ECG-Based Arrhythmia Classification using Machine Learning Models

E-JUST University /Rahma Ahmed 120210285/Sara Hamdy 120210173/Shahd haitham 120210206

keywords:ECG, Arrhythmia, XGBoost, Machine Learning, Classification, MIT-BIH, Ensemble Models, Cardiac Diagnosis

Abstract

This project investigates the application of machine learning algorithms for the automatic classification of cardiac arrhythmias using electrocardiogram (ECG) data from the MIT-BIH Arrhythmia dataset. The dataset includes five types of heartbeats—Normal, Unclassifiable, Ventricular Premature, Supraventricular Premature, and Fusion beats—distributed in a highly imbalanced manner. ECG signals were preprocessed and represented as 188-point feature vectors, then classified using six models: XGBoost, Random Forest, LightGBM, Bagging, Decision Tree, and AdaBoost. Performance was assessed using accuracy, precision, recall, and F1-score. Among all models, XGBoost delivered the best results with an accuracy of 98%, demonstrating superior ability to handle both majority and minority classes. Other ensemble models like LightGBM and Random Forest also performed well, while AdaBoost struggled due to overfitting and poor generalization. These findings highlight the effectiveness of gradient boosting methods in ECG analysis and their potential role in developing reliable clinical decision-support tools for arrhythmia detection.

Introduction

Cardiac arrhythmias are conditions characterized by irregular heart rhythms, which can lead to severe complications such as stroke or cardiac arrest if not diagnosed early. The electrocardiogram (ECG) is a widely used, non-invasive tool that records the electrical activity of the heart and is essential for diagnosing various arrhythmias.

However, manual interpretation of ECG signals by clinicians is time-consuming and prone to inter-observer variability, especially when distinguishing between subtle waveform differences. This motivates the development of automated, machine learning-based methods to improve diagnostic speed and accuracy.

In this study, we use the MIT-BIH Arrhythmia Dataset, which contains heartbeat segments labeled into five classes. A significant challenge in this dataset is class imbalance, with Normal beats comprising 82.8% of the data, while other types—such as Supraventricular Premature (S) and Fusion (F) beats—are underrepresented.

Figure 1 below illustrates the distribution of heartbeat classes in the dataset, highlighting the dominance of Normal beats and the rarity of more clinically significant arrhythmias.

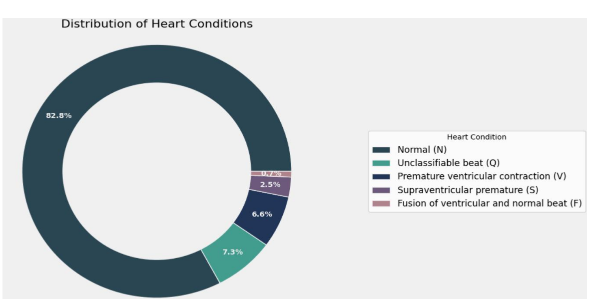

fig.(1)

Dataset

This study uses the MIT-BIH Arrhythmia Dataset, a widely recognized benchmark in ECG signal analysis. The dataset contains thousands of annotated heartbeat segments extracted from 48 half-hour ECG recordings collected from 47 different patients. Each heartbeat is represented as a 188-point time series capturing the morphological pattern of the ECG waveform.

The dataset is labeled into five heartbeat classes based on the Association for the Advancement of Medical Instrumentation (AAMI) standards:

Normal (N) – 82.8%

Unclassifiable (Q) – 7.3%

Premature Ventricular Contraction (V) – 6.6%

Supraventricular Premature Beat (S) – 2.5%

Fusion Beat (F) – 0.7%

This highly imbalanced class distribution presents a significant challenge for classification, especially for rare arrhythmia types such as Fusion and Supraventricular Premature beats.

The dataset is publicly available on Kaggle:
🔗https://www.kaggle.com/datasets/shayanfazeli/heartbeat

Methods

1. Data Preprocessing

The ECG data was obtained from the MIT-BIH Arrhythmia dataset, where each heartbeat is represented by 188 time-point features. The data was normalized to a uniform scale to improve model training consistency. An 80:20 stratified train-test split was used to maintain the original class distribution and ensure fair evaluation across all heartbeat types.

2. Classification Models

Six machine learning models were implemented and compared:

XGBoost: Gradient boosting framework with high accuracy and robustness.

Random Forest: Ensemble of decision trees using bagging.

LightGBM: Gradient boosting framework optimized for speed and memory efficiency.

Bagging Classifier: Uses bootstrap aggregation of decision trees to improve generalization.

Decision Tree: A basic tree-based model used as a baseline.

AdaBoost: Adaptive boosting model that emphasizes misclassified examples during training.

All models were trained using default hyperparameters. Evaluation metrics included accuracy, precision, recall, and F1-score, with particular attention to performance on minority classes (S and F).

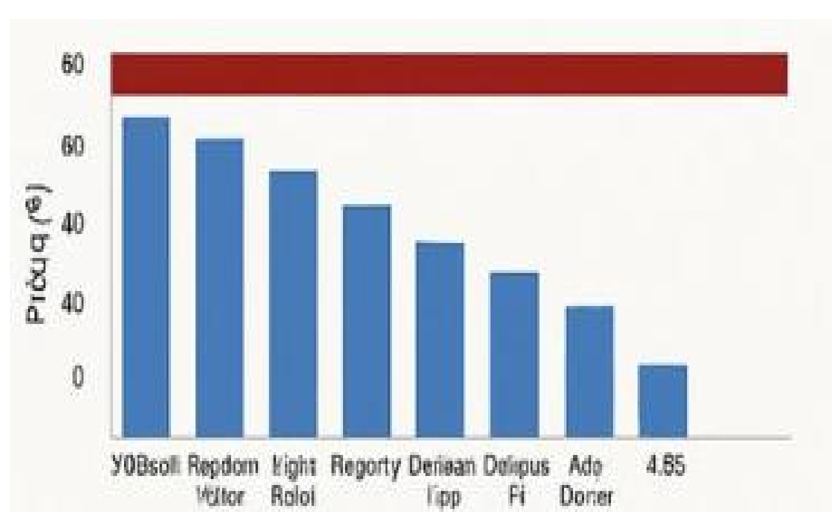

fig 2

Analysis

The XGBoost model outperformed all other classifiers, achieving an overall accuracy of 98%, with high precision and recall across most heartbeat classes. Random Forest, LightGBM, and Bagging also performed strongly, with accuracies around 97% and balanced classification for both majority and minority classes.

In contrast, the Decision Tree model showed lower overall performance (95%), especially for underrepresented classes. AdaBoost performed the worst, with an accuracy of only 45%, failing to generalize well due to overfitting on rare classes.

Key Observations:
XGBoost demonstrated excellent performance on both Normal and Unclassifiable beats.

LightGBM showed slightly better detection of rare classes like Supraventricular Premature (S) and Fusion (F) beats.

Confusion matrices revealed that most misclassifications occurred between morphologically similar heartbeat types.

Figure 3 shows representative ECG waveforms for each class, highlighting the subtle differences the models must learn to distinguish.

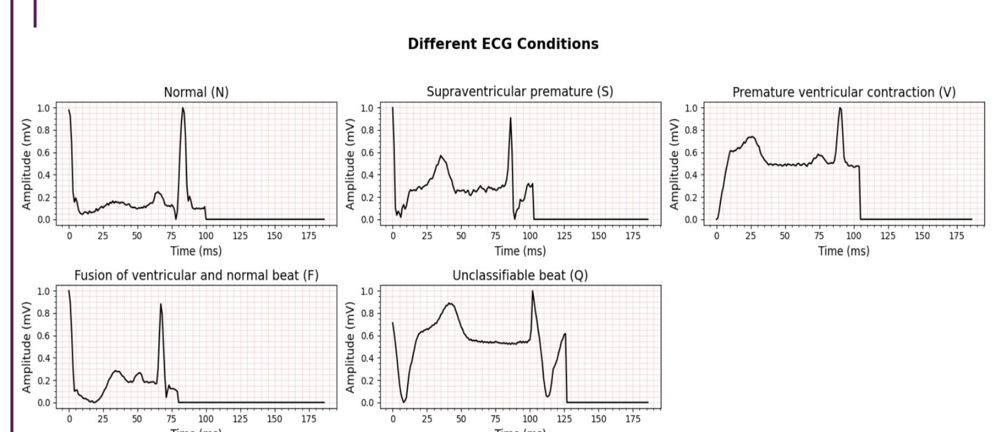

fig 3

Result

The detailed classification report for XGBoost revealed strong precision and recall values for the Normal (N) and Unclassifiable (Q) classes, with precision and recall close to or above 0.98. Performance was slightly lower for minority classes such as Fusion (F) and Supraventricular Premature (S) beats, consistent with their lower representation in the dataset.

	precision	recall	f1-score	support
Normal (N)	0.99	0.99	0.99	18118
Supraventricular premature (S)	0.79	0.78	0.78	556
Premature ventricular contraction (V)	0.95	0.95	0.95	1448
Fusion of ventricular and normal beat (F)	0.81	0.81	0.81	162
Unclassifiable beat (Q)	0.99	0.98	0.98	1608
accuracy			0.98	21892
macro avg	0.91	0.90	0.90	21892
weighted avg	0.98	0.98	0.98	21892

Fig 4

The confusion matrix for XGBoost illustrates the model's accurate predictions along the diagonal, with relatively few misclassifications. Most errors occurred between morphologically similar heartbeat classes, such as Fusion beats being confused with Normal or Ventricular beats.

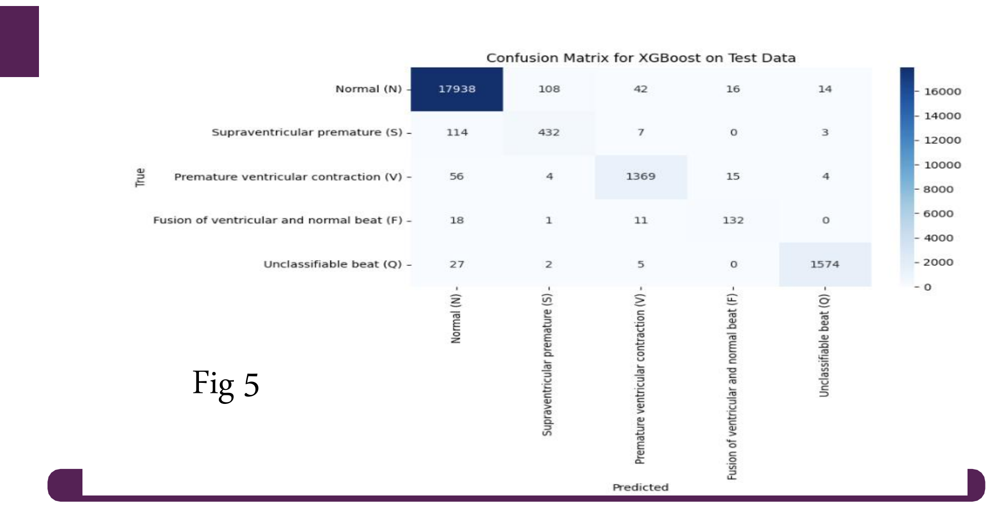

Fig 5

Conclusion

This project highlights the potential of machine learning, especially ensemble methods, for accurate ECG arrhythmia classification. XGBoost achieved the highest accuracy of 98%, outperforming other models in handling both common and rare heartbeat types. While other ensemble models also performed well, class imbalance remains a key challenge. These results support the use of gradient boosting models in developing fast, reliable tools for clinical arrhythmia detection.