JPPIPA 11(3) (2025)

Jurnal Penelitian Pendidikan IPA

Journal of Research in Science Education

http://jppipa.unram.ac.id/index.php/jppipa/index

Development of an Efficient 1D-CNN Model for Myocardial Infarction Classification Using 12-Lead ECG Signals

Ahmad Haidar Mirza[1*], R.M. Nasrul Halim[1], Muhammad Jidan Hasbiallah[2]

[1]Teknik Informatika, Fakultas Sains Teknologi, Universitas Bina Darma, Palembang, Indonesia.
[2]Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya, Palembang, Indonesia.

Received: December 31, 2024
Revised: February 10, 2025
Accepted: March 25, 2025
Published: March 31, 2025

Corresponding Author:
Ahmad Haidar Mirza
haidarmirza@binadarma.ac.id

DOI: 10.29303/jppipa.v11i3.10238

Abstract: Myocardial Infarction (MI) is a leading cause of global mortality, necessitating efficient diagnostic methods. This study develops a simplified one-dimensional Convolutional Neural Network (1D-CNN) model for classifying MI using 12-lead ECG signals from the PTB-XL dataset. The research focuses on reducing computational complexity by limiting convolutional layers while maintaining high accuracy. The proposed model processes ECG signals of varying lengths (600–1000 samples), identifying 700 samples as optimal, achieving an average accuracy of 96.18%, sensitivity of 82.84%, specificity of 97.63%, precision of 84.13%, and an F1-score of 82.68%. Leads V5 and V6 demonstrate superior performance in detecting MI, while other leads, such as I and AVL, require further optimization. By combining precise signal segmentation and an efficient CNN architecture, this model minimizes computational load without compromising performance, making it a strong candidate for real-time clinical applications. The findings highlight the importance of signal length optimization and simplified architecture in enhancing early MI detection.

Keywords: 12 lead; CNN; ECG; Myocardial infarction; PTB-XL

Introduction

Myocardial Infarction (MI) remains a leading cause of mortality worldwide, necessitating timely and accurate diagnostic solutions to reduce its global burden. Myocardial Infarction (MI), also known as a heart attack, is the most dangerous type of coronary heart disease, with the highest mortality rate (Darmawahyuni & Nurmaini, 2019). MI occurs when the oxygen supply to the heart muscle is insufficient, causing tissue damage and increasing the risk of death. Early detection through ECG signal observation is crucial to prevent death from heart attacks. The main clinical symptoms of MI are heart failure, angina pectoris, and arrhythmia (Strong et al., 2017). According to the World Health Organization (WHO), Ischemic Heart Disease (IHD), including MI, accounted for 16% of global deaths in 2019, totaling approximately 8.9 million fatalities (Xiong et al., 2022). In Indonesia, the 2018 Basic Health Research (Riskesdas) reported that 2.78 million individuals, or 15 per 1,000 people, suffer from heart and cardiovascular diseases, emphasizing the urgent need for effective diagnostic tools (National Institute of Health Research and Development, 2018). Early detection of MI is vital to enabling timely interventions and improving patient outcomes, especially in regions with limited healthcare resources. To detect heart disease, experts can use Electrocardiogram (ECG) signals (Hainaut & Gade, 2003).

Electrocardiograms (ECGs) are the standard diagnostic tool for detecting heart abnormalities, providing crucial information through characteristic waveforms such as P, QRS, and T waves. However, manual interpretation of ECG signals can be time-intensive and prone to errors, particularly in high-pressure clinical environments (Ricardo et al., 2009). Specifically, ECG records the heart's electrical activity from multiple perspectives, allowing it to differentiate various types of MI based on the infarction location in the myocardium (Meek & Morris, 2002). ECG signals are

How to Cite:
Mirza, A. H., Halim, R. N., & Hasbiallah, M. J. (2025). Development of an Efficient 1D-CNN Model for Myocardial Infarction Classification Using 12-Lead ECG Signals. *Jurnal Penelitian Pendidikan IPA*, *11*(3), 1167–1182. https://doi.org/10.29303/jppipa.v11i3.10238

recorded through electrodes placed on the patient's body and have specific wave characteristics, such as P, QRS, and T waves, which cardiologists use to assess heart conditions (Nurmaini et al., 2019). In patients experiencing MI, changes in ECG signals can be observed in the ST interval length, ST elevation, and T wave morphology alterations (Zimetbaum & Josephson, 2003). In contrast, in normal patients, the five main waves (P, QRS, and T) do not exhibit abnormal morphology (Partan, 2018). However, ECG signal observation can be challenging due to variations in morphology among different patients, influenced by differing physical conditions (Banerjee & Mitra, 2013).

Recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have demonstrated significant potential for automating ECG signal analysis. These models can detect MI with high accuracy, yet many existing approaches are computationally intensive and unsuitable for real-time applications in resource-constrained settings (Mezgec & Seljak, 2017; Rawi et al., 2022). Traditionally, 12-lead ECG is commonly used to detect MI, but the addition of posterior leads in the 15-lead ECG has been shown to increase sensitivity in detecting MI (Brady et al., 2000; Sokhanvar, 2006). Several deep learning techniques used in ECG signal processing for detecting Myocardial Infarction (MI) include Convolutional Neural Networks (CNN) (Baloglu et al., 2019), Stacked Autoencoders (Vincent et al., 2008; Nurmaini et al., 2020a; Zhang et al., 2019), Deep Belief Networks (DBN) (Huanhuan & Yue, 2014; Chakraborty et al., 2022), Deep Boltzmann Machines (DBM) (Mathews et al., 2018), Recurrent Neural Networks (RNN) (Übeyli, 2010; Hasbullah et al., 2023), and Long Short-Term Memory (LSTM) (Übeyli, 2010; Muraki et al., 2022; Mirza et al., 2023). Although CNNs demonstrate advantages in feature recognition, the challenge of time-consuming computations remains a concern. Some studies have shown that reducing the number of convolutional and fully connected layers can speed up computation time without sacrificing accuracy (Mezgec & Seljak, 2017; Wang et al., 2019)

This study addresses these challenges by developing a simplified 1D-CNN model tailored for MI classification using 12-lead ECG signals. Unlike previous works that prioritize complexity over practicality, the proposed model reduces convolutional layers to minimize computational load while maintaining diagnostic performance. Using the PTB-XL dataset—a publicly available, large-scale ECG database—this research evaluates the model's effectiveness across varying signal lengths and lead configurations (Wagner et al., 2020). The novelty of this study lies in optimizing signal processing and CNN architecture to achieve a balance between computational efficiency and diagnostic accuracy, making it feasible for integration into portable and real-time diagnostic systems.

The significance of this research is multifaceted. First, it addresses the critical need for early and accurate MI detection to save lives and reduce long-term complications. Second, the model's computational efficiency ensures practicality in diverse clinical and emergency settings. Third, its scalability enables implementation in portable devices, expanding access to advanced diagnostics in underserved regions. Finally, by demonstrating the potential of AI-driven ECG analysis, this study contributes to advancing artificial intelligence applications in healthcare, fostering innovation in cardiology and beyond (Baloglu et al., 2019; Nurmaini et al., 2020b). In summary, this study addresses the critical need for efficient and accurate MI classification using standard 12-lead ECGs, combining innovative model design with practical clinical implications.

Method

The research process is divided into 4 main steps: data collection, data preparation which consists of signal normalization, noise removal, segmentation, development of a 12-lead classifier model framework using 1D-CNN and model performance evaluation.

Data Collection

The ECG data utilized in this study is sourced from the publicly available, non-commercial PTB-XL database (Wagner et al., 2020). This dataset comprises 21,799 12-lead ECG clinical recordings, each lasting 10 seconds, collected from 18,869 patients. Among these, 5,469 recordings are classified as Myocardial Infarction (MI), while 9,514 are classified as normal. The data was recorded using a standard 12-lead setup (I, II, III, AVR, AVL, AVF, V1, V2, V3, V4, V5, V6) and includes 10 ECG class categories: anterior MI (AMI), anterolateral MI (ALMI), anteroseptal MI (ASMI), inferior MI (IMI), inferolateral MI (ILMI), inferior posterior MI (IPMI), infero-posterolateral MI (IPLMI), lateral MI (LMI), posterior MI (PMI), and the normal/healthy class. The raw signal data is stored in a specialized compressed format, maintaining high fidelity for clinical and research applications.

Data Preparation

The next step is the data preparation process, which includes normalization, noise removal, and data segmentation.

Normalization

The ECG dataset requires normalization because the features in the data have varying value ranges

(Figure 2), which could impact model performance. Normalization also reduces computation time by standardizing the data. In this study, Min-Max scaling was employed to normalize the amplitude of the ECG signals. This method scales the signal values to a specified range, specifically between 0 (lower bound, *lb*) and 1 (upper bound, *ub*), ensuring consistency across the dataset. The normalization process only affects the amplitude scale within this range without altering the signal's morphology. To implement this, the normalize_bound (signal, *lb* = 0, *ub* = 1) function from the Wafeworm Database (wfdb) library was utilized, which rescales the signal values using the formula:

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \times (ub - lb) + lb \quad (1)$$

where x represents the original signal values, and x_{min} and x_{max} are the minimum and maximum values of the signal, respectively. This ensures that all signals are scaled within the defined bounds while preserving their morphological integrity.

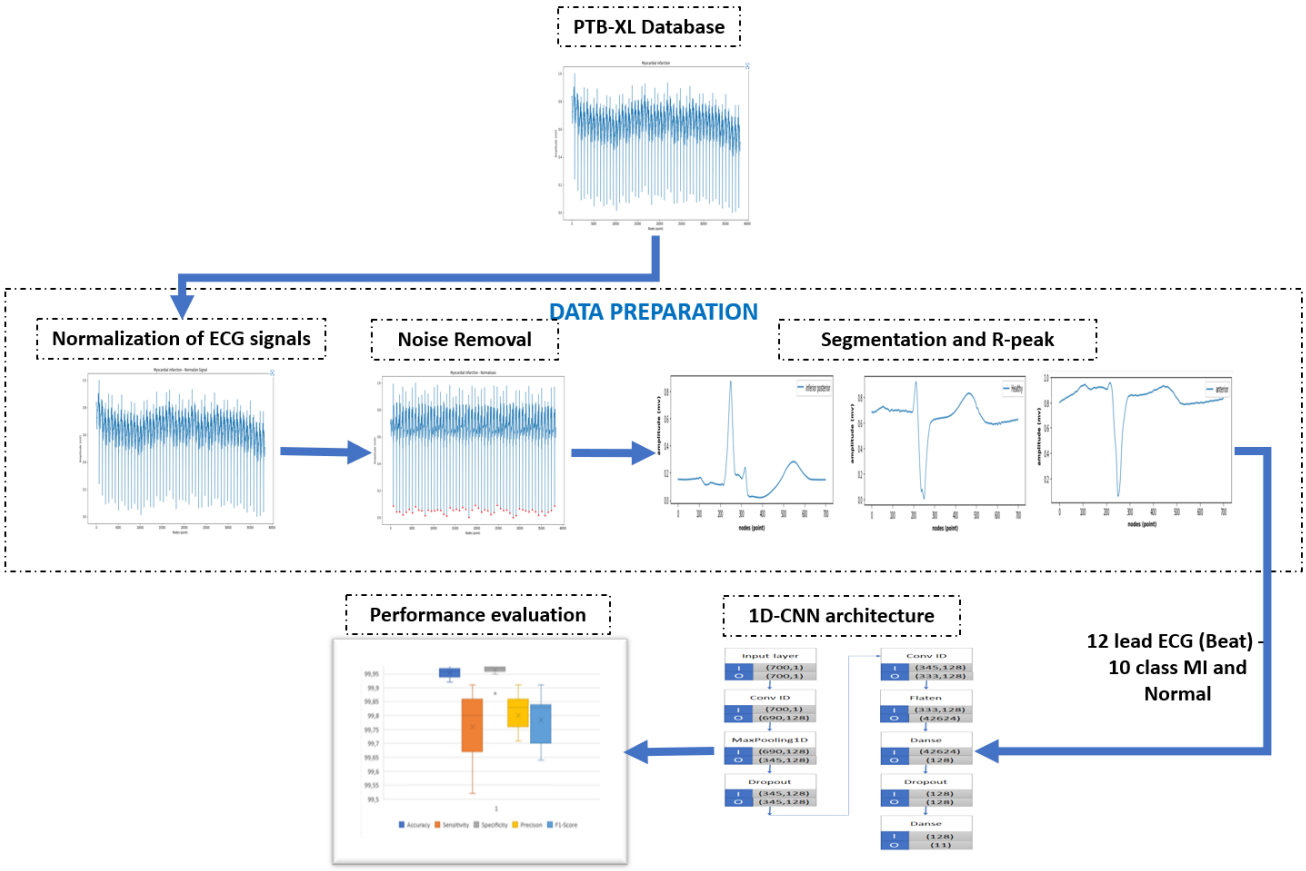

Figure 1. Research Methods

Noise Removal

ECG signals are often disrupted by various types of artifacts and interference from electrical channels (Wang et al., 2015). After normalization, the ECG signals require denoising to improve their quality and ensure accurate analysis. Noise and artifact removal is, therefore, a crucial step in processing raw ECG signals. In this study, ECG signal reconstruction from noise-contaminated signals was performed using the Discrete Wavelet Transform (DWT) method, which is widely used for analysis, noise removal, and signal and image compression.

The denoising process involved decomposing the signal into multiple levels of wavelet coefficients. Specifically, the signal was decomposed into five levels using the Symlet 5 (Sym5) wavelet, which is selected for its superior performance in preserving ECG morphology during denoising (Nurmaini et al., 2020b). At each decomposition level, the wavelet coefficients were subjected to soft thresholding to remove noise components. The threshold was calculated using the universal threshold formula: $\lambda=\sigma \sqrt{2\log N}$ where σ represents the noise standard deviation estimated from the detail coefficients at the first level of decomposition, and N is the length of the signal.

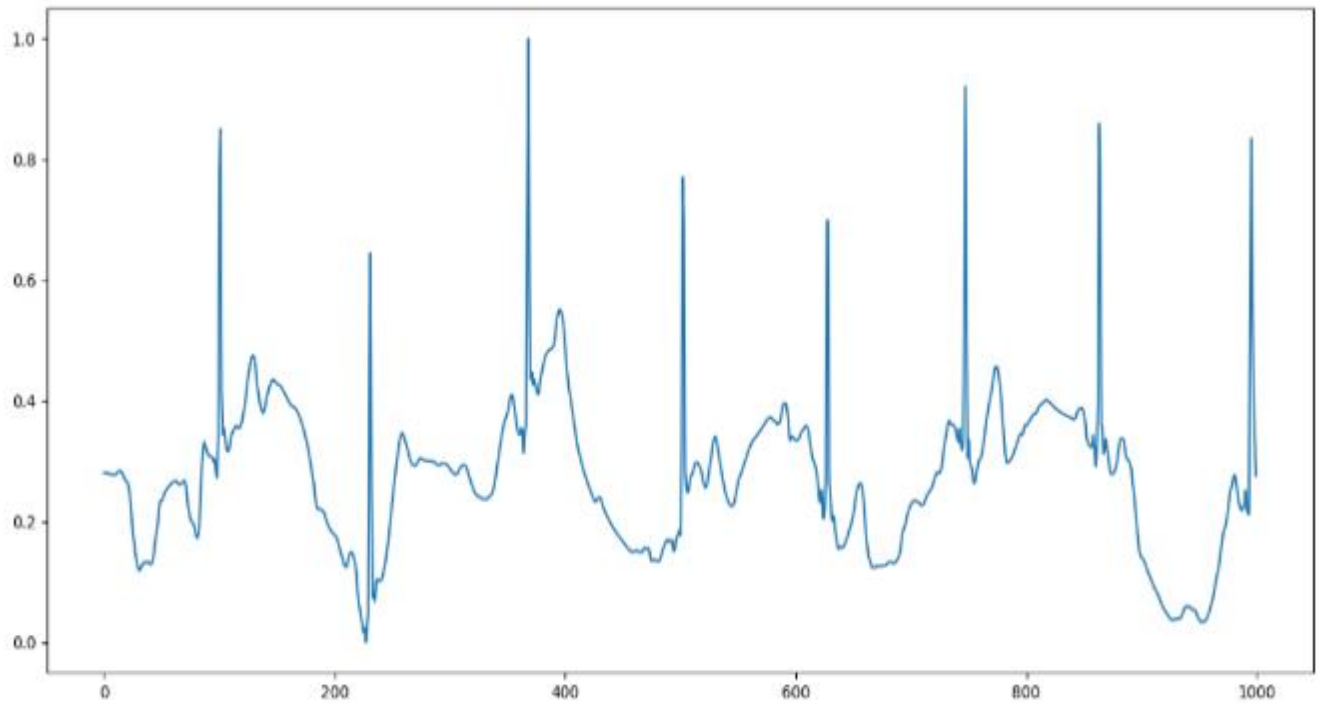

Figure 2. A sample of the PTB-XL dataset ECG signal after normalization

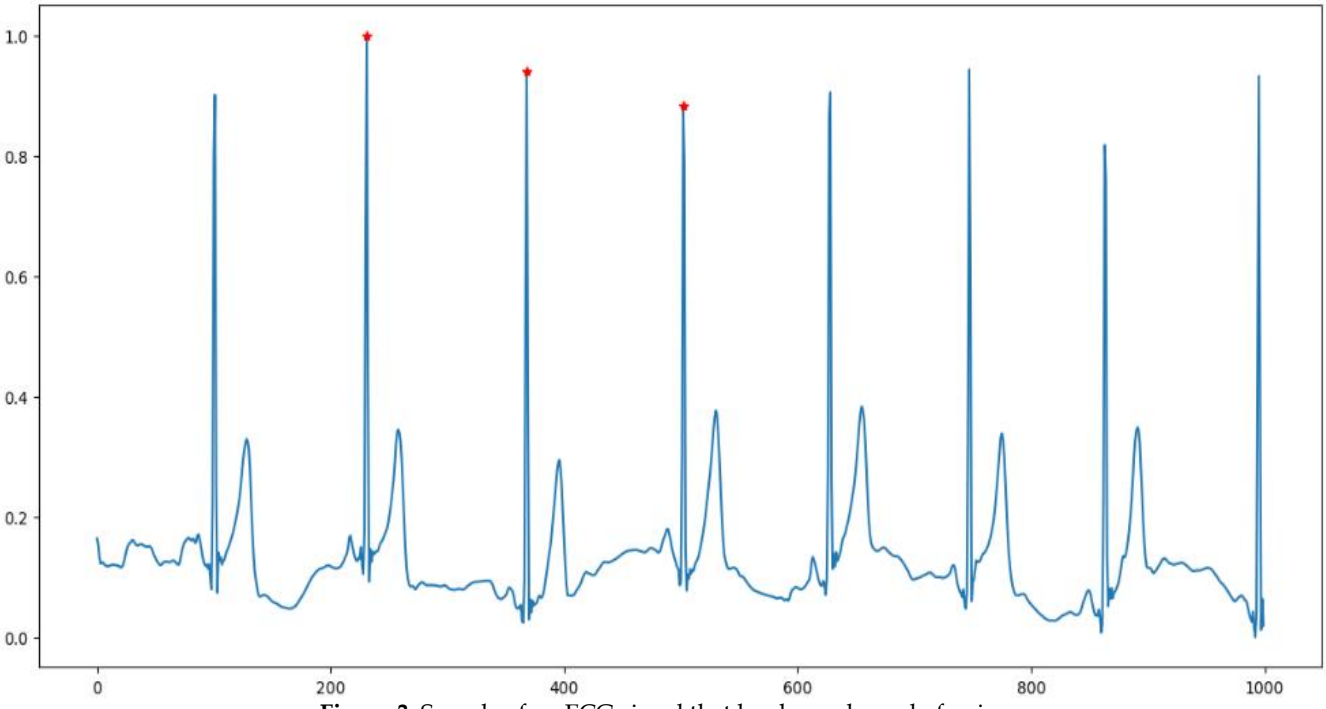

Figure 3. Sample of an ECG signal that has been cleaned of noise

Coefficients below the threshold were set to zero, while those above the threshold were shrunk to reduce noise while retaining essential signal information. After thresholding, the denoised signal was reconstructed by applying the inverse wavelet transform, yielding a cleaner signal with baseline wander corrected and aligned closer to zero, as illustrated in Figure 3. This approach effectively preserves the morphological characteristics of the ECG signal while minimizing noise.

Signal Segmentation

After the denoising stage, the ECG signal undergoes segmentation to identify the R-peak, the highest point in

the QRS complex within each ECG cycle. Accurate detection of the R-peak is essential for segmenting the ECG signal into meaningful intervals for analysis. In this study, the Pan-Tompkins algorithm was utilized to detect R-peaks with high precision, as it effectively identifies QRS complexes by analyzing slope, amplitude, and duration (Fariha et al., 2020). The sample of ECG signal segmentation for 9 MI classes and the normal class in the PTB-XL dataset is illustrated in Figure 4.

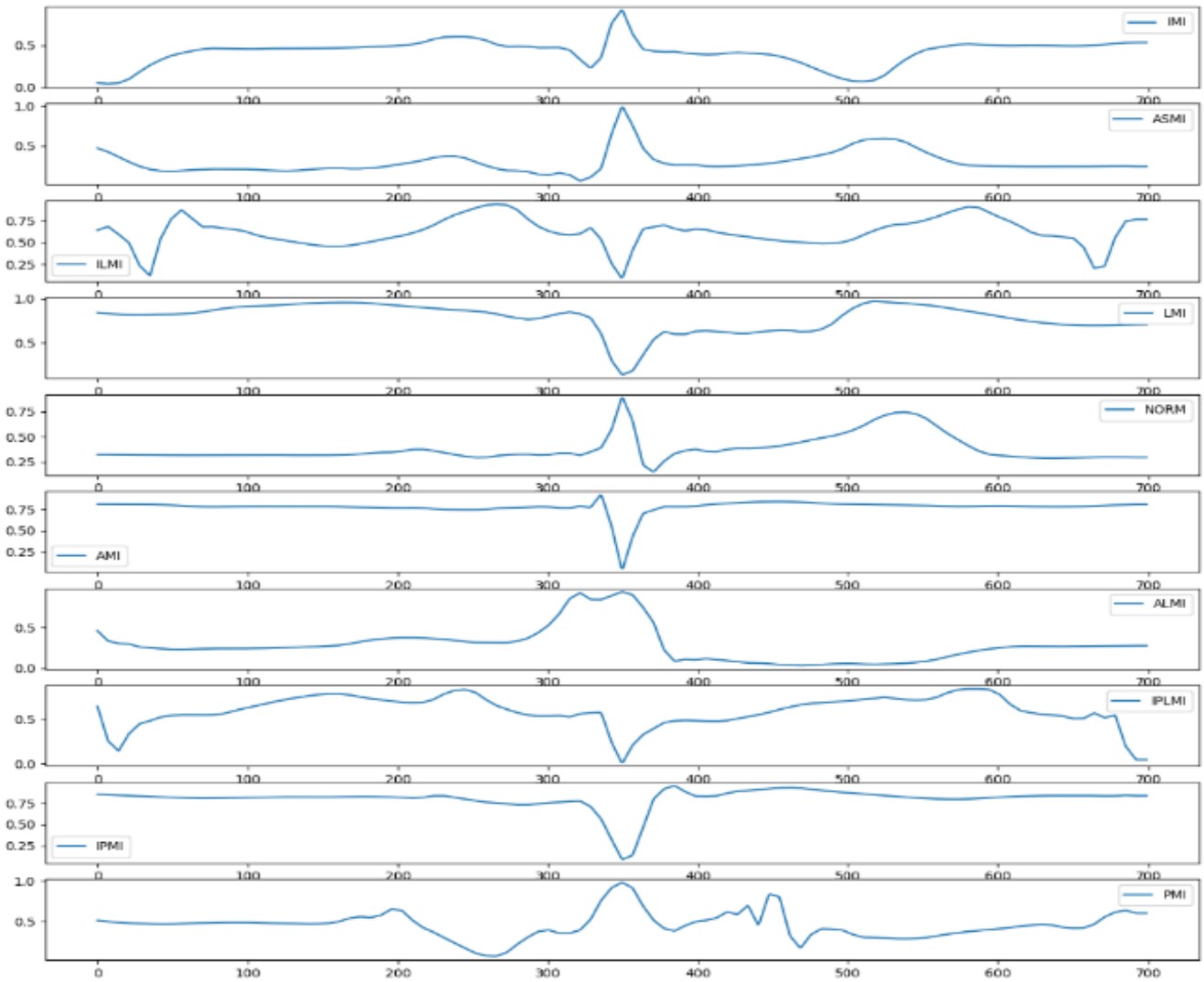

Figure 4. Sample 9 MI class and normal class in the 12-lead ECG signals of the PTB-XL dataset

Once the R-peaks were detected, the signal was segmented into specific durations relative to each R-peak to create uniform segments for classification purposes. Each segment includes a portion of the signal before and after the R-peak, defined as t_1 and t_2, respectively. The following durations were considered: t_1 = 0.20 s, t_2 = 0.40 s: Total signal length of 0.60 s, comprising 600 samples (200 samples for t_1, 400 samples for t_2). This duration was selected to focus on the QRS complex and the immediately surrounding regions; t_1 = 0.25, t_2 = 0.40 s: Total signal length of 0.65, comprising 650 samples (250 samples for t_1, 400 samples for t_2). This captures a slightly longer pre-R-peak region to account for potential variations in P waves; t_1 = 0.25 s, t_2 = 0.45 s: Total signal length of 0.70 s, comprising 700 samples (250 samples for t_1, 450 samples for t_2) (Mirza et al., 2022). The additional post-R-peak duration ensures complete coverage of the T wave for comprehensive analysis; t_1 = 0.30 s, t_2 = 0.50 s: Total signal length of 0.80 s, comprising 800 samples (300 samples for t_1, 500 samples for t_2). This longer segment is useful for capturing subtle variations in the ST segment and T wave; t_1 = 0.35 s, t_2 = 0.55 s: Total signal length of 0.90 s, comprising 900 samples (350 samples for t_1, 550 samples for t_2). This duration balances coverage of the P wave, QRS complex, and T wave; t_1 = 0.40 s, t_2 = 0.60 s: Total signal length of 1.00 s, comprising 1000 samples (400 samples for t_1, 600 samples for t_2). This provides the most comprehensive segment,

accommodating variability in ECG morphology across individuals.The choice of these durations was based on prior studies by Baloglu et al. (2019) and aimed to optimize the inclusion of diagnostically significant features, such as the P wave, QRS complex, and T wave, while maintaining manageable segment lengths for classification tasks. These segments were used to develop a robust classification model for myocardial infarction, leveraging features derived from signal morphology and dynamics. The results of ECG signal beat segmentation for each MI class and the Normal class, with varying signal lengths (Samples), can be seen in the Table 1.

Table 1. The number of beats in the 12-lead ECG PTB-XL dataset

Class	Signal Length					
	600	650	700	800	900	1000
IMI	24696	25095	25965	29497	29455	29402
ASMI	34130	34504	35739	40554	40516	40449
ILMI	6568	6613	6833	7772	7764	7759
LMI	5824	5874	6092	6900	6890	6890
NORM	49988	50298	52272	59119	59071	58975
AMI	24836	25089	26000	29450	29430	29350
ALMI	25897	26024	27015	30590	30560	30510
IPLMI	4996	5024	5209	5900	5890	5890
IPMI	5822	5839	6123	6860	6860	6860
PMI	1150	1150	1202	1350	1350	1350
TOTAL	183907	185510	192450	217992	217786	217435

The results of beat segmentation for the 12-lead ECG signal with varying signal lengths (Samples) can be seen in Table 2.

Table 2. The number of beats in 9 MI classes and the normal class in the PTB-XL dataset

LEAD	Signal Length					
	600	650	700	800	900	1000
I	17366	17606	17605	20671	20638	20579
II	14467	14620	14620	17170	17151	17107
III	14630	14722	14722	17305	17296	17277
V1	19409	19479	19478	22904	22891	22889
V2	18059	18141	18141	21339	21323	21320
V3	14952	15062	15062	17705	17699	17693
V4	11543	11650	13707	13690	13683	13655
V5	12214	12365	14548	14528	14505	14461
V6	15118	15329	18034	17995	17953	17896
AVR	16785	16846	16845	19816	19811	19797
AVL	15625	15828	15827	18588	18570	18522
AVF	13739	13862	13861	16281	16266	16239
TOTAL	152074	153284	160225	180151	179997	179749

Development of a 12-Lead Classifier Model Framework Using 1D-CNN

To determine the optimal classification model, a series of experiments were conducted to find the simple sequential classification model framework with the best accuracy. These experiments involved comparing 8 different 1D-CNN models with varying numbers of convolution layers, combined with different Maxpooling, Dropout, Flatten, and Dense layers in each model. The comparison of the 1D-CNN architectures used in the tests to determine the best classification model framework is shown in Table 3. Each test was conducted using a dataset that had previously undergone the segmentation process.

Table 3. The number of beats in 9 MI classes and the normal class in the PTB-XL dataset

Layer	Architecture							
	1	2	3	4	5	6	7	8
ConvID	1	1	2	2	3	3	4	4
Max-pooling	0	1	0	1	0	1	0	1
Dropout	1	2	1	2	1	2	1	2
Flatten	1	1	1	1	1	1	1	1
Danse	1	2	1	2	1	2	1	2
Test Accuracy	74.2	73.0	75.4	76.17	74.8	75.7	75.65	75.81
Avearge Time	136	138	144	151	168	168	410	435
Total Time	6812	1985	7190	7426	8399	8478	20481	21852

The 1D-CNN model consists of two fully connected layers, each with 650 Samples, and one node in the output layer. This model requires input in the form of three dimensions, including the number of samples (n samples) the number of features (n features), and timesteps. In this experiment, parameters such as input size, kernel size, activation function, number of epochs, and batch size were kept the same for each 1D-CNN model tested.

The training and testing processes were performed using lead V4 with 10 classes, including MI and Normal classes. The selection of lead V4 was based on research conducted by Baloglu et al. (2019), which identified lead V4 as the lead with the highest accuracy. The test data consisted of 185,510 beats, with 135,212 beats for the MI class and 50,298 beats for the Normal class. This data was split into 70% training data and 30% testing data (15% for validation and 15% for testing).

Based on the comparison of accuracy from various architectures and the number of CNN layers used, the 4th model with two convolution layers was chosen for this study. The 4th architecture was selected because it provided the best accuracy as well as shorter training time compared to the 5th, 6th, 7th, and 8th models (with an average training time of 151 seconds and a total training time of 7,426 seconds for 50 epochs). This indicates that fewer convolution layers can significantly reduce computation time without compromising accuracy.

The proposed 1D-CNN architecture (Table 4) consists of two convolution layers. The first layer uses 128 filters with an 11x1 kernel and the ReLU activation

function, producing feature maps that serve as input for the next process. The second layer uses 128 filters with a 13x1 kernel and the ReLU activation function, followed by a Flatten layer to convert the feature maps into a vector.

Table 4. Details of the proposed CNN architecture layers

Layer	Type	Output Shape	Parameter
conv1d_2	(Conv1D)	(None,690.128)	1536
max_pooling1d_1	(MaxPooling1D)	(None, 345.128)	0
dropout_2	(Dropout)	(None, 345.128)	0
conv1d_3	(Conv1D)	(None, 333.128)	213120
flatten_1	(Flatten)	(None, 42624)	0
dense_1	(Dense)	(None, 128)	5456000
dropout_3	(Dropout)	(None, 128)	0
dense_2	(Dense)	(None, 10)	1290

Next, there is a Dense layer with 128 samples using ReLU for initial classification, followed by a Dropout layer to prevent overfitting. The final layer is a Dense layer with a softmax activation function to generate output probabilities. The model is compiled using the categorical_crossentropy loss function, the Adam optimizer (learning rate 0.001), and accuracy as the metric. Training is performed for 50 epochs using the fit method, training the entire dataset to produce the best model.

Model Evaluation

After determining the classification model framework, the next step is to train and test the classification model using the segmented PTB-XL dataset. Training is the process of building the model, while testing is the process used to evaluate the model after training.

Model evaluation uses evaluation data that must be different from the data used to train the model. At this stage, training and testing are performed on 12-lead ECG signals with a total of 10 MI classes and the normal class, with a 70% training data and 30% testing data split (15% for validation and 15% for testing). Testing with this evaluation data will yield the actual accuracy of a trained model. However, accuracy is not the only metric to focus on during evaluation, as high accuracy can sometimes be misleading due to dataset imbalance. Therefore, other evaluation metrics are needed for a more comprehensive assessment. As a result, performance evaluation requires the use of additional metrics such as sensitivity, specificity, precision, F1-score, and the confusion matrix.

Result and Discussion

Subsequently, the model is compiled and trained using Keras, optimized with the Adam optimizer with a learning rate (lr) of 0.001 and a beta of 0.0001. The loss function used is "categorical_crossentropy", which is commonly employed for multi-class classification problems. The accuracy metric is used to evaluate the model's performance during the training process. The model will be trained using the fit function with data X (features) and Y (labels in categorical form) as training inputs. The number of epochs is set to 50, which corresponds to the number of iterations through the entire dataset. The batch size used is 128, meaning that the model will be trained using batches of 128 data points in each iteration.

Performance Evaluation

Table 5 contains the performance evaluation results of the model for each ECG signal length for 12 leads (I, II, III, V1, V2, V3, V4, V5, V6, AVR, AVL, AVF), divided into results from Training Acc., Validation Acc., and Testing Acc. The following table shows the performance metrics of accuracy for the proposed CNN model on the PTB-XL dataset at signal lengths of 600, 650, 700, 800, 900, and 1000 samples.

Table 5. Accuracy of the CNN Model with a signal length of 600 samples

Lead	Training Acc (%)	Validation Acc (%)	Testing Acc (%)
I	82.05	77.43	78.33
II	89.26	79.55	77.52
III	86.24	78.36	78.18
V1	81.25	79.09	80.41
V2	82.74	82.83	83.77
V3	81.38	82.03	81.47
V4	87.69	84.12	83.84
V5	89.18	83.25	83.77
V6	87.44	80.78	81.68

Table 6. Accuracy of the CNN model with a signal length of 650 samples

Lead	Training Acc (%)	Validation Acc (%)	Testing Acc (%)
I	90.13	77.85	76.66
II	90.43	81.40	81.17
III	84.68	75.15	75.57
V1	84.34	60.92	62.39
V2	84.99	82.07	81.41
V3	76.06	75.66	76.07
V4	87.83	83.35	81.14
V5	87.50	74.66	74.18
V6	88.86	83.30	82.27
AVR	82.16	80.85	80.46
AVL	86.00	68.13	69.15
AVF	91.15	72.31	73.60
Average	86.18	76.30	76.17

Based on the data from Tables 5 to 10, an analysis of the CNN model's performance was conducted at various signal lengths. The highest training accuracy (86.18%) was achieved at a signal length of 650 samples, while the

lowest (75.14%) occurred at 1000 samples. The trend suggests that increasing signal length generally reduces training accuracy, likely due to the increased complexity of longer signals, making them more challenging for the model to learn.

For validation accuracy, the highest value (80.7%) was recorded at 700 samples, while the lowest (76.3%) was observed at 650 samples, possibly indicating overfitting due to the significant gap between training and validation accuracy. Similarly, the highest testing accuracy (80.69%) was achieved at 700 samples, closely followed by 600 samples (80.68%), whereas the lowest (77.5%) occurred at 1000 samples, reflecting a decline in model performance as signal length increased.

Table 7. Accuracy of the CNN model with a signal length of 700 samples

Lead	Training Acc (%)	Validation Acc (%)	Testing Acc (%)
I	79.46	76.24	77.68
II	93.79	8379	82.92
III	85.80	80.09	79.74
V1	80.00	79.36	78.92
V2	82.53	82.69	83.02
V3	81.10	81.01	79.99
V4	84.66	80.89	79.08
V5	88.29	83.50	84.00
V6	89.97	82.58	83.47
AVR	81.55	80.02	80.96
AVL	86.87	77.73	78.10
AVF	86.22	80.54	80.37
Average	85.02	80.70	80.69

Table 8. Accuracy of the CNN model with a signal length of 800 samples

Lead	Training Acc (%)	Validation Acc (%)	Testing Acc (%)
I	76.11	75.62	77.88
II	93.37	82.63	82.41
III	78.21	76.58	76.68
V1	79.03	75.10	75.66
V2	83.08	81.99	82.64
V3	85.20	81.14	81.12
V4	85.99	83.08	82.28
V5	87.55	83.96	84.92
V6	87.96	81.39	81.57
AVR	78.59	80.30	80.45
AVL	83.66	77.47	77.44
AVF	84.66	79.38	77.86
Average	83.62	79.89	80.08

Overall, a signal length of 700 samples demonstrated the best performance, with the highest combined validation and testing accuracy. Shorter signal lengths (600–700 samples) proved more optimal for the CNN model, while longer ones (800–1000 samples) exhibited declining performance across all metrics, likely due to increased data complexity exceeding the model's capacity. The 700-sample signal was selected as the best, considering accuracy (96.18%), sensitivity (82.84%), specificity (97.63%), precision (84.13%), and F1-score (82.68%). Compared to other signal lengths, it provided an optimal balance between sensitivity and specificity while maintaining stable performance without overfitting or degradation in other metrics. Thus, this selection was based on a comprehensive evaluation beyond just accuracy, ensuring the model's generalization ability.

Table 9. Accuracy of the CNN model with a signal length of 900 samples

Lead	Training Acc (%)	Validation Acc (%)	Testing Acc (%)
I	73.85	73.78	77.23
II	89.33	82.59	81.71
III	77.58	75.72	75.10
V1	77.30	78.61	77.84
V2	79.95	80.58	81.46
V3	78.63	79.67	79.99
V4	81.32	78.87	80.10
V5	85.25	81.72	83.72
V6	85.31	80.60	81.57
AVR	77.90	79.83	79.88
AVL	81.15	74.32	74.80
AVF	87.71	80.74	80.99
Average	81.27	78.92	79.53

Table 10. Accuracy of the CNN model with a signal length of 1000 samples

Lead	Training Acc (%)	Validation Acc (%)	Testing Acc (%)
I	75.76	75.55	76.99
II	90.18	74.82	71.84
III	77.37	65.69	65.30
V1	78.30	77.15	77.43
V2	74.44	80.62	81.68
V3	78.80	80.47	80.73
V4	81.40	76.56	77.12
V5	82.75	81.94	81.44
V6	85.93	79.67	81.27
AVR	80.00	80.54	80.08
AVL	81.82	77.56	76.83
AVF	88.13	79.23	79.30
Average	75.14	77.21	77.50

A comparison of accuracy results across the six tables revealed that the best accuracy (training, validation, and testing) for the PTB-XL dataset was obtained with a 700-sample ECG signal (Table 7). Lead II had the highest training accuracy (93.79%), while Lead V1 had the lowest (80.00%). Lead V2 recorded the highest validation accuracy (82.69%), whereas Lead AVL had the lowest (77.73%). For testing accuracy, Lead V5 performed best (84.00%), while Lead V4 had the lowest (79.08%). In some leads, such as AVL, testing accuracy was lower than both training and validation accuracy, suggesting potential generalization issues. The

average accuracy across all leads was 85.02% for training, 80.70% for validation, and 80.69% for testing.

This study also observed that increasing signal length tended to decrease training accuracy for CNN models in MI classification based on 12-lead ECG signals. To confirm that this trend was statistically significant rather than a subjective observation, a one-way ANOVA was performed on accuracy results across different signal lengths (600, 650, 700, 800, 900, and 1000 samples). The ANOVA results indicated significant differences in accuracy between signal length groups. Further analysis using Tukey's post-hoc test identified a significant accuracy decline between 700-sample and 1000-sample signals ($p < 0.05$), confirming that model performance deteriorates as signal length increases.

Table 11. Performance matrix of 12-lead ECG signals in the PTB-XL dataset with a signal length of 700 samples

Lead	ACC (%)	SEN (%)	SPE (%)	PRE (%)	F1-Score (%)
I	95.54	79.48	97.20	80.06	78.63
II	96.58	85.53	97.84	86.63	85.80
III	95.95	83.42	97.53	84.02	83.40
V1	95.79	81.59	97.34	85.31	81.90
V2	96.61	82.56	97.87	86.38	82.54
V3	96.00	79.76	97.51	82.20	79.12
V4	96.34	81.19	97.76	83.69	81.35
V5	96.80	85.81	98.06	85.81	85.49
V6	96.69	85.77	97.91	86.99	86.11
AVR	96.20	82.36	97.59	83.87	82.54
AVL	95.62	81.53	97.31	81.25	81.05
AVF	96.08	85.12	97.63	83.37	84.16
Average	96.18	82.84	97.63	84.13	82.68

From Table 11, it can be concluded that the CNN model performs exceptionally well across all ECG leads. Lead V5 achieves the highest accuracy (96.80%) and specificity (98.06%), making it the most reliable lead for detecting ECG patterns and recognizing negative cases. Lead V6 records the highest precision (86.99%) and F1-Score (86.11%), indicating an optimal balance between sensitivity and precision for MI detection. The model maintains strong overall performance, with an average accuracy of 96.18%, sensitivity of 82.84%, specificity of 97.63%, precision of 84.13%, and an F1-Score of 82.68%. However, Lead I demonstrates the lowest performance across key metrics, with sensitivity (79.48%), precision (80.06%), and F1-Score (78.63%) being comparatively lower. Despite this, the model still effectively detects ECG patterns in this lead. Overall, the CNN model exhibits strong classification capability for MI and normal data, with the best performance observed in Leads V5 and V6.

This analysis highlights the leads with the highest and lowest accuracy and the factors influencing their performance. The top-performing leads, V5, V6, and II, benefit from both physiological and technical advantages. V5 and V6, positioned in the left precordial area, provide optimal detection of left ventricular activity, which is commonly affected by MI. Their close proximity to the heart's primary electrical source ensures stronger and cleaner signals. Lead II is particularly effective in capturing P, QRS, and T waves and is more resistant to noise than other limb leads. Conversely, Leads I, AVL, and AVR exhibit the lowest performance. Leads I and AVL are more sensitive to lateral wall activity, which is less relevant for most MI cases, and their electrode placement on the extremities makes them more prone to electrical interference. Lead AVR, being less physiologically informative, produces low-amplitude signals with more noise, making classification more challenging for the model.

Table 12. Performance metrics of CNN model for Myocardial Infarction (MI) subtypes classification in the PTB-XL dataset with a signal length of 700 samples

Performance Matrix	Class									
	IMI	ASMI	ILMI	LMI	NORM	AMI	ALMI	IPLMI	IPMI	PMI
Accuracy	88.79	88.95	96.62	99.88	90.09	98.80	98.91	99.91	99.88	99.99
Sensitivity	60.30	67.56	68.15	99.63	85.94	98.66	99.19	99.44	99.56	100
Specificity	94.64	93.61	99.18	99.90	91.49	98.81	98.86	99.92	99.88	99.99
Precision	69.38	70.75	66.35	96.93	78.81	92.77	94.21	97.35	96.59	98.20
F1-Score	63.77	68.90	66.02	98.24	82.12	95.61	96.62	98.36	98.03	99.07

The training and validation graphs are used to monitor the model's performance during the training process. These graphs typically display the changes in accuracy and loss of the model on the training and validation datasets over iterations or epochs. The following detailed Training and Validation (Accuracy and Loss) graphs for lead V5 at a signal length of 700 are shown in the Figure 5 below.

The training accuracy (train acc) and validation accuracy (val acc) graph (Figure 5) shows a comparison between the training accuracy and the validation accuracy of the model on lead V5, evaluated over 50 epochs. The training accuracy shows a consistent increase, reflecting the ongoing learning process and the model's ability to adjust parameters for the training data. On the other hand, the validation accuracy tends to fluctuate but generally shows an upward trend,

indicating an improvement in the model's ability to generalize to unseen data. By the 50th epoch, both curves exhibit a convergence pattern, signaling good alignment between training and validation performance and minimal overfitting. This phenomenon highlights the model's stability and efficiency in completing the given task.

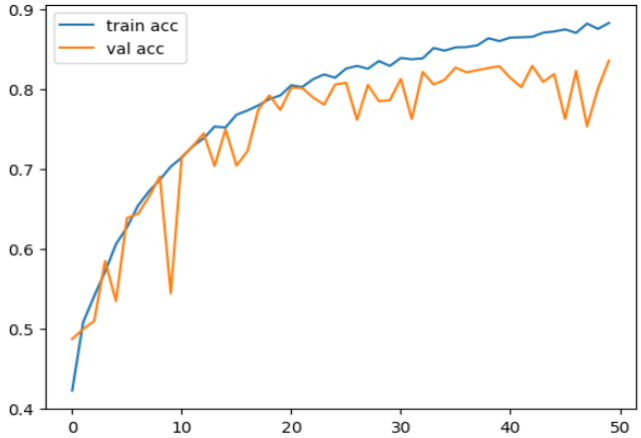

Figure 5. Graph of train accuracy and validation accuracy on lead V5

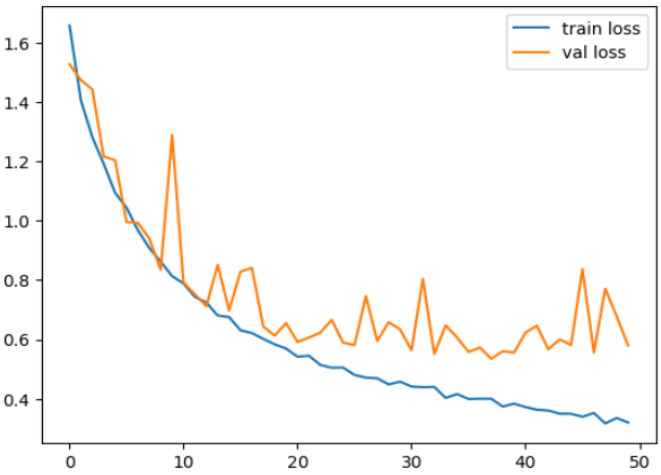

Figure 6. Graph of train loss and validation loss on lead V5

The train loss and validation loss graph (Figure 6) illustrates the comparison between train loss and validation loss on lead V5 over 50 epochs during the deep learning model training process. Initially, the train loss (blue line) decreases sharply, indicating rapid improvement as the model learns to minimize errors. The validation loss (orange line) fluctuates significantly in the early epochs, suggesting instability in the model's generalization ability. However, both loss curves show a decreasing trend as training progresses, indicating that the model is effectively optimizing its performance on both the training and validation datasets. By the end of training (epoch 50), the train loss and validation loss approach convergence, signaling minimal overfitting and the model's robustness in handling unseen data. The alignment between the two losses reinforces the effectiveness of the training strategy used.

The confusion matrix (Figure 7) displayed shows the classification performance of the model across 10 classes. The main diagonal (from the top left to the bottom right) represents the correct predictions (true positives) for each class. For example, the model correctly predicted 181 samples from the first class, 211 samples from the second class, and so on. The class with the highest number of correct predictions is the fifth class (573), indicating that the model performs very well on this class.

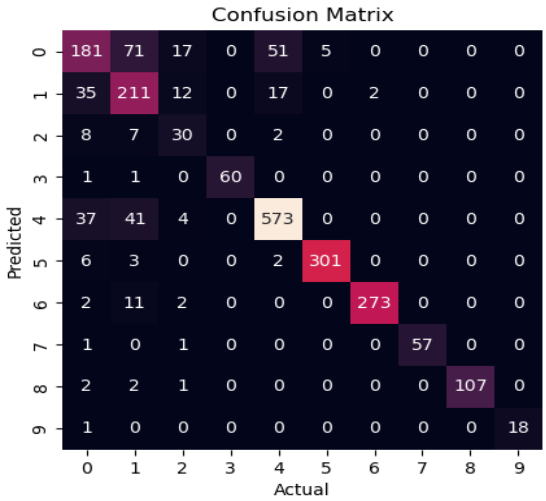

Figure 7. Confusion matrix on lead V5

However, some classes, such as the third class (30), have much lower correct predictions, which may indicate the model's difficulty in recognizing patterns from this class. The imbalance in the number of correct predictions across classes, along with the errors between classes, suggests that the model may need further refinement, such as through data adjustments or training parameter tuning, to improve accuracy for certain classes.

Table 13. Performance matrix of 12-lead ECG signals in the PTB-XL dataset at each signal length

Node	ACC (%)	SEN (%)	SPE (%)	PRE (%)	F1-Score (%)
600	95.38	82.57	97.59	84.43	82.50
650	95.22	77.69	97.02	81.43	78.28
700	96.18	82.84	97.63	84.13	82.68
800	96.06	81.91	97.55	83.86	81.82
900	95.91	81.08	97.44	82.87	80.83
1000	95.51	78.68	97.21	81.69	78.68

Based on Table 13, the highest accuracy (ACC) was achieved with a signal length of 700 Samples, with a value of 96.18%, indicating the overall best performance

in classification. The accuracy slightly decreased for other signal lengths, with the lowest value being 95.22% for 650 Samples. The accuracy difference between signal lengths is quite small (< 1%), showing the model's consistency in classification across different signal lengths.

The highest sensitivity (SEN) was also recorded at 700 Samples with a value of 82.84%, indicating the model's best ability to detect positive samples. However, the lowest sensitivity was observed at 650 Samples with a value of 77.69%, which may suggest that the model struggled to capture all positive cases at this signal length. This trend indicates that a signal length around 700 Samples is more optimal for sensitivity. Specificity (SPE) showed the highest value at 700 Samples, with a score of 97.63%, indicating the model's best ability to recognize negative samples. Specificity values for other signal lengths remained high (> 97%), indicating that the model consistently identified negative samples well, even with shorter or longer signal lengths.

The highest precision (PRE) was achieved at 700 Samples with a value of 84.13%, showing that the model at this signal length was the best at minimizing false positives. The lowest precision was recorded at 650 Samples with a value of 81.43%, though the difference remains relatively small compared to other signal lengths. The F1-Score, which is a combination of precision and sensitivity, was highest at 700 Samples with a value of 82.68%. The lowest F1-Score was observed at 650 Samples with a value of 78.28%, due to the lower sensitivity at this signal length. Overall, the best performance across all metrics was recorded at a signal length of 700 Samples, indicating that this is the optimal signal length for the CNN model in this study.

Performance Evaluation using K-Fold Cross Validation

K-Fold is a commonly used cross-validation technique for evaluating model performance. This technique splits the dataset into k equal parts (folds) and performs model training and testing k times. The model is trained using k-1 folds and tested on the remaining fold. This process is repeated k times, with each fold being used as the test set exactly once.

Table 14 shows the classification model performance on training, validation, and testing data based on various leads at a signal length of 700. Overall, the training accuracy has the highest average of 94.01%, reflecting the model's ability to learn patterns from the training data well. Lead V4 recorded the highest validation accuracy at 85.36%, followed by V5 with 85.55%, indicating better generalization ability compared to other leads. For testing accuracy, lead V2 achieved the highest value at 84.94%, followed by V4 at 83.65%, confirming the stability of the model's performance on unseen data. On the other hand, leads AVR and AVL showed low validation and testing accuracy, with averages below 80%, which may suggest that signals from these leads are less informative for the classification task.

Table 14. CNN model accuracy on PTB-X dataset test data with 5-fold cross validation

Lead	Traning Acc (%)	Validation Acc (%)	Testing Acc (%)
I	96.48	80.07	79.79
II	97.30	83.69	78.81
III	92.97	80.45	79.08
V1	90.00	80.37	80.00
V2	90.82	83.89	84.94
V3	90.44	84.19	82.46
V4	94.22	85.36	83.65
V5	96.62	85.55	80.89
V6	96.12	83.83	83.42
AVR	90.32	79.16	78.83
AVL	96.72	79.17	78.83
AVF	96.12	83.11	83.42
Average	94.01	82.40	81.18

Analysis of the average accuracy indicates that although the model performs well in training, there is significant variation between leads in terms of generalization ability. Leads such as V4 and V2 may be the focus for further analysis due to their consistent performance in validation and testing. In both graphs (Figure 8), it can be observed that all folds exhibit a similar pattern of increase, both for training accuracy and validation accuracy. Figure 8 shows that the model has stability across folds during the training and evaluation process. The training accuracy graph consistently increases as the epochs progress and reaches a value close to 0.9 in all folds by the end of the training. There is no indication of overfitting, as the training accuracy does not diverge significantly from the validation accuracy.

Validation accuracy tends to rise until it approaches 0.8 for most folds. However, there is slight variation at certain epochs, especially in specific folds (e.g., Fold 3, which shows more fluctuation). This fluctuation pattern may be caused by variations in data distribution within each fold, but overall, it shows that the model can learn effectively from the data. Both training accuracy and validation accuracy exhibit increasing trends and convergence, indicating that the model can learn patterns from the training data without losing generalization ability on the validation data. The average validation accuracy close to 0.8 reflects the model's good performance in processing the provided data

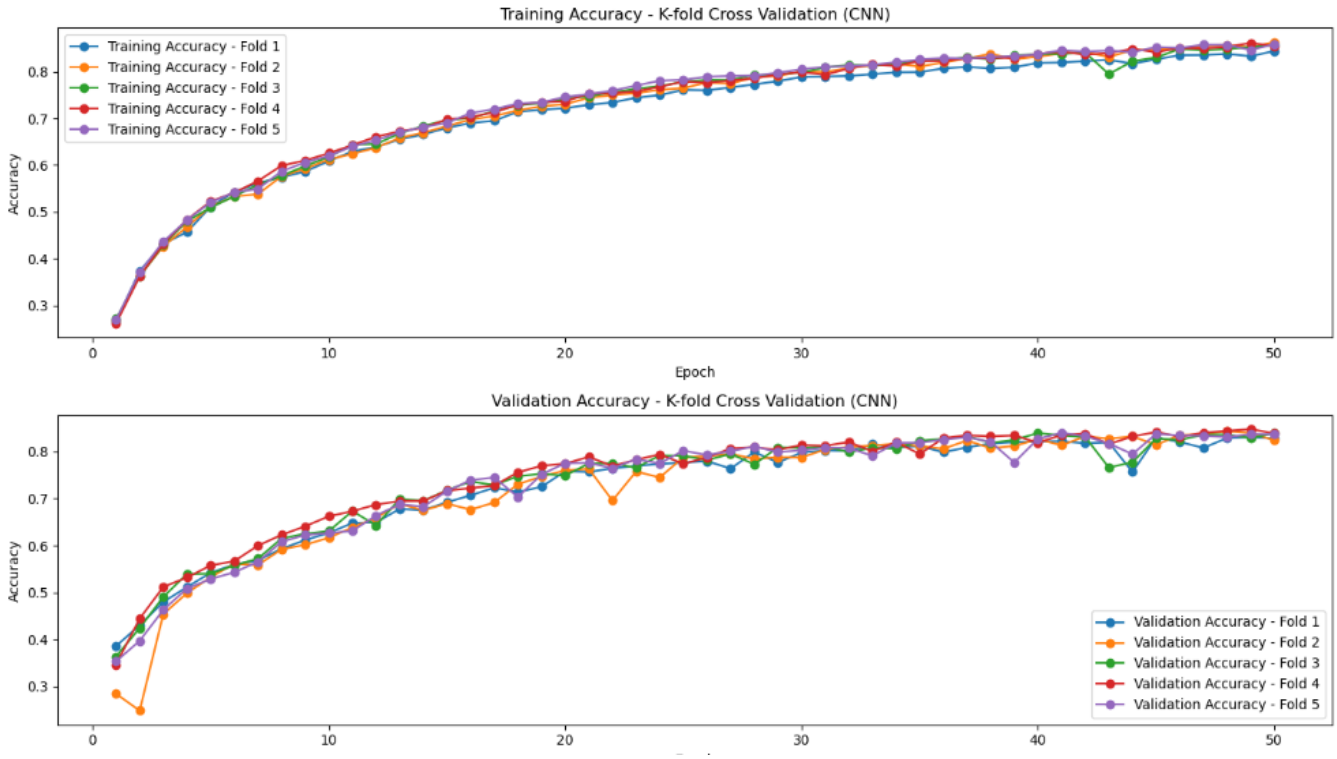

Figure 8. Training accuracy and validation accuracy graph in Lead V2 with 5-fold cross-validation

Validation accuracy tends to rise until it approaches 0.8 for most folds. However, there is slight variation at certain epochs, especially in specific folds (e.g., Fold 3, which shows more fluctuation). This fluctuation pattern may be caused by variations in data distribution within each fold, but overall, it shows that the model can learn effectively from the data. Both training accuracy and validation accuracy exhibit increasing trends and convergence, indicating that the model can learn patterns from the training data without losing generalization ability on the validation data. The average validation accuracy close to 0.8 reflects the model's good performance in processing the provided data.

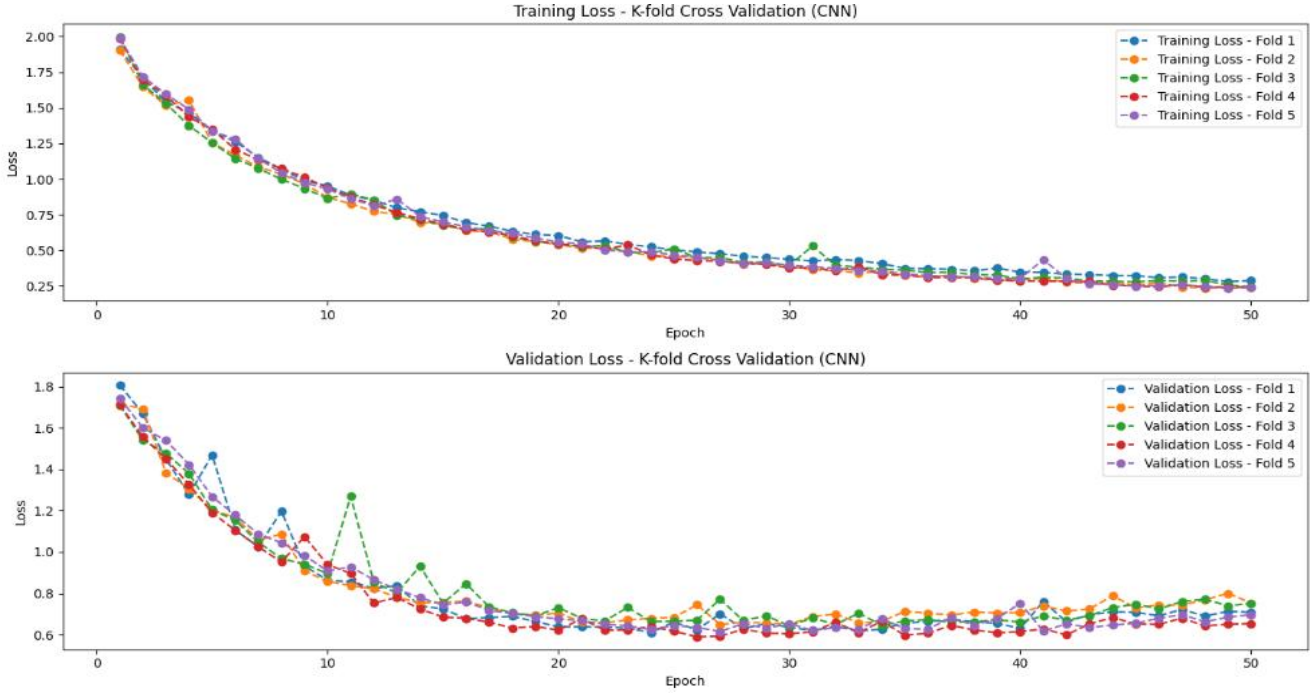

Figure 9. Training loss and validation loss graph in Lead V2 with 5-fold cross-validation

In the training loss graph (Figure 9), a consistent downward trend is observed across all leads. At the beginning of the training, the loss value is relatively high (around 2.0) but decreases significantly and stabilizes at a low value (~0.2) after 40 epochs. This pattern indicates that the model successfully learns the patterns from the training data, with the error decreasing as the epochs progress. The consistency across leads suggests that the model has good stability during the training process.

The validation loss graph (Figure 9) also shows a similar decreasing pattern, although there is some fluctuation in certain leads, especially during the early stages of training (0–20 epochs). After around 20 epochs, the validation loss tends to decrease and stabilize around ~0.6. This initial fluctuation may indicate partial overfitting, which is later controlled as the training progresses. The stability of the validation loss at the end of the training indicates that the model has developed good generalization ability.

Overall, the model demonstrates consistent and stable performance during training, marked by the decrease in both training loss and validation loss over time. Although there is some fluctuation in the validation loss at the beginning, the final results show that the model successfully reduces errors on both training and validation data. This suggests that the model has been well-trained and has a high potential for generalizing to new data.

Table 15 presents the performance evaluation of the classification model across various ECG leads using key metrics: accuracy (ACC), sensitivity (SEN), specificity (SPE), precision (PRE), and F1-score. On average, the model shows strong performance with an accuracy of 85.03%, sensitivity of 85.98%, specificity of 79.12%, precision of 97.58%, and F1-score of 98.17%. These metrics indicate reliable classification capabilities, particularly in precision and F1-score, reflecting high effectiveness in identifying positive cases with minimal false positives.

Among the leads tested, AVL stands out with the highest accuracy of 96.72%, indicating excellent classification ability. The highest sensitivity is achieved by lead V2 with 95.21%, showing optimal performance in detecting positive cases. Lead V5 records the highest specificity at 98.00%, demonstrating its ability to accurately identify negative cases. Several leads such as V1, V2, V4, and V5 achieve perfect precision (100%), meaning all positive predictions for these leads are accurate, and the same leads also achieve a perfect F1-score (100%), indicating a balanced precision and sensitivity.

Table 15. Evaluation results of 12 lead ECG signals on the PTB-XL dataset with 5-fold cross validation

Lead	ACC (%)	SEN (%)	SPE (%)	PRE (%)	F1-Score (%)
I	80.12	73.88	70.69	96.00	100
II	80.79	86.78	70.69	97.00	99.00
III	83.00	85.52	90.59	99.00	100
V1	80.00	90.00	80.23	100	100
V2	85.10	95.21	70.71	100	100
V3	82.23	97.02	73.60	99.00	100
V4	81.83	84.80	77.88	100	100
V5	81.00	84.80	98.00	100	100
V6	83.12	85.39	77.88	84.00	83.00
AVR	90.32	83.69	71.09	100	99.00
AVL	96.72	79.09	77.53	97.00	98.00
AVF	96.12	85.55	90.58	99.00	99.00
Average	85.03	85.98	79.12	97.58	98.17

Lead I shows the lowest accuracy (80.12%) and specificity (70.69%), suggesting limited ability to differentiate between positive and negative cases. Lead II and V2 also show low specificity (70.69%), indicating challenges in accurately identifying negative cases for these leads.

Overall, the model demonstrates excellent performance, especially in precision and F1-score for most leads, with standout results for AVL, V2, and V5. However, the low specificity in certain leads such as I, II, and V2 highlights areas that need optimization to improve the robustness and reliability of the model for clinical applications.

Table 16. Model performance matrix for MI and normal class of PTB-XL dataset with 5-fold cross validation

Performance Matrix	Class							
	IMI	ASMI	ILMI	LMI	ALMI	IPLMI	IPMI	PMI
Accuracy	80.32	85.05	83.62	100	85.91	99.32	100	100
Sensitivity	81.49	82.59	82.13	100	85.92	99.33	100	100
Specificity	82.62	81.42	99.21	100	96.32	97.73	83.13	98.83
Precision	83.50	83.65	91.43	78.78	96.00	96.00	95.93	95.92
F1-Score	80.00	84.63	82.31	97.00	78.00	98.00	98.00	100

Table 16 presents the performance evaluation results of several classification models based on five key metrics: accuracy, sensitivity, specificity, precision, and F1-score. Each metric reflects the model's ability to classify data effectively and efficiently. Based on the data, the PMI class demonstrates the best performance across all metrics, with accuracy, sensitivity, specificity, precision, and F1-score all achieving a perfect value of

100%. This indicates that the PMI class can perfectly identify both positive and negative data without any classification errors.

The LMI class also shows excellent results, particularly in accuracy, sensitivity, and F1-score, all reaching 100%, reflecting its ability to classify very precisely. However, despite its high performance, the LMI class has a lower precision value compared to PMI, at 78.78%, indicating slight imprecision in predicting positive cases. In contrast, the IMI and ASMI classes perform lower on several metrics, particularly F1-score and sensitivity, with values of 80.00 and 81.49% for IMI, and 82.59% for ASMI. Nevertheless, these classes remain competitive in terms of accuracy and specificity.

Overall, the PMI class excels in various aspects of classification performance, followed by the LMI class, which also shows excellent results. Other classes, although performing well, show some shortcomings in specific metrics such as precision and F1-score. This evaluation is important for selecting the most suitable model based on the specific needs of the desired classification application.

Conclusion

This study successfully developed an efficient 1D-CNN model for myocardial infarction (MI) classification using 12-lead ECG signals. The proposed model achieved an average accuracy of 96.18% with an optimal signal length of 700 samples, demonstrating high performance while reducing computational complexity. The evaluation results indicate that leads V5 and V6 perform best in detecting MI, whereas leads I and AVL require further optimization. Additionally, other evaluation metrics, including sensitivity (82.84%), specificity (97.63%), precision (84.13%), and F1-score (82.68%), confirm the model's balanced ability to classify MI cases accurately. These findings highlight that optimizing signal length and adopting a simplified CNN architecture can enhance efficiency without compromising accuracy. The proposed model has strong potential for real-time clinical applications, particularly in resource-limited healthcare settings. Future research could focus on optimizing the performance of lower-performing leads and integrating this model into portable medical devices for early MI detection.

Acknowledgments
Thank you to all parties who have helped a lot in this research.

Author Contributions
Conceptualization, methodology, validation, formal analysis, investigation, resources, data curation, A.H.M.; writing—original draft preparation, A.H.M. and R.M.N.H.; writing—review and editing, A.H.M., R.M.N.H., and M.J.H.; visualization, R.M.N.H. and M.J.H. All authors have read and agreed to the published version of the manuscript.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest

References

Baloglu, U. B., Talo, M., Yildirim, O., Tan, R. S., & Acharya, U. R. (2019). Classification of Myocardial Infarction with Multi-Lead ECG Signals and Deep CNN. *Pattern Recognition Letters*, *122*, 23–30. https://doi.org/10.1016/j.patrec.2019.02.016

Banerjee, S., & Mitra, M. (2013). Application of Cross Wavelet Transform for ECG Pattern Analysis and Classification. *IEEE Transactions on Instrumentation and Measurement*, *63*(2), 326–333. https://doi.org/10.1109/TIM.2013.2279001

Brady, W. J., Hwang, V., Sullivan, R., Chang, N., Beagle, C., Carter, C. T., Martin, M. L., & Aufderheide, T. P. (2000). A Comparison of 12-and 15-Lead ECGs in ED Chest Pain Patients: Impact on Diagnosis, Therapy, and Disposition. *The American Journal of Emergency Medicine*, *18*(3), 239–243. https://doi.org/10.1016/S0735-6757(00)90112-8

Chakraborty, A., Chatterjee, S., Majumder, K., Shaw, R. N., & Ghosh, A. (2022). A Comparative Study of Myocardial Infarction Detection from ECG Data Using Machine Learning. *Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2021*, 257–267. https://doi.org/10.1007/978-981-16-2164-2_21

Darmawahyuni, A., & Nurmaini, S. (2019). Deep Learning with Long Short-Term Memory for Enhancement Myocardial Infarction Classification. *2019 6th International Conference on Instrumentation, Control, and Automation (ICA)*, 19–23. https://doi.org/10.1109/ICA.2019.8916683

Fariha, M. A. Z., Ikeura, R., Hayakawa, S., & Tsutsumi, S. (2020). Analysis of Pan-Tompkins Algorithm Performance with Noisy ECG Signals. *Journal of Physics: Conference Series*, *1532*(1), 12022. https://doi.org/10.1088/1742-6596/1532/1/012022

Hainaut, C., & Gade, W. (2003). The Emerging Roles of BNP and Accelerated Cardiac Protocols in Emergency Laboratory Medicine. *American Society for Clinical Laboratory Science*, *16*(3), 166–179. https://doi.org/10.29074/ascls.16.3.166

Hasbullah, S., Zahid, M. S. M., & Mandala, S. (2023). Detection of Myocardial Infarction Using Hybrid Models of Convolutional Neural Network and Recurrent Neural Network. *BioMedInformatics*, *3*(2),

478–492. https://doi.org/10.3390/biomedinformatics3020033

Huanhuan, M., & Yue, Z. (2014). Classification of Electrocardiogram Signals with Deep Belief Networks. *2014 IEEE 17th International Conference on Computational Science and Engineering*, 7–12. https://doi.org/10.1109/CSE.2014.36

Mathews, S. M., Kambhamettu, C., & Barner, K. E. (2018). A Novel Application of Deep Learning for Single-Lead ECG Classification. *Computers in Biology and Medicine*, *99*, 53–62. https://doi.org/10.1016/j.compbiomed.2018.05.013

Meek, S., & Morris, F. (2002). Introduction. I—Leads, Rate, Rhythm, and Cardiac Axis. *BMJ*, *324*(7334), 415–418. https://doi.org/10.1136/bmj.324.7334.415

Mezgec, S., & Seljak, B. K. (2017). NutriNet: A Deep Learning Food and Drink Image Recognition System for Dietary Assessment. *Nutrients*, *9*(7), 657. https://doi.org/10.3390/nu9070657

Mirza, A. H., Nurmaini, S., & Partan, R. U. (2022). Automatic Classification of 15 Leads ECG Signal of Myocardial Infarction Using One Dimension Convolutional Neural Network. *Applied Sciences*, *12*(11), 5603. https://doi.org/10.3390/app12115603

Mirza, A. H., Nurmaini, S., Partan, R. U., & Herdiansyah, M. I. (2023). Myorcadial Infarction Classification with 15 Lead Electrocardiogram (ECG) Signal using Hybrid Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BILSTM). *2023 Eighth International Conference on Informatics and Computing (ICIC)*, 1–6. https://doi.org/10.1109/ICIC60109.2023.10381905

Muraki, R., Teramoto, A., Sugimoto, K., Sugimoto, K., Yamada, A., & Watanabe, E. (2022). Automated Detection Scheme for Acute Myocardial Infarction Using Convolutional Neural Network and Long Short-Term Memory. *PLoS One*, *17*(2), e0264002. https://doi.org/10.1371/journal.pone.0264002

National Institute of Health Research and Development. (2018). *National Report on Basic Health Research, RISKESDAS 2018*. Jakarta: Ministry of Health.

Nurmaini, S., Darmawahyuni, A., Mukti, A. N. S., Rachmatullah, M. N., Firdaus, F., & Tutuko, B. (2020a). Deep Learning-Based Stacked Denoising and Autoencoder for ECG Heartbeat Classification. *Electronics*, *9*(1), 135. https://doi.org/10.3390/electronics9010135

Nurmaini, S., Partan, R. U., & Rachmatullah, M. N. (2019). Deep Classifier on the Electrocardiogram Interpretation System. *Journal of Physics: Conference Series*, *1246*(1), 12030. https://doi.org/10.1088/1742-6596/1246/1/012030

Nurmaini, S., Tondas, A. E., Darmawahyuni, A., Rachmatullah, M. N., Partan, R. U., Firdaus, F., Tutuko, B., Pratiwi, F., Juliano, A. H., & Khoirani, R. (2020b). Robust Detection of Atrial Fibrillation from Short-Term Electrocardiogram Using Convolutional Neural Networks. *Future Generation Computer Systems*, *113*, 304–317. https://doi.org/10.1016/j.future.2020.07.021

Partan, R. (2018). Cardiac Arrhythmias Classification Using Deep Neural Networks and Principle Component Analysis Algorithm. *Int. J. Advance Soft Compu. Appl.*, *10*(2).

Rawi, A. A., Albashir, M. K., & Ahmed, A. M. (2022). Classification and Detection of ECG Arrhythmia and Myocardial Infarction Using Deep Learning: A Review. *Webology*, *19*(1), 1151–1170. https://doi.org/10.14704/WEB/V19I1/WEB19078

Ricardo, R. A., Bassani, R. A., & Bassani, J. W. M. (2009). A Simple Laboratory Method for Teaching How Electrocardiogram is Generated. *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany: Vol. 25/12 General Subjects*, 385–387. https://doi.org/10.1007/978-3-642-03893-8_111

Sokhanvar, S. (2006). Evaluation of ST Segment Deviation on 15-Lead ECG in Patients with Acute Coronary Syndrome (ACS). *Journal of Advances in Medical and Biomedical Research*, *14*(56), 32–39.

Strong, A., Sharma, G., Tu, C., Aminian, A., Rodriguez, J., & Kroh, M. D. (2017). A Population-Based Study of Early Postoperative Outcomes in Patients Undergoing Bariatric Surgery with Congestive Heart Failure. *Journal of the American College of Surgeons*, *225*(4), e55. https://doi.org/10.1016/j.jamcollsurg.2017.07.666

Übeyli, E. D. (2010). Recurrent Neural Networks Employing Lyapunov Exponents for Analysis of ECG Signals. *Expert Systems with Applications*, *37*(2), 1192–1199. https://doi.org/10.1016/j.eswa.2009.06.022

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and Composing Robust Features with Denoising Autoencoders. *Proceedings of the 25th International Conference on Machine Learning*, 1096–1103. https://doi.org/10.1145/1390156.1390294

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), 1–15. https://doi.org/10.1038/s41597-020-0495-6

Wang, J., Ye, Y., Pan, X., & Gao, X. (2015). Parallel-Type Fractional Zero-Phase Filtering for ECG Signal Denoising. *Biomedical Signal Processing and Control*, *18*, 36–41.

https://doi.org/10.1016/j.bspc.2014.10.012

Wang, R., Xu, J., & Han, T. X. (2019). Object Instance Detection with Pruned Alexnet and Extended Training Data. *Signal Processing: Image Communication*, *70*, 145–156. https://doi.org/10.1016/j.image.2018.09.013

Xiong, P., Lee, S. M.-Y., & Chan, G. (2022). Deep Learning for Detecting and Locating Myocardial Infarction by Electrocardiogram: A Literature Review. *Frontiers in Cardiovascular Medicine*, *9*, 860032. https://doi.org/10.3389/fcvm.2022.860032

Zhang, J., Lin, F., Xiong, P., Du, H., Zhang, H., Liu, M., Hou, Z., & Liu, X. (2019). Automated Detection and Localization of Myocardial Infarction with Staked Sparse Autoencoder and Treebagger. *IEEE Access*, *7*, 70634–70642. https://doi.org/10.1109/ACCESS.2019.2919068

Zimetbaum, P. J., & Josephson, M. E. (2003). Use of the Electrocardiogram in Acute Myocardial Infarction. *New England Journal of Medicine*, *348*(10), 933–940. https://doi.org/10.1056/NEJMra022700