Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Precise ECG Diagnosis and Validation of Educational Utility for Acute Myocardial Infarction Using Deep Learning and Explainable Artificial Intelligence

Jongkwang Kim
Kyungpook National University
Byungeun Shon
Kyungpook National University Hospital
Yongjin Kim
Kyungpook National University
Nayeon Kim
Kyungpook National University
Yejin Jang
Kyungpook National University
Jeongmin Lee
Kyungpook National University
Seong-Mi Yoon
Kyungpook National University
Dong Heon Yang
Kyungpook National University Hospital
Namkyun Kim
Kyungpook National University Hospital
Sungmoon Jeong
jeongsm00@gmail.com
Kyungpook National University

Article

Keywords: Artificial intelligence, Electrocardiograms, ST-segment elevation myocardial infarction, non-ST-elevation myocardial infarction, Explainable AI

Posted Date: April 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8217685/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on July 4th, 2026. See the published version at https://doi.org/10.1038/s41598-026-58956-3.

Precise ECG Diagnosis and Validation of Educational Utility for Acute Myocardial Infarction Using Deep Learning and Explainable Artificial Intelligence

Jongkwang Kim[1,2†], Byungeun Shon[1,2†], Yongjin Kim[3], Nayeon Kim[3], Yejin Jang[3], Jeongmin Lee[3], Seong-Mi Yoon[4,5], Dong Heon Yang[6,7], Namkyun Kim[6,7], Sungmoon Jeong[1,2*]

[1]Department of Medical Informatics, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

[2]Research Center for AI in Medicine, Kyungpook National University Hospital, Daegu, Republic of Korea

[3]College of Medicine, Kyungpook National University, Daegu, Republic of Korea

[4]Department of Forensic Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

[5]Preclinical Research Center, Daegu-Gyeongbuk Medical Innovation Foundation (K-MEDI hub), Daegu, Republic of Korea

[6]Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

[7]Division of Cardiology, Department of Internal Medicine, Kyungpook National University Hospital, Daegu, Republic of Korea

†These authors have contributed equally to this work and share first authorship

*** Correspondence:**
Sungmoon Jeong

jeongsm00@gmail.com

Keywords: Artificial intelligence, Electrocardiograms, ST-segment elevation myocardial infarction, non-ST-elevation myocardial infarction, Explainable AI

Abstract

Artificial intelligence (AI) holds significant promise for electrocardiogram (ECG) analysis, yet accurately detecting non-ST-segment elevation myocardial infarction (NSTEMI) and overcoming the "black box" nature of deep learning models remain persistent challenges. Here, we present a comprehensive deep learning framework capable of classifying STEMI, NSTEMI, and non-acute coronary syndrome (non-ACS) from 12-lead ECG images, while also localizing infarction sites. Utilizing ,2070 validated ECGs, our pipeline integrates ResNet for acute myocardial infarction detection, Faster R-CNN for ST-segment elevation localization, and an ensemble approach for final

classification. The model achieved a 98.3% AUROC for detection and an overall three-class accuracy of 93.6%, with high F1 scores for identifying infarction territories. To address interpretability, we developed an explainable AI (XAI) web viewer that visualizes detected regions. Furthermore, we evaluated the model's utility as an educational tool in a prospective pilot study with medical students. AI assistance significantly improved the students' overall diagnostic accuracy from 43% to 82% ($p<0.05$), with notable gains in identifying NSTEMI and complex STEMI subtypes. These findings demonstrate that our interpretable AI model not only supports clinical decision-making with high diagnostic precision but also serves as an effective educational aid for enhancing novice clinicians' proficiency.

1 Introduction

Recent advances in artificial intelligence (AI) have introduced both remarkable opportunities and significant challenges within the medical field, particularly in clinical diagnostics. AI has demonstrated substantial potential across diverse medical domains, including disease detection, treatment planning, and patient risk stratification [1–6]. Among these applications, the 12-lead electrocardiogram (ECG) represents a low-cost, non-invasive, and rapid diagnostic tool capable of identifying a wide spectrum of cardiac abnormalities, making it one of the most extensively explored modalities for AI-driven analysis. AI-based ECG models have achieved promising outcomes in predicting atrial fibrillation [7], identifying left ventricular dysfunction [8], and forecasting heart failure [9].

This study focuses on acute coronary syndrome (ACS), a critical cardiovascular condition that demands immediate intervention and encompasses acute myocardial infarction (AMI). AMI is further classified into non-ST-segment elevation myocardial infarction (NSTEMI) and ST-segment elevation myocardial infarction (STEMI). The underlying pathophysiology of AMI is primarily associated with the rupture or erosion of atherosclerotic plaques [10]. NSTEMI results from partial coronary artery occlusion, whereas STEMI arises from complete occlusion [11]. STEMI is particularly severe and necessitates urgent reperfusion therapy to limit myocardial necrosis and improve survival outcomes [12,13]. In emergency settings, the ECG remains indispensable for the diagnosis of ACS; however, its diagnostic performance is constrained by several factors. The absence of posterior leads, individual variations in coronary anatomy, and the presence of collateral circulation can diminish both sensitivity and specificity [14,15]. Furthermore, ECG interpretation for AMI is often challenging, even for experienced clinicians, including paramedics, emergency physicians, and cardiologists [16,17]. In cases of non-ST-segment elevation ACS (NST-ACS), where ST elevation is not apparent, additional biomarker testing and serial ECG monitoring are frequently required, leading to delayed diagnosis and treatment [18].

To overcome these diagnostic challenges, numerous AI-based studies have sought to detect AMI using ECG data [19–22]. Several efforts have focused on converting widely used paper-based ECGs into digital signals to extract and recognize characteristic patterns [23–25]. However, traditional image processing approaches are highly susceptible to clinical variables such as poor print quality, background noise, and electrode misplacement, leading to suboptimal diagnostic accuracy in real-world settings [26]. Deep learning, with its capacity for data-driven, end-to-end representation learning, has emerged as a promising approach to mitigate these limitations and effectively address noise-related artifacts.

This is a provisional file, not the final typeset article

Despite these advances, current deep learning models continue to face key methodological limitations. Most notably, existing algorithms predominantly focus on STEMI detection, which presents with more distinct ECG abnormalities, whereas robust models capable of accurately distinguishing the more subtle NSTEMI remain limited. Additionally, the intrinsic "black box" nature of machine learning models poses a major barrier to clinical adoption. The lack of interpretability regarding how these models generate diagnostic conclusions undermines clinician confidence, particularly in high-stakes environments where every decision may directly affect patient outcomes [27]. Understanding the basis of model predictions and identifying potential systematic biases are essential to ensure accountability among physicians who retain ultimate clinical responsibility [28]. To date, no clinically validated study has comprehensively employed deep learning to classify AMI from 12-lead ECG images while concurrently identifying and interpreting STEMI subtypes based on established medical criteria.

Therefore, this study proposes a methodological framework with several key objectives. First, we developed a deep learning model capable of accurately differentiating STEMI, NSTEMI, and non-ACS cases from 12-lead ECG images. Second, for ECGs predicted as STEMI, we implemented a classification module to precisely localize the infarction site by integrating medical domain knowledge of ECG changes associated with coronary artery territories. Third, we incorporated post-hoc explanation methods to establish an "explainable AI (XAI)" system that enhances transparency and clinical reliability through visual interpretation of model outputs. Finally, we evaluated whether the ECG-AMI Viewer—developed from this framework—can assist polyclinic students (PKs) in identifying myocardial infarction cases, thereby assessing its educational value. The overall structure of the proposed model is illustrated in Figure 1.

2 Materials and Methods

2.1 Study Design

This study comprised two main components: (1) retrospective data collection for the development and validation of an ECG-based AI model, and (2) a prospective pilot study to evaluate the educational utility of the developed AI system.

1. Data Composition

Between January 2011 and December 2019, 12-lead ECGs were obtained from emergency department patients diagnosed with AMI and from those without ACS. Patients with ST-STEMI and NSTEMI were categorized into the AMI group. ECGs from patients without coronary syndrome were randomly selected to form the control group. AMI classification was determined through a comprehensive review of laboratory findings, coronary intervention records, echocardiographic data from electronic health records, and final confirmation by cardiologists.

2. Pilot Study Design

A pilot study was conducted to assess the effectiveness of the ECG-AI model as both an educational and diagnostic support tool for polyclinic students (PKs).

2.1. Study Participants

The study was conducted at a mid-sized medical school in South Korea that follows a six-year, case-based learning curriculum emphasizing experiential learning in real clinical settings. Simulation-based education (SBE) is integrated throughout all years, with an increasing focus on complex clinical simulations, such as critical care, in the senior years. The study population consisted of fourth-year medical students with substantial clinical training who were nearing graduation. These participants had completed foundational ECG interpretation training and hospital-based clinical rotations, representing a cohort for whom proficiency in ECG interpretation is essential for the rapid identification of myocardial infarction in clinical practice.

Prospective convenience sampling was employed for participant recruitment. Based on prior studies with similar research designs [30,31], the target sample size was set at six to ten students to ensure adequate data representation for 12-lead ECG test items. Participation invitations were distributed via email to all fourth-year students, and the first ten respondents were enrolled. Two students with prior exposure to the ECG-AI system were excluded to avoid bias, yielding a final sample of eight participants.

2.2. Assessment Tools and Study Procedures

The ECG interpretation test consisted of 60 ECG images displayed on a computer monitor over a two-hour testing period. The duration was selected based on prior evidence indicating that clinicians require an average of 113 seconds to interpret a 12-lead ECG [32]; thus, 120 seconds per item was allotted to ensure sufficient review time.

The test set included 15 control ECGs, 30 STEMI ECGs, and 15 NSTEMI ECGs. To reflect real-world case diversity and clinical frequency, the 30 STEMI ECGs were subdivided into specific infarction types and randomly ordered: anterior wall (13 items), inferior wall (8 items), lateral wall (8 items), and suspected left main disease (6 items). Some ECGs demonstrated multi-territory infarction patterns, resulting in a total of 35 categorized STEMI sub-items.

The pilot study employed a pre-test/post-test design to evaluate the effect of AI assistance on ECG interpretation accuracy. The overall experimental framework is illustrated in Figure 2.

The study was conducted in two sequential phases:

Pre-test (baseline assessment): Participants interpreted 60 ECGs within two hours without AI assistance.

Post-test (AI-assisted assessment): One week later, participants interpreted a new set of 60 ECGs—matched in type and difficulty to the pre-test—within the same two-hour time frame, this time with AI assistance.

2.2 Dataset

Standard supine 12-lead ECGs were acquired at a paper speed of 25 mm/s and an amplitude setting of 10 mm/mV. For patients with AMI, the first ECG obtained upon arrival at the emergency department was selected as the MI ECG. The control group consisted of patients randomly selected from those with no history or clinical evidence of coronary syndrome. All ECGs were collected using the Philips ECG system (PageWriter TC30, TC70; Philips Medical Systems). Each ECG image comprised a grid background with electrical signals represented by fine tracing lines (Fig. 3(a)).

This is a provisional file, not the final typeset article

Between January 2011 and December 2019, a total of 2,070 ECG images were collected from the emergency department of Kyungpook National University Hospital. This dataset included 1,092 ECGs from patients with AMI (725 STEMI, 367 NSTEMI) and 978 ECGs from individuals without ischemic heart disease. Patients without ischemic heart disease were defined as those with no prior diagnosis of ischemic heart disease and whose final clinical assessment confirmed a non-ischemic etiology for their visit. The study protocol and data collection procedures were approved by the Institutional Review Board of Kyungpook National University Hospital (approval number: 2020-12-022-011).

2.3 Data Annotation

A supervised learning framework was employed for model training. Diagnosis labeling of 12-lead ECG images for STEMI, NSTEMI, and control groups was performed based on medical records and expert clinical judgment. In addition to ST-segment elevation (STE, ≥1 mm above baseline), regions showing lesser STE (<1 mm above baseline) were also annotated, as these subtle findings can serve as important features for distinguishing STEMI from NSTEMI.

Ground truth labels were generated using the labelImg Python package [29]. Human annotators manually marked each ECG lead for STE (regions where the ST segment rose ≥1 mm, including the QRS complex) and lesser STE (regions where the ST segment rose <1 mm, including the QRS complex) (Fig. 3(b)). The criterion for STE was defined as an elevation of at least 1 mm above the baseline at the J-point, following the American College of Cardiology/American Heart Association guidelines for STEMI diagnosis.

All annotations were validated against a gold standard established through consensus by three board-certified cardiologists to ensure diagnostic accuracy and labeling consistency across STEMI, NSTEMI, and control datasets.

2.4 Data Preprocessing

The ECG data used in this study were obtained as scanned images of paper-based printouts, with an original resolution of 1228 × 2460 pixels. A standard 12-lead ECG typically includes 12 short leads and one long rhythm lead (Lead II). In this study, the long rhythm lead was subdivided into four equal segments to enable independent detection of STE within each subregion. Unlike traditional ECG digitization techniques that require complex preprocessing steps such as grid line removal or waveform thinning, the proposed deep learning approach required only a single channel separation step. This minimal preprocessing process highlights the robustness of the model to noise and artifacts commonly present in printed ECG documents.

2.5 Model Architecture

The system proposed in this study is a multi-stage deep learning pipeline based on ECG images, designed to emulate a physician's diagnostic reasoning process. This design allows intuitive interpretation, even for users with limited experience in AI. The pipeline performs comprehensive AMI diagnosis through four main stages. First, it classifies ECG images to determine the presence of AMI. Second, it detects ST-STE within ECGs classified as MI. Third, based on these detection results, it differentiates between STEMI and NSTEMI and determines the infarction territory. Finally, it calculates the integrated probabilities for each diagnostic category.

Stage 1: Classification of Myocardial Infarction

In the first stage, 12-lead ECG images are analyzed to determine whether AMI is present, serving as an initial filter to distinguish AMI ECGs from control ECGs. To identify the optimal model architecture, several deep learning frameworks were evaluated, including ResNet (ResNet18, ResNet34, ResNet50), EfficientNet (b0–b7), and ConvNeXtV2 (base, small, tiny). ResNet introduces residual learning to facilitate the training of deeper networks and mitigate the vanishing gradient problem [33]. EfficientNet applies a compound scaling approach that uniformly adjusts network depth, width, and resolution to achieve high accuracy with minimal computational cost [34]. ConvNeXtV2 incorporates design principles from vision transformers into convolutional neural networks, providing improved efficiency and performance [35]. All models were initialized with weights pre-trained on the ImageNet dataset to enhance learning efficiency. Training was performed for 200 epochs with a batch size of 16 and a learning rate of 0.00005. To improve generalization, data augmentation was applied using the Albumentations library, including variations in brightness and contrast, hue and saturation adjustments, and the introduction of noise. The output of this stage provides the probabilities that the input ECG image belongs to either the control or AMI class ($P_{Control}$, P_{AMI}).

Stage 2: Detection of ST-segment Elevations

For ECG images classified as AMI, an object detection model—Faster R-CNN—was used to automatically identify regions of STE and lesser STE, which are essential for STEMI diagnosis. Faster R-CNN enhances detection speed and accuracy by introducing a Region Proposal Network (RPN) that efficiently generates candidate regions likely to contain relevant features [36]. The model employed a ResNet50 backbone and utilized weights pre-trained on the COCO dataset [37] to improve training efficiency. Cardiologists manually annotated bounding boxes on ECG images to establish ground-truth data for training. The model input consisted of 384 × 384 pixel ECG images, and its output included bounding box coordinates, class labels, and confidence scores for detected STE and lesser STE objects. To refine predictions and reduce redundancy, Soft Non-Maximal Suppression (Soft NMS) was applied during post-processing, ensuring that only the most reliable detections were retained.

Stage 3: Classification of STEMI & NSTEMI and Infarction Territory Determination

In the third stage, the model differentiates between STEMI and NSTEMI and determines the infarction territory using the ST-segment detection results from the previous stage.

For STEMI/NSTEMI classification, the outputs from the STE detection model were converted into a 12-dimensional binary feature vector using the Lead-STE Presence encoding method, indicating the presence or absence of STE in each lead. Based on this feature vector, the AutoGluon Tabular framework was employed to construct the optimal classification model. AutoGluon is an automated machine learning (AutoML) system that explores a diverse range of models—including LightGBM, CatBoost, XGBoost, RandomForest, ExtraTrees, and deep neural networks—and ensembles them to achieve high stability and predictive performance [38]. The output provides the conditional probabilities that the ECG image belongs to the STEMI or NSTEMI class (P_{STEMI}, P_{NSTEMI})

Infarction territory determination was performed using a rule-based logic grounded in established clinical ECG interpretation guidelines. This approach utilized the STE information encoded from all

This is a provisional file, not the final typeset article

12 leads to classify infarction territories. The specific diagnostic criteria used to determine infarction locations are summarized in Table 1.

Stage 4: Final Probability Calculation

The final stage integrated the outputs of the previous modules to compute the overall probabilities for control, STEMI, and NSTEMI classifications and to identify the infarction territory.

The final probabilities were calculated as follows:

$$P_{Final_Control} = P_{Control}$$

$$P_{Final_STEMI} = P_{AMI} \times P_{STEMI \mid AMI}$$

$$P_{Final_NSTEMI} = P_{AMI} \times P_{NSTEMI \mid AMI}$$

2.6 ECG Viewer

An additional component of this study was the ECG Viewer, a web-based application developed to assess the diagnostic capabilities of medical professionals and to facilitate ECG testing during the pilot study. The viewer visually displayed AI-detected regions of ST-STE (red bounding boxes) and lesser STE (blue bounding boxes) superimposed on the ECG image. Users could easily toggle between the AI-generated predictions and the actual diagnostic results for direct comparison. The interface also allowed control over bounding box visibility, and hovering over a box revealed its classification label and confidence score. These interactive features enabled medical professionals to intuitively understand the AI model's diagnostic rationale and contributed to mitigating the "black box" limitation commonly associated with deep learning systems.

2.7 Data Analysis

Evaluation Metrics for Deep Learning Algorithms

A comprehensive set of quantitative metrics was used to evaluate the performance of the developed deep learning model. This approach enabled assessment of predictive accuracy from multiple perspectives and verified model robustness under conditions of potential data imbalance. Performance evaluation was based on true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The following key metrics were calculated and analyzed.

Accuracy represented the proportion of correct predictions among all predictions, providing a general measure of overall model performance. Sensitivity (or Recall) denoted the proportion of actual positive samples correctly identified as positive, whereas specificity measured the proportion of actual negative samples correctly identified as negative. Precision indicated the proportion of true positive predictions among all positive predictions, and the F1-score, defined as the harmonic mean of Precision and Recall, reflected the balance between sensitivity and precision—particularly useful in imbalanced datasets.

To further evaluate performance across all threshold levels, both the receiver operating characteristic (ROC) curve and area under the ROC curve (AUC), as well as the precision–recall (PR) curve and

area under the PR curve (AUPRC), were employed. The ROC curve depicts the relationship between sensitivity and the false-positive rate (1 − specificity), while the PR curve illustrates the trade-off between Precision and Recall, providing complementary insights into classification performance.

$$Accuracy = \frac{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN}}$$

$$Sensitivity = Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$Specificity = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

$$F1 - Score = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

Analysis of Clinical Applicability of Deep Learning Algorithms

To assess the clinical applicability of the deep learning model, results were collected from fourth-year medical students interpreting 12-lead ECGs related to AMI with and without AI assistance. Data normality was examined using the Shapiro–Wilk test. Identification performance for STEMI, NSTEMI, control, and STEMI subtypes (anterior MI, inferior MI, lateral MI, and suspected left main disease) before and after AI assistance was compared using accuracy as the primary metric. The paired Wilcoxon signed-rank test was applied to assess differences in accuracy between pre- and post-AI conditions.

3 Results

AI Model Performance Evaluation

The proposed AI model demonstrated excellent performance in ECG-based AMI diagnosis. Model performance was evaluated across three primary diagnostic functions, with results for each stage summarized below.

In the initial stage of 12-lead ECG analysis for AMI classification, the ResNet18 model achieved the highest overall performance. It recorded an AUROC of 98.3% and an AUPRC of 98.7%, effectively distinguishing AMI from control ECGs. These findings indicate that even a relatively lightweight architecture such as ResNet18 is sufficient for accurate AMI classification within the current dataset. This also suggests that more complex architectures may further improve performance when trained on larger datasets. Detailed results are presented in Table 2, which highlights the reliability and consistency of the overall diagnostic pipeline.

Next, the object detection model was evaluated for its ability to identify STE regions in ECGs classified as myocardial infarction. The mean average precision (mAP), reflecting the overall detection performance, was 42.4%. Importantly, the AP50 (average precision at an intersection over

This is a provisional file, not the final typeset article

union [IoU] threshold of 0.5)—a clinically relevant indicator of localization accuracy—was 71.5%, demonstrating the model's capability to accurately identify and localize STE regions.

The third evaluation focused on STEMI classification performance. As shown in Table 3, the model achieved an AUROC of 91.3% and an AUPRC of 95.1%, with a sensitivity of 95.9%, confirming its strong ability to detect STEMI cases effectively.

Furthermore, as summarized in Table 4, the model demonstrated high accuracy in infarction site prediction among STEMI patients. The F1-scores for infarction territories were as follows: anterior (91.1%), inferior (84.2%), lateral (73.8%), and suspected left main disease (68.4%). Although a relatively higher number of false positives was observed for lateral MI and suspected left main disease, the model maintained high overall accuracy, sensitivity, and specificity, underscoring its potential for clinical application.

In the final stage, the model was evaluated for its ability to classify ECGs into control, STEMI, and NSTEMI categories. As shown in Table 5, classification was conducted using a one-vs-rest approach. The model achieved a macro-average accuracy of 93.6%. Despite the inherent diagnostic challenges posed by NSTEMI due to subtle or nonspecific ST-segment changes, the model successfully differentiated NSTEMI from other categories, suggesting strong clinical applicability.

Effect of Deep Learning AI Model on Polyclinic Students

A total of eight fourth-year polyclinic students participated in the pilot study to evaluate the educational and diagnostic impact of the deep learning (DL) algorithm. The median accuracy of participants' 12-lead ECG interpretations improved significantly from 43% without AI assistance to 82% with AI assistance ($p < 0.05$) (Fig. 4a). Accuracy in interpreting control ECGs increased from 48% without AI to 88% with AI support ($p = 0.012$). Similarly, identification performance improved for both STEMI (42% to 81%, $p = 0.011$) and NSTEMI (21% to 50%, $p = 0.012$) (Fig. 4b). The DL model enhanced diagnostic accuracy not only for STEMI, which typically presents with clearer ECG changes, but also for the more diagnostically challenging NSTEMI.

Identification Performance by STEMI Subtype

To further evaluate performance, ECGs were classified into specific STEMI subtypes: anterior MI, inferior MI, lateral MI, and suspected left main disease. In the test dataset, the distribution of these subtypes was as follows: anterior MI (37%), inferior MI (22%), lateral MI (22%), and suspected left main disease (17%). The ECG interpretation accuracy of students varied markedly across subtypes, with the lowest performance observed in suspected left main disease (median 17%) compared with inferior MI (median 83%).

In contrast, the DL algorithm achieved an average accuracy of 95% across all STEMI subtypes, substantially outperforming the students. Although students demonstrated higher accuracy for inferior MI than for suspected left main disease, the DL model consistently improved accuracy across all subtypes: anterior MI (58% to 80%, $p = 0.012$), inferior MI (83% to 91%, $p = 0.035$), lateral MI (21% to 62%, $p = 0.012$), and suspected left main disease (17% to 60%, $p = 0.011$). Overall, AI assistance led to significant improvements in ECG interpretation accuracy for all four STEMI subtypes, as shown in Figure 5, demonstrating the educational and diagnostic value of the model in enhancing clinical recognition of diverse infarction patterns.

4 Discussion

This study demonstrates the potential of deep learning–based AI models to improve the diagnostic accuracy of 12-lead ECGs for AMI. We developed a deep learning framework capable not only of identifying AMI but also of localizing the anatomical site of infarction in STEMI. By integrating explainable AI (XAI) techniques, the model provides transparency regarding its predictive rationale, addressing one of the most critical barriers to AI adoption in medicine. Furthermore, the pilot study confirmed the model's educational effectiveness, showing substantial improvement in ECG interpretation accuracy among polyclinic students.

The significance of this study lies in its comprehensive, clinically grounded approach to solving longstanding diagnostic challenges in real-world cardiology practice. Unlike many previous investigations that primarily targeted STEMI—with its distinct ECG manifestations—this model achieved high accuracy (95%) in distinguishing NSTEMI, a considerably more difficult entity to diagnose. This performance exceeds that of internal medicine specialists (50%) and AI-assisted specialists (56%) using the same dataset, as reported in prior research [18]. The model's capacity to recognize NSTEMI in the presence of subtle or nonspecific ECG alterations highlights its strong potential for early detection, timely intervention, and improved patient outcomes. Moreover, our model surpassed the AMI classification accuracy of the prior AI model [18] and achieved performance comparable to AI-assisted internal medicine specialists in STEMI detection. Importantly, it extended beyond simple AMI classification by accurately localizing infarction territories (anterior, inferior, lateral, and suspected left main disease). Such spatial diagnostic precision can enable cardiologists to formulate rapid reperfusion strategies and optimize decision-making in time-sensitive emergency contexts.

A particularly notable aspect of this study is the objective validation of both the clinical and educational utility of the model through the pilot experiment. The ECG interpretation accuracy of fourth-year polyclinic students increased markedly from 43% without AI assistance to 82% with AI support. These results demonstrate that AI can serve as a powerful adjunctive tool to mitigate diagnostic limitations among novice clinicians and, in turn, enhance patient safety. The improvement was especially striking for the recognition of ECGs suggestive of left main coronary artery disease—a condition characterized by low incidence and high interpretive difficulty—where accuracy rose from 17% to 60%. This finding provides tangible evidence that AI can effectively identify complex and atypical electrocardiographic patterns, offering valuable support to both trainees and healthcare professionals.

Finally, this study contributes a key methodological advancement by addressing the persistent "black box" problem of AI in clinical medicine. The lack of transparency in algorithmic decision-making has long hindered the trust and integration of AI tools into routine practice. By incorporating visual explanations through the ECG Viewer (Fig. 6), which highlights the specific leads and waveform segments used for predictions, the model enhances interpretability and allows clinicians to critically appraise AI outputs while retaining full clinical responsibility. This explainability not only reinforces physician trust but also represents a crucial step toward the safe and ethical implementation of AI-assisted diagnostics in cardiology.

Study Strengths

This is a provisional file, not the final typeset article

Granular Diagnosis and High Accuracy: The proposed model comprehensively classifies STEMI, NSTEMI, and non-ACS cases while precisely localizing infarction territories in STEMI, thereby enhancing its clinical applicability.

Use of Real-World Clinical Data: A key strength of this study lies in its use of authentic, paper-based ECG images for model training—data that inherently include real-world variability such as background noise and print quality degradation. This approach enabled the development of a robust and field-ready model, in contrast to previous systems trained solely on preprocessed digital signals.

Explainability (XAI) Integration: The "ECG Viewer" provides a visual representation of the medical evidence underpinning the model's diagnostic decisions, thereby improving interpretability and fostering clinician trust. This explainability feature represents a critical step toward greater clinical acceptance and responsible integration of AI in medical practice.

Validated Educational Utility: The pilot study statistically confirmed that the AI tool significantly improved diagnostic performance among medical trainees, highlighting its dual role as both a clinical decision support system and an educational resource.

Study Limitations

Single-Institution Dataset: The study utilized data exclusively from Kyungpook National University Hospital, which may reflect institution-specific characteristics related to race, epidemiology, and equipment (Philips ECG systems). This limitation may restrict the generalizability of findings to other hospitals, device types, or patient populations.

Retrospective Data Collection: Because data were retrospectively collected between 2011 and 2019, the potential for selection bias cannot be completely ruled out.

Small-Scale Pilot Study: The pilot study evaluating the educational impact of the AI system involved only eight participants. Although the improvement in diagnostic accuracy reached statistical significance, larger randomized controlled trials including participants of varying experience levels—such as medical students from different academic years, interns, residents, and paramedics—are warranted to confirm and generalize these findings.

Lack of Real-Time Clinical Validation: The current evaluation was performed in a controlled testing environment rather than a real-time clinical context. Future prospective studies are needed to assess how the AI model performs when integrated into emergency department workflows and to determine its effect on key clinical outcomes, such as reducing time to reperfusion therapy.

In conclusion, this study successfully developed a deep learning–based AI model capable of precise diagnosis and visual explanation of AMI using 12-lead ECG images. The model demonstrated substantial promise for both clinical application and medical education. Future large-scale, multi-institutional prospective studies are warranted to validate its generalizability and real-world clinical effectiveness. If confirmed, this system could play a pivotal role in enhancing diagnostic accuracy for cardiovascular emergencies and improving patient outcomes.

5 Conclusion

This study underscores the potential of AI to serve not merely as an automated diagnostic instrument but as a robust collaborative partner that supports both clinical decision-making and medical education. By training on authentic, noise-containing ECG images from real-world clinical environments, the model demonstrated high precision in distinguishing diagnostically challenging NSTEMI cases and accurately identifying infarction sites in STEMI. These findings highlight its capability to assist clinicians in making rapid, evidence-based decisions. Furthermore, the pilot study revealed a significant improvement in diagnostic accuracy among novice clinicians using AI assistance, suggesting the model's promise as an effective educational tool for bridging knowledge and experience gaps while promoting standardized diagnostic proficiency.

Importantly, the incorporation of XAI features directly addresses the longstanding "black box" concern by providing interpretability and transparency—key prerequisites for safe and responsible AI integration into clinical workflows. This explainability empowers clinicians to critically appraise AI recommendations while maintaining full professional accountability. Nonetheless, because the model was developed using data from a single institution, further research employing multi-institutional datasets is essential to ensure its generalizability and broader clinical applicability.

Ultimately, this study represents a pivotal step toward the realization of an era of augmented intelligence, in which AI complements rather than replaces clinicians—uniting computational precision with human expertise to advance diagnostic accuracy, consistency, and patient safety.

6 Abbreviations

AI: Artificial intelligence

ECG: Electrocardiogram

AMI: Acute myocardial infarction

STEMI: ST-segment elevation myocardial infarction

NSTEMI: Non-ST-segment elevation myocardial infarction

STE: ST-segment elevation

ACS: Acute coronary syndrome

NST: Non-ST-segment elevation

XAI: Explainable AI

PK: Polyclinic student

SBE: Simulation based education

RPN: Region Proposal Network

NMS: Non-Maximal Suppression

AutoML: Automated Machine Learning

This is a provisional file, not the final typeset article

TP: True positive

TN: True negative

FP: False positive

FN: False negative

AUC: Area Under the ROC Curve

PR: Precision–recall

AUPRC: Area Under the PR Curve

mAP: mean Average Precision

IoU: Intersection over Union

DL: Deep learning

RCT: Randomized Controlled Trial

7 Data availability

The dataset used for this study is not publicly available due to the possibility of compromising individual privacy but is available from the corresponding author on reasonable request.

8 Funding

This work was partly supported by the Institute of Information & Communications Technology Planning & Evaluation(IITP)-ITRC(Information Technology Research Center) grant funded by the Korea government(MSIT)(IITP-2025-RS-2020-II201808) And This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : RS-2022-KH130593).

9 Acknowledgments

The authors would like to thank Dong Heon Yang, Professor of Cardiology, Kyungpook National University Hospital, Daegu, Republic of Korea, for his valuable assistance in editing the manuscript.

10 Declarations

10.1 Ethics approval

This study was approved by the Institutional Review Board (2020-12-022-011) at Kyungpook National University Hospital, Daegu, Republic of Korea. All methods were carried out in accordance with the regulations of the Institutional Review Board. The Institutional Review Board of Kyungpook National University Hospital granted an exemption from the need to obtain informed consent due to the thorough de-identification of patient data. Clinical trial number: not applicable.

10.2 Competing interests

The authors declare no competing interests.

11 Author information

Jongkwang Kim and Byungeun Shon contributed equally to this work.

Authors and Affiliations

Department of Medical Informatics, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

Jongkwang Kim, Byungeun Shon, Sungmoon Jeong

Research Center for AI in Medicine, Kyungpook National University Hospital, Daegu, Republic of Korea

Jongkwang Kim, Byungeun Shon, Sungmoon Jeong

College of Medicine, Kyungpook National University, Daegu, Republic of Korea

Yongjin Kim, Nayeon Kim, Yejin Jang, Jeongmin Lee

Department of Forensic Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

Seong-Mi Yoon

Preclinical Research Center, Daegu-Gyeongbuk Medical Innovation Foundation (K-MEDI hub), Daegu, Republic of Korea

Seong-Mi Yoon

Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

Dong Heon Yang, Namkyun Kim

This is a provisional file, not the final typeset article

Division of Cardiology, Department of Internal Medicine, Kyungpook National University Hospital, Daegu, Republic of Korea

Dong Heon Yang, Namkyun Kim

Author contributions

JK designed, developed, methodology, analyzed the results, project administration, manuscript writing, manuscript review and revision. BS manuscript writing, manuscript review and revision, visualization, software, methodology, data curation. YK investigation, manuscript writing–review and editing, project management, methodology. NK investigation, manuscript writing–review, project administration, methodology. YJ investigation, manuscript writing–review, project management, methodology. JL investigation, manuscript writing–review, project management, methodology. SY investigation, manuscript review, project management, methodology. DY project administration, funding acquisition, investigation, manuscript review. NK project administration, designed, analyzed the results, manuscript review. SJ project administration, funding acquisition, investigation, conceptualization, manuscript review and revision.

Corresponding author

Correspondence to Sungmoon Jeong

12 References

1. Adler ED. et al. Improving risk prediction in heart failure using machine learning. *Eur J Heart Fail.* **22**(1), 139–47. https://doi.org/10.1002/ejhf.1628 (2020).

2. Oren O, Gersh BJ, Bhatt DL. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *Lancet Digit Health.* **2**(9), e486–8. https://doi.org/10.1016/S2589-7500(20)30160-6 (2020).

3. Shmatko A, Ghaffari Laleh N, Gerstung M, Kather JN. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat Cancer.* **3**(9), 1026–38. https://doi.org/10.1038/s43018-022-00436-4 (2022).

4. Truslow JG. et al. Cardiovascular risk assessment using artificial intelligence-enabled event adjudication and hematologic predictors. *Circ Cardiovasc Qual Outcomes.* **15**(6), e008007. https://doi.org/10.1161/CIRCOUTCOMES.121.008007 (2022).

5. Park JJ. et al. A machine learning derived risk score predicts mortality in East Asian patients with acute heart failure. *Eur J Heart Fail.* **25**, 2331–3. https://doi.org/10.1002/ejhf.3059 (2023).

6. Xu Z. et al. The efficacy of machine learning models for predicting the prognosis of heart failure: a systematic review and meta-analysis. *Cardiology.* **22**, 1–19. https://doi.org/10.1159/000538639 (2024).

546 7. Attia ZI. et al. An artificial intelligence-enabled ECG algorithm for the identification of patients
547 with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet.*
548 **394**(10201), 861–7. https://doi.org/10.1016/S0140-6736(19)31721-0 (2019).

549 8. Attia ZI. et al. Screening for cardiac contractile dysfunction using an ar tificial intelligence-enabled
550 electrocardiogram. *Nat Med.* **25**(1), 70–4. https://doi.org/10.1038/s41591-018-0240-2 (2019).

551 9. Akbilgic O. et al. ECG-AI: electrocardiographic artificial intelligence model for prediction of heart
552 failure. *Eur Heart J Digit Health.* **2**(4), 626–34. https://doi.org/10.1093/ehjdh/ztab080 (2021).

553 10. Arbustini E. et al. Plaque erosion is a major substrate for coronary thrombosis in acute
554 myocardial infarction. *Heart.* **82**(3), 269–272. https://doi.org/10.1136/hrt.82.3.269 (1999).

555 11. Bhat P., Dretler A., Gdowski M., Ramgopal R., Williams E. D. The Washington manual of
556 medical therapeutics. Saint Louis, MO, USA: Washington University. (2016).

557 12. CannonCP. et al. Relationship of symptom-onset-to-balloon time and door-to balloon time with
558 mortality in patients under going angioplasty for acute myocardial infarction. *JAMA.* **283**(22), 2941–
559 7. https://doi.org/10.1001/jama.283.22.2941 (2000).

560 13. Murray C. J. et al. Global, regional, and national disability-adjusted life years (DALYs) for 306
561 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: quantifying
562 the epidemiological transition. *Lancet.* **386**, 2145–2191. https://doi.org/10.1016/S0140-
563 6736(15)61340-X (2015).

564 14. Tragardh E, Claesson M, Wagner GS, Zhou S, Pahlm O. Detection of acute myocardial infarction
565 using the 12-lead ECG plus inverted leads versus the 16-lead ECG (with additional posterior and
566 right-sided chest electrodes). *Clin Physiol Funct Imaging.* **27**(6), 368–74.
567 https://doi.org/10.1111/j.1475-097X.2007.00761.x (2007).

568 15. McCabe JM. et al. Physician accuracy in interpreting potential ST segment elevation myocardial
569 infarction electrocardiograms. *J Am Heart Assoc.* **2**(5), e000268.
570 https://doi.org/10.1161/JAHA.113.000268 (2013).

571 16. Cook DA, Oh SY, Pusic MV. Accuracy of physicians' electrocardiogram interpretations: a
572 systematic review and meta-analysis. *JAMA Intern Med.* **180**(11), 1461–71.
573 https://doi.org/10.1001/jamainternmed.2020.3989 (2020).

574 17. Debrabant B. et al. STEMI, cardiogenic shock, and mortality in patients admitted for acute
575 angiography: associations and predictions from plasma proteome data. *Shock.* **55**(1), 41–47.
576 https://doi.org/10.1097/shk.0000000000001595 (2021).

577 18. Bo Eun P. et al. Signal-Guided Multitask Learning for Myocardial Infarction Classification Using
578 Images of Electrocardiogram. *Cardiology.* **6**, 1-10 https://doi.org/10.1159/000542399 (2024).

579 19. Liu W-C. et al. A deep learning algorithm for detecting acute myocardial infarction. *Euro*
580 *Intervention.* **17**, 765–773. https://doi.org/10.4244/eij-d-20-01155 (2021).

This is a provisional file, not the final typeset article

20. Gustafsson S. et al. Development and validation of deep learning ECG-based prediction of myocardial infarction in emergency department patients. *Sci Rep.* **12**, 19615. https://doi.org/10.1038/s41598-022-24254-x (2022).

21. Lee BT. et al. Usefulness of deep-learning algorithm for detecting acute myocardial infarction using electrocardiogram alone in patients with chest pain at emergency department: DAMI-ECG study. *J Cardiovasc Interv.* **2**, 100–112. https://doi.org/10.54912/jci.2022.0028 (2023).

22. Chang K-C. et al. Usefulness of multi-labelling artificial intelligence in detecting rhythm disorders and acute ST-elevation myocardial infarction on 12-lead electrocardiogram. *Eur Heart J Digit Health.* **2**, 299–310. https://doi.org/10.1093/ehjdh/ztab029 (2021).

23. Lence A. et al. Automatic digitization of paper electrocardiograms–A systematic review. *J. Electrocardiol.* **80**, 125–132. https://doi.org/10.1016/j.jelectrocard.2023.05.009 (2023).

24. Badilini F., Erdem T., Zareba W., Moss A. J. ECGScan: a method for conversion of paper electrocardiographic printouts to digital electrocardiographic files. *J. Electrocardiol.* **38**(4), 310–318. https://doi.org/10.1016/j.jelectrocard.2005.04.003 (2005).

25. Waits G. S., Soliman E. Z. Digitizing paper electrocardiograms: status and challenges. *J. Electrocardiol.* **50**(1), 123–130. https://doi.org/10.1016/j.jelectrocard.2016.09.007 (2017).

26. Kim J. et al. ECG data analysis to determine ST-segment elevation myocardial infarction and infarction territory type: an integrative approach of artificial intelligence and clinical guidelines. *Front. Physiol.* **15**, 1462847. https://doi.org/10.3389/fphys.2024.1462847 (2024).

27. Tjoa E, Guan C. A survey on explainable artificial intelligence (xai): toward medical xai. *IEEE Trans Neural Networks Learn Syst.* **32**(11), 4793–813. https://doi.org/10.1109/TNNLS.2020.3027314 (2020).

28. Ebrahimi, A. et al. Identification of patients' smoking status using an explainable AI approach: a Danish electronic health records case study. *BMC Med Res Methodol.* **24**(1), 114. https://doi.org/10.1186/s12874-024-02231-4 (2024).

29. GitHub (2015). Tzutalin. LabelImg. Git code. Available at:

30. Gormley GJ, Fenwick T. Learning to manage complexity through sim ulation: students' challenges and possible strategies. *Perspect Med Educ.* **5**(3), 138-146. https://doi.org/10.1007/s40037-016-0275-3 (2016).

31. Behrens CC, Dolmans DH, Driessen EW, Gormley GJ. Dancing with emotions: an interpretive descriptive study of facilitators' recognition and response to students' emotions during simulation. *Med Educ.* **59**(4), 439-448. https://doi.org/10.1111/medu.15554 (2024).

32. Chen K-W. et al. Artificial intelligence-assisted remote detection of ST-elevation myocardial infarction using a mini-12-lead electrocardiogram device in prehospital ambulance care. *Front. Cardiovasc. Med.* **9**, 1001982. https://doi.org/10.3389/fcvm.2022.1001982 (2022).

33. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. **12**, 770-778. https://ieeexplore.ieee.org/document/7780459#:~:text=10.1109/CVPR.2016.90 (2016).

34. Erickson N. et al. Autogluon-tabular: Robust and accurate automl for structured data. *arXiv*, 2003.06505. (2020). https://arxiv.org/abs/2003.06505#:~:text=https%3A//doi.org/10.48550/arXiv.2003.06505

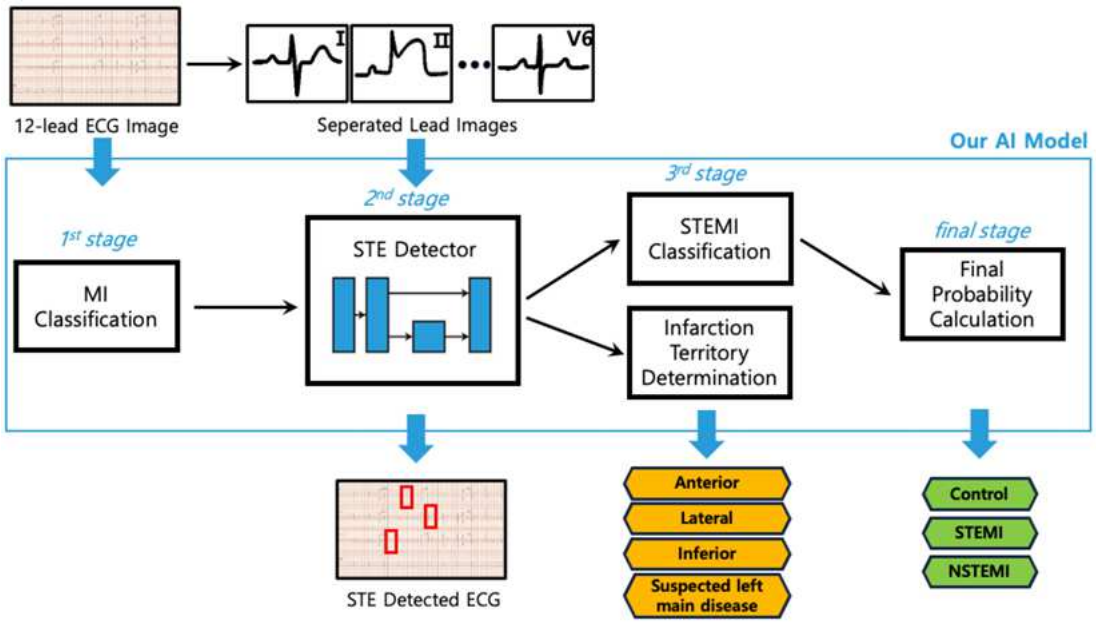
35. Woo S. et al. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. **1**, 16133-16142. https://www.researchgate.net/publication/373314265_ConvNeXt_V2_Co-designing_and_Scaling_ConvNets_with_Masked_Autoencoders#:~:text=10.1109/CVPR52729.2023.01548 (2023).

36. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv*, 1506.01497. https://arxiv.org/abs/1506.01497#:~:text=https%3A//doi.org/10.48550/arXiv.1506.01497 (2015).

37. Lin Y. et al. Microsoft COCO: Common Objects in Context. *arXiv*, 1405.0312. https://arxiv.org/abs/1405.0312#:~:text=https%3A//doi.org/10.48550/arXiv.1405.0312 (2014).

38. Tan M., Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv*, 1905.11946. https://arxiv.org/abs/1905.11946#:~:text=https%3A//doi.org/10.48550/arXiv.1905.11946 (2019).

13 Figure



This is a provisional file, not the final typeset article

FIGURE 1. Workflow of the AI system for ECG analysis based on 12-lead ECG images utilizing clinical evidence.

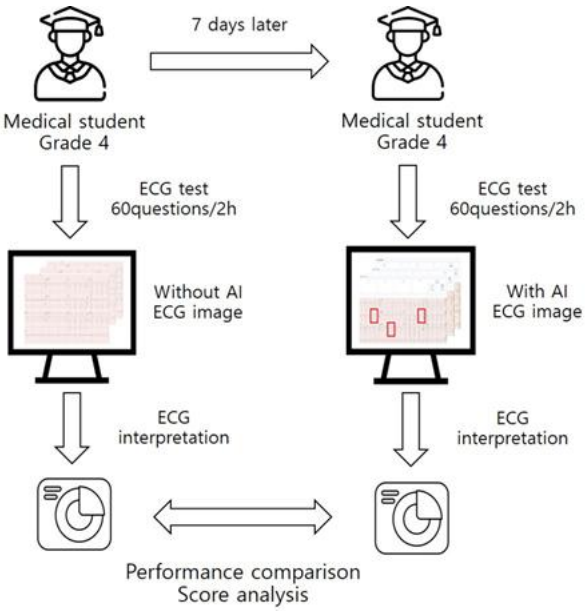


FIGURE 2. Study design for evaluating changes in ECG interpretation accuracy with and without AI assistance.

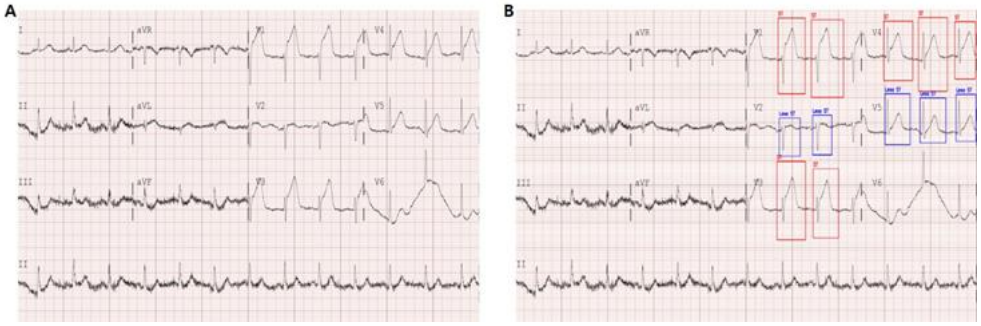


FIGURE 3. (A) Example of a digitized 12-lead ECG image. (B) Manual annotations of ST-segment elevation (STE; red) and lesser STE (blue) regions on a 12-lead ECG image.

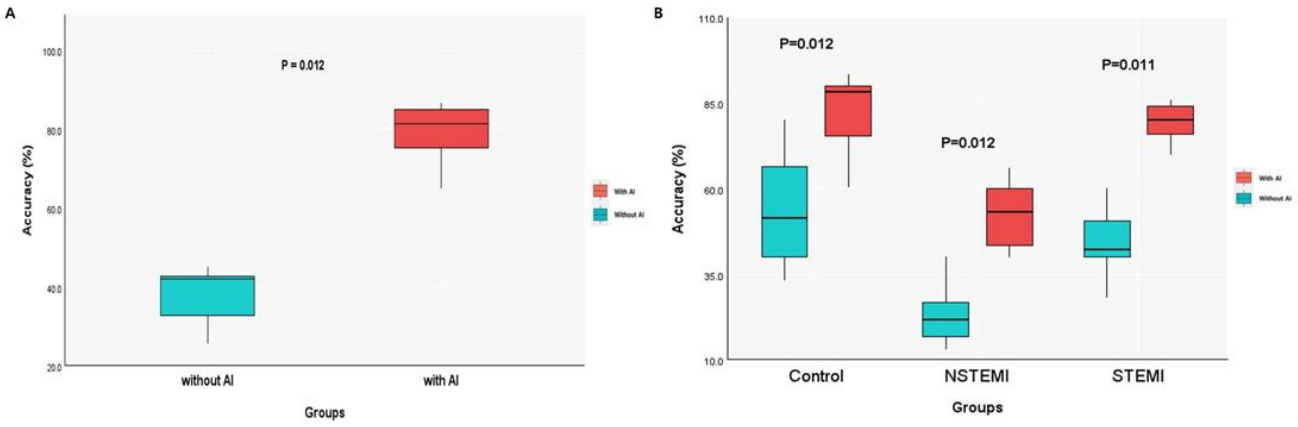


FIGURE 4. Comparison of improvements in 12-lead ECG interpretation accuracy and control/STEMI/NSTEMI identification performance among fourth-year medical students using a deep learning AI algorithm. (A) Overall ECG interpretation accuracy of all participants. (B) Classification accuracy by MI type and control category.

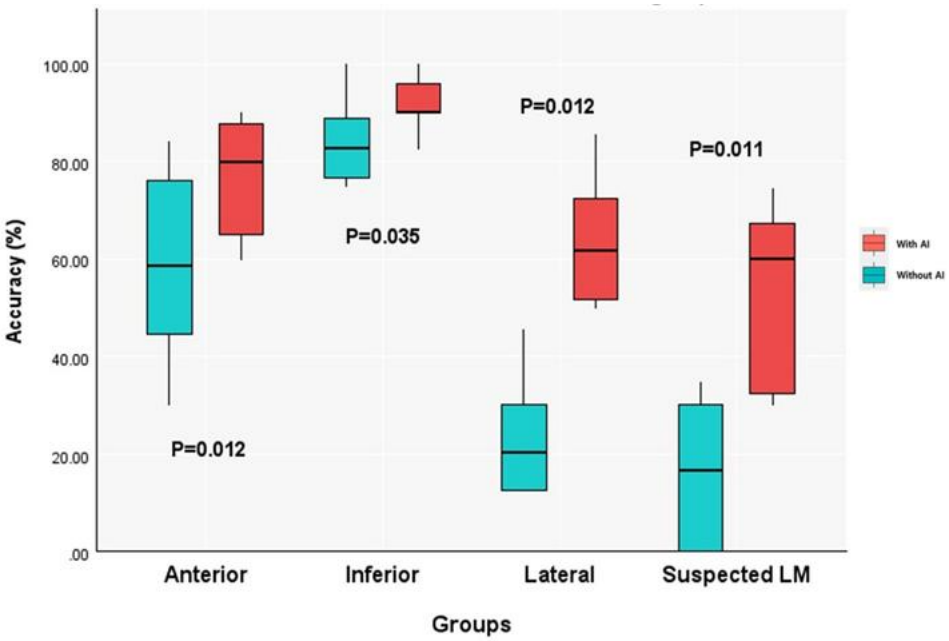


FIGURE 5. Comparison of improvements in ECG interpretation accuracy by STEMI subtype among polyclinic students using a deep learning AI algorithm.



FIGURE 6. Visualization of diagnostic rationale for the deep learning AI model using the ECG Viewer based on explainable AI (XAI).

This is a provisional file, not the final typeset article

14 Tables

TABLE 1. Clinical criteria for identifying infarction territories on the ECG

Infarction territory	Condition
Anterior	More than one lead shows STE among V1, V2, V3, and V4
Lateral	Two or more leads show STE among V5, V6, L1, and aVL
Inferior	Two or more leads show STE among L2, L3, and aVF
Suspected left main disease	STE is observed in aVR

TABLE 2. Performance of AMI classification

Model	AUROC	AUPRC	Accuracy	Sensitivity	Specificity	Precision	F1
Resnet18	0.983	0.987	0.94	0.917	0.964	0.966	0.941
Resnet34	0.979	0.987	0.944	0.922	0.969	0.971	0.946
Resnet50	0.98	0.986	0.944	0.917	0.974	0.976	0.946
ConvNextv2_tiny	0.963	0.962	0.925	0.899	0.954	0.956	0.927
ConvNextv2_small	0.938	0.962	0.915	0.867	0.969	0.969	0.915
ConvNextv2_base	0.905	0.939	0.872	0.757	0.999	0.999	0.862
Efficientnet_b0	0.971	0.981	0.935	0.908	0.964	0.966	0.936
Efficientnet_b1	0.978	0.985	0.935	0.904	0.969	0.97	0.936
Efficientnet_b2	0.978	0.986	0.947	0.913	0.985	0.985	0.948
Efficientnet_b3	0.976	0.983	0.925	0.894	0.959	0.961	0.926
Efficientnet_b4	0.971	0.977	0.935	0.89	0.985	0.985	0.935
Efficientnet_b5	0.977	0.985	0.944	0.922	0.969	0.971	0.946
Efficientnet_b6	0.952	0.966	0.896	0.844	0.954	0.953	0.895

TABLE 3. Performance of STEMI Classification

Class	AUROC	AUPRC	Accuracy	Sensitivity	Specificity	Precision	F1
STEMI	0.913	0.951	0.904	0.959	0.795	0.903	0.93

TABLE 4. Performance of infarction territory identification

Class	TP	FP	FN	TN	Accuracy	Sensitivity	Specificity	Precision	F1
Anterior	108	19	2	285	0.949	0.982	0.938	0.85	0.911
Lateral	38	21	6	349	0.935	0.864	0.943	0.644	0.738

Inferior	24	7	2	381	0.978	0.923	0.982	0.774	0.842
Suspected left main disease	13	10	2	389	0.971	0.867	0.975	0.565	0.684

TABLE 5. Performance of final classification (Control / STEMI / NSTEMI)

Class	AUROC	AUPRC	Accuracy	Sensitivity	Specificity	Precision	F1
Control	0.98	0.97	0.944	0.974	0.917	0.914	0.976
STEMI	0.948	0.874	0.913	0.862	0.941	0.887	0.927
NSTEMI	0.852	0.797	0.949	0.795	0.982	0.906	0.957
macro avg.	0.927	0.881	0.936	0.877	0.947	0.902	0.953

This is a provisional file, not the final typeset article