INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior de Tecnologia e Gestão

Université ULT

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

Mohamed Khalil CHAABANI

Dissertation presented to the School of Technology and Management of Bragança to obtain the master's degree in electrical and computer engineering.

Work oriented by:

Dr. João Paulo TEIXEIRA

Dr. Mohamed Aymen SLIM

May, 2025

Dedication

I begin by expressing my deepest and most heartfelt gratitude to my family. Their unwavering support, encouragement, and belief in me have been the foundation of my strength throughout this journey. From the very beginning, they have stood by my side, offering their wisdom, guidance, and unconditional love, even when the path was not always clear or easy.

Their sacrifices and understanding during the most challenging moments have been a constant reminder of what truly matters, and I am profoundly grateful for everything they have done to help me reach this milestone.

I want to extend my sincere thanks to my friends, for their companionship and support. Your presence has been a source of joy, laughter, and balance during this intense period. The moments we shared, whether through late-night conversations, shared meals, or simply being there, have made this experience far more enriching and memorable. Your encouragement, humour, and honesty have helped me stay focused and motivated, and I am truly thankful to have had you all along this journey.

Acknowledgment

First and foremost, I would like to thank my supervisors, Dr. João Paulo Teixeira and Dr. Mohamed Aymen Slim, for their invaluable guidance, encouragement, and expertise. Their insightful feedback and unwavering support have been instrumental in shaping this research and bringing it to fruition.

I am also grateful to my colleagues and professors at ESTIG, Instituto Politécnico de Bragança, and ULT, Université Libre De Tunis, whose collaboration and camaraderie made this journey enjoyable and enriching. Their support and encouragement have been invaluable throughout this process.

Lastly, I would like to express my sincere gratitude to all those who have supported me throughout my thesis journey and thank you all for being an integral part of my academic journey.

ABSTRACT

Cardiovascular diseases, including myocardial infarction, remain among the leading causes of mortality worldwide. Timely and accurate diagnosis is critical for effective treatment but often requires labour-intensive manual analysis of clinical-grade electrocardiograms (ECGs). This dissertation proposes a novel deep learning-based approach for binary classification of cardiac pathologies, using the PTB-XL dataset. The final model architecture integrates EfficientNetB3 for spatial feature extraction and a Linformer block to capture long-range dependencies between ECG leads, the results prove its adaptability for ECG image classification tasks.

Extensive experimentation and iterative model development were conducted to reach the final design. Early trials involved exploring different hyperparameter tuning like Optuna and Adam optimizer and a wide range of hyperparameter configurations, including different learning rates, dropout rates, batch sizes, and numbers of Linformer layers. These experiments were critical in finding the optimal combination of parameters that balanced computational efficiency and model accuracy. Comprehensive details of these trials and evaluations are provided in the report.

The preprocessing pipeline involves selecting the ECGs and converting them to 11 images (aVR was excluded) each representing a lead, converting RGBA ECG images to RGB format and applying normalization to ensure compatibility with model input requirements. This preprocessing step addresses the unique format of the dataset and prepares it for high-performance neural network training.

Initial results from the finalized model architecture have demonstrated promising performance, achieving an AUC (Area Under the Curve) of 85.02% and a F1-score of 78.94%. The achieved results are comparable to recent state-of-the-art models reported on the PTB-XL dataset, which typically range between 85% and 95% AUC for similar binary classification tasks. These results indicate that the model's AUC of 85.02% is promising but on the edge of the current state-of-the-art. These findings indicate strong potential for the system to support clinical decision-making by automating the classification of ECG data. Ongoing research aims to extend the current binary classification framework to multi-class scenarios, further enhancing its clinical applicability. Additionally, efforts are being made to improve the model for faster inference times, enabling real-time ECG analysis and improving its feasibility for deployment in healthcare settings.

Keywords: ECG classification, EfficientNet, Linformer, Manual and Dynamic hyperparameter tuning techniques, Adam Optimizer, Optuna, AUC, Myocardial Infarction.

RESUMO

As doenças cardiovasculares, incluindo o enfarte do miocárdio, continuam a ser uma das principais causas de mortalidade em todo o mundo. O diagnóstico atempado e preciso é fundamental para um tratamento eficaz, mas muitas vezes requer uma análise manual intensiva de electrocardiogramas (ECGs) de grau clínico. Esta dissertação propõe uma nova abordagem baseada em aprendizagem profunda para a classificação binária de patologias cardíacas, utilizando o conjunto de dados PTB-XL. A arquitetura final do modelo integra o EfficientNetB3 para extração de caraterísticas espaciais e um bloco Linformer para capturar dependências de longo alcance entre as leads do ECG, os resultados provam a sua adaptabilidade para tarefas de classificação de imagens de ECG.

Foi efetuada uma experimentação extensiva e um desenvolvimento iterativo do modelo para chegar à conceção final. Os primeiros ensaios envolveram a exploração de diferentes afinações de hiperparâmetros, como o Optuna e o optimizador Adam, e uma vasta gama de configurações de hiperparâmetros, incluindo diferentes taxas de aprendizagem, taxas de abandono, tamanhos de lotes e números de camadas Linformer. Estas experiências foram fundamentais para encontrar a combinação ideal de parâmetros que equilibrasse a eficiência computacional e a precisão do modelo. O relatório apresenta pormenores abrangentes sobre estas experiências e avaliações.

O processo de pré-processamento envolve a seleção dos ECGs e a sua conversão em 11 imagens (aVR foi excluída), cada uma representando uma Lead, a conversão das imagens de ECG RGBA para o formato RGB e a aplicação da normalização para garantir a compatibilidade com os requisitos de entrada do modelo. Essa etapa de pré-processamento aborda o formato exclusivo do conjunto de dados e o prepara para o treinamento de redes neurais de alto desempenho.

Os resultados iniciais da arquitetura do modelo finalizado demonstraram um desempenho promissor, alcançando uma AUC (Área sob a curva) de 85,02% e uma pontuação F1 de 78,94%. Os resultados alcançados são comparáveis aos modelos recentes do estado da arte reportados no conjunto de dados PTB-XL, que tipicamente variam entre 85% e 95% de AUC para tarefas de classificação binária semelhantes. Estes resultados indicam que a AUC do modelo de 85,02% é promissora, mas no limite do estado da arte atual. Estes resultados indicam um forte potencial para o sistema apoiar a tomada de decisões clínicas, automatizando a classificação dos dados de ECG. A investigação em curso visa alargar a atual estrutura de classificação binária a cenários multi-classe, melhorando ainda mais a sua aplicabilidade clínica. Além disso, estão a ser estão a ser feitos esforços no sentido de melhorar o modelo para obter tempos de inferência mais rápidos, permitindo a análise de ECG em tempo real e melhorando a sua viabilidade de implementação em ambientes de cuidados de saúde.

Palavras-chave: Classificação de ECG, EfficientNet, Linformer, Técnicas de ajuste de hiperparâmetros manual e dinâmico, Otimizador Adam, Optuna, AUC, Enfarte do Miocárdio.

SUMMARY

FIGURES INDEX

TABLES INDEX

ABBREVIATIONS AND ACRONYMS

- **AI** (Artificial Intelligence)
- **AMI** (Acute Myocardial Infarction)
- **AUC** (Area Under the Curve)
- **CNN** (Convolutional Neural Network)
- **CSV** (Comma-Separated Values)
- **CVD** (Cardiovascular Disease)
- **DL** (Deep Learning)
- **ECG** (Electrocardiogram)
- **FC** (Fully Connected)
- **LIME** (Local Interpretable Model-Agnostic Explanations)
- **LSTM** (Long Short-Term Memory)
- **MI** (Myocardial Infarction)
- **ML** (Machine Learning)
- **MLP** (Multi-Layer Perceptron)
- **NLP** (Natural Language Processing)
- **NN** (Neural Network)
- **PCI** (Percutaneous Coronary Intervention)
- **PNG** (Portable Network Graphics)
- **PTB** (Physikalisch-Technische Bundesanstalt)
- **RGBA** (Red Green Blue Alpha)
- **RNN** (Recurrent Neural Network)
- **SGD** (Stochastic Gradient Descent)
- **SHAP** (SHapley Additive exPlanations)
- **SVM** (Support Vector Machine)
- **TPE** (Tree-structured Parzen Estimator)

Chapter 1. Introduction

In recent years, advancements in artificial intelligence (AI) and deep learning (DL) have transformed medical diagnostics by offering innovative solutions for conditions requiring early and accurate detection. Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, with myocardial infarction (MI) being one of the most severe and life-threatening conditions. Electrocardiography (ECG) is a widely used, non-invasive diagnostic tool that records the heart's electrical activity, providing crucial information about cardiac health. However, traditional ECG interpretation methods rely heavily on human expertise, making them prone to variability, time consumption, and potential errors, particularly in high-volume or resource-limited healthcare settings.

This research demonstrates the potential of combining transfer learning with Transformer models for ECG classification by achieving high accuracy and AUC scores on validation trials. Furthermore, the results highlight the scalability of the approach, opening avenues for real-time, AI-driven cardiac diagnostics in clinical settings, particularly in remote or underserved areas where access to cardiology expertise may be limited.

1.1. Cardiovascular Diseases (CVDs)

Cardiovascular diseases encompass a group of disorders affecting the heart and blood vessels, being the leading cause of global mortality. According to (World Health Organization, 2021), CVDs are responsible for an estimated 17.9 million deaths annually, accounting for 31% of all global deaths. These disorders include coronary artery disease, cerebrovascular disease, rheumatic heart disease, and myocardial infarction (MI), among others.

Table 1. Types of Cardiovascular Diseases and Clinical Characteristics

Type of CVD	Description	Common Risk Factors	Symptoms and Complications
Coronary Artery Disease (CAD)	A condition where plaque buildup narrows the coronary arteries, reducing blood flow to the heart.	Hypertension, high cholesterol, smoking	Chest pain, heart attack, heart failure
Cerebrovascular Disease	Diseases of the blood vessels supplying the brain often lead to strokes.	Age, obesity, sedentary lifestyle	Numbness, paralysis, speech, and vision problems
Rheumatic Heart Disease	Damage to the heart valves caused by rheumatic fever, an inflammatory disease resulting from untreated streptococcal infection.	Inadequate healthcare access	Breathlessness, fatigue, swollen joints
Myocardial Infarction (MI)	Blockage of the blood flow to the heart, causing damage to heart muscle tissue.	Age, smoking, diabetes, high cholesterol	Chest pain, shortness of breath, fatigue

Table 1 provides a clear comparison of several types of cardiovascular diseases (CVDs), with a particular emphasis on the distinct characteristics and severe risks associated with myocardial infarction (MI). This analysis highlights MI as a particularly dangerous condition due to its unique risk factors and both acute and long-term health consequences.

Unlike chronic cardiovascular conditions that may progress gradually over time, MI is an acute medical emergency. As shown in Table 1, MI occurs when blood flow to the heart is blocked, typically by a blood clot, leading to rapid and potentially irreversible damage to heart muscle tissue. This blockage often results in severe chest pain, shortness of breath, and fatigue, all of which show a critical state requiring urgent medical intervention. Without timely treatment, the mortality risk is significant, marking MI as one of the deadliest CVDs.

Table 1 notes common risk factors for MI, such as age, smoking, diabetes, and high cholesterol. These factors are prevalent across populations, suggesting that MI risks span across age groups and demographics, especially as lifestyle factors like smoking and poor diet remain common. Since many of these risk factors are modifiable, table 1 highlights the importance of lifestyle changes and preventive healthcare measures to reduce MI incidence.

While MI is acute, its impact extends beyond the first episode. Survivors often experience significant heart muscle damage, which may lead to heart failure or other complications over time. Table 1 links MI with symptoms like shortness of breath and fatigue, which can persist and severely affect the patient's quality of life, limiting daily activities and increasing dependency on healthcare resources.

When compared with other forms of CVD in Table 1, MI is notable for both its immediate risk of death and the high likelihood of recurring health issues. For instance, although cerebrovascular diseases like stroke and rheumatic heart disease are also severe, MI presents a particularly acute combination of preventable risk factors and fatal outcomes, underscoring its danger compared to other CVDs.

Table 1 offers essential insights into why MI is so dangerous, emphasising its life-threatening symptoms, modifiable risk factors, and long-term complications.

It highlights the necessity for preventive measures and early intervention to address modifiable risk factors, thus potentially reducing MI's substantial health and economic burden.

1.2. Myocardial Infarction (MI)

Myocardial infarction, commonly known as a heart attack, is a life-threatening condition that occurs when blood flow to a part of the heart is blocked for an extended period, resulting in damage or death of heart muscle cells (Thygesen et al., 2018). This blockage is often caused by a clot in one of the coronary arteries.

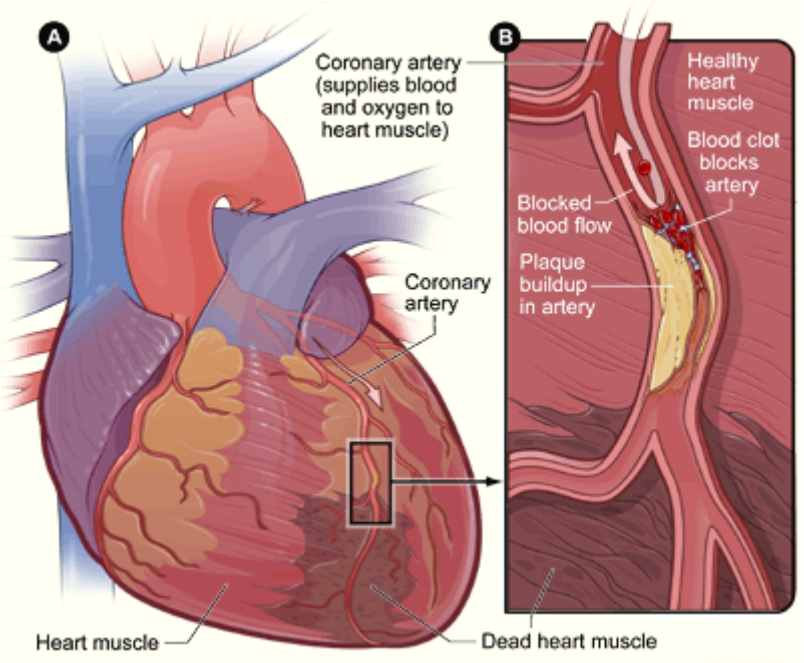


Figure 1. Heart with Obstructed Coronary Artery Causing MI (Wikipedia contributors, 2024)

Figure 1 illustrates the heart anatomy affected during myocardial infarction. Part A shows the overall heart, highlighting the coronary artery that supplies oxygen to the heart muscle. Part B zooms in on the obstructed artery, showing how plaque buildup and a blood clot block blood flow, leading to dead heart muscle tissue. This blockage is a primary cause of myocardial infarction and underscores the importance of rapid diagnosis and intervention to prevent extensive damage to the heart muscle.

The clinical symptoms of MI vary in severity and can include intense chest pain, shortness of breath, sweating, and radiating pain to the arms, neck, or back. Due to the rapid progression of MI and the potential for irreversible damage, early and exact diagnosis is crucial to managing patient outcomes effectively.

MI not only imposes a significant burden on individual health but also exerts strain on healthcare systems due to the need for emergency care, surgical interventions, and long-term cardiac rehabilitation for survivors.

The electrocardiogram (ECG) is the primary diagnostic tool for assessing heart function in suspected MI cases (Costa et al., 2021). It is a non-invasive, accessible test that records the heart's electrical signals. An ECG can reveal specific abnormalities associated with MI.

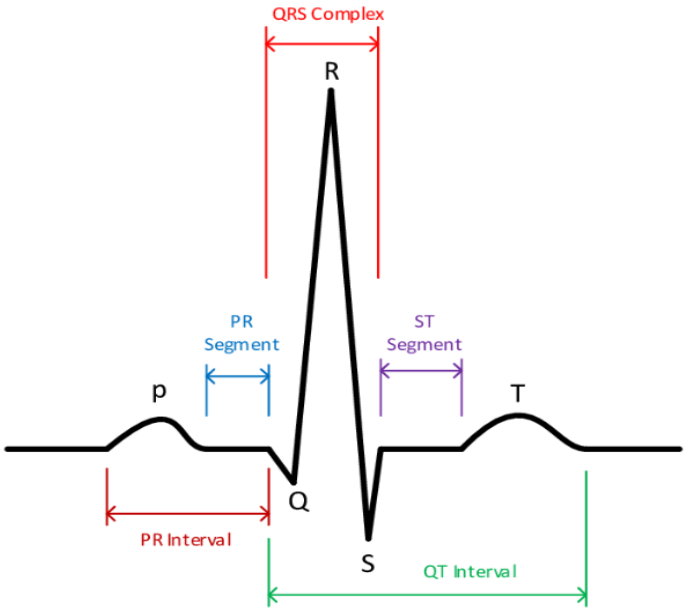


Figure 2. Explanatory ECG Diagram (Nguyen et al., 2020)

In figure 2 we can observe the segments of the ECG Diagram:

ST-segment Elevation: A standard indicator of acute MI, seen when a part of the heart muscle is deprived of oxygen.

T-wave Inversion: This can indicate ischemia, where the blood supply to the heart is restricted.

Pathological Q waves: Signals the presence of dead myocardial tissue due to prolonged ischemia.

Given the subtlety of some MI markers, ECG interpretation requires a high degree of expertise. Deep learning and machine learning models are increasingly being explored to aid in automated interpretation, potentially enhancing diagnostic accuracy and speed.

1.3. Motivation and Relevance

Myocardial infarction is a time-sensitive medical emergency. The outcome for patients experiencing MI heavily depends on how quickly and accurately the condition is diagnosed and treated. When blood flow to a part of the heart is restricted or entirely blocked, the heart muscle begins to suffer from oxygen deprivation, leading to damage or death of myocardial cells. This process can begin within minutes and, if left untreated, leads to irreversible cardiac injury that affects the heart's function, potentially resulting in heart failure or even death.

The medical community often refers to the first hour following an MI as the "golden hour." This period is critical because timely medical intervention, such as administering thrombolytics, percutaneous coronary intervention (PCI), or coronary artery bypass grafting (CABG), can significantly improve patient outcomes. Studies like (Smith et al., 2012) show that interventions performed within this golden hour can reduce mortality and long-term complications, preserving heart function and improving quality of life. Thus, the ability to diagnose MI quickly and accurately is not just a matter of clinical precision but one of life or death.

1.3.1. Limitations of Traditional Diagnostic Methods

The electrocardiogram (ECG) remains the most used diagnostic tool for detecting MI. ECGs record the electrical activity of the heart and can reveal certain patterns, such as ST-segment elevation, T-wave inversions, or pathological Q waves, which suggest an ongoing MI (O'Gara et al., 2013). However, interpreting these subtle changes requires considerable expertise. MI symptoms can be diverse, and the ECG abnormalities associated with it might not always be easily discernible.

In some cases, patients may not exhibit the classic signs of MI, such as chest pain. Instead, they may experience milder symptoms or even remain asymptomatic, especially in elderly patients or those with diabetes (Prochnau et al., 2019). This variability increases the risk of delayed diagnosis if clinicians rely solely on traditional visual cues in ECG readings.

Not all MIs present with apparent ST-segment elevation. NSTEMI cases, where the blockage is partial or affects smaller arteries, often present with less distinct ECG changes, making diagnosis challenging and increasing the risk of missed or delayed intervention.

Beyond ECG, physicians often use cardiac biomarkers, like troponin levels, and imaging studies, such as echocardiograms, to support diagnosis.

However, each of these diagnostic modalities has limitations:

Time to Obtain Biomarkers: Biomarkers like troponin are highly specific for cardiac injury, but they take several hours to appear in the bloodstream. Waiting for lab results can delay diagnosis, especially in emergency settings.

Resource and Time Constraints in Imaging: Imaging studies, while effective, are not always readily available in emergency settings and may take time to arrange, which could lead to delays in diagnosis and intervention.

These limitations underscore the need for more robust, efficient diagnostic methods that reduce reliance on human interpretation alone and enable clinicians to diagnose MI with higher speed and accuracy.

1.3.2. Role of Deep Learning and Automation in MI Detection

Deep learning (DL) and artificial intelligence (AI) present transformative possibilities for cardiology, particularly in MI diagnosis. Automated ECG analysis using Machine Learning (ML) algorithms can detect subtle changes in ECG patterns that may be overlooked by human eyes, especially in busy clinical environments. These models can identify even minor abnormalities by applying algorithms trained on vast datasets, enhancing diagnostic precision.

ML models, particularly those based on deep learning, can process extensive amounts of ECG data in a fraction of the time needed for human analysis. These systems can continuously monitor ECG patterns in real-time, allowing for rapid alerts if signs of MI are detected, which is crucial during the golden hour. Automated systems can help clinicians make faster decisions, improving patient outcomes by reducing the time to treatment.

Despite the promise of deep learning in healthcare, there are challenges to fully implementing automated diagnostic systems:

Data Quality and Variability: Deep learning algorithms are only as good as the data on which they are trained. ECG data can vary significantly based on patient age, underlying health conditions, and device quality. Ensuring that the model performs accurately across diverse populations is a key hurdle.

Integration into Clinical Workflows: Integrating AI-based systems into hospital infrastructures is complex. Automated solutions must be designed to work seamlessly with electronic health records (EHRs) and be accessible to clinicians at the point of care.

Interpretability: Clinicians often hesitate to adopt "black box" models that provide diagnostic recommendations without transparent reasoning. Developing interpretable AI systems where clinicians can understand how an algorithm arrived at a diagnosis is a priority to build trust in deep learning models but unfortunately it remains part of the future works.

1.3.3. Societal and Healthcare Impact

The implications of improved MI diagnostics extend beyond individual patient outcomes. On a societal level, reducing the incidence and severity of MI through faster diagnosis can alleviate some of the burden on healthcare systems. MI treatment, when delayed, is significantly resource-intensive and costly, often requiring emergency care, surgery, and long-term rehabilitation. Early diagnosis reduces the need for such extensive interventions, lowering healthcare costs and freeing resources for other patients.

Furthermore, the widespread adoption of accurate, automated MI diagnostics can improve healthcare access and equity. In regions with limited healthcare resources, remote clinics or underserved hospitals may lack cardiology specialists. AI-enabled diagnostic tools can help bridge this gap, providing these facilities with the tools they need to offer quality care and timely interventions to an at-risk patients. For example, recent studies have explored using smartwatches for passive detection of atrial fibrillation (Tison et al., 2018).

Chapter 2. State of the art

This section explores key studies and methodologies in the domain of deep learning for ECG classification, highlighting the evolution of techniques, challenges addressed, and areas for future exploration.

2.1. Machine Learning and Deep Learning

Machine learning (ML) is a subset of artificial intelligence (AI) that focuses on creating systems that can learn from data to make predictions or decisions without being explicitly programmed. ML algorithms work by identifying patterns in data, enabling models to improve over time based on experience. In healthcare, ML has shown promise in applications like image recognition, medical diagnostics, and predictive analytics.

ML can be broadly categorized into three types:

Supervised Learning: The model learns from labelled data, which includes input-output pairs. Common supervised learning tasks include classification and regression.

Unsupervised Learning: The model learns from unlabelled data by identifying underlying structures or patterns, such as clustering.

Reinforcement Learning: The model learns by interacting with an environment and receiving feedback, often through a reward-punishment system.

In the context of ECG analysis, supervised learning is widely used as it enables models to classify heartbeats as normal or abnormal, detect arrhythmias, or identify specific conditions such as myocardial infarction based on labelled ECG data.

Traditional ML approaches for ECG analysis include algorithms like Support Vector Machines (SVM), Random Forests, and k-Nearest Neighbours (k-NN). These models rely on separately engineered features derived from ECG signals, such as heart rate, wave amplitudes, and time intervals. Although these algorithms are effective for structured and interpretable data, they often struggle with complex, high-dimensional data like raw ECG signals, which contain intricate temporal and spatial dependencies that are challenging to capture with conventional feature engineering.

Deep learning (DL), a specialized subset of ML based on artificial neural networks, addresses these limitations by automatically learning hierarchical features from raw data. DL models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), excel in extracting spatial and temporal patterns directly from ECG signals, eliminating the need for a separate extensive feature extraction.

In ECG analysis, deep learning models have proven highly effective for tasks like arrhythmia detection, heart rate variability analysis, and MI detection. Here is an overview of popular deep learning architectures used in ECG analysis:

Convolutional Neural Networks (CNNs): CNNs are designed to process grid-like data, making them ideal for image and signal processing. In ECG analysis, CNNs automatically learn spatial features, such as waveform shapes and anomalies, by applying convolutional filters over the ECG data. CNNs can handle high-dimensional input data and capture local dependencies in ECG signals, which is beneficial for detecting the presence of MI.

Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs): RNNs are specialized for sequential data, allowing them to model temporal dependencies. LSTMs, a variant of RNNs, can learn long-term dependencies by overcoming the vanishing gradient problem, making them suitable for capturing the temporal sequence in ECG signals. This capability is valuable for diagnosing MI, as ECG abnormalities may persist over several heartbeats.

Transformers: Recently, transformer models have been adapted for ECG analysis. Originally developed for Natural Language Processing (NLP), transformers can capture long-range dependencies in sequential data. They can model relationships across multiple ECG leads, enabling a comprehensive understanding of the entire cardiac cycle. Unlike RNNs, transformers do not process data sequentially; instead, they analyse all data points in parallel, which improves efficiency.

Linformer: an efficient alternative to traditional Transformer models, designed to handle long-range dependencies in sequential data like. By approximating the self-attention mechanism, it reduces the computational complexity of processing large sequences from quadratic to linear. This makes it ideal for tasks that require both efficiency and accuracy while still capturing important relationships between ECG leads.

Despite their advantages, deep learning models for ECG analysis face several challenges:

Data Quality and Variability: ECG signals can vary significantly between patients, age groups, and recording devices. Deep learning models need diverse, high-quality training data to generalize well across different populations.

Computational Demands: Training deep learning models on ECG data is computationally intensive, especially with large networks like CNNs or transformers. The high-dimensional nature of ECG data also requires substantial memory and processing power, which can be a barrier for real-time applications.

Interpretability: Deep learning models are often criticized for their "black box" nature. In healthcare, clinicians need to understand how a model arrives at a diagnosis to ensure trust and reliability. Developing interpretable DL models is crucial for their adoption in clinical practice.

Overfitting: Due to the complexity of deep networks, there is a risk of overfitting, particularly when the training dataset is limited or unrepresentative. Regularization techniques, such as dropout and batch normalization, help mitigate this risk but do not eliminate it entirely.

This foundational overview has provided a comprehensive understanding of machine learning's role in ECG analysis, highlighting both traditional approaches and the transformative impact of deep learning. By leveraging the capacity of CNNs and other architectures to learn directly from ECG data, deep learning has introduced new possibilities for automated, accurate, and efficient MI detection. Nevertheless, the choice of model architecture remains critical, as it directly influences performance, interpretability, and clinical applicability.

2.2. Overview of Automated Diagnostic Systems in Cardiology

Automated diagnostic systems have been increasingly applied in cardiology, especially for detecting life-threatening conditions such as (Ribeiro et al., 2020). The literature reveals a growing interest in leveraging (DL) and (AI) to interpret electrocardiograms (ECGs), biomarkers, and other clinical data automatically. Artificial neural networks have been successfully applied to classify electroencephalogram signals (Rodrigues et al., 2010). This approach offers the potential to mitigate the challenges posed by traditional methods, such as subjective interpretation and the time needed for biomarker testing. Some studies (Acharya et al., 2018) have explored various algorithms, datasets, and feature extraction methods to improve the speed and accuracy of MI detection.

Deep learning, which can learn hierarchical representations directly from raw data, has emerged as a powerful alternative. Leveraging architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), researchers have developed models capable of accurately identifying and classifying various cardiac pathologies. These advancements have paved the way for automated, scalable, and precise ECG analysis, addressing challenges in real-time monitoring and early diagnosis of critical conditions like myocardial infarction, arrhythmias, and heart failure.

2.1.1. Deep Learning Models in ECG Detection

Table 2 provides a summary of notable studies in this area, including the deep learning models used, key features extracted, performance metrics, and limitations found in each study. This review will compare different approaches and provide a critical assessment of their effectiveness.

Table 2. Key studies and performance metrics of ML models for MI detection

Study	Model Type	Key Features Extracted	CVD	Performance Metrics	Limitations
(Khan et al., 2021)	Deep Learning (CNN LSTM)	ECG signal features	(MI)	Accuracy: 94%	High computational cost; requires large datasets
(Patel et al., 2024)	Machine Learning (Random Forest)	Heart rate variability, QRS complex	(MI)	Accuracy: 88%	Limited interpretability; may miss subtle patterns
(Cao et al., 2022)	(ResNet-18)	Pre-trained features from ECG images	(ARR)	Accuracy: 98%	Transfer learning may not capture all specifics
(Borghi et al., 2021)	Machine Learning (MLP)	Jitter and Shimmer parameters	Atrial Fibrillation (AFib)	Accuracy: 85%	Limited generalization; dependent on feature extraction

Due to their ability to accurately process sequential and complex ECG data, these studies show a clear trend toward using deep learning models, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. The integration of these advanced techniques has shown promising results in ECG classification tasks.

2.1.2. Advances and Limitations of Deep Learning Models in MI Detection

Convolutional Neural Networks (CNNs) have been widely adopted for analysing ECG signals due to their ability to identify spatial patterns in waveform data.(Khan et al., 2021) used a hybrid model combining CNN and LSTM architectures, achieving an accuracy of 94%. However, the performance of CNNs can be limited by high computational costs and the necessity for large datasets, which may hinder their generalization across diverse populations.

Long Short-Term Memory (LSTM) networks have shown promise in handling time-series data, such as ECG signals, by retaining long-term dependencies. (Xia et al., 2018) demonstrated the efficacy of LSTM in detecting atrial fibrillation with a sensitivity of 90%. However, LSTMs need substantial labelled data for training, which can limit their applicability in real-time scenarios. Additionally, the black-box nature of these models raises concerns about interpretability, which is critical in clinical settings.

Traditional deep learning models, such as Multi-Layer Perceptron (MLP) networks and Random Forests, have also been applied to detect myocardial infarction (MI). (Borghi et al., 2021) Employed an MLP model that achieved a sensitivity of 85%. However, while some machine learning models like CNNs or RNNs can automatically learn features, for models like MLP and Random Forests that rely on structured feature inputs, they require these features to be engineered and extracted beforehand, the effectiveness and generalizability of these models are thus heavily dependent on the quality and relevance of the chosen pre-extracted features.

Random Forest models, as used by (Patel et al., 2024), provide better interpretability by allowing clinicians to understand the importance of individual features. However, while traditional models offer transparency, they may compromise accuracy compared to deep learning approaches, showing a trade-off between interpretability and performance.

2.1.3. Comparative Analysis of Existing Models

The comparison of various deep learning models for myocardial infarction (MI) detection reveals distinct strengths and limitations based on accuracy, sensitivity, specificity, and interpretability. This section will delve deeper into these performance aspects, analysing how each model type contributes to or limits automated MI diagnostics.

Deep learning models, particularly Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, tend to achieve higher accuracy and sensitivity than traditional machine learning approaches.

For instance, studies employing CNNs reported accuracies above 90% when evaluated on controlled datasets. CNNs excel in capturing spatial and morphological features of ECG waveforms, which are crucial for detecting patterns associated with myocardial infarction. Similarly, LSTMs (Faust et al., 2018), designed to manage sequential data, perform well by keeping temporal dependencies, making them suitable for interpreting the dynamic nature of ECG signals. However, while these models achieve impressive sensitivity, critical in identifying MI cases, they can sometimes be overly sensitive to noise in the ECG signals, leading to false positives.

Specificity, or the ability of the model to correctly identify non-MI cases, is another crucial performance measure, especially to avoid unnecessary alarms in clinical settings. Traditional machine learning models, such as Support Vector Machines (SVM) and Random Forests, often perform better in specificity due to their reliance on explicitly defined features. These models tend to be less prone to overfitting compared to complex deep-learning models. Random Forests have shown moderate success in achieving balanced specificity and sensitivity, as their ensemble nature reduces variance, thereby providing more stable outputs.

However, these models may underperform in accurately identifying complex or nuanced ECG changes associated with early or atypical MI presentations, limiting their application in broader MI diagnostics.

A significant challenge in applying these models to clinical practice is their generalization across diverse patient data. Many studies in this area have trained models on small and homogenous datasets, which may not reflect the broader patient population with varied demographics and comorbidities. Despite their high accuracy in controlled datasets, deep learning models often face difficulties when applied to data outside their training set.

For instance, a CNN model trained on ECG data from one population may not perform as effectively when tested on another, especially if there are differences in underlying health conditions, age, or ECG device characteristics.

Computational requirements are critical, especially in emergency settings where timely diagnosis is crucial. CNNs and LSTMs typically demand more computational power due to their high parameter counts and iterative processes. This makes them challenging to implement in real-time applications, particularly on standard clinical hardware or portable devices used in pre-hospital settings.

In contrast, simpler models like SVMs and Random Forests are less resource-intensive and faster to deploy, making them more practical for real-time decision-making. However, these traditional models may sacrifice diagnostic accuracy for speed, particularly in cases where high granularity in ECG interpretation is needed.

In a clinical setting, model interpretability is crucial for gaining clinician trust and ensuring the safety of AI-assisted diagnostics. Deep learning models are often criticized for their "black box" nature, where decision-making processes are not easily interpretable. This is especially relevant in healthcare, where practitioners must understand the rationale behind diagnostic suggestions to make informed decisions. Traditional models like Random Forests offer more interpretability through feature importance analysis, allowing clinicians to see which ECG aspects influence the model's predictions. This feature-level transparency is less prominent in CNNs and LSTMs, hindering adoption despite their high accuracy. As a result, research has an ongoing trade-off between developing highly accurate models and those that are interpretable enough to be trusted in clinical applications.

Recent research like (Jia et al., 2023) trends indicate a move toward hybrid and ensemble approaches that combine the strengths of different models. By integrating traditional feature-based models with deep learning architectures, researchers aim to achieve both high accuracy and interpretability. For example, ensemble methods that combine the outputs of SVMs, Random Forests, and CNNs can leverage the high accuracy of deep learning models with the interpretability of traditional models, leading to a balanced solution. Hybrid models can also improve generalization across diverse datasets by incorporating both raw and engineered features. However, these approaches introduce added complexity in model design and training, and may require even more computational resources, thus presenting a challenge in real-time, resource-limited scenarios.

In addition to hybrid and ensemble approaches, methods that improve the explainability of machine learning models have gained significant attention in the field of medical diagnostics.

Such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) are two prominent methods used to improve the interpretability of machine learning models (Ningthoujam et al., 2025), especially in healthcare applications like ECG diagnosis, where explainability is crucial. These techniques help clinicians understand how a model reaches its decisions by identifying the contribution of individual features to the final prediction. LIME works by approximating the black-box model with a simpler, interpretable model locally around the prediction, whereas SHAP provides a unified measure of feature importance based on cooperative game theory. Both methods are beneficial in enhancing trust in machine learning models, making them more transparent and enabling clinicians to make better-informed decisions when using AI-assisted tools in cardiac diagnostics.

The literature on machine learning and deep learning applications in myocardial infarction detection underscores the advancements and remaining challenges in automated cardiac diagnostics. Deep learning models, particularly CNNs and LSTMs, exhibit high accuracy and sensitivity, demonstrating the capability to detect subtle ECG changes that may signal MI. However, issues such as computational requirements, limited interpretability, and dataset diversity continue to impede their practical application. Traditional models, while less accurate, still play a valuable role in contexts where interpretability and real-time responsiveness are prioritized.

In conclusion, the comparative analysis of machine learning and deep learning models for myocardial infarction detection highlights both the significant advancements made and the ongoing challenges in the field. Deep learning models, particularly CNNs and LSTMs, offer high accuracy and sensitivity, making them ideal for detecting subtle and complex ECG changes associated with MI. However, they are often hindered by issues related to computational requirements, limited interpretability, and difficulties in generalizing to diverse datasets. Traditional machine learning models, though less accurate, remain valuable due to their interpretability and ability to operate in real-time applications. Hybrid and ensemble models, which combine the strengths of both approaches, offer a promising path forward by balancing performance with interpretability and real-world applicability. As the field advances, further research into these hybrid solutions may provide the most robust and adaptable systems for MI detection.

2.5.2. Model Selection Rationale

Recent advancements in deep learning have demonstrated the capability of neural networks to achieve performance levels comparable to cardiologists in detecting and classifying arrhythmias from ambulatory electrocardiograms.

For instance, (Hannun et al., 2019) highlighted a deep neural network that achieved cardiologist-level accuracy in arrhythmia detection, highlighting the potential of these models to enhance diagnostic reliability.

Table 3 highlights the key strengths and weaknesses of LSTM (which was evaluated but not chosen), Transformer (which was considered but not used), and Linformer architectures in the context of handling sequence data.

Table 3. Comparison of Transformer, LSTM, and Linformer

Model	Architecture Type	Strengths	Weaknesses	Suitability for Task
LSTM	Recurrent Neural Network (RNN)	Effective for short-to-medium sequential data, lightweight	Struggles with long-range dependencies, slower due to sequential processing, gradient issues in long sequences	Limited, not suitable for long sequences with high-dimensional data
Transformer	Self-Attention Based	Handles long-range dependencies well, parallel processing allows faster training on large datasets	High computational cost, quadratic complexity with respect to sequence length	Suitable but resource-intensive
Linformer	Optimized Self-Attention	Maintains efficiency of Transformer while reducing computational complexity to linear	May sacrifice some accuracy on complex tasks due to approximation in attention	Best choice for handling large sequences efficiently

While LSTM are known for their simplicity and effectiveness in handling short to medium sequences, their limitations become apparent with longer sequences. Specifically, LSTM struggles with modelling long-range dependencies due to issues like gradient vanishing, which reduces their effectiveness on large datasets with high-dimensional sequential inputs. Moreover, the sequential processing of LSTM makes them computationally slower, which was a critical drawback for this project, where high efficiency and scalability were required.

On the other hand, Transformers offered significant improvements by allowing for parallel processing and capturing long-range dependencies through a self-attention mechanism(Vaswani et al., 2017). However, their computational cost scales quadratically with sequence length, making them expensive to train, particularly for lengthy sequences like high-resolution medical images or large time-series datasets.

Linformer addresses the high computational cost of Transformers by approximating the self-attention mechanism, reducing its complexity to linear with respect to sequence length. This efficiency enables Linformer to handle large and high-dimensional sequence data effectively, balancing the performance benefits of Transformers with reduced memory usage and processing time. Consequently, Linformer emerged as the preferred choice for this project, offering an optimal balance of accuracy and efficiency that was critical for the large-scale ECG data used (S. Wang et al., 2020).

Table 4 focuses on comparing CNN architectures for image classification, specifically EfficientNetB3, ResNet, and MobileNet, in terms of their strengths and suitability for handling high-resolution medical images. Recent advancements have demonstrated that convolutional neural networks can achieve cardiologist-level performance in arrhythmia detection (Rajpurkar et al., 2017).

Table 4. EfficientNetB3, ResNet, and MobileNet in Image Classification

Model	Architecture Type	Strengths	Weaknesses	Suitability for Task
EfficientNetB3	Compound Scaled CNN	High accuracy with efficient scaling, captures fine-grained details without high computational cost	Less adaptable for very low-power devices compared to MobileNet	Ideal choice for high-resolution medical images
ResNet	Residual Connections CNN	Robust, effective for deep networks, reliable performance on general image tasks	More resource-intensive, lacks scaling flexibility of EfficientNet	Suitable, but less efficient for high-dimensional medical data
MobileNet	Depth wise Separable CNN	Highly efficient on low-power devices, lightweight	Limited in performance for complex high-resolution tasks	Not suitable due to limited accuracy for high-res medical data

While a strong general-purpose model, ResNet was not selected due to its computational cost for general-purpose image classification and performs well on deep networks. However, its lack of a compound scaling strategy makes it less flexible in balancing accuracy and computational efficiency for high-dimensional medical imaging tasks. While ResNet achieved reasonable accuracy in initial tests, it was found to be more resource-intensive than EfficientNetB3, leading to a lower suitability for this project given the hardware constraints.

MobileNet, known for its efficiency, was not chosen due to its limited accuracy in high-resolution medical imaging for low-power applications. However, this lightweight design limits its performance on complex, high-resolution datasets. In testing, MobileNet was unable to match the accuracy of EfficientNetB3 and ResNet, making it unsuitable for high-precision medical imaging tasks. Therefore, EfficientNetB3 was selected as the optimal model for image feature extraction in this project, balancing high accuracy with efficient resource usage.

EfficientNetB3 leverages a unique compound scaling approach (Tan & Le, 2019), which scales depth, width, and resolution simultaneously, enabling it to maintain high accuracy without excessive computational cost. This makes EfficientNetB3 highly efficient for handling high-resolution images like ECG scans, where fine-grained details are crucial for accurate diagnosis.

2.4. Linformer and EfficientNetB3

Based on the comparative analysis and testing results, Linformer and EfficientNetB3 were selected as the optimal models for handling sequence data and image classification, respectively, in this project. Here is a summary of the rationale behind each choice:

Linformer for Sequence Data: Linformer combines the accuracy benefits of self-attention with a computationally efficient, linear-complexity design. It is well-suited for large-scale ECG sequence data, which requires efficient processing and the ability to capture long-range dependencies without excessive memory usage. Compared to LSTM and Transformer, Linformer strikes a balance between efficiency and performance, making it an ideal choice for handling high-dimensional medical sequence data in a resource-constrained environment.

EfficientNetB3 for Image Classification: EfficientNetB3's compound scaling strategy allows it to capture intricate details in high-resolution images without the computational burden that ResNet and other traditional CNNs require. Its balance of high accuracy and computational efficiency makes it optimal for medical imaging tasks, where precise feature extraction is essential for diagnosis, leading to its selection as the primary model for image classification in this project.

The model selection process aimed to identify architectures that maximize performance while adhering to computational efficiency, which is essential in processing large ECG datasets and high-resolution images. Linformer and EfficientNetB3 were chosen due to their unique capabilities in managing complex data types efficiently.

Linformer provides an efficient alternative to standard Transformers for sequential data, while EfficientNetB3 offers high accuracy for image classification without excessive resource requirements. Together, these models enable robust analysis and efficient data processing tailored to the needs of medical data applications in this project.

Chapter 3. MATERIALS AND METHODOLOGY

3.1. PTB-XL Dataset

The PTB-XL dataset is a large, publicly available dataset of clinical electrocardiogram (ECG) recordings, designed to support various research applications in ECG analysis, including deep learning and machine learning for medical diagnostics. Created by the Physikalisch-Technische Bundesanstalt (PTB), this dataset offers an extensive range of ECG data, which enables researchers to develop, validate, and benchmark new methods for automated ECG interpretation.

3.1.1. ECG Signals

The PTB-XL dataset comprises 21,837 clinical 12-lead ECG records collected from 18,885 patients. The recordings vary in length, with each ECG sample spanning to 10 seconds and sampled at a frequency of 100 Hz (Wagner et al., 2020). Each record in the dataset consists of 12 ECG leads, corresponding to standard clinical leads such as I, II, III, aVR, aVL, aVF, and six precordial (chest) leads (V1–V6). This extensive lead data provides a comprehensive view of cardiac electrical activity, offering a rich source for feature extraction and diagnostic assessment.

Each ECG record is stored in Waveform (.hea) files, which hold the raw ECG signals, and is accompanied by detailed metadata in a CSV file that includes patient information, recording specifics, and clinical annotations. The signals are stored in a standardized format to ensure consistency, facilitating reproducible research and interoperability across studies.

While the PTB-XL dataset includes all 12 standard ECG leads, researchers often selectively exclude certain leads for specific studies, particularly when reducing computational overhead or focusing on regions of cardiac activity. However, no leads are excluded by default in the dataset; all 12 are provided in their original form.

In many cases, leads such as the limb leads I, II, III, aVR, aVL, aVF and precordial leads V1 to V6 may be individually analysed or selected based on the study's specific requirements.

3.1.2. Clinical Annotations

A key feature of the PTB-XL dataset is its extensive set of clinical annotations, which provide ground truth labels for supervised learning and other diagnostic tasks. The annotations are structured into multiple categories, including diagnostic, form, rhythm, and beat annotations, each offering insights into several aspects of cardiac health:

Diagnostic Annotations: These are the primary annotations, covering a wide range of cardiac pathologies such as myocardial infarction, conduction disturbances, and arrhythmias. Diagnostic labels are assigned based on clinical criteria and organized into five hierarchical levels:

Superclasses: The highest level includes categories like Normal, Myocardial Infarction, Conduction Disturbance, Hypertrophy, and more.

Subclasses: Provides more specific pathology descriptors, for example, distinguishing between diverse types of myocardial infarction.

These hierarchical annotations are invaluable for researchers aiming to develop models that can generalize across a range of conditions or for studies focused on specific diagnoses.

Table 5. PTB-XL Diagnostic Superclasses and Subclasses

Superclass	Example Subclasses	Description
NORM	NORM	Normal ECG
MI	Anterior MI, Inferior MI, Lateral MI	Myocardial infarction (various locations)
STTC	ST Elevation, ST Depression, T Wave Inversion	ST-T abnormalities
CD	LBBB, RBBB, AV Blocks	Conduction disturbances

Superclass	Example Subclasses	Description
HYP	LVH, RVH	Ventricular hypertrophy
ISC	Ischemic Changes	Non-specific ischemia without infarction
RHY	AFib, AFlutter, PVC, SVT	Cardiac rhythm disorders
Others	Paced Rhythm, Early Repolarization, Low Voltage	Non-diagnostic or benign findings

Form Annotations: These labels describe waveform morphology, identifying patterns such as T-wave changes or QRS widening, which are often indicative of various cardiac conditions. Form annotations support the fine-grained classification of ECG signals based on waveform characteristics, aiding in nuanced analyses.

Rhythm Annotations: Rhythm labels provide information about the regularity and type of heart rhythm, for example sinus rhythm, atrial fibrillation. These annotations are essential for detecting arrhythmic events and assessing overall heart rate variability, a significant biomarker for cardiovascular health.

Beat Annotations: Beat labels denote individual heartbeat events, including the location of P, QRS, and T waves within the ECG signal. This detailed annotation allows for beat-to-beat analysis, which is particularly useful in high-resolution ECG assessments.

Each ECG record in the PTB-XL dataset can have multiple annotations from these categories, allowing for multi-label classification and supporting diverse research applications, from arrhythmia detection to waveform morphology analysis. The annotations are standardized and validated by clinical experts, ensuring their reliability for deep learning training purposes.

The structure of the PTB-XL dataset, with its rich composition of ECG leads, inclusive approach to lead data, and comprehensive clinical annotations, makes it a versatile and valuable resource for advancing ECG analysis and cardiovascular diagnostics.

3.2. Preprocessing Techniques

Effective preprocessing is essential to transform raw ECG signals from the PTB-XL dataset into a format that is optimized for deep learning algorithms.

The initial phase of preprocessing is focused on curating a dataset that accurately represents two distinct groups: healthy individuals and patients diagnosed with myocardial infarction (MI). The complete dataset comprises 21,837 clinical 12-lead ECG records obtained from 18,885 patients. To ensure a balanced dataset for subsequent analysis and model training, we selectively extract a subset of 7,086 ECG records, consisting of 3,543 records from healthy individuals and 3,543 from patients diagnosed with myocardial infarction. This balanced representation is crucial as it helps mitigate any potential bias that might arise from class imbalance, ensuring that both classes are equally represented. By maintaining this balance, the model can learn to distinguish effectively between the two conditions, ultimately enhancing the accuracy and reliability of the classification results. This approach also sets the foundation for more robust model development, facilitating fair evaluation across the two groups and supporting meaningful insights into the performance of the classification system.

3.2.1. Image Generation for ECG and aVR Lead Exclusion

The second step in preprocessing is to load the ECG signals and extract each lead's data from the dataset. The PTB-XL dataset provides recordings in .Hea (header) files, containing metadata and configurations for each ECG sample.

The .Hea file contains metadata about the ECG recording, including information such as:

Sampling Frequency: the sampling frequency of the PTB-XL dataset is 500 Hz and there is a down sampled version through a decimation process with 100 Hz.

Signal Information: The number of leads (12 in total) and names of each lead I, II, III, aVR, aVL, aVF, V1 to V6.

Amplitude Scaling: Specifies the scaling factors for each lead to convert digital units into physical units.

Recording Duration: Each ECG sample is 10 seconds long, which results in a matrix of size 5000 samples x 12 leads for each record (and 1000 x 12 for the 100 Hz).

In our case, the .Hea file provides metadata that allows us to interpret each signal correctly, and the data matrix in the file contains 12 columns (one per lead) with 1000 rows, where each row represents a sample point in time.

The following Python script reads a .Hea file, which contains metadata associated with the ECG records, and displays its contents for inspection. The .Hea file holds essential information, such as the number of leads, sampling frequency, and other parameters related to the ECG signal. This step is integral to understanding the structure of the data and ensuring that each record is correctly formatted before further processing.

Initially, the aVR lead was excluded from this analysis based on the limited research accessible at the time, which indicated its frequent neglect in routine clinical electrocardiogram (ECG) interpretation. This perspective was informed by suggestions within the literature that aVR often provides redundant information compared to other limb leads. However, further investigation has revealed that while aVR might be commonly overlooked, it can indeed offer crucial diagnostic insights in specific cardiac conditions. Therefore, future work should consider incorporating the aVR lead to potentially enhance the comprehensiveness of the ECG analysis.

```
import h5py
import matplotlib.pyplot as plt
import wfdb

data, info = wfdb.rdsamp(r'path to .hea file')
plt.plot(data)
plt.show()
```

Figure 3. Reading .hea file

After executing the code, the generated image represents an ECG diagram, where all the leads are stacked on top of each other in a single visual representation. This layout can make it difficult to analyse individual lead information effectively. In the context of our study, where precision and clarity are essential for accurate classification, we preferred to separate each lead into an individual image.

By doing so, we ensure that the distinct characteristics of each lead are preserved, improving the model's ability to learn and differentiate key features from the ECG signals. This separation also allows for better visualization and easier interpretation, as each lead can be examined independently for anomalies or significant patterns.

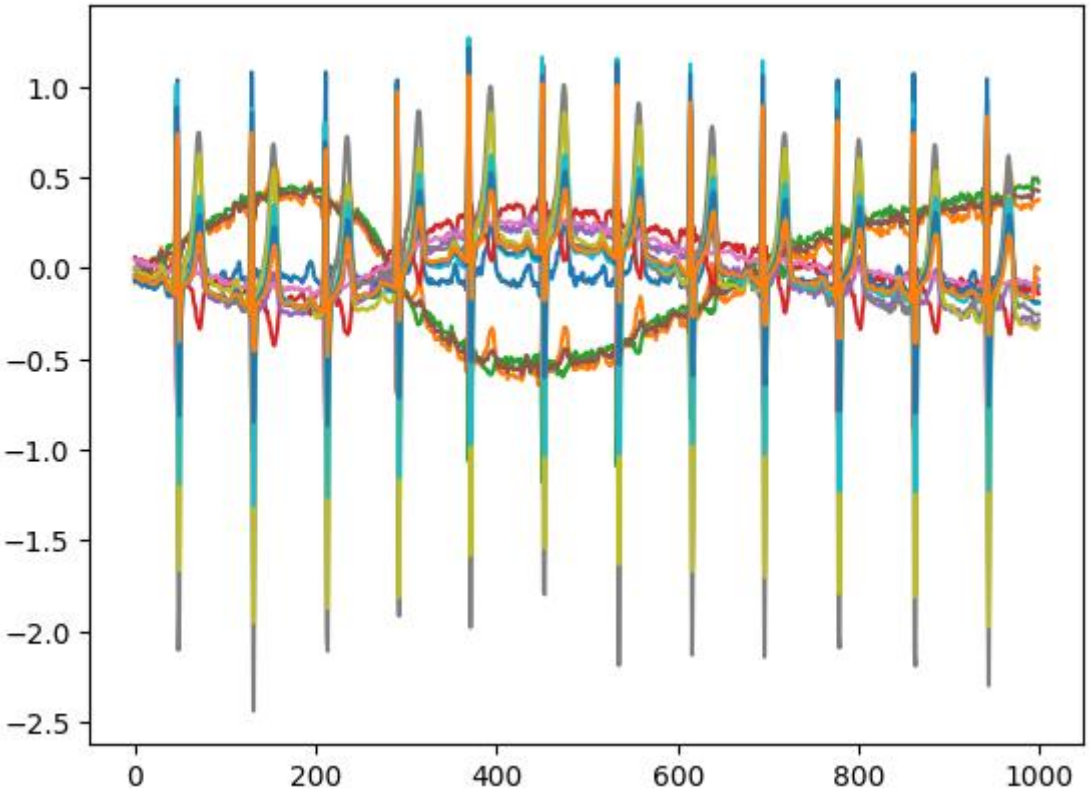


Figure 4. ECG diagram

3.2.2. Image Generation for Each Lead

The Python script below reads a .Hea file and extracts each ECG lead, omitting aVR as previously discussed. Each lead is visualized to verify the quality and integrity of the signals.

```
import wfdb
import matplotlib.pyplot as plt

# Read the sample data from the WFDB record
data, info = wfdb.rdsamp(r'path to .hea files')

# Get lead names from the info object
lead_names = info['sig_name']

# Plot each lead except 'aVR'
for i, lead_name in enumerate(lead_names):
    if lead_name != 'AVR':  # Skip 'aVR'
        plt.plot(data[:, i], label=lead_name)
        plt.title(f'Lead: {lead_name}')
        plt.xlabel('Sample Index')
        plt.ylabel('Amplitude')
        plt.show()
```

Figure 5. Lead separation and aVR exclusion

This step produces individual plots for each of the remaining 11 leads, which allows for a visual inspection of the signals to ensure they are correctly extracted and without noise or artifacts that could affect classification.

After loading the data, each lead's signal is converted into a separate image. This approach allows us to treat each lead's time-series data as a 2D image, facilitating deep learning approaches that are well-suited to image data. In this study, each ECG recording is converted into images corresponding to 11 leads, each representing the temporal amplitude variation of the heart's electrical activity.

The images generated for each lead are then stored as separate PNG files, with each file representing the time-series waveform of a lead. This provides a structured dataset that can be processed through convolutional neural networks (CNNs) or other deep learning architectures.

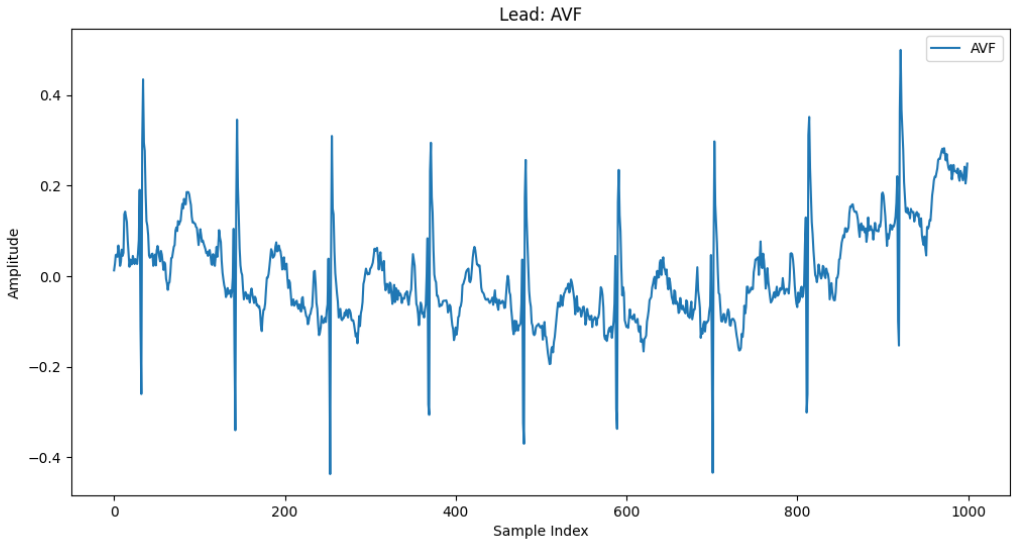


Figure 6. Sample of an AVF Lead

3.2.3. Resizing and Normalization

Once the lead images are generated, resizing and normalization are applied to standardize the input dimensions and scale across the dataset. Consistent image dimensions and pixel value scaling ensure uniformity, which is critical for model performance and comparability.

Resizing: Each generated image is compressed to a specific dimension 300 x 300 pixels, as required by EfficientNetB3. Resizing allows for compatibility with pre-trained models and reduces computational demands without losing significant detail.

Normalization: Image pixel values are normalized to a range between 0 and 1. This is a frequent practice in image preprocessing, ensuring that each feature has similar ranges and preventing any pixel value range from dominating the model's learning process.

```python
from PIL import image
import numpy as np

def preprocess_image(image_path, target_size(select your, target size)):
    # Open the image file
    img = Image.open(image_path).convert('RGB') # Convert RGBA to RGB
    # Resize the image
    img_resized = img.resize(target_size)
    # Normalize pixel values
    img_normalized = np.array(img_resized) / 255.0
    return img_normalized
```

Figure 7. Image normalization

This code snippet converts each image to RGB format (in case of RGBA), resizes it to 300 x 300 pixels, and normalizes pixel values. These standardized images can then be used as input for deep learning models.

3.2.4. Final Structure of the Processed Dataset

After completing the above steps, each ECG recording in the dataset is transformed into a collection of 11 images, one per lead, with consistent sizing and normalized pixel values. The final processed dataset is organized as follows:

Dataset structure: the dataset is composed of two directories Healthy and Myocardial Infarction each one contains 3543 folders.

File Structure: Each sample has a folder containing images for leads I, II, III, aVL, aVF, V1–V6.

Data Exclusions: The aVR lead is excluded from each sample for reasons outlined previously.

Image Specifications: Each lead image is a 300 x 300 pixel RGB image, with pixel values normalized to a range of 0 to 1.

This structured dataset provides an optimized input for deep learning, allowing for efficient training of models with a focus on the most informative data.

Each preprocessing step is designed to ensure that the input data is as informative, consistent, and optimized as possible for deep learning models. By transforming raw ECG signals into standardized images, we harness the power of image-based analysis with deep learning methods. This approach allows for efficient feature extraction and pattern recognition, which are critical for accurate classification. These preprocessing techniques collectively prepare the PTB-XL dataset to serve as a robust input for models tasked with distinguishing MI cases from healthy ECGs.

3.3. Model Architecture

In this study, we present a hybrid deep learning model specifically designed for the classification of Electrocardiogram (ECG) images. This model leverages a combination of EfficientNet-B3 as a feature extractor and a Linformer-based attention mechanism, which together capture both the local and global relationships within the ECG data. The proposed architecture is meticulously crafted to process the spatial features of individual ECG leads while also accounting for the intricate dependencies that exist between leads. These dependencies are critical for improving the model's accuracy in differentiating between healthy and myocardial infarction (MI) conditions.

3.3.1. EfficientNet-B3 Backbone

At the core of our model lies EfficientNet-B3, a state-of-the-art deep convolutional neural network known for its optimal balance between model complexity and computational efficiency. EfficientNet-B3 has been pretrained on the ImageNet dataset, which provides a rich and diverse set of image features, making it well-suited for transferring learned knowledge to the task of ECG image classification. This pretrained backbone enables the model to extract meaningful and robust features from each ECG lead, producing a 1536-dimensional feature vector that encapsulates both low and high level visual characteristics for each one of the 11 ECG leads.

The selection of EfficientNet-B3 over larger models such as EfficientNet-B7 was driven by its performance in achieving high accuracy while maintaining lower computational requirements. The trade-off between accuracy and computational resources is particularly important when working with high-dimensional ECG data, where the sheer volume of input features can strain system resources.

EfficientNet-B3 can process these data efficiently without compromising the model's capacity to capture detailed patterns and anomalies in the ECG signals.

3.3.2. Linformer Attention Block

To model the complex inter-lead dependencies inherent in ECG data, we introduced a Linformer attention block (S. Wang et al., 2020) . The Linformer, an advanced attention mechanism designed to overcome the computational bottleneck of traditional self-attention, employs a linear-time approximation, which allows it to process large-scale dependencies with significantly reduced computational complexity. This makes it a highly effective choice for handling the inter-lead relationships in ECG data, where each lead provides a time-series representation of the heart's electrical activity that can influence the interpretation of other leads.

The Linformer block operates by applying multi-head attention across the sequence of lead-based features extracted by EfficientNet-B3 in form of 11 feature vectors and its output is also a sequence of 11 feature vectors, each still 1536-dimensional, but now they contain information about how the leads relate to each other. Specifically, it uses four attention heads to process the sequences in parallel, followed by a feedforward layer with 128 units, which further refines the captured information. This configuration is designed to enhance the model's ability to capture long-range dependencies between leads while maintaining computational efficiency. The adoption of the Linformer's linear attention mechanism ensures that the model remains scalable even as the number of leads and the complexity of the input data increases. This feature is particularly advantageous when dealing with ECG data, where inter-lead correlations play a crucial role in the accurate classification of myocardial infarction.

3.3.3. Fully Connected Layers and Sigmoid Binary Classifier

The 11 feature vectors from the Linformer are then flattened into a single large vector (11 * 1536 dimensions). This flattened vector is passed through one or more fully connected layers to further process and reduce the dimensionality of the combined information. To mitigate overfitting, dropout regularization is applied to the FC layer with a dropout rate of 0.5. This helps to prevent the model from relying too heavily on specific features, ensuring that the network generalizes well to unseen data. The dropout rate was determined through rigorous hyperparameter optimization, ensuring an effective balance between model complexity and regularization.

The final fully connected layer outputs a single value (a logit). This value is then passed through a sigmoid function during evaluation to get a probability between 0 and 1, which is used for the binary classification either healthy or with myocardial infarction.

The model's ability to make these predictions with high accuracy is supported by the synergy between the EfficientNet-B3 backbone, which extracts spatial features from each lead, and the Linformer attention mechanism, which captures the relationships across leads.

Together, this hybrid architecture harnesses the power of both convolutional and attention-based models to process ECG images effectively. By combining the localized feature extraction capabilities of EfficientNet-B3 with the global attention provided by the Linformer, the model achieves superior performance in detecting subtle patterns and anomalies in ECG data. This design ensures that the model can handle the complexities of multi-lead ECG images while preserving crucial signal information, making it a powerful tool for ECG classification tasks.

In conclusion, the integration of EfficientNet-B3 and Linformer provides a balanced solution that captures both local and global features of ECG data, ensuring efficient processing and high accuracy. This model architecture is well-suited for real-world applications, where timely and accurate diagnoses of myocardial infarction can have significant clinical implications.

3.4. Hyperparameter Optimization

In this study, we utilized two distinct hyperparameter tuning strategies to optimize the performance of our hybrid deep learning model. The first technique involved the use of Optuna, a modern, efficient optimization framework for hyperparameters. Optuna employs state-of-the-art algorithms, such as Tree-structured Parzen Estimators (TPE), to efficiently explore the hyperparameter space and identify optimal configurations. This approach enabled us to automate the search process, quickly identifying well-performing combinations of hyperparameters, thus enhancing the model's performance while minimizing manual effort. Through Optuna, we were able to achieve significant improvements in model accuracy and generalization, highlighting the effectiveness of automated optimization.

The second tuning approach involved Manual and Dynamic hyperparameter tuning, this method required a more hands-on approach, where we adjusted hyperparameters manually based on empirical observations and prior knowledge of model behaviour. Dynamic tuning allows for flexibility during the training process, adapting parameters like learning rate, batch size, and dropout rate in real-time to respond to the model's evolving performance. This

technique is especially valuable in scenarios where automated tuning may not fully capture the nuances of model behaviour across different training epochs.

Manual tuning, while more time-consuming, allows for fine-grained control over the model's hyperparameters, particularly when dealing with complex datasets like ECG images.

The key difference between these two techniques lies in their approach:

> **Optuna** automates the search for optimal hyperparameters, relying on advanced algorithms to efficiently explore the hyperparameter space,
>
> **Manual tuning** requires more direct involvement from the researcher, adjusting parameters based on feedback from the model during training.

Both strategies offer unique advantages and were used to complement each other in our work, ensuring both thorough exploration and fine-tuning of the model.

Each of these hyperparameter tuning methods will be discussed in more detail in the subsequent sections, where we will describe the specific configurations, the search spaces, and other . This comprehensive tuning process played a crucial role in optimizing the model’s performance and ensuring its robustness in ECG classification tasks.

3.4.1. Hyperparameter Optimization with Optuna and Adam Optimizer

Optuna was seamlessly integrated into the model training pipeline to automate the search for optimal hyperparameter configurations. In this approach, a study was defined where 20 trials were generated, with each trial representing a unique combination of hyperparameters, such as learning rates, dropout rates, and attention block settings. The model was trained for 8 epochs in each trial to efficiently evaluate the impact of the hyperparameters, balancing the need for quick results with computational efficiency. Additionally, Optuna's dynamic pruning feature was utilized to terminate trials early if they were underperforming, thus conserving computational resources. As a result, the integration of Optuna allowed for the efficient exploration of the hyperparameter space, with 20 distinct trial outcomes produced, ultimately aiding in the identification of the most promising configurations for ECG classification.

Hyperparameter optimization was performed using Optuna, a framework designed for automatic hyperparameter tuning.

The following hyperparameters were considered for optimization:

Batch size: The batch size was optimized within the range of 8 to 32. Larger batch sizes can provide more stable gradient estimates, but they also require more memory, so the optimal batch size is a balance between performance and memory usage.

Learning rate: The learning rate was optimized between 5×10^{-5} and 5×10^{-4}. A lower learning rate helps to achieve finer convergence, while a higher rate speeds up training but may risk overshooting the optimal solution.

Linformer depth: The depth of the Linformer attention mechanism, ranging from 6 to 12 layers, was tuned. Deeper models can capture more complex relationships between ECG leads, but they also increase computational complexity.

Optuna's optimization process iteratively explored these hyperparameters to identify the configuration that maximized the model's performance on the validation set.

The Adam optimizer (Adaptive Moment Estimation) was chosen for training due to its adaptive learning rates, which provide efficient convergence even with complex models like ours (Kingma & Ba, 2014). Compared to traditional optimizers like SGD, Adam's ability to adjust learning rates based on first and second moments of the gradient ensures more stable training, especially for models with many parameters, like the EfficientNet-B3 backbone.

The optimizer was configured with a learning rate that was selected through Optuna's search process, ensuring that the model converged quickly without overshooting the optimal solution (Akiba et al., 2019). This allowed us to strike a balance between training speed and model accuracy.

3.4.2.Manual and Dynamic hyperparameter tuning techniques and Adam Optimizer

Hyperparameter optimization was conducted using Manual and Dynamic hyperparameter tuning techniques, an automatic optimization framework. The goal was to identify the best configuration for training parameters to achieve high performance while avoiding overfitting.

The key hyperparameters optimized are described below:

Batch size: The batch size was optimized within a range of 8 to 16, with the optimal value determined as 8 based on validation performance. Smaller batch sizes provided more granular updates, improving the model's ability to generalize, especially with a limited dataset.

Learning rate: The learning rate was optimized between 5×10^{-5} and 5×10^{-4}, with the optimal value selected as$1.5x10^{-4}$. This learning rate provided a good balance, ensuring stable and efficient convergence without overshooting the optimal solution.

Linformer depth: The depth of the Linformer attention block was tuned between 6 and 10 layers, and the best-performing model used 6 layers. A shallower Linformer depth reduced computational cost while still capturing essential inter-lead relationships in ECG signals.

Manual and Dynamic hyperparameter tuning technique's optimization process iteratively explored these hyperparameters to maximize the validation AUC, which served as the primary performance metric during tuning. The selected configuration achieved a high AUC and stable convergence across trials.

The Adam optimizer was selected for this training too, as it offers adaptive learning rates and efficient convergence for deep models with many parameters, such as EfficientNet-B3 (Kingma & Ba, 2014). The optimizer was configured with the learning rate obtained through Manual and Dynamic hyperparameter tuning process.

3.4.3. Fine-Tuning Strategy

Following the initial training phase, a carefully crafted fine-tuning strategy was employed exclusively during the Manual and Dynamic hyperparameter tuning techniques to enhance the model's performance and adaptability to the ECG dataset. This process involved progressively unfreezing the final five convolutional layers of the EfficientNet-B3 backbone, thereby allowing these layers to be retrained while keeping the earlier layers frozen. This hierarchical approach ensured that the model retained the valuable pretrained features from ImageNet in the

early layers, which capture general patterns such as edges and textures, while the newly unfrozen layers adapted specifically to the nuances of ECG signals.

The fine-tuning process was particularly crucial in enabling the model to discern subtle patterns in the ECG diagrams, such as variations in waveform morphology, that are indicative of myocardial infarction. This step allowed the model to learn domain-specific features that the initial frozen layers could not fully capture.

This strategy allowed for greater flexibility in optimizing the model's performance by dynamically adjusting hyperparameters such as learning rates, dropout rates, and the number of trainable layers. By integrating fine-tuning with these more flexible tuning methods, we ensured that the model not only converged efficiently but also achieved enhanced generalization across the validation dataset.

Overall, fine-tuning played a pivotal role in bridging the gap between leveraging general-purpose pretrained networks and crafting a high-performing, domain-specific ECG classification model. The combination of selective unfreezing and domain-specific training demonstrated its effectiveness in optimizing the model's capability to distinguish between healthy and myocardial infarction classes.

3.4.4. ReduceLROnPlateau and Early Stopping

To mitigate overfitting and promote efficient training, an early stopping mechanism and a ReduceLROnPlateau were implemented during both dynamic and Optuna-driven hyperparameter tuning techniques.

The ReduceLROnPlateau learning rate scheduler is a crucial technique for adapting the learning rate during training to improve convergence and prevent overfitting (Al-Kababji et al., 2022) . This scheduler dynamically reduces the learning rate by a factor of 0.5 whenever the validation AUC (Area Under the Curve) metric stagnates for a specified number of consecutive epochs, in this case 3 epochs. By monitoring the validation AUC, this method identifies when the model's performance has plateaued, indicating that further training with the current learning rate may lead to diminishing returns or potential overshooting of the optimal solution.

The primary advantage of the ReduceLROnPlateau scheduler is its ability to allow the model to continue training effectively without the risk of overshooting the optimal minima, especially during the later stages of training.

By lowering the learning rate after the plateau is detected, the scheduler enables finer adjustments to the weights, facilitating a more precise convergence toward the optimal solution. This approach not only helps in preventing the model from getting stuck in local minima but also enables it to navigate the complex landscape of the loss function more efficiently.

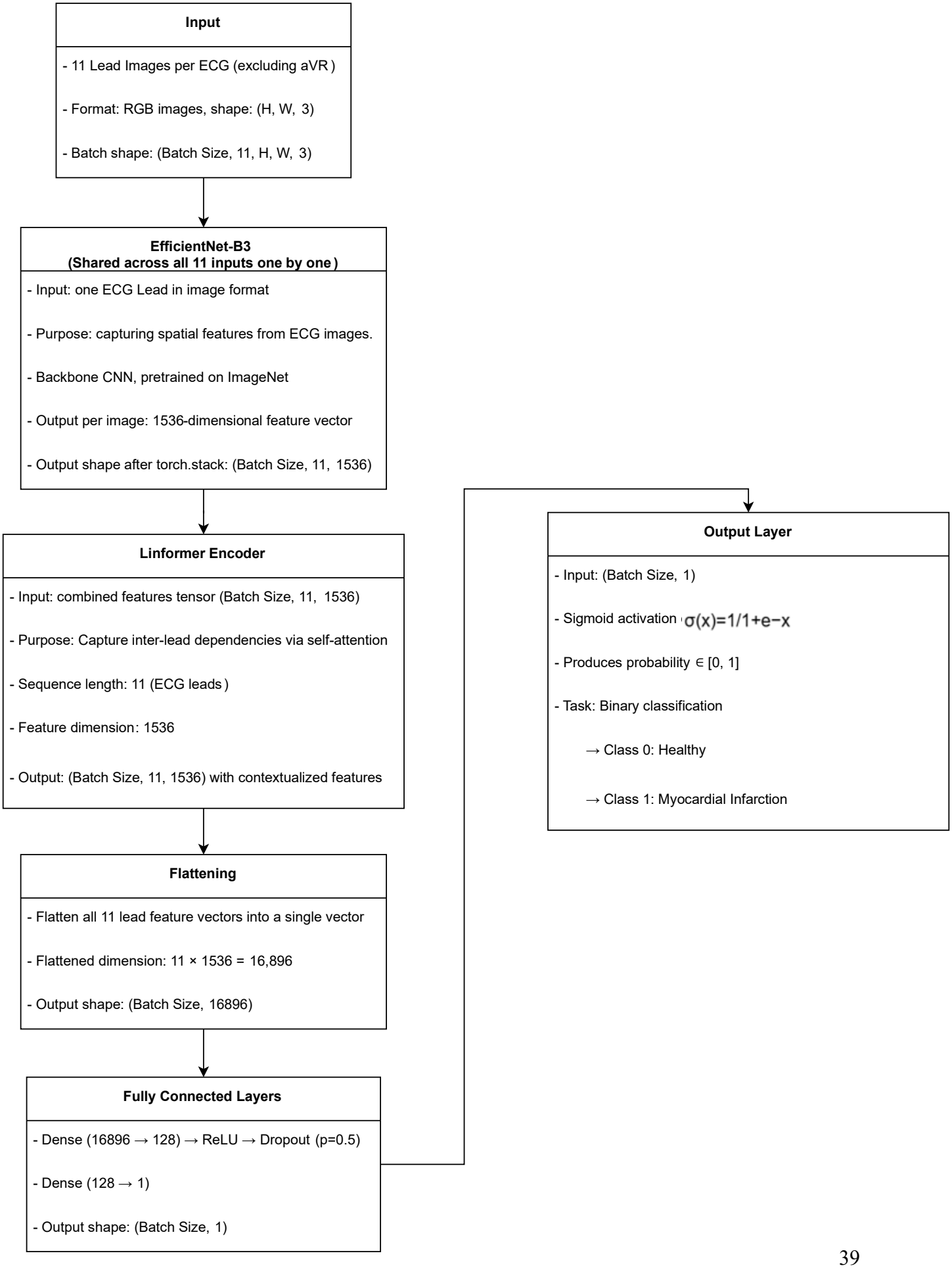
The ability to reduce the learning rate dynamically ensures that the model can maintain steady progress throughout training while avoiding drastic changes in the weight updates that could destabilize the learning process. It also contributes to faster convergence, as the model can focus on refining the learned features rather than making large, potentially erratic weight updates. Overall, this learning rate scheduler played a key role in optimizing the balance between training speed and model accuracy, enabling the model to adapt its learning rate based on the performance dynamics observed during training. This resulted in a more efficient training process and improved final model performance, particularly when handling complex datasets such as the ECG images in this study.

To maintain model generalization and avoid overfitting, an early stopping strategy was employed during training. This technique continuously monitored the validation AUC and halted the training process if no improvements were observed for five consecutive epochs.

The decision to stop training early allowed the model to avoid unnecessary training iterations, reducing the risk of overfitting and preserving computational resources. By leveraging this approach, the model maintained a balance between training efficiency and classification accuracy, ensuring robust performance for distinguishing between healthy and myocardial infarction ECGs.

By employing a patience parameter of 5 epochs, early stopping ensured that the model did not continue training unnecessarily when no further performance gains were achievable. This approach not only safeguarded against overfitting but also optimized the computational efficiency of the training pipeline by preventing redundant training cycles.

The use of early stopping proved instrumental in maintaining a balance between model complexity and generalization capability, contributing to the model's ability to effectively classify healthy and myocardial infarction ECG records without overfitting to the training data.


Input
- 11 Lead Images per ECG (excluding aVR)
- Format: RGB images, shape: (H, W, 3)
- Batch shape: (Batch Size, 11, H, W, 3)
EfficientNet-B3
(Shared across all 11 inputs one by one)
- Input: one ECG Lead in image format
- Purpose: capturing spatial features from ECG images.
- Backbone CNN, pretrained on ImageNet
- Output per image: 1536-dimensional feature vector
- Output shape after torch.stack: (Batch Size, 11, 1536)
Linformer Encoder
- Input: combined features tensor (Batch Size, 11, 1536)
- Purpose: Capture inter-lead dependencies via self-attention
- Sequence length: 11 (ECG leads)
- Feature dimension: 1536
- Output: (Batch Size, 11, 1536) with contextualized features
Flattening
- Flatten all 11 lead feature vectors into a single vector
- Flattened dimension: 11 × 1536 = 16,896
- Output shape: (Batch Size, 16896)
Fully Connected Layers
- Dense (16896 → 128) → ReLU → Dropout (p=0.5)
- Dense (128 → 1)
- Output shape: (Batch Size, 1)
Output Layer
- Input: (Batch Size, 1)
- Sigmoid activation σ(x)=1/1+e−x
- Produces probability ∈ [0, 1]
- Task: Binary classification
→ Class 0: Healthy
→ Class 1: Myocardial Infarction

Chapter 4. RESULTS AND DISCUSSION

Before delving into the results, it is essential to provide a detailed overview of the key evaluation metrics used in assessing model performance. These metrics offer valuable insights into several aspects of classification quality, including the model's ability to correctly identify myocardial infarction cases, maintain balanced predictions, and generalize well across unseen data. Understanding these metrics is crucial for interpreting and comparing the outcomes obtained during various training phases.

Accuracy: measures the proportion of correctly predicted instances (both positive and negative) out of the total number of instances.

$$Accuracy = \frac{True\ Positives\ (TP) + True\ Negatives\ (TN)}{Total\ Samples} \quad (1)$$

While accuracy is useful for balanced datasets, it may be misleading for imbalanced datasets.

Precision: also known as Positive Predictive Value, evaluates the proportion of correctly identified positive predictions among all positive predictions.

$$Precision = \frac{TP}{TP + False\ Positives\ (FP)} \quad (2)$$

High precision indicates a low false positive rate.

Recall: also called Sensitivity or True Positive Rate, measures the ability of the model to correctly identify positive instances.

$$Recall = \frac{TP}{TP + False\ Negatives\ (FN)} \quad (3)$$

High recall indicates a low false negative rate.

F1-Score: the harmonic mean of Precision and Recall, providing a balance between the two, especially relevant in cases of imbalanced datasets.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

A high F1-score indicates a good trade-off between Precision and Recall.

Specificity: Measures the model's ability to correctly identify negative cases. It is the proportion of true negatives out of all actual negatives, indicating how well the model avoids false positives. A high specificity means the model reliably excludes negative instances, reducing the risk of incorrectly flagging them as positive.

Area Under the Curve (AUC): evaluates the model's ability to distinguish between classes by plotting the True Positive Rate (Recall) against the False Positive Rate (1 - Specificity) at various thresholds. AUC ranges from 0 to 1, where 1 represents perfect discrimination, and 0.5 indicates random guessing.

Training Loss: Measures how well the model fits the training data. A steady decline indicates effective learning, while sudden drops or plateaus suggest potential overfitting or optimization issues.

Validation Loss: Reflects the model's ability to generalize to unseen data. Lower validation loss alongside reduced training loss is indicative of a robust model.

4.1. Hyperparameter Tuning using Optuna

This section presents the performance evaluation of several model training trials conducted using Optuna-driven hyperparameter tuning for ECG classification. The models leverage EfficientNetB3 as a feature extractor, combined with Linformer blocks to capture inter-lead dependencies. The objective of these trials was to optimize classification performance by exploring diverse hyperparameter configurations, including learning rates, dropout rates, and attention block parameters. Key evaluation metrics such as training loss, validation loss, AUC (Area Under the Curve) and F1-Score were employed to comprehensively assess the models' performance.

4.1.1. Results and Metrics analysis

This subsection presents a detailed analysis of the performance metrics obtained during the hyperparameter optimization process using Optuna. For each of the top-performing trial configurations identified by Optuna, we examine the trends of key evaluation metrics, including Area Under the Curve (AUC), accuracy, precision, recall, and F1 score, over the training epochs. This analysis aims to provide insights into the learning behaviour of different model configurations, highlighting their strengths, weaknesses, and stability throughout the training process. Visualizations of these metrics across epochs for representative trials are provided to facilitate a comparative understanding of their performance dynamics.

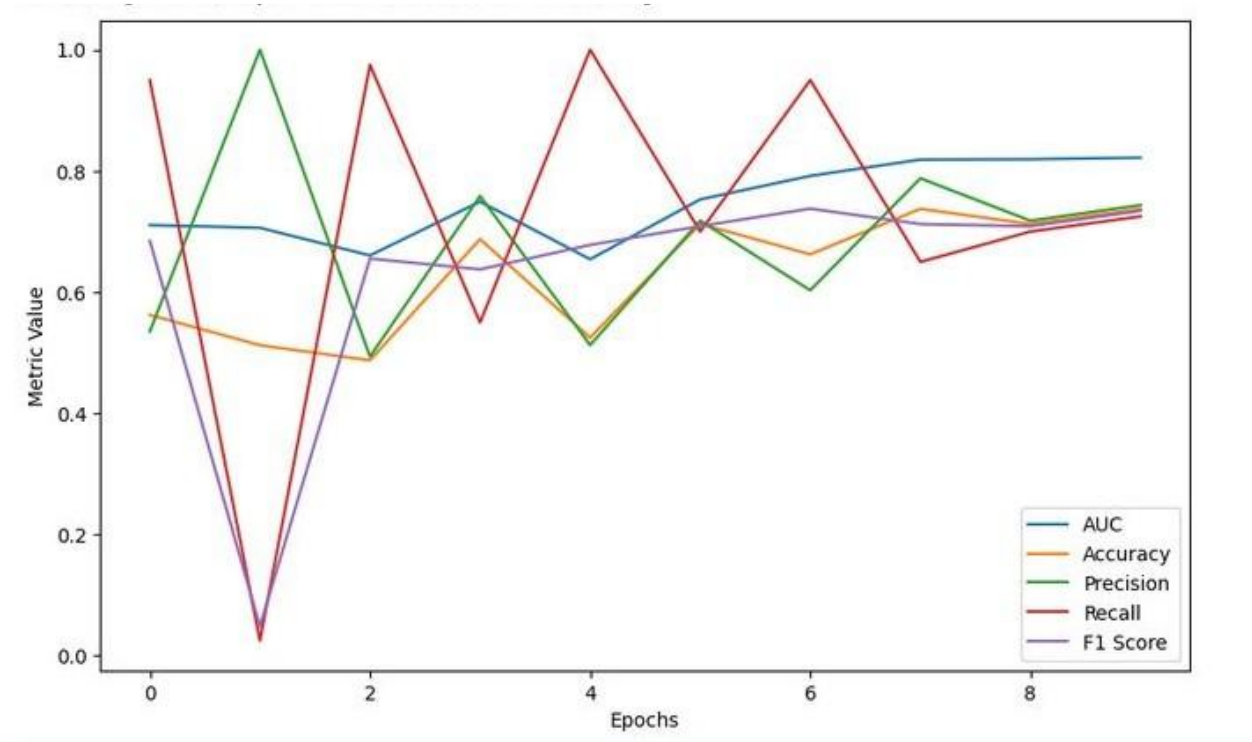


Figure 8. Trial 1 (Optuna)

Batch size = 15

Learning rate = 0.00010663035259145652

Linformer depth = 8

Metrics Analysis Over Epochs:

AUC (Area Under Curve): AUC starts at around 0.75 and remains relatively stable with slight improvements across the epochs, fluctuating between 0.7 and 0.8. This consistent AUC suggests that the model can distinguish between classes to a moderate degree, though it does not reach optimal separation performance (closer to 1.0). The stability in AUC implies that this model configuration retains class separability but might benefit from a higher learning rate or further adjustments to improve the AUC further.

Accuracy: Accuracy starts at around 0.58, rises to about 0.65 by epoch 3, and then gradually improves until it reach 0.77 by epoch 9. This suggests that the model seems to be learning progressively but hasn't converged to high accuracy. The stability toward later epochs indicates limited overfitting but suggests possible constraints imposed by the batch size or learning rate.

Precision: Precision exhibits sharp fluctuations, starting near 0.6, dropping drastically near epoch 2, then recovering to stabilize between 0.6 and 0.7.

The drop in precision could indicate instability in the model's handling of false positives, possibly due to the batch size of 15. This fluctuation may also reflect difficulties in handling class-specific patterns. The recovery in precision shows the model's capacity to adjust, though this erratic behaviour could be minimized with a smaller batch size or a slightly deeper Linformer block.

Recall: Recall begins near 0.65, dips to almost zero around epoch 2, and stabilizes at approximately 0.7 by epoch 8. The dip followed by recovery suggests temporary failure in detecting true positives. This could result from unstable gradient updates due to the learning rate or the Linformer depth configuration 8. The recovery is promising, demonstrating that the model adjusts its recall capabilities.

F1 Score: The F1 Score follows the fluctuations in precision and recall, starting near 0.65, dropping significantly around epoch 2, and recovering to stabilize around 0.75. The erratic behaviour highlights convergence challenges likely due to hyperparameter choices (batch size and depth). Despite this, the recovery indicates that the model eventually regains balanced performance.

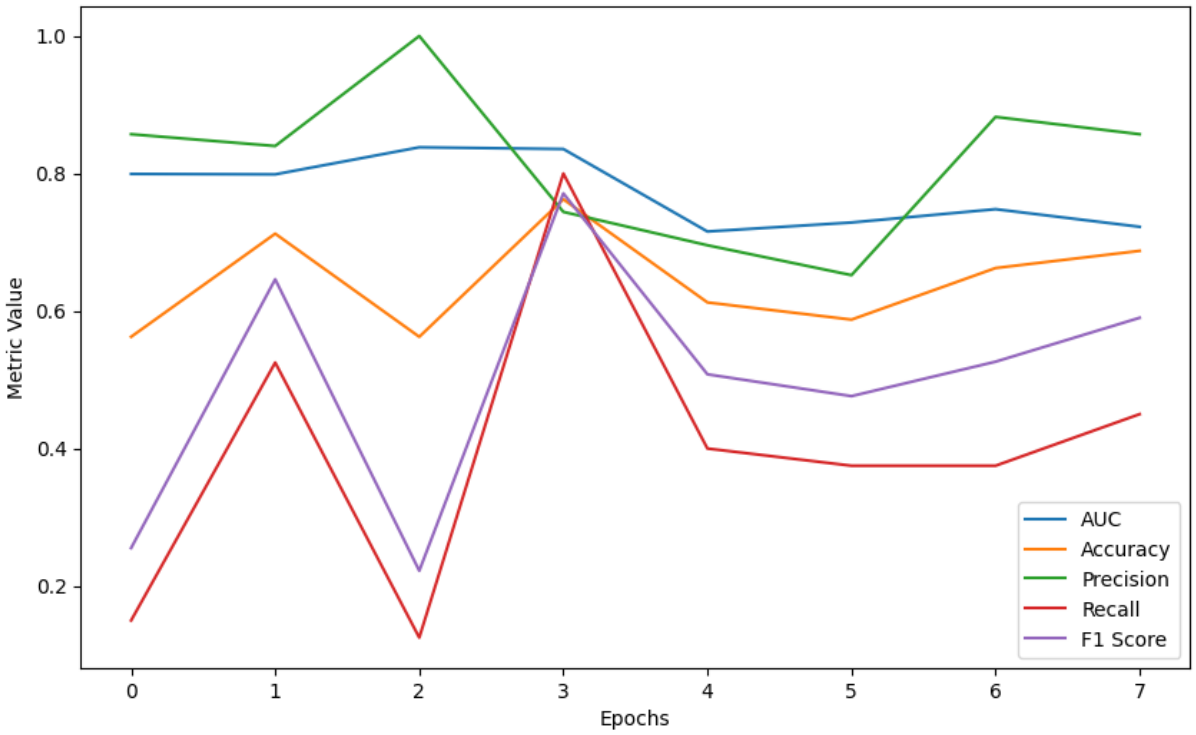


Figure 9. Trial 2 (Optuna)

Batch size = 18

Learning rate = 0.00006075

Linformer depth = 6

Metric Analysis Over Epochs:

AUC (Area Under Curve): The AUC starts around 0.75 and stays relatively stable across epochs, showing a slight decline and oscillation. The values are consistently between 0.7 and 0.8, which suggests the model has a moderate ability to distinguish between classes but lacks significant improvement. This stability might indicate early saturation, possibly due to the low learning rate or shallow Linformer depth.

Accuracy: Accuracy begins around 0.6, increases to approximately 0.65 by epoch 2, but then decreases slightly and stabilizes around 0.6 in later epochs. This plateau suggests limited learning and potential issues with generalization, possibly due to the low learning rate or limited model complexity with the Linformer depth set to 6.

Precision: Precision starts high at around 0.9, peaks at epoch 2, but then sharply drops around epoch 3. After this drop, it stabilizes around 0.7 for the remaining epochs. This initial peak followed by a sudden drop might indicate overfitting on certain positive instances early in training, after which the model struggles to maintain high precision due to insufficient capacity to generalize.

Recall: Recall begins low, showing an upward trend until a sharp peak at epoch 3, then declines and stabilizes around 0.3. This peak and subsequent drop suggest the model initially learns to identify positive instances but fails to maintain recall, potentially due to instability in parameter updates or a lack of generalization due to the limited Linformer depth and low learning rate.

F1 Score: The F1 Score, balancing precision and recall, starts at around 0.5, peaks at epoch 2, but then follows the recall trend, decreasing and stabilizing around 0.4. The F1 Score mirrors the fluctuations in precision and recall, indicating that the model has difficulty balancing these metrics consistently over epochs.

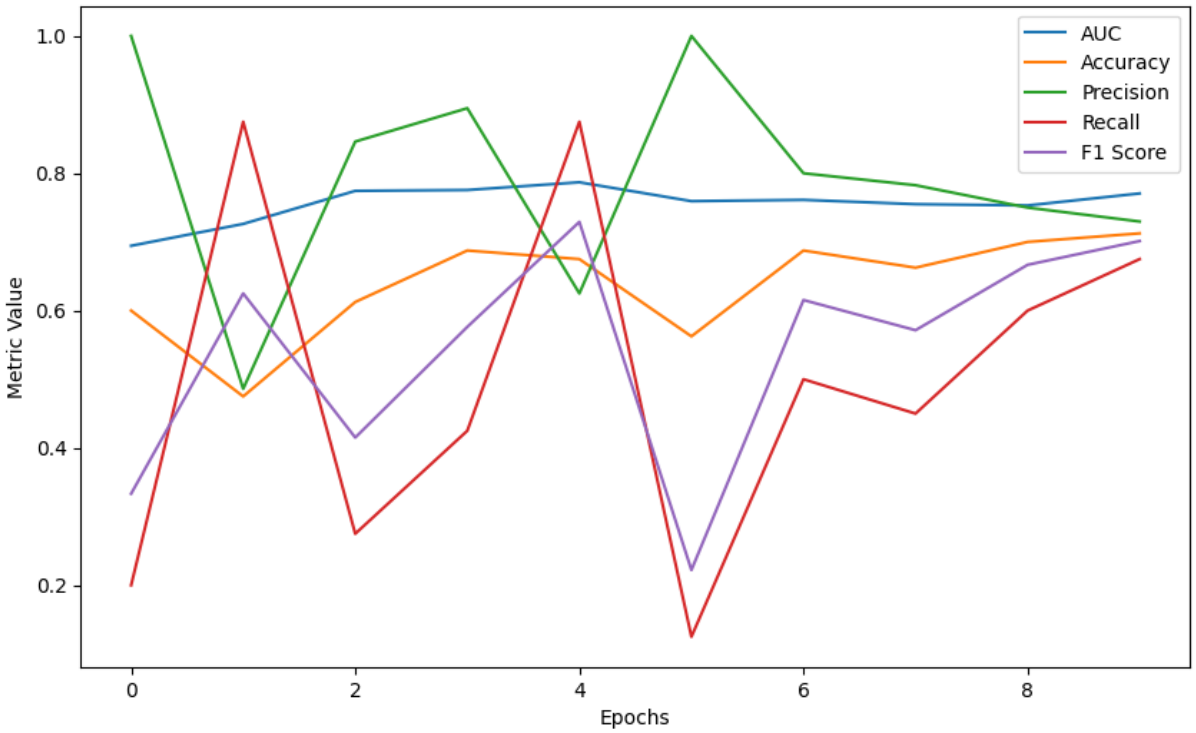


Figure 10. Trial 3 (Optuna)

Batch size = 22

Learning rate = 0.0000676

Linformer depth = 8

Metric Analysis Over Epochs:

AUC (Area Under Curve): The AUC remains stable, oscillating around 0.7 with minimal fluctuations. This stability suggests the model has a moderate and consistent capacity to separate classes but lacks significant improvement over epochs. The low learning rate might prevent the model from making more meaningful adjustments.

Accuracy: Accuracy begins at a moderate level, around 0.6, and gradually increases to about 0.7 by the final epochs. This slow but steady improvement suggests the model is learning progressively, though it may be limited by the chosen learning rate and Linformer depth, which restrict further gains in performance.

Precision: Precision exhibits a highly volatile pattern, starting near 1.0, dropping drastically in the first few epochs, then oscillating between 0.6 and 0.9 in later epochs.

The high initial precision suggests an initial bias towards identifying positives accurately, but the sharp fluctuations point to instability in identifying true positives consistently, potentially due to the low learning rate and depth.

Recall: Recall starts at a low level, improves dramatically by epoch 1, but then drops sharply and oscillates throughout training. The low learning rate may contribute to the erratic recall performance, as it limits the model's capacity to adjust sufficiently, leading to an unstable representation of positive cases.

F1 Score: The F1 Score follows a fluctuating pattern like recall, highlighting the model's struggle to balance precision and recall over time. Despite some improvements, the fluctuations imply difficulty in maintaining stable performance across epochs, reflecting an issue with convergence stability.

4.1.2. Discussion for Optuna models

Based on the metric analysis and conclusions for each configuration:

The first configuration (batch size 15, learning rate 0.000107, Linformer depth 8) showed the most stable AUC values, oscillating moderately around 0.7 to 0.8, indicating consistent class separability. This configuration reached around 0.77 accuracy, higher than others, indicating moderate but steady learning. Even though it had some instability but recovers quickly after epoch 5. By contrast, configurations with lower learning rates or larger batch sizes show more pronounced fluctuations in recall and precision, indicating difficulties with convergence stability.

The first configuration achieves a balanced performance with relatively steady AUC and accuracy around 0.77, with a learning rate of 0.000107, which is higher than some other configurations and contributes to improved convergence. This setup strikes a better balance between the need for convergence speed (from the higher learning rate) and stability in the metrics (from the moderate batch size and Linformer depth).

The first configuration "figure 8" (batch size 15, learning rate 0.000107, Linformer depth 8) is selected as the best-performing configuration due to its overall stability and balance across metrics. This setup offers a consistent AUC, moderate accuracy, and a quick recovery in precision and recall after initial instability, making it the most reliable choice. Adjustments to this configuration, such as further increasing the learning rate slightly or deepening the Linformer depth, could potentially yield further improvements.

Despite the promising results achieved with the selected configuration, it is crucial to acknowledge certain practical limitations encountered during the training process. The models were trained from 8 to 10 epochs (early stopping mechanism) as part of the 20 trials generated by Optuna's hyperparameter tuning process. Each trial demanded significant computational resources, with testing alone requiring approximately 7,564 minutes using an Intel(R) Celeron(R) N4120 CPU @ 1.10GHz 1.10 GHz and 8.00 GB RAM memory.

While these initial results provide a valuable starting point, further steps were necessary to validate and improve the model's performance. As a follow-up, trials were conducted using Manual and Dynamic hyperparameter tuning techniques to investigate whether they could yield superior results. This approach involved systematically exploring configurations such as adjusting learning rates, fine-tuning dropout rates, and selectively unfreezing layers.

The outcomes of these trials and their discussions will be presented in the next section, highlighting potential improvements and insights gained compared to the results obtained through Optuna-based optimization. These efforts underscore the commitment to refining the model for more robust and effective ECG classification performance.

4.2. Manual and Dynamic Hyperparameter Tuning

This analysis evaluates the performance of three model training trials designed for ECG classification. The architecture of each model leverages EfficientNetB3 as a feature extractor, combined with Linformer blocks for sequence modelling. The primary goal of these trials was to achieve optimal classification performance by dynamically tuning hyperparameters, unfreezing layers, and adapting dropout rates during training. The models were assessed using key performance metrics such as training loss, validation loss, AUC (Area Under the Curve), and F1-Score.

All three models shared a common core but differed slightly in their training dynamics:

EfficientNetB3 Base Model: Used as a feature extractor. In early epochs, its layers were frozen to accelerate training. Gradually, additional layers were unfrozen for fine-tuning.

Linformer Block: Processes multi-lead ECG image features as sequential data, capturing inter-lead correlations.

Dynamic Adjustments: Each trial included mechanisms to adjust learning rates, dropout rates, and fine-tune additional layers based on validation performance.

4.2.1. Results and Metrics analysis

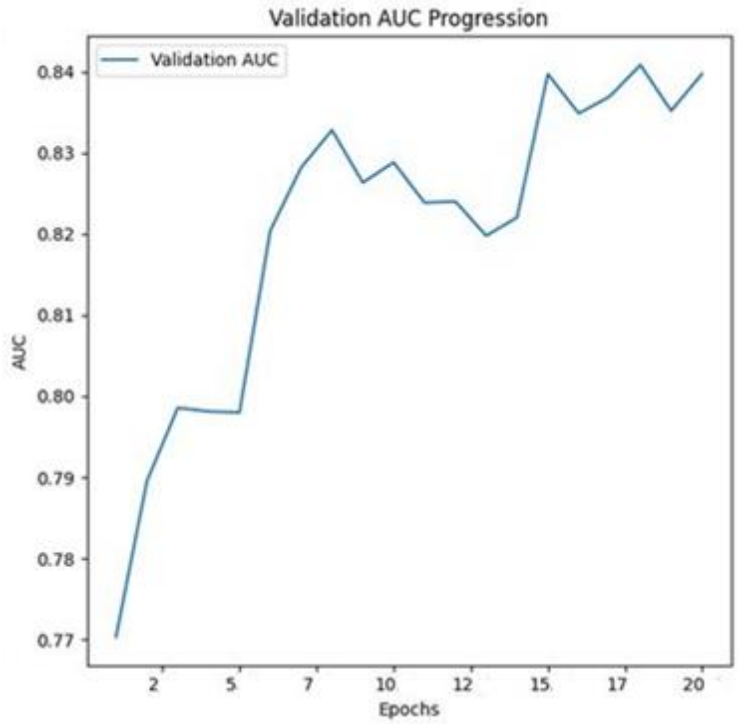


Figure 11. Trial 1 AUC

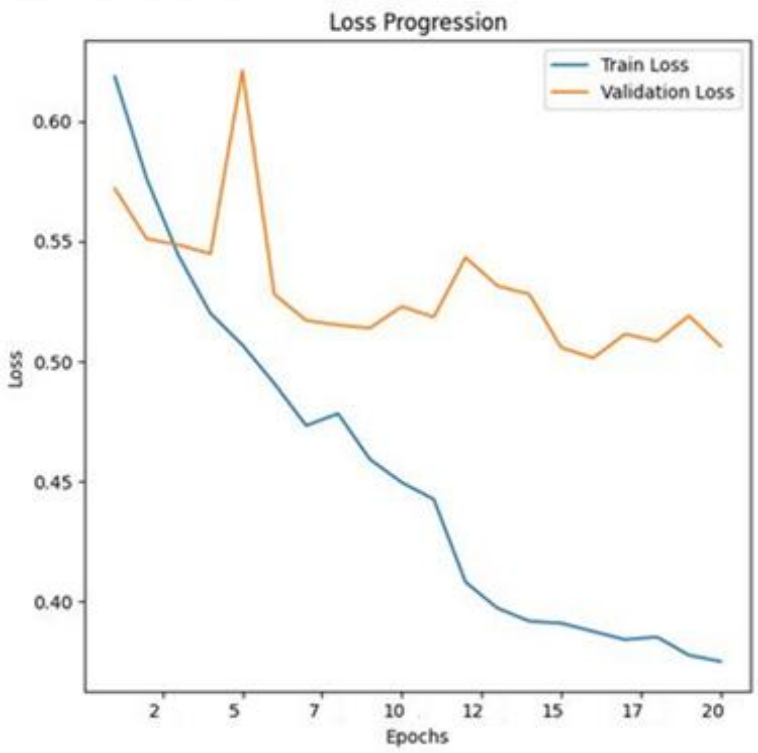


Figure 12. Trial 1 LOSS

Performance:

Best Validation AUC: 0.8408

Best F1-Score: 0.7732

Learning Behaviour:

Demonstrated consistent improvement in early epochs, with validation AUC and F1-Score increasing steadily.

Encountered performance dips in mid-training, requiring tuning adjustments to recover. The model regained momentum, achieving respectable metrics.

Observations:

Training loss reduced steadily, indicating effective optimization.

Validation loss exhibited fluctuations, suggesting challenges in generalization.

Performance plateaued in later epochs despite adjustments, indicating that the model may have reached its capacity with the given configuration.

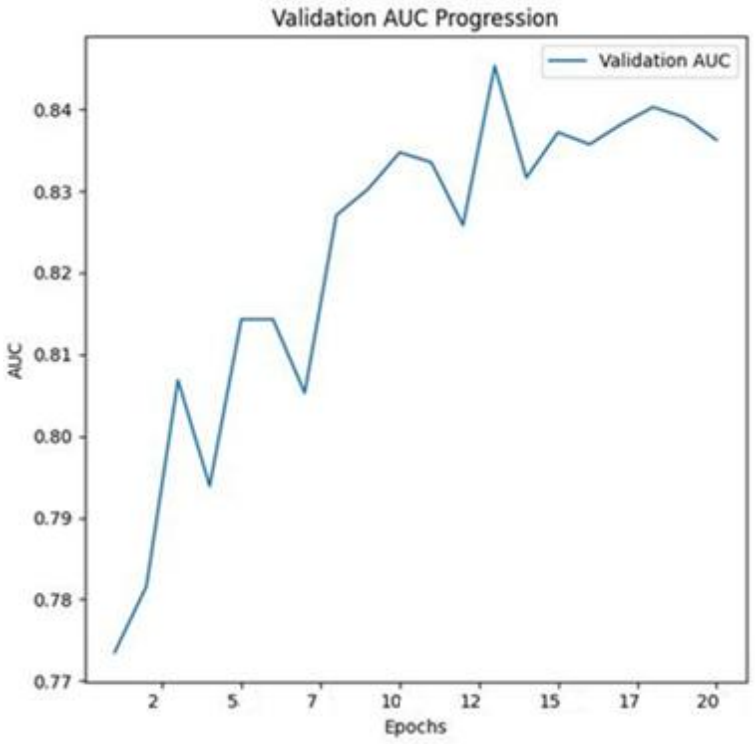


Figure 13. Trial 2 AUC

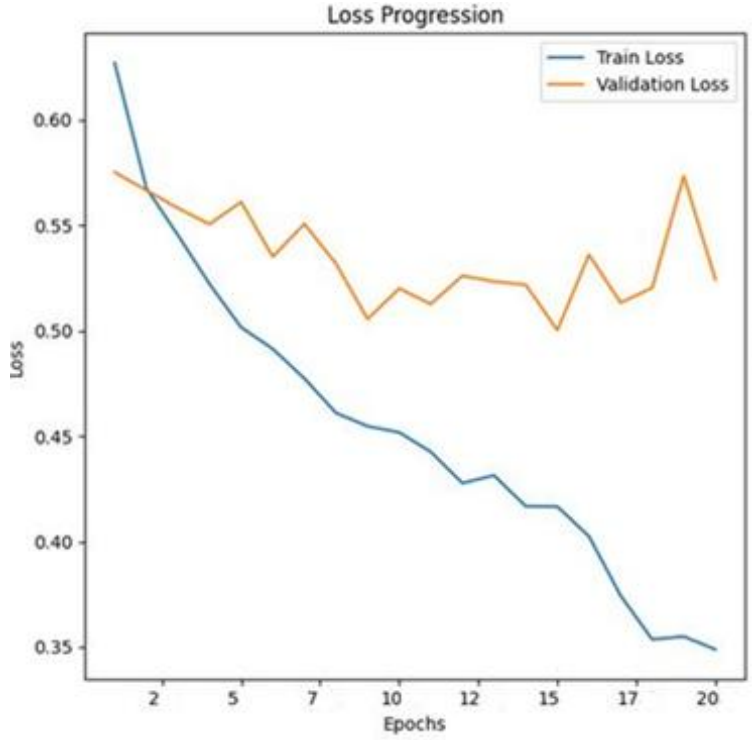


Figure 14. Trial 2 LOSS

Performance:

Best Validation AUC: 0.8453

Best F1-Score: 0.7718

Learning Behaviour:

Showed improvement in the first few epochs but exhibited more frequent performance dips compared to Trial 1.

Adjustments to the learning and dropout rates helped mitigate overfitting tendencies, but the model struggled to sustain validation performance improvements consistently.

Observations:

Validation loss increased intermittently during later epochs, highlighting potential overfitting or challenges in optimizing for unseen data.

Despite having a slightly higher AUC than Trial 1, the less stable learning dynamics made it less reliable.

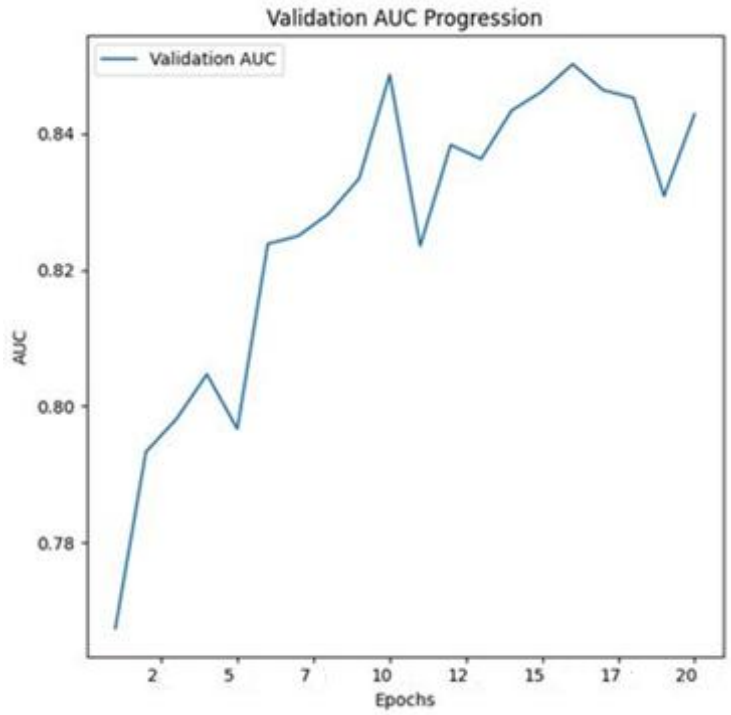


Figure 15. Trial 3 AUC

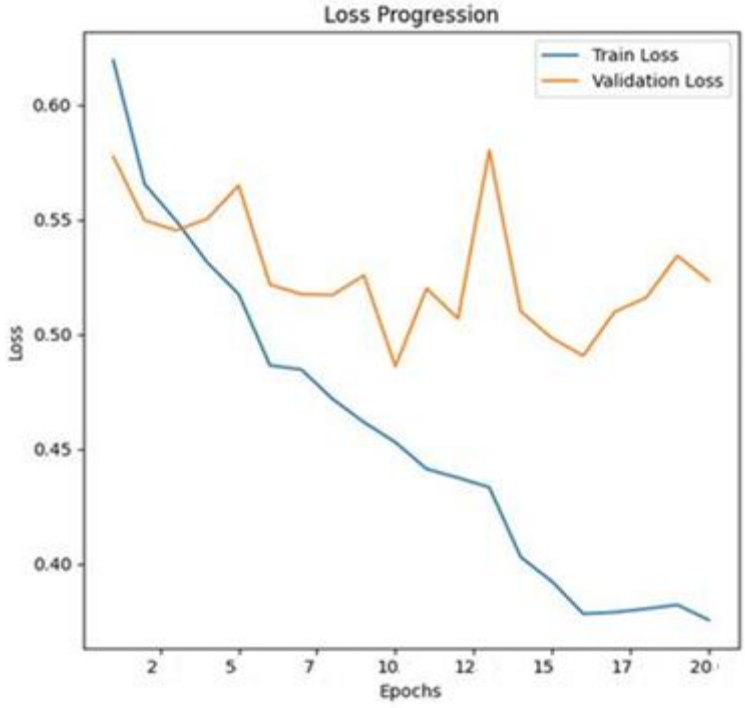


Figure 16. Trial 3 LOSS

Performance:

Best Validation AUC: 0.8502

Best F1-Score: 0.7894

Learning Behaviour:

Achieved the highest AUC among all trials, reflecting superior classification ability.

Demonstrated smoother training and validation curves with fewer sharp fluctuations, indicating better stability and generalization.

Showed consistent improvement even after tuning adjustments, indicating that the model adapted well to the dataset.

Observations:

Validation loss remained low and stable throughout the training, confirming the model's robustness, also achieved a balance between training loss reduction and validation performance, minimizing overfitting risks.

The model’s ability to generalize to unseen data makes it a strong candidate for real-world applications.

4.2.2.Discussion and selection of the best Model

Selecting the best configuration for any deep learning model requires careful analysis of its performance metrics and behaviour during training and validation. For this part of the study three trials were conducted, each with a similar architecture combining EfficientNetB3 as the feature extractor and a Linformer block for sequential modelling of multi-lead ECG data. The goal was to identify the configuration that offered the best balance of accuracy, AUC, stability, and generalization. Each trial reflected unique challenges and adjustments, making this comparative process both insightful and critical.

Table 6. Dynamic Hyperparameter Tuning Trials

Model	Hyperparameter Tuning	Characteristics	Metrics	
			AUC	F1-Score
Trial 1	Dynamic	Unfreezing base layers, and adapting dropout rates during training	0.8408	0.7732
Trial 2	Dynamic	Unfreezing base layers, and adapting dropout rates during training	0.8453	0.7718
Trial 3	Dynamic	Unfreezing base layers, and adapting dropout rates during training	0.8502	0.7894

The first trial demonstrated early promise, with validation AUC reaching a respectable 0.8408 and an F1-Score of 0.7732, one of the highest among all trials. The training process began steadily, with clear improvements in the initial epochs. However, the model faced fluctuations in validation metrics around mid-training. Adjustments to hyperparameters, such as lowering the learning rate and unfreezing additional layers, were necessary to regain performance stability. Although the model achieved competitive metrics, these fluctuations made it less reliable overall. The dips in validation performance hinted at potential challenges in balancing learning stability and generalization.

The second trial brought a slightly higher AUC of 0.8453, but the model exhibited more frequent instability. While the initial epochs showed steady progress, later phases were marked by oscillations in validation loss and metrics, which were harder to mitigate effectively. The F1-Score, although solid at 0.7718, was slightly lower than the first trial. This trial's instability raised concerns about its ability to maintain performance consistency under different conditions. Overfitting tendencies in the later epochs, as evidenced by increasing validation loss, suggested that the model was struggling to adapt to unseen data despite adjustments to the learning rate and dropout rate.

The third trial stood out with its balanced and robust performance. It achieved the highest AUC among all trials, peaking at 0.8502, and maintained stable validation curves throughout the training process. This trial was particularly notable for its smooth progression, with fewer sharp drops in metrics, which is often an indicator of better generalization. It also reached an F1-Score of 0.7894, which was slightly higher than the first two trials, so the overall stability and superior AUC and F1-score made it the most reliable configuration.

Unlike the earlier trials, this model adapted well to dynamic tuning, benefiting from adjustments to learning rate and dropout rate without significant disruption. The results suggest that Trial 3 effectively balanced complexity and generalization, making it a strong candidate for further exploration or deployment.

Trial 3 achieved a better AUC (0.8502) due to a well-balanced trade-off between dropout tuning and layer unfreezing. Unlike previous trials, where dropout was either too aggressive (causing underfitting) or too relaxed (leading to overfitting), Trial 3's 0.3 dropout rate allowed robust feature learning while preventing memorization of noise.

Dynamically adjusting hyperparameters played a crucial role in optimizing model performance. Trial 3 benefited the most, demonstrating stable validation loss and improved generalization, this adaptability is crucial for tasks like ECG classification, where the goal is not only high accuracy but also reliable generalization to diverse patient data.

4.3. Discussion

In this section, we present a comprehensive comparison of five distinct models used for ECG classification. The first model is the third trial's model developed as part of this work, which incorporates a combination of EfficientNetB3 and a Transformer block for feature extraction and classification.

The other four models are drawn from reputable scientific papers. Three of these models (Model 2, Model 3, and Model 4) are from (Narotamo et al., 2024) and represent state-of-the-art approaches in the field of ECG analysis, utilizing ResNet50, MobileNetV2, and AlexNet respectively. The fifth model (Model 5) is based on the findings from "Acute Myocardial Infarction Detection Using Deep Learning-Enabled Electrocardiograms" (Frontiers in Cardiovascular Medicine, published in 2024), which also focused on AMI detection using the PTB-XL dataset and employed a CNN backbone. By examining the performance, training strategies, and results of these models, we aim to assess the strengths and weaknesses of each, ultimately highlighting the potential benefits and limitations of our proposed model in comparison to existing solutions.

Table 7. Selected Models for the discussion

Models	Epochs	Dataset	Backbone	Results
Trial 3's Model	20 (early stopping applied)	7,086 ECGs (3,543 healthy, 3,543 myocardial infarction)	EfficientNetB3 + Linformer block	AUC (%) = 85.02 F1 (%) = 78.94
Model 2 (MI)	100 (early stopping applied)	PTB-XL Dataset 21 430 ECGs	ResNet50	Sens (%) = 68.54 Spec (%) = 73.73
Model 3 (MI)	100 (early stopping applied)	PTB-XL Dataset 21 430 ECGs	MobileNetV2	Sens (%) = 57.68 Spec (%) = 75.40
Model 4 (MI)	100 (early stopping applied)	PTB-XL Dataset 21 430 ECGs	AlexNet	Sens (%) = 64.56 Spec (%) = 74.22
Model 5 (AMI)	N/I	PTB-XL Dataset 21 430 ECGs	CNN Backbone	AUC (%) = 94.4 F1 (%) = 80.3

Both the trial's model and the other MI/AMI-detection models described in the table 7 focus exclusively on identifying myocardial infarction (or acute myocardial infarction), yet they differ notably in data utilization and architectural approach. In the trial's model, a hybrid architecture is employed that couples a pretrained EfficientNetB3 with a Transformer block. This model was trained for only 20 epochs, using early stopping, and was developed on a subset of the PTB-XL dataset containing 7,086 ECGs (evenly split between healthy and MI cases). The reported performance, with an AUC of 85.02% and an F1 score of 78.94%, demonstrates that even with a reduced portion of the dataset, the combination of an advanced, pretrained convolutional backbone and a Transformer to capture long-range dependencies can yield competitive results.

In contrast, Models 2, 3, and 4 from (Narotamo et al., 2024) utilize the entire PTB-XL dataset, which comprises 21,430 ECGs. These models were all trained for 100 epochs with early stopping and used classical CNN architectures: ResNet50, MobileNetV2, and AlexNet, respectively. The ResNet50-based model achieved a sensitivity of 68.54% and specificity of 73.73%, indicating a moderate ability to correctly identify MI cases while keeping false positives at a tolerable level. Meanwhile, the MobileNetV2 model exhibited lower sensitivity (57.68%) but higher specificity (75.40%), suggesting it is more conservative in flagging MI. The AlexNet-based model offered intermediate performance with sensitivity at 64.56% and specificity at 74.22%.

The fifth model, from (L. Wang et al., 2024), also utilized the PTB-XL dataset and employed a CNN backbone for feature extraction. It reported an impressive AUC of 94.4% and an F1-score of 80.3% on its validation set for AMI diagnosis. While the specific architectural details and training epochs for this CNN model aren't explicitly stated in the available information, its high performance metrics suggest a potentially more sophisticated or well-tuned CNN architecture compared to the ResNet50, MobileNetV2, and AlexNet models in this specific context.

While all these approaches target the same diagnostic task, the trial's model distinguishes itself in several ways. First, its hybrid design leverages the strengths of both a highly efficient pretrained CNN and a Transformer module to capture both fine-grained local features and global relationships in the ECG signal. This combination allowed the trial model to reach a respectable level of performance with significantly fewer training epochs, which may translate to lower computational requirements and faster convergence during training. However, it is

important to note that the trial model was developed using only a subset of the PTB-XL data, potentially limiting its exposure to the full spectrum of ECG variability present in the entire dataset. This reduced data size might affect its robustness or generalizability when compared to models trained on the full dataset.

On the other hand, Models 2, 3, and 4, built on the complete PTB-XL dataset, benefit from a larger, more diverse set of training samples. This advantage is reflected in their reported sensitivity and specificity metrics, which are critical for clinical applications. The high AUC and F1-score of Model 5 from (L. Wang et al., 2024) further underscore the potential of well-designed CNN architectures trained on the full PTB-XL dataset for accurate AMI detection.

In summary, while the trial's model and the other MI/AMI-specific models are all designed to detect myocardial infarction, they take different paths toward that goal. The trial's model, using a subset of the PTB-XL data and a hybrid CNN-Transformer architecture, demonstrates that a modern approach can achieve competitive performance in fewer epochs, offering potential advantages in training efficiency and computational cost. In contrast, the models that utilize the entire PTB-XL dataset (with architectures like ResNet50, MobileNetV2, AlexNet, and the CNN-Backbone in (L. Wang et al., 2024)) may provide better generalization and higher overall performance (as seen in Model 5), which is crucial in a clinical setting where missing an MI can have serious consequences.

Thus, if the application requires rapid deployment and lower resource usage with acceptable performance, the trial's hybrid model is very appealing. Conversely, if the primary concern is achieving maximum sensitivity and robustness across a broader range of patient data, the ResNet50, AlexNet-based approaches, and potentially the CNN architecture used in the fifth model might be better options, therefore, a part of our future work involves implementing the proposed hybrid model on the entire PTB-XL dataset to assess its performance and generalizability on a larger and more diverse collection of ECG recordings.

Chapter 5. Conclusion and Future work

This study highlights the transformative potential of deep learning in facilitating ECG-based cardiac diagnostics. By utilizing advanced architectures like EfficientNetB3 and the Linformer attention mechanism, we demonstrated the feasibility of classifying healthy individuals and those with myocardial infarction (MI) with high accuracy. Among the tested configurations, Trial 3 performed particularly well, achieving a validation AUC of 0.8502, an F1-Score: 0.7894, showcasing stable performance throughout training and validation reaching a training loss of ~ 0.3 and a validation loss < 0.5 . This outcome underscores the power of combining innovative models with efficient training strategies, such as dynamic hyperparameter tuning and layer unfreezing.

This model illustrates how AI can significantly contribute to the healthcare sector, particularly in remote or under-resourced settings. Automated ECG analysis can provide timely, accessible, and accurate diagnostics, potentially assisting clinicians or even operating autonomously in critical situations. However, during the realisation of this model we encountered some challenges. Limited computational resources and a reduced dataset constrained the scope of the study, focusing the analysis on binary classification instead of exploring multi-class scenarios. Despite these constraints, the results lay a solid foundation for future advancements.

Looking forward, several directions for improvement and expansion are possible. Firstly, to rigorously evaluate the model's generalizability and robustness, a key next step involves implementing and training the proposed hybrid architecture on the entire PTB-XL dataset, this will allow for a comprehensive assessment of its performance across a wider spectrum of ECG morphologies and patient demographics, potentially leading to improved sensitivity and overall diagnostic accuracy. Secondly, recognizing the critical importance of trust and interpretability in clinical applications, future work will focus on enhancing the explainability of the model's predictions, this will involve exploring techniques such as attention visualization from the Transformer block, gradient-based saliency maps applied to the EfficientNet features, or the integration of post-hoc explainability methods such as LIME and SHAP (Ningthoujam et al., 2025).

The goal is to provide clinicians with insights into the model's decision-making process, highlighting the specific ECG features that contribute most significantly to the classification of myocardial infarction. This increased transparency is crucial for fostering clinician understanding and ultimately facilitating the adoption of AI-powered diagnostic tools in real-world clinical practice. Further research will also explore the optimization of the model architecture and hyperparameters.

One other notable future improvement would be the incorporation of multi-class classification, enabling the model to detect a broader range of cardiac conditions, such as arrhythmias, ischemia, and ventricular hypertrophy. Expanding the model in this way could significantly enhance its clinical utility, transforming it from a specialized tool into a comprehensive diagnostic system capable of screening for multiple cardiac abnormalities simultaneously. While transitioning to a multi-class framework presents challenges such as the need for a larger, well-annotated dataset, and increased computational demands. Techniques like transfer learning could be employed to improve training efficiency. Transfer learning allows the model to initially learn simpler tasks before progressively adapting to more complex scenarios, thus mitigating the data and computational constraints.

The feasibility of multi-class classification has been supported by prior research. For example, (Hannun et al., 2019) demonstrated the capability of deep learning models to identify multiple arrhythmias from ECG data, while (Rajpurkar et al., 2017) achieved cardiologist-level performance in detecting a dozen arrhythmia types using a deep neural network. These studies show that multi-class ECG models are achievable and can perform comparably to human experts. Building on such findings, our model could be adapted to identify not only MI but also a variety of other cardiac pathologies, addressing a wider range of diagnostic needs.

The practical applications of such a model align closely with recent advances in AI-driven healthcare technologies. For instance, (Attia et al., 2019) trained a model to detect atrial fibrillation and other abnormalities from standard ECGs, showing that deep learning can facilitate early detection and treatment for a wider patient population. Following this approach, our future model could serve as an asset for healthcare providers by offering more inclusive diagnostics. Particularly in rural or under-resourced areas, a robust, multi-class ECG model could act as a frontline diagnostic tool, flagging potential issues for further review and enabling timely intervention for a variety of cardiac conditions.

As the field of artificial intelligence continues to advance, there is significant potential for integrating newer deep learning techniques into clinical practice. Recent advancements in self-supervised learning techniques, such as TransEHR, highlight the potential for improved analysis of clinical time series data (Xu et al., 2023).These techniques can enhance the model's ability to learn from unlabelled data, which is particularly valuable in medical settings where annotated datasets may be limited.

In conclusion, while this study has achieved good results using deep learning Model for Myocardial Infarction diagnostics, the potential for future work is vast. Expanding the scope to handle multi-class classification, improving model efficiency, also external validation on independent datasets will be essential to confirm the model's performance and clinical utility, and addressing computational constraints and enhancing the explainability of the model's predictions are key objectives that will shape the next steps in AI-driven cardiac care. With continued research and advancements in machine learning, Models like these could play a crucial role in improving patient outcomes, enhancing healthcare accessibility, and saving lives.

References

Acharya, U. R., Hagiwara, Y., Koh, J. E. W., Oh, S. L., Tan, J. H., Adam, M., & Tan, R. S. (2018). Entropies for automated detection of coronary artery disease using ECG signals: A review. *Biocybernetics and Biomedical Engineering*, *38*(2), 373–384. https://doi.org/10.1016/j.bbe.2018.03.001

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. https://doi.org/10.1145/3292500.3330701

Al-Kababji, A., Bensaali, F., & Dakua, S. P. (2022). *Scheduling Techniques for Liver Segmentation: ReduceLROnPlateau Vs OneCycleLR*.

Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., Carter, R. E., Yao, X., Rabinstein, A. A., Erickson, B. J., Kapa, S., & Friedman, P. A. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet*, *394*(10201), 861–867. https://doi.org/10.1016/S0140-6736(19)31721-0

Borghi, P. H., Borges, R. C., & Teixeira, J. P. (2021). Atrial fibrillation classification based on MLP networks by extracting Jitter and Shimmer parameters. *Procedia Computer Science*, *181*, 931–939. https://doi.org/10.1016/j.procs.2021.01.249

Costa, R., Winkert, T., Manhães, A., & Teixeira, J. P. (2021). QRS Peaks, P and T Waves Identification in ECG. *Procedia Computer Science*, *181*, 957–964. https://doi.org/10.1016/j.procs.2021.01.252

Faust, O., Shenfield, A., Kareem, M., San, T. R., Fujita, H., & Acharya, U. R. (2018). Automated detection of atrial fibrillation using long short-term memory network with RR interval signals. *Computers in Biology and Medicine*, *102*, 327–335. https://doi.org/10.1016/j.compbiomed.2018.07.001

Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, *25*(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3

Jia, J., Liang, W., & Liang, Y. (2023). *A Review of Hybrid and Ensemble in Deep Learning for Natural Language Processing. arXiv:2312.05589v1*

Khan, A. H., Hussain, M., & Malik, M. K. (2021). Cardiac Disorder Classification by Electrocardiogram Sensing Using Deep Neural Network. *Complexity*, *2021*(1). https://doi.org/10.1155/2021/5512243

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. *arXiv:1412.6980*

Cao, M., Tran, D., Duong, T. H., Le, H., Pham, T. T., & Nguyen, V. H. (2023). ECG Heartbeat classification using deep transfer learning with Convolutional Neural Network and STFT technique. *Journal of Physics: Conference Series, 2547*(1), 012031. https://doi.org/10.1088/1742-6596/2547/1/012031

Narotamo, H., Dias, M., Santos, R., Carreiro, A. V., Gamboa, H., & Silveira, M. (2024). Deep learning for ECG classification: A comparative study of 1D and 2D representations and multimodal fusion approaches. *Biomedical Signal Processing and Control*, *93*, 106141. https://doi.org/10.1016/j.bspc.2024.106141

Nguyen, T.-N., Nguyen, M.-H., Tran, H., & Nguyen, T.-H. (2020). *A Deep Learning Framework for Inter-Patient ECG Classification*. *19, 74–84. Doi: 10.12928/telkomnika.v18i2.14403*

Ningthoujam, A. S., Sharma, S., & Nandi, A. (2025). *Explainable AI Based Coronary Heart Disease Prediction: Enhancing Model Transparency in Clinical Decision Making*. https://doi.org/10.1101/2025.03.20.644456

O'Gara, P. T., Kushner, F. G., Ascheim, D. D., Casey, D. E., Chung, M. K., de Lemos, J. A., Ettinger, S. M., Fang, J. C., Fesmire, F. M., Franklin, B. A., Granger, C. B., Krumholz, H. M., Linderbaum, J. A., Morrow, D. A., Newby, L. K., Ornato, J. P., Ou, N., Radford, M. J., Tamis-Holland, J. E., … Zhao, D. X. (2013). 2013 ACCF/AHA Guideline for the Management of ST-Elevation Myocardial Infarction. *Journal of the American College of Cardiology*, *61*(4), e78–e140. https://doi.org/10.1016/j.jacc.2012.11.019

Patel, S. J., Yousuf, S., Padala, J. V, Reddy, S., Saraf, P., Nooh, A., Fernandez Gutierrez, L. M. A., Abdirahman, A. H., Tanveer, R., & Rai, M. (2024). Advancements in Artificial Intelligence for Precision Diagnosis and Treatment of Myocardial Infarction: A Comprehensive Review of Clinical Trials and Randomized Controlled Trials. *Cureus*. https://doi.org/10.7759/cureus.60119

Prochnau, D., Glück, G., Holz, I., Lascho, C., & Eggers, R. (2019). Painless myocardial infarction - a clinical challenge in the emergency department. *CJEM*, *21*(4), 557–559. https://doi.org/10.1017/cem.2019.11

Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). *Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks*. *arXiv:1707.01836*

Ribeiro, A. H., Ribeiro, M. H., Paixão, G. M. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., Ferreira, M. P. S., Andersson, C. R., Macfarlane, P. W., Meira, W., Schön, T. B., & Ribeiro, A. L. P. (2020). Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature Communications*, *11*(1), 1760. https://doi.org/10.1038/s41467-020-15432-4

Rodrigues, Miguel, P., Teixeira, & Paulo, J. (2010). Classification of Electroencephalogram signals using Artificial Neural Networks. *2010 3rd International Conference on Biomedical Engineering and Informatics*, 808–812. https://doi.org/10.1109/BMEI.2010.5639941

Smith, S. W., Dodd, K. W., Henry, T. D., Dvorak, D. M., & Pearce, L. A. (2012). Diagnosis of ST-Elevation Myocardial Infarction in the Presence of Left Bundle Branch Block With the ST-Elevation to S-Wave Ratio in a Modified Sgarbossa Rule. *Annals of Emergency Medicine*, *60*(6), 766–776. https://doi.org/10.1016/j.annemergmed.2012.07.119

Tan, M., & Le, Q. V. (2019). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. *arXiv:1905.11946*

Thygesen, K., Alpert, J. S., Jaffe, A. S., Chaitman, B. R., Bax, J. J., Morrow, D. A., & White, H. D. (2018). Fourth Universal Definition of Myocardial Infarction (2018). *Journal of the American College of Cardiology*, *72*(18), 2231–2264. https://doi.org/10.1016/j.jacc.2018.08.1038

Tison, G. H., Sanchez, J. M., Ballinger, B., Singh, A., Olgin, J. E., Pletcher, M. J., Vittinghoff, E., Lee, E. S., Fan, S. M., Gladstone, R. A., Mikell, C., Sohoni, N., Hsieh, J., & Marcus, G. M. (2018). Passive Detection of Atrial Fibrillation Using a Commercially Available Smartwatch. *JAMA Cardiology*, *3*(5), 409–416. https://doi.org/10.1001/jamacardio.2018.0136

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. *arXiv:1706.03762*

Wagner, P., Strodthoff, N., Bousseljot, R.-D., Kreiseler, D., Lunze, F. I., Samek, W., & Schaeffter, T. (2020). PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, *7*(1), 154. https://doi.org/10.1038/s41597-020-0495-6

Wang, L., Fan, Y., Xiao, C., & Ouyang, W. (2024). Advancements and Applications of Convolutional Neural Network Models in Cardiovascular Disease: A Comprehensive Review. *Journal of Clinical and Nursing Research*, *8*(10), 167–174. https://doi.org/10.26689/jcnr.v8i10.7827

Wang, S., Li, B. Z., Khabsa, M., Fang, H., & Ma, H. (2020). *Linformer: Self-Attention with Linear Complexity*. *arXiv:2006.04768.*

Wikipedia contributors. (2024). *Management of acute coronary syndrome — Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/w/index.php?title=Management_of_acute_coronary_syndrome&oldid=1261932756

World Health Organization. (2021). Cardiovascular diseases (CVDs) fact sheet. *World Health Organization(2021). [Online]. https://aho.org/fact-sheets/cardiovascular-diseases-cvds-fact-sheet/*

Xia, Y., Wulan, N., Wang, K., & Zhang, H. (2018). Detecting atrial fibrillation by deep convolutional neural networks. *Computers in Biology and Medicine*, *93*, 84–92. https://doi.org/10.1016/j.compbiomed.2017.12.007

Xu, Y., Xu, S., Ramprassad, M., Tumanov, A., & Zhang, C. (2023). TransEHR: Self-Supervised Transformer for Clinical Time Series Data. In *Proceedings of Machine Learning Research PMLR 225:623-635 [Online]. https://proceedings.mlr.press/v225/xu23a/xu23a.pdf*

Paper Publication and Presentation certificates

ICDTHT´25

February 13 - 15
Salinas, Ecuador

International Conference on Demographic Transition, Health and Technologies

Acceptance Letter

Dear **Mohamed Khalil Chaabani**,

CeDRI, SusTEC, Instituto Politécnico de Bragança,

from **Portugal**

On behalf of the **ICDTHT´25 - International Conference on Demographic Transition, Health and Technologies**, I am pleased to inform you that your submission "**Article 164: Binary Classification of Cardiac Pathologies using Deep Learning: A PTB-XL Dataset Approach**" has been accepted as a Full Paper for publication and oral presentation in this conference. So, you are cordially invited to participate and present the paper in our conference to be held in **UPSE - Universidad Estatal Península de Santa Elena, in Salinas, Ecuador**, between **February 13 to 15, 2025**. We sincerely hope that you will join us in making a success. We look forward to seeing you there.

UPSE - Universidad Estatal Península de Santa Elena, in Salinas, Ecuador, February 13, 2025

(Pedro Miguel Afonso Gaspar)

Organization

International Association for Science and Research

UPSE
UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA

Certificate

Mohamed Khalil Chaabani from **Portugal**, **CeDRI, SusTEC, Instituto Politécnico de Bragança**, participated and presented the paper **Article 164: Binary Classification of Cardiac Pathologies using Deep Learning: A PTB-XL Dataset Approach** at the **ICDTHT´25 International Conference on Demographic Transition, Health and Technologies**, held between **February 13 to 15, 2025** in **UPSE - Universidad Estatal Península de Santa Elena, in Salinas, Ecuador**.

UPSE - Universidad Estatal Península de Santa Elena, in Salinas, Ecuador, February 13, 2025

(Pedro Miguel Afonso Gaspar)

Organization

International Association for Science and Research

UPSE
UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA