Inter-Patient Multi-Label Localization of Myocardial Infarction Territories Using 12-Lead ECG Signals

Ankita Shukla
Department of Computer Engineering, Aligarh Muslim University, India
shuklankita321@gmail.com (corresponding author)

Muhammed Izharuddin
Department of Computer Engineering, Aligarh Muslim University, India
izharuddinmuhammed67@gmail.com

Received: 30 April 2026 | Revised: 31 May 2026 | Accepted: 5 June 2026

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.19707*

ABSTRACT

The localization of Myocardial Infarction (MI) from Electrocardiogram (ECG) signals remains challenging due to inter-patient variability, class imbalance, and limited interpretability of current methods. This study presents a one-dimensional convolutional network, MobileECGNet, for multi-label, territory-wise MI localization using 12-lead ECG signals. The architecture incorporates depthwise separable convolutions, squeeze-and-excitation blocks, and temporal attention to capture inter-lead dependencies and temporal waveform characteristics. Experiments were conducted on the PTB-XL dataset using an inter-patient evaluation protocol with stratified splitting to preserve label distribution. ECG signals were standardized to fixed-length segments to ensure consistent input representation. The proposed method achieved a macro-AUC of 0.93, indicating strong localization performance. In addition, to analyze lead-wise contributions, a SHAP-based explainable AI technique was used, which exhibits alignment with clinically relevant ECG leads. Overall, the results demonstrate that the proposed framework provides an effective and interpretable solution for multi-label MI localization problems.

Keywords-deep learning; myocardial infarction localization; ECG analysis; interpretability

I. INTRODUCTION

MI ranks among the most common causes of mortality worldwide; therefore, obtaining adequate early diagnosis and accurate identification of the target area of the heart is essential for effective management [1, 2]. The 12-lead Electrocardiogram (ECG) is the main non-invasive diagnostic modality, revealing spatial information on heart function. Specific deviations in the ST segment, T-wave inversions, and pathological Q waves are associated with different areas of infarction and are commonly applied in clinical assessments [3, 4].

The exploitation of high-dimensional data, such as PTB-XL [5], has contributed to the development of automated ECG classification by deep learning methodologies. However, most studies to date have concentrated on MI detection or generalized classification of ECG without focusing on proper localization. For instance, in [6], cardiologist-level accuracy was reached for arrhythmia detection, and in [7], large-scale models were trained with multi-diagnostic capacity. Although effective, these approaches are mainly used for diagnostic purposes and not for anatomical localization.

To go beyond detection, in recent years, multi-label ECG classification has been investigated. In [8], it was shown that Convolutional Neural Networks (CNNs) can extract meaningful representations from the PTB-XL dataset. However, many methods, such as [9], have developed hybrid deep learning models for classification tasks, but did not address MI localization. ML-Net [10] is a lightweight multi-channel convolutional framework for automated myocardial infarction detection using 12-lead ECG signals. ML-ResNet [11] is based on residual learning for the detection and localization of myocardial infarction from multi-lead ECG recordings. In [12], a multi-scale ResNet with an attention mechanism was proposed to improve MI localization by enhancing the representation of discriminative features from ECG signals. In addition, evaluation strategies often rely on intra-patient splits, which may lead to optimistic performance due to data leakage, whereas inter-patient evaluation provides a more realistic but challenging assessment. Furthermore, increasing model complexity may hinder deployment in real-time or resource-constrained clinical settings. To address these limitations, this work proposes an attention-based convolutional framework for MI localization.

The main contributions of this work are:

- Modeling of co-occurring infarct territories using a multi-label formulation from 12-lead ECG signals.
- Integration of depthwise separable convolutions with channel and temporal attention for effective spatial–temporal feature learning.
- A stratified inter-patient evaluation protocol with SHAP-based analysis, ensuring reliable generalization and interpretability.

II. METHOD

A. Problem Formulation

MI localization is formulated as a multi-label classification problem, in which each ECG recording may correspond to one or more anatomical territories. Given a 12-lead ECG signal $X \in \mathbb{R}^{12\times T}$, the objective is to predict the presence of MI across four regions (anterior, inferior, lateral, and posterior). The ground truth label vector y is defined in (1) and represents the presence or absence of MI in each territory:

$$y = [y_1, y_2, y_3, y_4] \in \{0,1\}^4 \quad (1)$$

where $y_i \in \{0,1\}$ indicates the label for class i and the model learns a mapping function as expressed in (2), from the input signal to class-wise outputs:

$$f: \mathbb{R}^{12\times T} \rightarrow [0,1]^4 \quad (2)$$

Probability outputs are generated for each class as:

$$\hat{y}_i = \sigma(z_i), i \in \{1,2,3,4\} \quad (3)$$

here, $\sigma(\cdot)$ is the sigmoid function, z_i denotes the raw output (logit) for class i, and $\hat{y}_i \in [0,1]$ represents the predicted probability. Applying a threshold to $\hat{y}$ results in multi-label output predictions.

B. Dataset and Inter-Patient Evaluation

For the localization of MI, the PTB-XL dataset [5] was used, with its fundamental characteristics summarized in Table I. The SCP-ECG statements were coded and annotated with clinically relevant anatomical territories (anterior, inferior, lateral, and posterior) and mapped in Table II. This mapping results in a multi-label formulation, where each ECG recording may correspond to one or more affected regions. To conduct an unbiased evaluation, an inter-patient evaluation protocol with patient-wise partitioning was applied. The dataset is defined as $D = \{(X_i, y_i, p_i)\}$, where X_i denotes the ECG signal, the corresponding multi-label annotation, and p_i is the patient identifier. The training and validation sets were constructed to ensure patient-wise separation as:

$$p_i \notin \{ P_{\text{train}} \cap P_{\text{val}} \} \quad (4)$$

In addition, stratified splitting was employed to preserve label distribution across sets, improving training stability under class imbalance.

TABLE I. PTB-XL DATASET SUMMARY

Attribute	Value
Number of recordings	21,837 ECG recordings
Number of patients	18,885 subjects
Sampling frequency	500 Hz
Recording duration	5 seconds
Annotation standard	SCP-ECG
MI subclasses used	AMI, ASMI, ALMI, IMI, ILMI, IPMI, LMI
Target territories	Anterior, Inferior, Lateral, Posterior
Evaluation protocol	Inter-patient (patient-disjoint)

TABLE II. SCP-ECG TO MI TERRITORY MAPPING

Territory	SCP-ECG Codes
Anterior	AMI, ASMI, ALMI
Inferior	IMI, ILMI, IPMI
Lateral	LMI, ALMI, ILMI
Posterior	PMI, IPMI

C. Signal Preprocessing

All ECG signals were first processed using z-score normalization [7], which reduces the inter-sample variability and stabilizes training. The normalized signal $\tilde{X}$ was computed as:

$$\tilde{X} = \frac{X-\mu}{\sigma+\epsilon} \quad (5)$$

where X denotes the input signal, μ and σ are its mean and standard deviation, and ϵ is a small constant for numerical stability.

D. Proposed Model

As illustrated in Figure 1, the proposed MobileECGNet processes a 12-lead ECG signal $X \in \mathbb{R}^{12\times T}$, where T denotes the number of temporal samples to perform multi-label MI localization. The network transforms the input into hierarchical feature representations, X_1, X_2, and X_{se}, followed by attention-based aggregation into a compact feature vector $f \in \mathbb{R}^{64}$, which is finally mapped to class probabilities $\hat{y} \in [0,1]^4$.

1) Convolutional Feature Extraction

The input ECG signal is first processed using depthwise separable convolution [13] as:

$$X_1 = \text{DWConv}_1(X),\ X_2 = \text{DWConv}_2(X_1) \quad (6)$$

where X represents the input ECG signal, and $X_1 \in \mathbb{R}^{32\times T}$ and $X_2 \in \mathbb{R}^{64\times T}$ denotes intermediate feature maps. Here, $\text{DWConv}(\cdot)$ denotes depthwise separable convolution, which decomposes standard convolution into channel-wise filtering and pointwise fusion. This design efficiently explores temporal patterns (e.g., the QRS complex and ST segment) while modeling cross-lead relationships, which are essential for MI detection.

2) Channel Attention (SE Block)

To model the lead-wise importance of each feature channel, a squeeze-and-excitation mechanism [14] is applied. A global descriptor is first computed by averaging each channel over time as:

$$s_c = \frac{1}{T}\sum_{t=1}^{T} X_2(c,t) \quad (7)$$

where s_c represents the average activation of the channel c, and the collection of all such values forms the descriptor vector $s \in \mathbb{R}^{64}$. This descriptor is then used to generate channel-wise weights through a nonlinear transformation:

$$w = \sigma(W_2 \cdot \mathrm{ReLU}(W_1 s)) \quad (8)$$

where W_1 and W_2 are learnable parameters, $w \in \mathbb{R}^{64}$ denotes channel weights, and $\sigma(\cdot)$ is the sigmoid function. These weights recalibrate the feature map via element-wise multiplication $X_{se} = X_2 \odot w$, allowing the network to emphasize diagnostically relevant leads (e.g., II, III, aVF for inferior MI) while suppressing less informative ones. A dropout layer with probability $p = 0.2$ is applied to improve generalization.

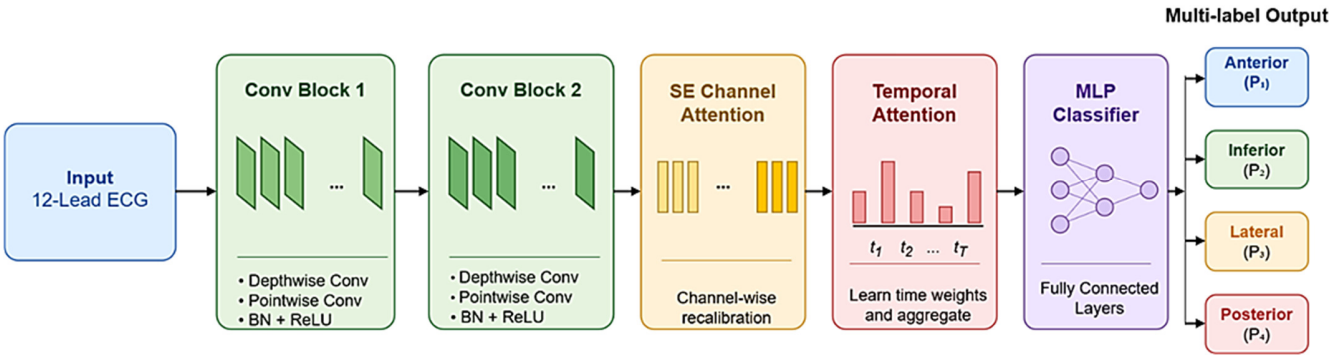


Fig. 1. Proposed MobileECGNet architecture for MI localization

3) Temporal Attention Mechanism

Temporal attention [15] is applied to focus on informative ECG segments. The attention weights are computed as:

$$\alpha = \mathrm{Softmax}(\mathrm{Conv1D}(X_{se})) \quad (9)$$

where $\alpha \in \mathbb{R}^T$ denotes attention weights over time. This allows the model to emphasize clinically significant regions such as ST-segment deviations and suppress irrelevant signal portions. These weights are then used to aggregate the feature map into a compact representation, as defined in (10). Here, $X_{se}(:,t)$ denotes the feature vector across all channels at the time step t, and α_t is the corresponding attention weight. This weighted summation highlights clinically relevant segments (e.g., ST abnormalities) while reducing noise, producing a fixed-dimensional feature vector $f \in \mathbb{R}^{64}$.

$$f = \sum_{t=1}^{T} \alpha_t X_{se}(:,t) \quad (10)$$

4) Feature Aggregation and Classification

The aggregated feature vector $f \in \mathbb{R}^{64}$ is passed through a fully connected classifier:

$$z = W_2 \cdot \mathrm{ReLU}(W_1 f) \quad (11)$$

$$\hat{y} = \sigma(z) \quad (12)$$

where $z \in \mathbb{R}^4$ represents logits and $\hat{y}$ denotes predicted probabilities for the four MI territories. The final predictions are obtained using sigmoid activation (12), which enables independent multi-label predictions, where $\hat{y} \in [0,1]^4$ denotes the predicted probabilities for each class. This formulation allows independent prediction of multiple infarct regions in a multi-label setting.

E. Loss Function

To address class imbalance in multi-label MI localization, Focal Binary Cross-Entropy (FBCE) loss [16] is used:

$$L = \frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{4} \alpha \, (1 - p_{ic})^{\gamma} \, l_{ic} \quad (13)$$

where $l_{ic} = \mathrm{BCEWithLogits}(z_{ic}, y_{ic})$ represents the classification loss for class c and sample i, $y_{ic} \in \{0,1\}$ is the ground truth, z_{ic} is the predicted logit, and $p_{ic} = \exp(-l_{ic})$ represents prediction confidence. The parameters α and γ control class balancing and focusing, respectively. Compared to standard binary cross-entropy, the modulating term $(1 - p_{ic})^{\gamma}$ reduces the contribution of well-classified samples and emphasizes misclassified or low-confidence samples. This is particularly useful in ECG-based MI localization, where certain territories are underrepresented. As a result, the loss improves robustness to class imbalance and promotes balanced learning across all classes.

F. Architectural Design Justification

As shown in Figure 1, the architecture progressively transforms the 12-lead input into compact feature representations. An initial depthwise convolution with $k = 7$, $s = 1$, $p = 3$ preserves temporal resolution and captures broader waveform context ($12 \rightarrow 12$), followed by a pointwise 1×1 convolution for cross-lead feature interaction ($12 \rightarrow 32$). A second depth-wise convolution with $k = 5$, $s = 1$, $p = 2$ captures localized temporal patterns, followed by expansion ($32 \rightarrow 64$) to enhance feature capacity.

A squeeze-and-excitation block ($r = 16$) models channel-wise importance, enabling adaptive weighting of clinically relevant leads, while dropout ($p = 0.2$) improves generalization. Temporal attention focuses on informative ECG segments and aggregates features into a fixed 64-dimensional vector.

Finally, fully connected layers ($64 \rightarrow 32 \rightarrow 4$) map features to MI territories. Overall, the combination of depthwise separable convolutions, channel attention, and temporal attention enables efficient modeling of temporal dynamics and inter-lead dependencies for inter-patient multi-label MI localization.

III. EXPERIMENTATION

A. Experimental Setup

The proposed MobileECGNet was evaluated on the PTB-XL dataset [5], following an inter-patient evaluation protocol [8]. Model training involves mini-batch gradient descent with shuffled batches to ensure stable convergence and improved generalization. Experiments were performed on an NVIDIA GPU (e.g., T4/RTX series) using Python with deep learning libraries (e.g., PyTorch). Supporting libraries, including NumPy, SciPy, scikit-learn, and Matplotlib, were used for data processing, evaluation, and visualization.

B. Hyperparameter Tuning and Configuration

The training configuration followed the settings summarized in Table III. The model was optimized using AdamW [17] with a learning rate of 10^{-3}, a weight decay of 10^{-4}, and a batch size of 64.

Early stopping based on validation AUC was employed to prevent overfitting. To address class imbalance, FBCE loss was used, where the parameters γ and α were tuned empirically on the validation set. Hyperparameters were selected based on macro-AUC and macro F1-score, while class-specific thresholds were optimized using validation predictions to improve multi-label performance.

TABLE III. MODEL HYPERPARAMETERS

Parameter	Value
Optimizer	AdamW
Learning rate	1×10^{-3}
Weight decay	1×10^{-4}
Batch size	64
Loss function	Focal BCE
Early stopping metric	AUC
Input length	2500 samples
Sampling frequency	500 Hz

C. Evaluation Metrics

Model performance was assessed using standard multi-label metrics (precision, recall, F1-score, and AUC) based on common evaluation practices [18]. Due to class imbalance across MI territories, macro-averaged metrics are primarily reported. In addition, AUC curves across training epochs are analyzed to assess the learning behavior of the model, showing rapid convergence followed by stable optimization.

IV. RESULTS

A. Territory-Wise Performance and Training Behavior Analysis

Table IV summarizes the performance by territory. The model demonstrates strong discriminative capability, with a validation AUC that settles at about 0.93, indicating that the MI patterns can effectively be separated between territories. The model's relatively high AUC indicates that it can appropriately rank positive and negative samples in terms of decision thresholds. This can be attributed to the joint influence of depth-wise convolutional layers and attention, with which both global waveform structures and local clinically relevant segments are captured, enhancing robustness to inter-patient variability.

TABLE IV. TERRITORY-WISE PERFORMANCE

Territory	Precision	Recall	F1-score
Anterior	0.89	0.91	0.90
Inferior	0.67	0.65	0.66
Lateral	0.20	0.09	0.12
Posterior	0.64	0.62	0.63

MI in the anterior region achieves maximum performance, with an F1 score of 0.90, indicating a more clearly defined and dependable ECG signature in this region. The proposed model captures these patterns very well with the hierarchical feature extraction and channel attention, focusing on relevant leads. Inferior and posterior territories display moderate F1 scores (0.66 and 0.63) due to inter-patient heterogeneity and overlapping ECG features in these areas. Despite the fact that temporal attention enables focusing on informative segments, the small discrepancies between these territories hinder accurate discrimination.

In contrast, the lateral territory scores are similarly lower (precision 0.20, recall 0.09, F1-score 0.12), which can be attributed mostly to severe class imbalance and fine morphological abnormalities. Such an imbalance limits the macro F1-score and illustrates the challenges of multi-label MI localization in real inter-patient settings.

The training behavior is illustrated in Figure 2, which shows the AUC curves over epochs. The validation AUC increases rapidly during the initial epochs and gradually stabilizes around 0.93 as training progresses, indicating effective learning. The close alignment between training and validation curves indicates good generalization performance, while the lack of a substantial gap shows stable optimization performance and no sign of overfitting.

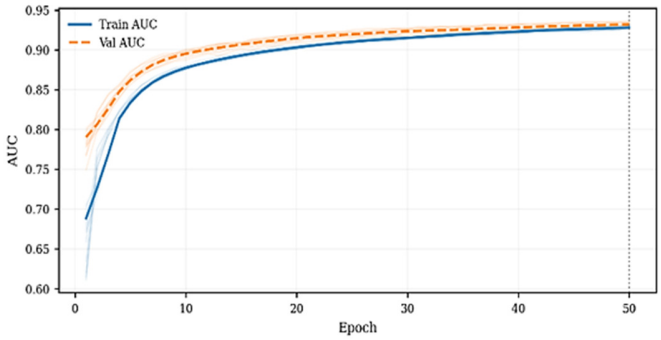


Fig. 2. AUC curve over training epochs.

B. Comparison with State-of-the-Art Methods

Table V presents a quantitative comparison with recent PTB-XL-based inter-patient MI localization methods. The proposed framework achieved a Macro-AUC of 0.9322 and a Macro-F1 score of 0.5800, outperforming MCA-net [19] and Spa-Tem MI [20] in both evaluation metrics. The performance values for MCA-net and Spa-Tem MI were obtained from their respective published studies under PTB-XL-based inter-patient evaluation settings. Since different studies employ variations in preprocessing strategies, localization granularity, class definitions, and experimental protocols, only the most comparable localization-oriented metrics were considered for reference. The improved performance of the proposed method demonstrates the effectiveness of the territory-wise multi-label learning framework for inter-patient localization of MI using 12-lead ECG signals.

TABLE V. COMPARISON WITH INTER-PATIENT MI LOCALIZATION METHODS ON PTB-XL

Model	Macro-AUC	Macro-F1
MCA-Net [19]	0.8316	0.5463
Spa-Tem MI [20]	0.8097	0.5263
Proposed Method	**0.9322**	**0.5800**

C. SHAP-Based Model Interpretability

Model interpretability was analyzed using SHAP [21], where Figure 3 presents the mean ± standard deviation (SD) of lead-wise importance values across the 12 standard ECG leads (I, II, III, aVR, aVL, aVF, V1–V6). The mean SHAP value reflects the average contribution of each lead to the prediction, while the SD indicates variability across samples.

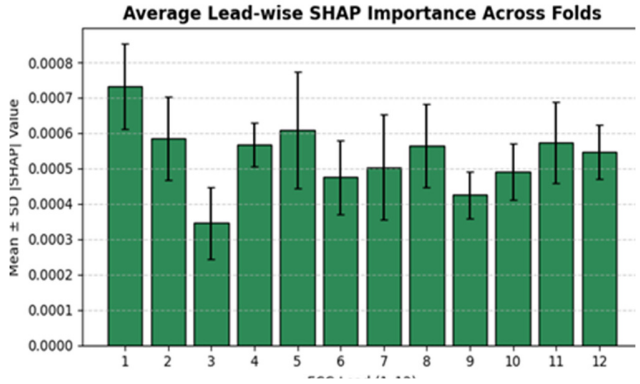


Fig. 3. SHAP-based interpretability of ECG leads.

As observed in Figure 3, Lead I shows the highest importance, followed by Leads V5, II, V2, and V6, further validating their strong influence on model predictions. This aligns with clinical knowledge, suggesting that such leads provide more clinically actionable insights concerning the anterior region and would yield improved performance for anterior MI. Other leads, such as III and V3, demonstrate relatively lower SHAP values, confirming lower contributions. These leads, however, contribute less to providing an approximation of distinct infarction patterns in this dataset and consequently have poorer performance on some territories.

Leads from the inferior and posterior area exhibit somewhat more modest SHAP values, indicating a more nuanced and overlapping ECG picture of these infarct locations. In addition, small error bars (SD) in most leads also reflect that the model provided consistent features, which reflect the robustness of model representations. Generally, stronger SHAP indicates more clinically useful leads, and moderate results describe complex or overlapped results, while low results imply lower discriminative contribution, and low scores indicate low variability (SD) in the models.

V. CONCLUSION

This study presented MobileECGNet, an attention-based deep learning framework for inter-patient multi-label MI localization using 12-lead ECG signals. Unlike several earlier studies that focused mainly on binary MI detection or coarse localization, this framework performs territory-wise multi-label localization under a strict inter-patient evaluation setting using the PTB-XL dataset. The architecture integrates depthwise separable convolutions with channel and temporal attention mechanisms to effectively capture inter-lead and temporal ECG characteristics while maintaining computational efficiency.

Experimental results demonstrated that the proposed method achieved a Macro-AUC of 0.9322 and a Macro-F1 score of 0.5800, showing improved performance compared to recent PTB-XL-based inter-patient MI localization methods reported in the literature. The model demonstrated strong performance for anterior MI localization, while moderate results in inferior and posterior territories reflected challenges associated with inter-patient variability and overlapping ECG patterns. In addition, SHAP-based interpretability analysis showed strong alignment with clinically relevant ECG leads, improving model transparency and clinical relevance.

Overall, the proposed framework contributes an interpretable and computationally efficient solution for territory-wise MI localization under realistic inter-patient conditions. Future work will focus on improving localization performance for underrepresented infarction territories through advanced imbalance-handling strategies and multi-scale lead interaction modeling. Further validation on larger multi-modal clinical datasets and real-time clinical deployment scenarios will also be explored to improve robustness and practical applicability.

DECLARATION OF COMPETING INTERESTS

The authors declare no conflicting interests that could have influenced the results of this study.

ACKNOWLEDGMENT

The authors acknowledge the PTB-XL dataset providers.

DATA AVAILABILITY

The dataset used in this study is available at [22].

DECLARATION OF GENERATIVE AI USE

Generative AI tools were used only for language refinement and editing assistance. All scientific content and conclusions were verified by the authors.

REFERENCES

[1] "Leading causes of death," *World Health Organization*. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death.

[2] D. P. Zipes, P. Libby, R. O. Bonow, D. L. Mann, G. F. Tomaselli, and E. Braunwald, Eds., *Braunwald's heart disease: a textbook of cardiovascular medicine*, 11th ed. Elsevier, 2019.

[3] G. S. Wagner and D. G. Strauss, *Marriott's practical electrocardiography*, 12th ed. Lippincott Williams & Wilkins, 2014.

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

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

[6] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, pp. 65–69, Jan. 2019, https://doi.org/10.1038/s41591-018-0268-3.

[7] A. H. Ribeiro *et al.*, "Automatic diagnosis of the 12-lead ECG using a deep neural network," *Nature Communications*, vol. 11, no. 1, Apr. 2020, Art. no. 1760, https://doi.org/10.1038/s41467-020-15432-4.

[8] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1519–1528, May 2021, https://doi.org/10.1109/JBHI.2020.3022989.

[9] K. N. Chaithra, Neelappa, M. T. Sreedevi, and R. Asha, "Hybrid Deep Learning Models for Accurate ECG Classification in Cardiovascular Disease Diagnosis," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 26683–26688, Oct. 2025, https://doi.org/10.48084/etasr.12209.

[10] Y. Cao *et al.*, "ML-Net: Multi-Channel Lightweight Network for Detecting Myocardial Infarction," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3721–3731, July 2021, https://doi.org/10.1109/JBHI.2021.3060433.

[11] C. Han and L. Shi, "ML–ResNet: A novel network to detect and locate myocardial infarction using 12 leads ECG," *Computer Methods and Programs in Biomedicine*, vol. 185, Mar. 2020, Art. no. 105138, https://doi.org/10.1016/j.cmpb.2019.105138.

[12] Y. Cao *et al.*, "Detection and Localization of Myocardial Infarction Based on Multi-Scale ResNet and Attention Mechanism," *Frontiers in Physiology*, vol. 13, Jan. 2022, https://doi.org/10.3389/fphys.2022.783184.

[13] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." arXiv, Apr. 17, 2017, https://doi.org/10.48550/arXiv.1704.04861.

[14] S. Hooda and R. K. Tripathy, "SERN-AwGOP: Squeeze-and-Excitation Residual Network With an Attention-Weighted Generalized Operational Perceptron for Atrial Fibrillation Detection," *IEEE Access*, vol. 13, pp. 34844–34853, 2025, https://doi.org/10.1109/ACCESS.2025.3544493.

[15] S. Batool, M. Ghafoor, I. A. Taj, and W. Abdul, "A hybrid spatial and temporal attention driven network for left ventricular function assessment using echocardiography," *Scientific Reports*, vol. 15, no. 1, Dec. 2025, Art. no. 43029, https://doi.org/10.1038/s41598-025-21486-5.

[16] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, Oct. 2020, https://doi.org/10.1109/TPAMI.2018.2858826.

[17] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization." arXiv, Jan. 04, 2019, https://doi.org/10.48550/arXiv.1711.05101.

[18] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, July 2009, https://doi.org/10.1016/j.ipm.2009.03.002.

[19] W. Pan, Y. An, Y. Guan, and J. Wang, "MCA-net: A multi-task channel attention network for Myocardial infarction detection and location using 12-lead ECGs," *Computers in Biology and Medicine*, vol. 150, Nov. 2022, Art. no. 106199, https://doi.org/10.1016/j.compbiomed.2022.106199.

[20] J. Yu, J. Gao, N. Wang, P. Feng, B. Zhou, and Z. Wang, "Spa-Tem MI: A Spatial–Temporal Network for Detecting and Locating Myocardial Infarction," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023, https://doi.org/10.1109/TIM.2023.3301047.

[21] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, Long Beach, CA, USA, 2017, vol. 30.

[22] P. Wagner, N. Strodthoff, R.-D. Bousseljot, W. Samek, and T. Schaeffter, "PTB-XL, a large publicly available electrocardiography dataset." PhysioNet, https://doi.org/10.13026/6SEC-A640.