Binary Classification of Cardiac Pathologies using Deep Learning: A PTB-XL Dataset Approach

Mohamed Khalil Chaabani [1,2][0009-0006-1813-6944], Nathan Guerreiro[1,3][0009-0007-2879-2248], Luiz Ribeiro[1,3][0009-0002-2452-7264], Luiz E. Luiz[1,3][0000-0001-9305-091X], Mohamed Aymen Slim[2][0009-0003-3369-5398], João Paulo Teixeira[1][0000-0002-6679-5702]

[1] CeDRI, SusTEC, Instituto Politécnico de Bragança, 5300-253 Bragança, Portugal.
{a60546, a61448, a58865}@alunos.ipb.pt
{luiz.luiz, joaopt}@ipb.pt

[2] Private Higher Polytechnic Institute, Private University of Tunis, 32 Bis, 1002 Avenue Kheireddine Pacha, Tunis, Tunisia.
khalil.chaabani@yahoo.com
mohamed-aymen.slim@ult-tunisie.com

[3] Universidade Tecnologica Federal do Paraná, 80230-901 Curitiba, Brasil.
{nathanguerreiro, luizotavioribeiro, lluiz}@utfpr.edu.br

Abstract. Cardiovascular diseases, such as myocardial infarction, are among the leading causes of death worldwide. Accuracy and time are crucial for diagnosing these conditions and for effective treatment, usually requiring time-consuming manual analysis of clinical-grade electrocardiogram (ECG). This paper presents a novel deep learning-based method for binary classification of cardiac pathologies using the PTB-XL dataset. The model integrates EfficientNetB3 for spatial feature extraction and a Linformer block to capture long-range dependencies between leads. Preprocessing involves converting RGBA ECG images to RGB format and normalizing them to meet the requirements of the inputs of the layers. Initial experiments have shown promising results, achieving an AUC (Area Under the Curve) of 86.06%. Further work includes tests to optimize the model's performance based on different key metrics, including accuracy and precision.

Keywords: ECG classification, Linformer, Optuna, EfficientNet, Adam Optimizer.

1 Introduction

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, responsible for approximately 17.9 million deaths annually [1]. Among these, myocardial infarction (MI), commonly known as a heart attack, is one of the most critical conditions, occurring when blood flow to a portion of the heart is blocked for a prolonged period, causing irreversible damage to the heart muscle. Timely and accurate detection of MI is essential for improving patient outcomes, as delayed diagnosis can lead to fatal consequences, including heart failure and sudden cardiac arrest [2].

The electrocardiogram (ECG) is the gold standard for diagnosing MI and other heart conditions. It measures the heart's electrical activity and can reveal abnormalities, such as ST-segment elevation, which are indicative of MI.

However, the subtle characteristics of MI on ECG, such as minor deviations in ST segments or T-wave inversions, can be challenging for physicians to interpret, especially in cases where the changes are not pronounced. This challenge is compounded when clinicians must analyze large volumes of ECG data under time constraints, which can increase the likelihood of diagnostic errors [3].

Machine learning (ML), particularly deep learning models, has emerged as a powerful tool for automating ECG interpretation, reducing the time and effort required for diagnosis while enhancing accuracy [4]. Deep learning models can detect subtle patterns in ECG signals that clinicians may miss, providing a significant advantage in diagnosing conditions like MI. By learning from vast datasets, these models can identify and quantify minute ECG changes characteristic of MI, helping to overcome the limitations of manual interpretation. Furthermore, in emergency settings where rapid diagnosis is critical, ML models can support clinicians by providing real-time analysis, potentially improving response times and patient outcomes [5].

This study uses a large publicly available dataset of 12-lead ECGs, the PTB-XL dataset, to develop a deep learning model for binary classification of cardiac conditions. Specifically focusing on distinguishing between healthy individuals and those diagnosed with MI. Our model integrates EfficientNetB3 for feature extraction with a Linformer block to capture long-range dependencies between ECG leads, providing a robust solution for detecting MI based on subtle ECG characteristics.

The rest of this paper is organized as follows: Section 2 presents the related works, identifying prior research on ECG classification and highlighting the approach's novelty. Section 3 outlines the methodology, including dataset preprocessing and model architecture. Section 4 discusses the experimental results, evaluating the model's performance across key metrics. Finally, Section 5 concludes with the author's insights and potential future directions.

2 Related Work

Recent advancements in deep learning have revolutionized the automatic detection of cardiovascular diseases, including myocardial infarction (MI), through electrocardiogram (ECG) analysis. Several methodologies have been proposed, primarily focusing on convolutional neural networks (CNNs) and their variants [6]. Although these approaches have shown promising results, this work introduces a novel combination of EfficientNetB3 for feature extraction and a Linformer block to model the dependencies between ECG leads. This section reviews existing approaches and highlights why the proposed method stands out.

2.1 CNN-Based Models

CNNs have been widely applied to ECG classification due to their ability to extract features from raw data automatically. For instance, [7] developed a 9-layer CNN model to classify MI based on ECG signals, achieving an accuracy of over 90%.

However, their model only processed single-lead ECG data, limiting its ability to capture the inter-lead relationships crucial for detecting subtle signs of MI in multi-lead ECGs. In contrast, the proposed model utilizes the full 12-lead ECG dataset, and by incorporating Linformer, will capture the long-range dependencies between leads, offering a more holistic view of the heart's electrical activity.

2.2 CNN-LSTM Hybrid Approaches

Several works have combined CNNs with long short-term memory (LSTM) networks to improve the temporal modelling of ECG signals. For example, [8] proposed a CNN-LSTM model to classify ECG arrhythmias. While this approach benefits from LSTM's ability to manage time-series data, it suffers from computational inefficiency when dealing with long sequences, such as full 12-lead ECGs. The LSTM layer adds complexity, making real-time clinical applications less feasible.

By contrast, the proposed model replaces the LSTM with a Linformer block, which reduces computational complexity by projecting ECG sequences into a lower-dimensional space while retaining important temporal dependencies. This makes the proposed approach more scalable and efficient for handling large datasets like PTB-XL.

2.3 Transformer-Based Approaches

Recently, transformers have gained popularity in time-series analysis due to their ability to model global relationships within data. For instance, [9] applied a transformer model to physiological time-series data, achieving superior performance compared to CNNs. However, transformers require significant computational resources due to their quadratic complexity.

The proposed use of Linformer addresses this issue by reducing the transformer's computational burden, allowing us to capture lead-to-lead relationships efficiently without compromising accuracy [10]. This makes Linformer an ideal choice for large-scale ECG analysis, as it retains the strengths of transformers while being computationally lighter.

2.4 EfficientNet and Linformer

EfficientNet and Linformer are both modern neural network architectures, each designed with efficiency and performance in mind, but they serve distinct roles.

EfficientNet is a convolutional neural network (CNN) architecture introduced by [11] to achieve high accuracy with lower computational costs. It uses a compound scaling method, balancing the network's width, depth, and resolution in a way that scales the model uniformly without dramatically increasing parameters or computational load. This efficiency makes EfficientNet ideal for tasks like image classification, where high performance and minimal resource consumption are crucial.

Linformer is a Transformer-based model designed to improve the efficiency of self-attention mechanisms, particularly for long-sequence data, introduced by [10]. Unlike standard transformers, which have quadratic complexity due to the self-attention mechanism, Linformer approximates this attention with a linear complexity, reducing memory and computational requirements. This makes Linformer well-suited for tasks where modeling relationships in long sequences is essential, such as natural language processing or, in your case, capturing inter-lead dependencies in ECG data.
EfficientNetB3 significantly improves over traditional CNNs, particularly in model scaling and efficiency. Its design enables the model to achieve high accuracy with fewer parameters, which is critical for medical applications where overfitting to small datasets is a concern [11]. Combining EfficientNet with Linformer leverages spatial feature extraction and temporal modeling, ensuring that subtle ECG patterns, often missed by traditional CNNs or CNN-LSTM hybrids, are detected. This combination improves generalization while maintaining computational efficiency, making it well-suited for real-time MI detection in clinical settings.

3 Methodology

The electrocardiogram (ECG) is a non-invasive tool that measures the heart's electrical activity through different leads, each providing a view of specific heart regions.

The standard 12-lead ECG includes six limb leads (I, II, III, aVL, aVR, and aVF) and six precordial (chest) leads (V1 to V6).

Each lead focus on a specific area of the heart: leads I, aVL, lateral wall of the left ventricle; leads II, III, and aVF, inferior wall of the heart; leads V1-V2, septal region; leads V3-V4, anterior wall of the left ventricle; leads V5-V6, central section; and aVR, generally used to detect global heart ischemia, it does not provide specific insights into individual regions affected by myocardial infarction and is often less relevant for detecting regional MI [2].

Myocardial infarction (MI) is typically identified by specific patterns, such as ST-segment elevation or T-wave inversions, in the leads that correspond to the affected region.

For example, ST-segment elevation in leads II, III, and aVF often indicates an infarction in the inferior wall, while changes in leads V1-V4 suggest anterior or septal infarction [3].

3.1 Dataset: PTB-XL

PTB-XL includes 21,837 clinical 12-lead ECG records from 18,885 subjects. These include both control subjects and pathologic subjects, with a variety of cardiac conditions recorded, such as myocardial infarction, bundle branch block, and dysrhythmiasthologies: The dataset labels cover a broad range of 71 diagnostic classes organized hierarchically. Pathologies include myocardial infarction, conduction disorders, hypertrophy, and other ECG abnormalities .

Recording Duration: Each ECG sample is 10 seconds long, which results in a matrix of size 5000 samples x 12 leads for each record.

Sampling Frequency were recorded at a sampling frequency of 100 Hz, allowing detailed temporal resolution of each heartbeat for more precise analysis .Recording Conditions: The ECGs were recorded under resting conditions in a controlled clinical setting, ensuring standardization across the dataset and making it suitable for diagnostic analysis [12] .

This study focuses on the binary classification of MI versus healthy ECGs. The dataset includes ECG signals from 11 leads, excluding the aVR lead from the standard 12 leads. aVR lead is generally considered less informative for diagnosing MI, as it does not provide useful information on the regions of the heart that are typically affected by ischemic events. According to [13], aVR is less relevant in MI diagnosis as it rarely shows specific changes that can be attributed directly to infarction.

Due to computational limitations, initial experiments were performed using a reduced dataset of 200 healthy ECG and 200 MI ECG selected from the PTB-XL dataset and that dataset was divided into two sets 70% training and 30% validation. This allowed to conduct early tests and refine the model despite limited hardware resources.

3.2 Preprocessing

Preprocessing the ECG images was crucial to ensure the data was in a format suitable for the model. The raw ECG signals were converted to PNG format, resized to 150x150 pixels, and normalized using ImageNet statistics. The reduced image size was chosen primarily to fit the available computational resources, as working with full-resolution ECG images would have significantly increased memory and processing requirements.

Unlike typical image classification tasks, data augmentation was not applied during preprocessing. Augmenting medical data, especially ECG signals, can lead to the distortion of clinically key features, which could result in incorrect model predictions. Studies like [14] have shown that even minor changes in ECG morphology can significantly impact diagnostic outcomes, making augmentation inappropriate for this type of data. For example, flipping or rotating ECG images can misrepresent the direction of electrical impulses, thus making the ECG data unreliable for training.

3.3 Model Architecture

The model leverages two powerful components: EfficientNetB3 for spatial feature extraction and Linformer for modelling long-range dependencies between ECG leads.

EfficientNetB3 was selected due to its state-of-the-art performance in terms of accuracy and computational efficiency [11]. It uses a compound scaling method, which balances network depth, width, and resolution, allowing the model to perform well on smaller datasets without overfitting. This is particularly important for medical applications like MI detection, where overfitting to the training set can lead to poor generalization on unseen data. Using EfficientNet ensures that the model captures detailed spatial patterns in ECG images, such as subtle variations in the ST-segment and T-wave, which are crucial for MI diagnosis.

As for the Linformer, transformers have shown enormous potential for time-series data, but their quadratic complexity makes them computationally expensive [10]. Linformer addresses this by approximating the attention mechanism with a low-rank projection, reducing computational complexity while retaining the ability to model global relationships between ECG leads. Linformer is well-suited for ECG analysis, where inter-lead dependencies provide critical information about the heart's function. Combining EfficientNet and Linformer allows for extracting both local spatial features and long-range dependencies, creating a more comprehensive model for MI detection.

3.4 Hyperparameter Tuning with Optuna

Optuna is an open-source, automatic hyperparameter optimization framework designed to improve machine learning models by finding optimal hyperparameters efficiently. Optuna uses an approach known as *define-by-run*, which enables users to dynamically define search spaces for hyperparameters within their code, making it flexible and highly adaptable for complex models.

Optuna supports various optimization algorithms, including *Tree-structured Parzen Estimator* (TPE) and *multi-objective optimization*, allowing it to explore hyperparameters effectively for both single-objective and multi-objective problems. Additionally, Optuna integrates features like *pruning*, which stops trials early if they show inferior performance, thereby reducing computation time [15].

Optuna was used to tune the model's hyperparameters because it efficiently navigates large search spaces. The search space included batch size, learning rate, and whether to freeze the base layers of EfficientNet. The primary tuning metric was AUC (Area Under Curve), which comprehensively evaluates the model's performance across all classification thresholds. Twenty trials were conducted, each involving five epochs of training and validation to identify the best configuration.

AUC (Area Under the Curve) of the ROC curve was chosen as the main evaluation metric because it measures how well the model distinguishes between classes. It balances sensitivity and specificity, which is crucial in medical diagnoses where both false positives and false negatives can have profound consequences.

3.5 Training Process

The training process was conducted using the Adam optimizer (Adaptive Moment Estimation), which was selected because it combines the advantages of both AdaGrad and RMSProp.

Specifically, Adam adapts the learning rate for each parameter based on the estimates of lower-order moments, allowing for fast and robust convergence across complex architectures like the proposed, which involves both EfficientNet and Linformer. Adam is particularly well-suited for large datasets and non-stationary objectives, such as ECG signal classification, where the underlying signal characteristics may vary between different ECG leads and patients.

The EarlyStopping mechanism was implemented to monitor the validation performance and halt training if no improvement was observed over 4 epochs. This is critical in preventing overfitting, especially when dealing with medical datasets that often have fewer samples than general image datasets. Overfitting in such cases can lead to the model memorizing the training data, thus performing poorly on unseen data.

The ReduceLROnPlateau scheduler was also used to adjust the learning rate dynamically during training. When the validation performance plateaued, the learning rate gets reduced. This technique helps the model make more precise adjustments in later stages of training, improving the overall fit without oscillating around suboptimal values. This approach is especially useful in complex deep-learning architectures, where early rapid updates are beneficial, but finer tuning is required as the model approaches convergence.

4 Results

This section presents the outcomes of experiments conducted to identify the optimal hyperparameters for our ECG classification model using Optuna. The focus was maximizing the Area Under the Receiver Operating Characteristic Curve (AUC), a crucial metric for evaluating the model's ability to distinguish between healthy and myocardial infarction (MI) ECGs.

The best testing configuration achieved an AUC of 86.06% with a batch size of 14 and a learning rate of 9.85e-5, while freezing the base layers of EfficientNetB3, as shown in Fig. 1.

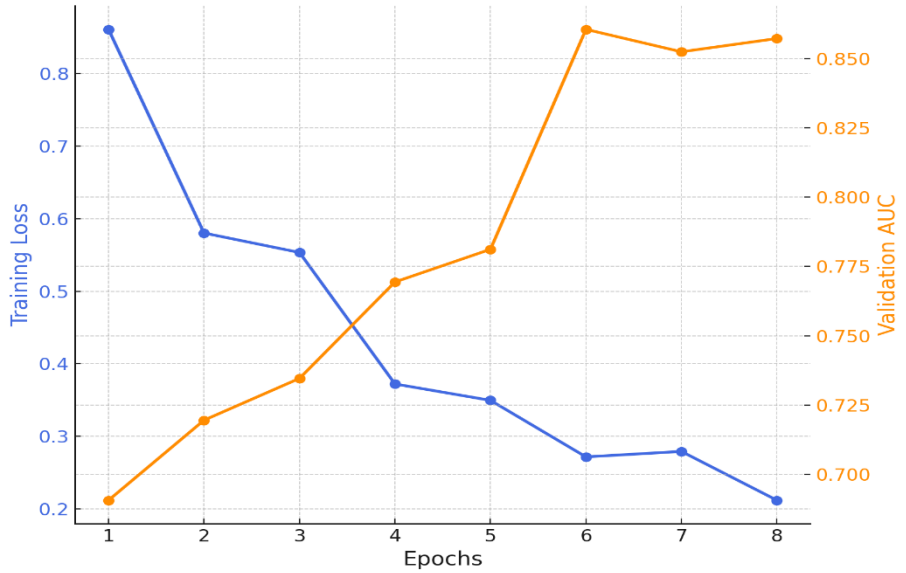

Fig. 1. Training Loss and Validation AUC over Epochs

Furthermore, Table 1 summarizes the trials' results, focusing on the most notable configurations encountered. These findings suggest that a larger batch size and maintaining frozen base layers are beneficial for enhancing model performance. To comprehensively evaluate its efficacy, further experiments will fine-tune the model and explore additional metrics, including accuracy, precision, recall, and F1-score.

Table 1. Samples of trials and the best trial yet.

Trial	AUC %	Batch Size	Learning Rate	Freeze Base Layers	Observations
0	76.56%	8	8.49e-05	Yes	The model plateaued early, likely due to limited batch size. Freezing base layers helped prevent overfitting.
1	77.06%	9	1.10e-04	Yes	A slight improvement, but early stopping was triggered due to plateauing.
2	86.06%	14	9.85e-05	Yes	Best performance achieved with a larger batch size and stable learning rate.
3	84.19%	14	1.21e-04	No	Unfreezing the base layers led to overfitting, resulting in a slight drop in AUC.

5 Conclusion

This study presents a novel deep-learning model for the binary classification of MI using the PTB-XL dataset. Combining EfficientNetB3 for spatial feature extraction and Linformer blocks for temporal dependency modelling, the model can efficiently capture long-range relationships between ECG leads, which is critical for detecting myocardial infarction.

The main objective of this study is to demonstrate the effectiveness of combining EfficientNetB3 and Linformer for detecting myocardial infarction (MI) from ECG signals. Despite promising results, achieving these outcomes required considerable computational resources; training the model once took approximately 6,846 minutes using an Intel(R) Celeron(R) N4120 CPU @ 1.10GHz 1.10 GHz and 8.00 GB RAM memory. This extensive training time highlights the need for powerful computational setups so for the future works leveraging the distributed computing resources is essential.

The best trial achieved promising results, with an AUC of 0.8606. These results suggest that our model effectively captures the intricate patterns in ECG signals indicative of myocardial infarction.

Further works aim to extend this work beyond binary classification to cover a broader range of cardiac pathologies, transforming the model into a comprehensive diagnostic tool.

Both EfficientNet and Linformer have shown remarkable flexibility in adapting to more complex tasks. EfficientNet is highly scalable, which allows it to efficiently manage multi-class classification by adjusting its depth, width, and resolution. Similarly, the Linformer architecture provides an efficient method for modelling long-range dependencies, making it suitable for multi-class problems involving ECG data.

By leveraging these capabilities, we are confident that future work will enable the model to classify a wider range of cardiac conditions accurately. The goal is to provide clinicians with a tool that detects myocardial infarction and identifies other critical cardiovascular diseases. This would improve diagnostic precision and contribute to better patient outcomes.

Acknowledgements

This work was supported by national funds through FCT/MCTES (PIDDAC): CeDRI, UIDB/05757/2020 (DOI: 10.54499/UIDB/05757/2020) and UIDP/05757/2020 (DOI: 10.54499/UIDP/05757/2020); and SusTEC, LA/P/0007/2020 (DOI: 10.54499/LA/P/0007/2020).

References

[1] World Health Organization, “Cardiovascular diseases (CVDs) fact sheet,” *World Health Organization(2021)*, 2021.

[2] K. Thygesen *et al.*, "Fourth Universal Definition of Myocardial Infarction (2018)," *J Am Coll Cardiol*, vol. 72, no. 18, pp. 2231–2264, Oct. 2018, doi: 10.1016/j.jacc.2018.08.1038.

[3] R. Costa, T. Winkert, A. Manhães, and J. P. Teixeira, "QRS Peaks, P and T Waves Identification in ECG," *Procedia Comput Sci*, vol. 181, pp. 957–964, 2021, doi: 10.1016/j.procs.2021.01.252.

[4] P. H. Borghi, R. C. Borges, and J. P. Teixeira, "Atrial fibrillation classification based on MLP networks by extracting Jitter and Shimmer parameters," *Procedia Comput Sci*, vol. 181, pp. 931–939, 2021, doi: 10.1016/j.procs.2021.01.249.

[5] G. H. Tison *et al.*, "Passive Detection of Atrial Fibrillation Using a Commercially Available Smartwatch.," *JAMA Cardiol*, vol. 3, no. 5, pp. 409–416, May 2018, doi: 10.1001/jamacardio.2018.0136.

[6] Rodrigues, P. Miguel, Teixeira, and J. Paulo, "Classification of Electroencephalogram signals using Artificial Neural Networks," in *2010 3rd International Conference on Biomedical Engineering and Informatics*, IEEE, Oct. 2010, pp. 808–812. doi: 10.1109/BMEI.2010.5639941.

[7] U. R. Acharya *et al.*, "Entropies for automated detection of coronary artery disease using ECG signals: A review," *Biocybern Biomed Eng*, vol. 38, no. 2, pp. 373–384, 2018, doi: 10.1016/j.bbe.2018.03.001.

[8] Y. Xia, N. Wulan, K. Wang, and H. Zhang, "Detecting atrial fibrillation by deep convolutional neural networks," *Comput Biol Med*, vol. 93, pp. 84–92, Feb. 2018, doi: 10.1016/j.compbiomed.2017.12.007.

[9] Y. Xu, S. Xu, M. Ramprassad, A. Tumanov, and C. Zhang, "TransEHR: Self-Supervised Transformer for Clinical Time Series Data," 2023. *AwardID:2029004*.

[10] S. Wang, B. Z. Li, M. Khabsa, H. Fang, and H. Ma, "Linformer: Self-Attention with Linear Complexity," Jun. 2020.

[11] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 2019.

[12] P. Wagner *et al.*, "PTB-XL, a large publicly available electrocardiography dataset," *Sci Data*, vol. 7, no. 1, p. 154, May 2020, doi: 10.1038/s41597-020-0495-6.

[13] B. J. Drew *et al.*, "Practice Standards for Electrocardiographic Monitoring in Hospital Settings," *Circulation*, vol. 110, no. 17, pp. 2721–2746, Oct. 2004, doi: 10.1161/01.CIR.0000145144.56673.59.

[14] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks," Jul. 2017.

[15] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, NY, USA: ACM, Jul. 2019, pp. 2623–2631. doi: 10.1145/3292500.3330701.