ECG Heartbeat Categorization

Nguyn Kiên Trung BA12-178

February 2025

1 Introduction

I will collect this data from Kaggle. This dataset is composed from famous dataset in heartbeat classification, the MIT-BIH Arrhythmia Dataset. I have applied this dataset with two models are DNN and RNN.

2 Data Analysis

The dataset consists of five types of heartbeats: Normal, Supraventricular, Ventricular, Fusion, and Unknown. These categories are encoded as 0, 1, 2, 3, and 4, respectively. Each sample contains a sequence of 187 values representing the heartbeat signal.

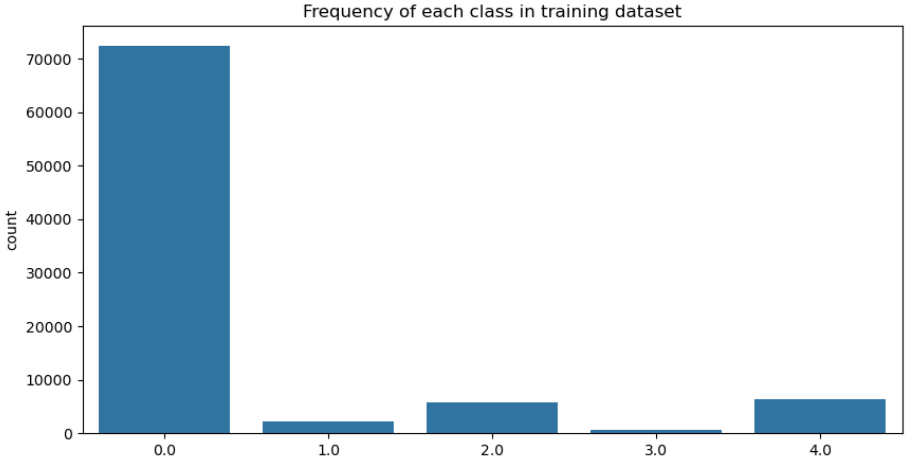

Figure 1: Frequency of each class in training dataset

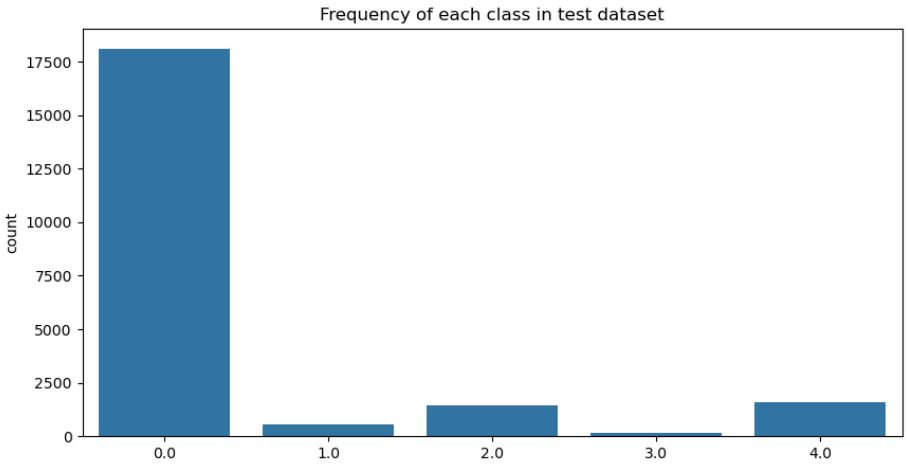

Figure 2: Frequency of each class in test dataset

We can see in figure one and two, the model has a normal label that is larger than the other labels by quite a large margin. This leads to data imbalance and causes overfitting. To fix this, I decided to create more mocks to balance the minority classes.

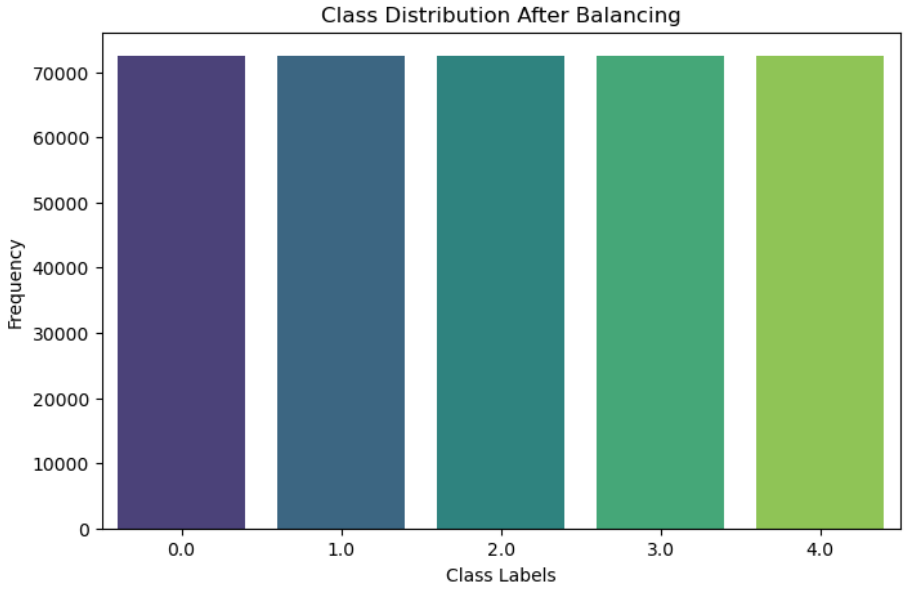

Figure 3: Class Distribution After Balancing

3 Models

3.1 DNN Model

I have used fully connected layers, dropout and regularization to improve the stability and generalization of the model. The model consists of four dense layers, the first three of which use ReLU activation function for feature extraction and the last layer applies softmax for classification. L2 regularization and dropout are used to prevent overfitting.

3.2 LSTM Model

I have implemented LSTM to capture temporal dependencies in sequential data. The model consists of two LSTM layers. Batch normalization and dropout are applied to stabilize the training process and prevent overfitting. The final classification is done through two fully connected layers followed by ReLU activation for feature learning and softmax for multi-class classification.

4 Results of two models

This is the confusion matrix of the two models LSTM and DNN.

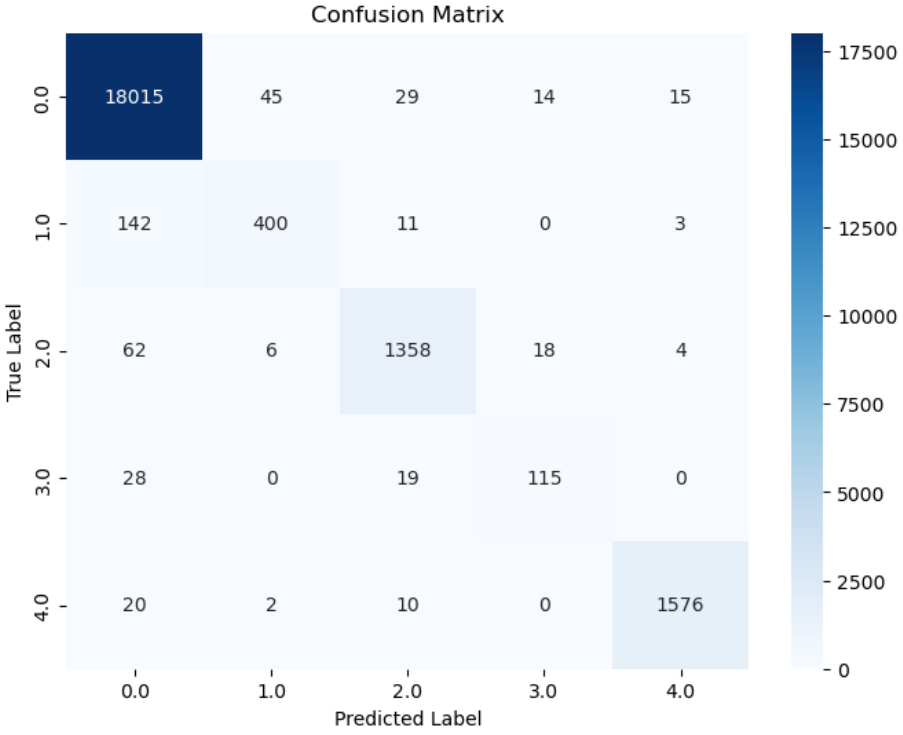

Figure 4: Confusion Matrix of LSTM

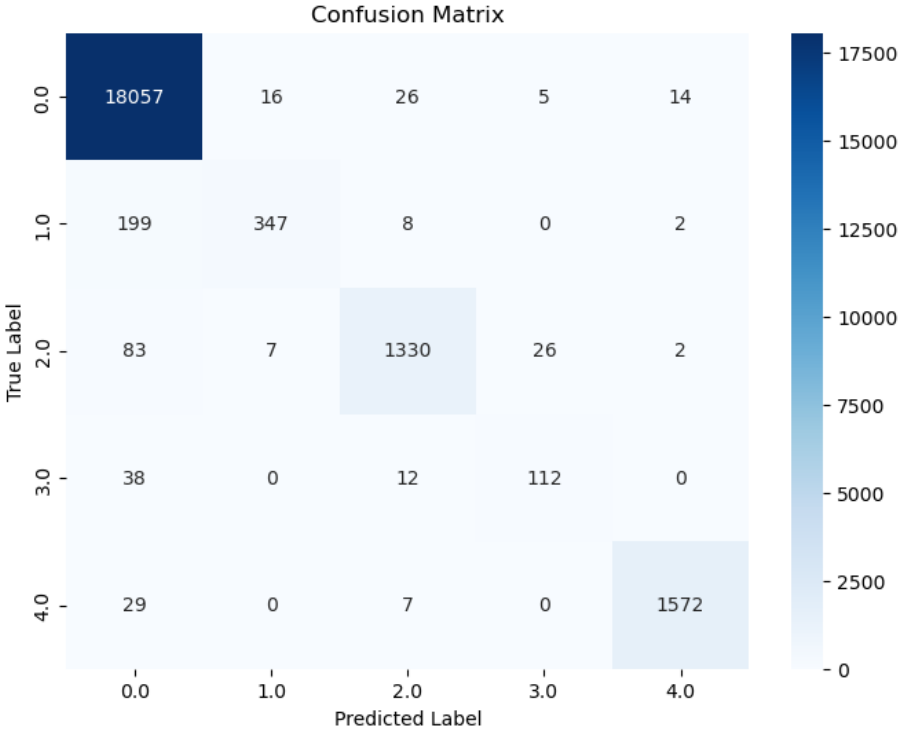

Figure 5: Confusion Matrix of DNN

Based on the results of the two confusion matrices we can see that the LSTM model is slightly better than the DNN model.