A deep learning algorithm for detecting acute myocardial infarction

Wen-Cheng Liu[1], MD; Chin-Sheng Lin[1], MD, PhD; Chien-Sung Tsai[2], MD; Tien-Ping Tsao[3], MD; Cheng-Chung Cheng[1], MD; Jun-Ting Liou[1], MD; Wei-Shiang Lin[1], MD; Shu-Meng Cheng[1], MD, PhD; Yu-Sheng Lou[4], MS; Chia-Cheng Lee[5,6], MD; Chin Lin[4,7,8]*, PhD

1. Division of Cardiology, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.; 2. Division of Cardiovascular Surgery, Department of Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.; 3. Division of Cardiology, Heart Centre, Cheng Hsin Hospital, Taipei, Taiwan, R.O.C.; 4. Graduate Institute of Life Sciences, National Defense Medical Center, Taipei, Taiwan, R.O.C.; 5. Planning and Management Office, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.; 6. Division of Colorectal Surgery, Department of Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.; 7. School of Medicine, National Defense Medical Center, Taipei, Taiwan, R.O.C.; 8. School of Public Health, National Defense Medical Center, Taipei, Taiwan, R.O.C.

Abstract

Background: Delayed diagnosis or misdiagnosis of acute myocardial infarction (AMI) is not unusual in daily practice. Since a 12-lead electrocardiogram (ECG) is crucial for the detection of AMI, a systematic algorithm to strengthen ECG interpretation may have important implications for improving diagnosis.

Aims: We aimed to develop a deep learning model (DLM) as a diagnostic support tool based on a 12-lead electrocardiogram.

Methods: This retrospective cohort study included 1,051/697 ECGs from 737/287 coronary angiogram (CAG)-validated STEMI/NSTEMI patients and 140,336 ECGs from 76,775 non-AMI patients at the emergency department. The DLM was trained and validated in 80% and 20% of these ECGs. A human-machine competition was conducted. The area under the receiver operating characteristic curve (AUC), sensitivity, and specificity were used to evaluate the performance of the DLM.

Results: The AUC of the DLM for STEMI detection was 0.976 in the human-machine competition, which was significantly better than that of the best physicians. Furthermore, the DLM independently demonstrated sufficient diagnostic capacity for STEMI detection (AUC=0.997; sensitivity, 98.4%; specificity, 96.9%). Regarding NSTEMI detection, the AUC of the combined DLM and conventional cardiac troponin I (cTnI) increased to 0.978, which was better than that of either the DLM (0.877) or cTnI (0.950).

Conclusions: The DLM may serve as a timely, objective and precise diagnostic decision support tool to assist emergency medical system-based networks and frontline physicians in detecting AMI and subsequently initiating reperfusion therapy.

*Corresponding author: No.161 Min-Chun E. Rd, Sec. 6, Neihu, Taipei 114, Taiwan, R.O.C.
E-mail: xup6fup0629@gmail.com

Abbreviations

ALT	alanine aminotransferase
AMI	acute myocardial infarction
AST	aspartate aminotransferase
AUC	area under the ROC curve
BMI	body mass index
CAD	coronary artery disease
CAG	coronary angiogram
CK	creatine kinase
CKD	chronic kidney disease
COPD	chronic obstructive pulmonary disease
Cr	creatinine
cTnI	conventional cardiac troponin I
DLM	deep learning model
DM	diabetes mellitus
ECG	electrocardiogram
ED	emergency department
eGFR	estimated glomerular filtration rate
GLU	glucose
Hb	haemoglobin
HDL	high-density lipoprotein cholesterol
HF	heart failure
hsTnI	high-sensitivity cardiac troponin I
HTN	hypertension
IRA	infarct-related artery
K	potassium
LAD	left anterior descending artery
LCx	left circumflex artery
LDL	low-density lipoprotein
LMCA	left main coronary artery
Na	sodium
NSTE-ACS	non-ST-segment elevation acute coronary syndrome
NSTEMI	non-ST-segment elevation myocardial infarction
PLT	platelet count
PRROC	precision-recall receiver operating characteristic curve
PTB	Physikalisch-Technische Bundesanstalt
RCA	right coronary artery
ROC	receiver operating characteristic
STEMI	ST-segment elevation myocardial infarction
TC	total cholesterol
TG	triglycerides
WBC	white blood cell count

Introduction

Acute myocardial infarction (AMI) remains a major public health issue despite global advances in diagnosis and management[1]. AMI refers to the evidence of acute myocardial injury detected by abnormal cardiac biomarkers with necrosis in a clinical setting consistent with myocardial ischaemia. The categories of ST-segment elevation myocardial infarction (STEMI) and non-ST-segment elevation acute coronary syndrome (NSTE-ACS) based on the presentation of a 12-lead electrocardiogram (ECG) have customarily been included in the concept of acute coronary syndrome (ACS)[2]. Patients with

symptoms suggestive of myocardial ischaemia and ST-segment elevation on the ECG require timely reperfusion therapy to reduce cardiac morbidity and mortality[3]. Likewise, patients with non-ST-segment elevation myocardial infarction (NSTEMI) considered to be in the very high or high risk categories require an immediate/early invasive strategy to prevent a worse prognosis[4].

However, prompt management depends on rapid recognition and precise diagnosis. Despite the established criteria for the diagnosis of AMI, it remains a critical challenge for emergency physicians to recognise rapidly. Previous studies reported that the rate of misdiagnosis of AMI at first medical contact ranged from 2 to 30%[5-7]. Failure to identify high-risk ECG findings in patients with AMI results in lower quality care and higher adverse outcomes. One of the leading causes of missed identification in the diagnostic process was incorrect interpretation of a diagnostic test[8,9]. Systematic processes to improve ECG interpretation may therefore have important implications for improving diagnosis. Since the principal diagnostic tool for AMI is a 12-lead ECG, a more detailed analysis of the ECG may significantly speed up this process.

The current artificial intelligence revolution that started with a deep learning model (DLM) has provided us with an unprecedented opportunity to improve the healthcare system, and it has been proven to be effective in medical applications[10-12]. Additionally, DLMs were confirmed to be superior to cardiologists in ECG interpretation when they were trained by large annotated ECG data sets[13,14]. To our knowledge, the available and applicable ECG databases of AMI were relatively small. Our study aimed to develop a DLM to detect AMI in a timely, objective and precise manner by a 12-lead ECG. More than 100,000 AMI-associated ECGs were recruited and learned by the DLM. Facilitated by the system's powerful computing ability, the performance of the trained model was compared with that of physicians, including cardiologists and an emergency physician. The diagnostic power for STEMI and NSTEMI by the DLM and conventional cardiac troponin I (cTnI) was also evaluated.

Methods

STUDY DESIGN

This was a single-centre, case-control study. The data were provided by the Tri-Service General Hospital, Taipei, Taiwan, and the retrospective design was ethically approved by the institutional review board (IRB No. 2-107-05-168). An electronic health system was built for collecting ECGs and medical records. The study period was from January 2012 to December 2018.

STUDY POPULATION

AMI patients presenting to the emergency department (ED) who received a coronary angiogram (CAG) to rule in type I AMI and to confirm the infarct-related artery (IRA) of STEMI were recruited[2]. AMI patients with no electronic ECG available, right side ECG, posterior ECG and pacemaker rhythm were excluded. Non-AMI patients presenting to the ED during the same period were recruited, while excluding those with a history of AMI or

any elevated cTnI during their ED stay. The definitions of AMI, STEMI, NSTEMI, non-AMI, and non-STEMI in this study are provided in **Supplementary Table 1** and **Supplementary Appendix 1**. The AMI cases were divided into development (80%) and validation (20%) cohorts by date. The ECGs in the development cohort were excluded from the validation cohort. There was no overlap of patients between these two cohorts.

ADJUDICATED FINAL DIAGNOSIS

Adjudication of the final diagnosis was performed by three board-certified interventional cardiologists who did not participate in the human-machine competition and who retrospectively and independently reviewed the AMI cases according to the clinical presentations, serial ECGs, serial cTnI levels and angiographic findings to make the final diagnosis of STEMI and NSTEMI, as recommended in the current guidelines[2-4]. In situations of disagreement about the diagnosis, cases were reviewed and adjudicated in consensus meetings.

DATA COLLECTION AND IMPLEMENTATION OF THE DLM

Data collection and DLM implementation are shown in **Supplementary Appendix 1** and **Supplementary Figure 1**. ECG recordings were collected using a Philips 12-lead ECG machine (PH080A), and the DLM was based on ECG12Net, which had previously been developed[14]. The output of the DLM was the probability of STEMI, NSTEMI, and non-AMI.

HUMAN-MACHINE COMPETITION

We evaluated the performance of participanting physicians using a competition set of 450 ECGs, which included 174 STEMI, 138 NSTEMI, and 138 non-AMI ECGs. The STEMI ECGs, based on the IRA, were further classified into the left main coronary artery (LMCA), left anterior descending artery (LAD), left circumflex artery (LCx), or right coronary artery (RCA). Five cardiologists and one emergency attending physician participated in the competition. In addition, the Philips 12-lead algorithm was also included to detect AMI in the competition[15]. The physicians had no access to any patient information and no knowledge of the data. Their responses were entered into an online standardised data entry program. We calculated the sensitivities, specificities, and kappa values to compare their results with those of the DLM.

STATISTICAL ANALYSIS

The study cohort was divided into training, validation, and competition sets. We presented their characteristics as the means and standard deviations, numbers of patients, or percentages where appropriate. They were compared using either the Student's t-test or the chi-square test, as appropriate. The statistical analysis was performed using R software version 3.4.4 (R Foundation for Statistical Computing, Vienna, Austria).

All analyses were based on ECGs but not patients. A significance level of p<0.05 was used throughout the analysis. The primary analysis was to evaluate the performance of the DLM, the

physicians and the Philips algorithm for STEMI detection in the human-machine competition. The receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) were applied to evaluate the competition results. We also used precision-recall ROC (PRROC) to evaluate the model performance in hypothetical real-world situations. Because the proportions of STEMI, NSTEMI, and non-AMI were distorted in the competition set, we re-weighted the samples based on the incidences in the real world (0.1%, 0.2%, and 99.7% of STEMI, NSTEMI, and non-AMI cases, respectively)[16-18]. The secondary analysis was performed on the whole validation cohort. We included more clinical information, such as patient characteristics and laboratory tests, to improve the model performance. A multivariable logistic regression model was used to integrate the DLM and clinical information. A series of logistic regression models identified the effects of different clinical information on the performance of STEMI and NSTEMI detection. The AUC was applied to evaluate the changes in model performance. The research interests, model comparison and statistical methods in this study are summarised in detail in **Supplementary Table 2**.

Results

BASELINE CHARACTERISTICS OF THE COHORTS

There were 1,051 ECGs before CAG from 737 STEMI patients, 697 ECGs before CAG from 287 NSTEMI patients and 140,336 ECGs from 76,775 non-AMI patients in this study. The development and validation cohorts included records from 58,056 and 19,743 patients, respectively. The characteristics and laboratory data are shown in **Supplementary Table 3**, and a detailed description is shown in **Supplementary Appendix 2**.

PREDICTION OF STEMI, NSTEMI AND NON-AMI

The results of the human-machine competition are summarised in **Figure 1**. The AUC of the DLM in the human-machine competition was 0.976 for STEMI detection, with a corresponding sensitivity and specificity of 89.7% and 94.6%, respectively. In contrast, the sensitivities and specificities for STEMI detection among the physicians and the Philips algorithm ranged from 60.5-92.6% and 76.0-97.5%, respectively, which were lower than those of the DLM. The PRROC analysis demonstrated the feasibility of an automatic ECG screening system, which revealed that the AUC of the DLM for STEMI detection was 0.586 in the hypothetical real world. The DLM achieved 63.2% precision and 50.3% recall using the appropriate cut-off point. These values were significantly better than those of all the physicians and the Philips algorithm.

Performance rankings and consistency analysis of STEMI detection among the DLM, the physicians and the Philips algorithm in the human-machine competition were carried out **(Figure 2)**. The DLM achieved the best global performance (kappa=0.645) **(Figure 2A)**, whereas the physicians had relatively better STEMI detection but poor discrimination of NSTEMI and non-AMI. The consistency analysis of AMI detection among the DLM, the physicians and the Philips algorithm is shown in the heatmap **(Figure 2B)**.
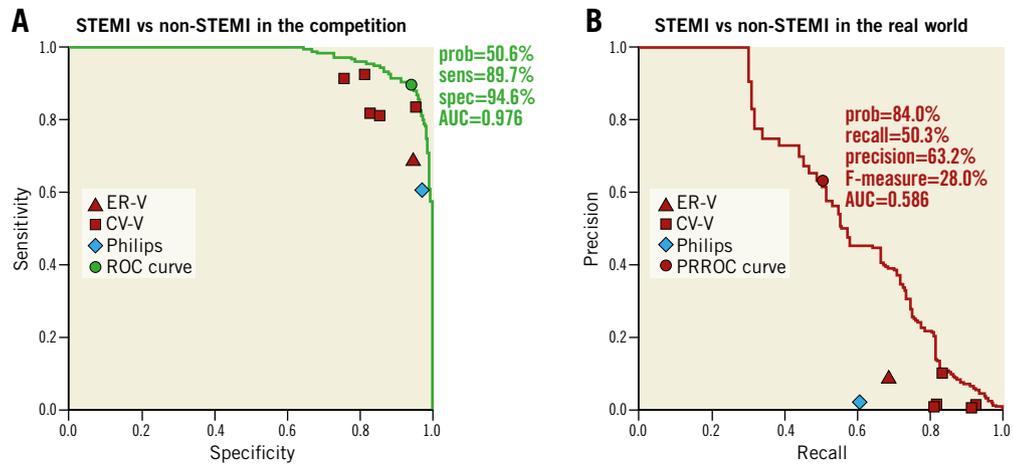
Figure 1. *Performance comparison for STEMI detection in the human-machine competition. The area under the receiver operating characteristic curve (AUC) was generated by the prediction of the DLM. The triangles, the square and the diamond denote the cardiologists, the emergency physician and the Philips algorithm, respectively. A) The ROC curve in the competition set (STEMI=174, NSTEMI=138, and non-AMI=138). B) The precision-recall ROC curve in the revised proportion of the hypothetical real world (STEMI=0.1%, NSTEMI=0.2%, and non-AMI=99.7%).*

ANALYSIS OF IRA OF STEMI

The DLM achieved the best global performance (kappa=0.629) for the IRA detection of STEMI **(Supplementary Figure 2)**. As shown in **Supplementary Figure 3**, after exclusion of LMCA and LCx, the AUC of the DLM for anterior STEMI detection was 0.975, with a corresponding sensitivity of 92.6%, which outperformed all participating physicians. Moreover, the AUC of the DLM in inferior STEMI detection was 0.974, with
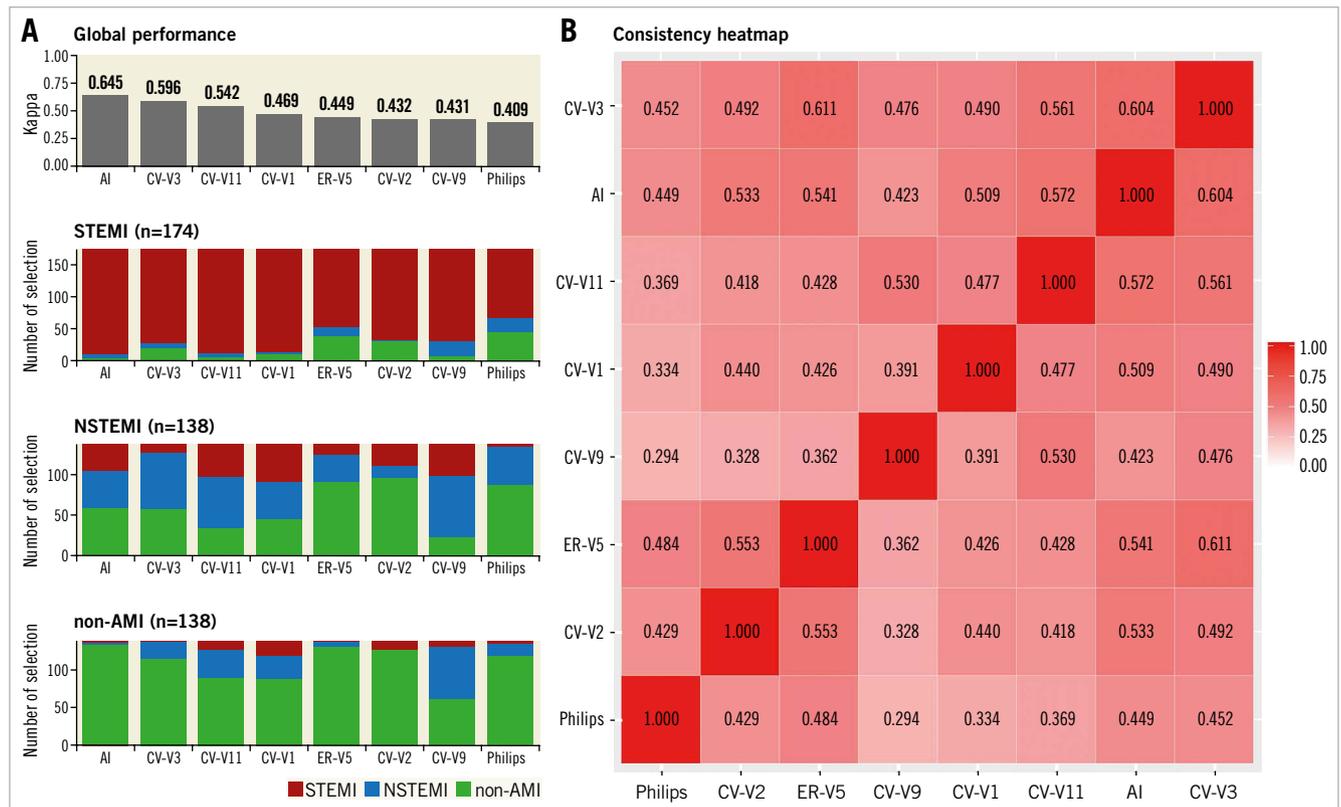
Figure 2. *Performance rankings and consistency analysis of STEMI detection among the DLM, the physicians and the Philips algorithm in the human-machine competition. A) Global performance rankings based on the class-3 kappa values. V(X) denotes (V) visiting staff with (X) years of experience. B) Consistency analysis as a heatmap coloured based on the values; the values in each cell were the kappa values of each pair.*

a corresponding sensitivity of 84.8%, which was better than all but one best physician. In the combined detection of anterior and inferior STEMI, the DLM had better performance than all physicians (AUC, 0.975; sensitivity, 89.4%).

INTERPRETATIONS OF STEMI ECGs BY THE DLM AND PHYSICIANS

Selected examples of STEMI ECGs in the human-machine competition are shown in **Figure 3**. A typical STEMI ECG with an IRA of the LAD **(Figure 3A)** was consistently detected by both the DLM and the physicians. One STEMI ECG with an IRA of the RCA **(Figure 3B)** was misdiagnosed by the DLM but correctly recognised by the best cardiologists. One STEMI ECG with an IRA of the RCA **(Figure 3C)** was misdiagnosed by both the DLM and the best cardiologists. The DLM correctly detected the ECG **(Figure 3D)** as STEMI with an IRA of the LAD, which was misdiagnosed by the best cardiologists.

Among the 138 NSTEMI ECGs, 58 ECGs were detected as non-AMI by the DLM, with an accuracy of 58.0%, which was worse than that of the best cardiologist (75.4%). This discrepancy was due to a more conservative AMI diagnostic strategy by the DLM. In contrast, among 138 non-AMI ECGs, the specificity of 96.4% of the DLM was much better than those of the two best cardiologists (82.6% and 64.5%). After adjustment for the specificities, the misdiagnosis of NSTEMI by the DLM was obviously less than that by the best cardiologists **(Table 1)**. Nevertheless, the DLM offered the best performance in AMI detection under the standardisation of the best cardiologists. The ECG lead-specific analyses for the detection of STEMI and the corresponding IRA are shown in **Supplementary Figure 4**, and a detailed description is shown in **Supplementary Appendix 2**.

LOGISTIC REGRESSION ANALYSIS OF STEMI AND NSTEMI

Univariate and multivariate logistic regression analyses in the development cohort revealed that male sex, prior CAD, cTnI, haemoglobin, total cholesterol and low-density lipoprotein (LDL) levels were independent risk factors for STEMI and NSTEMI detection **(Supplementary Figure 5)**.

DIAGNOSTIC VALUE ANALYSIS

We evaluated the performance of the DLM after adjusting for significant patient characteristics, disease histories, and laboratory data
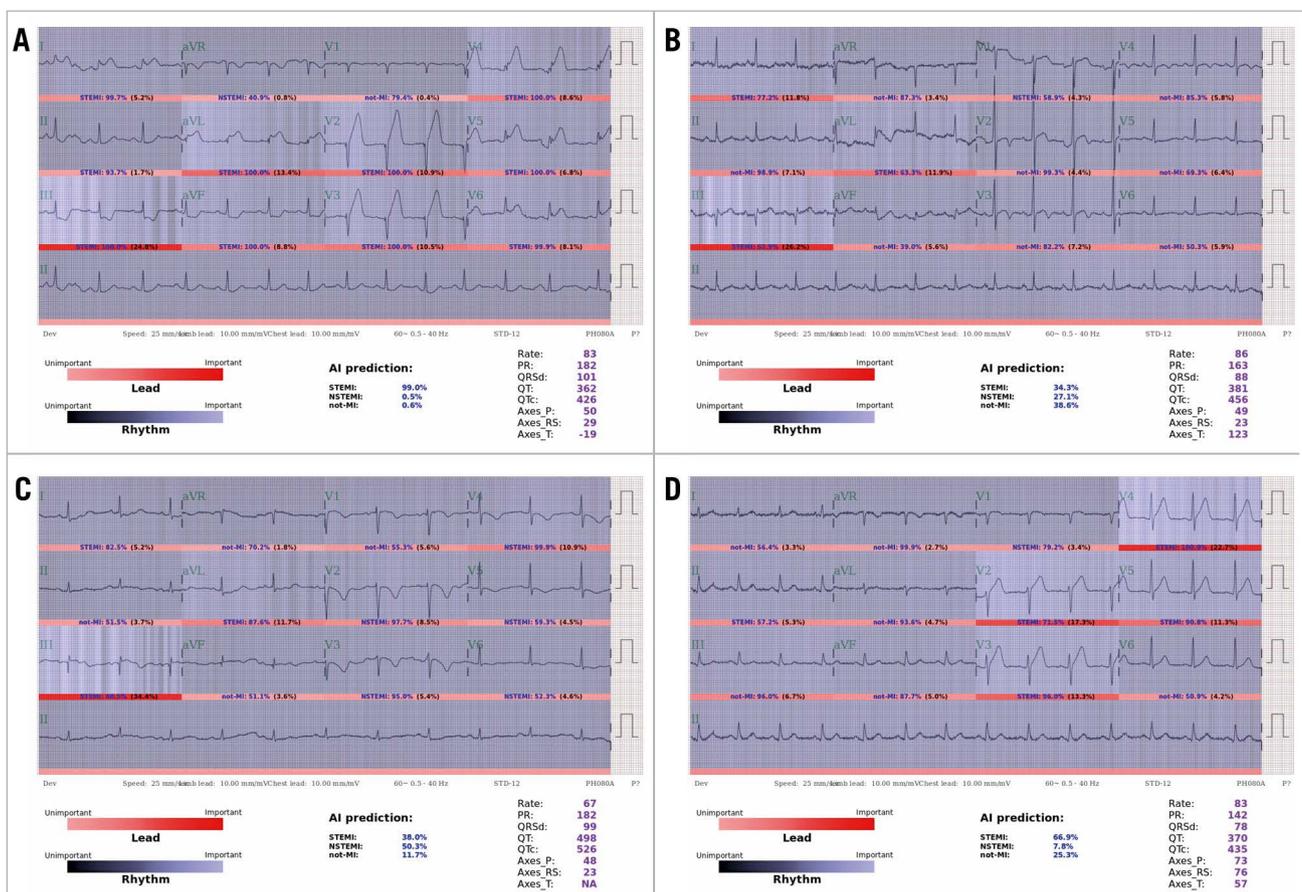
Figure 3. *Interpretations of selected STEMI ECGs by the DLM and physicians in the human-machine competition. A) Both the DLM and the best cardiologists consistently detected STEMI. B) The DLM misdetected STEMI, which was correctly detected by the best cardiologists. C) Both the DLM and the best cardiologists misdetected STEMI. D) The DLM correctly detected STEMI, which was misdetected by the best cardiologists.*

Table 1. Maximum sensitivity of the DLM for a specific specificity.

	Revised item[a]	Sensitivity[b] (STEMI)	Sensitivity[c] (NSTEMI)	Specificity[d]
DLM (original)	0.000	164/174 (94.3%)	80/138 (58.0%)	133/138 (96.4%)
CV-V3		146/174 (83.9%)	80/138 (58.0%)	114/138 (82.6%)
DLM (specificity=82.6%)	0.450	166/174 (95.4%)	108/138 (78.3%)	114/138 (82.6%)
CV-V11		162/174 (93.1%)	104/138 (75.4%)	89/138 (64.5%)
DLM (specificity=64.5%)	0.612	166/174 (95.4%)	123/138 (89.1%)	89/138 (64.5%)

[a] The revised item was used to modify the probability of non-AMI given by DLM. For example, if an original probability of STEMI/NSTEMI/non-AMI was 0.220/0.310/0.470, then the prediction was defined as non-AMI according to the largest probability. However, the revised item was used to make DLM become more sensitive, which was used to modify the probability of non-AMI as 0.470-0.450=0.020 as the first situation. Therefore, the new prediction of this case was defined as NSTEMI according to the largest revised probability (0.220/0.310/0.020). [b] The sensitivity of STEMI was defined as the percentage of STEMI that was correctly identified as STEMI. [c] The sensitivity of NSTEMI was defined as the percentage of NSTEMI that was correctly identified as NSTEMI. [d] The specificity is defined as the percentage of non-AMI cases that was correctly identified as non-AMI.

to ensure consistency across a wide range of putative confounding variables in the validation cohort. The DLM had significantly better performance than cTnI in detecting STEMI, with an AUC of 0.997 with a corresponding sensitivity and specificity of 98.4% and 96.9%, respectively **(Figure 4A)**. However, cTnI had significantly better performance than the DLM in detecting NSTEMI. The AUC for NSTEMI detection by the combination of the DLM and the first recorded cTnI increased to 0.978, with a corresponding sensitivity and specificity of 91.6% and 96.7%, respectively, which was better than that of the DLM (0.877) or cTnI (0.950) individually **(Figure 4B)**. Using the DLM independently was sufficient to detect STEMI, and the addition of patient characteristics did not significantly improve its performance. However, cTnI was found to improve the diagnostic accuracy for NSTEMI better than any additional characteristics **(Supplementary Figure 6)**.

Discussion

In this study, we established a DLM to detect STEMI precisely through ECG analysis, which applied a deep convolutional network to extract notable ECG features with a development cohort of more than 100,000 ECGs. All AMI patients were validated by CAG, and the corresponding IRA of STEMI was identified. Most importantly, our DLM performed better than the physicians in STEMI detection, with a high sensitivity of 89.7% and specificity of 94.6%.

The application of deep learning technology in the cardiovascular field for arrhythmias, dyskalaemia, and valvular heart disease has recently grown in popularity[13,14,19-21]. However, no large-scale study has been designed for AMI detection. Previous DLMs for AMI detection by a 12-lead ECG mainly used the Physikalisch-Technische Bundesanstalt (PTB) diagnostic ECG database[22,23]. These studies may be limited because they were not further validated. Moreover, comparisons between the DLM and physicians were lacking. In comparison with previous studies, we enrolled the largest number of clinically validated ECGs for development and validation. Additionally, we further confirmed the role of cTnI in assisting with NSTEMI detection by our DLM. All these results highlight the strengths of the current study.

The sensitivity and specificity for STEMI detection by the DLM were better than those of the physicians. ECG is the timeliest tool among all objective detection methods for AMI. However,
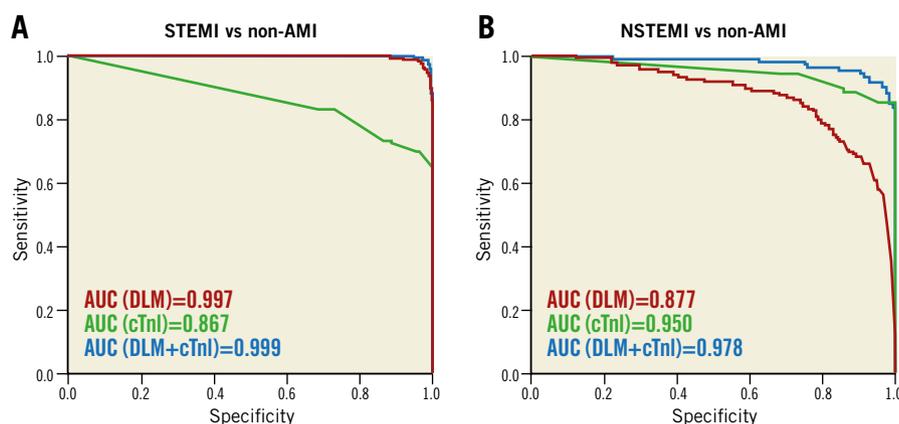
Figure 4. *Comparison of the diagnostic value between the DLM and cTnI in the validation cohort. The area under the receiver operating characteristic curve (AUC) was generated from the logistic regression analysis using the validation cohort. The p-values represent the comparison among the DLM, cTnI and the DLM plus cTnI. A) Regarding STEMI detection: DLM vs cTnI, p<0.001; DLM vs DLM+cTnI, p=ns. B) Regarding NSTEMI detection: cTnI vs DLM, p<0.01; cTnI+DLM vs cTnI, p<0.05.*

the low sensitivity and the disagreement in interpreting ECGs between physicians remain issues. The sensitivity of subjective interpretation for AMI detection using a 12-lead ECG ranged only from 61 to 74%, with a specificity ranging from 72 to 89%[24-26]. In contrast, previous prehospital computer algorithm interpretations for STEMI had a sensitivity of approximately 69%[27,28]. Our DLM provided extraordinary performance that could support decision-making systems in clinical practice.

The DLM could objectively identify STEMI based on analysing and learning a large number of ECGs. Moreover, subtle ECG changes in the earliest phase of STEMI, which are easily missed by physicians, could be correctly recognised by the DLM. Nevertheless, prior MI or cardiomyopathy might mislead the DLM owing to baseline ST-T changes. Therefore, information regarding previously available ECGs or the history of cardiovascular disease may be needed to strengthen the capacity of the DLM for STEMI detection further.

The performance of our DLM on the detection of STEMI equivalents and STEMI mimics was further evaluated. STEMI equivalents, including de Winter sign, Wellens' syndrome, hyperacute T-waves, ST elevation in the lead aVR with diffuse ST depression, ST elevation in the presence of bundle branch block, and posterior wall AMI, representing coronary occlusion without meeting the traditional ST elevation criteria, were crucial for timely recognition[29,30]. Additionally, high take-off T presentations, such as hyperkalaemia,

benign early repolarisation, left ventricular hypertrophy and Brugada syndrome, which mimick STEMI, were usually misdiagnosed, leading to false initiation of primary PCI[31,32]. Our study demonstrated that the DLM exhibited excellent diagnostic power in the detection of STEMI equivalents (except for type 1 Wellens' syndrome) and provided extraordinary differentiating capacity in the detection of high take-off T **(Supplementary Figure 7, Supplementary Figure 8)**. Further prospective and large ECG validation data sets are needed to confirm the discriminating abilities of the DLM.

Our DLM has several potential clinical applications. First, the DLM could provide decision support and a high-risk alarm system for AMI that could help to reduce medical errors in the ED resulting from intense time pressures or heavy workloads and harried staff during busy working hours. Second, the DLM could be incorporated into ECG machines in ambulances to facilitate telemedicine and shorten the decision time before initiation of reperfusion therapies. Third, our DLM could be applied in rural and remote areas and places lacking experts to facilitate ECG interpretation and promote diagnostic accuracy, thereby initiating timely management and improving the prognosis of STEMI patients. Finally, the DLM could be incorporated into a wearable device for AMI detection, especially for patients with an extremely high risk of atherosclerotic cardiovascular disease. Accordingly, our DLM exhibits diagnostic benefits and may improve the quality of health care in the near future **(Central illustration)**.
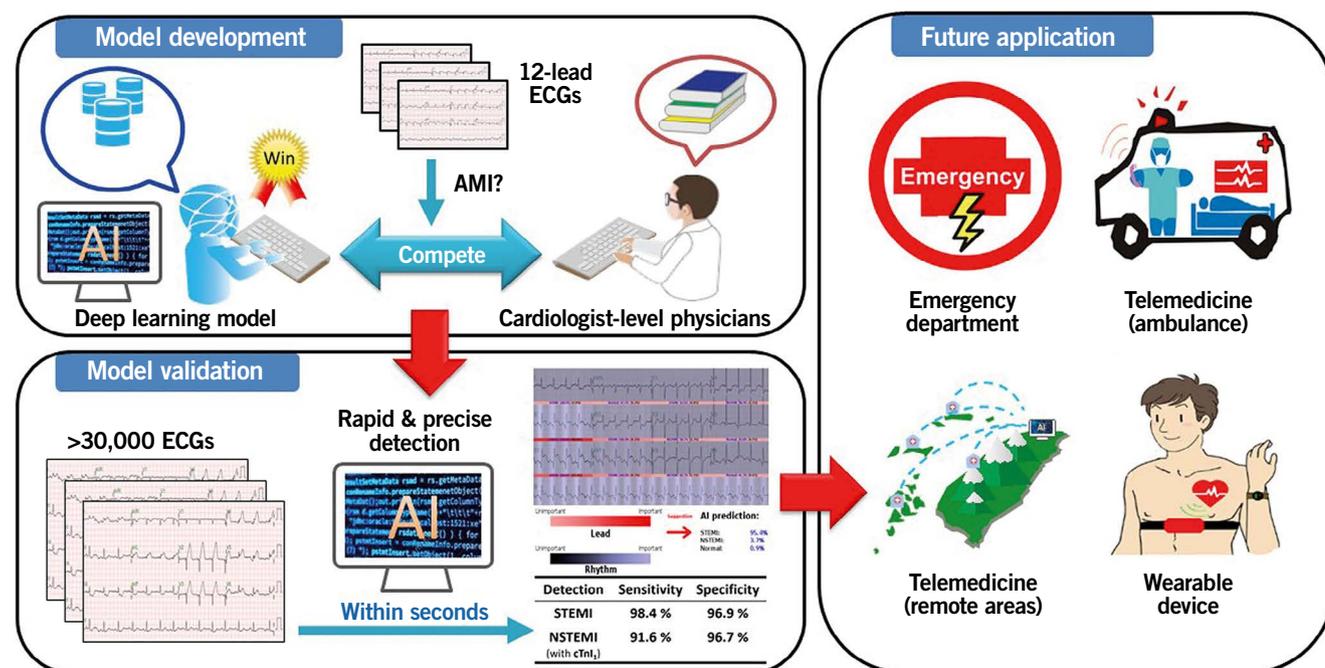
Central illustration. *Schematic diagram of the development, validation and future application of the current deep learning model for detecting AMI. The DLM learned from more than 100,000 ECGs was developed and trained. Compared with cardiologist-level physicians. The DLM exhibited the best performance in the detection of STEMI. The validated model achieved excellent diagnostic power with a sensitivity of 98.4% and a specificity of 96.9% for STEMI detection. With the ability of real-time detection, precise diagnosis and early alarm, the application of DLM for STEMI detection, including in-hospital, pre-hospital settings, telemedicine and wearable devices, would improve the quality of health care of cardiovascular disease in the near future.*

Limitations

Some limitations of this study should be noted. First, the human-machine competition was based on a well-designed retrospective study. A real-world prospective study should be conducted to verify the clinical impact of the DLM. Second, only six attending physicians participated in the competition with the DLM. Although their performance in AMI detection was relatively consistent with that in previous studies, comparisons should be made with more physicians to confirm the superiority of the DLM[33]. Third, the studied patients were enrolled from only one academic medical centre, although the diagnosis and management of AMI was based on the guidelines. Multicentre validation is needed to confirm the value and application of this study. Fourth, there were fewer NSTEMI cases than STEMI cases, which may limit the capacity for NSTEMI detection by our DLM. Fifth, during the study period, cTnI rather than hsTnI was used for AMI diagnosis. Sixth, information on prior ECGs to improve diagnostic performance was not available in our DLM system. Seventh, the impacts of coronary collateral flow on ST-T changes during AMI and the performance of the DLM in the detection of STEMI were not analysed. Finally, only patients in the ED were enrolled, which may have led to selection bias and constrained the generalisability of the results.

Conclusions

We established an optimal DLM to detect STEMI based on a 12-lead ECG with better accuracy than physicians. Integration of a DLM may assist frontline physicians in recognising AMI in a timely and precise manner to prevent delayed diagnosis or misdiagnosis of AMI and thereby provide prompt reperfusion therapy. Further prospective validation with pre-hospital and in-hospital ECG tests is needed to confirm the performance of our DLM.

Impact on daily practice

STEMI can now be recognised using this cardiologist-level algorithm, achieving real-time STEMI diagnosis and early alarms. A comprehensive ecosystem has been established including in-hospital, pre-hospital and wearable devices, improving the quality of care in AMI.

Funding

Conflict of interest statement

The authors have no conflicts of interest to declare.

References

1. GBD 2017 Congenital Heart Disease Collaborators. Global, regional, and national burden of congenital heart disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Child Adolesc Health.* 2020;4:185-200.

2. Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA, White HD; ESC Scientific Document Group. Fourth universal definition of myocardial infarction (2018). *Eur Heart J.* 2019;40:237-269.

3. Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, Caforio ALP, Crea F, Goudevenos JA, Halvorsen S, Hindricks G, Kastrati A, Lenzen MJ, Prescott E, Roffi M, Valgimigli M, Varenhorst C, Vranckx P, Widimsky P; ESC Scientific Document Group. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J.* 2018;39:119-77.

4. Collet JP, Thiele H, Barbato E, Barthélémy O, Bauersachs J, Bhatt DL, Dendale P, Dorobantu M, Edvardsen T, Folliguet T, Gale CP, Gilard M, Jobs A, Jüni P, Lambrinou E, Lewis BS, Mehilli J, Meliga E, Merkely B, Mueller C, Roffi M, Rutten FH, Sibbing D, Siontis GCM; ESC Scientific Document Group. 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. *Eur Heart J.* 2021;42:1289-367.

5. McCarthy BD, Beshansky JR, D'Agostino RB, Selker HP. Missed diagnoses of acute myocardial infarction in the emergency department: results from a multicenter study. *Ann Emerg Med.* 1993;22:579-82.

6. Masoudi FA, Magid DJ, Vinson DR, Tricomi AJ, Lyons EE, Crounse L, Ho PM, Peterson PN, Rumsfeld JS; Emergency Department Quality in Myocardial Infarction Study Investigators. Implications of the failure to identify high-risk electrocardiogram findings for the quality of care of patients with acute myocardial infarction: results of the Emergency Department Quality in Myocardial Infarction (EDQMI) study. *Circulation.* 2006;114:1565-71.

7. Wu J, Gale CP, Hall M, Dondo TB, Metcalfe E, Oliver G, Batin PD, Hemingway H, Timmis A, West RM. Editor's Choice - Impact of initial hospital diagnosis on mortality for acute myocardial infarction: A national cohort study. *Eur Heart J Acute Cardiovasc Care.* 2018;7:139-48.

8. Kachalia A, Gandhi TK, Puopolo AL, Yoon C, Thomas EJ, Griffey R, Brennan TA, Studdert DM. Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers. *Ann Emerg Med.* 2007;49: 196-205.

9. Schenkel S. Promoting patient safety and preventing medical error in emergency departments. *Acad Emerg Med.* 2000;7:1204-22.

10. Estava A, Kuprel B, Novoa R, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist level classification of skin cancer with deep neural networks. *Nature.* 2017;542: 115-8.

11. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88.

12. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA.* 2016;316: 2402-10.

13. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med.* 2019;25:65-9.

14. Lin CS, Lin C, Fang WH, Hsu CJ, Chen SJ, Huang KH, Lin WS, Tsai CS, Kuo CC, Chau T, Yang SJ, Lin SH. A Deep-Learning Algorithm (ECG12Net) for Detecting Hypokalemia and Hyperkalemia by Electrocardiography: Algorithm Development. *JMIR Med Inform.* 2020;8:e15931.

15. The Philips 12-lead Algorithm Physician's Guide. Available at: http://incenter. medical.philips.com/doclib/enc/fetch/577817/577818/12-Lead_Algorithm_ Physician_s_Guide_for_Algorithm_Verion_PH080A%2C_(ENG).pdf%3Fnodeid%3 D3325283%26vernum%3D-2. Last accessed 17/08/2021.

16. Widimsky P, Wijns W, Fajadet J, De Belder M, Knot J, Aaberge L, Andrikopoulos G, Baz JA, Betriu A, Claeys M, Danchin N, Djambazov S, Erne P, Hartikainen J, Huber K, Kala P, Klinceva M, Kristensen SD, Ludman P, Ferre JM, Merkely B, Milicic D, Morais J, Noc M, Opolski G, Ostojic M, Radovanovic D, De Servi S, Stenestrand U, Studencan M, Tubaro M, Vasiljevic Z, Weidinger F, Witkowski A, Zeymer U; European Association for Percutaneous Cardiovascular Interventions. Reperfusion therapy for ST elevation acute myocardial infarction in Europe: description of the current situation in 30 countries. *Eur Heart J.* 2010;31:943-57.

17. O'Gara PT, Kushner FG, Ascheim DD, Casey DE Jr, Chung MK, de Lemos JA, Ettinger SM, Fang JC, Fesmire FM, Franklin BA, Granger CB, Krumholz HM,

Linderbaum JA, Morrow DA, Newby LK, Ornato JP, Ou N, Radford MJ, Tamis-Holland JE, Tommaso CL, Tracy CM, Woo YJ, Zhao DX. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2013;61:e78-140.

18. Jernberg T. Swedeheart annual report 2015. Karolinska University Hospital, Huddinge 2016. Available at: https://www.ucr.uu.se/swedeheart/dokument-sh/arsrapporter-sh/aeldre-arsrapporter-older-reports/arsrapport-2015/swedeheart-annual-report-2015

19. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM, Asirvatham SJ, Scott CG, Carter RE, Friedman PA. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nat Med.* 2019;25:70-4.

20. Kwon JM, Lee SY, Jeon KH, Lee Y, Kim KH, Park J, Oh BH, Lee MM. Deep Learning–Based Algorithm for Detecting Aortic Stenosis Using Electrocardiography. *J Am Heart Assoc.* 2020;9:e014717.

21. Kwon JM, Kim KH, Akkus Z, Jeon KH, Park J, Oh BH. Artificial intelligence for detecting mitral regurgitation using electrocardiography. *J Electrocardiol.* 2020;59:151-7.

22. Gupta A, Huerta E, Zhao Z, Moussa I. Deep Learning for Cardiologist-Level Myocardial Infarction Detection in Electrocardiograms. In: Jarm T, Cvetkoska A, Mahnic-Kalamiza S, Miklavcic D (eds). 8th European Medical and Biological Engineering Conference. EMBEC 2020. IFMBE Proceedings, vol 80. Springer, Cham. https://doi.org/10.1007/978-3-030-64610-3_40

23. Strodthoff N, Strodthoff C. Detecting and interpreting myocardial infarction using fully convolutional neural networks. *Physiol Meas.* 2019;40:015001.

24. Asch FM, Shah S, Rattin C, Swaminathan S, Fuisz A, Lindsay J. Lack of sensitivity of the electrocardiogram for detection of old myocardial infarction: a cardiac magnetic resonance imaging study. *Am Heart J.* 2006;152:742-8.

25. Trägårdh E, Claesson M, Wagner GS, Zhou S, Pahlm O. Detection of acute myocardial infarction using the 12-lead ECG plus inverted leads versus the 16-lead ECG (with additional posterior and right-sided chest electrodes). *Clin Physiol Funct Imaging.* 2007;27:368-74.

26. McCabe JM, Armstrong EJ, Ku I, Kulkarni A, Hoffmayer KS, Bhave PD, Waldo SW, Hsue P, Stein JC, Marcus GM, Kinlay S, Ganz P. Physician accuracy in interpreting potential ST-segment elevation myocardial infarction electrocardiograms. *J Am Heart Assoc.* 2013;2:e000268.

27. Garvey JL, Zegre-Hemsey J, Gregg R, Studnek JR. Electrocardiographic diagnosis of ST segment elevation myocardial infarction: an evaluation of three automated interpretation algorithms. *J Electrocardiol.* 2016;49:728-32.

28. de Champlain F, Boothroyd LJ, Vadeboncoeur A, Huynh T, Nguyen V, Eisenberg MJ, Joseph L, Boivin JF, Segal E. Computerized interpretation of the pre-hospital electrocardiogram: predictive value for ST segment elevation myocardial infarction and impact on on-scene time. *CJEM.* 2014;16:94-105.

29. Lawner BJ, Nable JV, Mattu A. Novel patterns of ischemia and STEMI equivalents. *Cardiol Clin.* 2012;30:591-9.

30. Tzimas G, Antiochos P, Monney P, Eeckhout E, Meier D, Fournier S, Harbaoui B, Muller O, Schläpfer J. Atypical Electrocardiographic Presentations in Need of Primary Percutaneous Coronary Intervention. *Am J Cardiol.* 2019;124:1305-14.

31. Somers MP, Brady WJ, Perron AD, Mattu A. The prominent T wave: electrocardiographic differential diagnosis. *Am J Emerg Med.* 2002;20:243-51.

32. Wang K, Asinger RW, Marriott HJ. ST-segment elevation in conditions other than acute myocardial infarction. *N Engl J Med.* 2003;349:2128-35.

33. Willems JL, Abreu-Lima C, Arnaud P, van Bemmel JH, Brohet C, Degani R, Denis B, Gehring J, Graham I, van Herpen G, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med.* 1991;325:1767-73.

Supplementary data

Supplementary Appendix 1. Methods including: the definition of AMI, STEMI, NSTEMI, non-AMI and non-STEMI; data collection; timelines of door-to-balloon/CAG, door-to-ECG, and ECG-to-balloon/CAG; implementation of the deep learning model (DLM); implementation details of the DLM; training details; data augmentation; model visualisation and summary of the research interests, model comparison and statistical methods.

Supplementary Appendix 2. Results including: the baseline characteristics of the cohorts, ECG lead-specific analysis, and discussion.

Supplementary Figure 1. Architecture of the DLM.

Supplementary Figure 2. Performance rankings of infarct-related artery detection of STEMI among DLM, physicians and the Philips algorithm in the human-machine competition.

Supplementary Figure 3. Performance comparison for anterior (LAD), inferior (RCA), and combined anterior and inferior (LAD+RCA) STEMI detection in the human-machine competition.

Supplementary Figure 4. ECG lead-specific analyses for the detection of STEMI, STEMI-LAD and STEMI-RCA.

Supplementary Figure 5. Univariate and multivariate logistic regression analysis of STEMI, and NSTEMI in the development cohort.

Supplementary Figure 6. Comparison of the diagnostic value among additional demographic variables, cTnI and DLM in the validation cohort.

Supplementary Figure 7. The test examples of the detection of STEMI equivalents by the DLM.

Supplementary Figure 8. The test examples of the detection of high take-off T ECG by the DLM.

Supplementary Table 1. The definitions of AMI, STEMI, NSTEMI, non-AMI and non-STEMI.

Supplementary Table 2. The research interests, model comparison and statistical methods.

Supplementary Table 3. Corresponding patient characteristics and laboratory results of STEMI, NSTEMI, and non-AMI ECGs in the development and validation cohorts.

The supplementary data are published online at:
https://eurointervention.pcronline.com/
doi/10.4244/EIJ-D-20-01155

Supplementary data

Supplementary Appendix 1. Methods

The definitions of AMI, STEMI, NSTEMI, non-AMI and non-STEMI [2-4]

The definitions of AMI, STEMI, NSTEMI, non-AMI, and non-STEMI in this study are summarised in detail in

Supplementary Table 1.

Data collection

ECG recordings were collected using a Philips 12-lead ECG machine (PH080A). The ECG signal was recorded

in a digital format. The sampling frequency was 500 Hz, with 10 seconds recorded in each lead. Patient

characteristics and laboratory tests were collected from our electronic medical records. The laboratory data

collected closest to the time of the ECG were assigned to each ECG record. Because the ECG records were

sometimes conducted within a relatively short time period, some ECGs from the same patients shared the same

patient characteristics and laboratory data.

Timelines of door-to-balloon/CAG, door-to-ECG, and ECG-to-balloon/CAG

In STEMI patients without cardiac arrest, endotracheal intubation, or mechanical support, the mean door-to-

balloon time was 65.7 min, the mean door-to-ECG time was 3.9 min, and the mean ECG-to-balloon time was

60.8 min. In STEMI patients with cardiac arrest, endotracheal tube intubation, or mechanical support, the mean

door-to-balloon time was 205.6 min, the mean door-to-ECG time was 25.9 min, and the mean ECG-to-balloon

time was 178.7 min. In NSTEMI patients, the mean door-to-CAG time was 629.6 min, the mean door-to-ECG

time was 6.7 min, and the mean ECG-to-CAG time was 622 min.

Implementation of the deep learning model (DLM)

We developed a DLM with 82 convolutional layers and an attention mechanism. The technology details, such

as the model architecture, data augmentation, and model visualisation, have been described previously [14]. We

used the same architecture to train two new DLMs for AMI detection and infarct-related artery (IRA) analysis

of STEMI. The first DLM was trained via full samples with three categories, including STEMI, NSTEMI, and non-AMI, and the output of this model was a class-3 softmax output. The second DLM was trained via STEMI ECGs, and the output of this model was a class-4 softmax output for IRA analysis.

The standard input format of the DLM was a length of 1,024 numeric sequences, but the original length of our 12-lead ECG signals was 5,000. In the training process, we randomly cropped a length of 1,024 sequences as input. For the inference stage, nine overlapping lengths of 1,024 sequences based on interval sampling were used to generate a prediction and were averaged as the final prediction. Due to the scarcity of AMI cases in our study, an oversampling process was implemented to ensure that rare samples were adequately recognised. The settings for the training model were as follows: (1) Adam optimiser with standard parameters ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) and a batch size of 36 for optimisation; (2) a learning rate of 0.001; and (3) a weight decay of 10^{-4}. The 100th epoch model was used as the final model, and the presented performance in the validation set was only evaluated once.

Implementation details of the DLM

The DLM architecture

The architecture of our DLM was based on ECG12Net, which was previously used for serum K+ concentration estimation [14]. Supposing that a standard 12-lead ECG signal comprised 12 sequences of N numbers (N = 1,250 in our database), the ECG signal sequence X = [x1,1, x1,2, …, x1,N; x2,1, x2,2, …, x2,N; …; x12,1, x12,2, …, x12,N] was used as the input, and the output was a one-hot encoder of AMI categories (STEMI, NSTEMI, and non-AMI) and the IRA of STEMI (STEMI-LMCA, STEMI-LAD, STEMI-LCx, and STEMI-RCA).

For example, a label of STEMI is encoded as [1,0,0], and a label of NSTEMI is encoded as [0,1,0]. Each output label corresponded to a segment of the input. Because the ECG information was mostly provided by morphologic changes with shift invariance, convolutional layers with weight sharing were used to adapt to this

situation and reduce the hazard of overfitting. We therefore developed a 12-channel sequence-to-sequence model to conduct this task as a revision of DenseNet. The complete architecture of the DLM is shown in **Supplementary Figure 1**. We defined a "dense unit" as a neural combination as follows: (1) a batch normalisation layer to normalise input data, (2) a rectified linear unit (ReLU) layer for non-linearisation, (3) a 1×1 convolution layer with 4K filters to reduce the dimensions of the data, (4) a batch normalisation layer for normalisation, (5) a ReLU layer for non-linearisation, (6) a 3×1 convolution layer with 4K filters to extract features, (7) a batch normalisation layer for normalisation, (8) a ReLU layer for non-linearisation, and (9) a 1×1 convolution layer with K filters to extract features. K was a model constant that was set at 32 in all our experiments. After using a dense unit to extract features, we used the dense connectivity resulting from direct connections from any layer to all subsequent layers to build a "dense block". We designed a model with five dense blocks comprising 3,3,6,6, and three dense units.

Dense blocks cannot be concatenated when the size of the feature maps changes. Thus, a pooling block was used to concatenate each dense block for downsampling in our architecture. This block included a dense unit with a 2×1 stride and an average pooling layer with a 2×1 kernel size and stride, which was used for downsampling. Each dense block was concatenated by the pooling block to integrate the features of the previous blocks.

A length of 864 numeral sequences was used as the input in our experiment. We designed an ECG lead block with 80 trainable layers, the architecture of which is shown in **Supplementary Figure 1A**. The input data were passed through a batch normalisation layer, followed by a convolution layer, another batch normalisation layer, a ReLU layer, and a pooling layer. The initial convolution layer comprised K convolution filters with a kernel size of 7×1 and a stride of 2×1. Next, the data were passed through a series of dense blocks and a pooling block, resulting in a $16\times1\times864$ array. A ReLU layer, a batch normalisation layer, and a global pooling layer were followed by the last dense block. Finally, a fully connected layer with k output was created for follow-up use, where k is the number of categories, which was equal to 3 in the first AMI detection model and 4 in the second

IRA analysis model of STEMI. This ECG lead block was used to extract 864 features from each ECG lead, making a basic output prediction based on each lead. **Supplementary Figure 1B** shows how ECG12Net integrated all the information from the ECG to make an overall prediction. ECG12Net comprised 12 ECG lead blocks corresponding to lead sequences. We designed an attention mechanism based on a hierarchical attention network to concatenate these blocks, increasing the interpretive power of ECG12Net. The attention block comprised a batch normalisation layer followed by a fully connected layer and then two combinations of a batch normalisation layer, a ReLU layer, and a fully connected layer. The first and second fully connected layers each contained $8/k$ neurons. Attention scores were calculated for each ECG lead and then integrated for standardisation by a linear output layer. The standardised attention scores were used to weight the 12 ECG lead outputs by simple multiplication. The 12 weighted outputs were summed and converted into a softmax output layer to provide the final prediction value. The above model using ECG information was named ECG12Net, which contained 82 trainable layers. The m-log-loss function was used to calculate model loss. A dropout layer was added only in the fully connected layer, and the dropout rate was set to 0.5.

Training details

The 12-lead ECG signal sequences were first trained by the 12 ECG leads separately. Due to the seriously uneven distribution in STEMI, NSTEMI, and non-AMI, an oversampling process was implemented to improve performance by ensuring that rare samples were adequately recognised. We sampled 12 STEMI ECGs, 12 NSTEMI ECGs, and 12 non-AMI ECGs in each batch. This process sufficiently considered rare STEMI and NSTEMI cases so as not to be skewed by the overwhelming number of normal cases. We used the software package MXNet version 1.3.0 to implement ECG12Net. The settings used for the training model were as follows: (1) Adam optimiser with standard parameters ($\beta_1 = 0{:}9$ and $\beta_2 = 0{:}999$) and a batch size of 36 for optimisation; (2) initial learning rate set at 0.001 and lowered by 10 three times when validation loss plateaued after an epoch; and (3) a weight decay of 10^{-4}. Because the sampling rate of our machine is 500 Hz, our 12-lead ECG signal includes 12 numeral sequences with 5,000 digits. However, the standard input format of ECG12Net was a length of 1,024 numeric sequences. We randomly cropped a length of 1,024 sequences as input in the

training process. During the inference stage, the nine overlapping lengths of 1,024 sequences based on interval sampling (X_1 to X_{1024}, X_{498} to X_{1521}, X_{995} to X_{2018}, X_{1492} to X_{2515}, X_{1989} to X_{3012}, X_{2486} to X_{3509}, X_{2983} to X_{4006}, X_{3480} to X_{4503}, and X_{3977} to X_{5000}) were used to generate predictions and averaged as the final prediction. The 100th epoch model was used as the final model, and the model performance in the validation set was verified only once.

Data augmentation

A previous study reported severe overfitting in an atrial fibrillation detection task and suggested a series of data augmentations to improve model performance. In the current study, the problem of overfitting was due to the large number of parameters in the deep learning architecture (~3 million trainable parameters) relative to the sample size. The first step in tackling this issue was to resize the sequence length by adjusting heart rate. We randomly resampled a broader range of heart rates in a uniform distribution from 0.8 *HR* to 1.2 *HR*, where *HR* was the original heart rate for each sample. The second step was to randomly crop a length of 1,024 sequences as input. The third step was to add a random variable drawn from a Gaussian distribution with a mean of 0 and a standard deviation of 0.1. Fourth, time points were selected uniformly and at random, and the ECG signal values within a 50 ms vicinity of these points were set at 0. This method was called dropout burst. Finally, we set six random ECG lead sequences to 0 in the combined training step. We observed that the final DLM only used information from a few ECG leads to make a prediction and inferred that the model had ceased to learn features from the other ECG leads because it had perfectly predicted all the data in the training set. This approach forced the DLM to learn all the abnormal ECG leads.

Model visualisation

To interpret the network predictions, we conducted heatmaps to visualise the ECG rhythms and leads using class activation mappings (CAMs) and attention mechanisms based on the global average pooling (GAP) architecture in the last network, which was used at the end of each ECG lead. In addition, the various

contributions each ECG lead made to the final prediction were weighted by the attention mechanisms, which were used to visualise the importance of each ECG lead.

Summary of the research interests, model comparison and statistical methods

The research interests, model comparison and statistical methods in this study are summarised in detail in **Supplementary Table 2**.

Supplementary Appendix 2. Results

The baseline characteristics of the cohorts

The characteristics and laboratory data are shown in **Supplementary Table 3**. Patients in the validation cohort were significantly older, had more comorbidities, had impaired estimated glomerular filtration rates and alanine aminotransferase, lower cTnI, and higher glucose and low-density lipoprotein cholesterol levels than those in the development cohort. The development/validation cohorts consisted of 860/191, 559/138, and 109,904/30,432 STEMI, NSTEMI, and non-AMI ECGs, respectively. The LAD and RCA were the most commonly identified IRAs in STEMI. Patients with STEMI were more likely to be male, to be overweight, to have prior coronary artery disease (CAD), and to have higher cTnI and more impaired lipid profiles than those in the non-AMI group. Patients with NSTEMI were more likely to be male, be older and have prior CAD and more comorbidities, higher cardiac biomarkers, and more impaired lipid profiles than those in the non-AMI group.

### *ECG lead-specific analysis*

The ECG lead-specific analyses for the detection of STEMI and the corresponding IRA are shown in **Supplementary Figure 4**. ECG leads were specifically analysed for the detection of STEMI in the hypothetical real world. Leads III, V2, aVL, and V3 demonstrated better performance than the other leads for the detection of STEMI, with the AUCs of 0.913, 0.913, 0.911, and 0.908, respectively. For the detection of the IRA of STEMI, lead-specific ROC curve analysis on the IRA of STEMI demonstrated that the best performances for

the LAD were V4, V2, and V3 with AUCs of 0.970, 0.955, and 0.953, respectively, and those for the RCA were aVL, lead III, and aVF with AUCs of 0.995, 0.978, and 0.966, respectively.

Discussion

With the aid of the first recorded cTnI, the DLM exhibited an excellent diagnostic yield with an AUC of 0.978 for NSTEMI detection, which was significantly better than those of the DLM or cTnI alone, with AUCs of 0.877 and 0.949, respectively. The universal diagnosis of NSTEMI is derived from the clinical presentation, 12-lead ECG, and cardiac troponin levels. To date, biomarker measurement for myocardial injury, preferably high-sensitivity cardiac troponin, was mandatory in all patients with suspected NSTEMI due to its high sensitivity and specificity [4]. However, several concerns should be considered in current practice. First, the guidelines suggest that the second cardiac troponin assessment be performed 1-3 hours after the first blood test in unconfirmed cases. Repeated time-consuming laboratory tests might delay the diagnosis. Second, cardiac troponin levels might be perturbed in some clinical conditions other than AMI. Combined with the information of the first recorded cTnI, the DLM allows rapid and powerful NSTEMI detection in patients at high or very high risk.

Regarding NSTEMI detection, DLM showed less sensitivity than the cardiologists. Several points should be clarified. Among the 58 NSTEMI ECGs unrecognised by the DLM, there were several atypical ECG presentations, including intraventricular conduction disorders, ventricular hypertrophy, poor R wave progression, or baseline variants. Even experienced cardiologists could not identify some of these ECGs. Moreover, overdiagnosis of NSTEMI by ECG is commonplace in clinical practice, which may partially explain the high sensitivity and low specificity of the performance of the physicians in this study. With the aid of the DLM with its high specificity in the detection of NSTEMI, physicians could exclude NSTEMI early, which reduced subsequent lab tests, ED observation time and guided physicians to differentiate it from other diagnoses unrelated to AMI. As a result, it was worthwhile to increase the ECG training data along with the first-record cTnI to enhance the capacity of the DLM in NSTEMI detection in the future.
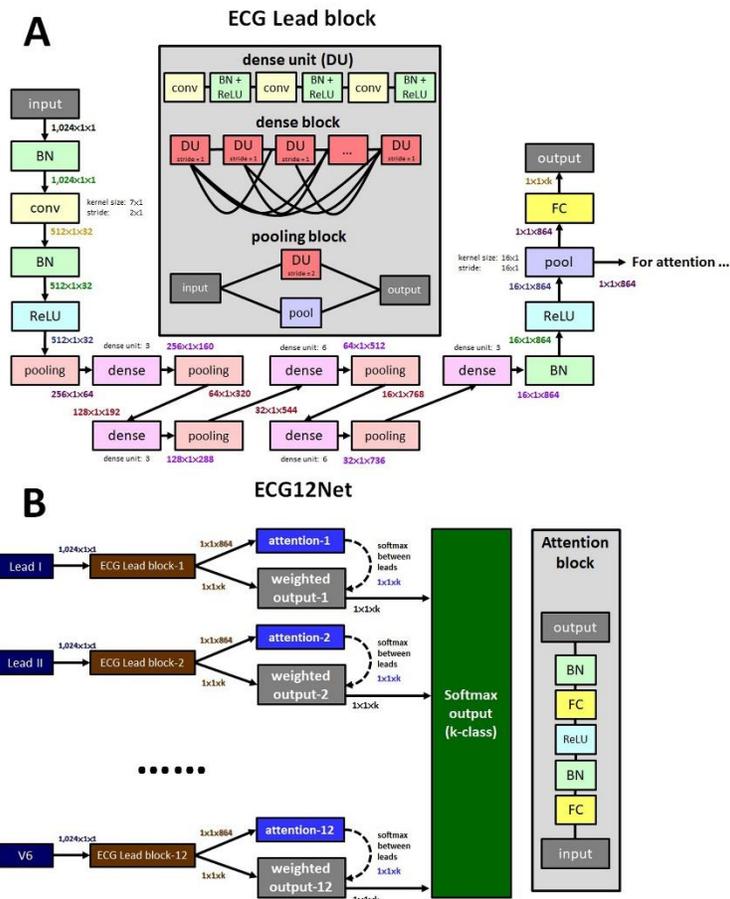
Supplementary Figure 1. Architecture of the DLM.

A) Electrocardiography (ECG) lead block with 80 trainable layers.

B) The DLM integrated all the information from the ECG leads to make an overall prediction. The bold and coloured words denote the output dimensions of the layers and the black words signify the important role for the layers. The model constant K was equal to 32 for all the dense blocks and pooling blocks.

BN: batch normalisation; Conv: convolution; FC: fully connected; ReLU: rectified linear unit
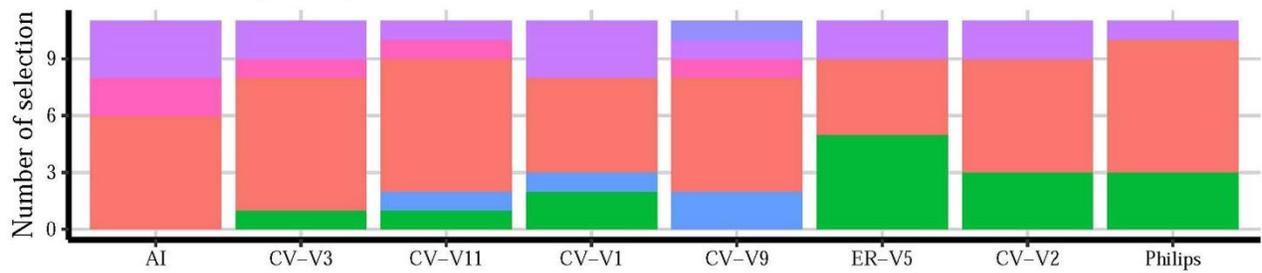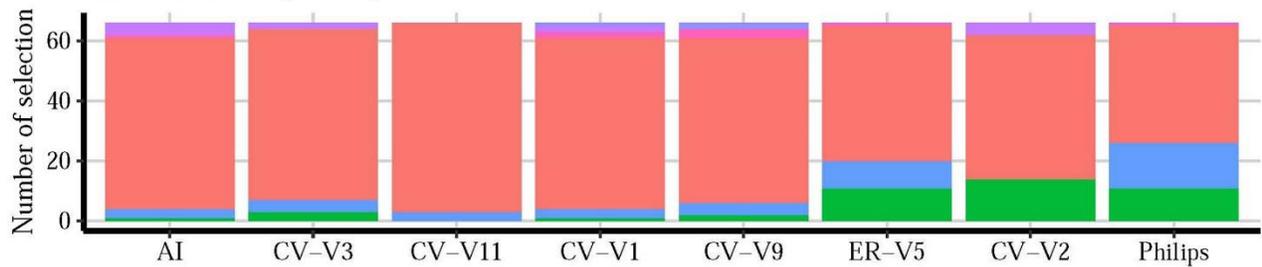
Global performance
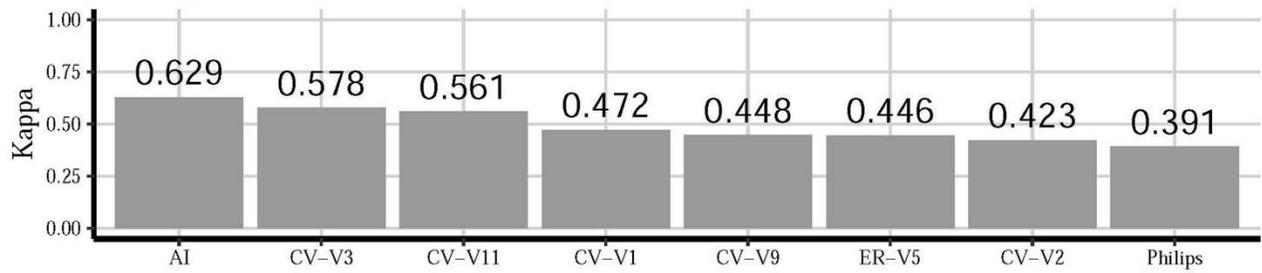
STEMI–LMCA (n = 2)
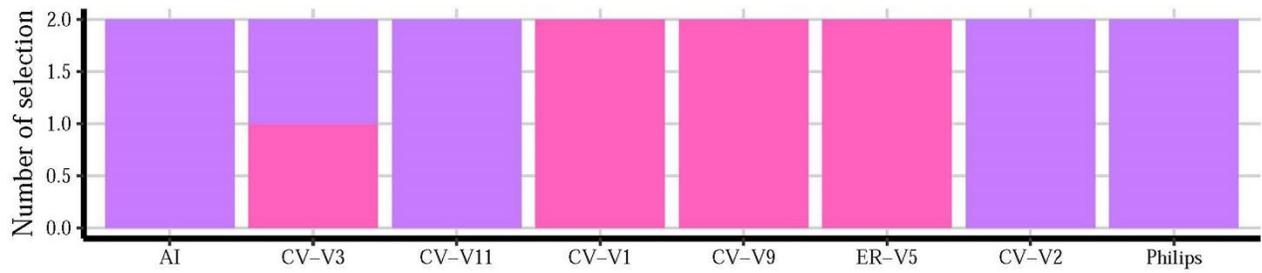
STEMI–LAD (n = 95)
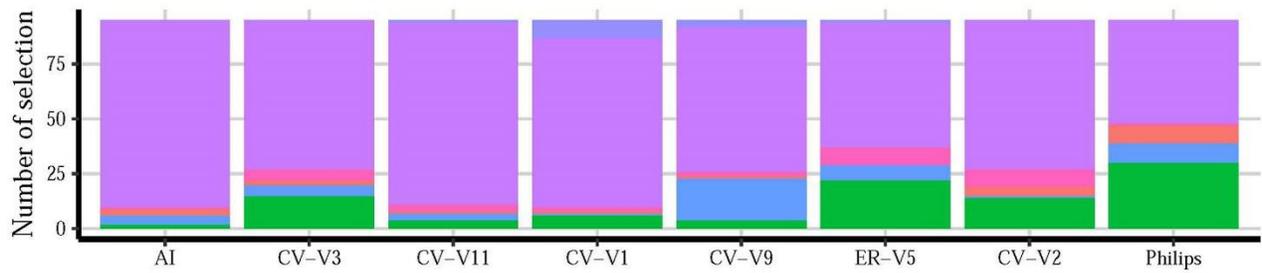
STEMI–LCx (n = 11)

STEMI–RCA (n = 66)
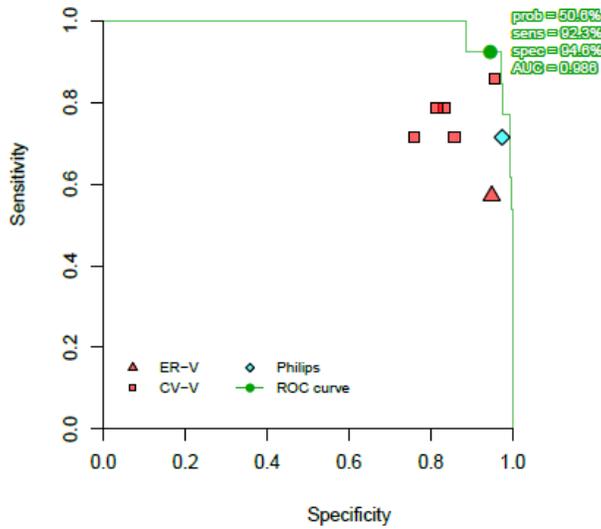
STEMI–LMCA   STEMI–LAD   STEMI–LCx   STEMI–RCA   NSTEMI   not–AMI
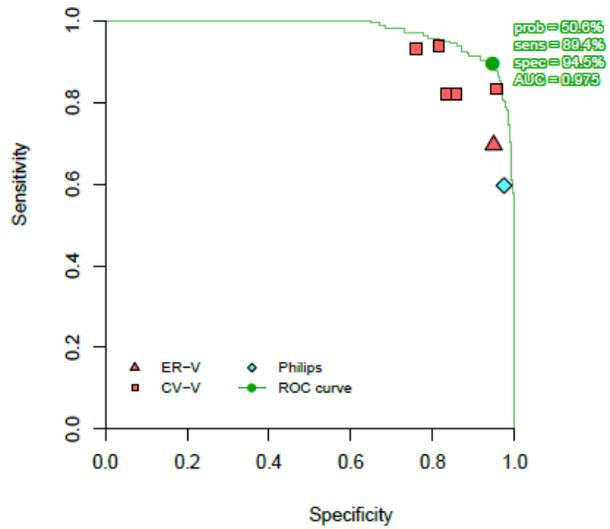
Supplementary Figure 2. Performance rankings of infarct-related artery detection of STEMI among DLM, physicians and the Philips algorithm in the human-machine competition.

Global performance rankings based on the 6-class kappa values. V(X) denoted the (V) visiting staff with (X) years of experience. The infarct-related arteries of STEMI were classified into the LMCA, LAD, RCA and LCx. LAD: left anterior descending artery; LCx: left circumflex artery; LMCA: left main coronary artery; RCA: right coronary artery
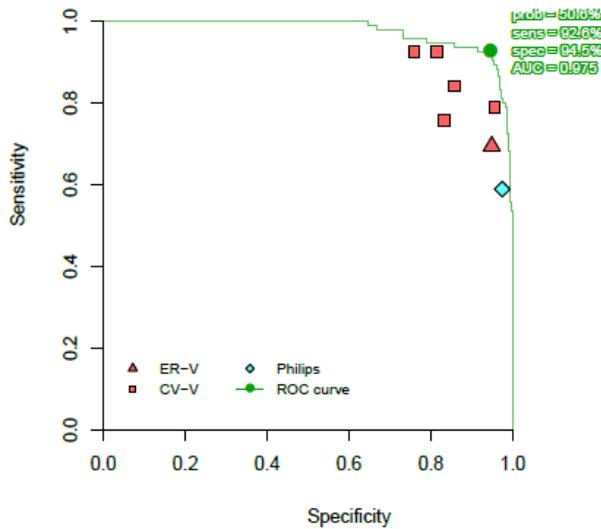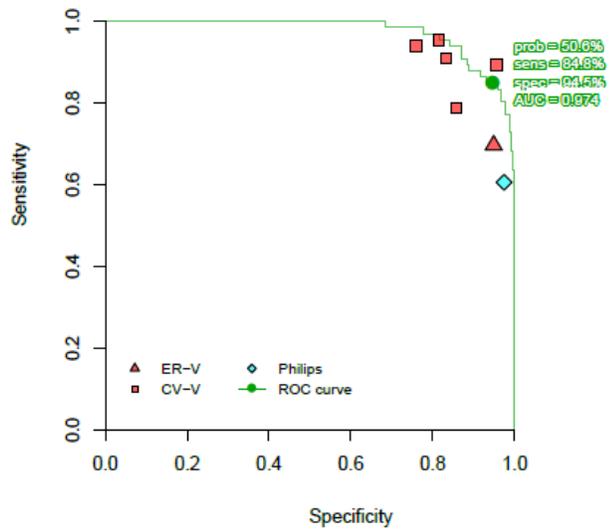
STEMI–LCx/LMCA vs. non–STEMI in competition

prob = 50.6%
sens = 92.3%
spec = 84.6%
AUC = 0.986

ER–V
CV–V
Philips
ROC curve

STEMI–LAD/RCA vs. non–STEMI in competition

prob = 50.6%
sens = 89.4%
spec = 94.5%
AUC = 0.975

ER–V
CV–V
Philips
ROC curve

STEMI–LAD vs. non–STEMI in competition

prob = 50.6%
sens = 92.6%
spec = 94.5%
AUC = 0.975

ER–V
CV–V
Philips
ROC curve

STEMI–RCA vs. non–STEMI in competition
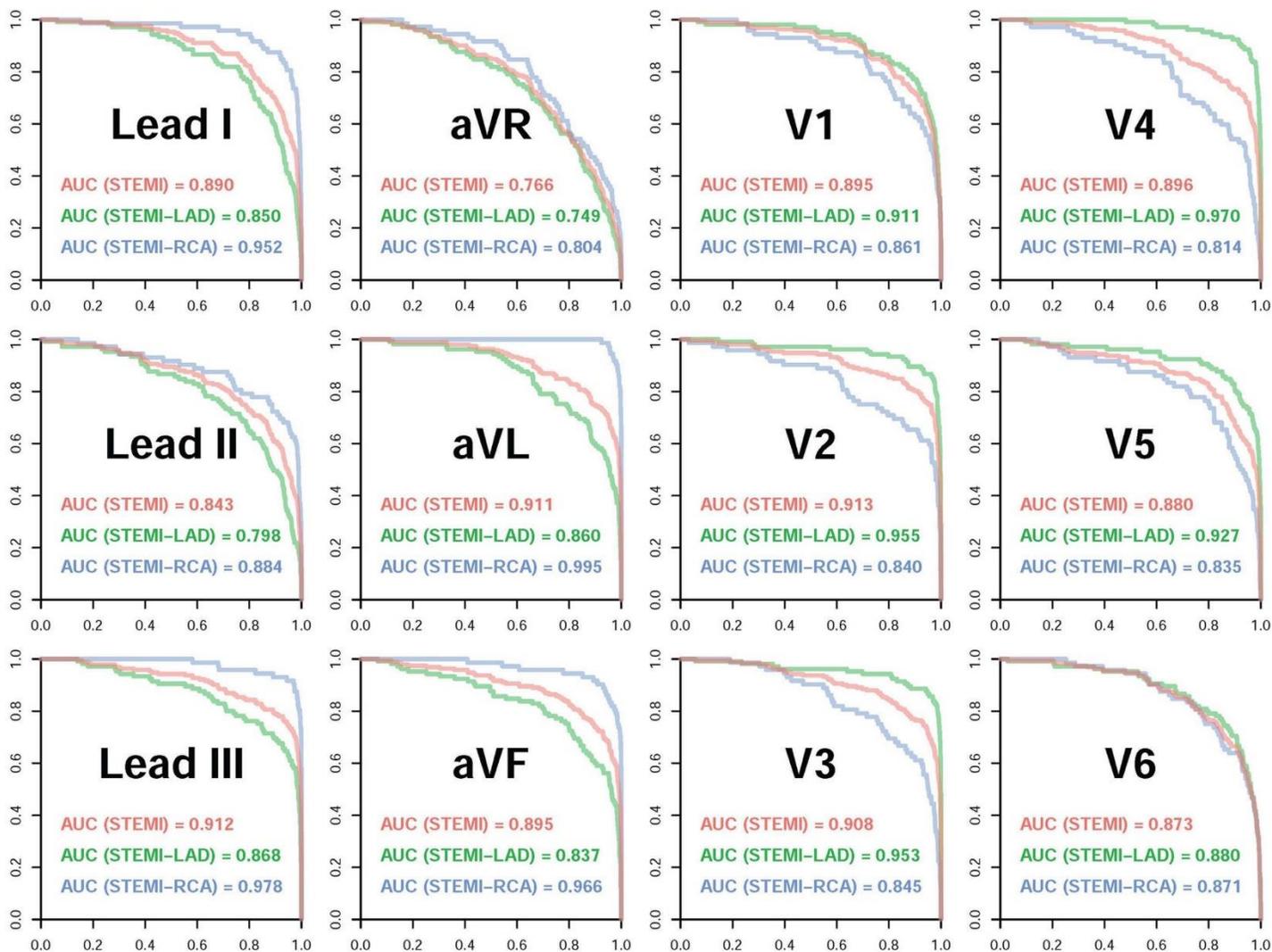
prob = 50.6%
sens = 84.8%
spec = 94.5%
AUC = 0.974

ER–V
CV–V
Philips
ROC curve

Supplementary Figure 3. Performance comparison for anterior (LAD), inferior (RCA), and combined anterior and inferior (LAD+RCA) STEMI detection in the human-machine competition.

The area under the receiver operating characteristic curve (AUC) was generated by the prediction of the DLM.

The triangles, the square and the diamond denote the cardiologists, the emergency physician and the Philips algorithm, respectively.

Supplementary Figure 4. ECG lead-specific analyses for the detection of STEMI, STEMI-LAD and STEMI-RCA.
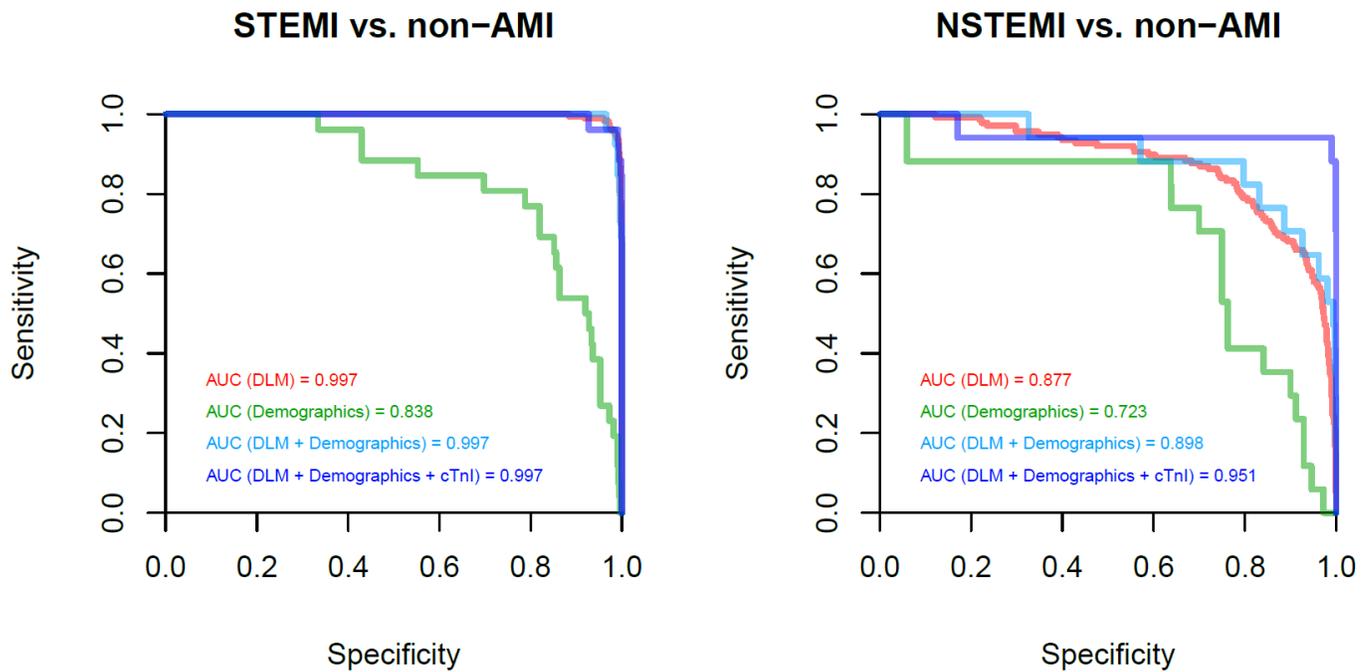
The receiver operating characteristic (ROC) curves with the specificity on the x-axis and the sensitivity on the y-axis were generated by the DLM for the detection of STEMI and the corresponding IRA in the revised proportion of the hypothetical real world (STEMI = 0.1%, NSTEMI = 0.2%, and non-AMI = 99.7%). The controls were the non-AMI samples.

AUC: area under the ROC curve
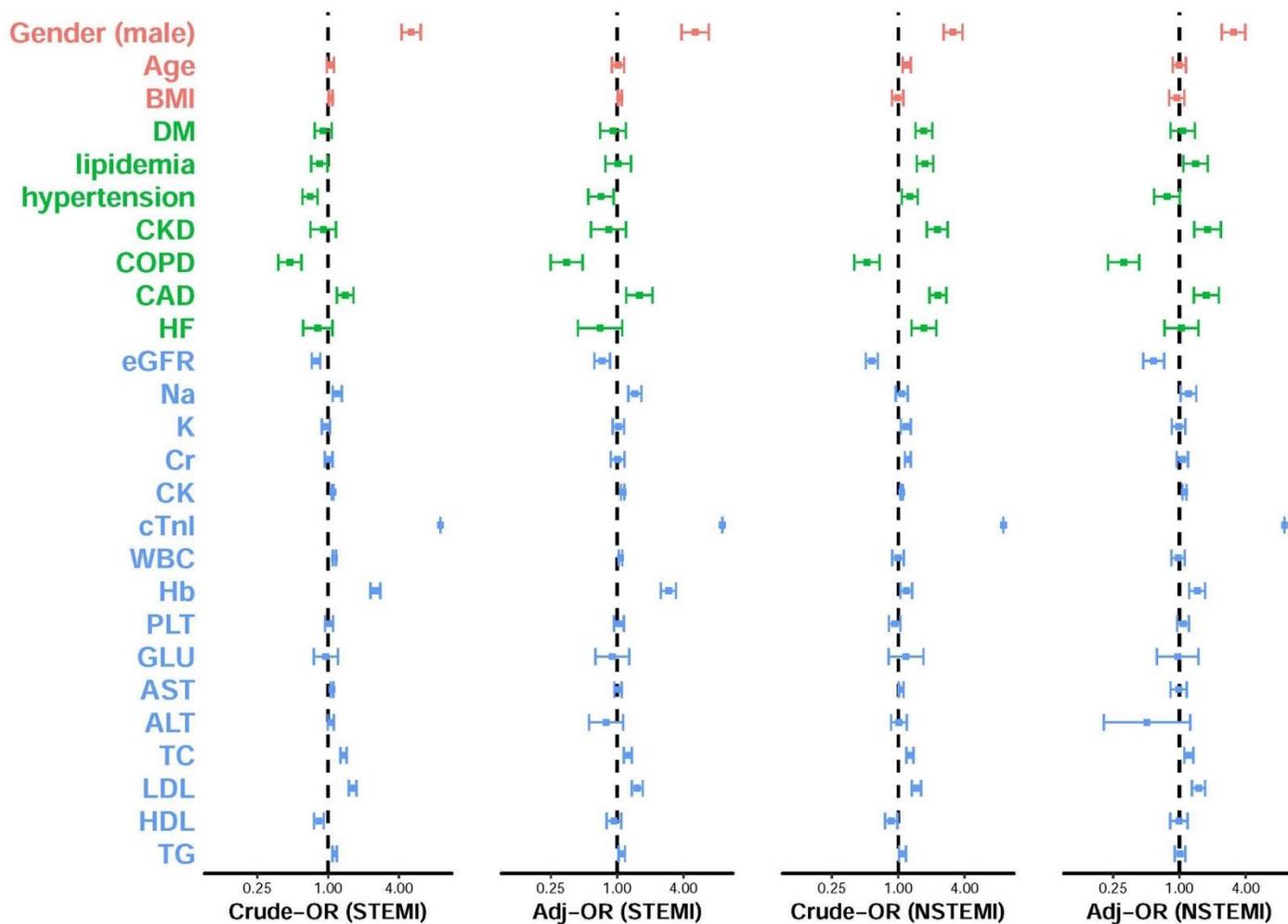
Supplementary Figure 5. Univariate and multivariate logistic regression analysis of STEMI, and NSTEMI in the development cohort.

The controls in all analyses were non-AMI samples. The adjusted variables included gender, age, body mass index, and all disease histories. The continuous variables were standardised by the mean and standard deviation. The units of each continuous variable were one standard deviation.
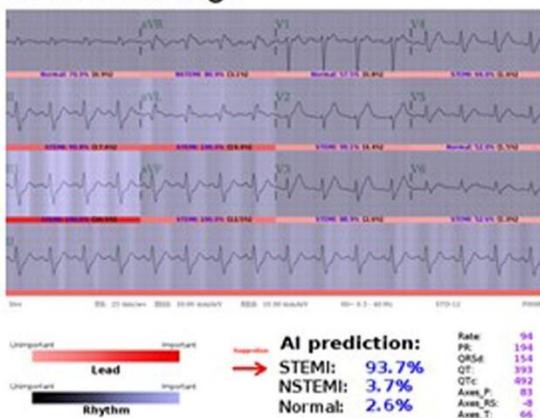
Supplementary Figure 6. Comparison of the diagnostic value among additional demographic variables, cTnI and DLM in the validation cohort.

The receiver operating characteristic (ROC) curves were generated from the logistic regression analysis using the development cohort. Patient demographic variables to predict (5A) STEMI and (5B) NSTEMI included gender, age, BMI, CAD, eGFR, and Hb. (5A) DLM vs DLM + Demographics vs DLM + Demographics + cTnI, p=ns; DLM or DLM + Demographics or DLM + Demographics + cTnI vs Demographics, p<0.0001. (5B) DLM vs Demographics, p<0.05; DLM + Demographics + cTnI vs DLM + Demographics, p=0.08.

AUC: area under the ROC curve

A. de Winter sign

AI prediction:
STEMI: 93.7%
NSTEMI: 3.7%
Normal: 2.6%

Rate: 94
PR: 194
QRSd: 154
QT: 393
QTc: 492
Axes_P: 83
Axes_RS: -8
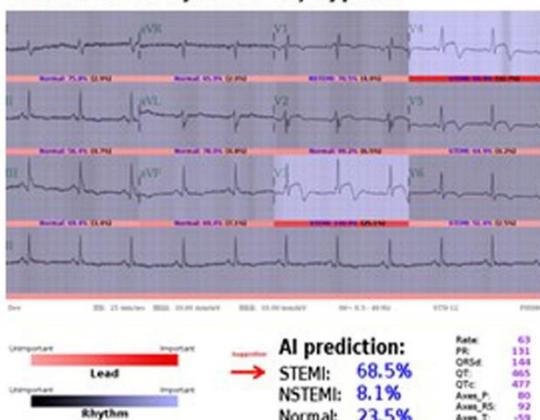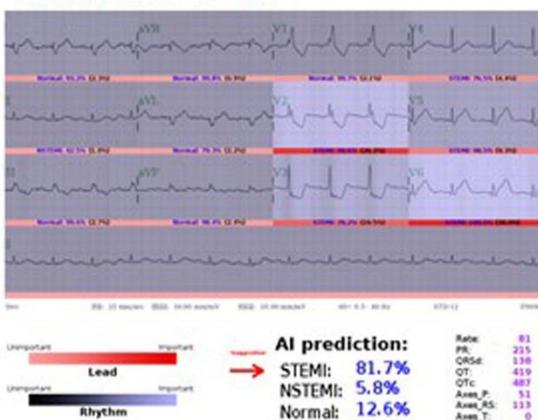Axes_T: 66

B. Wellens' syndrome, type I

AI prediction:
STEMI: 1.3%
NSTEMI: 29.0%
Normal: 69.7%

Rate: 92
PR: 139
QRSd: 93
QT: 366
QTc: 453
Axes_P: 60
Axes_RS: 41
Axes_T: 80

C. Wellens' syndrome, type II

AI prediction:
STEMI: 68.5%
NSTEMI: 8.1%
Normal: 23.5%

Rate: 63
PR: 131
QRSd: 144
QT: 465
QTc: 477
Axes_P: 80
Axes_RS: 92
Axes_T: -59

D. Posterior wall MI

AI prediction:
STEMI: 81.7%
NSTEMI: 5.8%
Normal: 12.6%

Rate: 81
PR: 215
QRSd: 138
QT: 419
QTc: 487
Axes_P: 51
Axes_RS: 113
Axes_T: 0

E. aVR ST elevation

AI prediction:
STEMI: 80.5%
NSTEMI: 14.2%
Normal: 5.3%

Rate: 58
PR: 104
QRSd: 125
QT: 439
QTc: 432
Axes_P: 159
Axes_RS: -75
Axes_T: 46

F. Hyperacute T

AI prediction:
STEMI: 83.8%
NSTEMI: 11.2%
Normal: 5.0%

Rate: 71
PR: 161
QRSd: 125
QT: 400
QTc: 435
Axes_P: 25
Axes_RS: -50
Axes_T: 62

G. LBBB

AI prediction:
STEMI: 88.3%
NSTEMI: 6.6%
Normal: 5.2%

Rate: 90
PR: 147
QRSd: 120
QT: 383
QTc: 469
Axes_P: 66
Axes_RS: 89
Axes_T: -36

H. RBBB

AI prediction:
STEMI: 52.4%
NSTEMI: 6.1%
Normal: 41.5%

Rate: 71
PR: 164
QRSd: 151
QT: 401
QTc: 523
Axes_P: 67
Axes_RS: 61
Axes_T: 35

Supplementary Figure 7. The test examples of the detection of STEMI equivalents by the DLM. STEMI equivalents including de Winter sign, Wellens' syndrome, posterior wall MI, ST elevation in lead aVR with diffuse ST depression, hyperacute T-waves and ST elevation in the presence of bundle branch block. The prediction rate of STEMI in each example of STEMI equivalent ECG by the DLM is shown in each figure.
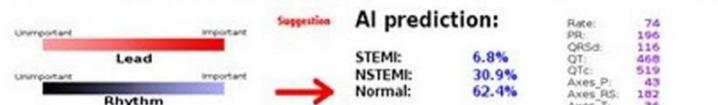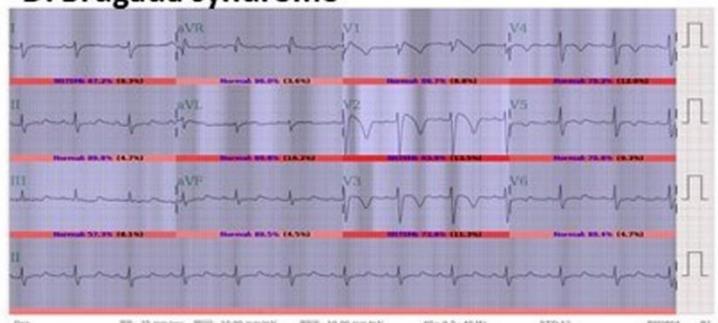
Supplementary Figure 8. The test examples of the detection of high take-off T ECG by the DLM. High take-off T ECG including hyperkalaemia, benign early repolarisation, left ventricular hypertrophy, and Brugada syndrome. The prediction rate of STEMI in each example of high take-off T ECG by the DLM is shown in each figure.

Supplementary Table 1. The definitions of AMI, STEMI, NSTEMI, non-AMI and non-STEMI.

Groups	Definition and inclusion in this study
AMI	AMI included symptoms of myocardial ischaemia, the ECG presentation and the elevated cTnI (above the 99th percentile of the upper reference limit of healthy individuals), which included both STEMI and NSTEMI
STEMI	AMI patients with ST-segment elevation on ECG who were validated by CAG
NSTEMI	AMI patients without ST-segment elevation on ECG who were validated by CAG
Non-AMI	Patients with a normal cTnI series during an ED stay who had neither STEMI nor NSTEMI
Non-STEMI	NSTEMI and non-AMI

AMI: acute myocardial infarction; CAG: coronary angiogram; cTnI: conventional cardiac troponin I; ECG: 12-lead electrocardiogram; ED: emergency department; NSTEMI: non-ST-segment elevation myocardial infarction; STEMI: ST-segment elevation myocardial infarction

Supplementary Table 2. The research interests, model comparison and statistical methods.

Figures	Purpose	Comparison	Methods
Figure 1	To compare the performance between the DLM and physicians in detecting STEMI by ECGs in the human-machine competition.	STEMI vs non-STEMI	AUC-ROC curve, PRROC curve with sensitivity (recall), specificity, and positive predictive value (precision).
Figure 2	To compare the performance of STEMI detection among the DLM, the physicians and the Philips algorithm.	DLM vs physicians, Philips algorithm	The performance (kappa value) and consistency analysis.
Figure 4A	To compare the performance of the DLM, cTnI and the DLM plus cTnI in detecting STEMI in the validation cohort	STEMI vs non-AMI	AUC-ROC curve.
Figure 4B	To compare the performance of the DLM, cTnI and the DLM plus cTnI in detecting NSTEMI in the validation cohort.	NSTEMI vs non-AMI	AUC-ROC curve.

AMI: acute myocardial infarction; AUC-ROC: area under the receiver operating characteristic curve; CAG: coronary angiogram; cTnI: cardiac troponin I; DLM: deep learning model; ECG: 12-lead electrocardiogram; NSTEMI: non-ST-segment elevation myocardial infarction; STEMI: ST-segment elevation myocardial infarction

Supplementary Table 3. Corresponding patient characteristics and laboratory results of STEMI, NSTEMI, and non-AMI ECGs in the development and validation cohorts.

	Development cohort				Validation cohort				*p*-value#
	STEMI (n=860)	NSTEMI (n=559)	non-AMI (n=109,904)	*p*-value	STEMI (n=191)	NSTEMI (n=138)	non-AMI (n=30,432)	*p*-value	
STEMI location									
STEMI-LMCA	21 (2.4%)				3 (1.6%)				
STEMI-LAD	420 (48.8%)				105 (55.0%)				
STEMI-LCx	87 (10.1%)				11 (5.8%)				
STEMI-RCA	332 (38.6%)				72 (37.7%)				
Gender (male)	688 (83.8%)	420 (76.2%)	55,453 (50.5%)	<0.001	150 (82.9%)	84 (62.2%)	15,484 (50.9%)	<0.001	0.369
Age (years)	61.8±13.8	64.3±13.8	60.9±19.6	<0.001	62.9±14.6	65.9±13.7	62.6±20.2	0.165	<0.001
BMI (kg/m²)	25.9±4.5	24.4±3.9	24.5±8.8	0.009	26.9±4.7	25.0±4.9	24.5±6.0	0.043	0.575
Disease history									
CAD	197 (24.0%)	188 (34.1%)	20,275 (18.4%)	<0.001	133 (73.5%)	95 (70.4%)	7,439 (24.4%)	<0.001	<0.001
HF	50 (6.1%)	66 (12.0%)	8,099 (7.4%)	<0.001	21 (11.6%)	33 (24.4%)	2,972 (9.8%)	<0.001	<0.001
DM	176 (21.4%)	187 (33.9%)	25,429 (23.1%)	<0.001	39 (21.5%)	50 (37.0%)	7,675 (25.2%)	0.004	<0.001
HTN	249 (30.3%)	243 (44.1%)	42,081 (38.3%)	<0.001	67 (37.0%)	83 (61.5%)	14,177 (46.6%)	<0.001	<0.001
CKD	68 (8.3%)	101 (18.3%)	9,929 (9.0%)	<0.001	8 (4.4%)	26 (19.3%)	2,332 (7.7%)	<0.001	<0.001
Hyperlipidaemia	198 (24.1%)	219 (39.7%)	30,087 (27.4%)	<0.001	34 (18.8%)	53 (39.3%)	8,579 (28.2%)	<0.001	0.007
COPD	85 (10.4%)	62 (11.3%)	21,600 (19.7%)	<0.001	24 (13.3%)	19 (14.1%)	7,090 (23.3%)	<0.001	<0.001
Laboratory test									
Na (mEq/L)	137.3±3.2	136.9±3.6	136.6±4.5	<0.001	137.1±2.7	135.9±3.4	135.8±4.7	0.005	<0.001
K (mEq/L)	3.9±0.6	4.0±0.6	3.9±0.5	0.006	3.8±0.5	4.0±0.6	3.9±0.5	0.008	0.211
eGFR (mL/min)	74.2±26.3	63.8±30.7	82.5±37.0	<0.001	74.2±26.5	64.3±37.4	81.0±35.0	<0.001	<0.001
Cr (mg/dl)	1.3±1.3	1.9±2.2	1.3±1.6	<0.001	1.3±0.9	2.3±2.6	1.2±1.3	<0.001	<0.001
CK (ng/mL)	389.8±650.7	296.1±325.4	131.7±409.0	<0.001	348.9±597.0	252.5±310.7	122.5±306.9	<0.001	<0.001
cTnI (ng/mL)	60.6±598.7	224.8±1,121.7	0.0±0.0	<0.001	4.8±16.6	2.7±6.5	0.0±0.0	<0.001	0.015
WBC (10³/ul)	11.1±3.6	8.8±3.0	8.9±4.5	<0.001	11.2±3.2	9.3±2.8	8.8±4.6	<0.001	0.125
Hb (gm/dl)	14.6±1.9	13.2±2.4	12.9±2.3	<0.001	14.7±1.7	13.2±2.7	12.9±2.3	<0.001	0.120
PLT (10³/ul)	228.5±64.0	221.0±74.6	227.0±81.9	0.425	228.4±90.7	216.5±52.9	210.1±74.9	0.015	<0.001
GLU (gm/dl)	193.9±85.3	219.4±126.3	198.7±114.8	0.631	166.0±13.1	215.8±85.5	241.1±128.5	0.462	<0.001
AST (U/L)	54.0±85.3	45.6±104.5	32.6±81.3	<0.001	51.3±65.0	36.4±37.1	33.0±91.3	0.075	0.590
ALT (U/L)	41.3±73.4	34.2±78.9	32.8±93.1	0.215	44.6±21.3	39.0±40.6	79.0±200.9	0.762	<0.001
TC (gm/dl)	172.0±40.9	168.4±37.5	148.8±47.7	<0.001	173.6±36.8	162.8±38.3	147.6±48.0	<0.001	0.081
LDL (gm/dl)	111.4±33.7	106.8±33.8	89.7±36.3	<0.001	116.4±33.2	103.2±28.0	95.9±38.2	<0.001	<0.001
HDL (gm/dl)	38.7±9.0	39.2±9.4	41.2±14.4	<0.001	41.5±10.4	35.3±9.8	42.0±15.0	0.007	0.295
TG (gm/dl)	153.4±148.7	137.0±73.4	118.0±127.8	<0.001	120.3±55.8	157.7±96.2	116.6±160.7	0.043	0.354

The hypothesis test between the development cohort and the validation cohort.

ALT: alanine aminotransferase; AST: aspartate aminotransferase; BMI: body mass index; CAD: coronary artery disease; CK: creatine kinase; CKD: chronic kidney disease; COPD: chronic obstructive pulmonary disease; Cr: creatinine; cTnI: conventional cardiac troponin I; DM: diabetes mellitus; eGFR: estimated glomerular filtration rate; GLU: glucose; Hb: haemoglobin; HDL: high-density lipoprotein cholesterol; HF: heart failure; HTN: hypertension; K: potassium; LAD: left anterior descending artery; LCx: left

circumflex artery; LDL: low-density lipoprotein cholesterol; LMCA: left main coronary artery; Na: sodium; PLT: platelet; RCA: right coronary artery; TC: total cholesterol; TG: triglyceride; WBC: white blood cell count