A Comparative Study of Random Forest for ECG Arrhythmia Classification

Doan Ngoc Linh
Student ID: 22BA13190

I. INTRODUCTION

Electrocardiogram (ECG) signals play an important role in monitoring and diagnosing cardiovascular diseases because they directly reflect the heart's electrical activity over time. Early detection of cardiac rhythm abnormalities can significantly reduce the risk of serious complications, including heart failure and stroke. However, in clinical practice, ECG interpretation often requires specialized expertise and is time-consuming, especially when the number of patients is large or when continuous monitoring is needed. In this study, we employ a Random Forest model for ECG heartbeat classification using the MIT-BIH Arrhythmia Dataset. The obtained results are compared with those reported in previous studies to clarify the suitability of a classical machine learning approach for this task.

II. PROBLEM STATEMENT

This report addresses a multi-class classification problem for electrocardiogram (ECG) signals at the heartbeat level. Each input sample is a processed ECG beat represented as a fixed-length feature vector consisting of 187 numerical values. The objective is to predict the corresponding class label among five heartbeat categories. The goal is to develop a model that can learn the relationship between ECG features and class labels and generalize well to unseen ECG samples.

III. DATA

A. Dataset Description

The dataset used in this study is the MIT-BIH Arrhythmia Dataset, a widely adopted benchmark for ECG signal classification. It contains preprocessed ECG heartbeat segments, where each sample corresponds to a single heartbeat. Representing the data at the beat level allows the model to focus on morphological patterns of individual heartbeats, which is essential for distinguishing different arrhythmia types.

B. Data Structure

Each ECG sample is represented as a fixed-length feature vector of 187 numerical values corresponding to normalized ECG amplitude values over time. The last column represents the class label. The dataset is divided into a training set containing 87,554 samples and a test set containing 21,892 samples. Using a separate test set enables unbiased evaluation of the model's generalization performance.

C. Class Distribution

A key characteristic of this dataset is severe class imbalance, where normal heartbeats dominate the dataset. This imbalance may bias models toward the majority class. Therefore, in addition to accuracy, class-wise metrics such as precision, recall, and F1-score are reported, along with macro F1-score to better evaluate minority class performance.

IV. METHOD

A. Random Forest Model

A Random Forest classifier is employed to solve the ECG heartbeat classification problem. Random Forest is an ensemble learning method that combines multiple decision trees to reduce variance and improve generalization. Each tree is trained using a bootstrap sample of the training data, while a random subset of features is selected at each split. The final prediction is obtained by majority voting across all trees. Class weighting is applied during training to mitigate the impact of class imbalance.

B. Hyperparameter Experiments

To analyze the influence of hyperparameters, several configurations were evaluated. The investigated hyperparameters include the number of trees ($n\_estimators$) and the maximum tree depth (max_depth).

TABLE I
PERFORMANCE COMPARISON OF RANDOM FOREST MODELS

Experiment	n_estimators	max_depth	Accuracy	Macro F1
1	100	None	0.9730	0.86
2	200	None	0.9734	0.86
3	200	20	0.9719	0.85
4	300	30	0.9729	0.86

V. RESULTS AND DISCUSSION

VI. RESULTS AND DISCUSSION

The experimental results of the Random Forest classifier under different hyperparameter configurations are presented and analyzed in this section. All models were evaluated on an independent test set consisting of 21,892 ECG heartbeat samples. Test accuracy and macro F1-score were selected as the primary evaluation metrics to assess both overall performance and robustness under severe class imbalance.

A. Overall Classification Performance

Table II summarizes the overall test accuracy and macro F1-score achieved by the Random Forest models with different values of *n_estimators* and *max_depth*.

TABLE II
PERFORMANCE COMPARISON OF RANDOM FOREST MODELS

Experiment	n_estimators	max_depth	Accuracy	Macro F1
Exp-1	100	None	0.9730	0.86
Exp-2	200	None	0.9734	0.86
Exp-3	200	20	0.9719	0.85
Exp-4	300	30	0.9729	0.86

All configurations achieved high classification performance, with test accuracy ranging from 97.19% to 97.34%. Among the tested settings, Exp-2 (*n_estimators* = 200, *max_depth* = None) achieved the highest test accuracy of 0.9734 while maintaining a stable macro F1-score of approximately 0.86.

Increasing the number of trees from 100 to 200 resulted in a slight improvement in accuracy, indicating improved ensemble stability. However, further increasing the number of trees to 300 did not yield significant additional gains. Similarly, restricting the maximum depth of the trees led to a small reduction in both accuracy and macro F1-score, highlighting the trade-off between model complexity and generalization capability.

B. Class-wise Performance Analysis

To provide deeper insight into the classification behavior of the model, the confusion matrix of the best-performing configuration (Exp-2) is reported in Table III. This analysis helps illustrate how the model performs across different heartbeat classes.

TABLE III
CONFUSION MATRIX FOR THE BEST RANDOM FOREST MODEL (EXP-2)

True / Pred	0	1	2	3	4
0	18101	10	3	3	1
1	224	331	0	0	1
2	168	0	1261	16	3
3	53	0	13	96	0
4	92	0	3	0	1513

The results indicate excellent classification performance for the majority class (Class 0), with near-perfect recall. More importantly, the Random Forest model demonstrates reasonable discrimination capability for minority arrhythmia classes, as reflected by consistent macro F1-scores across experiments. This confirms the effectiveness of class weighting in mitigating the impact of severe class imbalance.

C. Comparison with the Original Study

When compared with the results reported by Kachuee et al., a noticeable performance improvement can be observed. The deep convolutional neural network (CNN) proposed in the original study achieved an overall classification accuracy of approximately 93.4%, whereas the Random Forest model in this work achieved an accuracy of approximately 97.3%.

This performance gain can be attributed to several factors. First, Random Forest is a robust ensemble learning method that reduces variance by aggregating multiple decision trees. Given the relatively large size of the MIT-BIH Arrhythmia Dataset and the structured feature representation used in this study, Random Forest is able to effectively capture nonlinear relationships without requiring complex representation learning.

Second, explicit class weighting was applied during training to address class imbalance, leading to improved classification performance for minority arrhythmia classes and more balanced macro F1-scores. In contrast, deep learning models trained on raw ECG signals are more sensitive to data imbalance and require careful architectural and optimization design.

Third, the CNN model in the original study relies heavily on convolutional architecture design and hyperparameter tuning. Random Forest, on the other hand, requires fewer hyperparameters, is less sensitive to training instability, and demonstrates strong performance even with limited tuning.

It should be noted that this comparison is not strictly equivalent, as the two models were trained under different experimental setups and data representations. While deep learning approaches remain advantageous for end-to-end learning from raw signals and may generalize better in more complex scenarios, the results of this study demonstrate that Random Forest provides a competitive and computationally efficient alternative for ECG heartbeat classification.