ECG Heartbeat Classification Using Logistic Regression

Nguyen The Khai - 23bi14205 - DS

***Abstract*—This work implements a logistic regression classifier for ECG heartbeat categorization using the MIT-BIH Arrhythmia dataset. The model achieves 78.51% overall accuracy with strong performance on majority classes but struggles with severe class imbalance. Results establish a linear baseline for comparison with deep learning approaches reported in literature.**

I. Data Exploration

The MIT-BIH Arrhythmia dataset contains 87,554 training and 21,892 test samples, each represented by 187 temporal features corresponding to resampled ECG heartbeat signals. Features are normalized to [0, 1] range. The classification task involves five categories following AAMI EC57 standards:

- Normal (Class 0),
- Supraventricular ectopic beat (Class 1)
- Ventricular ectopic beat (Class 2)
- Fusion beat (Class 3)
- Unknown beat (Class 4)

Severe class imbalance characterizes the dataset. Normal beats dominate at 82.77% of training data, while minority classes exhibit extreme imbalance ratios. Class 3 (Fusion) represents only 0.73% of samples with an imbalance ratio of 1:113.06 relative to the majority class. Classes 1, 2, and 4 show moderate imbalance at ratios of 1:32.60, 1:12.52, and 1:11.27 respectively. Data quality inspection reveals no missing values or duplicate samples, indicating clean preprocessed data.

II. Preprocessing

The training set was split into 90% training (78,798 samples) and 10% validation (8,756 samples) using stratified sampling to preserve class distribution across splits. Standardization via z-score normalization was applied, centering features to zero mean and unit variance. This transformation addresses inter-patient amplitude variability inherent in ECG signals and facilitates numerical stability in optimization.

The standardization parameters were computed exclusively on the training set and applied identically to validation and test sets to prevent data leakage. Post-normalization verification confirmed mean ≈ 0 and standard deviation $= 1$ on training data, validating correct implementation.

III. Model Implementation and Training

Logistic regression with One-vs-Rest (OvR) multi-class strategy was selected as a baseline classifier. The model employs L2 regularization with $C = 1.0$ and uses the lbfgs solver with maximum 1,000 iterations. Built-in class weighting addresses imbalance by assigning loss penalties inversely proportional to class frequencies.

Training converged successfully, achieving 78.93% training accuracy and 79.20% validation accuracy. The minimal train-validation gap indicates absence of overfitting, suggesting the model generalizes appropriately within its representational capacity. However, the moderate absolute performance suggests linear decision boundaries may be insufficient for capturing complex ECG morphological patterns.

IV. Evaluation Results

Test accuracy reached 78.51%, closely matching validation performance and confirming model stability. Table I presents detailed per-class metrics.

TABLE I
Per-Class Performance Metrics

Class	Precision	Recall	F1-Score	Support
Normal (0)	0.969	0.792	0.871	18118
SVEB (1)	0.333	0.565	0.419	556
VEB (2)	0.306	0.651	0.416	1448
Fusion (3)	0.119	0.778	0.207	162
Unknown (4)	0.728	0.912	0.809	1608
Macro Avg	0.491	0.739	0.544	21892
Weighted Avg	0.885	0.785	0.820	21892

Class 0 (Normal) achieves excellent precision (0.969) but moderate recall (0.792), with 2,778 false negatives primarily confused with Class 2 (VEB). This suggests morphological similarity between certain normal and ventricular beats that linear models cannot disambiguate.

Minority classes show characteristic precision-recall imbalance. Classes 1-3 exhibit low precision (0.119-0.333) but reasonable recall (0.565-0.778), indicating the model predicts these classes liberally to compensate for limited training examples. Class 3 (Fusion) performs worst with F1-score of 0.207, despite 77.8% recall, due to extreme precision degradation (0.119).

Class 4 (Unknown) surprisingly achieves strong performance (F1 = 0.809), suggesting this class possesses distinct morphological characteristics that linear classifiers can capture despite imbalance.

The macro F1-score of 0.544 substantially trails the weighted F1-score of 0.820, quantifying the performance gap between majority and minority classes. Confusion matrix analysis reveals 3,778 total misclassifications, with Class 0 contributing 3,778 errors through false negatives distributed across other categories.

V. Conclusion

This work establishes a logistic regression baseline achieving 78.51% accuracy on MIT-BIH Arrhythmia classification. Performance substantially underperforms the 93.4% accuracy reported by Mohammad Kachuee et al. using residual CNNs, indicating temporal feature learning provides significant discriminative value beyond linear classification of raw signals.

The 14.9 percentage point gap validates the necessity of nonlinear models for ECG classification. Severe class imbalance remains the primary challenge, with minority class F1-scores ranging from 0.207 to 0.419 despite balanced loss weighting. Future work should explore data augmentation, advanced sampling techniques, or ensemble methods to improve minority class performance while maintaining computational efficiency advantages of simple models.