UNIVERSITY OF SCIENCE AND TECHNOLOGY OF HANOI

HC18 Challenge Using UNet++

Vũ Hải Thiên Long
22BI13268

Contents

1. Introduction

During pregnancy, ultrasound imaging is used to measure fetal biometrics. One of these measurements is the fetal head circumference (HC). The HC can be used to estimate the gestational age and monitor growth of the fetus. The HC is measured in a specific cross section of the fetal head, which is called the standard plane. The dataset for this challenge contains a total of 1334 two-dimensional (2D) ultrasound images of the standard plane that can be used to measure the HC.

2. Dataset

The data is divided into a **training set** of **999 images** and a **test set** of **335 images**. The size of each *2D ultrasound image* is **800 by 540 pixels** with a pixel size ranging from *0.052 to 0.326 mm*. The *pixel size* for each image can be found in the *CSV files*.

The **training set** also includes an image with the *manual annotation* of the **head circumference (HC)** for each HC, which was made by a *trained sonographer*. The *CSV file* **training_set_pixel_size_and_HC** includes the **head circumference measurement** (in *millimeters*) for each *annotated HC* in the **training set**. All *filenames* start with a number. There are **999 images** in the **training set**, but the *filenames* only go to **805**. Some *ultrasound images* were made during the same *echoscopic examination* and have therefore a very *similar appearance*. These images have an *additional number* in the *filename* in between "„*and*"*HC*".

- Number of training sample: 999
- Number of test sample: 335
- Image size: 800 x 540 pixels
- Classes: ['N': 0, 'S': 1, 'V': 2, 'F': 3, 'Q': 4]

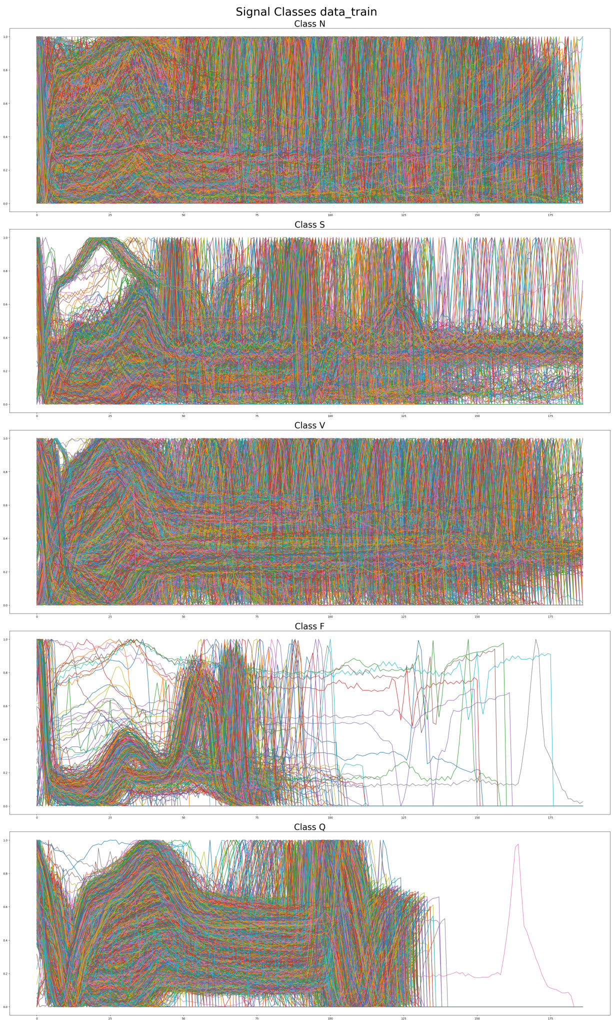

Figure 2.1: Sample ECG signal from the MIT-BIH Arrhythmia Dataset of all classes in train dataset

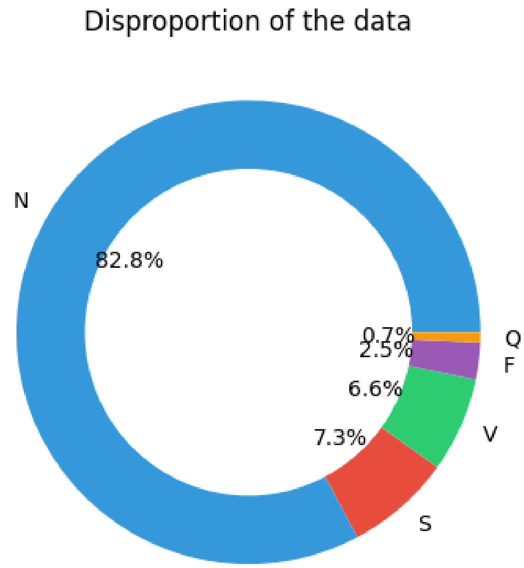

Figure 2.2: ECG class percentage in train dataset

The dataset is imbalanced, with most samples belonging to the 'N' class (normal heartbeats), which may bias the classification algorithm toward the majority class and hinder learning of minority class features. However, despite the lower quantity of 'F' and 'Q' class signals, their distinct shapes could lead to better classification performance.

2.1 Data processing

Due to the imbalance in the dataset, with the 'N' class (normal heartbeats) having significantly more samples, I applied under-sampling to balance the dataset. This process involves randomly selecting a subset of 'N' class signals, reducing their number to match the average number of signals in the other classes. Under-sampling helps prevent the model from becoming biased toward the majority class ('N') and ensures it can effectively learn the features of the minority classes.

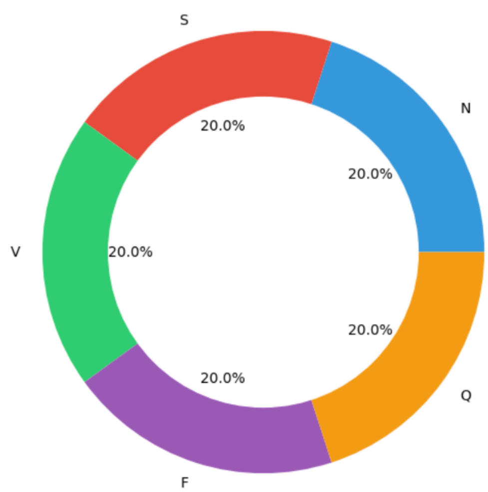

Figure 2.3: Dataset after Resampling

3. Method

3.1 Introduction to the Method

In this study, I use a **Convolutional Neural Network (CNN)** for ECG heartbeat classification, as CNNs automatically extract important features from raw ECG signals. The architecture includes convolutional, pooling, and fully connected layers for effective pattern recognition. Also CNNs are ideal for handling sequential data like ECG signals, automatically learning relevant features.

- **Convolutional Layers**:
  - 128 filters with a kernel size of 11
  - 64 filters with a kernel size of 3
  - ReLU activation
- **Pooling Layers**:
  - Max pooling with a pool size of 3 and strides of 2
- **Fully Connected Layers**:
  - 64 and 32 units in dense layers
- **Output Layer**:
  - Softmax activation with 5 units (classifies into 5 categories: N, S, V, F, Q)

4. Evaluation

To evaluate the performance of the model, I used the following metrics:

- **Accuracy**: The proportion of correctly classified samples to the total number of samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision**: The proportion of true positive samples to the total number of samples predicted as positive.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall**: The proportion of true positive samples to the total number of positive samples.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1-score**: The harmonic mean of precision and recall, providing a balance between the two metrics.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Confusion matrix**: A table showing the number of true positive, true negative, false positive, and false negative samples for each class.
- **ROC curve**: A graphical representation of the true positive rate (sensitivity) versus the false positive rate (1-specificity) for different threshold values.
- **Precision-recall curve**: A graphical representation of precision versus recall for different threshold values.

Class	**Accuracy**	**Precision**	**Recall**	**F1-score**	**ROC AUC**
N	**0.99**	0.98	0.99	0.99	0.99
S	**0.68**	0.84	0.75	0.85	0.97
V	**0.94**	0.96	0.95	0.98	1.00
F	**0.70**	0.84	0.76	0.84	0.99
Q	**0.99**	0.99	0.99	0.99	1.00

Table 4.1: Model performance metrics

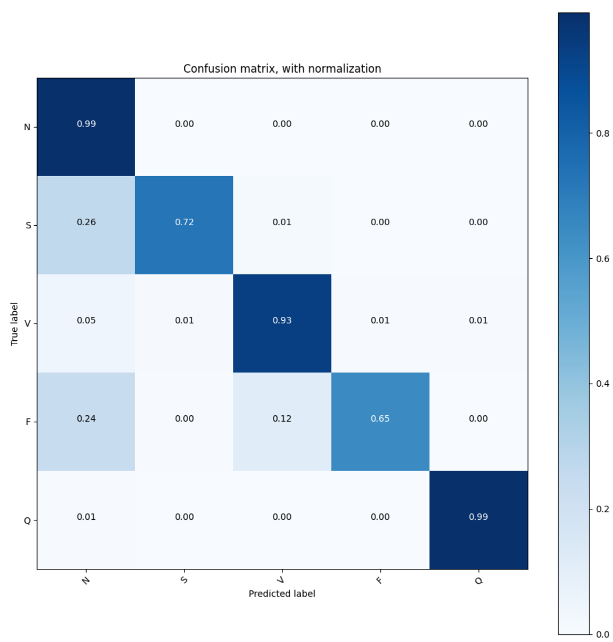

Figure 4.1: Confusion matrix

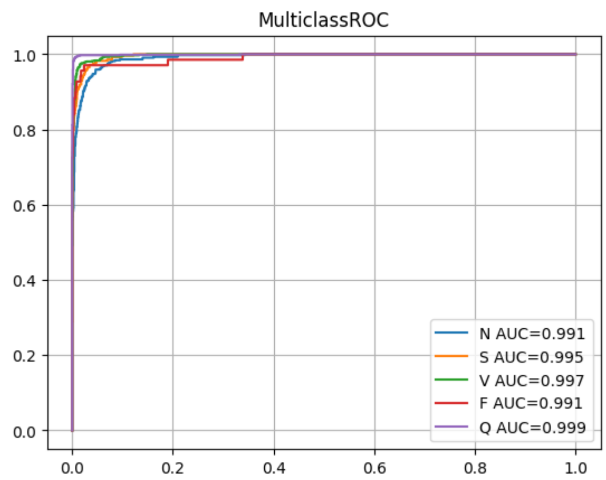

Figure 4.2: ROC curve

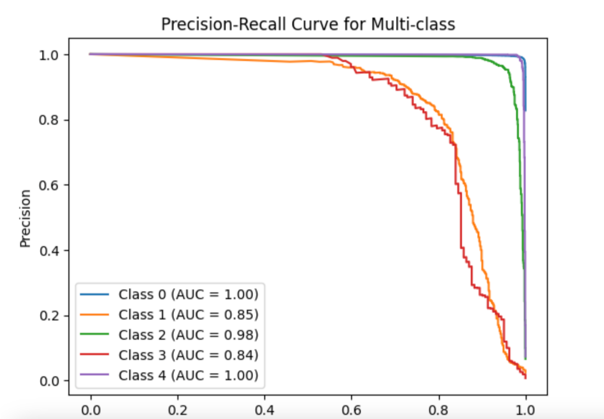

Figure 4.3: Precision-recall curve

5. Discussion

High Accuracy: The model achieved high accuracy, especially for the minority classes ('F' and 'Q'). This is because the ECG signals for these classes have clear and distinct patterns, making it easier for the model to recognize them.

Class Imbalance Issue: The dataset is imbalanced, with most samples belonging to the 'N' class (normal heartbeats). This could have made the model biased towards predicting 'N'. To address this, we used under-sampling to balance the classes, but this reduced the number of 'N' samples, potentially causing the model to miss some important details from normal heartbeats.

Simple Preprocessing: The preprocessing steps were simple, such as normalizing and extracting R-peaks from the ECG signals. This worked well for the dataset because the signals were relatively clean and had clear patterns. However, in real-world data, ECG signals often have noise and more variation, which might affect the model's performance.

Distinct Signal Patterns: Classes like 'Q' had high performance because their ECG signals have unique and recognizable shapes, which made them easier to classify. Other classes, like 'S' and 'V', might need more fine-tuning or more data to improve accuracy.

Future work:

Test the model on more diverse and real-world ECG data to see how well it generalizes.

Experiment with other techniques like over-sampling or more complex preprocessing methods to handle noisy signals and data imbalance better.

6. Conclusion

This study presents a deep learning-based ECG heartbeat classification approach using a Convolutional Neural Network (CNN). The model demonstrated high accuracy, precision, recall, and F1-score for all classes, along with strong performance in the ROC and precision-recall curves, showcasing its ability to distinguish between classes and balance precision and recall. These results suggest CNNs are effective for ECG classification, with future work focused on optimizing the model and exploring other deep learning techniques.

7. References

[1] Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh. "ECG Heartbeat Classification: A Deep Transferable Representation." arXiv preprint arXiv:1805.00794 (2018).