Report Study of ECG Heartbeat Categorization

Nguyen Dinh Tung – 22BI13455

Feb 27, 2025

1 Introduction

This report presents the results of my study on ECG heartbeat categorization taken from Kaggle. The data used in this study is the MIT-BIH Arrhythmia Database. I tried to analyze the data and apply CNN to classify 5 types of heartbeat.

2 Data Analysis and Preprocessing

As the figure below shows: the data was clean regarding missing values but observed a significant imbalance in class distribution—class 0 (normal heartbeats) far outnumbered the others.

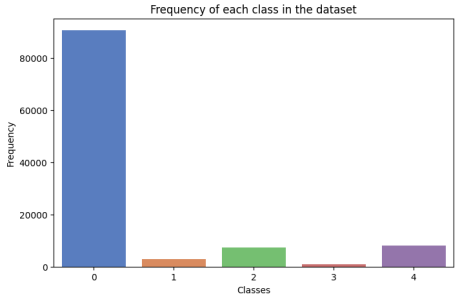

Figure 1: Frequency of each class in the dataset

Preprocessing

- **Cleaning data:** Check for missing and duplicate entries.
- **Normalize the features:** Ensure all input features contribute equally to the learning process.
- **Reshape:** Reshape the data by adding a new axis to transform it into the shape `(samples, time_steps, channels)`.

3 Model Evaluation

Class	Precision	Recall	F1-score	Support
0	0.99	1.00	0.99	18118
1	0.93	0.74	0.83	556
2	0.95	0.96	0.96	1448
3	0.92	0.68	0.78	162
4	0.99	0.99	0.99	1608
Accuracy	-	-	0.98	21892
Macro avg	0.96	0.87	0.91	21892
Weighted avg	0.98	0.98	0.98	21892

Table 1: Classification Report for CNN

4 Result

When comparing my CNN model to the original MIT-BIH paper, which reported an accuracy of 95.9%, my CNN achieves 98% accuracy, surpassing the paper's result. While the model excels at identifying the majority class (0) with near-perfect precision (0.99) and recall (1.00), it struggles more with minority classes such as 1 and 3, which show lower recall (0.74 and 0.68, respectively). As a result, the macro-average F1-score stands at 0.91, indicating that techniques like SMOTE or class weighting could be employed to further improve performance on underrepresented classes, bringing the model closer to a more balanced detection of arrhythmias.