A Report of ECG Heartbeat Categorization

Le Quoc Trung

Feb 26, 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 will analyze the data and apply CNN to classify 5 types of heartbeats.

2 Data Analysis

The dataset used in this study is the MIT-BIH Arrhythmia Database, comprising five types of heartbeats: Normal, Supraventricular, Ventricular, Fusion, and Unknown, encoded as 0, 1, 2, 3, and 4, respectively. Each data sample consists of 187 values representing the heartbeat signal.

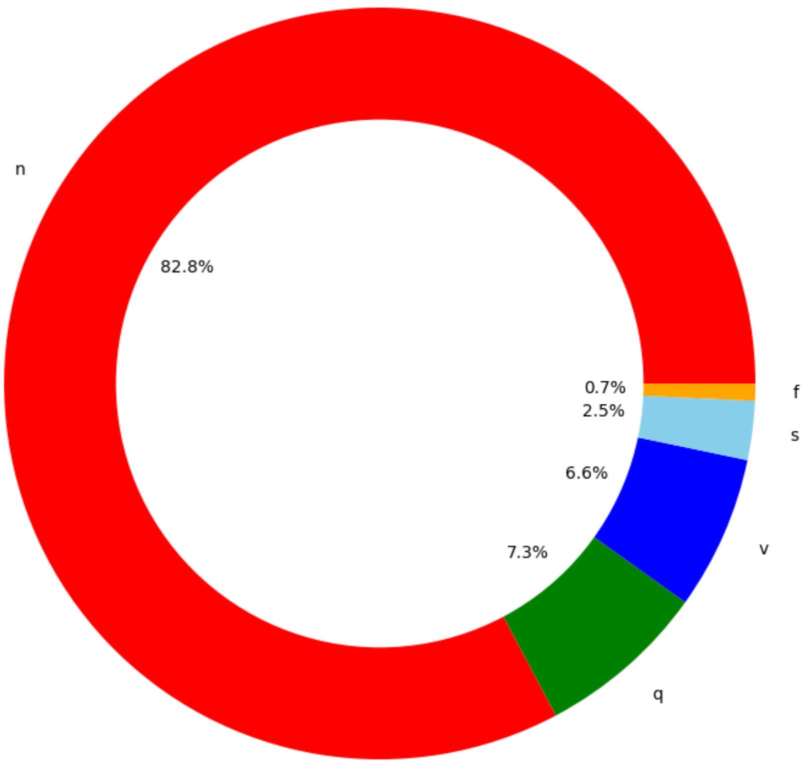

Figure 1: Data Distribution

As seen in the figure above, the current dataset has a significant imbalance between the Normal class and the other classes. This imbalance can cause the model to overfit, leading to inaccuracies in classification.

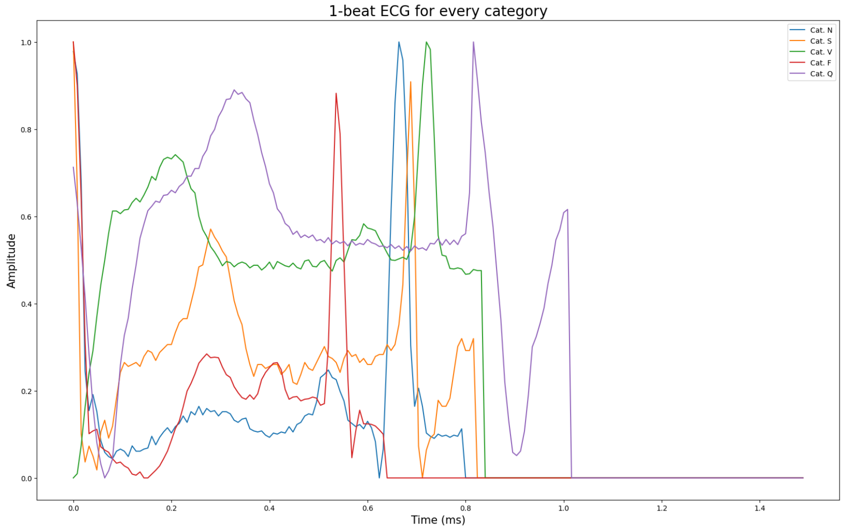

Figure 2: ECG of each class

This is a graph of ECG (electrocardiogram) signals for a single heartbeat from each different heartbeat type, providing a visual representation of the data. There are five distinct colors representing the five different heartbeat types in each case. To balance the dataset, I applied data augmentation techniques:

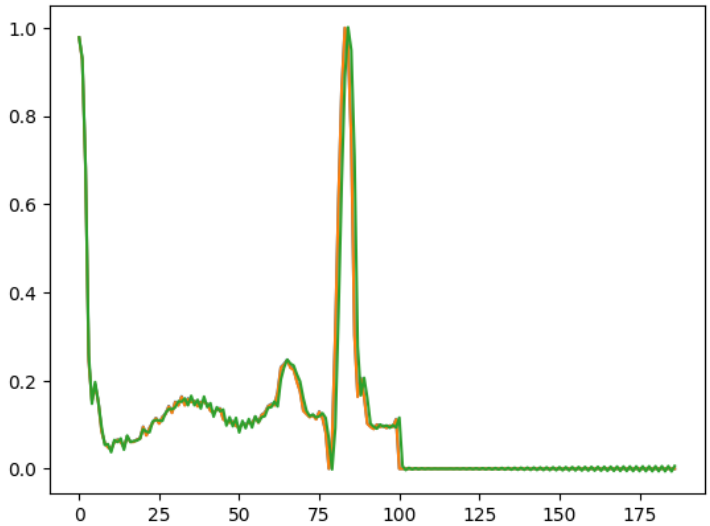

Figure 3: Data Augmentation

I enhanced the dataset using various data augmentation techniques, including time stretching or compression of the signal, amplitude scaling, and generating three transformed versions of the original signal. These methods help mitigate overfitting caused by the severe data imbalance, allowing the model to learn more generalized representations for underrepresented cases.

I will randomly select 800 samples from each class for testing to ensure a fair and accurate evaluation of the model's classification performance.

3 Model

I designed a convolutional neural network inspired by the original paper, utilizing residual connections to enhance feature extraction and prevent vanishing gradients. Similar to the original approach, my model consists of multiple convolutional layers with skip connections, followed by fully connected layers for classification. This design allows for more efficient learning and improved generalization in ECG heartbeat classification.

4 Results and Comparison

Table 1: Classification Report

Class	Precision	Recall	F1-Score	Support
0	0.81	1.00	0.89	800
1	1.00	0.83	0.91	800
2	0.97	0.95	0.96	800
3	0.99	0.93	0.96	800
4	1.00	0.99	1.00	800
Accuracy		0.94		4000
Macro Avg	0.95	0.94	0.94	4000
Weighted Avg	0.95	0.94	0.94	4000

Table 2: Performance Comparison

Method	Accuracy	Precision	Recall
Your Model	94%	95%	94%
Original Paper	93.4%	95.2%	95.1%

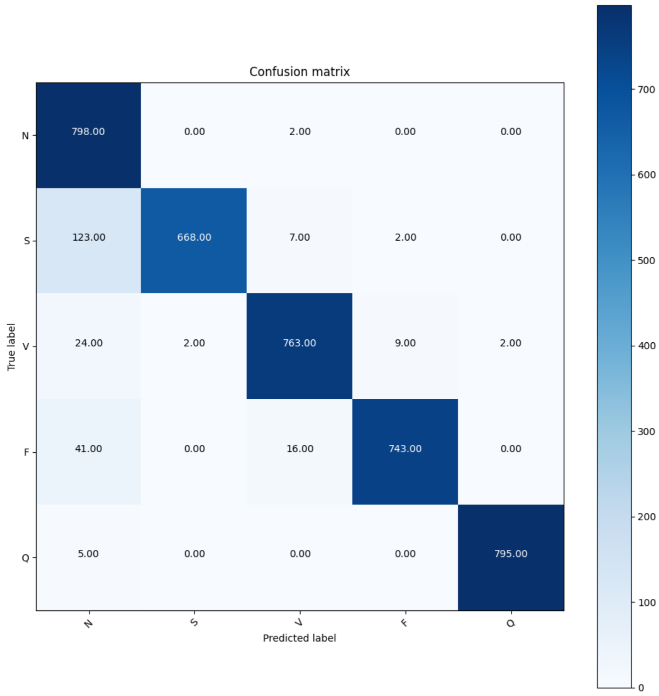

Figure 4: Confusion Matrix