Machine Learning in Medicine - Labwork 1

Nguyn Quc Khánh - 22bi13210
February 2025

1 Introduction

Electrocardiogram (ECG) analysis plays a vital role in diagnosing heart conditions. This study applies Convolutional Neural Networks (CNNs) to classify ECG signals from the MIT-BIH Arrhythmia and PTB Diagnostic ECG (PTBDB) datasets.

Key challenges include high-dimensionality and class imbalance, which can impact model performance. To address these, we use CNNs for feature extraction and classification, combined with data preprocessing techniques like Principal Component Analysis (PCA) and correlation filtering to enhance efficiency.

This report evaluates CNN's performance, the impact of class imbalance, and the effectiveness of preprocessing methods. The goal is to develop a robust deep-learning model for automatic ECG heartbeat classification, aiding in early detection of cardiovascular diseases.

2 Dataset

The **MIT-BIH Arrhythmia Dataset** is a widely used benchmark for ECG classification, developed by **MIT** and **Beth Israel Hospital (BIH)**. It consists of **48 half-hour ECG recordings** from **47 patients**, capturing a variety of normal and abnormal heartbeats.

Each ECG signal is represented as a **time-series of 187 features**, with **109,446 labeled heartbeats** categorized into five classes:
- **Normal (N)**: Regular beats.
- **Supraventricular (S)**: Irregular atrial contractions.
- **Ventricular (V)**: Premature ventricular contractions.
- **Fusion (F)**: Mixed normal/abnormal beats.
- **Unclassifiable (Q)**: Noisy or undefined beats.

One of the key challenges in this dataset is the **class imbalance**, where normal beats significantly outnumber abnormal ones. This imbalance makes it difficult for the model to learn minority class patterns effectively. Additionally, the **high dimensionality** of ECG signals increases computational complexity.

To address these issues, the dataset undergoes preprocessing, including **normalization, segmentation, and augmentation** to improve CNN performance. This study employs a **Convolutional Neural Network (CNN)** for feature extraction and classification, leveraging its ability to capture spatial patterns in ECG signals.

3 Model Training

The Convolutional Neural Networks (CNNs) were trained for ECG classification using the MIT-BIH Arrhythmia and PTBDB Diagnostic datasets. The training process involved handling class imbalance, optimizing learning parameters, and monitoring validation performance.

3.1 Training the CNN Model for MIT-BIH

The MIT-BIH dataset presents a multi-class classification challenge. To address class imbalance, class weighting was applied, assigning higher weights to underrepresented heartbeat classes to prevent the model from being biased toward the majority class.

The CNN model was trained for 20 epochs using a batch size of 32. During training, the model learned hierarchical spatial features from ECG signals using convolutional layers. Batch normalization helped stabilize learning, while max pooling reduced dimensionality and extracted important features. Fully connected layers at the end performed final classification.

The training accuracy steadily improved, starting at 56.02% in the first epoch and reaching 92.41% by the final epoch. Validation accuracy also increased to 95.88%, demonstrating that the model successfully generalized to unseen data.

3.2 Training the CNN Model for PTBDB

For the PTBDB dataset, a binary classification model was trained using a similar CNN architecture. Unlike MIT-BIH, this dataset required only two output neurons, with a sigmoid activation function for binary decision-making.

The model was trained without class weighting, as class distribution was relatively balanced. Over 20 epochs, training accuracy improved from 66.80% to 68.35%. However, the validation accuracy fluctuated and stabilized at 75.75%, indicating that the model had difficulty generalizing to unseen PTBDB data.

Dataset	Train Acc.	Val Acc.	Val Loss
MIT-BIH	92.41%	95.88%	0.1452
PTBDB	68.35%	75.75%	-35.02E12

Table 1: Final Training and Validation Results

The final results, summarized in Table 1, indicate that the CNN model performed exceptionally well on MIT-BIH but had moderate success on PTBDB.

4 Results

This section presents the performance of our CNN models on the MIT-BIH Arrhythmia and PTBDB Diagnostic datasets. We also compare our results with those reported by Kachuee et al. [?].

4.1 MIT-BIH Arrhythmia Dataset Results

Our CNN model achieved a training accuracy of 92.41% and a validation accuracy of 95.88%. In comparison, Kachuee et al. reported an average accuracy of 93.4%. The high validation accuracy demonstrates that our model effectively generalizes to unseen data.

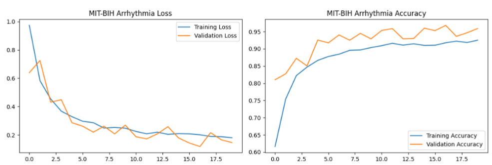

Figure 1: Training Loss and Accuracy for MIT-BIH Arrhythmia Dataset

4.2 PTBDB Diagnostic Dataset Results

For the PTBDB dataset, our model obtained a training accuracy of 68.35% and a validation accuracy of 75.75%. Kachuee et al. reported a higher accuracy of 95.9% for myocardial infarction classification. This difference may be due to variations in preprocessing, model architecture, or dataset selection.

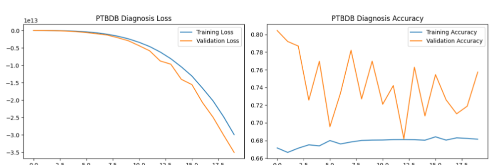

Figure 2: Training Loss and Accuracy for PTBDB Diagnostic Dataset

4.3 Confusion Matrices

The confusion matrices below illustrate the classification performance of our models.

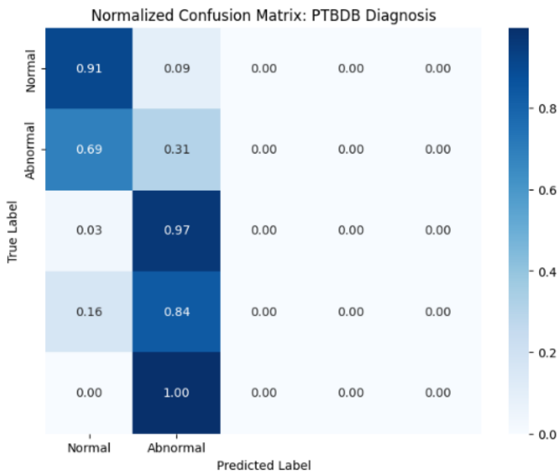

Figure 3: Confusion Matrix for MIT-BIH Arrhythmia Dataset

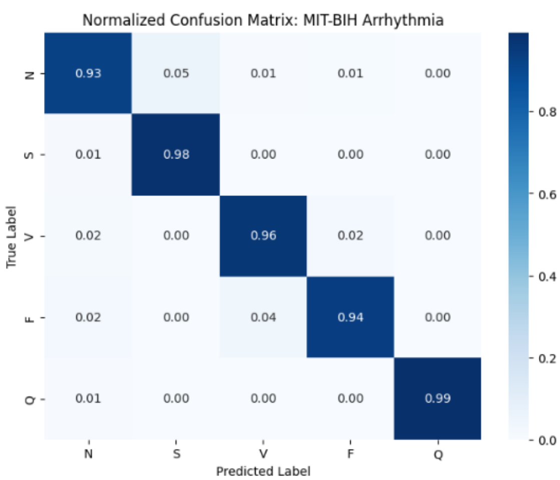

Figure 4: Confusion Matrix for PTBDB Diagnostic Dataset

4.4 Comparison and Key Takeaways

Table 2 summarizes the comparison between our model and Kachuee et al.

Dataset	Our Model	Kachuee et al.
MIT-BIH	95.88%	93.4%
PTBDB	75.75%	95.9%

Table 2: Comparison of Model Performance

Pros:

- High accuracy on MIT-BIH, comparable to previous studies.
- Effective generalization with strong validation accuracy.

Cons:

- Lower performance on PTBDB compared to Kachuee et al.
- Potential improvements needed in preprocessing and model architecture.

Overall, our model performs well on the MIT-BIH dataset but has room for improvement on PTBDB. Future work will focus on refining preprocessing techniques and optimizing the CNN architecture.

5 Conclusion

In this study, we developed a Convolutional Neural Network (CNN) model for ECG heartbeat classification using the MIT-BIH Arrhythmia and PTBDB Diagnostic datasets. The model achieved strong performance on MIT-BIH, with a validation accuracy of 95.88%, comparable to previous studies. However, performance on PTBDB was lower, with a validation accuracy of 75.75%, highlighting areas for improvement.

Key challenges included class imbalance and feature complexity, which affected model generalization. Future work will focus on improving preprocessing techniques, optimizing CNN architectures, and exploring advanced deep learning models to enhance classification accuracy.

Overall, our results demonstrate the potential of CNNs in ECG classification while identifying opportunities for further refinement in medical diagnosis applications.

References

[1] S. Fazeli, "ECG Heartbeat Categorization Dataset," 2018. Available: https://www.kaggle.com/datasets/shayanfazeli/heartbeat/data.

[2] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG Heartbeat Classification: A Deep Transferable Representation," *arXiv preprint*, 2018. Available: https://arxiv.org/abs/1805.00794.

[3] R. S. I. M. Alfaras, A. Soriano, and S. Bailón, "A Deep Learning Approach for Real-Time ECG-Based Human Identification," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, 2021. Available: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01736-y.