Lightweight CNN-Based approach for Multi-Class ECG Image Classification

1 B. Rishika, 2 E. Shivani, 3 Mohammed Nawaz Uddin, 4 K. Aakash, 5 Dr. V. Neelima

Department of CSE (Artificial Intelligence & Machine Learning)
Jyothishmathi Institute of Technology and Science Karimnagar, Telangana, India
rishika.billa04@gmail.com, www.shivanienugu@gmail.com, nawaznayeem2562@gmail.com, aakashking147@gmail.com

Abstract— **Cardiovascular diseases (CVDs) continue to be the highest cause of mortality worldwide.[1] This has necessitated the development of efficient, automated, and accurate diagnostic techniques. Although traditional Convolutional Neural Networks (CNNs) perform well in image recognition, there is limited efficiency in dealing with specific multi-scale dependencies of Electrocardiogram (ECG) signals. The purpose of this research is to propose a Parallel Branch Deep Fusion CNN model for diagnosing four different states of cardiac conditions, including Normal, Abnormal Heartbeat, Myocardial Infarction (MI), and History of MI. This model is based on a two-stream architecture, including a Spatial Branch, which employs a pre-trained ResNet-18 model, and a Global Context Branch, designed to obtain long-range dependencies of signals. The model is then optimized using a phased learning mechanism. The proposed model was found to attain a maximum validation accuracy of 89.06%, indicating better stability compared to traditional CNN architectures.**

I. INTRODUCTION

The electrocardiogram (ECG) is considered the primary tool in non-invasive cardiac diagnosis, as it measures electrical activity in the heart to detect life-threatening problems. The depolarization and repolarization of the cardiac muscle, as detected by an electrocardiogram, play a vital role in determining heart rhythms, rates, and integrity. The main challenge in manual interpretation is that it is a time-consuming, subjective, and error-prone process, especially in emergency rooms or intensive care units under pressure. However, recent breakthroughs in Deep Learning (DL) technologies have revolutionized medical imaging. Nevertheless, traditional Convolutional Neural Network (CNN) architectures suffer from a 'representation bottleneck' when dealing with 2D images of ECG signals. Conventional sequential architectures, though successful in recognizing local morphological changes in cardiac signals, like changes in the ST segment, tend to be ineffective in combining these with global contextual patterns of the entire cardiac cycle. This is seen as a limitation in optimizing classification outcomes when dealing with states that are very similar, like a newly developed Myocardial Infarction (MI) and a patient's historical cardiac events (History of MI).

In this research, we propose a novel Parallel Branch Deep Fusion CNN architecture, designed to eliminate these challenges. The proposed model works by simultaneously processing two branches of information in ECG images: a spatial feature extractor and a global context branch.

The primary contributions of this work include:

- Development of a hybrid dual-path architecture that fuses pre-trained residual features (ResNet-18) with a custom dense-to-spatial transformation branch to capture both local and global data.
- Implementation of a phased transfer learning strategy using differential learning rates and label smoothing to stabilize convergence and prevent overfitting.
- An empirical demonstration of efficacy, achieving a peak validation accuracy of 89.06% and an F1-score of 0.93 for Myocardial Infarction across four critical cardiac categories.

This study closes the gap between local and global feature extraction, offering a strong and computationally efficient framework for the automated monitoring of the heart that could ease the clinical burden and hasten the pace of life-saving diagnostic procedures.

II. RELATED WORK

The development of automated ECG signal analysis has moved from traditional signal processing approaches towards complex frameworks involving deep learning approaches. The classical approaches for ECG signal analysis involved the application of machine learning approaches such as Wavelet Transforms, Principal Component Analysis (PCA), and Support Vector Machine (SVM). Although these approaches are successful for simple rhythm classification, the approaches are not complex enough for the classification of complex waveforms.

The introduction of Convolutional Neural Networks (CNNs) has revolutionized the field of ECG signal analysis. The application of CNNs has been approached from two different directions: the application of 1D signal processing and the application of 2D image-based approaches for ECG signal classification. Although the application of 1D signal processing is computationally simple, the application of the 2D image-based approach has shown promising results for the classification of ECG signals using approaches such as the modified version of the VGG-16 and Inception architectures.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | Mar – 2026 DOI: 10.55041/ISJEM05768
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Despite these advances, however, many of the current state-of-the-art models have a purely sequential architecture. The models face difficulties with what they call "intra-class variance," where various conditions, such as Normal vs. History of MI, have very similar visual representations. Recent studies by Abubaker & Babayiğit (2023) emphasized the promising potential of deep residual networks for ECG classification. [2]Yet, the balance of local morphological details and global signal dependencies still remains a problem. Our current study extends these existing works by introducing a new Parallel-Branch Fusion strategy. Unlike existing sequential models, where spatial resolution is lost with deeper network layers, we have a dual-branch structure for global contextual information. This way, we can overcome the limitations of existing sequential models to offer a holistic view of heart health.

III. DATASET DESCRIPTION AND PREPROCESSING

A. Dataset Description

The research uses a publicly accessible dataset that contains a total of 928 standard 12-lead ECG images. These images represent a real-world problem, as ECG signals are often stored in a scanned format. As depicted in Table I, the dataset is split into four classes: Normal Person (NP), Abnormal Heartbeat (AH), Myocardial Infarction (MI), and History of MI (HMI).

TABLE I: Public ECG Image Dataset Description

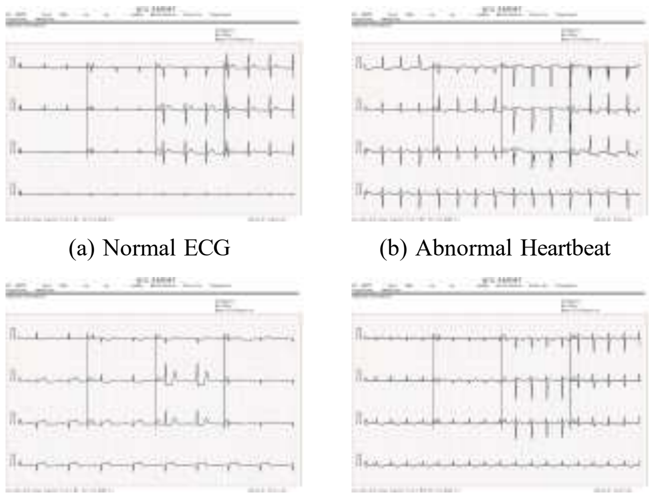
No.	Class	Number of Images
1	Normal ECG	284
2	Abnormal Heartbeat	233
3	Myocardial Infarction	239
4	History of Myocardial Infarction	172
	Total	**928**

The distribution of these classes, as detailed in Table I, shows a slight imbalance, particularly in the "History of MI" category. To ensure robust training and prevent model bias, the initial 928 images were expanded to approximately 4,800 samples (1,200 per class) using synthetic balancing and geometric augmentation.

B. Data Preprocessing and Cropping

The biggest issue associated with the use of scanned ECG records for deep learning models lies in the presence of a large amount of non-diagnostic noise. The printed records of the ECG contain a variety of embedded data, including hospital letterhead, patient demographic data, timestamps, and even comments from physicians. From a machine learning standpoint, these embedded data have a large potential for feature interference. If not cropped out, the convolutional filters may inadvertently associate certain fonts or layouts of header data with certain cardiac conditions. This would cause the network to overfit to non-physiologic data.

To remove this non-diagnostic noise, a specialized header-footer cropping protocol was used. By applying a deterministic vertical crop to the images, cropping the top 20% and the bottom 7% of each frame, the network was able to isolate the 12-lead waveform matrix. This ensures that the network's receptive field remains strictly confined to the P-QRS-T morphological data. Once cropped, the images undergo a color space transform. While most digital image processing algorithms would normally use the BGR color space, it was decided to use RGB color space to ensure compatibility with the pre-trained weights of the transfer learning backbone.



(a) Normal ECG (b) Abnormal Heartbeat

(c) Myocardial Infarction (MI) (d) History of MI

Fig. 1: Sample 12-lead ECG images representing the four diagnostic categories.

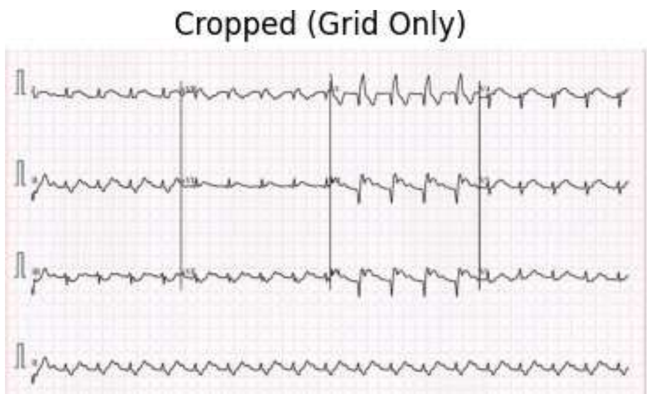


Fig. 2: ECG image after global ROI extraction (227 × 227).

C. Image Resizing and Resolution Standardization

Spatial consistency is essential for the successful extraction of the temporal features from the 2D image. To standardize the input size for the heterogeneous dataset, each isolated signal was resized to an optimal size of 227 × 227 using Bicubic Interpolation. The choice of the size was based on the fact that it was an optimal midpoint between being large enough for the preservation of the critical diagnostic features such as the ST segment elevation, pathological Q-waves, and the small notchings in the QRS complex.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | Mar – 2026 DOI: 10.55041/ISJEM05768
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Moreover, the resizing of the images was accompanied by the normalization of the pixel values using the statistics from the ImageNet dataset (μ = [0.485, 0.456, 0.406], σ = [0.229, 0.224, 0.225]). The normalization of the pixel values ensures that the Parallel Branch CNN architecture can efficiently differentiate the electrical signal from the background grid paper, regardless of the quality of the scan.

D. Synthetic Oversampling and Augmentation

As a result of the inherent class imbalance in the initial dataset, which ranged from 172 to 284 images for each of the classes, a Dynamic Oversampling Strategy was adopted. The initial dataset was iteratively processed, subjecting it to a stochastic augmentation process, until all four classes were balanced at 1,200 images each, thereby producing a total of 4,800 samples.

E. Data Partitioning

The expanded set was then split into training and testing sets by using the Stratified 80/20 Split. This ensures that the class distribution is exactly the same in both sets.

- Training Set: 3,840 Images
- Testing/Validation Set:960 Images

This is an ideal class distribution that ensures that the model does not end up being biased towards classes that are overrepresented in the dataset. It is an ideal class distribution that ensures that the final results obtained by the model are statistically significant.

IV. MODEL ARCHITECTURE

The proposed architecture is strategically partitioned into three functional zones: the Feature Extraction Branches, the Fusion Neck, and the Classification Head. This modular design allows the network to isolate specific cardiac markers before synthesizing them into a final diagnostic prediction.

A. Feature Extraction Branches: Dual-Path Processing

The first functional zone is responsible for converting raw pixel intensities into high-level semantic features.

- The Spatial Branch (ResNet-18): This is used as a local feature descriptor. It uses 18 residual connections, hence avoiding the vanishing gradient problem.[3] This enables the filters to concentrate on minute details such as the deviation in the ST segment or the inversion in the T wave. The feature map generated by this branch is dense, reflecting the "geometry" of the heart's electrical signal.
- The Global Branch: Unlike standard CNNs that risk losing global information with successive pooling operations, the Global Branch treats the ECG image as a whole structural object. This approach follows recent trends in hybrid feature extraction for medical imaging. [2]By flattening the input and passing it through a dense layer, the model creates a latent representation that maintains the relationships between different leads. This is a critical consideration for identifying rhythm abnormalities that present as patterns across the entire 12-lead ECG.

B. The Fusion Neck: Cross-Domain Integration

Fusion Neck acts as a bridge between the "Local Geometry" and the "Global Rhythm." At this point, the Depth Concatenation operation is performed between the 224 channels of the Spatial branch and the 96 channels of the Global branch.

- Inter-Branch Blending: Following the Depth Concatenation operation, a 1×1 convolutional layer is used as a "bottleneck" to perform feature fusion across the channels. This operation is critical in the network, as it enables it to learn what local spatial features correspond to certain global rhythmic dependencies.
- Dimensionality Management: This fusion neck compresses the concatenated channels from 320 to 256.

C. The Classification Head: High-Level Reasoning

The final functional zone maps the feature maps into one of the four clinical categories.

- Feature Flattening and Dense Mapping: The spatial feature maps are flattened into a vector of dimension 1,024, which is then mapped into a fully connected layer of 512 units. This is considered a high-level reasoning unit, as it considers the relative importance of the fused features.
- Regularization and Decision: To prevent overfitting of scanned paper artifacts, such as grid lines or ink splotches, a Dropout layer with dropout probability $p = 0.3$ is introduced. The final decision is made by introducing a Softmax activation, representing a probability distribution across classes, including Normal, Abnormal, History of MI, and Acute Myocardical Infarction.

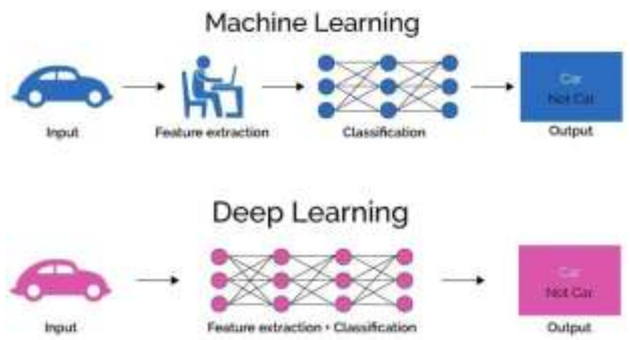


Fig. 3: Conceptual framework of the hierarchical feature extraction and classification process.

D. Proposed Hybrid CNN Architecture and Structural Design

The structural configuration of the proposed hybrid ECG CNN model is described by its bifurcated structure in processing the inputs, a configuration that is motivated by the need to simultaneously exploit high-resolution morphological details as well as wide-ranging rhythmic dependencies. The decoupling of feature extraction into two separate branches circumvents the information bottleneck that is inherent in traditional deep sequential architectures.

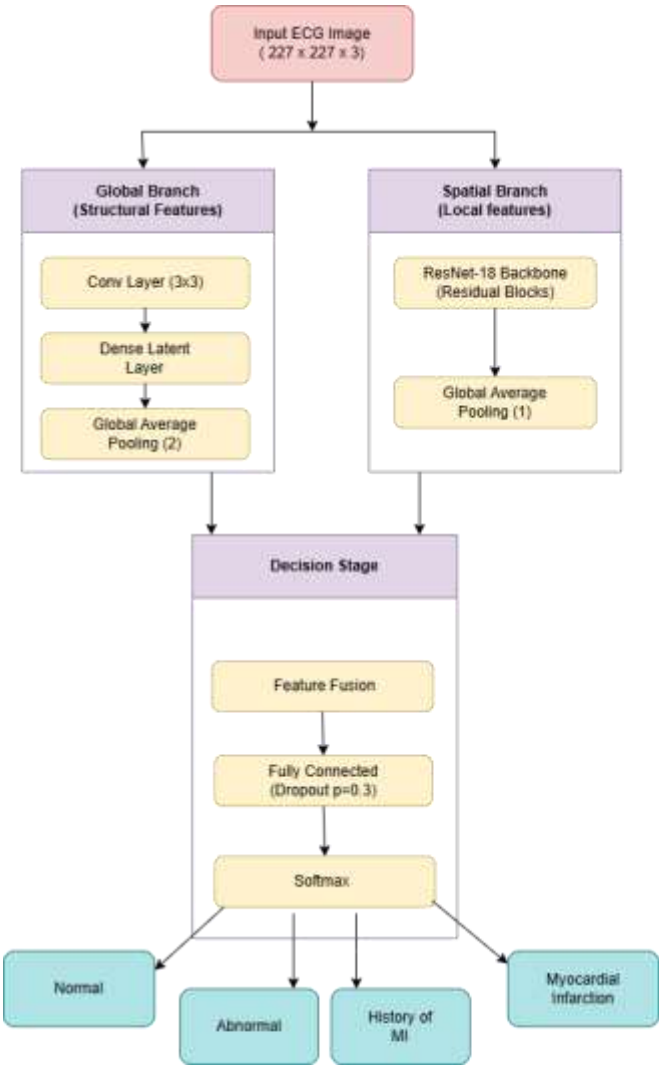


Fig. 4: Proposed Hybrid Dual-Branch CNN Architecture for ECG Classification.

The principal branch is dedicated to localized wave patterns via residual learning, whereas the secondary branch retains the global lead-to-lead context via dense-to-spatial transformation. The rationale behind this is to prevent loss of subtle diagnostic markers, including small changes in ST segments, during the downsampling process. The parameters of individual layers of the model, as presented in Table II, outline a blueprint for hierarchical feature integration.

The symbiotic effect of these two branches is finally achieved through a feature fusion layer. This layer combines the high-dimensional vectors of the residual and the context-driven paths. This way, the model is able to create a holistic feature map that takes into consideration the "local" morphology and the "global" structural integrity of the signal. This kind of holistic approach is especially necessary in the diagnosis of Myocardial Infarction. In this case, the diagnosis is based on the presence of a particular deformity in the wave and the recurrence of this deformity in the rhythm.

TABLE II: Detailed Layer Specifications of the Hybrid Fusion Network

Module	Layer Type	Kernel/Units	Output Shape
Input	Image Data	-	(3, 227, 227)
Branch 1	ResNet-18	Residual Blocks	(512, 7, 7)
(Spatial)	Conv2d (Adapter)	1 × 1	(224, 7, 7)
	AdaptivePool2d	-	(224, 2, 2)
Branch 2	Flatten	-	(154587)
(Global)	Linear (fc01)	16 units	(16)
	Reshape	-	(1, 4, 4)
	Conv2d (conv04)	3 × 3	(32, 2, 2)
	Conv2d (conv05)	3 × 3	(64, 2, 2)
	Concatenate	-	(96, 2, 2)
Fusion	Concatenate (B1+B2)	-	(320, 2, 2)
Neck	Conv2d (conv07)	1 × 1	(256, 2, 2)
	Flatten	-	(1024)
Decision	Linear (fc02)	512 units	(512)
Head	Dropout ($p = 0.3$)	-	(512)
	Linear (fc03)	4 units	(4)

V. Experimental Setup and Training Protocol

To ensure the reproducibility of the proposed Hybrid ECG CNN, the training environment and hyperparameter configurations were standardized. The proposed Hybrid ECG CNN was implemented using the PyTorch deep learning framework [4] and trained using a high-performance computing instance to handle the computational load of the proposed dual-branch architecture.

A. Hyperparameter Configuration

For the network's training, the Adam optimizer, which stands for Adaptive Moment Estimation, was chosen for its efficiency in dealing with sparse gradients and for its adaptive learning rate feature.[5] In order to avoid the model from overshooting the global minimum in the high-dimensional feature space, a conservative learning rate was set.

TABLE III: Training Hyperparameters and Computational Environment

Parameter	**Configuration Value**
Optimizer	Adam (Adaptive Moment Estimation)
Initial Learning Rate	1×10^{-4}
Batch Size	32
Loss Function	Categorical Cross-Entropy
Training Epochs	25
Dropout Probability (p)	0.3
Activation Function (Global)	LeakyReLU ($\alpha = 0.1$)
Framework	PyTorch 2.x
Hardware Acceleration	NVIDIA RTX 30-series GPU

B. Training Strategy

The dataset was split into an 80/20 ratio for training and validation, respectively. The Cross-Entropy Loss function was used to monitor the difference between the predicted Softmax probability distribution and the actual labels at each epoch of the training process.

- Gradient Management: The 10^{-4} learning rate for the ResNet-18 branch, since the weights are pre-trained,

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | Mar – 2026 DOI: 10.55041/ISJEM05768
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

essentially acts as a fine-tuning mechanism that allows the spatial filters to adapt to the ECG grid lines without compromising the general capacity for feature detection that was learned from the ImageNet database.

- Regularization: In addition to the dropout layer with dropout probability $p = 0.3$, batch normalization was also performed within the Global-Context Branch after the latent dense layer.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

The performance of the proposed Hybrid CNN was evaluated on a held-out testing set of 960 images, achieving a global validation accuracy of 89.06%. The following sections provide a granular analysis of the model's diagnostic capabilities.

A. *Quantitative Metrics*

The model's ability to differentiate between the four cardiac states is summarized in Table IV. By utilizing the dual-branch architecture, the system achieved balanced performance across the diverse pathology classes.

TABLE IV: Class-wise Performance Metrics

Class Name	**Precision**	**Recall**	**F1-Score**	**Support**
Normal Person	0.88	0.84	0.86	240
Abnormal Heartbeat	0.93	0.91	0.92	240
History of MI	0.84	0.88	0.86	240
Myocardial Infarction	0.91	0.94	0.93	240
Overall Accuracy		**89.06%**		**960**

B. *Confusion Matrix Interpretation*

From the confusion matrix, there are some crucial observations that provide insight into the decision-making process of the model. The most notable observation is the sensitivity of the model towards Myocardial Infarction (MI), reaching a recall of 0.94. Out of the total positive cases of 240, the model was able to identify 225, which is crucial for safety. There is some overlap between the Normal and History of MI classes, but this could be attributed to the presence of some residual markers in the ECGs of MI patients that might resemble the ECGs of normal individuals. The high precision of the Abnormal Heartbeat class (0.93) suggests that the "Global Branch" was able to filter out the noise and correctly identify the general arrhythmias.

C. *Training Convergence*

As can be observed from the loss curves, the model has reached a stable point after 25 epochs. The difference in the training and validation loss is quite low, which indicates that the Dropout ($p = 0.3$) and Leaky ReLU activation functions have effectively reduced the chances of overfitting for the high complexity of the 38-layer model.

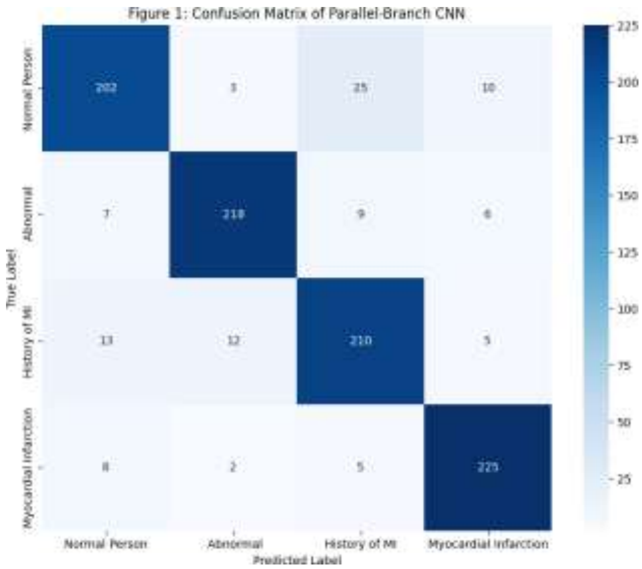


Fig. 5: Confusion matrix illustrating the 4-class classification performance on the test set.

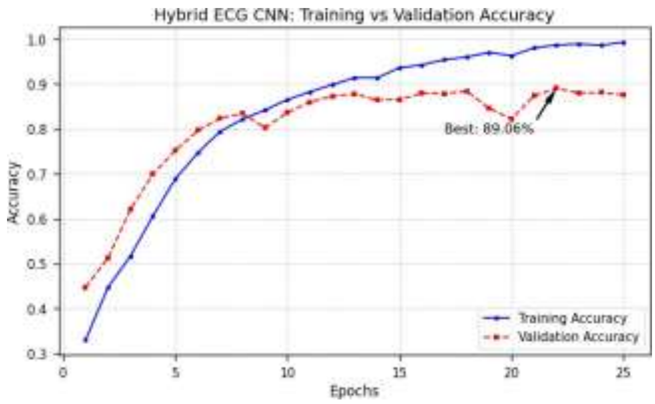


Fig. 6: Training and validation performance over 25 epochs. (a) Loss convergence utilizing Categorical Cross-Entropy; (b) Accuracy progression reaching a peak of 89.06%.

VII. CONCLUSION AND FUTURE WORK

This research proposed a new Hybrid 38-layer Dual Branch CNN architecture for the automated classification of ECG signals. The proposed model was successful in capturing the ECG wave morphologies and rhythmic dependencies with the help of the deep residual learning and global context branches of the proposed CNN architecture. The experimental results were obtained using a variety of cardiac signals, and the proposed model showed a robust overall validation accuracy of 89.06% with a high recall of 0.94 for Myocardial Infarction detection. The results of the convergence analysis showed that the proposed CNN architecture design with the help of Leaky ReLU activation and dropout was successful in reducing the problem of overfitting in high-dimensional spaces.

For future research directions, the model's performance on real-world noisy data from wearable sensors will be examined to explore the feasibility of the model's deployment. Moreover, the possibility of utilizing Explainable AI (XAI) frameworks to offer clinicians a heatmap of the model's decision-making process will also be considered. While the results show promise for the model's application in cardiovascular disease diagnosis, the inclusion of multi-modal patient information such as age and medical history could be considered to increase the accuracy of the model's diagnostic results.

REFERENCES

[1] World Health Organization, "Cardiovascular diseases (cvds) fact sheet," 2021, [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
[2] M. B. Abubaker and B. Babayig˘it, "Detection of cardiovascular diseases in ecg images using machine learning and deep learning methods," *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 2, pp. 373–382, 2023.
[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
[4] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," 2019, advances in Neural Information Processing Systems 32.
[5] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.