ECG Heartbeat Classification using an LSTM Network on the MIT-BIH Dataset

Pham Minh Duc (22BA13083)
Machine Learning in Medicine 2026
University of Science and Technology of Hanoi (USTH)
Email: ducpm.22ba13083@usth.edu.vn

***Abstract*—This report investigates a deep learning baseline for classifying electrocardiogram (ECG) heartbeat segments into five distinct categories using the MIT-BIH Arrhythmia dataset. The proposed workflow covers data inspection, preprocessing, temporal reshaping, and supervised training with a Long Short-Term Memory (LSTM) network. Model evaluation is conducted on an independent test set using overall accuracy, macro-level evaluation metrics, a normalized confusion matrix, and class-wise precision and recall. The results indicate strong predictive performance for dominant classes, whereas minority classes remain difficult to identify due to pronounced class imbalance. Possible enhancement strategies, including loss re-weighting, data augmentation, and hybrid CNN–LSTM models, are discussed.**

***Index Terms*—ECG, MIT-BIH, Arrhythmia Detection, LSTM, Deep Learning, Classification**

I. Introduction

Electrocardiogram (ECG) signals provide a non-invasive means of monitoring cardiac electrical activity and play a crucial role in the diagnosis of heart rhythm disorders. Traditional ECG analysis relies heavily on expert interpretation, which is labor-intensive and subject to inter-observer variability. Consequently, automated ECG classification methods based on machine learning and deep learning have gained significant attention.

In this practical work, we construct an end-to-end ECG heartbeat classification framework based on a Long Short-Term Memory (LSTM) neural network. The primary objective is to establish a reproducible baseline model and to systematically evaluate its performance and limitations on the MIT-BIH Arrhythmia dataset.

II. Dataset

A. Class Distribution

Fig. 1 illustrates the number of samples per class in the training and test sets. The dataset is highly imbalanced, with the *Normal* class accounting for the majority of samples. Minority classes such as *Supraventricular* and *Fusion* contain far fewer examples, which may negatively impact model performance.

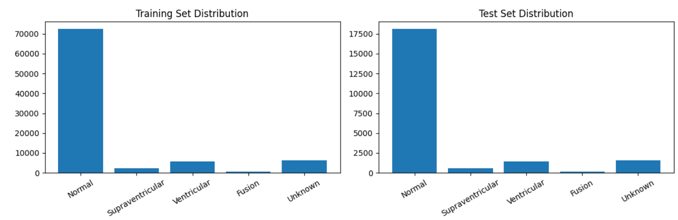

Fig. 1: Class distribution in the training and test sets.

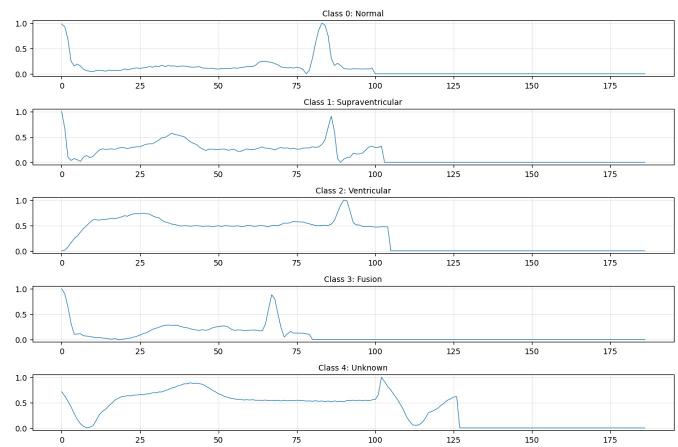

Fig. 2: Representative ECG heartbeat waveforms for each class.

B. Representative ECG Heartbeats

Fig. 2 presents one representative heartbeat waveform from each class. Although different arrhythmia types exhibit distinct morphological patterns, some minority-class waveforms remain visually similar to the normal class, increasing classification difficulty.

III. Methods

A. Preprocessing

The dataset is divided into training and validation subsets using a stratified 85/15 split to maintain the original class proportions. Feature normalization is applied using a `StandardScaler` fitted exclusively on the training data to

TABLE I: LSTM network architecture.

Layer	Output Shape	Description
Input	$(T, 1)$	ECG heartbeat sequence
LSTM (64)	(64)	Temporal modeling
Dropout (0.3)	(64)	Regularization
Dense (32)	(32)	Feature projection
Dense (5)	(5)	Softmax classification

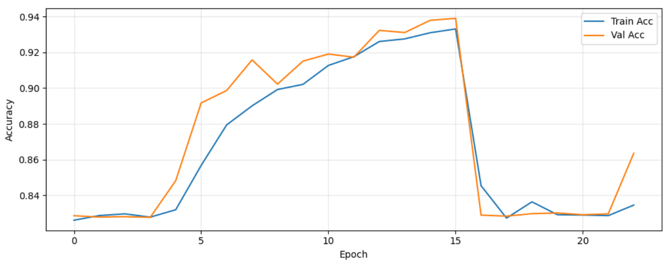

Fig. 3: Training and validation accuracy across epochs.

avoid information leakage. Each ECG heartbeat is then reshaped into a temporal sequence of dimension $(T, 1)$, enabling direct input into the LSTM network.

B. Model Architecture

A compact LSTM-based classification model is adopted. The network architecture comprises:

- One LSTM layer for temporal feature modeling,
- A dropout layer to reduce overfitting,
- A fully connected hidden layer with ReLU activation,
- A softmax output layer producing probabilities for five classes.

C. Training Setup

Model training is performed using the Adam optimization algorithm with a learning rate of 10^{-3} and sparse categorical cross-entropy as the loss function. Early stopping based on validation loss is employed to prevent overfitting and to restore the best-performing model parameters.

IV. Results

A. Learning Curves

The evolution of training and validation accuracy and loss over epochs is depicted in Figures 3 and 4. Rapid convergence during early epochs suggests effective learning of temporal patterns. The stabilization of validation metrics confirms the effectiveness of early stopping in selecting an optimal model.

B. Confusion Matrix

Figure 5 presents the normalized confusion matrix obtained on the test set. The classifier performs exceptionally well on the dominant *Normal* class. However, minority classes are frequently misclassified, most often as normal heartbeats, highlighting the influence of class imbalance.

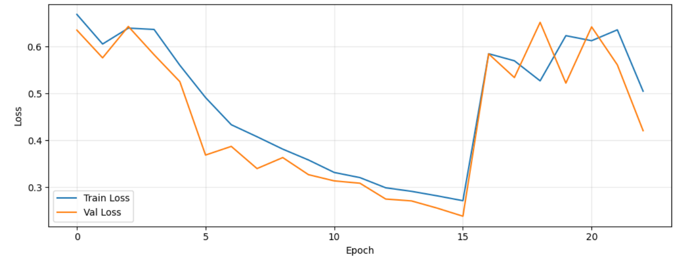

Fig. 4: Training and validation loss across epochs.

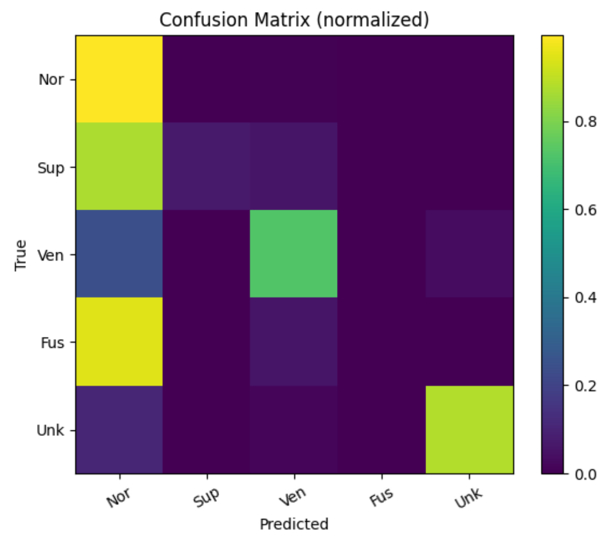

Fig. 5: Normalized confusion matrix on the test set.

TABLE II: Per-class classification report on the test set.

Class	Precision	Recall	F1-score	Support
Normal	0.94	0.99	0.97	18118
Supraventricular	0.91	0.07	0.13	556
Ventricular	0.87	0.72	0.79	1448
Fusion	0.00	0.00	0.00	162
Unknown	0.95	0.88	0.91	1608
Accuracy		0.94		21892
Macro Avg	0.73	0.53	0.56	21892
Weighted Avg	0.93	0.94	0.92	21892

C. Classification Report

Table II reports precision, recall, and F1-score for each heartbeat class. Despite the high overall accuracy, macro-averaged scores remain relatively low, indicating limited predictive capability for rare classes and emphasizing the imbalance-related limitations of the model.

V. Discussion

The results demonstrate that the LSTM-based model is capable of effectively learning temporal characteristics of ECG heartbeat signals, resulting in strong overall classification accuracy. Nevertheless, the severe imbalance among classes substantially reduces performance on underrepresented heartbeat types, particularly *Fusion* and *Supraventricular* beats. These shortcomings are clearly observed in the confusion matrix as well as in macro-averaged evaluation metrics.

Several directions may be explored to improve performance on minority classes:

- Employing class-weighted or focal loss functions to better handle imbalanced data,
- Applying data augmentation techniques to enrich rare heartbeat patterns,
- Utilizing hybrid CNN–LSTM architectures for enhanced feature extraction,
- Introducing attention mechanisms to focus on diagnostically relevant signal regions.

VI. Conclusion

This study proposed an end-to-end ECG heartbeat classification approach based on an LSTM neural network trained on the MIT-BIH Arrhythmia dataset. While the baseline model achieves strong performance for majority classes, the results reveal that class imbalance remains a major obstacle to reliable arrhythmia detection. Future work will concentrate on imbalance-aware learning strategies and the adoption of more advanced deep learning architectures to further enhance classification robustness.