Complex Features

Electrocardiography (ECG) plays a crucial role in diagnosing heart diseases by interpreting the electrical activity of the heart. Feature extraction techniques are used to analyze and classify ECG signals [5], which can be grouped into five categories: QRS, statistical, morphological, wavelet, and other features. The ECG signal consists of five major deflections: P, Q, R, S, and T waves, along with a minor U wave as shown in figure 2. The P wave represents atrial depolarization, while the QRS complex (Q, R, and S waves) corresponds to ventricular depolarization. The T wave indicates ventricular repolarization, and the U wave follows the T wave. The QRS complex is particularly important for ECG analysis, with features like R wave duration, amplitude, and QRS wave area being commonly used for classification [6]. These features help in the effective diagnosis of heart conditions by providing distinct information about the heart's electrical activity.

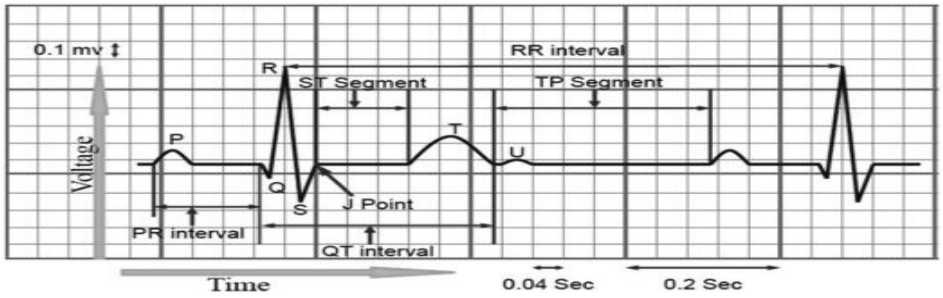

Figure 2: Standard fiducially points in the ECG (P, Q, R, S, T, and U) together with clinical feature

V. Various ECG Classification Methods

ECG classification methods aim to identify and analyze patterns in ECG signals for diagnosing heart conditions. These methods leverage advanced algorithms and techniques to improve accuracy and efficiency in cardiac diagnosis. The primary methods include:

5.1 Artificial Neural Networks- Artificial Neural Networks (ANNs) are mathematical models inspired by biological neural networks, used to simulate human cognition. They consist of processing elements (neurons) connected in a specific pattern (neural architecture), with learning algorithms to adjust connection weights and activation functions to produce outputs. The Multilayer Perceptron (MLP) is a type of supervised ANN, where the network learns by mapping input data to output based on historical data.

5.2 Convolutional Neural Network (CNN)- CNN (Convolutional Neural Network) is a feedforward network where information flows from input to output in one direction. Inspired by the brain's visual cortex, CNNs consist of convolutional and pooling layers that process data, typically in the form of grayscale or RGB images. After the input layer, CNNs may include several convolutional or pooling layers, with activation functions. For classification tasks, one or more fully connected (FC) layers are used, and the final layer produces the output.

5.3 Recurrent Neural Network (RNNs) - Recurrent Neural Networks is a class of neural networks that has directed connections between units of an individual layer, which will exhibit the unit's temporal behavior. It means that the state of the current instance will affect the state of the next instance [7]. Function softmax provides a normalized probability distribution over the possible classes, σ is the logistic sigmoid function, and W is a weight matrix. By using h as the input to another RNN, RNNs can be stacked, creating deeper architectures. RNN is an extension of an Artificial Neural Network (ANN) whose weights are shared across time. RNN is the most proper learning model for learning sequential input data and the time-series data classification where the feedback and the present value is fed again into the network and the output contains the adding of values in the memory.

5.4 Long Short-Term Memory (LSTM)- LSTM is a specific type of traditional RNN designed for temporal sequences and the long-range dependencies. A memory block places information and updates them across time-steps based on the input and output gates. The gates control the input and output flow of information to a memory cell. LSTM is a specialized variant of the traditional Recurrent Neural Network (RNN) designed to address the challenge of learning long-range dependencies in sequential data.

5.5 K nearest Neighbor (KNN)- kNN classifies feature vectors according to the labels of the closest training samples in the feature space. For an unknown feature vector, the distances from this vector to all vectors in the training set are calculated using a distance measure such as the Euclidean distance. Then, an unknown feature vector is assigned to the class in which the closest k samples mostly belong to. Thus, a kind of majority voting approach is applied. The value of k is a positive integer and is known to be a strongly influencing factor for the accuracy of the classification.

5.6 Support Vector Machine (SVM)- SVM is a widely used tool for solving binary classification problems because of its outstanding generalization performance. The main idea of the SVM is to find a maximum margin between the training data and the decision boundary [10]. Support vectors, which are the training samples that are closest to the decision boundary, are used for margin maximization. The SVM can be regarded as either a linear or nonlinear classifier according to the type of its kernel function

5.7. Decision Tree (DT)- DT learning aims to map observations about an item to a conclusion. This conclusion can be either a possible target class label or a target value. According to the difference in this conclusion, DT structures are called classification or regression trees. In addition to common decision tree approaches, there are some more specific decision tree structures that are used frequently for ECG classification. The Random Forest Tree is a type of ensemble classifier that uses many decision trees [11]. In this approach, multiple decision trees are trained with subsets of training data.

VI. Databases Available for Analysis of ECG Signals

Various publicly available ECG databases are used for analysis and evaluation:

- MIT-BIH Arrhythmia Database- Widely used for ECG analysis (details in Section 8.1).
- Physionet PTB Diagnostic ECG Database- Contains 549 records from 290 individuals (52 healthy, 148 sick) with 15 simultaneous signals at 1000 samples/second.
- QT Database- Includes ECG recordings from several MIT-BIH databases, plus others from the European Society of Cardiology and sudden death patients.
- Apnea-ECG Database- 70 recordings with ECG and apnea annotations at 100 Hz sampling rate.
- Non-Invasive Fetal ECG Database- 55 recordings from one subject during pregnancy for testing signal separation algorithms.
- Creighton University Ventricular Tachyarrhythmia Database- 35 single-channel recordings of ventricular tachycardia, flutter, and fibrillation at 250 Hz.
- AHA Database- 80 two-channel ECG recordings at 250 Hz.
- Fantasia Database- Records from 40 subjects (young and elderly) at 250 Hz.
- BIDMC Congestive Heart Failure Database- 20 hours of ECG recordings from 15 subjects at 250 Hz.
- European ST-T Database- 90 annotated recordings with two signals, sampled at 250 Hz.
- Long-Term ST Database- 86 long ECG recordings from 80 subjects, sampled at 250 Hz.
- INCART Database- 75 recordings with 12 leads, sampled at 257 Hz from 32 holter contacts.

These databases cover a wide range of ECG applications, from arrhythmia classification to heart failure monitoring.

VII. Application Fields for ECG Analysis & Classification

This paper reviews various fields of ECG signal analysis, including disease classification, heartbeat type detection, biometric identification, and emotion recognition.

- Disease Classification- ECG signals are crucial for diagnosing heart diseases, particularly arrhythmia, and enabling early detection to prevent severe health issues.
- Heartbeat Type Detection- Involves differentiating various ECG beats, aiding in accurate analysis and diagnosis of heart conditions.
- Biometric Identification- Uses ECG as a unique biometric identifier for individual recognition, enhancing security.
- Emotion Recognition- ECG signals can also be used to detect emotions, contributing to affective computing and human-machine interaction.
- Clinical Diagnostics and Monitoring- Vital for diagnosing and monitoring heart conditions like arrhythmias and myocardial infarctions.
- Telemedicine and Remote Monitoring- Portable ECG devices enable real-time monitoring and alerts in remote healthcare settings.
- Personalized Medicine- Tailoring treatment plans based on ECG analysis to improve patient outcomes.
- Sports Medicine and Fitness- Monitoring athletes' heart health and detecting cardiac issues early.
- Emergency Care- Quick ECG analysis helps paramedics make critical decisions before hospital arrival.
- Research and Drug Development- Studies the cardiac effects of drugs during clinical trials.
- Education- Assists in training healthcare professionals with simulated ECG data.
- Veterinary Cardiology- Used to monitor animal heart health.
- Public Health- Helps in epidemiological studies and prevention strategies for heart diseases.

VIII. Conclusions and Future Scopes

ECG signal preprocessing and feature extraction are critical steps in enhancing the accuracy of heart disease diagnosis, especially in the context of detecting cardiac abnormalities like arrhythmias. Our analysis indicates that effective preprocessing techniques, such as noise reduction, baseline correction, and signal normalization, are pivotal in improving the quality of raw ECG signals. These methods help mitigate the effects of interference and artifacts, ensuring that the signals are more reliable and suitable for further analysis. Once the signal is preprocessed, the next crucial step is feature extraction, which involves identifying key characteristics of the ECG signal that can be used for classification. Techniques such as wavelet transforms, time-domain analysis, and frequency-domain methods are employed to extract features that capture the essential patterns of heart activity. These extracted features provide the input necessary for machine learning classifiers to distinguish between normal and abnormal heart rhythms with high accuracy. The combination of preprocessing and feature extraction techniques significantly enhances the performance of classification models, particularly in arrhythmia detection. Future research should aim at automating the preprocessing steps, exploring hybrid approaches for feature extraction, and focusing on real-time classification to improve clinical applicability. By addressing these aspects, heart disease diagnosis could become more efficient, scalable, and accessible, leading to better patient outcomes and more effective early detection systems.

References:-

[1] V., C., R., M.., S., A., M., M. "ECG Signal Feature Extraction and SVM Classifier Based Cardiac Arrhythmia Detection." undefined (2023). doi: 10.1109/ICEEICT56924.2023.10157789.
[2] Mohammad, Mominur, Rahman.,Ashhadul, Islam., Skander, Charni., Halima, Bensmail., Thomas, Hilbel., Samir, Brahim, Belhaouari. (2023). 3. A Novel Feature Extraction Technique for ECG Arrhythmia Classification Using ML. doi: 10.1109/dasc/picom/cbdcom/cy59711.2023.10360505
[3] S. T.Sanamdikar, N. M. Karajanagi, A. V. Kulkarni and S. D. Kamble, "KPCA and SVR-based Cardiac Arrhythmia Classification on Electrocardiography Waves," 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2023, pp. 315-321, doi: 10.1109/ICSSIT55814.2023.10061047.
[4] Y. Kaya, H. Pehlivan, and M. E. Tenekeci, "Effective ECG beat classification using higher order statistic features and genetic feature selection," J. Biomed. Res., Vol. 28, pp. 7594–603, Aug. 2017.
[5] HemantAmhia and A. K. Wadhwani "Designing an Optimum andReduced Order Filter for Efficient ECG QRS Peak Detection and Classification of Arrhythmia Data", Hindawi Journal of Healthcare Engineering Volume 2021, Article ID 6542290, 17 pages https://doi.org/10.1155/2021/6542290
[6] H. Kaur and R. Rajni, "On the detection of cardiac arrhythmia with principal component analysis," J. Wireless Pers. Commun., Vol. 97, pp. 5495–509, Dec. 2017.
[7] M.Ramkumar , C.GaneshBabu , Vinoth Kumar K , Hepsiba D , A. Manjunathan .R.Sarath Kumar "ECG Cardiac arrhythmias Classification using DWT, ICA and MLP Neural Networks" International Conference on Robotics and Artificial Intelligence (RoAI) 2020 Journal of Physics: Conference Series 1831 (2021) 012015 IOP Publishing doi:10.1088/1742-6596/1831/1/012015
[8] P., Malleswari.,Srinivas, Padala., Matta, venkata, Pullarao., M., R., Sankar., Y., Mounika. (2023). 8. Deep Learning Frameworks for Cardiovascular Arrhythmia Classification. International Journal on Recent and Innovation Trends in Computing and Communication, doi: 10.17762/ijritcc.v11i11s.8067
[9] Rizal, Arifin.,Satria, Mandala. (2023). 9. Study of Arrhythmia Classification Algorithms on Electrocardiogram Using Deep Learning. Sinkron : jurnaldanpenelitianteknikinformatika, doi: 10.33395/sinkron.v8i3.12687
[10] Subramanian, K., &Prakash, N. K. (2020). Machine Learning based Cardiac Arrhythmia detection from ECG signal. 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). doi:10.1109/icssit48917.2020.9214077
[11] Ansari Y, Mourad O, Qaraqe K, Serpedin E. Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017-2023. Front Physiol. 2023 Sep 15;14:1246746. doi: 10.3389/fphys.2023.1246746. PMID: 37791347; PMCID: PMC10542398.
[12] Li H, Yuan D, Wang Y, Cui D, Cao L. Arrhythmia Classification Based on Multi-Domain Feature Extraction for an ECG Recognition System. Sensors. 2016; 16(10):1744. https://doi.org/10.3390/s16101744
[13] VedavathiGauribidanurRangappa, SahaniVenkataAppalaVaraprasad Prasad .AlokAgarwal, "Classification of Cardiac Arrhythmia stages using Hybrid Features Extraction with K-Nearest Neighbour classifier of ECG Signals", International Journal of Intelligent Engineering and Systems, Vol.11, No.6, 2018 DOI: 10.22266/ijies2018.1231.03

[14] Barhatte, A. S., Ghongade, R., &Thakare, A. S. (2015). QRS complex detection and arrhythmia classification using SVM. 2015 Communication, Control and Intelligent Systems (CCIS). doi:10.1109/ccintels.2015.7437915
[15] Kaya, Yasin&Pehlivan, Hüseyin&Tenekeci, Mehmet. (2017). Effective ECG beat classification using higher order statistic features and genetic feature selection. Biomedical Research. 28. 7594-7603.