ECG Arrhythmia Classification using Stacking Ensemble Learning

Pham Minh Hieu

University of Science and Technology of Hanoi

Abstract—This report presents a stacking ensemble approach for ECG heartbeat arrhythmia classification using the MIT-BIH Arrhythmia Dataset. Instead of relying on a single classifier or deep neural networks, the proposed system combines multiple tree-based models—Extra Trees, Random Forest, LightGBM, and XGBoost—into a two-level stacking framework. Out-of-fold class probability predictions are used as meta-features and fed into an XGBoost meta-classifier. The ensemble is evaluated using stratified cross-validation and achieves strong generalization performance on the unseen test set, demonstrating the effectiveness of ensemble learning for non-linear ECG signal classification.

I. INTRODUCTION

Electrocardiogram (ECG) analysis plays a critical role in diagnosing cardiac arrhythmias. While deep learning models such as Convolutional Neural Networks (CNNs) have achieved remarkable performance, they often require extensive computational resources and large-scale tuning.

Ensemble learning offers an alternative by combining multiple strong but diverse classifiers to capture complex, non-linear patterns in ECG signals. In this study, we implement a stacking ensemble framework using classical machine learning models to classify ECG heartbeats into five arrhythmia categories.

II. DATASET DESCRIPTION

The dataset is derived from the MIT-BIH Arrhythmia Database and consists of pre-segmented heartbeat signals.

A. Data Structure

The dataset is provided in two CSV files:

- `mitbih_train.csv`
- `mitbih_test.csv`

Each sample contains 188 columns:

- **Columns 0–186**: ECG signal amplitudes sampled at 125 Hz
- **Column 187**: Target class label (0–4)

B. Class Distribution

The dataset exhibits severe class imbalance, with Normal beats dominating the data. No explicit resampling is applied; instead, ensemble diversity and probabilistic stacking are used to improve minority class recognition.

III. METHODOLOGY

A. Noise Augmentation

To improve robustness and reduce overfitting, Gaussian noise is added to the training data:

$$X' = X + \mathcal{N}(0, 0.01)$$

This simulates realistic signal variations and improves model generalization.

B. Base Models

Four heterogeneous tree-based classifiers are used as base learners:

- Extra Trees Classifier (ET)
- Random Forest (RF)
- Light Gradient Boosting Machine (LGBM)
- Extreme Gradient Boosting (XGBoost)

All base models are trained to output class probability vectors using `predict_proba`.

C. Stacking Strategy

A stratified K-fold cross-validation scheme ($K = 7$) is used to generate out-of-fold predictions for each base model. For each training sample, predicted class probabilities from all base models are concatenated to form the meta-feature vector:

$$\mathbf{z}_i = [p_i^{ET}, p_i^{RF}, p_i^{LGB}, p_i^{XGB}]$$

D. Meta-Model

The meta-classifier is an XGBoost model trained on the stacked probability features. This allows the meta-model to learn optimal combinations of base model outputs and resolve inter-model disagreements.

IV. EXPERIMENTAL SETUP

A. Cross-Validation

Stratified K-Fold cross-validation ensures class distribution consistency across folds. Accuracy is used as the primary validation metric during base-model training.

B. Evaluation Metrics

Final evaluation on the test set includes:

- Overall Accuracy
- Macro-averaged Precision
- Macro-averaged Recall
- Macro-averaged F1-score

V. Results and Discussion

A. Classifiaction result

Table I reports the class-wise performance of the stacking ensemble model on the test set.

TABLE I
Class-wise Classification Performance

Class	Precision	Recall	F1-score
0	0.98	0.99	0.99
1	0.89	0.76	0.82
2	0.96	0.94	0.95
3	0.84	0.75	0.76
4	0.99	0.97	0.98
Macro Avg	0.93	0.88	0.90
Weighted Avg	0.98	0.98	0.98
Overall accuracy	0.98		

B. Confusion Matrix Analysis

Figure 1 presents the confusion matrix of the stacking ensemble model evaluated on the test set.
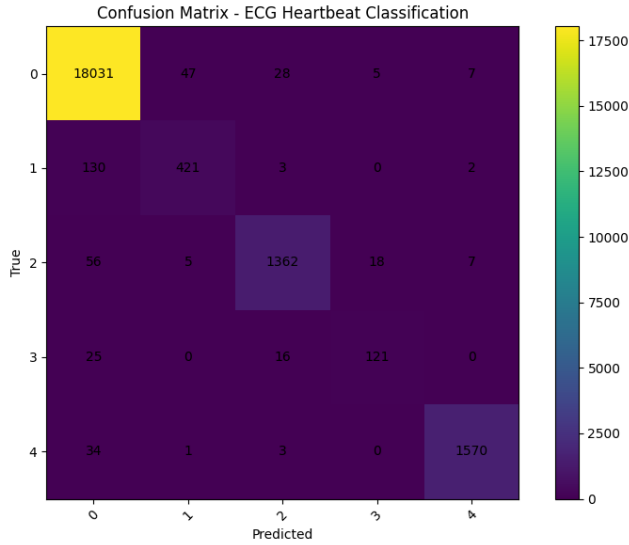

Fig. 1. Confusion matrix of the stacking ensemble model on the MIT-BIH test set.

VI. Conclusion

This study demonstrates that stacking ensemble learning is an effective and computationally efficient solution for ECG arrhythmia classification. By combining multiple tree-based classifiers and leveraging probabilistic meta-features, the proposed approach achieves strong performance without relying on deep neural networks(in fact, deep neural network is better when the dataset is too large or the data is unstructured tabular, otherwise our boost family is still the best :))).