Interpretable Deep Learning for Prehospital ST-Elevation Myocardial Infarction (STEMI) Detection from 12-Lead ECG

Junmo An[1], Richard E Gregg[1], Dillon J Dzikowicz[1], Ben Bailey[1]

[1]Philips, Cambridge, Massachusetts, USA

Abstract

Rapid and accurate detection of ST-elevation myocardial infarction (STEMI) in prehospital settings is essential for timely triage and reperfusion therapy. We developed an end-to-end Multi-Scale Hybrid Transformer (MS-Hybrid Transformer) to classify study-defined STEMI from 10-second 12-lead ECGs and evaluated its generalizability on an independent prehospital dataset. Development used hospital-acquired ECGs from Long Beach Medical Center (259 STEMI, 571 No-STEMI) with ECG-specific randomized data augmentation and patient-independent 10-fold cross-validation. Models from the 10 folds were ensembled and externally evaluated on prehospital ECGs from Carolinas HealthCare System (147 STEMI, 341 No-STEMI). The ensemble achieved a sensitivity of 0.755, specificity of 0.833, precision of 0.661, F1-score of 0.705, accuracy of 0.809, AUROC of 0.853, and AUPRC of 0.732. Gradient-weighted Class Activation Mapping (Grad-CAM) provided temporally localized explanations of model predictions, with beat-averaged visualizations highlighting clinically relevant ST-T waveform regions across multiple leads. These findings demonstrate good discrimination on an independent prehospital cohort and support accurate and interpretable STEMI detection using the MS-Hybrid Transformer.

1. Introduction

ST-elevation myocardial infarction (STEMI) is a time-critical cardiovascular emergency requiring rapid diagnosis and timely reperfusion, most commonly through primary percutaneous coronary intervention (PCI) in the cardiac catheterization laboratory [1]. Early identification of STEMI in prehospital settings, such as at the scene of an emergency, in an ambulance, or during transport by emergency medical services (EMS), can facilitate early activation of the cardiac catheterization team and reduce delays to reperfusion therapy [2]. In these settings, the standard 12-lead ECG is the primary noninvasive tool for identifying ST-segment elevation patterns that may trigger evaluation for acute coronary occlusion and reperfusion therapy.

Automated ECG interpretation can support rapid clinical decision-making, particularly when expert interpretation is not immediately available. However, reliable automated STEMI detection remains challenging, with previous prehospital studies demonstrating substantial variation in the diagnostic performance of computerized ECG interpretation algorithms [3]. ST-segment and T-wave morphology varies substantially across patients, while baseline abnormalities, conduction disturbances, electrode placement, and noise may mimic or obscure ischemic changes. These challenges may be further amplified in prehospital environments, where ECG acquisition conditions are less controlled.

Deep learning (DL) has been applied to automated 12-lead ECG analysis across adult and pediatric populations, including arrhythmia, abnormality, and age-group classification [4], [5], [6], [7]. Convolutional neural networks (CNNs) are effective for extracting local morphological patterns such as QRS complexes and ST-T changes. Nevertheless, conventional CNN architecture primarily emphasizes local temporal features and may be less effective at representing longer-range relationships distributed throughout a multi-lead ECG. Transformer architectures complement convolutional processing by using self-attention to model global temporal dependencies. Hybrid architectures combining CNNs and Transformers have shown promise for detecting myocardial ischemia, Chagas disease, and Brugada syndrome from 12-lead ECGs by integrating localized morphological features with broader temporal dependencies [8], [9], [10].

Despite their strong predictive performance, DL models often lack transparency, making it difficult to determine which ECG regions drive individual predictions. Previous ECG studies have addressed this limitation using SHapley Additive exPlanations (SHAP) to identify influential waveform regions [5], [7]. In this study, we used beat-averaged Gradient-weighted Class Activation Mapping (Grad-CAM) as a complementary approach to provide class-specific, temporally localized attribution of salient ECG regions. For STEMI detection, this permits qualitative inspection of whether post-hoc attribution maps overlap with clinically relevant waveform components, including ST-segment and T-wave morphology.

In this study, we developed an end-to-end Multi-Scale Hybrid (MS-Hybrid) Transformer for binary STEMI

detection from 10-second 12-lead ECGs. The model was trained and internally evaluated using hospital-acquired ECGs and subsequently tested on an independent prehospital STEMI dataset. ECG-specific randomized data augmentation was incorporated to improve robustness to signal variability, and Grad-CAM with beat-level aggregation was used to characterize temporal attribution of model predictions.

2. Methods

2.1. Study Data

The model development dataset consisted of 10-second 12-lead ECG recordings acquired with 500 Hz at Long Beach Medical Center (Long Beach, CA, USA). The dataset included 259 STEMI and 571 No-STEMI ECGs. External evaluation was performed using an independent prehospital ECG dataset obtained with 500 Hz from a regional STEMI system of care at Carolinas HealthCare System (Charlotte, NC, USA). The external dataset contained 147 STEMI and 341 No-STEMI ECGs. The two datasets were maintained independently, with the prehospital dataset used only for final external evaluation. This design enabled assessment of model generalization from hospital-acquired ECG data to ECGs collected in the prehospital environment. The classification task was formulated as binary discrimination between STEMI and No-STEMI.

2.2. Data Augmentation

ECG-specific randomized data augmentation was applied exclusively during model training to increase waveform diversity and improve robustness to variations encountered in clinical ECG acquisition. The augmentation strategy included temporal shifting to simulate signal alignment variability, amplitude scaling to account for variations in signal magnitude, baseline wander to approximate low frequency physiological and acquisition related drift, random point dropout to simulate transient signal degradation, additive Gaussian noise to represent electronic and environmental noise, and cutout to randomly mask a contiguous temporal segment of the ECG and improve robustness to localized signal loss. Representative examples of these augmentation techniques are shown in Figure 1. These transformations were designed to expose the model to physiological and technical variations while preserving diagnostically relevant morphology. Augmentation parameters were randomly sampled during training so that each ECG could be presented with combinations of transformations. Validation and test ECGs were not augmented. This augmentation strategy follows the RandAugment based approach used for the STEMI model development, including time shifting, scaling, baseline wander, random point dropout, Gaussian noise, and cutout [4].

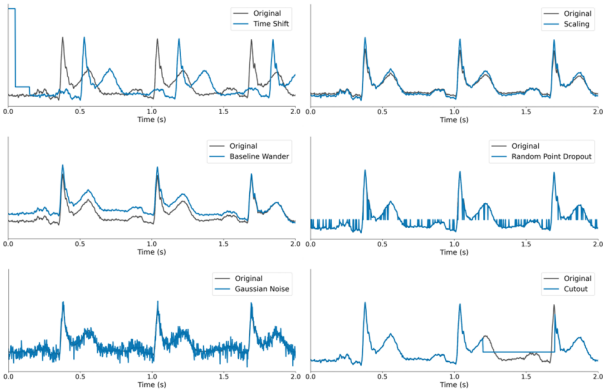

Figure 1. Data augmentation techniques incorporating time shifting, scaling, baseline wander, random point dropout, Gaussian noise, and cutout applied to lead II.

2.3. Multi-Scale Hybrid Transformer

Figure 2 illustrates the architecture of the proposed MS-Hybrid Transformer for STEMI detection. The model begins with two parallel one-dimensional convolutional (Conv1D) branches using wide (31-sample) and narrow (7-sample) kernels to capture ECG patterns at different temporal scales. Each branch is followed by batch normalization and rectified linear unit (ReLU) activation. Their outputs are concatenated and downsampled by max pooling. A CNN reducer then applies two additional Conv1D layers with batch normalization, ReLU activation, dropout, and temporal downsampling, yielding a sequence of 20 128-dimensional patch embeddings. A learnable classification (CLS) token is prepended, and learnable positional embeddings and dropout are applied before processing by four Transformer encoder layers with eight self-attention heads and a multilayer perceptron (MLP) dimension of 256. The final 128-dimensional CLS representation is passed through layer normalization and a linear classification layer to produce STEMI and No-STEMI logits. The model was trained using class-weighted cross-entropy loss and optimized with the AdamW optimizer (learning rate = 1×10^{-4}, weight decay = 1×10^{-2}), with dropout and ECG data augmentation used for regularization.

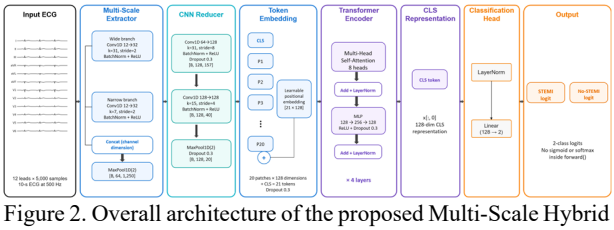

Figure 2. Overall architecture of the proposed Multi-Scale Hybrid Transformer for STEMI detection.

2.4. Model Training and Evaluation

The development dataset was partitioned into 10 patient-independent folds. Across 10 training loops, eight folds were used for training and two for validation, with the split rotated across loops. The independent Carolinas prehospital dataset was reserved as a holdout set and was not used for model development. One model was obtained from each training loop, and the 10 resulting models were ensembled for final external evaluation. Performance was assessed using sensitivity, specificity, precision (positive predictive value, PPV), F1-score, accuracy, area under the receiver operating characteristic curve (AUROC), and area under the precision-recall curve (AUPRC). All model training and explainability analyses were performed on a desktop workstation equipped with an Intel® Core™ i7-7700 CPU (3.60 GHz), 64 GB of RAM, and an NVIDIA GeForce RTX 3080 Ti GPU.

2.5. Explainability

Gradient-weighted Class Activation Mapping (Grad-CAM) was used to generate class-specific, temporally localized attribution maps for STEMI predictions. Grad-CAM was applied to the final Conv1D layer of the CNN reducer, where gradients of the target-class output were used to weight convolutional feature maps and interpolate them to the temporal resolution of the original ECG.

To reduce beat-to-beat variability, R-peaks were detected from lead II using a minimum inter-peak interval of 500 ms and an amplitude threshold of the signal mean plus 0.3 times its standard deviation. For each R-peak, 1-s windows spanning 400 ms before to 600 ms after the peak were extracted from both the ECG and Grad-CAM map. Valid windows were aligned and averaged point-by-point within each recording to obtain a representative mean beat and beat-averaged Grad-CAM map. The resulting maps were examined for emphasis on clinically relevant regions, particularly the ST segment and T wave. Grad-CAM was used to characterize temporal attribution rather than lead-specific importance or causal physiological relationships.

3. Results

The MS-Hybrid Transformer was evaluated on the independent prehospital dataset using an ensemble of models obtained from the 10 training folds. As summarized in Table 1, the model achieved a sensitivity of 0.755, specificity of 0.833, precision of 0.661, F1-score of 0.705, accuracy of 0.809, AUROC of 0.853, and AUPRC of 0.732. These findings indicate good discrimination on the independent prehospital cohort, with relatively balanced sensitivity and specificity despite differences between the development and external clinical datasets.

Table 1. Performance of the MS-Hybrid Transformer for STEMI detection on the independent prehospital test dataset.

Metric	Performance
Sensitivity	0.755
Specificity	0.833
Precision (PPV)	0.661
F1-score	0.705
AUROC	0.853
AUPRC	0.732
Accuracy	0.809

As shown in Figure 3, Grad-CAM was used to examine temporal regions contributing to STEMI predictions. The representative correctly classified STEMI ECG demonstrated localized model attribution within clinically relevant ST-T waveform regions. Thin lines represent individual R-peak-aligned beats, while the thick line represents the averaged beat, reducing transient variability and highlighting recurrent waveform patterns. Grad-CAM importance is indicated by color, with red representing higher attribution and blue representing lower attribution. Increased attribution was observed around the ST-segment and T-wave portions of the cardiac cycle across multiple leads, suggesting that the classifier utilized temporally localized waveform characteristics associated with STEMI rather than nonspecific global ECG features.

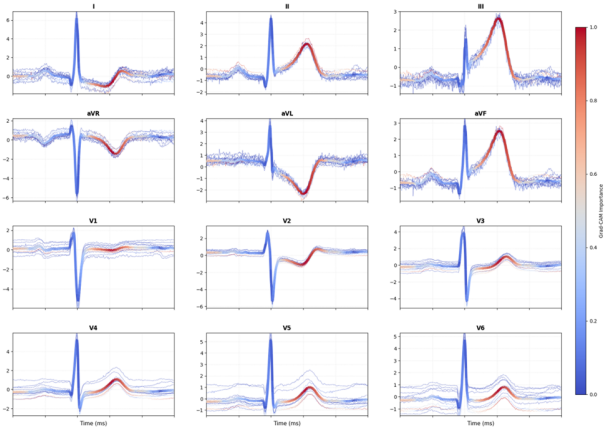

Figure 3. Beat-averaged Grad-CAM visualization for a representative correctly classified STEMI ECG across all 12 leads, showing temporal attribution of waveform regions associated with the model prediction.

4. Discussion and Future Work

This study developed and externally evaluated an interpretable MS-Hybrid Transformer for STEMI detection from 10-second 12-lead ECGs. The architecture combines multi-scale convolutional feature extraction with Transformer-based sequence modeling to capture both local morphology and temporal relationships.

A key strength of the study is evaluation on an independent prehospital ECG dataset, where acquisition conditions may differ from those of hospital-based development data. This provides a more rigorous assessment of model generalization than internal validation alone. The multi-scale convolutional front-end captures ECG features at different temporal resolutions, while the Transformer models broader temporal context. Randomized ECG augmentation was further used to improve robustness to variations in signal acquisition.

Beat-averaged Grad-CAM analysis frequently emphasized temporal regions corresponding to ST-T waveform morphology in representative STEMI ECGs, consistent with clinically relevant ischemic patterns. However, the Grad-CAM implementation provides temporal rather than lead-specific attribution and should not be interpreted as evidence of causal physiological relationships.

Several limitations should be considered. First, the study included data from a limited number of clinical sources, and the external evaluation was performed using a single prehospital cohort. Therefore, broader validation across geographically and demographically diverse EMS systems and institutions is needed to establish generalizability. Second, STEMI is clinically heterogeneous, and ECG morphology may vary according to culprit artery, infarct location, baseline conduction abnormalities, and timing of ECG acquisition relative to coronary occlusion. Performance was not evaluated separately across these clinically important subgroups. Third, the study was retrospective and did not assess the effect of model predictions on real-time prehospital triage, catheterization laboratory activation, or treatment delays. Finally, Grad-CAM provides qualitative temporal attribution and does not establish causal relationships or quantitative lead importance.

Future work will focus on larger multi-center validation, subgroup-specific performance analysis, and prospective evaluation in real-world prehospital workflows. Additional studies should also assess model robustness to diverse acquisition conditions and artifacts encountered in EMS environments and evaluate whether explainability outputs can improve clinician confidence and support timely STEMI decision-making.

References

[1] S. V. Rao *et al.*, "2025 ACC/AHA/ACEP/NAEMSP/SCAI guideline for the management of patients with acute coronary syndromes: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines," *Circulation*, vol. 151, no. 13, pp. e771–e862, 2025.

[2] J. S. Shavadia *et al.*, "Association between cardiac catheterization laboratory pre-activation and reperfusion timing metrics and outcomes in patients with ST-segment elevation myocardial infarction undergoing primary percutaneous coronary intervention: a report from the ACTION Registry," *JACC Cardiovasc. Interv.*, vol. 11, no. 18, pp. 1837–1847, 2018.

[3] J. L. Garvey, J. Zegre-Hemsey, R. Gregg, and J. R. Studnek, "Electrocardiographic diagnosis of ST segment elevation myocardial infarction: an evaluation of three automated interpretation algorithms," *J. Electrocardiol.*, vol. 49, no. 5, pp. 728–732, 2016.

[4] J. An, R. E. Gregg, and S. Borhani, "Effective data augmentation, filters, and automation techniques for automatic 12-lead ECG classification using deep residual neural networks," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2022, pp. 1283–1287.

[5] J. An, B. Bailey, Y.-H. Zhang, and D. J. Dzikowicz, "Left bundle branch block detection in 12-Lead ECG using End-to-End deep learning with explainability," 2024, pp. 1–4.

[6] J. An, B. Bailey, and R. E. Gregg, "Paradigm Shift from Feature-Based Machine Learning to End-to-End Deep Residual Neural Networks for Pediatric Age Classification from 12-Lead ECG," in *2023 Computing in Cardiology (CinC)*, IEEE, 2023, pp. 1–4.

[7] J. An *et al.*, "Automated Pediatric Ventricular Pre-Excitation (Wolff-Parkinson-White Syndrome) Detection in 12-Lead ECG Using Deep Learning and SHAP-Based Interpretability," in *2025 Computing in Cardiology (CinC)*, 2025, pp. 1–4.

[8] J. An, R. E. Gregg, H. Lee, D. J. Dzikowicz, and B. Bailey, "Multi-Scale Hybrid Transformer for Detecting Transient Myocardial Ischemia Using 10-Second 12-Lead STAFF III ECGs," in *2026 48th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2026, pp. 4436–4441.

[9] H. Lee and J. An, "Hybrid Convolution-Transformer Network Integrating Time and Frequency Domains from 12-Lead ECG for Chagas Disease Detection," 2025, pp. 1–4.

[10] J. An, Z. Li, D. J. Dzikowicz, B. Bailey, and M. Silka, "Brugada Detection from 12-Lead ECG Using a Multi-Scale Hybrid Transformer with Explainability," in *2026 Computing in Cardiology (CinC)*, 2026, pp. 1–4.

Address for correspondence:

Junmo An, PhD
222 Jacobs St,
Cambridge, MA 02141, USA
junmo.an@philips.com