ECG Heartbeat Classification using a Convolutional Neural Network

Nguyen Hoang Tung
Student ID: 22BA13318
University of Science and Technology of Ha Noi (USTH)
Email: tungtimo0808@gmail.com

Abstract—**This work studies ECG heartbeat classification using deep learning on the PTB Diagnostic ECG dataset. A 1D CNN is used to classify heartbeats into normal and abnormal classes. The model achieves about 92% accuracy and an AUC of 0.98, showing strong performance.**

I. EXPLORATORY DATA ANALYSIS

A. Dataset Overview

In this study, we use the PTB Diagnostic ECG dataset for ECG heartbeat classification. The dataset is provided in two separate files: one for normal heartbeats and one for abnormal heartbeats.

Each file contains ECG signals with the following structure:

- Each row represents one ECG heartbeat.
- Each column represents one time step of the ECG signal.
- There are 188 signal columns and 1 label column.

Therefore, each sample has a total of 189 columns:

- 188 values corresponding to the ECG signal.
- 1 value indicating the class label.

After loading the data:

- The normal dataset contains approximately 10,500 samples.
- The abnormal dataset contains approximately 4,000 samples.
- In total, the dataset consists of about 14,500 ECG heartbeats.
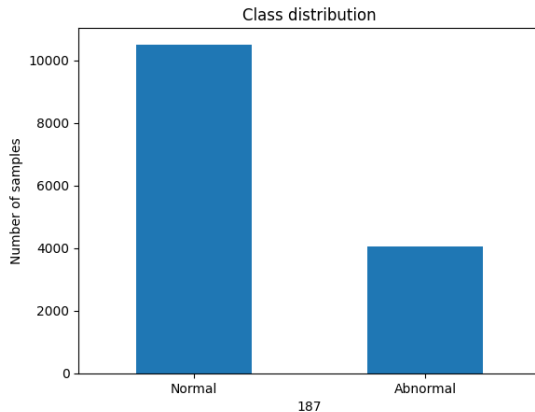
B. Class Distribution

Fig. 1. Class distribution of the PTB Diagnostic ECG dataset.

Figure 1 shows the number of samples in each class in the PTB Diagnostic ECG dataset.

From the figure, it can be observed that the dataset is imbalanced:

- Normal class: approximately 10,500 samples.
- Abnormal class: approximately 4,000 samples.

This class imbalance is an important issue to consider when training and evaluating the classification model, as it may bias the model toward the majority class if not properly handled.
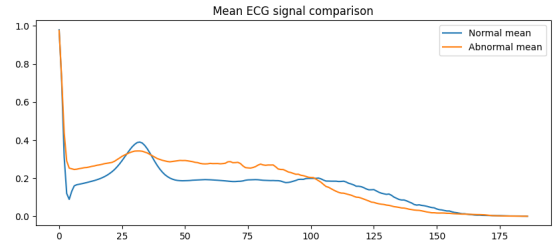
C. Mean ECG Signal Comparison

Fig. 2. Mean ECG signal comparison between normal and abnormal classes.

Figure 2 shows the average ECG signal for both normal and abnormal classes.

Although both signals share a similar overall shape, there are clear differences in amplitude and waveform in several regions of the signal. This indicates that the raw ECG signals contain discriminative information that can be effectively exploited for classification.

D. Conclusion and Insights from EDA

From the exploratory data analysis, we can conclude that:

- The dataset consists of 1D ECG signals with 188 time steps.
- The total number of samples is about 14,500.
- The dataset is imbalanced, with many more normal samples than abnormal ones.
- There are visible differences between the average normal and abnormal ECG signals.

Based on these observations, we can draw several important insights:

- **Class imbalance is an important challenge.** Because the number of normal samples is much larger, a simple

model may be biased toward predicting the normal class. Therefore, accuracy alone is not enough, and we also need to look at recall and F1-score, especially for the abnormal class.

- **Raw ECG signals already contain useful patterns.** The difference between the mean signals shows that the waveform shape and amplitude already carry important information. This means that the model does not need handcrafted features and can learn directly from raw signals.
- **Local patterns in time are important.** The ECG signal contains sharp peaks and local changes such as the QRS complex. This suggests that convolutional layers are suitable for capturing these local patterns.
- **The problem is suitable for deep learning.** The signal is long (188 time steps) and has complex shapes. This makes the problem difficult for simple methods, but suitable for deep learning models such as CNN.
- **High recall for the abnormal class is important.** In real medical applications, missing an abnormal case is more dangerous than a false alarm. Therefore, during evaluation, we should focus on recall for the abnormal class, not only overall accuracy.

These insights confirm that the dataset is suitable for training a deep learning model and also support the choice of a CNN-based architecture and the use of medical-oriented evaluation metrics.

II. MODEL IMPLEMENTATION

In this work, we implement a 1D Convolutional Neural Network (1D CNN) for binary classification of ECG heartbeats into normal and abnormal classes. The model is implemented in Python using TensorFlow and Keras.

A. Data Splitting and Input Format

After loading and merging the normal and abnormal datasets, the data is shuffled and split into training and test sets using an 80/20 ratio. Stratified splitting is applied to keep the same class distribution in both sets. The test set is kept completely separate and is only used for the final evaluation. The input data is reshape from $(N, 188)$ to $(N, 188, 1)$ in order to be compatible with 1D convolution layers, where where N is the number of samples and 188 is the number of time steps.

B. Handling Class Imbalance

Because the dataset is imbalanced, class weights are computed using the balanced strategy. These weights are used during training to reduce the bias toward the majority class and help the model learn both classes more fairly.

C. Network Architecture

The proposed 1D CNN architecture consists of the following components:

- A first Conv1D layer with 32 filters to extract low-level features from the ECG signal.
- A MaxPooling1D layer to reduce the temporal resolution.

- A second Conv1D layer with 64 filters followed by another MaxPooling1D layer.
- A third Conv1D layer with 128 filters to learn higher-level representations.
- A Global Average Pooling layer to convert feature maps into a fixed-length feature vector.
- A fully connected Dense layer with 64 units and ReLU activation.
- A Dropout layer to reduce overfitting.
- A final Dense layer with a Sigmoid activation function to output the probability of the abnormal class.

The total number of trainable parameters in the model is approximately 43,000, which makes the network lightweight and efficient.

D. Training Configuration

The model is trained using the Adam optimizer and the binary cross-entropy loss function, which is suitable for binary classification problems. The batch size is set to 128 and the maximum number of epochs is 30. During training, 20% of the training data is used as a validation set for monitoring the training process.

Early stopping is applied based on the validation AUC score. The training process is stopped if the validation AUC does not improve for 5 consecutive epochs, and the best model weights are restored. The validation set is only used for model selection, while the final performance is reported on the independent test set.

E. Prediction and Decision Rule

After training, the model outputs a probability score for each sample in the test set. A threshold of 0.5 is used to convert this probability into a class label: if the output is greater than 0.5, the sample is classified as abnormal; otherwise, it is classified as normal.

III. EXPERIMENTAL RESULTS

A. Classification Performance on the Test Set

The performance of the proposed model is evaluated on the independent test set using standard metrics including precision, recall, F1-score, and accuracy. Table I summarizes the classification results, while Table **??** shows the confusion matrix of the model.

TABLE I
CLASSIFICATION REPORT ON THE TEST SET.

Class	Precision	Recall	F1-score	Support
Normal (0)	0.79	0.95	0.86	809
Abnormal (1)	0.98	0.90	0.94	2102
Accuracy			0.92	2911
Macro Avg	0.89	0.93	0.90	2911
Weighted Avg	0.93	0.92	0.92	2911

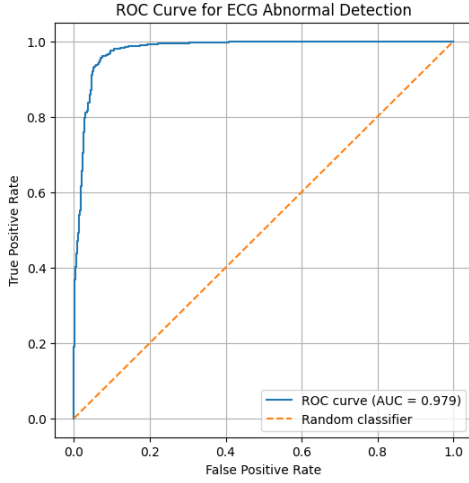## B. Evaluation on the Test Set Using ROC Curve

Fig. 3. ROC curve of the proposed 1D CNN model on the test set.

In addition to standard classification metrics, the Receiver Operating Characteristic (ROC) curve is used to evaluate the model performance on the test set. The ROC curve is plotted using the predicted probabilities of the test samples, and the Area Under the Curve (AUC) is computed. As shown in Figure 3, the proposed model achieves an AUC of approximately 0.98, which indicates a strong ability to distinguish between normal and abnormal ECG heartbeats.

IV. COMPARISON

In this section, we compare the performance of our proposed method with a reference approach reported in the literature. We use the work of Kachuee et al. (2018), which also evaluates ECG classification on the PTB Diagnostic ECG dataset, as the main reference.

A. Quantitative Comparison

Table II shows the comparison between our method and the reference method in terms of accuracy, precision, and recall.

TABLE II
COMPARISON WITH THE REFERENCE METHOD ON THE PTB DATASET.

Method	Accuracy (%)	Precision (%)	Recall (%)
Kachuee et al. (2018)	95.9	95.2	95.1
Proposed CNN (ours)	92.0	98.0	90.0

From the table, it can be seen that our method achieves a lower overall accuracy compared to the reference method. However, our model reaches a very high precision of 98% for the abnormal class, which means that most samples predicted as abnormal are indeed correct. The recall of the abnormal class is 90%, which is lower than the 95.1% reported in the reference method.

B. Discussion

The difference in performance can be explained by the difference in model complexity and training strategy. The method proposed by Kachuee et al. uses a deeper convolutional neural network with residual connections and applies transfer learning from a large arrhythmia dataset. In contrast, our model is a lightweight 1D CNN with about 43k parameters and is trained directly on the PTB dataset without any pretraining.

Although our method does not outperform the state-of-the-art approach, it still achieves competitive performance with a much simpler architecture. This shows that even a relatively small CNN trained on raw ECG signals can provide good classification results.

Moreover, considering the high AUC value (approximately 0.98) achieved by our model, the ranking ability of the classifier is very strong. The lower accuracy and recall are mainly due to the fixed decision threshold (0.5), which could be further optimized in future work.

Overall, this comparison confirms that the proposed method provides a good trade-off between performance and model complexity.

V. CONCLUSION

In this work, we studied the problem of ECG heartbeat classification using a deep learning approach. We used the PTB Diagnostic ECG dataset to classify heartbeats into normal and abnormal classes and proposed a lightweight 1D Convolutional Neural Network trained directly on raw ECG signals.

Experimental results on the independent test set show that the proposed model achieves an accuracy of 92% and an AUC of approximately 0.98. The high AUC value indicates that the model has a strong ability to distinguish between normal and abnormal heartbeats, while the classification results confirm that most abnormal cases are correctly detected.

We also compared our method with a state-of-the-art approach and showed that although our model does not outperform more complex deep architectures, it still achieves competitive performance with a much simpler and more efficient structure. This makes the proposed approach suitable for practical applications where computational efficiency is important.

In future work, the performance can be further improved by using deeper architectures, applying transfer learning, or optimizing the decision threshold to increase the recall for abnormal cases.

REFERENCES

[1] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG heartbeat classification: A deep transferable representation," *Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI)*, 2018, pp. 443–444.
[2] S. Fazeli, "ECG Heartbeat Categorization Dataset," Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/shayanfazeli/heartbeat