Journal of Theoretical and Applied Information Technology
15th July 2024. Vol.102. No. 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A MACHINE LEARNING-BASED OPTIMIZED FRAMEWORK FOR DETECTION OF ARRHYTHMIA FROM ECG DATA

MS.K.SHILPA[1] DR.T.ADILAKSHMI[2]

[1]Research Scholar, Department of CSE, UCE, Osmania University,Hyderabad,India

[2]Professor and Head, Department of CSE, Vasavi College of Engineering, Osmania University, Hyderabad,India

Email: shilpamtech555@gmail.com, t_adilakshmi@staff.vce.ac.in

ABSTRACT

Heart diseases are causing health issues for people across the globe due to various reasons, including lifestyle changes. With the emergence of artificial intelligence (AI), it is possible to have learning-based approaches for the automatic detection of several types of heart diseases. Many existing approaches followed data-driven techniques for the diagnosis of heart diseases. Some methods focused on ECG data, which has the potential to support the detection of different kinds of diseases, particularly arrhythmia. The literature shows that machine learning models result in deteriorated performance unless specific optimizations support them. Motivated by this fact, we proposed a machine-learning framework that exploits many classification models for detecting arrhythmia and classification. The proposed framework is subjected to multiple optimizations in terms of preprocessing, feature engineering, and hyperparameter tuning. To develop an optimized machine learning approach, we proposed two algorithms known as Feature Selection and Hyperparameter Optimization (FSHO) and Learning-based Arrhythmia Detection and Classification (LbADC). We used our empirical study's benchmark dataset, known as the MIT-BIH Arrhythmia dataset. The experimental results reveal that the proposed optimizations and machine learning framework could improve arrhythmia diagnosis and classification performance. The proposed optimizations of our framework achieved 96.8% accuracy in multi-class classification.

Keywords : *Healthcare, Machine Learning, Feature Engineering, Heart Disease Prediction, Arrhythmia Diagnosis*

1. INTRODUCTION

According to the World Health Organization, there has been an increasing incidence of heart disease across the globe. As per the third sustainable development goal of the United Nations, health is given the highest importance. There are different kinds of heart diseases that people suffer [1]. Heart stroke and sudden cardiac arrest are some of the severe life-threatening health incidents related to the heart. However, such incidents do not occur all of a sudden. They occur due to prolonged negligence of specific symptoms linked to heartbeat and health. There are many ways in which hot health can be measured. One such important measure or modality is known as ECG, which doctors widely use to know possible heart issues of a given patient. Arrhythmia is one of the hot diseases which can be diagnosed with the help of ECG signals. A person suffering from such ailment may have symptoms like irregular heartbeat beat, reduced ability, chest pain, dizziness, and shortness of breath. However, these symptoms are not apparent in some patients. If the patient is not aware of these symptoms or ignores them, ultimately, it may result in auto failure, stroke, or even death of the person. Therefore, asymptomatic cases present challenges in health care [2].

With a simple ECG, it is possible that a doctor can diagnose heart-related issues early. With the emergence of artificial intelligence, including machine learning and deep learning models, it is now possible to have learning-based approaches to detect heart diseases. There are many literature findings in this regard. An ECG is necessary to identify potentially fatal arrhythmias. Review includes feature extraction, signal decomposition, and machine learning for precise detection. Real-time applications of deep learning appear promising [3]. Detecting cardiac arrhythmias is essential for patient supervision. Usage of random forest could achieve 85.58% accuracy using feature selection and machine learning [6]. The role of ECG in the categorization of cardiac irregularities is significant. With 162 records, it employs SVM and achieves high accuracy [9]. A unique method for feature extraction from ECG data that uses SCM to identify cardiac arrhythmias is explored, and 2% more accuracy than prior techniques is achieved, with HOS being preferred for categorizing arrhythmias [15]. The literature shows that machine learning models result in deteriorated performance unless specific optimizations support them. The research problem and the proposed research design are linked to a machine learning-based optimized framework for detecting arrhythmia from ECG data. Our contributions to this paper are as follows.

1. We proposed a machine learning framework that exploits many classification models to detect and classify arrhythmia.
2. We proposed two algorithms: feature Selection and Hyperparameter Optimization (FSHO) and learning-based Arrhythmia Detection and Classification (LbADC).
3. We built an application to evaluate our framework using the MIT-BIH Arrhythmia dataset benchmark dataset.

The remainder of the paper is structured as follows: Section 2 reviews a literature review on recent machine learning models in arrhythmia diagnosis. Section 3 presents the proposed methodology and required procedures and algorithms. Section 4 presents the results of multiclass classification with different optimizations exhibited by many machine learning models. Section 5 includes our work and the scope for future research.

2. RELATED WORK

Kumari *et al*. [1] observed that an ECG is required to identify arrhythmias. Research on MIT-BIH and BIDMC data employs SVM and DWT. Attains a classification accuracy of 95.92 percent. Subramanian and Prakash [2] used to assess arrhythmias; noise is filtered, and signals are segmented for precise feature extraction (R-R Interval, BPM, P wave). Robust with a variety of training data, SVM reaches 91% accuracy. Sahoo *et al*. [3] found that an ECG is necessary to identify potentially fatal arrhythmias. Review includes feature extraction, signal decomposition, and machine learning for precise detection. Real-time applications of deep learning appear promising. Pham *et al*. [4] used ML and CDSS to achievehigh accuracy in rhythm classification. Cardiac death and arrhythmias such as Afb, Af, and Vfb suggest stroke. Ketu and Mishra [5] improved accuracy using SMOTE; the study examines ML methods for HD identification. Heart disease's high death rate demands clever remedies.

Singh [6] detected cardiac arrhythmias essential for patient supervision. This study use random forest to achieve 85.58% accuracy using feature selection and machine learning. Devi and Kalaivani [7] explored an Internet of Things-based ECG monitoring system essential for prompt arrhythmias identification, supporting remote patient monitoring and precise categorization. Yadav *et al*. [8] used MIMIC-III data, an objective framework was developed for classifying cardiac arrhythmias, and Random Forest achieved an AUC of 0.9787. Kumari *et al*. [9] described the role of ECG in categorizing cardiac irregularities. With 162 records, it employs SVM and achieves enormous accuracy. Jahan *et al*. [10] analyzed the accuracy of AF diagnosis using machine learning methods such as SVM, RF, and XGBoost.AF detection is essential to the world's medical treatment.

Sangaiah*et al*. [11] précised cardiac arrhythmia diagnosis is essential to averting unexpected fatalities. ECG signal improvement, feature extraction, and HMM classification are all included in the framework. Chang *et al*. [12] tested 12-lead ECG data; the deep-learning LSTM model correctly identified 12 cardiac rhythms, facilitating a faster diagnosis. Hus and Cheng [13], with high accuracy in arrhythmia classification, WBSP for ECG data is proposed. Duration and QRS similarity are two important retrieved characteristics. Tithi *et al*. [14] compared six techniques emphasizing machine learning for ECG anomaly detection. High accuracy is achieved by dividing the data for testing and training. Marinho *et al*. [15] steed out a unique method for feature extraction from ECG data that uses SCM to identify cardiac arrhythmias, leading to 2% more accuracy than prior techniques, with HOS being preferred for categorizing arrhythmias. Future work on ensemble classifiers and preprocessing methods will concentrate on enhancing system performance.

Afadar*et al*. [16] created a high-performing algorithm for classifying heart arrhythmias based on ECG data. Random Forest produced the best results when SVM, Naïve Bayes, and Random Forest were applied to a professional dataset. Further research will focus on integrating image-based categorization and expanding to more types of arrhythmias. Rao and Martis [17] explored, with an accuracy of 85.1%, the methodology includes QRS complex detection, median beat computation, PCA for dimensionality reduction, and decision tree classification. This approach supports doctors in their clinical practice and has room for development; future iterations might incorporate characteristics related to the RR interval. Baygin*et al*. [18]used HIT-based characteristics and created a model for classifying arrhythmias, and accuracy rates for diagnosing seven- and four-class arrhythmias were 92.95% and 97.18%, respectively. The model is appropriate for practical healthcare applications and is computationally efficient. Ibtehaz*et al*. [19] presented VFPred, a method that uses machine learning and signal processing to identify ventricular fibrillation (VF) in ECG data. With a brief 5-second signal, VFPred demonstrated excellent sensitivity and specificity. The research aims to build wearable monitoring devices and broaden this method to identify additional arrhythmias. Manju and Nair [20], for the patient's benefit, cardiac arrhythmias must be detected early. Utilizing pre-processed ECG data, this work proposes a 96.68% accurate approach for diagnosing arrhythmia types by utilizing XGBoost for feature reduction and SMOTEENN for balance. The new approach facilitates early detection and may be further refined using real-time data for analysis.

Pandey *et al*. [21] contrasted to single classifiers, the ensemble SVM classifier presented in this work achieves 94.4% accuracy in categorizing heartbeats from MIT-BIH arrhythmia data. Mincholе*et al*. [22]focused on an early diagnosis and found that the electrocardiogram (ECG) is essential. Large-scale electronic health records have led to a boom in machine learning in the medical field. Deep knowledge, in particular, helps with patient screening, but it needs more physiological understanding. Like DNNs, ML analyses ECGs effectively, although it might be challenging to interpret the results. Prakash and Ari [23] classified automatic ECG arrhythmias as essential for reducing the death rate of cardiac patients. The CNN and DTCWT-random forest techniques accurately categorize five different beat patterns. Islam *et al*. [24] identified that ECG readings have progressed thanks to intense machine learning. SMOTE-Tomek data balancing is used to improve performance with a higher level of accuracy for the single-lead ECG model CAT-Net. Zhao *et al*. [25] approach detects contactless arrhythmias in everyday situations using mm-Wave radar. With ensemble learning, a custom encoder-decoder model attains better performance.

Dhyani *et al*. [26] used SVM and 3D DWT to analyze ECG signals. SVM outperforms CSVM and WSVM, achieving high accuracy on the CPSC 2018 dataset. The method shows potential for clinical use and excels at accurately classifying arrhythmias. Lee and Kim [27] proposed that the steering wheel single-lead ECG device be used to track the cardiac health of the driver. The suggested method guarantees steady ECG readings in the face of noise. Farooq *et al*. [29] observed that as the world's population, urbanization, and globalization increase, so does public attention to health care. ECGs are classified by a LabVIEW-based system, which helps in early diagnosis and lessens hospital crowding. Pandey *et al*. [30] proposed a novel RBM model for arrhythmia classification from ECG data made possible by sophisticated deep learning algorithms, which yield excellent accuracy results. The literature shows that machine learning models result in deteriorated performance unless specific optimizations support them.

3. PROPOSED SYSTEM

This section presents the proposed system, including the machine learning framework, feature engineering, hyperparameter optimization, algorithms, and evaluation methodology.

3.1 Problem Definition

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Providing ECG data of a patient and developing a machine learning framework with optimizations to detect different kinds of arrhythmia automatically are challenging problems to consider.

3.2 Proposed Framework

We proposed a machine learning-based framework for the detection and classification of arrhythmia usingECG data. The framework is based on supervised learning, which requires training and test samples to complete its functionality. The framework is illustrated as shown in Figure 1. The data set used in this study is known as the MIT-BIH Arrhythmia dataset [31]. This is a widely used benchmarking dataset for Arrhythmia detection. The dataset is subjected to preprocessing, as illustrated in Figure 2. The preprocessing transforms the data into a format that can be used for machine learning. Once preprocessing is completed, the data set is divided into 80% and training under 20% for testing. Afterward, the training data will be subjected to feature engineering. Feature engineering is a process in which all the features are extracted from the given data set, and the features that can contribute to the class label selection process are identified. Unless contributing features are selected, machine learning algorithms or any classification algorithm shows deteriorated performance. This is the significance of feature selection in this research. Apart from feature selection, we also used an approach for hyperparameter optimization. This kind of optimization enables machine learning models to use parameters with appropriate values to improve the accuracy of the models.

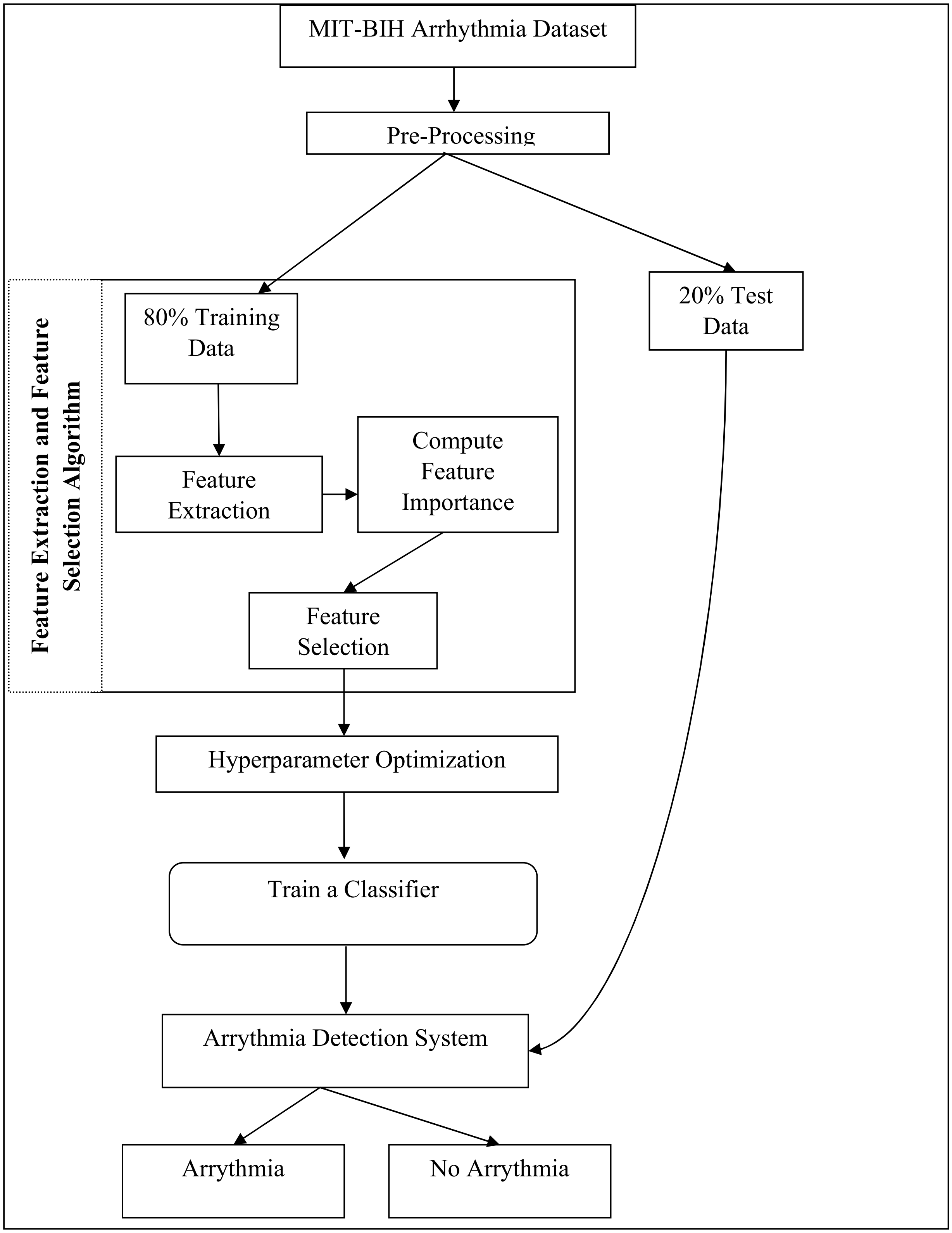


Figure 1: Proposed ML-Based Framework For Arrhythmia Detection From ECG Data

Journal of Theoretical and Applied Information Technology
15th July 2024. Vol.102. No. 13
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In this research, multiple machine-learning models are used to detect arrhythmia. Two optimizations, feature selection, and hyperparameter optimization, help the machine learning models leverage their performance in detecting arrhythmia from ECG data. The machine learning models are configured to perform multi-class classification. Five classes are involved in the data, which means that each test sample can be classified as one of those five class labels. Table 1 shows different classes under corresponding class labels.

Class Label	Description	Class Index
N	Normal	1
/	Paced Beat	2
R	Right Bundle Branch Block Beat	3
L	Left Bundle Branch Block	4
V	Premature Ventricular Beat	5

Table 1 Shows Different Classes And Corresponding Class Labels.

By classifying the ECG of each patient into one of the five classes, the proposed system can detect the presence of heart disease, known as arrhythmia, and also classify the kind of disease. Therefore, this research is more beneficial as it provides artificial intelligence-based diagnosis, which can help doctors. Thus, the proposal system can act as a clinical decision support system that can be integrated with applications used in health care units. The proposed architecture enables a pipeline that generates arrhythmia prediction from raw ECG data. A data pre-processing and feature extraction receives the raw ECG input, transforms the signal, and extracts the features required for arrhythmia detection. This system produces the prediction using a scalable machine-learning classifier.

3.3 Pre-Processing

Pre-processing is an important optimization in this research for improving the performance of machine learning models. As shown in Figure 2, pre-processing involves several steps.

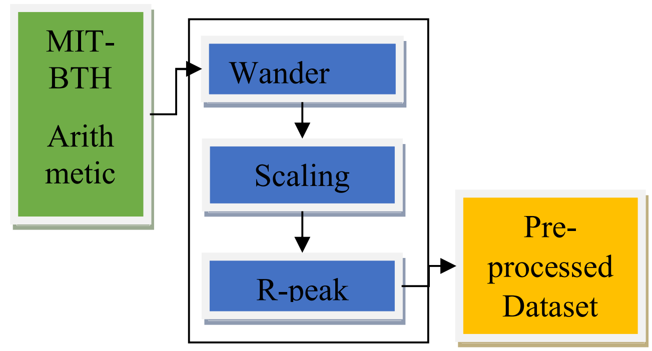


Figure 2: Pre-Processing Pipeline

In the ECG signals, baseline wander is some noise that may occur due to body movements, respiration, perspiration, and poor electrode contact. Unless the magnitude of the wander is removed, it may hinder the performance of ML models. This paper removes wander using a low-pass Butterworth filter by applying it to ECG samples in the dataset with 50Hz as the cutoff frequency. Afterward, the ECG signals are further subjected to scaling between 0 and 1 with the help of the min-max approach to reduce discrepancy among ECG samples due to diversified sensor hardware used by healthcare units for measuring ECG signals. The given dataset has R-peak values associated with ECG. As the proposed system does not need a human expert, it is essential to detect R-peak values. Thus, the pre-processing of given samples is completed.

3.4 Feature Selection

Feature selection in this research is another vital optimization that enables the proposal system to exploit optimal features for improving the prediction performance of machine learning models. Raw ECG data set denoted as γ is considered for experiments. It is expressed as in Eq. 1, where w represents the window length in seconds, v_s represents the sampling frequency in Hz, f represents the Heart Rate Variability (HRV) characteristics, and n is the variable subset of features. Eq. 2 shows the value of Δ denoting the feature selection process.

$$\gamma = [\gamma_1 \gamma_2 \dots \dots . . \gamma_{w \times v_s}] \xrightarrow{\Delta} [f_1 f_2 \dots \dots . . f_n] \quad (1)$$

$$\Delta = [\alpha, \beta, k, \lambda] \quad (2)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Because HRV properties are more relevant than other generic non-linear features like Shannon entropy, Renyi entropy, and Tsallis entropy, they were considered for this investigation [20]. According to Gokul H *et al.* [21], characteristics in the frequency domain need Fast Fourier Transformations, which are difficult to extract in the first place. Time-domain properties may be retrieved quickly, as they are statistical. R-peaks are converted to 14 time-domain characteristics by Δ.

Aiming to identify the ideal set of features that would maximize the performance of the underlying classifier while minimizing the number of features selected, feature selection was done because the classifier's prediction time is directly proportional to the number of features implemented. Consequently, the classifier's computational cost and processing time were decreased, leading to an ideal feature subset. As per ANOVA, or analysis of variance [23], every independent variable and the predictor variable were compared using the F-test. A metric known as an F-statistic or F-value is used to assess if there is a substantial difference in variance between the means of two populations. The predictive power increases with the F-statistic value. Figure 3 displays the computed F-values for the characteristics retrieved from both datasets. Eq 3 shows the features in descending order of their F-values, indicating each feature's significance in the F Data.

$$F^{Data} = \{f | f_i, F - value(f_i) > F - value(f_{i+1}), 1 \leq i \leq 14\} \quad (3)$$

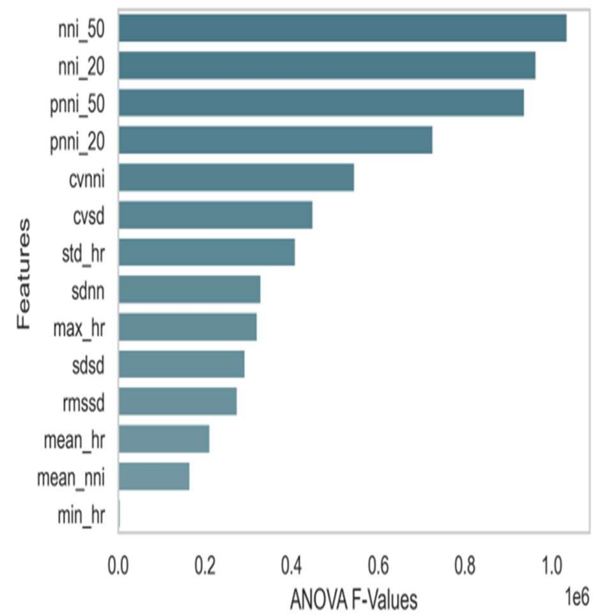


Figure 3: Illustrates Features And Their Importance

The computation of future importance enables the system to find the importance of each future. Once feature importance is known, it is possible to choose the best features for machine learning models based on a threshold value.

3.2 Training and Testing

The training data is used to train the models in the training phase. Figure 4 shows the complete pipeline of the training process. The training data is subjected to preprocessing, as discussed earlier. Then, feature extraction is carried out to identify all the possible features. The feature selection approach is applied to find the best contributing features from all the extracted features. With the final feature set, the given machine learning model is trained. Once the model training is completed, the model will be saved for future reuse.

Journal of Theoretical and Applied Information Technology
15th July 2024. Vol.102. No. 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

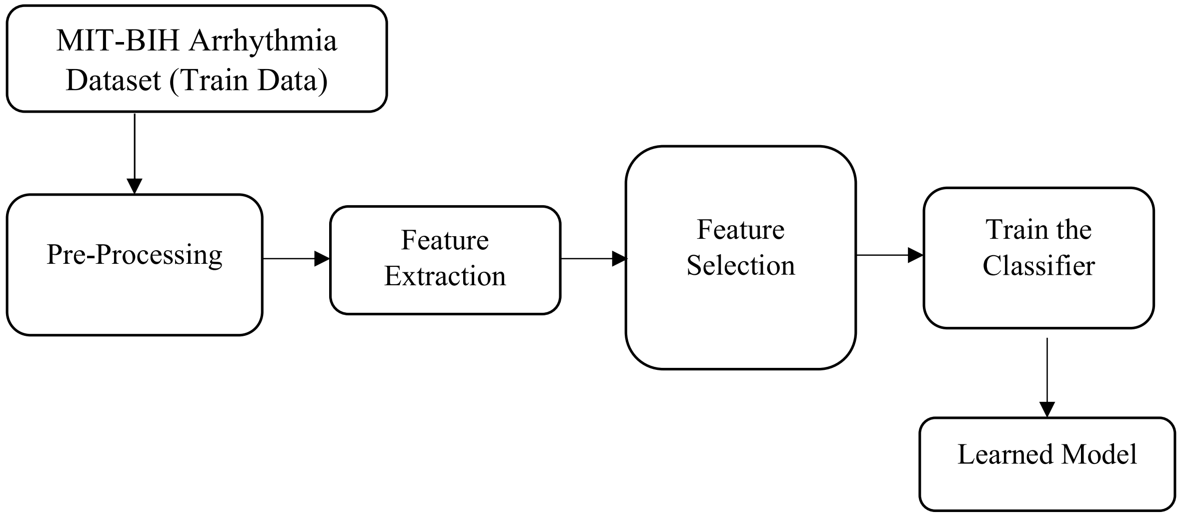


Figure 4:Pipeline Of The Training Phase

Many AI systems in the healthcare industry cannot achieve clinical applicability or generalizability. Because local clinical and administrative methods vary and sites differ technically, generalization can be challenging [24]. We employed a benchmark dataset for our empirical study. We could test and train the models on varied subsampled datasets by building a solid train-test method. The models were also evaluated using the same dataset they were trained on to create a point of comparison.

$F_n^A = \{f : f_i | 1 \leq i \leq n, f_i \in F^A\}, 1 \leq n \leq 14$
(4)

$F_n^C = \{f : f_i | 1 \leq i \leq n, f_i \in F^C\}, 1 \leq n \leq 14$
(5)

The models were trained on many feature subsets to select the best subset based on pipeline inference time, memory costs, and performance. The classifier in T_{AA} and $T_{CC}$was trained using feature subsets F_n(n = 4, 6, 8, 10, 12, 14),Hich were the most significant features selected by ANOVA. Based on accuracy, the best feature subsets for $T_{AA} F_{best}^A$ and $T_{CC} F_{best}^C$were identified. Two subsets of features obtained from $F_{best}^A \cup F_{best}^C$ and $F_{best}^A \cap F_{best}^C$, were used to train T_{AA} and T_{CC}. For T_{AA} and T_{CC}a 5-fold cross-validation was used.

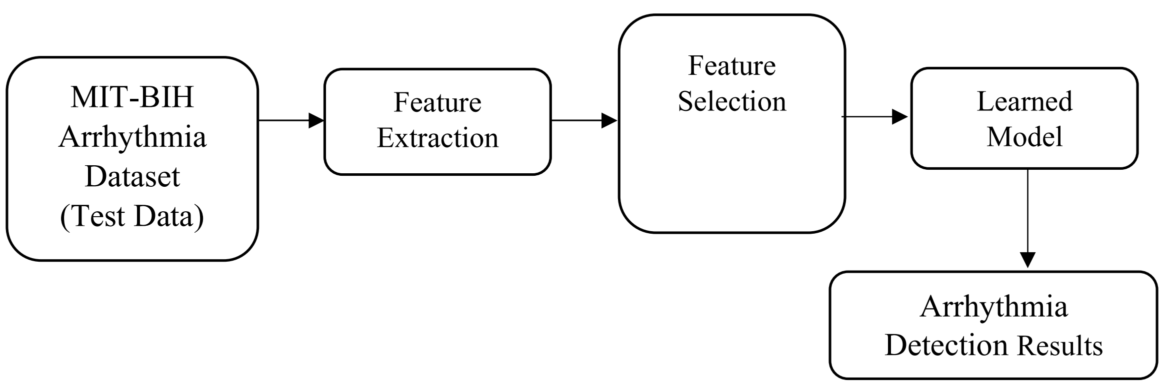


Figure 4:Pipeline Of The Testing Phase

In the testing phase, the given test data is subjected to feature extraction and selection. Once features are selected, the selected feature subset is given to a previously trained model for analyzing each test sample to predict the presence or absence of arrhythmia. The proposed system also supports multiclass classification, where for each ECG test data, the system not only detects arrhythmia but also classifies it.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.4 Proposed Algorithms

In this research, we proposed two algorithms. The former is meant for optimizing machine learning models, while the latter is intended for performing multiclass classification.

Algorithm 1: Feature Selection and Hyperparameter Optimization (FSHO)

Input: MIT-BIH Arrhythmia dataset D, threshold th, models pipeline M

Output: Selected features F', optimized models M'

1. Begin
2. Initialize feature importance vector V
3. D'←Preprocess(D)
4. F←ExtractFeatures(D')
5. V←FindFeatureImportance(ANOVA, F)
6. F'←SelectBestFeatures(F, V, th)
7. For each model m in M
8. m'←GridSearch(hyperparameters, D')
9. Update M'
10. End For
11. End

Algorithm 1: Feature Selection and Hyperparameter Optimization (FSHO)

Algorithm 1 inputs MIT-BIH Arrhythmia dataset D, threshold th, and models pipeline M. The algorithm performs preprocessing of given data. Preprocessing is one of the optimizations made on the data in this paper to improvethe performance of machine learning models. The preprocessing includes wander removal, scaling, and R-peak detection for each ECG sample in the given data set. After preprocessing is completed, the algorithm extracts all features from the dataset. However, using all features may result in mediocre performance when using machine learning models. To overcome this problem, the algorithm computes each feature's importance and finally chooses the best-performing features based on the given threshold. Apart from feature selection, the algorithm also performs hyperparameter optimization on all the machine-learning models in the pipeline. Grid search-based hyperparameter optimization is performed on each machine learning model to optimize its performance in arrhythmia detection.

Algorithm 2: Learning-based Arrhythmia Detection and Classification (LbADC)

Input: MIT-BIH Arrhythmia dataset D, models pipeline M

Output: Arrhythmia detection results R, performance statistics P

1. Begin
2. (F, M)←FSHOAlgorithm(D, th, M)
3. (T1, T2)←SplitData(D)
4. For each m in M
5. Train m using T1 and F
6. Persist model m
7. End
8. For each persisted model m in M
9. R←ArrhythmiaDetection(T2, m)
10. P←Evaluate(R, ground truth)
11. Display R
12. Display P
13. End For
14. End

Algorithm 2:Learning-based Arrhythmia Detection and Classification (LbADC)

Algorithm 2 takes the given dataset and machine learning models pipeline as input. This algorithm makes use of the FSHO (Algorithm 1) algorithm to perform two operations. The first operation, known as feature selection, is based on the features available in the dataset. The second operation, hyperparameter tuning, is carried out for each machine-learning model. Then, the dataset is split into a training set denoted as T1 and a test set denoted as T2. In the training process, each model is trained with the training data and corresponding features. Once the training is completed, the model will persist for future reuse. Afterward, the testing process is iterative in nature, which exploits each learned model to classify the test samples with multi-class classification. Finally, the algorithm returns the detection results and also performance statistics.

3.5 Evaluation Methodology

Since we used a learning-based approach (supervised learning), metrics derived from the confusion matrix, shown in Figure 5, are used for evaluating our methodology.

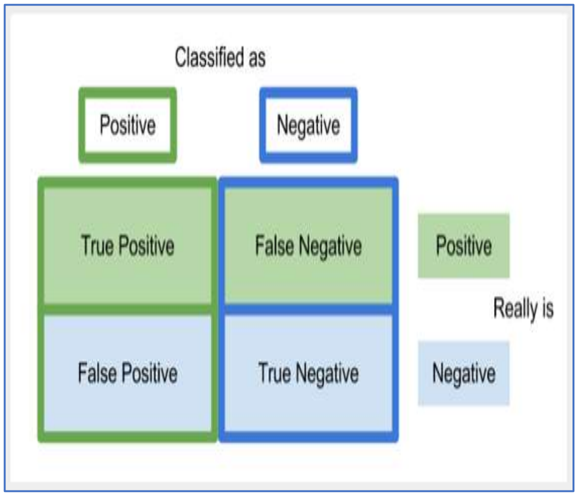


Figure 5: Confusion Matrix

Our method's predicted labels are compared with ground truth based on the confusion matrix to arrive at performance statistics. Eq. 6 to Eq. 9 express different metrics used in the performance evaluation.

$$\text{Precision (p)} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Recall (r)} = \frac{TP}{TP+FN} \quad (7)$$

$$\text{F1-score} = 2 * \frac{(p * r)}{(p+r)} \quad (8)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

The measures used for performance evaluation result in a value that lies between 0 and 1. These metrics are widely used in machine learning research.

4. EXPERIMENTAL RESULTS

This section presents the experimental results of the proposed framework and underlying algorithms. Experiments are made with the MIT-BIH Arrhythmia dataset [31]. The prototype is implemented using the Python programming language under machine learning libraries. This section provides observations made in terms of exploratory data analysis, multiclass classification of different machine learning models, and performance comparison.

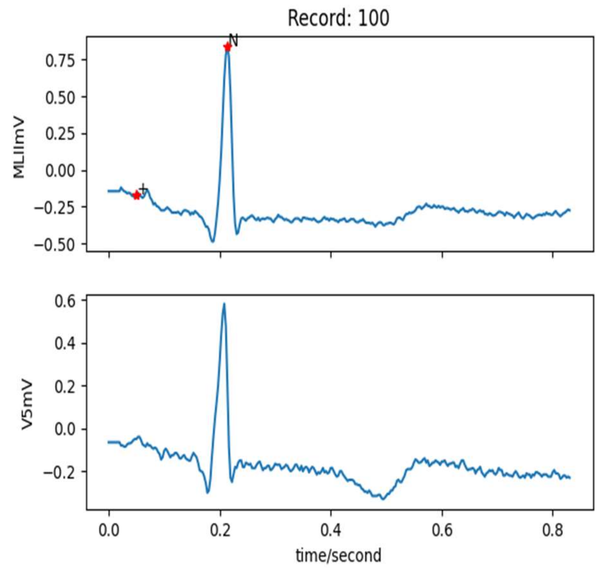


Figure 6: A Sample ECG Record

As presented in Figure 6, a sample ECG record reflects the heartbeat variations in the temporal domain.

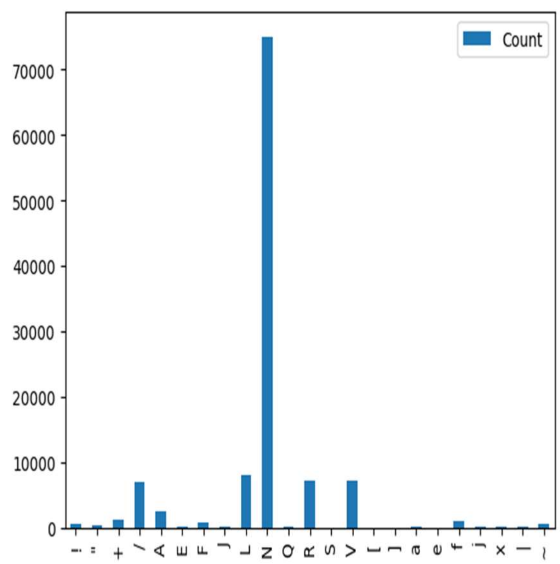


Figure 7: Data Distribution For All Annotations In The Dataset

Figure 7 shows the data distribution dynamics of a given ECG sample with different annotation values.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

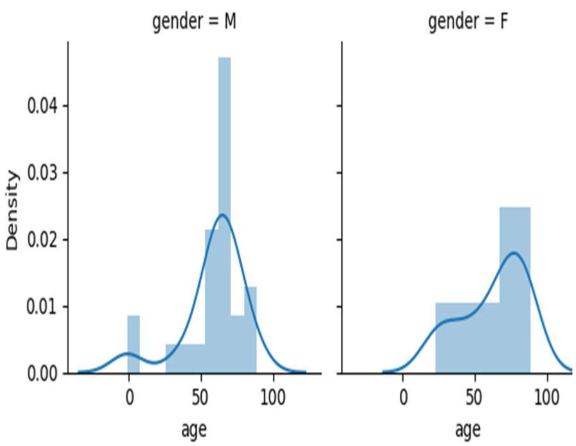


Figure 8: Gender And Age-Wise Data Distribution

Figure 8 shows the data distribution age-wise for both male and female patients. The data set includes more male patients between 60 and 70 years of age, while more female patients are between 60 and 80 years of age.

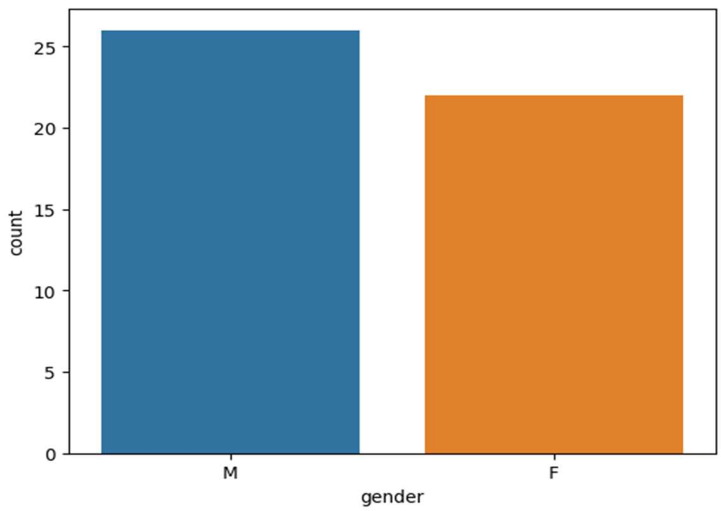
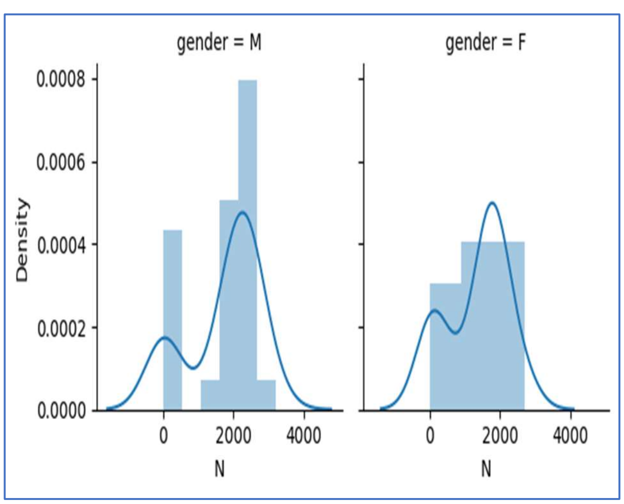
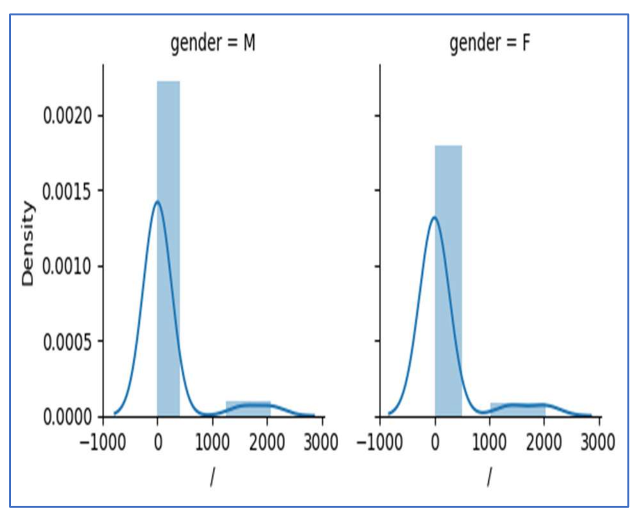


Figure 9: Gender-Wise Data Distribution

As shown in Figure 9, the details of the dataset are provided in terms of the number of ECG samples available for both male and female patients.





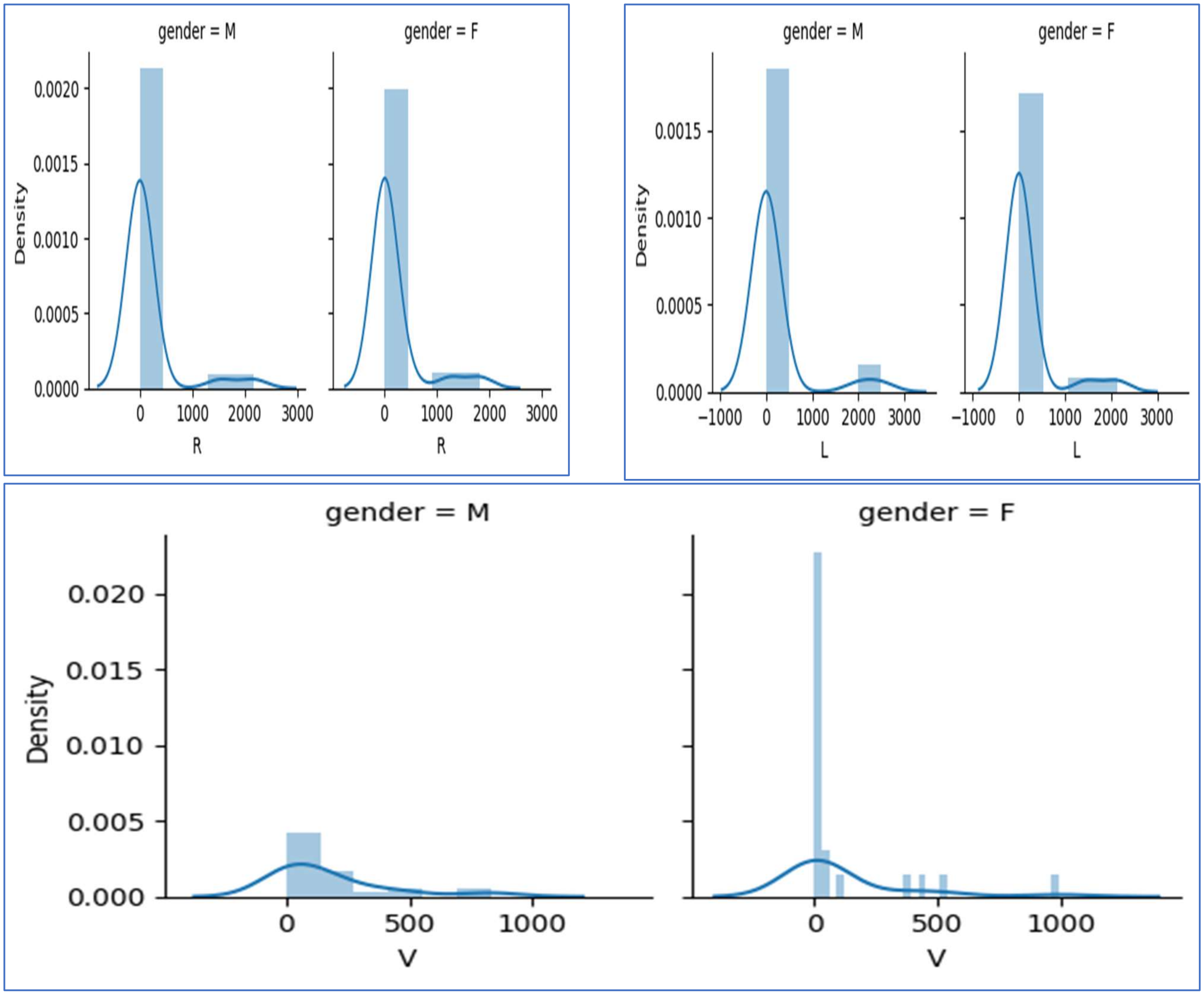
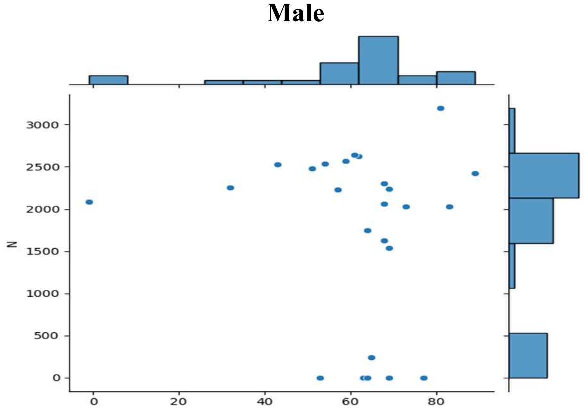
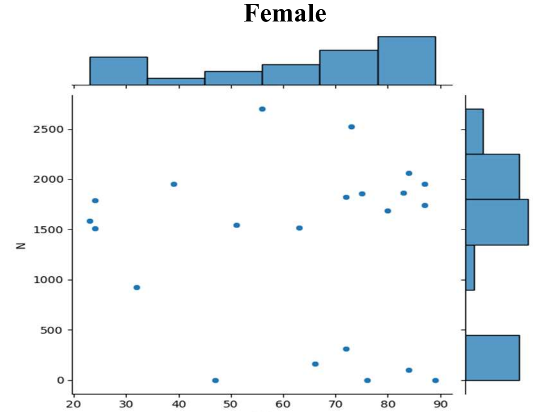
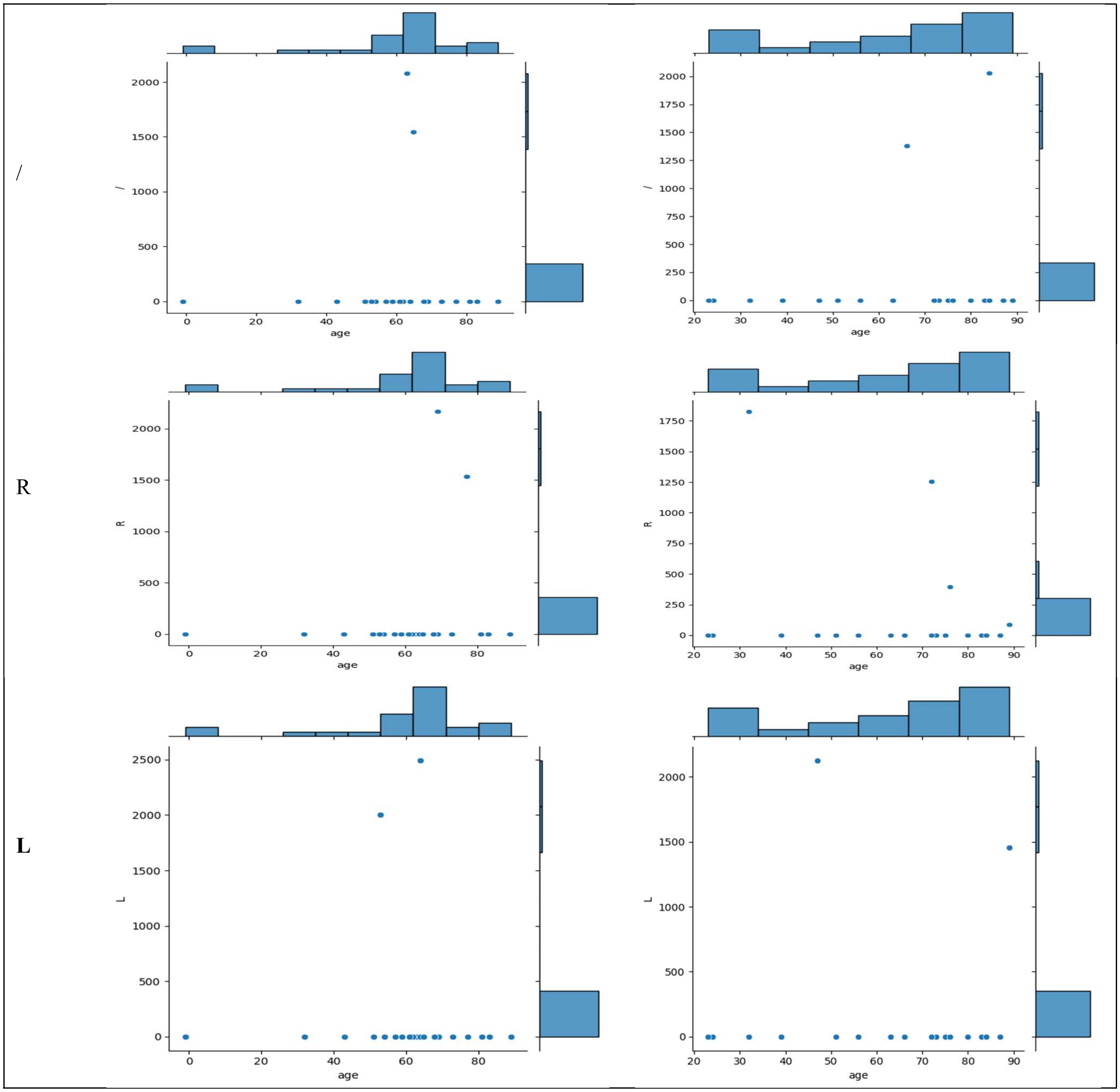


Figure 10: Data Distribution For All Classes Involved

For every class of arrhythmia disease, the density of samples is visualized for male and female patients in Figure 10.

Class	Male	Female
N		

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



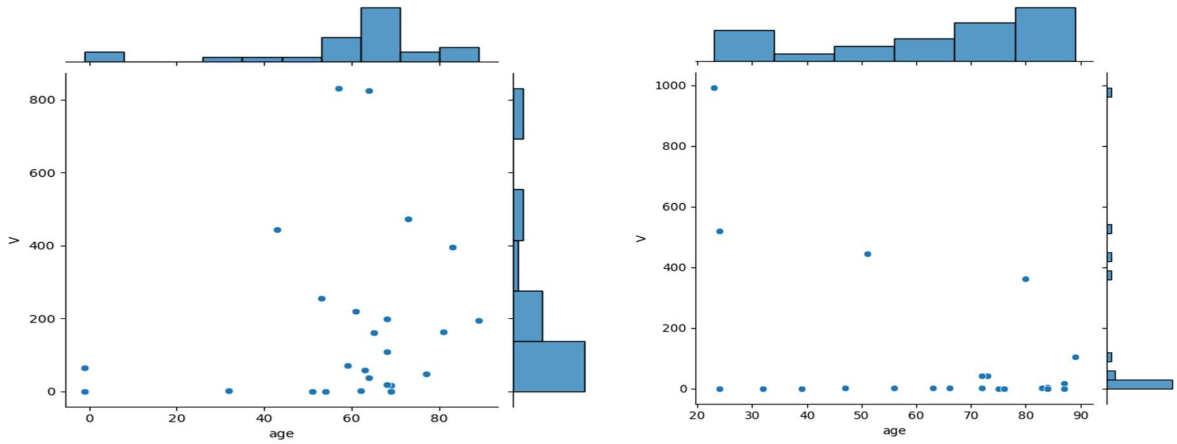
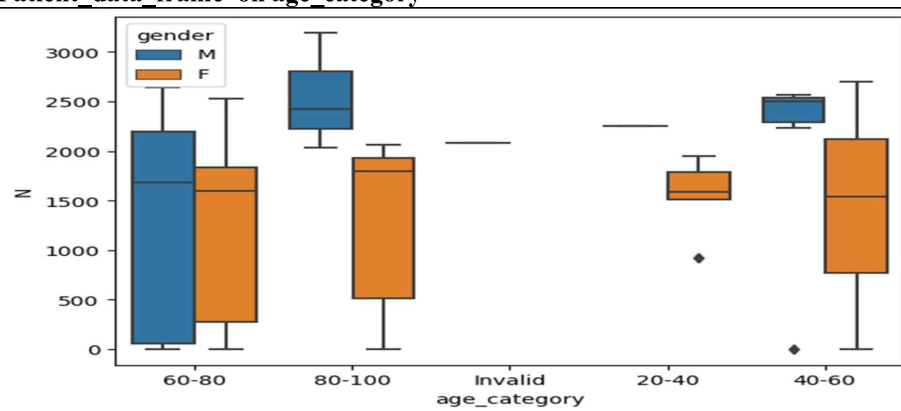
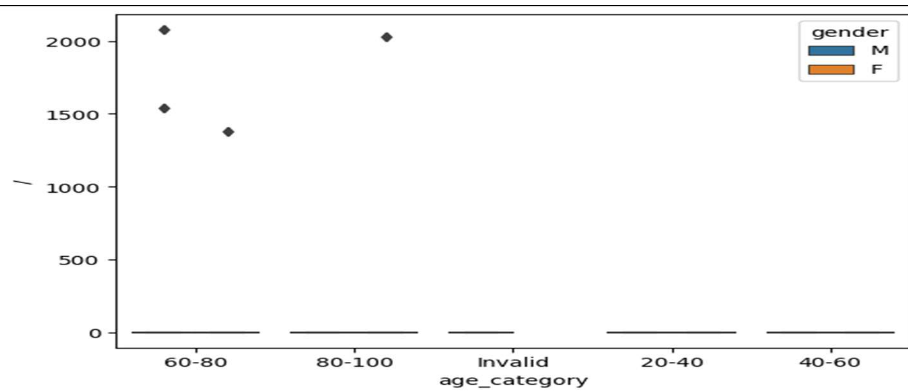
V	

Figure 11: Gender wise and age wise data distribution for all classes

The gender wise on the age wise data distribution dynamics are visualized in Figure 11 for all possible arrhythmia classes.

Class Name	Patient_data_frame on age_category
N	
/	

Journal of Theoretical and Applied Information Technology
15th July 2024. Vol.102. No. 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

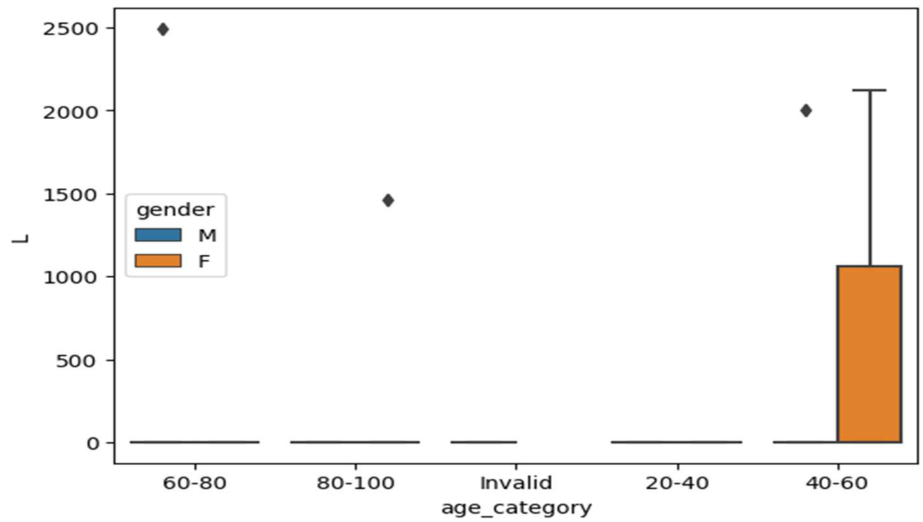
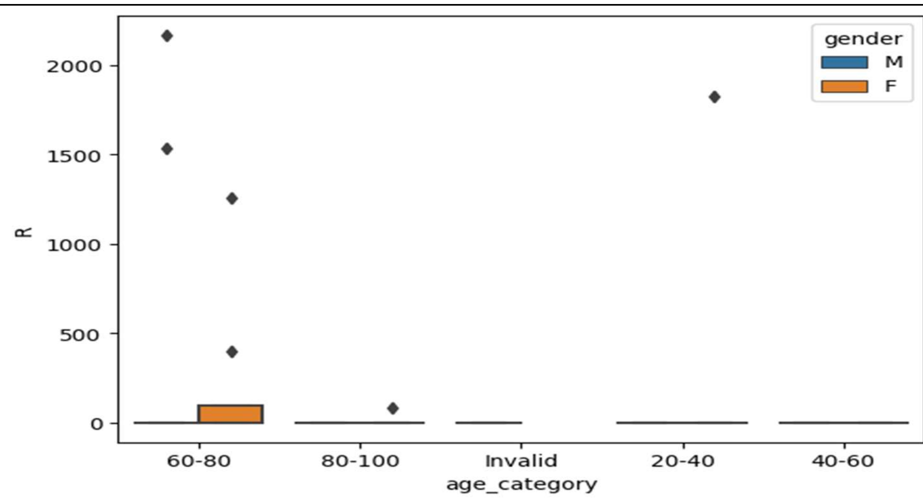
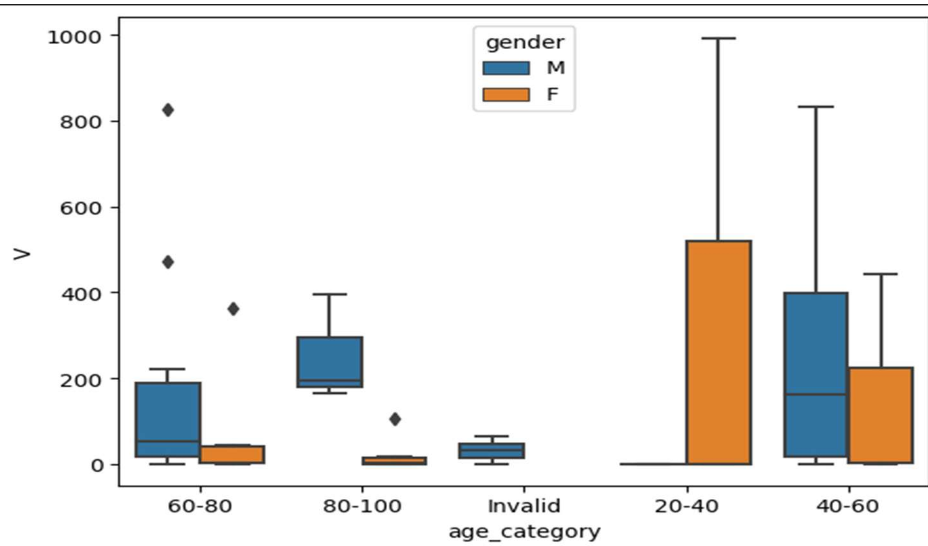
<table>
<tr><td>L</td><td>
</td></tr>
<tr><td>R</td><td>
</td></tr>
<tr><td>V</td><td>
</td></tr>
</table>

Figure 12: Category Of Age Gender-Wise For Each Class In The Data Set

As presented in Figure 12, the male and female population in the dataset are visualized for each class of arrhythmia against different age groups.

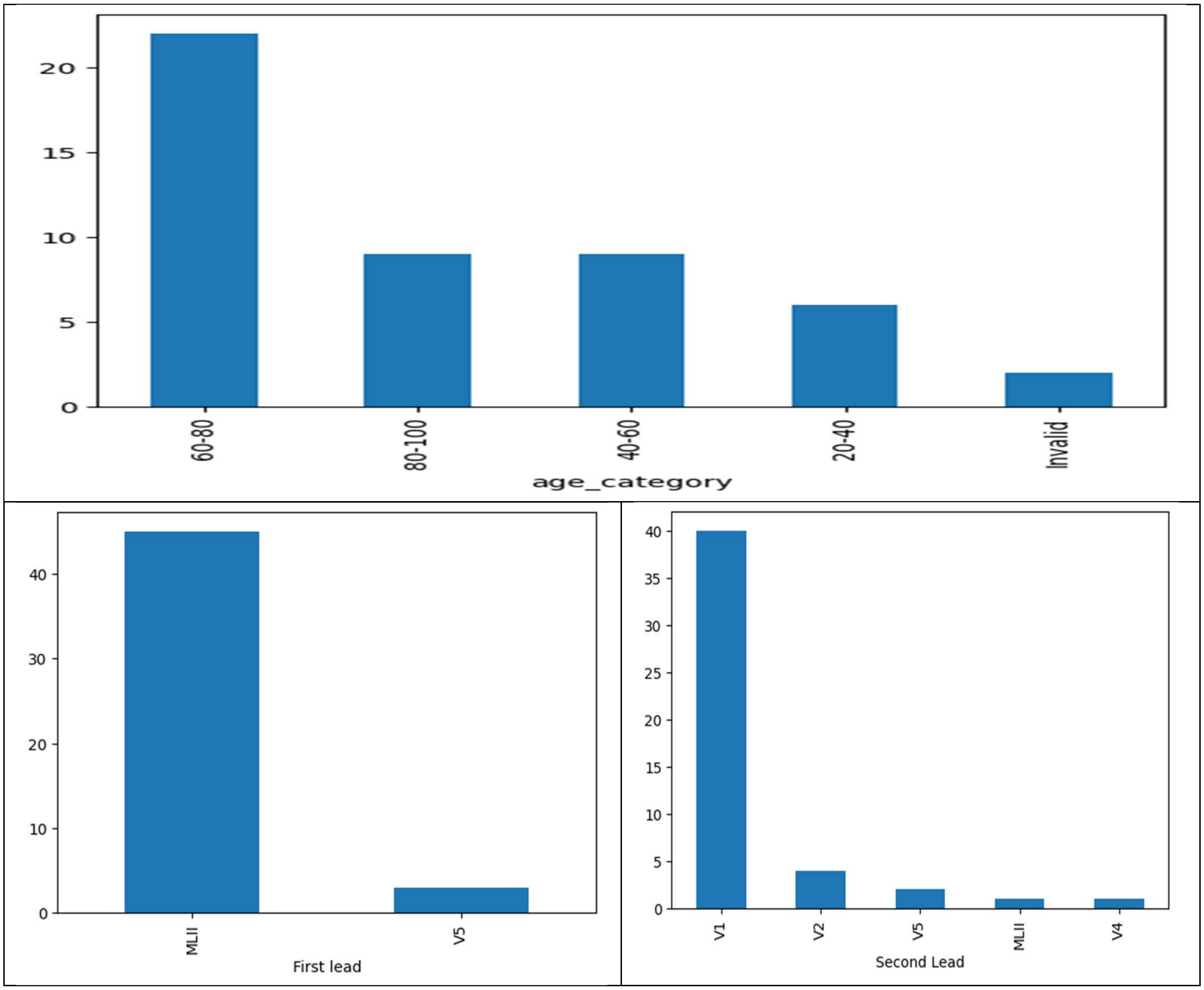


Figure 13: Showing Age Category Besides First Lead And Second Lead Details

As presented in Figure 13, different areas of ECG signals are presented along with first lead and second lead details and age categories.

4.2Data Preprocessing

This section presents the results of different methods used for pre-processing of given ECG samples.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

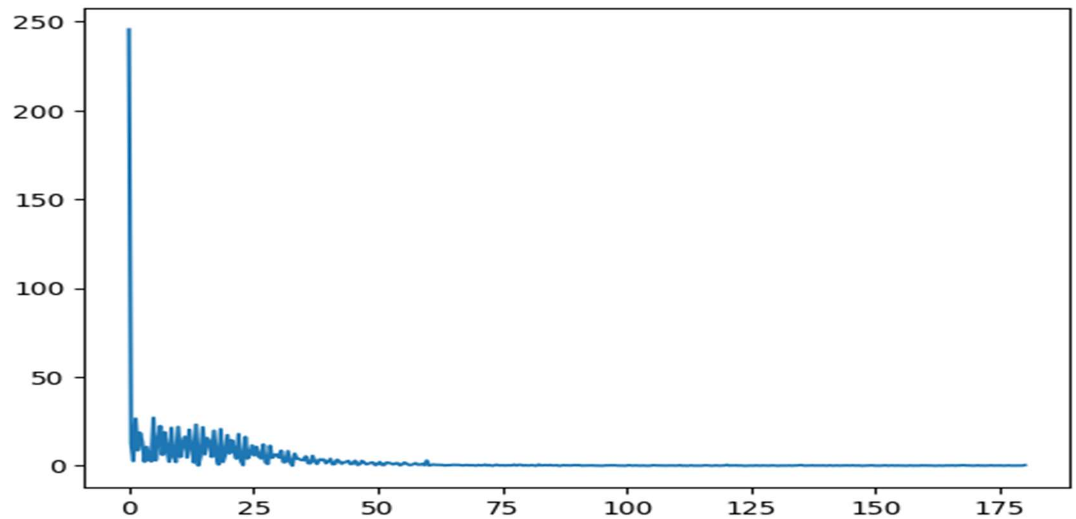


Figure 14: Frequencies Present In The Given ECG Signal

As presented in Figure 14, an ECG sample is visualized reflecting different frequencies linked to heart beat of given patient.

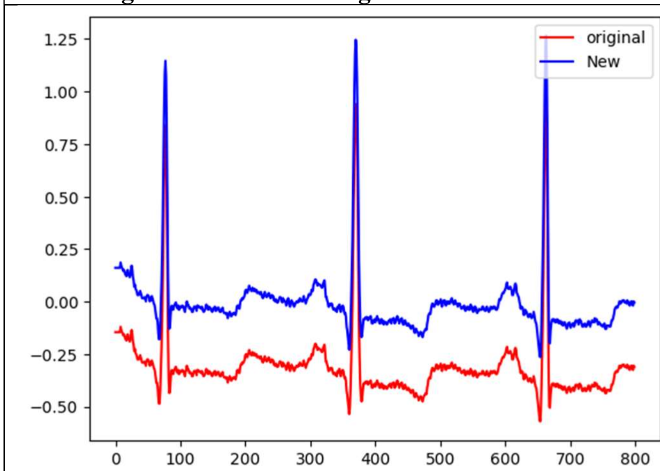
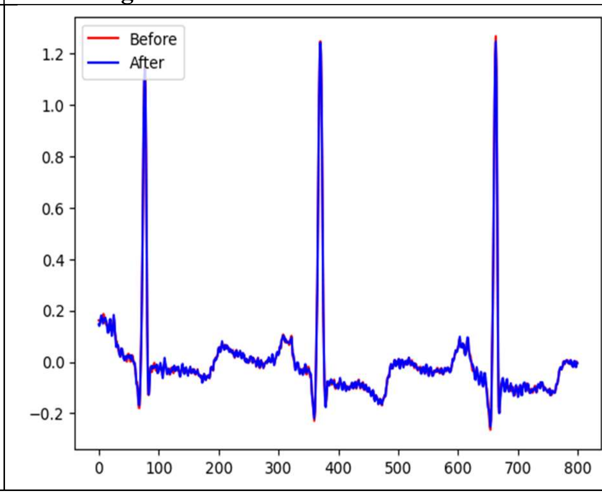
Removing Baseline Wandering	**Removing Power Line Interference**
	

Figure 15: Shows Preprocessing Results Of Wandering Removal And Power Line Interference Removal

As presented in Figure 15, preprocessing results are visualized reflecting wander removal and also removal of power line interference. Power line interference is the interference associated with voltage in the ECG sample which has 50 Hz frequency. So, the interference can be detected and delimited as part of preprocessing.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

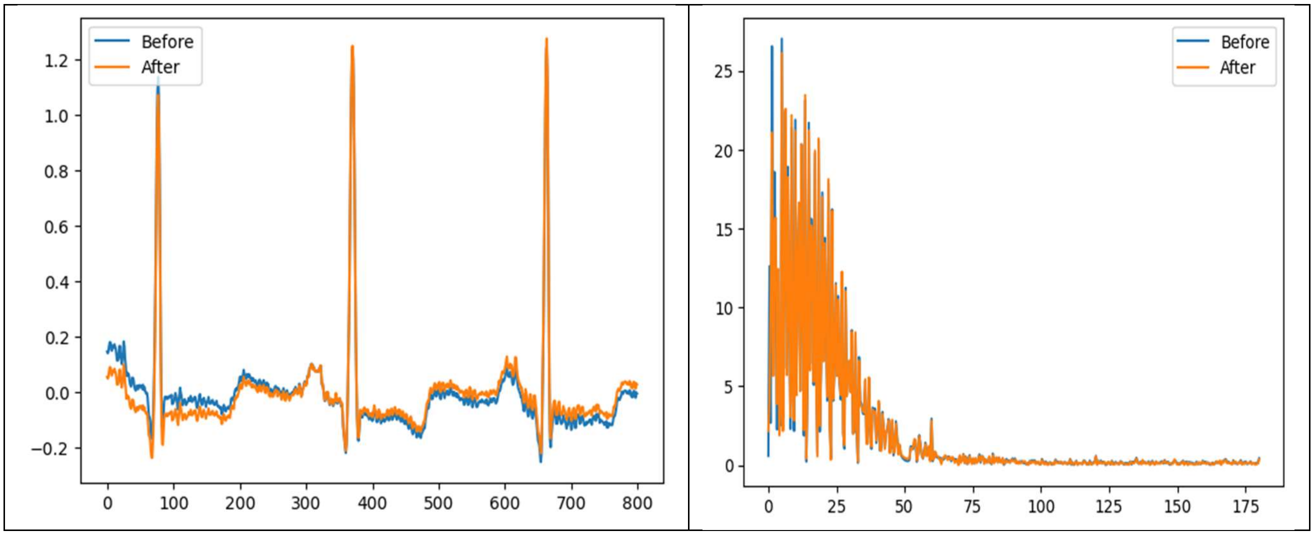


Figure 16: Result Of High Pass Filter

High-pas filter is a filter used to reduce noise in ECG sample. A high pass filter reduces low frequency interference signals that are due to breathing moments and these signals are generally between 0.05 Hz and 1.0 Hz.

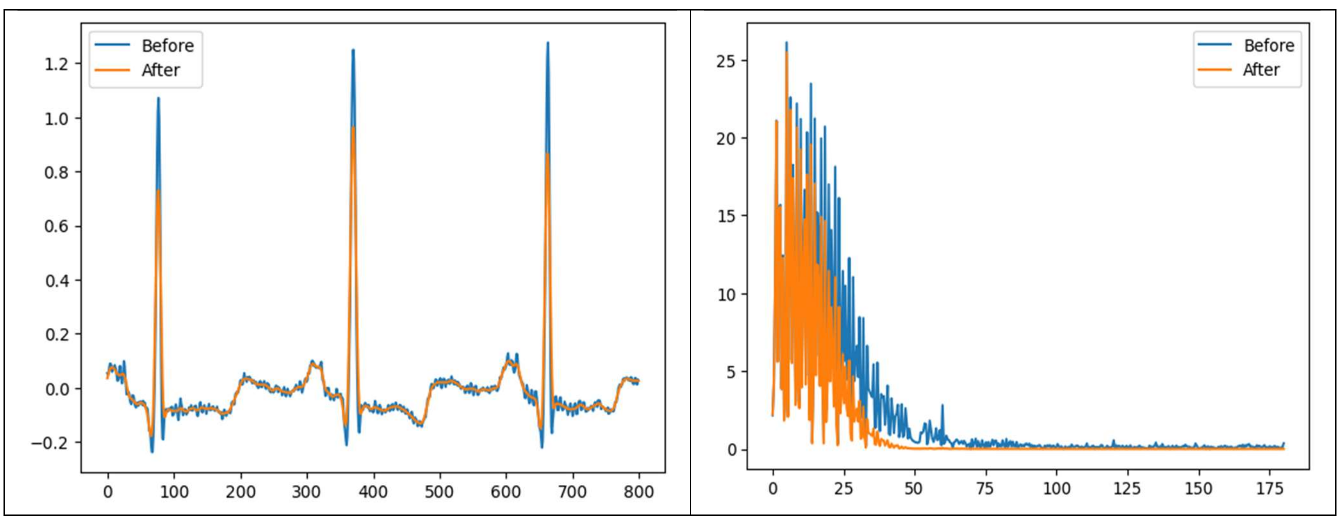


Figure 17: Result Of Low Pass Filter

Has presented in figure 17 a low pass filter is used to deal with interference in ECG sample. In fact, low pass filters are capable of reducing high frequency interference caused by muscle tremors under these frequencies are between 5 Hz and 450 Hz.

Journal of Theoretical and Applied Information Technology
15th July 2024. Vol.102. No. 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

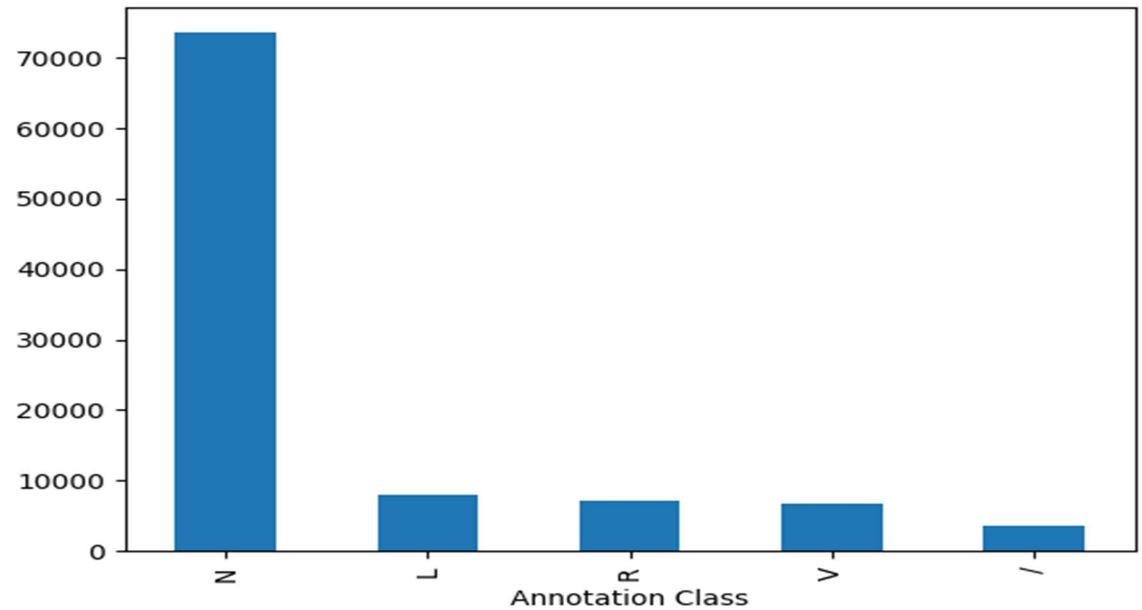
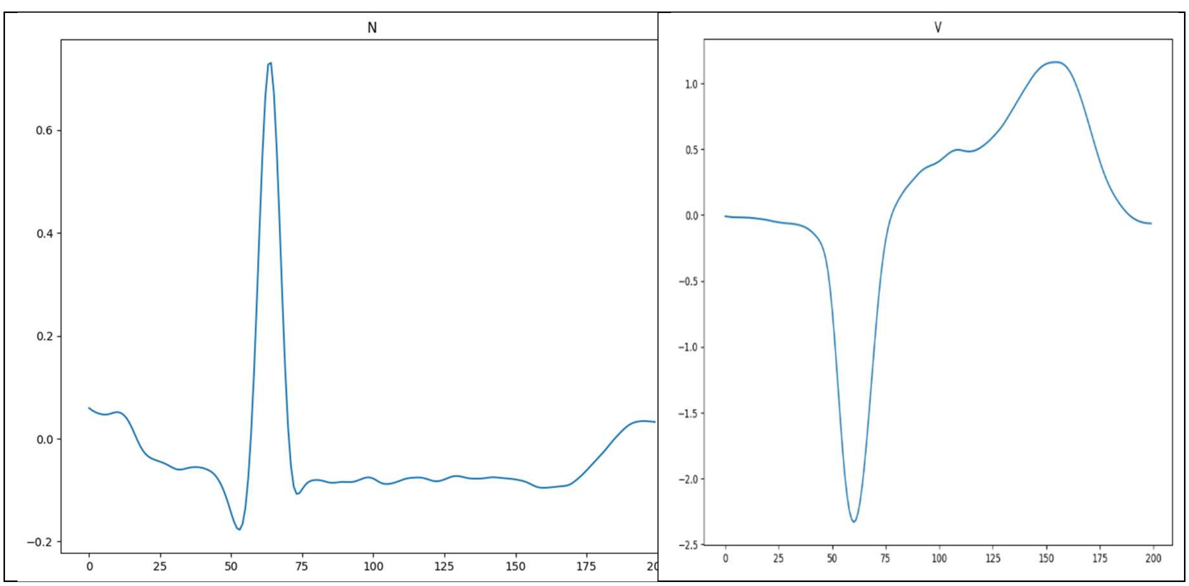


Figure 18: The Five Classes And Their Data Distribution Dynamics

As presented in figure 18 the data distribution of a 5 courses of arrhythmia of ECG samples is visualized.

4.3 Machine Learning Results

This section presents the results of machine learning models used in this research. The models are evaluated, and the observations are presented in terms of precision, recall, F1-score, and accuracy.



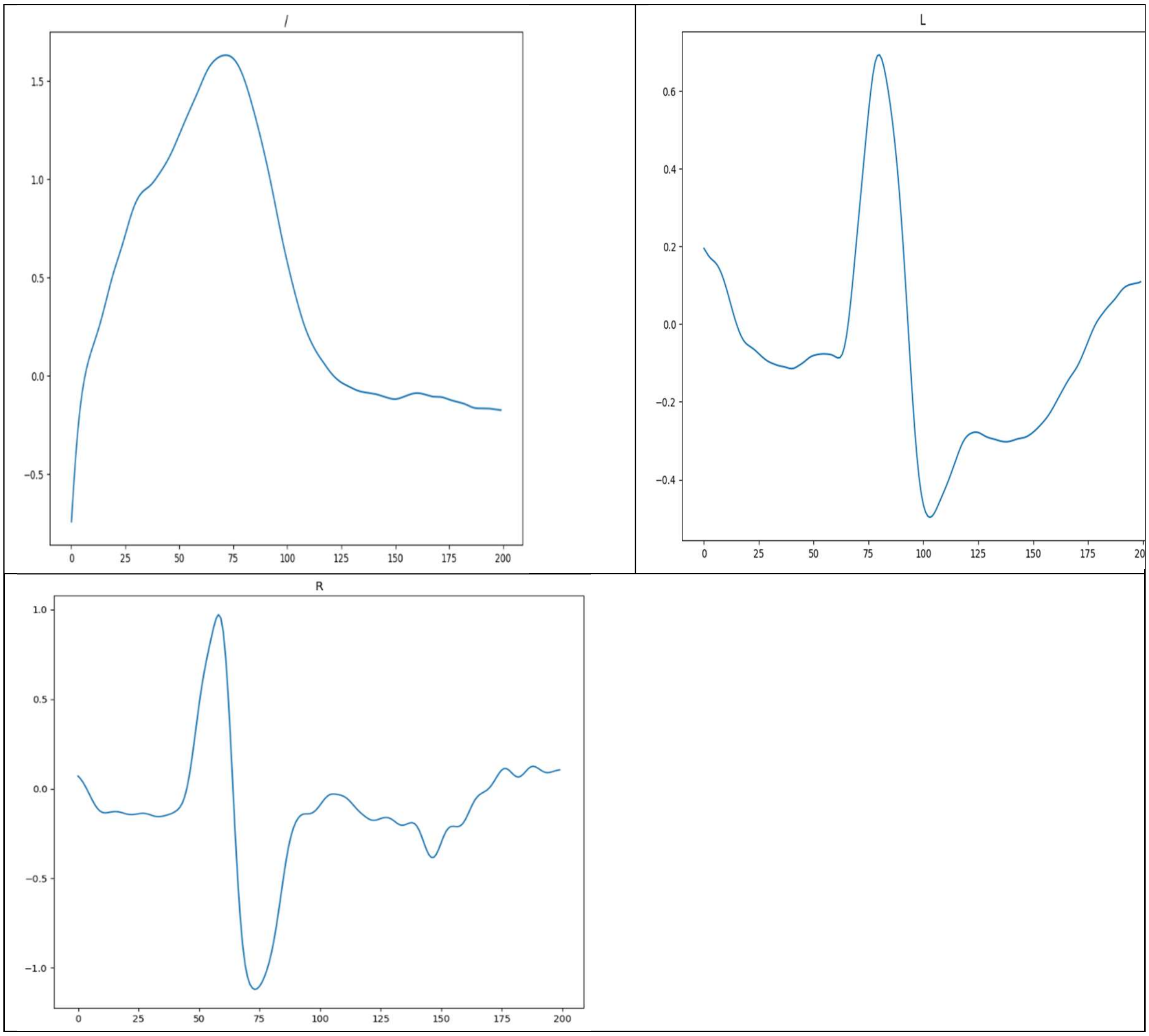


Figure 19: Shows The ECG Signal For Each Class

As presented in Figure 19, ECG sample for every arrhythmia class are provided. Each sample represents a specific class of arrhythmia.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

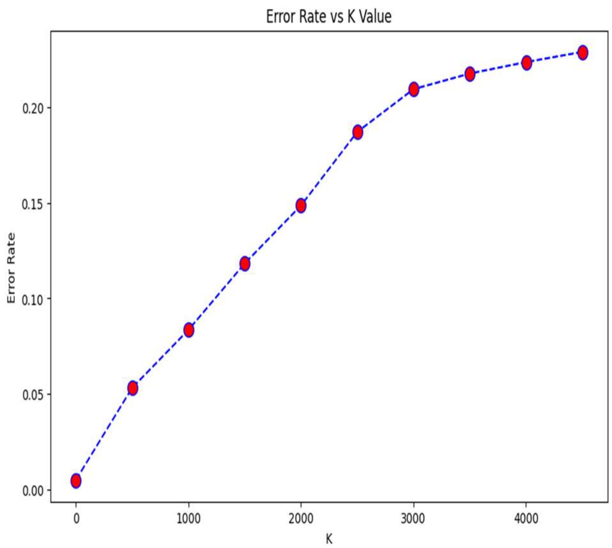


Figure 20: Shows The Error Rate Against The K Value For The KNN Model

The results of the KNN model are presented in Figure 20. As the k value is increased, there is an increase in the error rate.

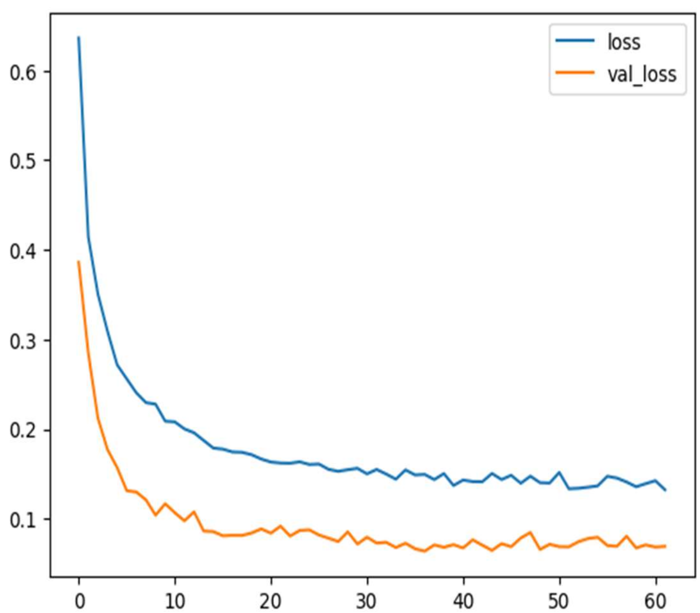


Figure 21: Loss Value Dynamics Exhibited By The ANN Model

The results of the ANN model in terms of training loss and validation loss are provided against the number of iterations. A low loss value indicates better performance.

4.4 Multi-Class Classification Results

The machine learning models used in this paper's empirical study are evaluated using multi-class classification. As part of multiclass classification, each test sample is labelled with one of the 5 arrhythmia classes.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Diagnosis Class	**Precision**	**Recall**	**F-1 Score**
Normal	0.84	0.76	0.79
Paced Beat	0.85	0.86	0.85
Right Bundle Branch Block Beat	0.97	0.69	0.81
Left Bundle Branch Block	0.72	0.74	0.73
Premature Ventricular Beat	0.77	0.89	0.83

Table 2: Multi-Class Classification Results Of The LR Model

As presented in Table 2, the results of the LR Model are provided for all 5 arrhythmia classes. Observations are made for every arrhythmia class of the ECG sample.

Diagnosis Class	**Precision**	**Recall**	**F-1 Score**
Normal	0.91	0.89	0.9
Paced Beat	0.89	0.95	0.92
Right Bundle Branch Block Beat	0.97	0.93	0.95
Left Bundle Branch Block	0.84	0.92	0.88
Premature Ventricular Beat	0.96	1	0.98

Table 3: Multi-Class Classification Results Of The KNN Model

As presented in Table 3, the results of the KNN Model are provided for all 5 arrhythmia classes. Observations are made for every arrhythmia class of the ECG sample.

Diagnosis Class	Precision	Recall	F-1 Score
Normal	0.93	0.97	0.95
Paced Beat	0.97	0.96	0.96
Right Bundle Branch Block Beat	0.91	0.87	0.89
Left Bundle Branch Block	0.87	0.94	0.9
Premature Ventricular Beat	0.89	0.96	0.92

Table 4: Multi-class classification results of the RF model

As presented in Table 4, the results of the RF Model are provided for all 5 arrhythmia classes. Observations are made for every arrhythmia class of the ECG sample.

Diagnosis Class	**Precision**	**Recall**	**F-1 Score**
Normal	0.93	0.86	0.89
Paced Beat	0.94	0.88	0.91
Right Bundle Branch Block Beat	0.91	0.93	0.92
Left Bundle Branch Block	0.92	0.91	0.91
Premature Ventricular Beat	0.96	0.87	0.91

Table 5: Multi-Class Classification Results Of The DT Model

As presented in Table 5, the results of the DT Model are provided for all 5 arrhythmia classes. Observations are made for every arrhythmia class of the ECG sample.

Diagnosis Class	Precision	Recall	F-1 Score
Normal	0.92	0.79	0.85
Paced Beat	0.91	0.86	0.88
Right Bundle Branch Block Beat	0.87	0.94	0.9
Left Bundle Branch Block	0.9	0.87	0.88
Premature Ventricular Beat	0.91	0.94	0.92

Table 6: Multi-Class Classification Results Of The SVM Model

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

As presented in Table 6, the results of the SVM Model are provided for all 5 arrhythmia classes. Observations are made for every arrhythmia class of the ECG sample.

Diagnosis Class	Precision	Recall	F-1 Score
Normal	0.94	0.88	0.91
Paced Beat	0.76	0.91	0.83
Right Bundle Branch Block Beat	0.73	0.95	0.82
Left Bundle Branch Block	0.59	0.92	0.72
Premature Ventricular Beat	0.84	0.89	0.86

Table 7: Multi-Class Classification Results Of The GNB Model

As presented in Table 7, the results of the GNB Model are provided for all 5 arrhythmia classes. Observations are made for every arrhythmia class of the ECG sample.

Diagnosis Class	Precision	Recall	F-1 Score
Normal	0.87	0.92	1
Paced Beat	0.84	0.88	0.96
Right Bundle Branch Block Beat	0.87	0.96	1
Left Bundle Branch Block	0.86	0.94	0.98
Premature Ventricular Beat	0.91	0.97	0.99

Table 8: Multi-Class Classification Results Of The ANN Model

As presented in Table 8, the results of the ANN Model are provided for all 5 arrhythmia classes. Observations are made for every arrhythmia class of the ECG sample.

4.5 Performance Comparison

Arrhythmia Detection Model	Avg Precision	Avg Recall	Avg F-1 Score	Avg Accuracy
Logistic Regression	**0.84**	**0.76**	**0.8**	**0.843**
KNN	**0.92**	**0.94**	**0.93**	**0.968**
Decision Tree	**0.93**	**0.89**	**0.91**	**0.92**
Random Forest	**0.91**	**0.93**	**0.92**	**0.961**
SVM	**0.9**	**0.87**	**0.88**	**0.891**
Gaussian Naive Bayes	**0.76**	**0.91**	**0.83**	**0.81**
ANN	**0.87**	**0.94**	**0.9**	**0.924**

Table 9: Performance Comparison Among All Models

Table 9 presents the results of Arrhythmia detection and classification using different machine learning models. Since this is a multi-class classification, the average of observations of all classes is considered.

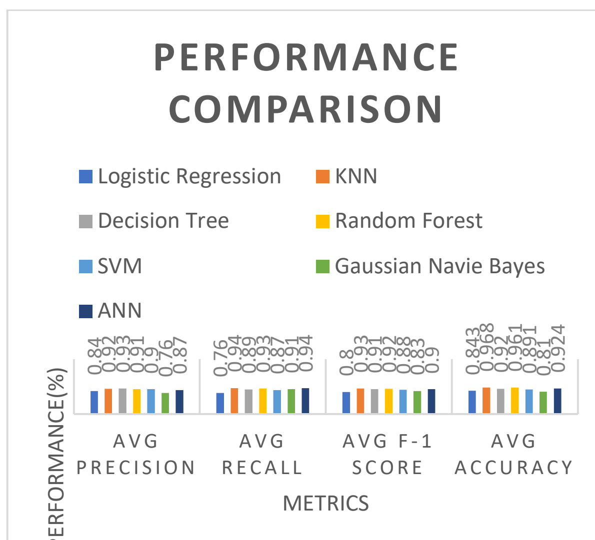


Figure 22: Performance comparison among all machine learning models

As presented in Figure 22, the results of experiments with different machine learning models are visualized. The average of all measures is computed for the observations of each arrhythmia class. Each model showed a different level of performance due to their mode of functionality. The LR model showed 84% as average precision, 76% as average recall, 80% as average F1 score, and 84.3% as average accuracy. The KNN model exhibited 92% precision, 94% recall, 93% F1-score, and 96.8% accuracy. The DT model showed 93% precision, 89% recall,and 91% F1-score with 92% accuracy. The RF model showed 91% precision, 93% recall, 92% F1-score, and 96.1% accuracy. The SVM model exhibited 90% precision,87% recall, 88% F1-score, and 89.1% accuracy. The GNB model showed 76% precision, 91% recall, 83% F1-score, and 81% accuracy. The ANN model exhibited 87% precision, 94% recall, 90% F1-score, and 92.4% accuracy. The practical implications of this work are that the proposed framework has been provisioned to exploit optimizations along with machine learning models to detect arrhythmia with multi-class classification more effectively.

5. CONCLUSION AND FUTURE WORK

In this paper, we propose a machine learning framework that exploits many classification models for detecting and classifying arrhythmia. Multiple optimizations in preprocessing, feature engineering, and hyperparameter tuning support the proposed framework. To develop an optimized machine learning approach, we proposed two algorithms known as Feature Selection and Hyperparameter Optimization (FSHO) and Learning-basedArrhythmia Detection and Classification (LbADC). We used our empirical study's benchmark dataset known as the MIT-BIH Arrhythmia dataset. The experimental results reveal that the proposed optimizations and machine learning framework could improve arrhythmia diagnosis and classification performance. The proposed optimizations of our framework achieved 96.8% accuracy in multi-class classification. This paper has certain limitations. The framework is evaluated with only one ECG data set. To generalize the findings, evaluating it with multiple datasets is important. Moreover, this work is limited to machine learning models. Deep learning models improve the learning process and performance in arrhythmia diagnosis. These limitations will be investigated in future endeavours.

REFERENCES

[1] MOHAMMED M. FARAG. (2022). A Self-Contained STFT CNN for ECG Classification and Arrhythmia Detection at the Edge. IEEE. 10(.), pp.94469-94486. [Online]. Available at: Digital Object Identifier 10.1109/ACCESS.2022.3204703 [Accessed].
[2] Ebrahimi, Z., Loni, M., Daneshtalab, M., &Gharehbaghi, A. (2020). A Review on Deep Learning Methods for ECG Arrhythmia Classification, pp.1-23.
[3] Hammad, M., Iliyasu, A. M., Subasi, A., Ho, E. S. L., & El-Latif, A. A. A. (2021). A Multitier Deep Learning Model for Arrhythmia Detection. IEEE Transactions on Instrumentation and Measurement, pp.70, 1–9.
[4] Teplitzky, B. A., McRoberts, M., & Ghanbari, H. (2020). Deep learning for comprehensive ECG annotation. Heart Rhythm, 17(5), pp.881–888.
[5] Altan, G., Allahverdi, N., & Kutlu, Y. (2018). A Multistage Deep Learning Algorithm for Detecting Arrhythmia. 2018 1st International Conference on Computer Applications & Information Security (ICCAIS), pp.1-5.
[6] Panganiban, E. B., Paglinawan, A. C., Chung, W. Y., &Paa, G. L. S. (2021). ECG diagnostic support system (EDSS): A deep learning neural network based classification system for detecting ECG abnormal rhythms from a low-powered wearable biosensors. Sensing and Bio-Sensing Research, 31, pp.1-15.
[7] Mahmud, T., Fattah, S. A., & Saquib, M. (2020). DeepArrNet: An Efficient Deep CNN Architecture for Automatic Arrhythmia Detection and Classification From Denoised ECG Beats. IEEE Access, 8, pp.1-13.
[8] Kanani, P., &Padole, M. (2020). ECG Heartbeat Arrhythmia Classification Using Time-Series Augmented Signals and Deep Learning Approach. Procedia Computer Science, 171, pp.524–531.
[9] Mousavi, S., Afghah, F., Khadem, F., & Acharya, U. R. (2021). ECG Language processing (ELP): A new technique to analyze ECG signals. Computer Methods and Programs in Biomedicine, 202, pp.1-8.
[10] Lih, O. S., Jahmunah, V., San, T. R., Ciaccio, E. J., Yamakawa, T., Tanabe, M., … Acharya, U. R. (2020). Comprehensive

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

electrocardiographic diagnosis based on deep learning, 103, pp.1-8.
[11] Cai, W., & Hu, D. (2020). QRS Complex Detection Using Novel Deep Learning Neural Networks. IEEE Access, 8, pp.97082–97089.
[12] Dang, H., Sun, M., Zhang, G., Qi, X., Zhou, X., & Chang, Q. (2019). A novel deep arrhythmia-diagnosis network for atrial fibrillation classification using electrocardiogram signals. IEEE Access,7, pp.75577–75590.
[13] Tuncer, T., Dogan, S., Pławiak, P., & Rajendra Acharya, U. (2019). Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals. Knowledge-Based Systems, pp.1-19.
[14] Hammad, M., Kandala, R. N. V. P. S., Abdelatey, A., Abdar, M., Zomorodi-Moghadam, M., Tan, R. S., … Pławiak, P. (2021). Automated detection of shockable ECG signals: A review. Information Sciences, 571, pp.580–604.
[15] Rahhal, M. M. A., Bazi, Y., AlHichri, H., Alajlan, N., Melgani, F., & Yager, R. R. (2016). Deep learning approach for active classification of electrocardiogram signals. Information Sciences, 345, pp.340–354.
[16] Petmezas, G., Haris, K., Stefanopoulos, L., Kilintzis, V., Tzavelis, A., Rogers, J. A., … Maglaveras, N. (2021). Automated Atrial Fibrillation Detection using a Hybrid CNN-LSTM Network on Imbalanced ECG Datasets. Biomedical Signal Processing and Control, 63, pp.1-9.
[17] Pourbabaee, B., Roshtkhari, M. J., & Khorasani, K. (2017). Deep Convolutional Neural Networks and Learning ECG Features for Screening Paroxysmal Atrial Fibrillation Patients. IEEE Transactions on Systems, Man, and Cybernetics: Systems, pp.1–10.
[18] Ghosh, S. K., Tripathy, R. K., Paternina, M. R. A., Arrieta, J. J., Zamora-Mendez, A., & Naik, G. R. (2020). Detection of Atrial Fibrillation from Single Lead ECG Signal Using Multirate Cosine Filter Bank and Deep Neural Network. Journal of Medical Systems, 44(6), pp.1-15.
[19] Radhakrishnan, T., Karhade, J., Ghosh, S. K., Muduli, P. R., Tripathy, R. K., & Acharya, U. R. (2021). AFCNNet: Automated detection of AF using chirplet transform and deep convolutional bidirectional long short term memory network with ECG signals. Computers in Biology and Medicine, 137, pp.1-13.
[20] Wu, Z., Lan, T., Yang, C., & Nie, Z. (2019). A Novel Method to Detect Multiple Arrhythmias Based on Time-Frequency Analysis and Convolutional Neural Networks. IEEE Access, 7, pp.170820–170830.
[21] Shadhon Chandra Mohonta, Mohammod Abdul Motin, Dinesh Kant Kumar. (2022). Electrocardiogram based arrhythmia classification using wavelet transform with deep learning model. Elsevier. 37, pp.1-8.
[22] S. Sowmya , Deepa Jose. (2022). Contemplate on ECG signals and classification of arrhythmia signals using CNN-LSTM deep learning model. Elsevier. 24, pp.1-7.
[23] Hui Wen Loh, Chui Ping Ooi , Shu Lih Oh. (2023). Deep neural network technique for automated detection of ADHD and CD using ECG signal. Elsevier. 241, pp.1-8.
[24] Sanjay Kumar, Abhishek Mallik, Akshi Kumar. (2023). Fuzz-ClustNet: Coupled fuzzy clustering and deep neural networks for Arrhythmia detection from ECG signals. Elsevier. 153, pp.1-9.
[25] Santanu Sahoo, Pratyusa Dash, B.S.P. Mishra, Sukanta Kumar Sabut. (2022). Deep learning-based system to predict cardiac arrhythmia using hybrid features of transform techniques. Elsevier. 16, pp.1-10.
[26] MD SHOFIQUL ISLAM, MD NAHIDUL ISLAM, NORAMIZA HASHIM. (2022). Detection of heart arrhythmia based on UCMFB and deep learning technique. IEEE. 10, pp.58081-58096.
[27] Mohammad Reza Mohebbian, Hamid Reza Marateb, Khan A. Wahid. (2023). Semi-supervised active transfer learning for fetal ECG arrhythmia detection. Elsevier. 3, pp.1-9.
[28] Devender Kumar, AbdolrahmanPeimankar, Kamal Sharma. (2022). Deepaware: A hybrid deep learning and context-aware heuristics-based model for atrial fibrillation detection. Elsevier. 221, pp.1-11.
[29] Devender Kumar, Sadasivan Puthusserypady, Helena Dominguez. (2023). An investigation of the contextual distribution of false positives in a deep learning-based atrial fibrillation detectio. Elsevier. 211, pp.1-12.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[30] Sannino, G., & De Pietro, G. (2018). A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. Future Generation Computer Systems, 86, pp.446–455.

[31] MIT-BIH Arrhythmia Database. Retrieved from https://physionet.org/content/mitdb/1.0.0/

[32] Task Force of the European Society of Cardiology theNorth American Society of Pacing Electrophysiology.“Heart Rate Variability”. In: Circulation 93.5 (1996),pp. 1043–1065. DOI: 10.1161/01.CIR.93.5.1043.

[33] Gokul H. et al. Gait Recovery System for Parkinson’sDisease using Machine Learning on Embedded Platforms. 2020. arXiv: 2004.05811

[34] Ellen R Girden. ANOVA: Repeated measures. 84. Sage,1992.

[35] C. J. Kelly, A. Karthikesalingam, M. Suleyman, et al.“Key challenges for delivering clinical impact withartificial intelligence”. In: BMC Med 17 195 (2019).