Arrhythmia Data Classification Using QRS Detection and Support Vector Machines

Deepak Rathore[1], Nitin Choudhary[2]
[1]Sanjeev Agrawal Global Educational (SAGE) University, Bhopal, India, 462026
[2] Sanjeev Agrawal Global Educational (SAGE) University, Bhopal, India, 462026
[1]deepak9981711@gmail.com, [2]nitin.c@sageuniversity.edu.in

Abstract. *Classification of arrhythmia patterns is an open field of research. One of the most widely used biological signals crucial for the detection of heart conditions is the electrocardiogram (ECG). Within the processing of ECG signals, one of the most critical aspects is the interpretation and acquisition of the QRS complex. Both the diagnosis of cardiac rhythm abnormalities and the assessment of heart rate variability depend heavily on the R wave (HRV). In this paper, the modified Pan-Tompkins method with an FIR filter is used for baseline wandering and noise reduction. Temporal statistical features are extracted using the QRS peak detection algorithm. The QRS interval, RR peak interval, QRS deviation, kurtosis, and skewness of the QRS are utilized as the features extracted from the ECG under test. This paper compares the classification accuracy of the standard QRS interval thresholding-based binary approach and the proposed support vector machine (SVM)-based classification method. The proposed optimized K-nearest neighbors (KNN) algorithm outperforms tree and thresholding-based classifiers.*

Keywords: *ECG, Arrhythmia Classification, QRS Peak Detection, R peak, FIR Filter, Pantompkins, SVM*

INTRODUCTION

The non-linear biological signal known as an electrocardiogram (ECG) mimics the functioning of the heart. One important aspect of heart health monitoring is the classification of arrhythmia data and QRS peak identification. Recent research aims to use support vector machine (SVM) classifiers to address this issue. Numerous approaches for efficient feature extraction and classification are highlighted in recent studies, emphasizing the significance of preprocessing ECG signals. Thus, it is crucial to analyze the ECG signal in clinical cardiac tests used to check for various heart abnormalities. For the past thirty years, digital signal processing has been extensively utilized to evaluate ECG signals. Accurate feature extraction from ECG signals requires effective preprocessing. V., C., et al. [1], 2023, stated that distinguishing between normal and arrhythmic beats significantly improves detection accuracy through R-peak identification. Rahman et al. (2023) [2] introduced a novel method for feature extraction that employs random selection to generate relevant features from available data, achieving an impressive accuracy of 99.79% in arrhythmia classification. S. T. Sanamdikar et al. [3] recently studied how utilizing preprocessed ECG data demonstrates that SVM classifiers outperform classical classifiers, with classification accuracies reaching 97.5%. When kernel-PCA and SVM are combined, feature extraction is enhanced, improving the accuracy of classification for arrhythmia recognition. A novel artificial intelligence-based method for QRS peak identification and ECG data classification has been proposed by

Hemant Amhia et al. [5]. A lower-order IIR filter design is suggested for the low-pass smoothing of ECG signal data. To create the reduced-order filter, the filter coefficient is optimized using min-max optimization.

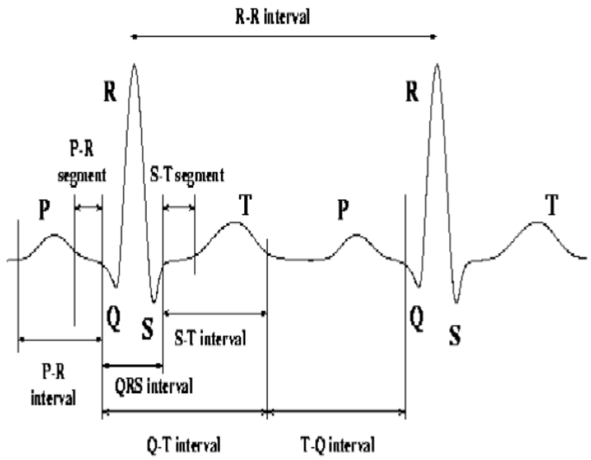


Fig. 1: Two cycle Normal ECG waveform.

The majority of heart pathology can be understood by examining the ECG signal. Heart rate and ECG readings serve as the evaluation metrics for a healthy heart. Cardiac arrhythmia refers to any non-linearity found in the patient's recorded ECG signal. Figure 1 depicts a normal ECG rhythm. It illustrates the key elements of an ECG wave, which include P, Q, R, S, and T. Specific intervals, such as the P-R, S-T, and Q-T intervals, are shown in the standard scalar electrocardiographic waveform. Thus, the ultimate aim of this paper is to use QRS peak detection for feature extraction to be applied in ECG classification.

II ECG DATASET

A filtered ECG signal obtained from the MIT-BIH arrhythmia database is used by the Hilbert transformation to perform the R-peak detection task, as noted by Hemant et al. [5]. Compared to previously reported results, the suggested method offers a higher percentage of precision, sensitivity, and productivity. To conduct this study in the MATLAB environment, a total of thirty MIT-BIH ECG arrhythmia records have been used, as shown in Figure 2. Each record lasts for thirty minutes and fifty-five seconds. The length has been rounded to the nearest second, so it cannot be precisely thirty minutes and fifty-six seconds due to cumulative rounding errors. Heart rates are calculated in beats per minute across three R-R intervals [5].

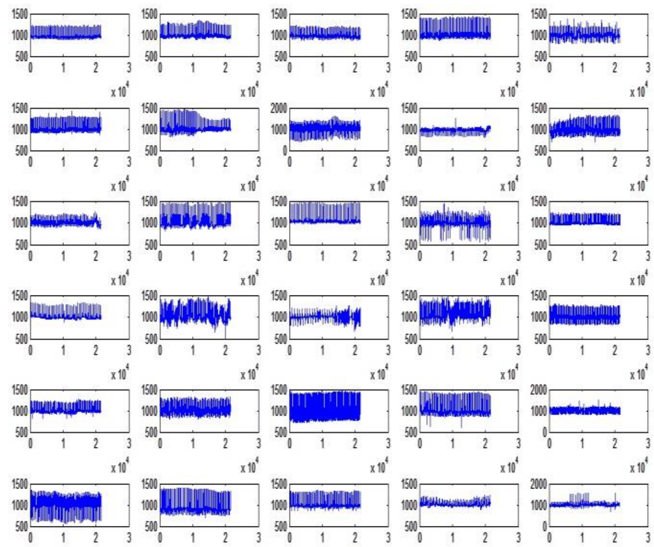


Fig. 2: Database used in the research.

III LITERATURE REVIEW

Various research efforts have been dedicated to enhancing the classification accuracy of ECG patterns for heart disease diagnosis, particularly in detecting arrhythmias. This section reviews significant contributions to this field. V., C. et al. [1] focused on detecting cardiac arrhythmias and assessing the risk of coronary heart disease (CHD) by emphasizing the preprocessing of ECG signals. They applied feature extraction methods and used a support vector machine (SVM) classifier to identify cardiac arrhythmias. The processed ECG signals were classified into arrhythmic and normal categories using the SVM technique. Additionally, k-nearest neighbor (KNN) was employed, achieving a classification accuracy of 97.5%. Mohammad, Mominur, and Rahman et al. [2] introduced a novel feature selection approach named Random Feature Extractor (RFE) to generate new features from existing ECG characteristics. They tested their method using the MIT-BIH dataset and achieved an impressive classification accuracy of 99.79%. S. T. Sanamdikar et al. [3] combined SVM classifiers with kernel principal component analysis (PCA) for ECG signal identification. The three-stage process involved noise removal using a low-pass filter, feature extraction via higher-order averages and kernel PCA, and ECG wave classification using the SVM algorithm. Their results, validated using the MIT-BIH dataset, highlighted the effectiveness of this method. Y. Kaya et al. [4] developed a three-step method combining feature extraction, dimensionality reduction, and machine-learning classification. They extracted statistical features, reduced dimensions using PCA, ICA, and genetic algorithms (GAs), and applied decision trees, SVM, neural networks, and KNN for heartbeat classification. Hemant Amhia et al. [5] proposed an artificial intelligence-based method for QRS peak identification and ECG classification. They used a lower-order IIR filter for noise reduction and introduced a statistical QRS thresholding method based on heart rate variability (HRV), achieving a classification accuracy of 76.667% with a fuzzy classifier. H. Kaur et al. [6] combined continuous wavelet transformation (CWT) and extended Kalman filtering to enhance signal clarity and reduce noise. They proposed a method using principal component analysis for

R-peak and QRS complex identification on de-noised signals. M. Ramkumar et al. [7] used discrete wavelet transform (DWT) for ECG signal preprocessing, ICA for feature extraction, and multi-layer perceptron (MLP) neural networks for classification, demonstrating effective arrhythmia organization. P. Malleswari et al. [8] used DWT for signal preprocessing and divided the signal using classifiers such as decision trees, logistic regression, and SVM variants. The approach achieved high classification precision when tested on the MIT-BIH database. Rizal, Arifin et al. [9] explored Long Short-Term Memory (LSTM) techniques for arrhythmia classification, focusing on optimizing correctness, specificity, and sensitivity. Subramanian, K. et al. [10] applied SVM techniques to classify arrhythmias, segmenting ECG signals into shorter parts and extracting features like R-R intervals and BPM for classification. Ansari Y et al. [11] highlighted deep learning models for ECG anomaly detection, emphasizing large datasets and model refinement. Li H et al. [12] proposed a multi-domain feature discovery algorithm using wavelet transforms and kernel ICA, optimized using genetic algorithms. Vedavathi Gauribidanur Rangappa et al. [13] used the Pan-Tompkins method for QRS peak detection and KNN for classification, achieving high accuracy with the MIT-BIH dataset. Barhatte, A. S. et al. [14] employed wavelet frequency distribution and SVM for feature extraction and heartbeat classification. In conclusion, SVM and KNN are widely used for ECG classification, but there remains significant scope for further improvements in accuracy and real-time application.

IV VARIOUS ECG CLASSIFICATION METHODS

This paper proposes the application of SVM-based machine learning classification for predicting normal or abnormal ECG patterns. An efficient FIR filter-based Pan-Tompkins method is introduced for preprocessing and baseline wandering correction. The FIR filter is used for low-pass smoothing of the ECG signal data. A QRS peak detection method is employed for the temporal statistical feature extraction from the ECG data. Additionally, a Bayesian optimization technique is proposed to improve classification accuracy while minimizing the order of the FIR filter design. The block diagram of the proposed ECG classification method is shown in Figure 3.

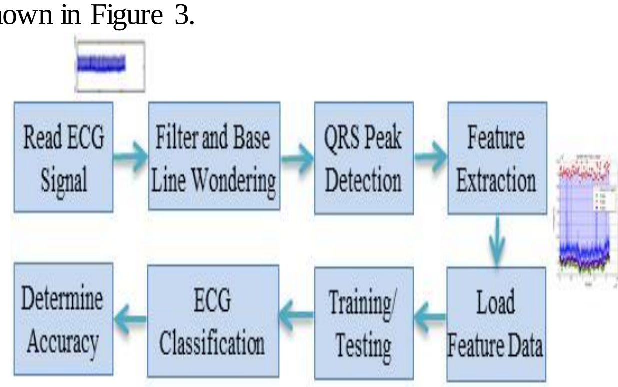


Fig. 3: Sequential processes of ECG peak detection and classification.

To remove power line noise, the differential of the filtered signal will be utilized to eliminate muscle artifacts, while baseline noise and motion artifacts will be addressed using an FIR filter with a Kaiser-Bessel window. The processed ECG signal is generated using the initial differential of the ECG signal.

The SVM-based classifiers are compared, as illustrated in the flowchart of the proposed QRS peak detection and ECG pattern classification shown in Figure 4.

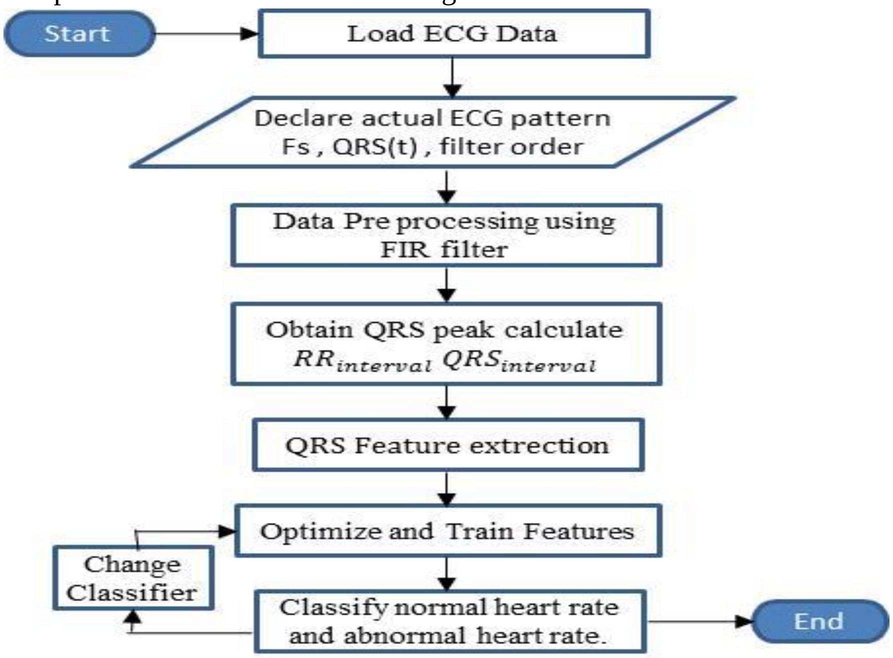


Fig. 4: Flow Chart of the proposed QRS Peak Detection and Disease Detection Algorithm.

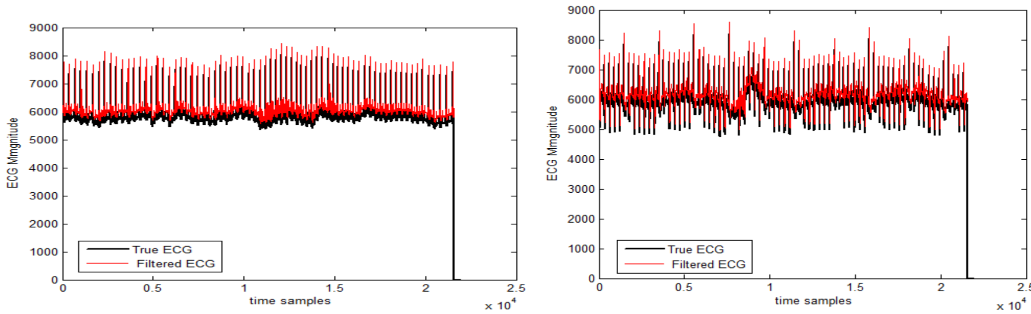
The proposed method compares the QRS interval thresholding method from [5] with the proposed SVM-based classifiers in terms of accuracy.

V RESULTS AND DISCUSSION

This research suggests the application of SVM-based machine learning classification for the identification of normal or abnormal ECG patterns. A productive Pan-Tompkins technique based on FIR filters is proposed for preprocessing and baseline wandering correction. The ECG signal data is low-pass smoothed using the FIR filter. Temporal statistical information is extracted from the ECG data using a QRS peak detection algorithm. The ECG data is recorded at a rate of 360 samples per second. Details and descriptions of the MIT-BIH database are provided.

5.1 FIR Filtering Results

Initially, Figure 5 provides a visual depiction of the outcome comparison for the proposed FIR filter design. The results of the suggested ECG signal preprocessing are plotted for the entire length of the ECG samples and are presented in the figure. Figure 5(a) shows the FIR filtered results of the 100m.mat ECG data, while Figure 5(b) displays the results for the 105m.mat ECG data. It is observed that the magnitude is enhanced.



(a) for 100m.mat signal (b) for 105m.mat signal

Fig. 5: Reults of FIR Filtering of ECG with magnitude enhancment.

5.2 Results of QRS Peak Detection

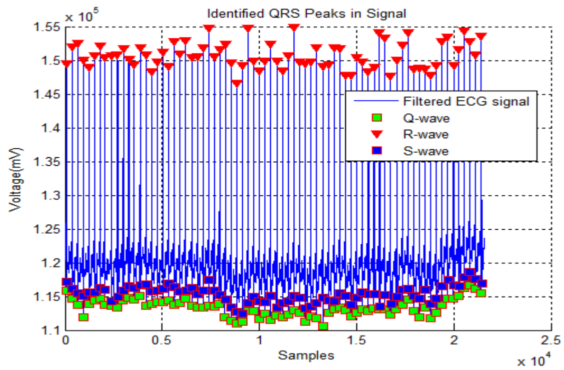
The findings of QRS peak detection for all four input ECG waves in Figure 6 are presented in the following subsection. It is evident that the proposed method effectively detects the Q, R, and S peaks for each of the four examples. The ECG signals from channels 100, 101, 104, and 105 are considered for QRS identification, and the results are represented in Figures 6(a)–6(d). These temporal samples are plotted on the x-axis, while the voltage magnitude of the ECG signal is plotted on the y-axis. It can be observed that the proposed approach ensures efficient recognition of ECG peaks, assisting in the accurate detection of Q, R, and S peaks.

5.3 Features detection and Parametric Evaluation

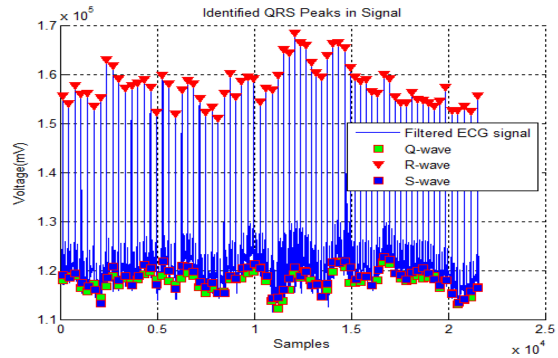
The main goal is to demonstrate how the classification efficiency of the proposed method has improved. To determine the efficiency parameter, accuracy is calculated. When the algorithm fails to detect a real beat, a false negative (Fn) is produced. A false positive (Fp) occurs when a false beat is detected. The accurately identified beat using the proposed approach is termed the true positive (Tp), while true negatives (Tn) represent accurately undetected beats. Classification performance is assessed in terms of accuracy.

Performance accuracy is the most intuitive metric, as it is a simple ratio of correctly predicted observations to all observations made.

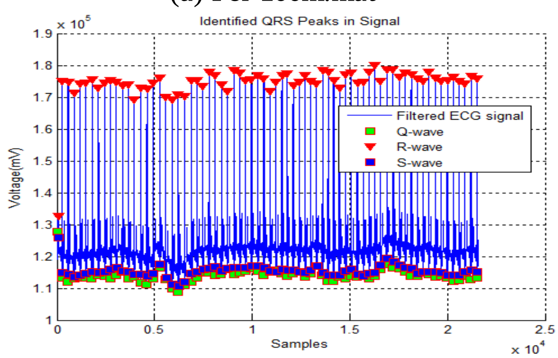
Hemant [5] has used this formula to calculate the accuracy of classification. The remainder of the paper presents the results of ECG pattern classification and feature extraction.



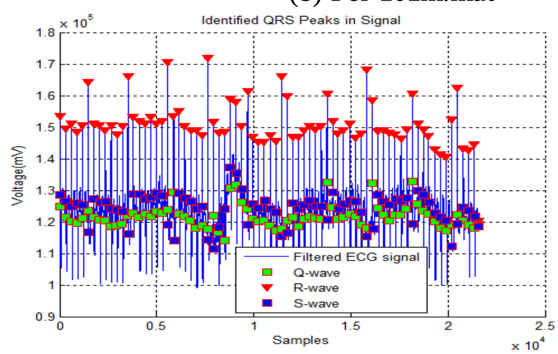
(a) For 100m.mat



(b) For 101m.mat



(c) for 104m.mat

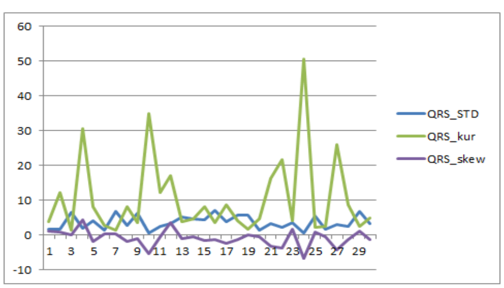


(d) for 105m.mat

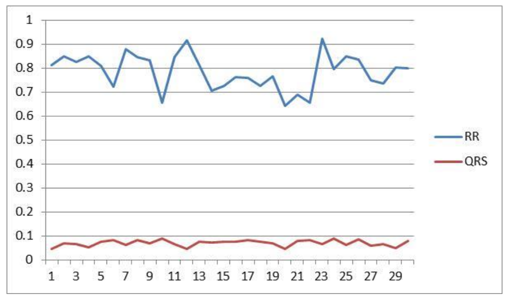
Fig. 6: Results of the QRS peak detection for four ECG signals with the proposed FIR filter method.

5.3 Feature extraction

This paper have used the five features as R-R time interval as $RR_{interval}$, QRS complexes interval $QRS_{interval}$, the deviation of QRS regions QRS_{STD}, Kurtosis value of QRS regions QRS_{krts}, and Skewness of QRS complexes QRS_{Skew}.



(a) QRS complex features

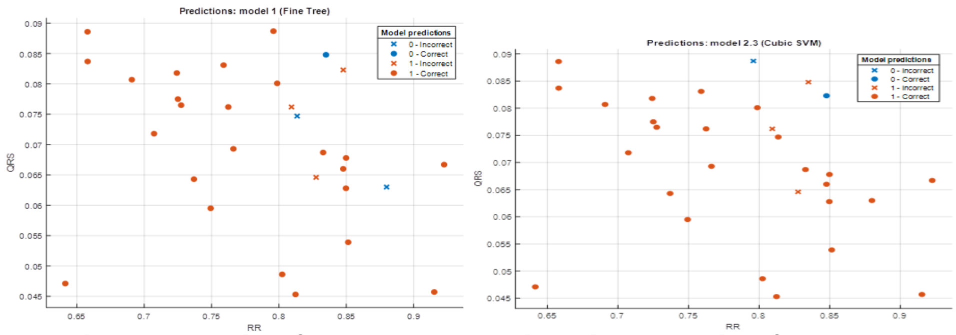


(b) RR and QRS Intervals

Fig. 7: Results of the Feature extraction for the CG Arrhythmia data.

5.4 Results of Classification

The classification results of the fine tree and cubic SVM methods are shown in the scatter plots in Figure 8.



(a) with Fine Tree classifier (A=83.3%) (b) Cubic SVM classifier (A=86.7%)

Fig. 8: Results of the Scatter plots of classifiers.

5.4.1 Results of Optimizable KNN

In this paper, to improve classification accuracy, it is proposed to use the Bayesian optimization-based KNN classification. The estimation error is less than 0.1, as shown in the results in Figure 9.

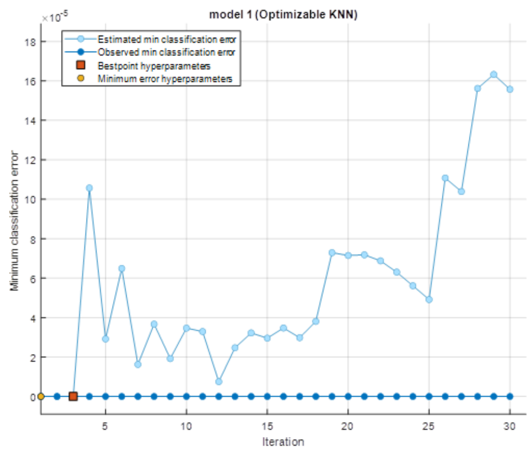


Fig. 9: The estimation error plotted for 30 iteration of Bayesian Optimization for KNN.

TABLE I. STATE OF ART RESULT COMPARISON

Classification rule	Percentage Accuracy
Hemant et al [5]	76.667 %
V., C., et al. [1],	97.7 %
Fine Tree Classifier	83.3 %
Cubic SVM Classifier	86.7 %
Proposed Optimized KNN	100 %

The optimization method employs 10 neighbors with Jaccard distance matrices over 30 iterations. The main advantage of using Bayesian optimization-based KNN classification is its ability to balance exploitation of well-known parameters with exploration of new parameter combinations to determine the ideal KNN hyperparameters. This approach leads to better performance and quicker convergence. Additionally, because Bayesian optimization incorporates uncertainty into the process, it can effectively manage noisy ECG data, increasing its resilience to variations in data quality.

VI CONCLUSION AND FUTURE SCOPE

A critical component is the comprehension and acquisition of the QRS complex. The R wave is crucial for both the identification of cardiac rhythm disorders and the evaluation of heart rate variability (HRV). In this study, baseline wandering and noise reduction are achieved by using the modified Pan-Tompkins approach with an FIR filter. The QRS peak detection algorithm is used to extract temporal statistical information. The features extracted from the ECG being tested include the RR peak interval, QRS deviation, skewness of QRS, kurtosis, and QRS interval. The use of an extended feature set is helpful in improving classification accuracy. The proposed SVM-based classification is compared with the traditional QRS interval thresholding approach in terms of accuracy. It is concluded that the suggested optimized SVM performs better with 100% accuracy over data without cross-validation, compared to tree-based classifiers (83.6% accuracy) and HRV thresholding (76.7% accuracy). Designing scalable Bayesian optimization techniques to manage large-scale ECG datasets will be a challenge in the future. It is difficult to incorporate multiple objectives, such as accuracy and processing cost, within the Bayesian optimization framework.

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.