Department of Information and Communication Technology

ML IN MEDICAL REPORT

Practice 1

ECG Heartbeat Categorization

22BI13352 - Vu Hoang Mai Nhi - Data Science

Lecturer:
Tran Giang Son

Academic year: 2022 - 2025
Hanoi, February 2025

1 Introduction

Electrocardiogram (ECG) signals are essential for identifying cardiovascular disease as they record the electrical activity of the heart. The precise classification of ECG heartbeats is crucial for the early identification of arrhythmias and other cardiac irregularities. Conventional procedures for identifying heartbeats depend on experienced medical professionals visually examining the ECG data, which are long, arbitrary, and prone to inaccuracies. Consequently, automated heartbeat categorization approaches have been created to increase the precision, effectiveness, and consistency of ECG analysis.

In this report, I employ two machine learning techniques, XGBoost and convolutional neural networks (CNNs), to categorize ECG heartbeats. XGBoost, a powerful gradient boosting technique, is applied for its speed in processing structured data, while CNNs apply deep learning to uncover spatial patterns from ECG waveforms. By evaluating the performance of different models, I seek to evaluate their usefulness in improving automated heartbeat categorization, ultimately contributing to more reliable and efficient clinical diagnosis.

2 Background

2.1 ECG

An electrocardiogram (ECG) is a noninvasive diagnostic technique that measures the heart's electrical activity over time. It consists of waveforms that indicate different stages of the cardiac cycle, including the P wave, QRS complex, and T wave. Analyzing ECG data is critical for diagnosing arrhythmias, ischemia, and other cardiovascular problems. However, manual interpretation of ECG data can be time-consuming and prone to mistakes. Machine learning (ML) techniques have been progressively employed to automate ECG categorization, boosting diagnostic accuracy and efficiency.

2.2 XGBoost

XGBoost (Extreme Gradient Boosting) is an enhanced gradient boosting technique extensively used for classification jobs due to its speed, scalability, and strong predictive performance. It generates an ensemble of decision trees using gradient-boosting techniques, reducing mistakes by repeatedly upgrading weak learners. XGBoost is highly good at processing structured data with tabular information, making it suited for ECG classification when variables such as heart rate variability and waveform characteristics are retrieved and utilized as inputs.

2.3 CNNs

Convolutional Neural Networks (CNNs) are a family of deep learning models developed for processing grid-like input, such as pictures and time-series signals.

In ECG classification, CNNs automatically learn spatial and temporal characteristics from raw waveform data without needing manual feature extraction. By using convolutional layers, pooling procedures, and fully connected layers, CNNs may successfully capture patterns in ECG data that correlate to distinct heartbeat types. This ability makes CNNs particularly effective for ECG analysis, as they can learn complicated representations straight from the data, enhancing classification accuracy.

This work employs both XGBoost and CNNs to evaluate their performance in ECG heartbeat classification and analyze their potential for enhancing automated diagnosis in the medical profession.

3 Dataset

In this practice, I use ECG Heartbeat Categorization Dataset from Kaggle as data source for labeled ECG records for practicing classification tasks purpose. This dataset consists of 5 types of heartbeats: Normal (N), Supraventricular (S), Ventricular (V), Fusion (F) and Unknown (Q), which are encoded to 0, 1, 2, 3 and 4 respectively. Each data contains 187 values of the heartbeat signal.

- Number of Samples: 109446
- Number of Categories: 5
- Sampling Frequency: 125Hz
- Data Source: Physionet's MIT-BIH Arrhythmia Dataset "cite"
- Classes: ['N': 0, 'S': 1, 'V': 2, 'F': 3, 'Q': 4]

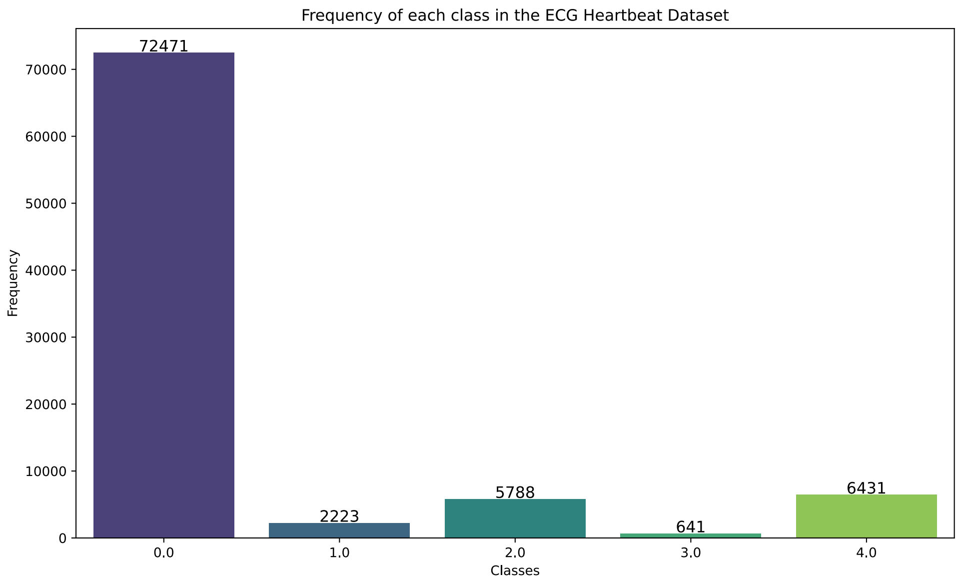

Figure 1: Frequency of each class in the ECG Heartbeat Dataset

As we can see from the histogram in Figure 1, the class distribution in the ECG heartbeat training dataset from Kaggle is severely skewed. Class 0 leads the dataset with 72,471 samples, substantially outnumbering the other classes. Class 1, Class 2, Class 3, and Class 4 have substantially fewer samples, with 2,223, 5,788, 641, and 6,431 occurrences, respectively. This imbalance may influence model performance, necessitating approaches such as resampling or weighted loss functions to increase classification accuracy.

To fix the class imbalance in the ECG heartbeat dataset and reduce the threat of overfitting, I applied the Synthetic Minority Over-sampling Technique (SMOTE) to balance the classes.

SMOTE produces synthetic samples for minority classes by interpolating between existing instances, considerably boosting their representation without duplicating data. Unlike random oversampling, which may lead to overfitting by reproducing previous samples, or undersampling, which may result in loss of essential information, SMOTE helps establish a more general decision boundary. This feature assures the model learns substantial patterns from all classes, enhancing its capacity to classify minority heartbeat types appropriately.

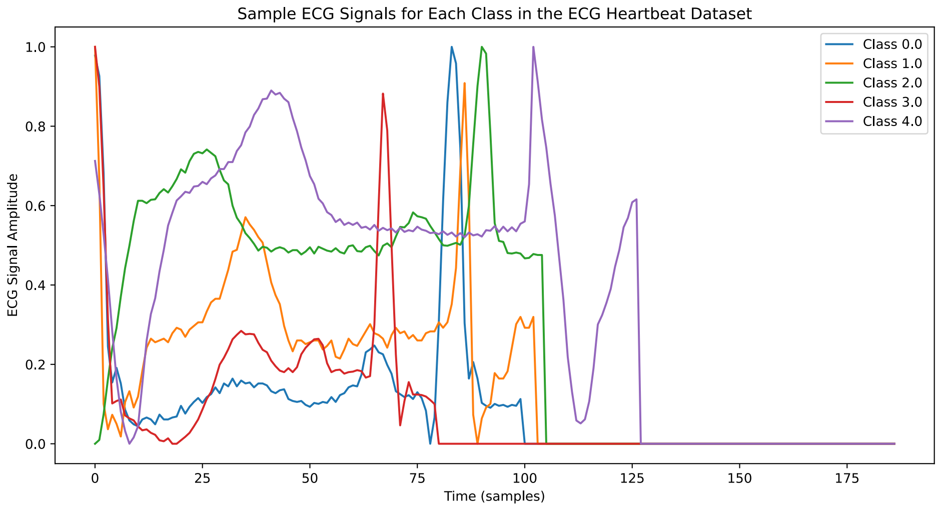

Figure 2: Sample ECG Signals for Each Class

After resampling using SMOTE, the sample ECG signals for each class in the MIT-BIH Arrhythmia Dataset show distinct waveform patterns for different heartbeat categories. Each class exhibits unique characteristics in terms of amplitude and shape, reflecting the variations in cardiac activity. The resampling process ensures a balanced representation of all classes, allowing the model to learn from diverse ECG patterns. However, it is important to validate that the synthetic samples generated by SMOTE retain meaningful physiological features to avoid introducing unrealistic patterns that could affect model performance.

4 Model

4.1 XGBoost

4.1.1 Method

In the XGBoost classification, the dataset was initially standardized with **StandardScaler** to guarantee that all features possess a standard distribution, enhancing model convergence. The XGBoost classifier was subsequently trained using a multi-class softmax goal to predict one of the five ECG heartbeat classifications. The model's performance was assessed using **log loss**, while training progress was tracked via an evaluation set. Following training, predictions were conducted on the test dataset, resulting in a classification accuracy of **X%**. The outcomes were subsequently examined using a classification report, yielding precision, recall, and F1-score insights for each category.

4.1.2 Evaluation

Class	**Precision**	**Recall**	**F1-score**
N	0.99	0.98	0.98
S	0.62	0.80	0.70
V	0.93	0.95	0.94
F	0.59	0.85	0.70
Q	0.98	0.98	0.98
Accuracy			**0.97**

Table 1: XGBoost Model performance metrics

The XGBoost classifier scored an **overall test accuracy of 96.84%**, exhibiting high performance in ECG heartbeat classification. The precision, recall, and F1-score for each class give more insights into the model's efficacy:

- **Class N and Class Q** had high precision and recall (98-99%), indicating the model can accurately identify and classify regular and specific abnormal heartbeats with minimal misclassification.
- **Class V** also performed well, with an F1 Score of 0.94, meaning the model can effectively detect this heartbeat type.
- **Class S and Class F**, representing minority categories, showed lower performance, with F1-scores of 0.70. Despite SMOTE balancing the dataset, these classes remain more challenging to classify, likely due to overlapping features or limited distinguishing characteristics.

- The **macro average F1 Score (0.86)** suggests that performance varies across classes, while the **weighted average F1 Score (0.97)** confirms that the model performs well overall, benefiting from the majority class influence.

While XGBoost demonstrates high accuracy, improvements such as fine-tuning hyperparameters, using feature engineering, or applying more advanced data augmentation techniques may further enhance classification, particularly for minority classes.

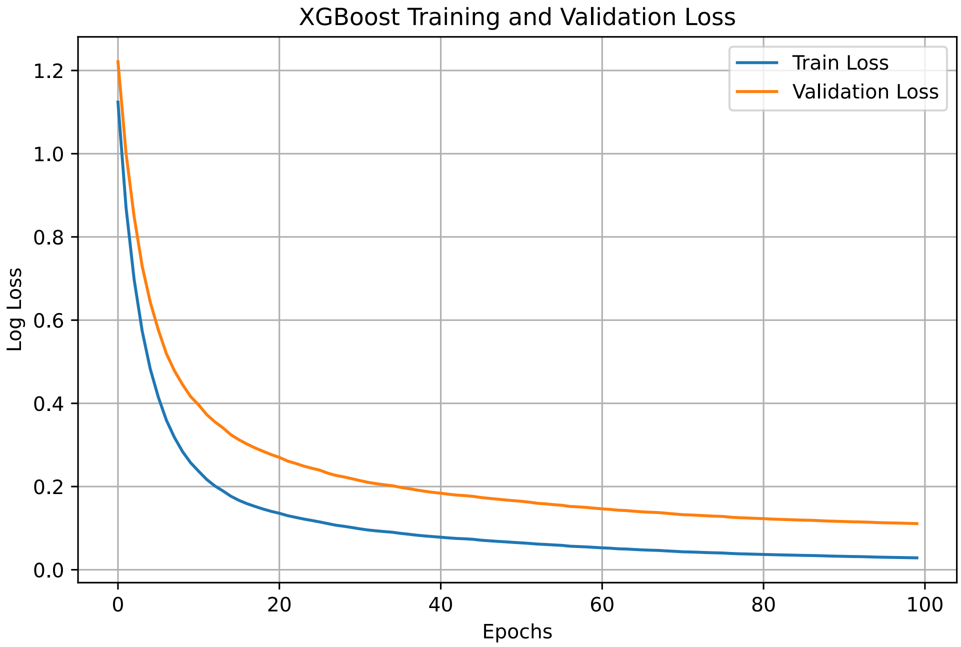

Figure 3: XGBoost Model Loss

The training loss for the XGBoost model declines consistently during the training phase, signifying that the model is proficiently acquiring patterns from the data. However, the validation loss, while decreasing, does so at a slower rate. This decrease indicates that the model may not generalize to unknown data as effectively as to the training set. The increasing disparity between training and validation loss raises concerns regarding potential overfitting when the model gets too tailored to the training data and fails to perform comparably on novel samples. Overfitting can lead to high accuracy on the training set but poor performance when meeting real-world ECG data. To alleviate this problem, we should investigate strategies such as hyperparameter adjustment, early halting, dropout regularization, and boosting data diversity through augmentation or further preprocessing to enhance generalization.

4.2 CNNs

4.2.1 Method

The CNN model was constructed with numerous convolutional and pooling layers to extract significant information from ECG data while integrating dropout layers to avoid overfitting. The model was trained using the Adam optimizer with a learning rate of 0.0001 and categorical cross-entropy loss. Early stopping was adopted to increase generalization, minimizing wasteful training after validation loss stopped improving. The input data was reshaped to satisfy CNN criteria, and labels were one-hot encoded for multi-class classification. After training for up to 20 epochs, the model showed high performance, demonstrating its capacity to successfully understand complicated patterns in ECG data.

4.2.2 Evaluation

Class	**Precision**	**Recall**	**F1-score**
N	0.99	0.97	0.98
S	0.57	0.84	0.68
V	0.93	0.95	0.94
F	0.60	0.87	0.71
Q	0.98	0.99	0.98
Accuracy			**0.97**

Table 2: CNN Model performance metrics

The CNN model scored an exceptional **test accuracy of 96.72%**, suggesting high overall performance. The accuracy, recall, and F1-score values are substantial across most classes, notably for class N and class Q, where the model displays remarkable classification performance. The **macro average F1-score of 0.86** implies that the model maintains a reasonable balance across all classes despite the class imbalance.

For other classes, the model demonstrates considerable gains in recall, notably for **class S (84%)** and **class F (87%)**, indicating it effectively recognizes a more significant proportion of these samples than XGBoost. However, the accuracy for **class S (57%)** remains lower, indicating some misclassification into other groups. Despite this, the model performs better in a recall, which is vital in medical applications because detecting all cases of a condition is more important than eliminating false positives.

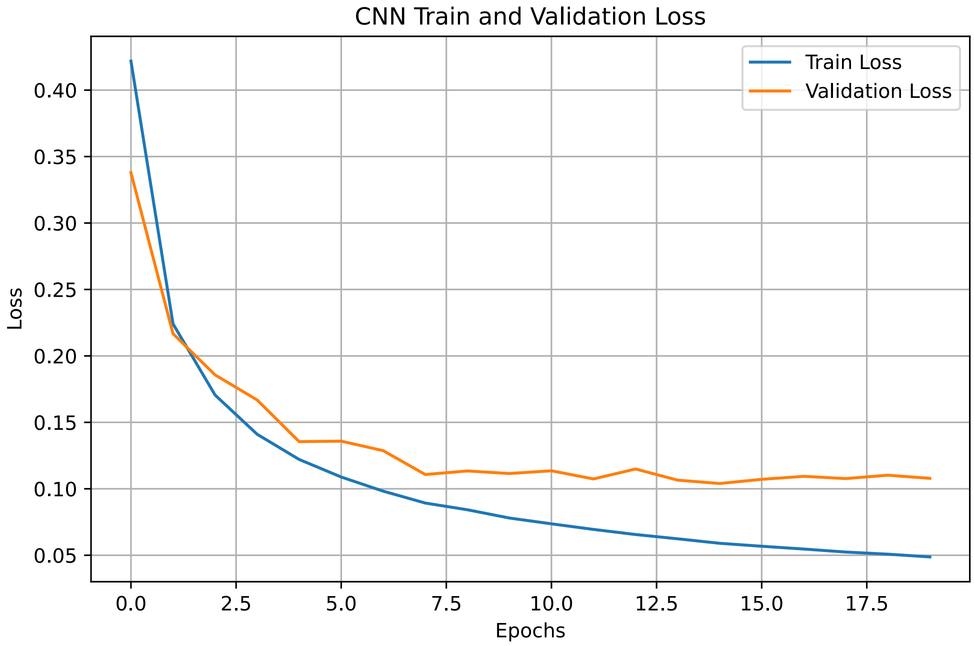

Figure 4: CNN Model Loss

The training loss for the CNN model continuously lowers across epochs, demonstrating that the model is effectively learning from the data, suggesting that the convolutional layers successfully collect essential properties in the ECG signals. The validation loss initially follows a similar declining pattern but swings slightly after a few epochs. These oscillations may indicate minor overfitting when the model gets overly specialized in the training data and fails to maintain consistent performance on unseen samples. However, dropout layers and early stopping help avoid severe overfitting, guaranteeing that the model does not continue training once performance declines. Despite modest changes, the overall trend of diminishing loss implies that the model generalizes well to new data, giving it a solid strategy for ECG heartbeat categorization.

4.3 Comparision

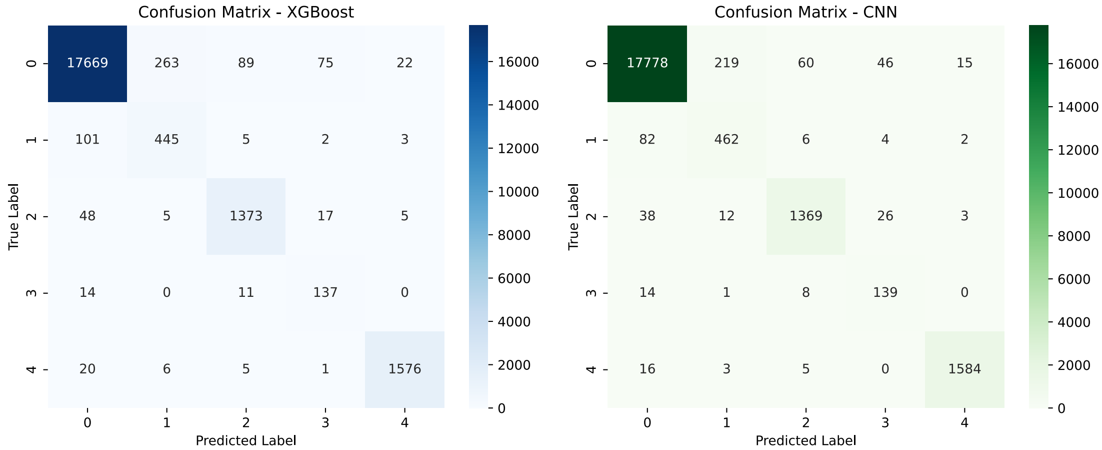

Figure 5: Confusion Matrices of two models

Both models exhibit good performance, with many correctly identified samples for **class N**, the majority class. However, CNN demonstrates somewhat improved classification accuracy across most classes, as demonstrated in the lower number of misclassified samples compared to XGBoost. Specifically, CNN exhibits fewer misclassifications in **class N** and **class V**, showing a more substantial capacity to identify these groups: **class S** and **class F** display misclassification in both models. However, CNN shows a slight improvement in differentiating **class S**.

5 Conclusion

In this ECG classification assignment, both CNN and XGBoost demonstrated outstanding accuracy, with CNN reaching 96.72% and XGBoost earning 96.84%. While their overall performance is comparable, considerable disparities exist in how they approach minority classes and generalize the data.

XGBoost fared well in terms of accuracy, delivering less erroneous optimistic projections. However, it struggled with memory for many minority classes, such as class S and class F, indicating that it missed some of these requirements. On the other hand, CNN displayed stronger recall, particularly for minority groups, showing it is better at detecting all instances of specific arrhythmias, even if some misclassifications occur. This increased memory is crucial in medical applications, where overlooking a valid case may have devastating effects.

Additionally, the confusion matrices demonstrate that CNN decreases misclassification for minority classes relative to XGBoost. This rise is likely related to CNN's propensity to capture spatial and temporal patterns in ECG data, making it well-suited for time-series classification tasks. However, CNN also indicated more significant fluctuations in validation loss, indicating a need for careful tuning to limit overfitting.

Given the medical context, where detecting all possible arrhythmias is crucial, CNN is suggested due to its improved recall performance for minority classes.

References

[1] Kachuee, M., Fazeli, S., Sarrafzadeh, M. (2018). ECG Heartbeat Classification: A deep transferable representation (pp. 443–444). https://doi.org/10.48550/arXiv.1805.00794

[2] ECG Heartbeat Categorization Dataset. (2018, May 31). Kaggle. ttps://www.kaggle.com/datasets/shayanfazeli/heartbeat/data