CardioClassify: Advanced Machine Learning for ECG-Based Arrhythmia Classification

Nikita B. Emberi, Shantanu Joshi, Aakash Kotha

1 Introduction

1.1 Objective

Arrhythmia, a group of conditions characterized by irregular electrical activity in the heart, is responsible for a significant proportion of cardiovascular-related fatalities worldwide [1]. Early and accurate detection of arrhythmias is critical for timely medical intervention and can drastically reduce the risk of severe outcomes such as cardiac arrest or stroke [2]. Despite advancements in diagnostic technologies, arrhythmia classification remains a complex problem due to the variability in electrocardiogram (ECG) signal patterns and the overlapping features among different arrhythmia types [3].

The CardioClassify project aims to address this challenge by building a comprehensive machine-learning framework capable of accurately classifying five distinct arrhythmia types from ECG signals. By leveraging both traditional machine learning techniques, such as Random Forest and XGBoost, and deep learning models like CNN 1D, this project seeks not only to achieve high classification performance but also to provide insights into the most critical ECG features contributing to the predictions. A rigorous optimization process, including parameter tuning via Optuna, further refines the CNN 1D architecture. This dual approach enables a systematic comparison of traditional and advanced methods, shedding light on the trade-offs and advantages of each in arrhythmia classification.

1.2 Problem Statement

The accurate identification of arrhythmias from ECG signals is a cornerstone of effective cardiac care. However, manual interpretation of ECG data is both labor-intensive and highly dependent on the expertise of the clinician. This process is inherently prone to human error, particularly when subtle variations in heart rhythms are involved, which may lead to misdiagnosis or delayed intervention. Traditional algorithmic approaches, while offering some automation, often fail to capture the complexity and nuances of ECG signals, further limiting their diagnostic utility.

Recent advancements in machine learning have demonstrated the potential to revolutionize arrhythmia detection. Studies such as Acharya et al. [4] highlighted the effectiveness of deep convolutional networks in achieving superior accuracy compared to traditional methods by learning directly from labeled ECG signals. Similarly, Yildirim et al. [5] demonstrated the capability of deep learning models to differentiate between multiple arrhythmia classes, supporting their applicability in clinical settings. Hannun et al. [7] and Rajpurkar et al. [8] further emphasized the viability of machine learning techniques by achieving cardiologist-level performance in arrhythmia detection. These promising results underscore the urgent need for robust machine learning frameworks, such as the one proposed in CardioClassify, to automate and enhance arrhythmia classification, ultimately aiding in faster and more accurate clinical decision-making.

GitHub Project: https://github.com/shanjos7030/STA-221-CardioClassify

1.3 Research Questions

To achieve the objectives of this study, we aim to address the following research questions:

1. How accurately can machine learning models classify five distinct arrhythmia types based on ECG signals?
2. What key features in ECG signals play the most significant role in distinguishing between different arrhythmia types?
3. How do the selection of machine learning algorithms, feature engineering techniques, and CNN 1D ablation influence the performance of arrhythmia classification models?

1.4 Project Scope and Research Motivation

- **Enhancing Diagnosis:** Early and accurate detection of arrhythmias can significantly improve patient outcomes by enabling timely interventions and reducing the risk of severe complications, such as cardiac arrest. Automating ECG classification addresses limitations of traditional manual interpretation, which is time-intensive and prone to human error.
- **Advancing ML Techniques:** This project applies both traditional machine learning models (e.g., Random Forest, XGBoost) and advanced neural networks (e.g., CNN 1D) to classify arrhythmias, leveraging recent advancements in time-series data analysis for robust ECG signal processing.
- **Optimization and Innovation:** By conducting ablation studies using Optuna, we optimize model performance through strategic parameter tuning, ensuring models are tailored for the complexity of ECG signal patterns and classification requirements.
- **Team Development:** The project offers a unique learning opportunity for the team to build expertise in biomedical data processing, feature engineering, and designing machine learning pipelines for specialized applications like ECG signal analysis.
- **Real-World Impact:** Successful classification models have significant implications for clinical diagnostics and remote patient monitoring, contributing to the development of accessible, technology-driven solutions for arrhythmia management.

2 Literature Review

In recent years, the application of machine learning techniques to ECG-based arrhythmia classification has witnessed remarkable advancements. To design an effective model, we review the following key studies that address similar challenges, highlighting methodologies, features, and innovations relevant to arrhythmia detection.

Acharya et al. [4] explored the efficacy of deep convolutional networks for arrhythmia detection using labeled ECG signals, demonstrating a significant improvement in accuracy compared to traditional methods. Their study highlights CNNs' ability to automatically extract high-level features, eliminating the need for manual feature engineering. The authors emphasize the importance of high-quality annotated datasets for achieving robust classification results, a limitation we also address in our approach.

Yildirim et al. [5] investigated deep learning architectures, particularly CNNs, for long-duration ECG signal classification. The study focused on feature extraction and signal segmentation, showing that deep learning models outperform conventional approaches in handling complex physiological data. Their results reinforce the applicability of CNNs to ECG data analysis and inspire the design of our CNN 1D framework to effectively handle intricate ECG signal patterns.

Singh and Krishnan [6] provided a comprehensive review of ECG signal feature extraction techniques, emphasizing their application in healthcare diagnostics. The authors discussed various methodologies, including time-domain, frequency-domain, and time-frequency analysis, essential for capturing the key characteristics of ECG signals. Their work highlights the role of robust feature extraction in improving the accuracy of machine learning models for arrhythmia detection. This aligns closely with our focus on optimizing feature engineering for traditional machine learning models like Random Forest and XGBoost.

Hannun et al. [7] expanded the scope of machine learning in arrhythmia detection by using a deep neural network trained on an extensive dataset of over 91,000 single-lead ECG records. Their approach achieved

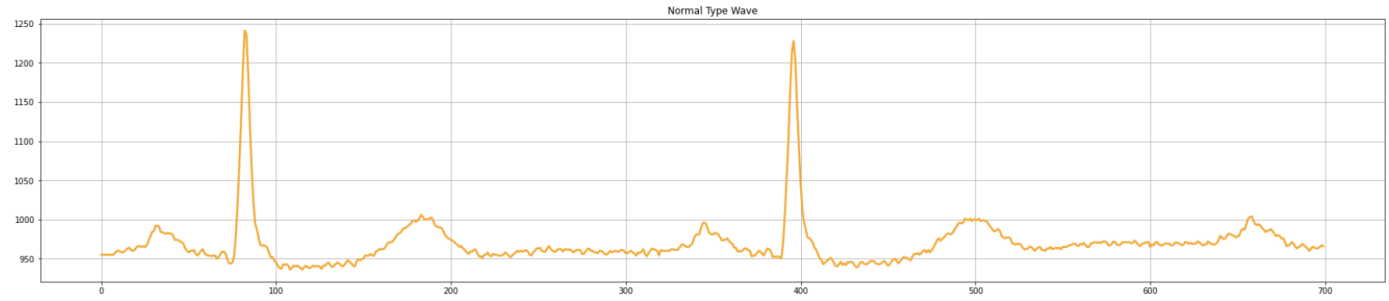

Figure 1: First 700 samples for the first patient in the dataset.

cardiologist-level performance in classifying arrhythmias, underlining the role of large datasets and robust evaluation metrics. The emphasis on dataset diversity and scalability directly informs our efforts to optimize classification performance through comprehensive data preprocessing and feature extraction.

Zhang et al. [9] examined the effectiveness of combining CNNs and RNNs for arrhythmia detection, leveraging CNNs for spatial feature extraction and RNNs for temporal sequence modeling. Their hybrid approach achieved superior accuracy compared to standalone models, demonstrating the benefits of combining spatial and temporal feature extraction techniques. This study underscores the potential of hybrid methodologies, offering insights into future extensions of our work, while complementing our current exploration of CNN 1D performance through ablation studies.

3 Dataset

3.1 Dataset Description

For this project, we utilize the MIT-BIH Arrhythmia Database [10], a widely recognized collection of 48 contiguous ECG (electrocardiogram) excerpts obtained from 47 subjects. Each excerpt is approximately 30 minutes long and consists of two-channel recordings that comprehensively capture the heart's electrical activity. These recordings provide detailed waveforms, revealing different stages of the cardiac cycle and offering insights into both normal and abnormal rhythms.

The dataset includes recordings from inpatients (around 60%) and outpatients (approximately 40%) at Boston's Beth Israel Hospital, covering an age range of 23 to 89 years. Of the 47 subjects, 25 (53%) are male. This diversity ensures a balanced representation of both common and rare arrhythmias, making the dataset invaluable for arrhythmia research.

Each recording is digitized at a sampling frequency of 360 Hz, meaning 360 samples are captured per second for each channel. This high-resolution data allows for precise analysis of the electrical signals generated by the heart. However, the version of this dataset used in this project, obtained from Kaggle, has been resampled to a lower frequency of 125 Hz for computational efficiency and specific modeling purposes.

The recordings are annotated by expert cardiologists, with consensus achieved after independent review by two or more experts. Each heartbeat is labeled into five distinct arrhythmia classes: Normal (N), Left Bundle Branch Block (L), Right Bundle Branch Block (R), Premature Ventricular Contraction (V), and Atrial Premature Contraction (A). These annotations label approximately 110,000 heartbeats, providing a clinically validated reference for arrhythmia classification and model development.

Figure 1 illustrates the first 700 samples of an ECG waveform, showcasing the high granularity of the dataset and the clarity of the electrical signal, critical for arrhythmia classification.

3.2 Segmentation

To prepare the data for analysis, each ECG signal was segmented into overlapping windows of 2000 samples, with a 1000-sample overlap. Each segment typically contains multiple heartbeats, and the annotation for the segment is determined based on the individual beat annotations within it. If all beats in the segment are of type N (Normal), the segment is labeled as Normal. Conversely, if even a single beat in the segment is of

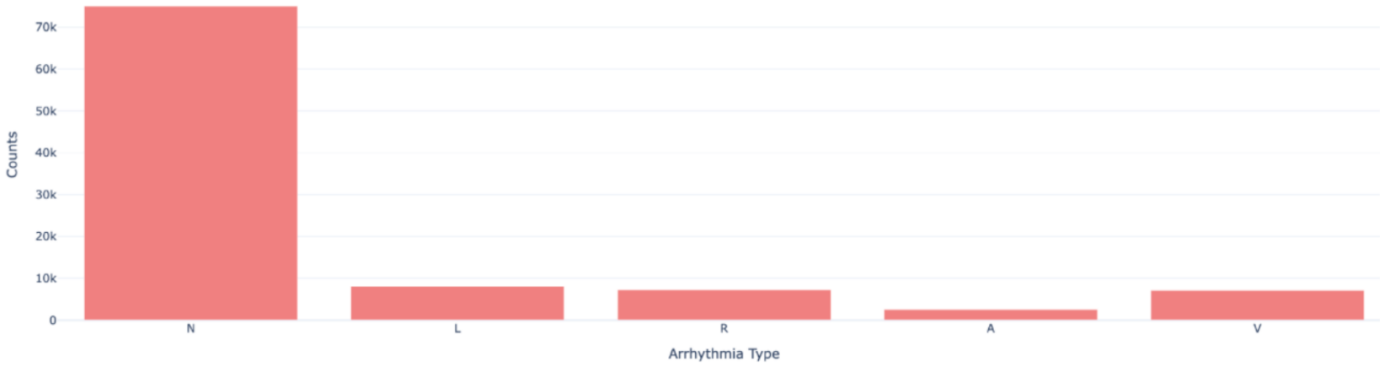

Figure 2: Bar plot illustrating class distribution across five arrhythmia categories.

type A (Atrial premature beat), the segment is labeled as having an Atrial premature beat. Importantly, the dataset specifies that an individual cannot have beats with different types of abnormal rhythms within a single recording, eliminating the possibility of annotation conflicts. This segmentation ensures the inclusion of critical features like the QRS complex, P-wave, and T-wave, enabling the model to learn from both local and contextual signal patterns.

3.3 Exploratory Data Analysis

Following the segmentation of ECG signals into overlapping windows of 2000 samples, an analysis of the class distribution was conducted. Table 1 presents the distribution of the five heartbeat classes across the dataset. The class distribution in Figure 2 highlights a significant imbalance, with the *Normal* class dominating the dataset at 75%, and the *Atrial Premature Contraction* class accounting for only 2.5%.

Table 1: Distribution of ECG Segments into 5 Classes of Arrhythmia

Type	**Description**	**Number of Samples**
N	Normal: Regular heartbeats without any rhythm abnormality.	75,052
L	Left bundle branch block: Beats showing a blockage in the left bundle branch, affecting the left side of the heart.	8,075
R	Right bundle branch block: Beats with a blockage in the right bundle branch, affecting the heart's right side.	7,259
A	Atrial premature beat: Early beats originating in the atria (upper chambers of the heart).	2,546
V	Ventricular premature beat: Early beats originating in the ventricles (lower chambers of the heart).	7,130

A detailed visual inspection of the ECG waveforms revealed distinct patterns corresponding to different heartbeat classes, as well as the presence of noise artifacts in certain segments. Figure 1 illustrates the clear QRS complex of a *Normal* beat, emphasizing the importance of preserving these features during preprocessing. However, noise, including high-frequency interference, was observed in some signals, necessitating robust noise removal techniques.

This inspection informed the design of preprocessing steps, which included balancing the dataset to address class imbalance, applying noise removal filters to enhance signal quality, and normalization to ensure uniform amplitude scaling across segments.

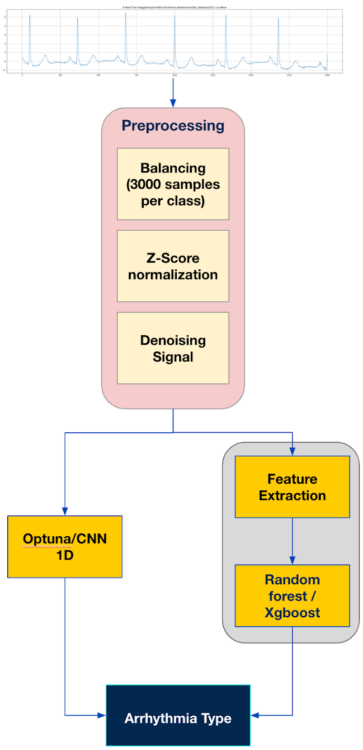

Figure 3: ML Pipeline for preprocessing, feature extraction, and classical and deep learning model training.

4 Methodology

For classical machine learning, we extracted features initially and fed them to Random Forest and XGBoost. For Deep Learning, we perform an ablation study using 1D CNN. Figure 3 illustrates the complete pipeline for preprocessing, feature extraction, and the training of both classical and deep learning models.

4.1 Preprocessing

A rigorous three-step preprocessing pipeline was implemented to transform the raw ECG data into a structured and balanced dataset for machine learning:

4.1.1 Class Balancing

To address this imbalance, oversampling techniques were applied to the minority classes (L, R, V, A), increasing their representation to 3,000 samples each. This resulted in a balanced dataset of 15,000 samples, ensuring equitable representation across all heartbeat types and mitigating potential class bias in the model training.

4.1.2 Denoising

A wavelet-based denoising approach using Symlet wavelet ('sym4')[1] effectively removed noise from ECG signals. Applying discrete wavelet transform (DWT) and soft thresholding preserved essential waveform features while filtering out high-frequency interference and baseline wander.

4.1.3 Z-Score Normalization

Signals were standardized by subtracting the mean and dividing by the standard deviation, ensuring uniform amplitude scaling and facilitating direct comparison across different ECG segments.

[1] https://wavelets.pybytes.com/wavelet/sym4/

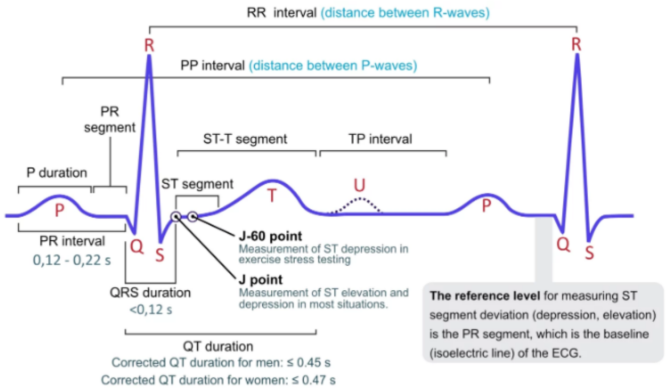

Figure 4: Features as shown for two consecutive ECG beats.

4.2 Classical Machine Learning (Random Forest and XGBoost)

4.2.1 Experiments

For the classical machine learning models, the preprocessed ECG signals are first passed through the `ecg_peaks` and `ecg_delineate` functions from the NeuroKit library[2], which extract the key peaks and offsets in the ECG signal. Specifically, the P peaks, Q peaks, S peaks, T peaks, P onsets, P offsets, T onsets, R onsets, and R offsets are identified. These extracted features are then used to derive eight time-domain features, as outlined in Table 2. These features, which capture critical physiological intervals, include the PP interval, SS interval, PR segment, QRS duration, ST segment, P duration, ST-T segment, and QQ interval. Figure 4 illustrates the features for two consecutive ECG beats. The Random Forest(100 estimators, Gini impurity) and XGBoost (learning-rate=0.3, maximum-depth=6)models were implemented using the `scikit-learn` and `XGBoost` libraries[3], respectively, with default parameters.

Table 2: Formula for features derivation from the values extracted using NeuroKit

Feature	Formula
PP Interval	diff(P peaks)
SS Interval	diff(S peaks)
PR Segment	(R onset - P offset)
QRS Duration	(R offset - R onset)
ST Segment	(T onset - S peaks)
P Duration	(P offset - P onset)
ST-T Segment	(T offset - S peaks)
QQ Interval	diff(Q peaks)

4.2.2 Assumptions

In this study, it was assumed that the extracted features from ECG signals (e.g., PR Interval, QRS Duration, Ventricular Rate, QTc Bazett) were consistent and representative across the entire dataset, ensuring stable relative importance across folds during cross-validation. The optimized Random Forest and XGBoost models were presumed to generalize effectively to unseen data, as validated through k-fold ($k = 5$) cross-validation, minimizing risks of overfitting to specific folds or subsets. Additionally, the feature engineering process, including preprocessing and scaling, was assumed to be invariant and free from dataset-specific biases, ensuring consistency across all experiments and enabling a valid comparison of model performance.

[2] https://neuropsychology.github.io/NeuroKit/
[3] https://scikit-learn.org/stable/, https://xgboost.readthedocs.io/en/stable/

4.3 Deep Learning (1D CNN)

4.3.1 CNN 1D Architecture Optimization with Optuna

The proposed method leverages a detailed ablation study of the CNN 1D architecture, optimized using Optuna, a state-of-the-art hyperparameter optimization framework. The primary goal of this study was to systematically explore the impact of various architectural and training parameters on the performance of the model for ECG-based arrhythmia classification.

Optuna[4] is an open-source framework designed for efficient and automated hyperparameter optimization. It employs advanced techniques such as Tree-structured Parzen Estimators (TPE) and Multi-Objective Optimization to intelligently search the parameter space, dynamically prioritizing promising configurations and pruning poor-performing trials to save computational resources. Optuna's ability to adaptively refine the search space during optimization makes it highly effective for deep learning models like CNNs, where the parameter space can be extensive.

Given the complexity of CNN 1D models and the large number of hyperparameters to optimize, traditional methods like grid search or random search would have been computationally expensive and inefficient. Optuna was chosen to:

- Dynamically explore a wide range of hyperparameters critical to the CNN 1D architecture.
- Prune suboptimal trials early, ensuring faster convergence to optimal configurations.
- Efficiently handle two experimental setups with different levels of parameter constraints.

4.3.2 Experiments

The CNN 1D ablation study was conducted in two experimental phases, each leveraging Optuna to identify optimal hyperparameter configurations. In the initial experimental phase, a comprehensive hyperparameter optimization was conducted using Optuna to explore the intricate relationships between various model parameters. The study systematically investigated the impact of Conv1D layer configurations, ranging from one to three layers, with filter counts progressively scaled from 16 to 128. Kernel sizes were examined across a spectrum from 3 to 15, allowing for nuanced feature extraction capabilities. Dropout regularization was explored with rates varying between 0.2 and 0.5 to assess its effect on model generalization. The training dynamics were further probed by varying batch sizes between 32 and 128, and learning rates spanning from 10^{-5} to 10^{-1}. The training duration was also flexibly adjusted, with epoch counts ranging from 10 to 60, providing a comprehensive exploration of the model's learning potential.

Building upon the insights from the first experiment, the second phase implemented a more focused optimization strategy by fixing certain foundational parameters. A consistent batch size of 36 and a learning rate of 0.001 were maintained, with a fixed training duration of 60 epochs to ensure stable model convergence. Despite these fixed parameters, the study maintained a comprehensive approach to architectural optimization. The remaining parameters—Number of Conv1D Layers, Number of Filters, Kernel Sizes, and Dropout Rate—were systematically optimized within the same ranges as established in Experiment 1. This approach allowed for a more targeted investigation of critical architectural elements while preserving the breadth of the initial exploratory phase. By maintaining core training parameters constant, the experiment could more precisely isolate and understand the impact of architectural variations on model performance.

4.3.3 Assumptions

In the proposed method, the following assumptions were maintained to ensure consistency and validity throughout the ablation study and the performance evaluation of the CNN 1D architecture:

- **Stationarity of Hyperparameter Impact:** It was assumed that the impact of each hyperparameter on the CNN 1D model's performance was consistent across all folds of cross-validation and that no external factors (e.g., hardware variability or software configurations) affected the optimization process.

[4] https://optuna.org/

- **Fixed Experimental Conditions:** During Experiment 2, the batch size, learning rate, and number of epochs were fixed to control for variability, assuming that these parameters did not significantly impact the exploration of other hyperparameters (e.g., number of layers, filters, kernel size, and dropout rate). These fixed parameters were chosen based on prior knowledge and preliminary experiments, assuming these values provided a stable foundation for further tuning.
- **Model Generalizability:** The best-performing architecture identified in each experiment was assumed to generalize well across unseen data, as validated through k-fold cross-validation. This assumption was critical to ensure the reliability of the optimization process.
- **Consistency and Independence of Optimization Framework:** It was assumed that Optuna's optimization and pruning mechanisms functioned as intended, dynamically adjusting the search space and terminating poorly performing trials without bias, while maintaining the independence of each trial. This ensured that the outcomes of one trial did not influence subsequent trials, validating the stochastic and unbiased nature of the optimization process.

5 Results

To ensure the robustness and generalizability of the models, 5-fold cross-validation was employed for all the models. The Random Forest classifier achieved an accuracy of 72%, with class-wise performance varying across the five classes. The highest F1-score was observed for class 1 (arrhythmias), and the weighted F1-score for the model was 0.7175, with a precision of 0.7171 and recall of 0.7219. The Random Forest model demonstrated strong discrimination ability with an AUROC of 0.9236 and an AUPRC of 0.8041. On the other hand, the XGBoost classifier achieved a slightly lower accuracy of 70%. Despite similar performance for class 1, the model's overall weighted F1-score was 0.6939, with a precision of 0.6937 and recall of 0.6978. The XGBoost model exhibited an AUROC of 0.9088 and an AUPRC of 0.7603, indicating competitive but slightly lower performance compared to the Random Forest classifier.

For both experiments using CNN 1D, Optuna conducted 20 trials to identify the best-performing architecture. Among the top three architectures listed for each experiment, Architecture 1 was determined to be the best-performing configuration (based on validation data accuracy). The top configuration from each experiment was further validated using 5-fold cross-validation to ensure the model's generalizability and robustness. The results of the ablation study on the CNN 1D architecture, showing the top three architectures from each experiment, are presented in Table 3.

Table 3: Top Architectures from Each Experiment

Experiment	Architecture	Conv1D Layers	Filters	Kernel Sizes	Dropout Rate	Batch Size	Learning Rate	Number of Epochs	Validation Data Accuracy
Experiment 1	Architecture 1	2	[80, 64]	[5, 7]	0.3201	96	0.00116	30	93.87%
Experiment 1	Architecture 2	2	[80, 96]	[13, 5]	0.2981	96	0.0007	25	93.17%
Experiment 1	Architecture 3	2	[64, 48]	[5, 7]	0.4274	32	0.0001	30	92.5%
Experiment 2	Architecture 1	3	[64, 64, 96]	[9, 3, 7]	0.2036	36 (Fixed)	0.001 (Fixed)	60 (Fixed)	96.71%
Experiment 2	Architecture 2	3	[32, 128, 32]	[9, 11, 5]	0.2145	36 (Fixed)	0.001 (Fixed)	60 (Fixed)	96.2%
Experiment 2	Architecture 3	3	[48, 80, 128]	[11, 11, 11]	0.2274	36 (Fixed)	0.001 (Fixed)	60 (Fixed)	95.45%

The 1D CNN model from Experiment 1, where all hyperparameters were optimized using Optuna, achieved an overall accuracy of 97% on test data. The model exhibited a weighted F1-score of 0.9714, with precision and recall values of 0.9714 and 0.9713, respectively. It demonstrated excellent discrimination ability, achieving an AUROC of 0.9983 and an AUPRC of 0.9940. The class-wise accuracies for this experiment were as follows: Class 0 - 94.52%, Class 1 - 99.21%, Class 2 - 98.92%, Class 3 - 94.62%, and Class 4 - 98.34%. Architecture 1 defines as the best-performing model for this experiment, and the test loss and accuracy plots (Figure 5) are based on the best fold from this architecture.

In contrast, Experiment 2, where certain parameters were fixed (as described in the methodology) while the remaining hyperparameters were randomized, achieved a slightly higher overall accuracy of 98% on test data. This model reported a weighted F1-score of 0.9806, with precision and recall values of 0.9801 and 0.9803, respectively. The class-wise accuracies for this experiment were: Class 0 - 95.29%, Class 1 - 99.21%, Class 2 - 99.51%, Class 3 - 97.50%, and Class 4 - 98.78%. It demonstrated equivalent excellent discrimination ability,

achieving an AUROC of 0.9989 and an AUPRC of 0.9962, reflecting a marginally better performance compared to Experiment 1. Here, Architecture 1 again is defined as the best-performing model for this experiment, and the test loss and accuracy plots (Figure 5) are based on the best fold from this architecture.

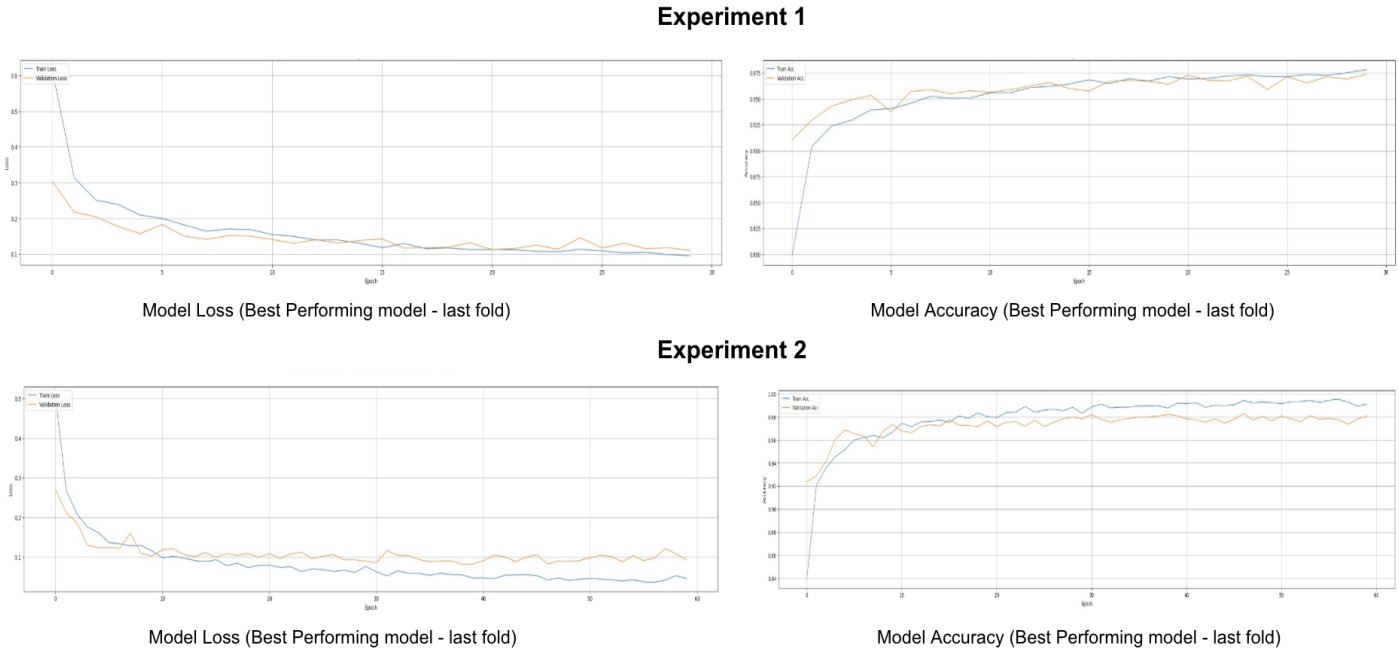

Figure 5: Plots for training and validation loss and accuracy over epochs for the best-performing architecture in Experiment 1 and Experiment 2 CNN 1D.

The plots depicted in Figure 5 show the progression of train and validation loss, as well as train and validation accuracy, over epochs for the best fold of the best-performing architecture (Architecture 1) in Experiment 1 and Experiment 2 for CNN 1D. Both loss and accuracy metrics demonstrate steady improvement and convergence, with losses stabilizing and remaining closely aligned, indicating effective learning and strong generalization. Similarly, accuracy metrics show rapid improvement during initial learning, followed by stabilization with minimal deviation, underscoring a balanced, robust model with no significant overfitting or divergence.

Table 4: Performance for Classical (Random Forest and XGBoost) and CNN 1D Ablation Study

Model	Accuracy	F1 Score	Precision	Recall	AUROC	AUPRC
Random Forest	0.72	0.7175	0.7171	0.7219	0.9236	0.8041
XGBoost	0.70	0.6939	0.6937	0.6978	0.9088	0.7603
CNN 1D Experiment 1	0.97	0.9714	0.9714	0.9713	0.9983	0.9940
CNN 1D Experiment 2	**0.98**	**0.9806**	**0.9801**	**0.9803**	**0.9989**	**0.9962**

Table 4 shows the result for Classical (Random Forest and XGBoost) and CNN 1D methods. The results provide an answer to the first research question, which investigates how accurately machine learning models can classify five distinct arrhythmia types based on ECG signals. As observed from the results, the CNN 1D model outperforms the Random Forest and XGBoost models in terms of all metrics. Using these, we addressed our third research question which describes the influence of machine learning algorithm selection and CNN 1D ablation on the performance of arrhythmia classification models.

6 Discussion

For discussion, we focus on the Random Forest model as it outperforms XGBoost in terms of accuracy to discuss the importance of the features. Feature importance analysis of the Random Forest model reveals that certain

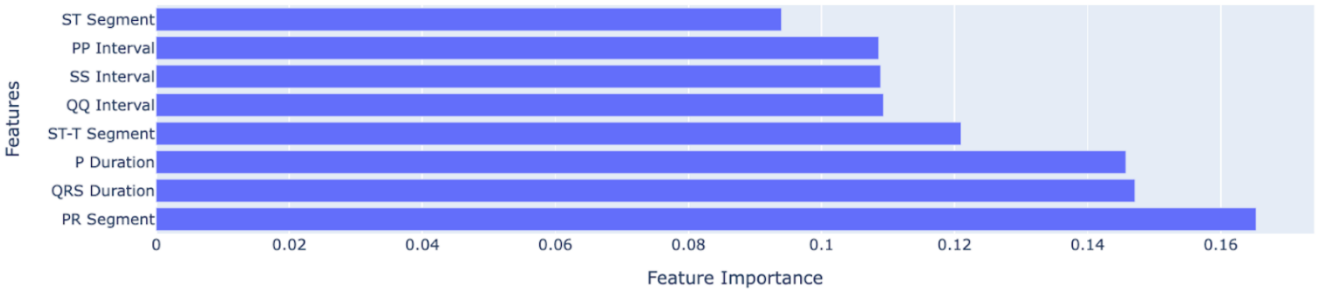

Figure 6: Feature importance for the Random Forest model.

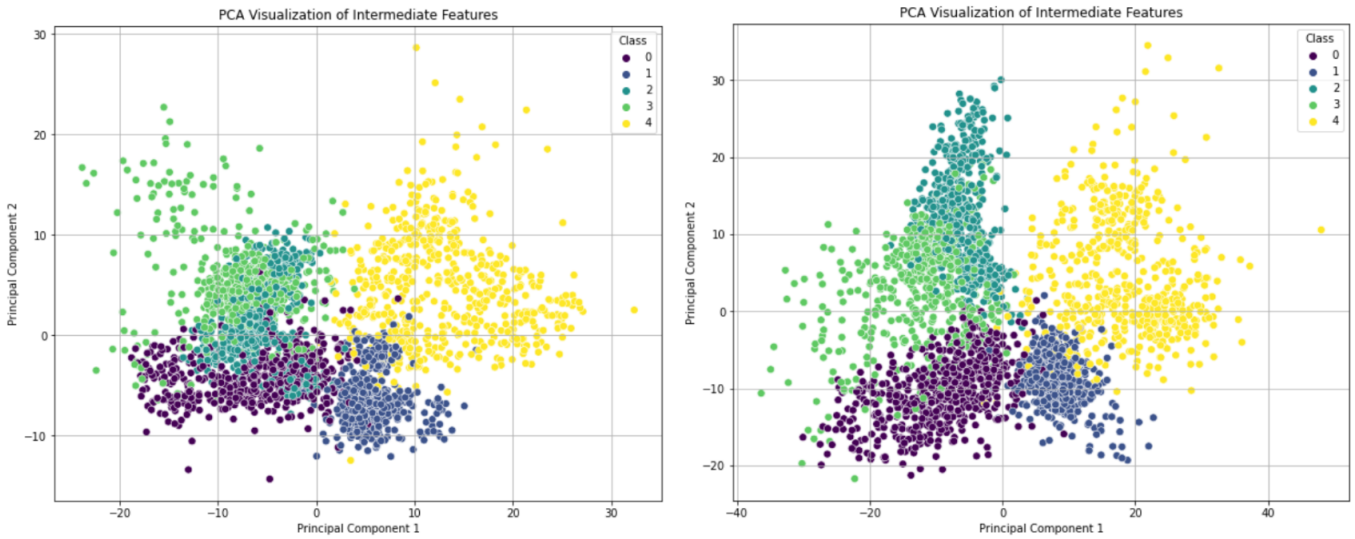

Figure 7: PCA Visualization of Intermediate Features from Experiment 1 (Left) and Experiment 2 (Right).

ECG-derived features contribute significantly to its performance. The PR Segment (0.1653) has the highest importance, followed by the QRS Duration (0.1471) and P Duration (0.1458). These features are crucial for capturing key temporal and morphological aspects of the ECG signal. The feature importance for all extracted features is shown in Figure 6. The above analysis addresses the second research question, which investigates which key features in ECG signals play the most significant role in distinguishing between different arrhythmia types.

The plots depicted in Figure 7 illustrate the **PCA (Principal Component Analysis)** visualization of the intermediate features extracted from the last layer of the CNN 1D model for the best-performing architecture in both experiments. This analysis helped us evaluate how well the models distinguished between the different classes in a reduced two-dimensional feature space.

The PCA visualization revealed distinct clusters for each class (0–4), showcasing the CNN 1D model's ability to learn discriminative features for arrhythmia classification. Class 4 (yellow) forms a clearly separated cluster, while neighboring classes, particularly 1, 2, and 3, show some overlap due to shared characteristics in ECG signal data, though their central regions remain distinct. Compared to the first experiment, the clusters in the second experiment are more compact and separable, reflecting improved feature learning and model stability with the fixed parameter setup. This visualization reinforces the effectiveness of the CNN 1D architecture in extracting meaningful features, especially under the refined optimization setup of the second experiment. The reduced overlap and clearer separability of classes compared to the first experiment highlight the impact of fixed parameters in guiding the model toward more robust feature representations.

By leveraging Optuna's advanced optimization techniques, the ablation study provided critical insights into the relationship between CNN 1D hyperparameters and model performance, leading to a well-optimized architecture for arrhythmia classification. The proposed method introduces significant methodological innovations

in ECG-based arrhythmia classification through a comprehensive approach to architectural exploration and hyperparameter optimization. By conducting a systematic ablation study on CNN 1D architectures, the research moved beyond traditional static design approaches, meticulously examining the impact of critical architectural components such as layer count, filter configurations, and kernel sizes. This exploration was conducted through two experimental setups that provided nuanced insights into hyperparameter influences. Optuna's dynamic and adaptive pruning mechanism enabled more efficient computational resource allocation, allowing for a broader and deeper exploration of hyperparameter configurations within realistic time constraints. These methodological innovations not only enhanced arrhythmia classification performance but also established a robust, generalizable framework for applying systematic hyperparameter optimization to time-series classification.

The results of our study demonstrate that CNN significantly outperforms classical machine learning models, such as Random Forest and XGBoost, in detecting arrhythmia from ECG data. This can be attributed to the inherent limitations of feature extraction in classical approaches, which rely solely on handcrafted features. While these features capture general trends, they may overlook subtle but critical details within the heartbeat signals. In contrast, the CNN effectively learns intricate patterns and temporal dependencies directly from the raw ECG data, enabling it to identify nuanced variations that are indicative of arrhythmias. This ability to automatically extract and prioritize relevant features from the data contributes to the superior accuracy of the CNN model compared to the classical machine learning methods.

7 Conclusion

In conclusion, the study highlights the superior performance of the CNN model in detecting arrhythmia compared to classical machine learning approaches like Random Forest and XGBoost. The CNN's ability to automatically extract intricate patterns and temporal dependencies from raw ECG data allows it to capture subtle variations that handcrafted features in classical models often overlook. This capability underscores the potential of deep learning in analyzing complex physiological signals. Looking ahead, the application of CNNs to 1D signals like ECG holds promising potential for advancing the early detection and diagnosis of various diseases like Silent Heart Attacks, paving the way for more accurate, non-invasive, and real-time healthcare solutions.

References

[1] Albert CM, Stevenson WG. The Future of Arrhythmias and Electrophysiology. Circulation. 2016 Jun 21;133(25):2687-96. doi: 10.1161/CIRCULATIONAHA.116.023519.

[2] Pham BT, Le PT, Tai TC, Hsu YC, Li YH, Wang JC. Electrocardiogram Heartbeat Classification for Arrhythmias and Myocardial Infarction. Sensors (Basel). 2023 Mar 9;23(6):2993. doi: 10.3390/s23062993.

[3] 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.

[4] Acharya, U. R., Fujita, H., Lih, O. S., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. Information Sciences, 405, 81–90. https://doi.org/10.1016/j.ins.2017.04.012.

[5] Yildirim, "O., Talo, M., Ay, B., Baloglu, U. B., Aydin, G., & Acharya, U. R. (2018). Arrhythmia detection using deep convolutional neural network with long duration ECG signals. Computers in Biology and Medicine, 102, 411–420. https://doi.org/10.1016/j.compbiomed.2018.09.009.

[6] Singh, A. K., & Krishnan, S. (2020). Comprehensive review of ECG signal feature extraction methods for healthcare diagnostics. Biomedical Signal Processing and Control, 62, 102074. https://doi.org/10.1016/j.bspc.2020.102074.

[7] Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine, 25(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3.

[8] Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., Ng, A. Y., & Irvin, J. (2017). Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. arXiv preprint. https://arxiv.org/abs/1707.01836.

[9] Zhang, Y., Pan, Y., Li, H., & Song, S. (2020). A hybrid deep learning model for arrhythmia detection based on spatial and temporal features. Biomedical Signal Processing and Control, 58, 101870. https://doi.org/10.1016/j.bspc.2019.101870.

[10] Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng Med Biol Mag. 2001;20(3):45-50. doi: 10.1109/51.932724.