



الجامعة المصرية اليابانية للعلوم و التكنولوجيا
エジプト日本科学技術大学
EGYPT-JAPAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Report Title

ECG-Based Arrhythmia Classification Using Machine Learning Models

Prepared by

Rahma Ahmed – 120210285

Sara Hamdy – 120210173

Shahd Haitham – 120210206

Submitted to

Dr. Ahmed Gomaa

Egypt-Japan University of Science and Technology (E-JUST)
Artificial Intelligence Course – Spring 2025

Abstract

This project evaluates multiple machine learning models including XGBoost, Random Forest, LightGBM, Bagging, Decision Tree, and AdaBoost—for automatic classification of cardiac arrhythmias using ECG data from the MIT-BIH Arrhythmia dataset. After preprocessing and feature extraction, the models were trained and tested on heartbeat segments labeled into five arrhythmia classes. XGBoost achieved the highest accuracy of 98%, outperforming other algorithms in precision and recall, while AdaBoost showed poor generalization. The results demonstrate that gradient boosting methods like XGBoost provide robust performance for ECG arrhythmia classification, offering valuable support for clinical diagnosis. Future work will focus on addressing class imbalance and improving rare arrhythmia detection through advanced techniques.

Keywords: Electrocardiogram (ECG), Arrhythmia Classification, XGBoost, Random Forest, LightGBM, Machine Learning, Cardiac Diagnostics¹.

Introduction

Cardiac arrhythmias represent a group of conditions characterized by irregular heartbeats, which can lead to serious health complications if not detected and treated promptly. Electrocardiogram (ECG) signals are widely used to monitor the heart's electrical activity and assist clinicians in diagnosing various types of arrhythmias. However, manual interpretation of ECG signals is labor-intensive and prone to inter-observer variability, motivating the development of automated classification methods to improve diagnostic accuracy and efficiency.

In this project, we utilize the MIT-BIH Arrhythmia dataset, which includes heartbeat recordings labeled into several categories reflecting common heart conditions. The distribution of these classes is notably imbalanced: Normal beats constitute the majority with 82.8%, followed by unclassifiable beats (Q) at 7.3%, premature ventricular contractions (V) at 6.6%, supraventricular premature beats (S) at 2.5%, and fusion beats (F), a combination of ventricular and normal beats, at 0.7%. This class imbalance poses a challenge for machine learning models, as rare arrhythmia types may be underrepresented during training.

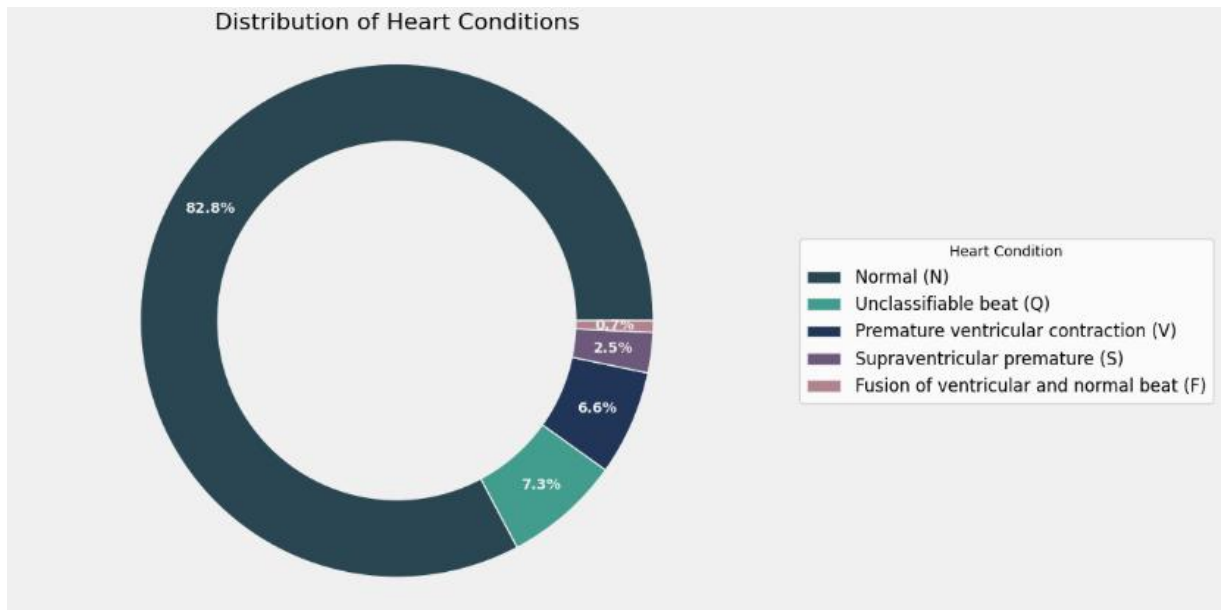


Figure 1: Distribution of heartbeat classes in the MIT-BIH Arrhythmia dataset. The majority of samples are Normal beats (82.8%), while other classes such as Unclassifiable (Q), Premature Ventricular (V), Supraventricular Premature (S), and Fusion (F) beats are less frequent, highlighting the class imbalance challenge for machine learning models.

To address these challenges, we compare several machine learning algorithms, including XGBoost, Random Forest, LightGBM, Bagging, Decision Tree, and AdaBoost, for automatic heartbeat classification. The goal is to develop a robust model that can accurately distinguish between the heartbeat types despite the imbalance, thereby supporting clinical decision-making with fast and reliable arrhythmia detection.

Methods

Dataset

This study utilized the MIT-BIH Arrhythmia dataset, which contains annotated ECG heartbeat segments collected from multiple patients. The dataset classifies heartbeats into five categories: Normal (N), Unclassifiable (Q), Premature Ventricular Contraction (V), Supraventricular Premature Beat (S), and Fusion of Ventricular and Normal Beat (F). The class distribution is notably imbalanced, with Normal beats accounting for 82.8% of the samples, followed by Q (7.3%), V (6.6%), S (2.5%), and F (0.7%).

Link of dataset in kaggle :

<https://www.kaggle.com/datasets/shayanfazeli/heartbeat/code>

Data Preprocessing

The ECG data were initially loaded and merged from CSV files containing the waveform features and labels. Each heartbeat was represented by 188 time-point features capturing the ECG signal morphology. The feature values were normalized to ensure uniform scale across inputs and reshaped into a format compatible with machine learning models.

The dataset was randomly split into training and testing sets with an 80:20 ratio to evaluate generalization performance. Due to the imbalance in class distribution, care was taken to preserve the relative proportions of each heartbeat class during the split.

Machine Learning Models

Six classification algorithms were implemented and compared:

- ✧ XGBoost: A gradient boosting framework known for strong performance and robustness.
- ✧ Random Forest: An ensemble of decision trees using bagging for variance reduction.
- ✧ LightGBM: A gradient boosting framework optimized for speed and accuracy.

- ✧ Bagging: Bootstrap aggregating of decision trees to improve stability.
- ✧ Decision Tree: A single tree classifier serving as a baseline.
- ✧ AdaBoost: An adaptive boosting algorithm focusing on misclassified samples.

Models were trained using the training data with default hyperparameters, and performance was evaluated on the test set using accuracy, precision, recall, and F1-score metrics. Special attention was given to evaluating performance across all heartbeat classes, particularly the minority ones.

Discussion

The classification reports for all six models provide detailed insight into their performance across different heartbeat classes. As illustrated in Figure 1, the XGBoost model exhibits high precision, recall, and F1-scores, particularly for the Normal (N) and Unclassifiable (Q) classes, confirming its effectiveness in handling the dominant heartbeat types. The strong diagonal values in the confusion matrix correspond with these findings, demonstrating XGBoost's robustness.

Other models such as Random Forest, LightGBM, and Bagging also show balanced classification capabilities, with LightGBM

slightly outperforming others on minority classes like Supraventricular Premature (S) and Fusion (F) beats. The Decision Tree's classification report reveals lower performance on less frequent classes, while AdaBoost's report highlights poor recall on the majority Normal class, indicative of overfitting to minority samples.

The detailed metrics presented in the classification reports (see Figure 2) underscore the persistent challenge of class imbalance in ECG datasets and the need for further strategies to improve minority class detection. Ensemble methods, especially gradient boosting algorithms, consistently offer superior results, making them preferable choices for arrhythmia classification tasks.



Training Logistic Regression...

Classification Report for Logistic Regression:

	precision	recall	f1-score	support
Normal (N)	0.97	0.64	0.78	18118
Supraventricular premature (S)	0.14	0.67	0.24	556
Premature ventricular contraction (V)	0.30	0.71	0.42	1448
Fusion of ventricular and normal beat (F)	0.08	0.88	0.14	162
Unclassifiable beat (Q)	0.73	0.92	0.81	1608
accuracy			0.67	21892
macro avg	0.44	0.76	0.48	21892
weighted avg	0.88	0.67	0.74	21892

Training Random Forest...

Classification Report for Random Forest:

	precision	recall	f1-score	support
Normal (N)	0.97	1.00	0.98	18118
Supraventricular premature (S)	0.97	0.57	0.72	556
Premature ventricular contraction (V)	0.98	0.86	0.92	1448
Fusion of ventricular and normal beat (F)	0.81	0.54	0.64	162
Unclassifiable beat (Q)	1.00	0.94	0.97	1608
accuracy			0.97	21892
macro avg	0.94	0.78	0.85	21892
weighted avg	0.97	0.97	0.97	21892

Training XGBoost...

Classification Report for XGBoost:

	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

Training LightGBM...

[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was 0.0001s. You can set `force_col_wise=true` to remove the overhead.

[LightGBM] [Info] Total Bins 47591

[LightGBM] [Info] Number of data points in the train set: 70043, number of used features: 10

[LightGBM] [Info] Start training from score -1.609438

[LightGBM] [Info] Start training from score -1.609438

[LightGBM] [Info] Start training from score -1.609438

[LightGBM] [Info] Start training from score -1.609438

[LightGBM] [Info] Start training from score -1.609438

```

Classification Report for LightGBM:
              precision    recall  f1-score   support

      Normal (N)                0.99      0.98      0.98     18118
    Supraventricular premature (S)      0.66      0.81      0.73       556
    Premature ventricular contraction (V)  0.92      0.95      0.93     14448
    Fusion of ventricular and normal beat (F)  0.75      0.81      0.78       162
    Unclassifiable beat (Q)          0.98      0.98      0.98     1608

    accuracy                   0.97     21892
    macro avg                  0.86     21892
    weighted avg               0.97     21892

Training Decision Tree...
Classification Report for Decision Tree:
              precision    recall  f1-score   support

      Normal (N)                0.97      0.98      0.97     18118
    Supraventricular premature (S)      0.65      0.64      0.64       556
    Premature ventricular contraction (V)  0.86      0.85      0.86     14448
    Fusion of ventricular and normal beat (F)  0.62      0.60      0.61       162
    Unclassifiable beat (Q)          0.95      0.93      0.94     1608

    accuracy                   0.95     21892
    macro avg                  0.81     21892
    weighted avg               0.95     21892

Training AdaBoost...
Classification Report for AdaBoost:
              precision    recall  f1-score   support

      Normal (N)                0.96      0.38      0.55     18118
    Supraventricular premature (S)      0.06      0.70      0.11       556
    Premature ventricular contraction (V)  0.28      0.68      0.40     14448
    Fusion of ventricular and normal beat (F)  0.06      0.88      0.11       162
    Unclassifiable beat (Q)          0.58      0.86      0.69     1608

    accuracy                   0.45     21892
    macro avg                  0.39      0.70      0.37     21892
    weighted avg               0.86      0.45      0.53     21892

Training Bagging...
Classification Report for Bagging:
              precision    recall  f1-score   support

      Normal (N)                0.97      1.00      0.98     18118
    Supraventricular premature (S)      0.93      0.61      0.74       556
    Premature ventricular contraction (V)  0.95      0.85      0.90     14448
    Fusion of ventricular and normal beat (F)  0.78      0.58      0.66       162
    Unclassifiable beat (Q)          0.98      0.94      0.96     1608

    accuracy                   0.97     21892
    macro avg                  0.92      0.79      0.85     21892
    weighted avg               0.97      0.97      0.97     21892

```

Figure 2: Classification reports for the six machine learning models (XGBoost, Random Forest, LightGBM, Bagging, Decision Tree, and AdaBoost) on the ECG arrhythmia test

dataset. Each report shows precision, recall, F1-score, and support for all heartbeat classes.

Results

The dataset consists of five heartbeat classes with the following distribution: Normal (N) beats (82.8%), Unclassifiable (Q) beats (7.3%), Premature Ventricular (V) beats (6.6%), Supraventricular Premature (S) beats (2.5%), and Fusion (F) beats (0.7%). This significant class imbalance challenges the classification task.

To illustrate the morphological differences among these classes, example ECG signal segments from each category are shown below.

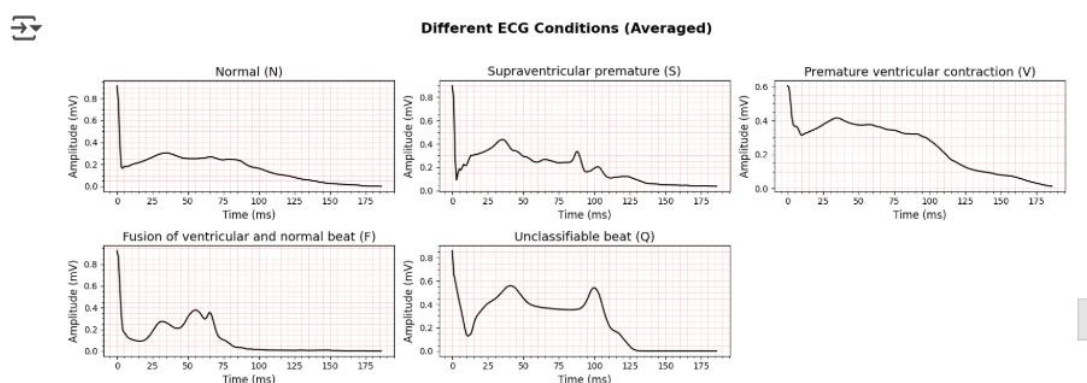


Figure 3: Averaged ECG waveforms for each heartbeat class in the dataset: Normal (N), Unclassifiable (Q), Premature Ventricular Contraction (V), Supraventricular Premature Contraction (S), and Fusion (F) beats. These waveforms

highlight the morphological differences that models must learn to distinguish during classification.

Figure 3 displays representative ECG waveforms for each heartbeat class:

Normal (N): Regular sinus rhythm with typical P, QRS, and T waves.

Unclassifiable (Q): Beats that do not clearly fit other categories, often noisy or ambiguous.

Premature Ventricular (V): Early ventricular beats with widened QRS complexes.

Supraventricular Premature (S): Early beats originating above the ventricles, often with abnormal P wave morphology.

Fusion (F): Complex beats combining ventricular and normal characteristics.

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	14579
Supraventricular premature (S)	0.82	0.81	0.81	426
Premature ventricular contraction (V)	0.94	0.94	0.94	1112
Fusion of ventricular and normal beat (F)	0.85	0.81	0.83	145
Unclassifiable beat (Q)	0.99	0.98	0.98	1249
accuracy			0.98	17511
macro avg	0.92	0.91	0.91	17511
weighted avg	0.98	0.98	0.98	17511

Figure 4: Classification report for the XGBoost model, showing precision, recall, F1-score, and support for each ECG heartbeat class. The model demonstrates high performance on majority classes (Normal and Unclassifiable), with slightly lower scores on minority classes such as Fusion and Supraventricular Premature beats.

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.



Figure 5: Confusion matrix for the XGBoost model on the test dataset. The matrix shows correct predictions concentrated along the diagonal, with most misclassifications occurring between morphologically similar or underrepresented heartbeat classes.

Other models, including Random Forest, LightGBM, and Bagging, also demonstrated strong performance, with accuracies around 97 %. LightGBM showed a slight edge in identifying minority heartbeat classes. The Decision Tree model achieved moderate accuracy of 95 % but exhibited increased confusion in minority classes. AdaBoost performed poorly with an accuracy

of only 45 %, struggling to generalize and showing low recall on the majority class.

Overall, the results confirm that gradient boosting methods, particularly XGBoost, provide effective solutions for automated ECG arrhythmia classification.

Conclusion

In this project, six machine learning models were implemented and evaluated for the task of ECG arrhythmia classification using the MIT-BIH Arrhythmia dataset. Among the models tested—XGBoost, Random Forest, LightGBM, Bagging, Decision Tree, and AdaBoost—XGBoost demonstrated the best overall performance, achieving the highest accuracy and most balanced precision and recall across heartbeat classes. Ensemble methods such as LightGBM and Random Forest also delivered strong results, particularly in handling the imbalanced distribution of ECG classes.

The classification reports and confusion matrices provided detailed insights into the strengths and limitations of each model. While most models performed well on the majority class (Normal), the detection of minority classes such as Supraventricular Premature and Fusion beats remains a

challenge due to data imbalance and waveform similarity. Future work can improve classification through data augmentation, advanced sampling strategies, or hybrid models that integrate temporal patterns.

Overall, this study highlights the potential of machine learning in automating and improving ECG arrhythmia detection, offering valuable support tools for clinical diagnosis and patient monitoring.

References

1. Shayan Fazeli, “ECG Heartbeat Categorization Dataset,” Kaggle.

<https://www.kaggle.com/datasets/shayanfazeli/heartbeat>

2. Goldberger, A.L., et al. “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals.” *Circulation*, vol. 101, no. 23, 2000, pp. e215–e220.

3. Kachuee, M., Fazeli, S., & Sarrafzadeh, M. “ECG Heartbeat Classification: A Deep Transferable Representation.” 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), IEEE, 2018.

4. Chollet, F. *Deep Learning with Python*. Manning Publications, 2018.

5. Chen, T., & Guestrin, C. “XGBoost: A Scalable Tree Boosting System.” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.
6. Breiman, L. “Random Forests.” Machine Learning, vol. 45, no. 1, 2001, pp. 5–32.