

Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction

Received: 24 January 2023

Accepted: 11 May 2023

Published online: 29 June 2023

 Check for updates

Salah S. Al-Zaiti ^{1,2,3,4}✉, Christian Martin-Gill ^{2,5}, Jessica K. Zègre-Hemsey ⁶, Zeineb Bouzid ³, Ziad Faramand⁷, Mohammad O. Alrawashdeh ^{8,9}, Richard E. Gregg¹⁰, Stephanie Helman¹, Nathan T. Riek ³, Karina Kraevsky-Phillips ¹, Gilles Clermont¹¹, Murat Akcakaya³, Susan M. Sereika¹, Peter Van Dam¹², Stephen W. Smith ^{13,14}, Yochai Birnbaum ¹⁵, Samir Saba^{4,5}, Ervin Sejdic^{16,17} & Clifton W. Callaway^{2,5}

Patients with occlusion myocardial infarction (OMI) and no ST-elevation on presenting electrocardiogram (ECG) are increasing in numbers. These patients have a poor prognosis and would benefit from immediate reperfusion therapy, but, currently, there are no accurate tools to identify them during initial triage. Here we report, to our knowledge, the first observational cohort study to develop machine learning models for the ECG diagnosis of OMI. Using 7,313 consecutive patients from multiple clinical sites, we derived and externally validated an intelligent model that outperformed practicing clinicians and other widely used commercial interpretation systems, substantially boosting both precision and sensitivity. Our derived OMI risk score provided enhanced rule-in and rule-out accuracy relevant to routine care, and, when combined with the clinical judgment of trained emergency personnel, it helped correctly reclassify one in three patients with chest pain. ECG features driving our models were validated by clinical experts, providing plausible mechanistic links to myocardial injury.

The electrocardiogram (ECG) diagnosis of acute coronary syndrome (ACS) in patients with acute chest pain is a longstanding challenge in clinical practice^{1–4}. Guidelines primarily focus on ST-segment elevation (STE) for discerning patients with ST-elevation myocardial infarction

(STEMI) versus other forms of ACS^{5–8}. A biomarker-driven approach is recommended in the absence of STE on the presenting ECG. This diagnostic paradigm has two important limitations. First, around 24–35% of patients with non-STEMI have total coronary occlusion, referred

¹Department of Acute & Tertiary Care Nursing, University of Pittsburgh, Pittsburgh, PA, USA. ²Department of Emergency Medicine, University of Pittsburgh, Pittsburgh, PA, USA. ³Department of Electrical & Computer Engineering, University of Pittsburgh, Pittsburgh, PA, USA. ⁴Division of Cardiology, University of Pittsburgh, Pittsburgh, PA, USA. ⁵University of Pittsburgh Medical Center, Pittsburgh, PA, USA. ⁶School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ⁷Department of Emergency Medicine, Northeast Georgia Health System, Gainesville, GA, USA. ⁸School of Nursing, Jordan University of Science and Technology, Irbid, Jordan. ⁹Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA. ¹⁰Advanced Algorithm Development Center, Philips Healthcare, Cambridge, MA, USA. ¹¹Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA, USA. ¹²Division of Cardiology, University Medical Center Utrecht, Utrecht, The Netherlands. ¹³Department of Emergency Medicine, Hennepin Healthcare, Minneapolis, MN, USA. ¹⁴Department of Emergency Medicine, University of Minnesota, Minneapolis, MN, USA. ¹⁵Division of Cardiology, Baylor College of Medicine, Houston, TX, USA. ¹⁶Department of Electrical & Computer Engineering, University of Toronto, Toronto, ON, Canada. ¹⁷Artificial Intelligence for Health Outcomes at Research & Innovation, North York General Hospital, Toronto, ON, Canada. ✉e-mail: ssa33@pitt.edu

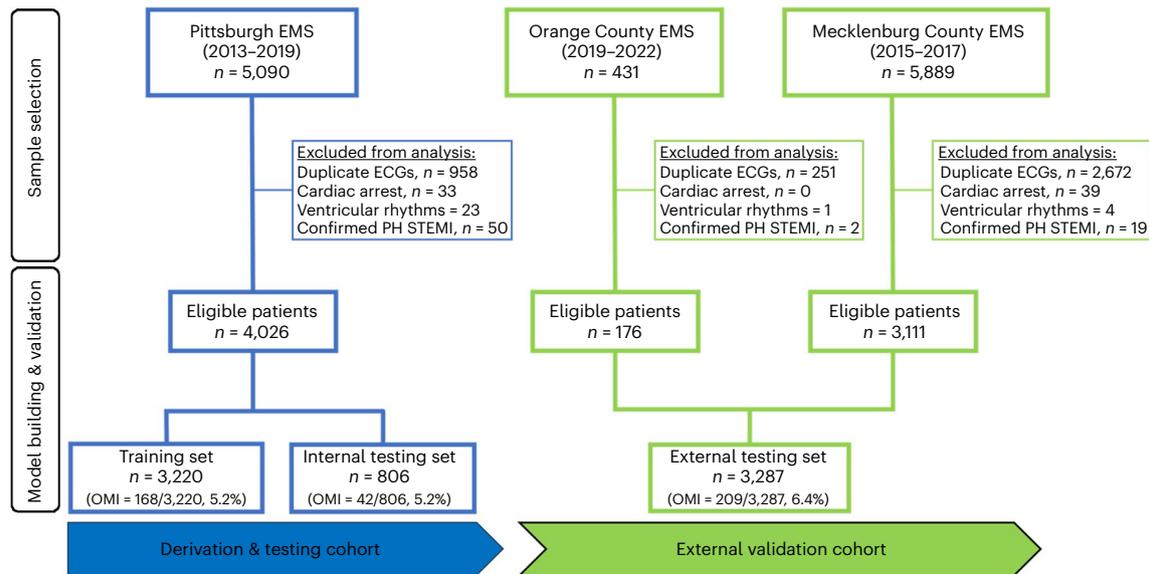


Fig. 1 | Cohort and sample selection. This flow diagram shows patient inclusion and exclusion criteria in each cohort as well as the dataset partition for training, internal testing and external validation cohorts. Exclusions are not mutually exclusive. EMS, Emergency Medical Services; PH, pre-hospital.

to as occlusion myocardial infarction (OMI), and require emergent catheterization^{9–13}. This vulnerable group, in contrast to ACS with an open artery (Extended Data Fig. 1), suffers from unnecessary diagnostic and treatment delays that are associated with higher mortality^{14–17}. This excess risk can be mitigated with enhanced diagnostic criteria. Although important ECG signatures of OMI are frequently described in the literature^{18–21}, they are subtle, involve the entire QRST complex and are spatial in nature (that is, changes diluted across multiple leads)^{22–24}. Visual inspection of ECG images by clinical experts is, thus, suboptimal and leads to a high degree of variability in ECG interpretation^{25–27}.

The second limitation is that cardiac biomarkers, including conventional or high-sensitivity troponin (hs-cTn), cannot differentiate OMI until peak level is reached, which is too late to salvage myocardium. Positive troponin results (>99th percentile limit) come with a high false-positive rate, and approximately one-third of patients remain in a biomarker-indeterminate ‘observation zone’ after serial sampling^{28,29}. More importantly, ~25% of acute myocardial infarction cases have a negative initial hs-cTn, which is observed in both the STEMI and OMI subgroups³⁰. Consequently, 25–30% of patients with OMI are not treated in a timely fashion, and around 63% (interquartile range, 38–81%) of patients evaluated for chest pain at the emergency department are admitted to the hospital because of an inconclusive initial assessment³¹. These diagnostic limitations have created a costly, inefficient clinical practice paradigm where most patients with chest pain are over-monitored, whereas some patients with OMI have delayed diagnosis and treatment, potentially contributing to the 14–22% excess risk of mortality seen in the non-STE ACS (NSTEMI-ACS) group^{15,32,33}.

In our previous work, we designed prototype algorithms for artificial intelligence (AI)-enabled ECG analysis and demonstrated the clinical feasibility of screening for ACS in the pre-hospital setting^{34,35}. Here we describe, to our knowledge, the first multi-site, prospective, observational cohort study to evaluate the diagnostic accuracy of machine learning for the ECG diagnosis and risk stratification of OMI at first medical contact and in the absence of a STEMI pattern (Extended Data Fig. 2). Our intelligent models were derived and externally validated on 7,313 patients with chest pain from multiple clinical sites in the United States. The results demonstrate the superiority of machine learning in detecting subtle ischemic ECG changes indicative of OMI in the absence of a STEMI pattern, outperforming practicing clinicians and other widely used commercial ECG interpretation software. We identified the

most important ECG features driving our model’s classifications and identified plausible mechanistic links to myocardial injury. Our derived OMI risk score provides enhanced rule-in and rule-out accuracy when compared to the HEART score, helping correctly reclassify one in three patients with chest pain. The benefits of this new clinical pathway in terms of clinical outcomes should be evaluated in prospective trials.

Results

Sample characteristics

After excluding patients with cardiac arrest, ventricular tachyarrhythmias, confirmed pre-hospital STEMI and duplicate ECGs, our derivation cohort included 4,026 consecutive patients with chest pain (age 59 ± 16 years, 47% females, 5.2% OMI). The two external validation cohorts together included 3,287 patients (age 60 ± 15 years, 45% females, 6.4% OMI) (Fig. 1 and Table 1). Most patients in the derivation and validation cohorts were in normal sinus rhythm (>80%), and around 10% were in atrial fibrillation. Around 3% of patients had left bundle branch block (BBB), and ~10% had ECG evidence of left ventricular hypertrophy (LVH). The derivation and validation cohorts were similar in terms of age, sex, baseline clinical characteristics and 30-d cardiovascular mortality. The validation cohort, however, had more Black and Hispanic minorities and a slightly higher rate of ACS and OMI.

Algorithm derivation and testing

The positive class for model training was the presence of OMI, defined as a culprit coronary artery with a thrombolysis in myocardial infarction (TIMI) flow grade of 0–1, as adjudicated from charts by independent reviewers blinded to all ECG analyses. A TIMI flow grade of 2 with severe coronary narrowing (>70%) and peak fourth-generation (not high sensitivity) troponin of 5–10 ng ml^{-1} was also indicative of OMI. The negative class for model training was the absence of OMI, which included all other non-ACS etiologies or those with non-coronary occlusive ACS subtypes.

Input data for model training was based on pre-hospital 12-lead ECGs. We selected 73 morphological ECG features out of 554 temporal–spatial metrics using a hybrid data-driven and domain expertise approach¹⁸. Using these features, 10 classifiers were trained to learn ischemic patterns between ACS and non-ACS groups and to estimate the probability of OMI. We chose these classifiers to maximize the chance of finding the best-fitting approach for learning the mathematical representation relating complex ECG data to underlying physiology.

Table 1 | Baseline demographic and clinical characteristics

	Derivation and testing cohort (n=4,026)	External validation cohort (n=3,287)
Age (years)	59±16 (18–102)	60±15 (21–100)
Sex		
Male	2,122 (53%)	1,814 (55%)
Female	1,904 (47%)	1,473 (45%)
Race		
White	1,698 (42%)	1,326 (40%)
Black	1,328 (33%)	1,544 (47%)
Others	52 (1.3%)	40 (1%)
Unknown	948 (24%)	377 (12%)
Ethnicity		
Not Hispanic	3,043 (76%)	2,850 (87%)
Hispanic / Latino	19 (1%)	116 (3.5%)
Unknown	964 (23%)	321 (9.5%)
Past medical history		
Hypertension	2,767 (69%)	2,090 (64%)
Diabetes	1,146 (29%)	1,067 (33%)
High cholesterol	1,520 (38%)	1,376 (42%)
Current smoker	1,244 (31%)	802 (25%)
Known CAD	1,388 (35%)	964 (30%)
Prior myocardial infarction	930 (23%)	929 (29%)
Prior PCI	963 (24%)	134 (4%)
Prior CABG	357 (10%)	470 (14%)
ECG and laboratory findings		
Sinus rhythm	3,496 (87%)	2,614 (80%)
Atrial fibrillation	354 (9%)	352 (11%)
Left BBB	94 (2.3%)	114 (3.5%)
Right BBB	237 (5.9%)	215 (6.6%)
ECG-LVH	383 (9.5%)	467 (14.2%)
cTnI positive (initial)	330 (8.2%)	736 (22.4%)
cTnI positive (serial testing)	729 (18.1%)	1,177 (35.8%)
Medical therapy		
PCI (any stent)	300 (7.5%)	245 (7.5%)
Emergent PCI (<90 min)	144 (3.6%)	157 (4.8%)
Total LAD occlusion	91 (2.3%)	94 (2.9%)
Total LCX occlusion	63 (1.6%)	88 (2.7%)
Total RCA occlusion	101 (2.5%)	102 (3.1%)
CABG	34 (0.8%)	30 (0.9%)
Study outcomes		
Confirmed ACS	550 (13.7%)	537 (16.3%)
OMI	210 (5.2%)	209 (6.4%)
Other acute myocardial infarction (NOMI)	240 (6.0%)	220 (6.7%)
Unstable angina	100 (2.5%)	108 (3.3%)
30-d cardiovascular death	137 (3.4%)	111 (3.4%)

Values are mean±s.d. (minimum–maximum) or n (%). CABG, coronary artery bypass graft; NOMI: non-occlusion myocardial infarction; PCI: percutaneous coronary intervention.

The random forest (RF) model achieved the best bias–variance tradeoff for training and internal testing. We compared the RF against the ECG interpretation of practicing clinicians and against the performance of a commercial ECG interpretation system that is cleared by the US Food and Drug Administration (FDA) for ‘Acute MI’ diagnosis. On the hold-out test set, the RF model (area under the receiver operating characteristic (AUROC) 0.91 (95% confidence interval (CI) 0.87–0.96)) outperformed both practicing clinicians (AUROC 0.79 (95% CI 0.73–0.76), $P < 0.001$) and the commercial ECG system (AUROC 0.78 (95% CI 0.70–0.85), $P < 0.001$) (Fig. 2a).

Next, we used probability density plots for OMI(+) and OMI(–) classes to denote the optimal separation margins for risk prediction. As recommended by guidelines⁶, we defined a risk score to identify patients at low risk (OMI score <5), intermediate risk (OMI score 5–20) and high risk (OMI score >20), with these cutoffs yielding excellent separation between classes (log-rank chi-square, 133.04; degrees of freedom = 2; $P < 0.001$) (Fig. 2b, left). Our OMI score classified 74.4% of patients as low risk and 4.6% as high risk. Using the low-risk group in a rule-out strategy yielded a sensitivity of 0.91 and a negative predictive value (NPV) of 0.993, with an overall missed event rate of 0.5%. Using high-risk class for a rule-in strategy yielded a specificity of 0.976 and a positive predictive value (PPV) of 0.514, with an overall false discovery rate of 2%. Finally, we compared this OMI score to the HEART score, which uses patient history, ECG data, age, risk factors and troponin values (Fig. 2b, right). Our OMI score, which is based on ECG data alone, classified 66% more patients as low risk than the HEART score, with a similar false-negative rate <1%, and classified fewer patients as high risk and with much higher precision (51% versus 33%). The OMI score also triaged 50% fewer patients as intermediate risk and still got better discrimination for OMI detection (11.2% versus 5.6%).

Model explainability

We used Tree SHAP algorithms to generate an importance ranking that explains the output of the RF model based on SHAP values estimated for the top 25 features (Fig. 3a). The features with the greatest impact on classification output included slight ST-depression in leads VI, V2, I and aVL; slight ST-elevation in leads III and V4–V6; loss of concave pattern in anterior leads; T wave enlargement in II and aVF and T flattening or inversion in I and aVL; prolonged $T_{\text{peak}}-T_{\text{end}}$ interval; T axis deviation; increased repolarization dispersion; and distorted directions of activation and recovery patterns. Most of these ECG patterns can be mechanistically linked to cardiac ischemia, suggesting their clinical value as plausible features for OMI detection.

To better visualize these global ECG patterns detected by our model, we created pooled population median beats for the OMI(+) class ($n = 414$ ECGs) and superimposed these median beats on the pooled population median beats of patients with normal sinus rhythm and OMI(–) status ($n = 9,072$ ECGs) (Fig. 3b). Findings from this figure support the patterns described by SHAP values above. Specifically, OMI is associated with ST-depression and T flattening in V1–V2, I and aVL; slight ST-elevation in the anterior leads with loss in concave pattern; peaked T wave in inferior leads; $T_{\text{peak}}-T_{\text{end}}$ prolongation (seen in many leads); global repolarization dispersion (seen as peaked T in some leads and flattening in others); T axis deviation (away from the left ventricle); and distorted activation and recovery patterns (seen in the horizontal plane as loss of R wave progression in pre-cordial leads with increased T wave discordance). Due to prevalent multi-vessel disease in this cohort, these OMI patterns remained relatively consistent regardless of culprit location.

Nevertheless, to examine local explainability of feature importance, we used force plots on individual cases to identify the features that met the contribution threshold of the RF model on a given ECG. These force plots were also examined by study investigators to further corroborate on the clinical validity of model predictions. Extended Data Fig. 3 shows a selected example of a 12-lead ECG with its corresponding force plot for the local features contribution.

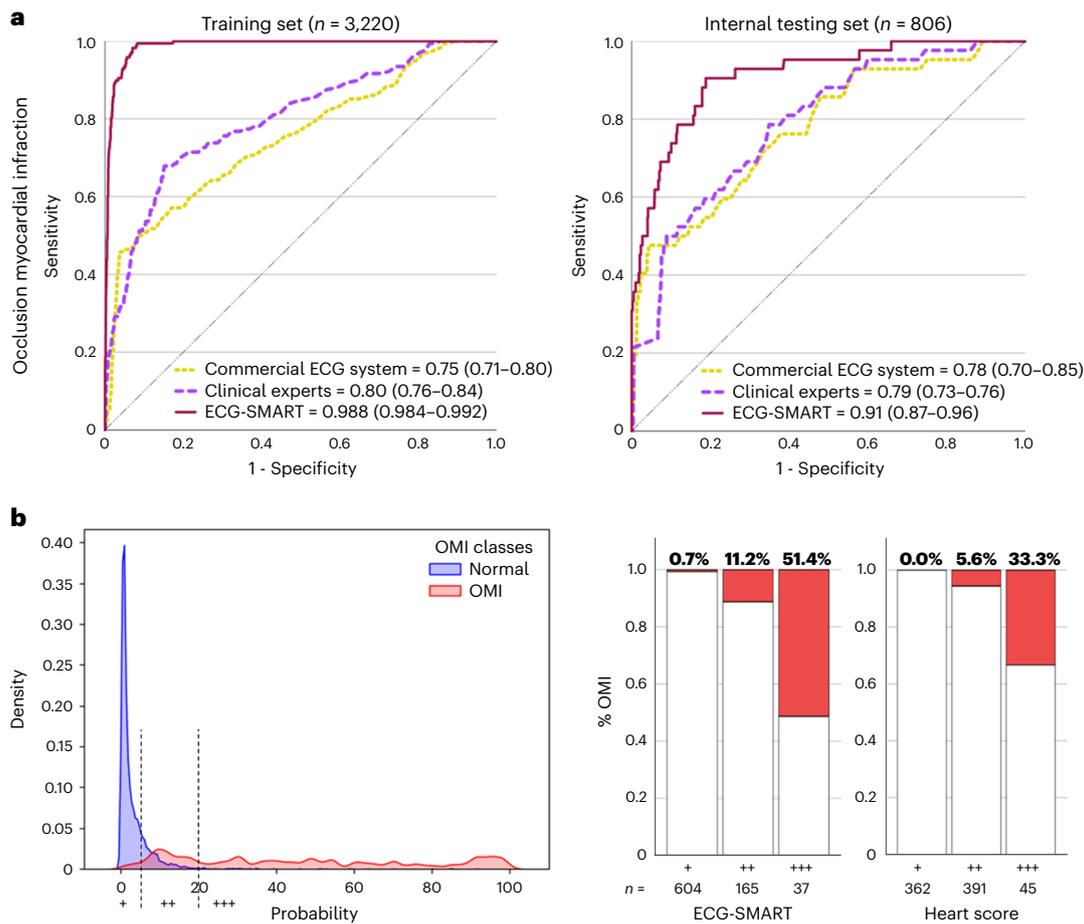


Fig. 2 | Algorithm derivation and testing. This figure shows the classification performance of the machine learning model against other reference standards for detecting OMI (a), the probability density plots of OMI(+) and OMI(-) classes as denoted by the machine learning model, along with optimal cutoffs of low risk,

intermediate risk and high risk (b, left), and distribution of patients in low risk (+), intermediate risk (++) and high risk (+++) as per the machine learning model and HEART score (b, right).

External validation

We tested the final lock-out model on 3,287 patients from two independent external clinical sites. Machine learning engineers were blinded to outcome data from other sites, and the pre-populated model predictions were independently evaluated by the clinical investigators. Our model generalized well and maintained high classification performance (AUROC 0.87 (95% CI 0.85–0.90)), outperforming the commercial ECG system (AUROC 0.75 (95% CI 0.71–0.79), $P < 0.001$) and practicing clinicians (AUROC 0.80 (95% CI 0.77–0.83), $P < 0.001$) (Fig. 4a). Our OMI risk score was a strong predictor of OMI, independent from age, sex and other coronary risk factors (odds ratio (OR) 10.60 (95% CI 6.78–16.64) for high-risk class and OR 2.85 (95% CI 1.91–4.28) for intermediate-risk class) (Fig. 4b). This risk score triaged 69% of patients in the low-risk group at a false-negative rate of 1.3% and identified 5.1% of patients as high risk at acceptable true-positive rate >50%. The overall sensitivity, specificity, PPV and NPV for the OMI rule-in and rule-out strategy were 0.86 (95% CI 0.81–0.91), 0.98 (95% CI 0.97–0.99), 0.54 (95% CI 0.46–0.62) and 0.99 (95% CI 0.98–0.99), respectively. This diagnostic accuracy remained relatively similar across subgroups based on age, sex, race, comorbidities and baseline ECG findings, indicating the lack of aggregation bias (Fig. 4c). In comparison, the sensitivity, specificity, PPV and NPV for ECG overread by practicing clinicians were 0.58, 0.93, 0.36 and 0.97 and, for the commercial ECG system, 0.79, 0.80, 0.22 and 0.98, respectively.

Next, we evaluated the incremental gain of our derived risk score in reclassifying patients at first medical contact (Fig. 5). Initial assessment by emergency personnel was based on the modified HEAR

(history, ECG, age and risk factors) score to triage patients into low-risk, intermediate-risk and high-risk groups³⁶. At baseline, emergency personnel triaged 48% of patients as low risk with an NPV of 99.0% and triaged 3% of patients as high risk with a PPV of 54.1%. Nearly 50% of patients remained in an indeterminate observation zone. Applying our OMI risk score would help triage 45% more patients as low risk while keeping the NPV at 98.8% and would help detect 85% more patients with OMI while keeping PPV at 50.0%. The OMI score would also help reduce the number of patients in the indeterminate observation zone by more than half. These numbers translate into a net reclassification improvement (NRI) index of 41% (95% CI 33–50%). To validate this incremental clinical utility, we manually reviewed ECGs reclassified correctly as OMI(+) (Extended Data Fig. 4). Many of these ECGs showed subtle or non-specific changes that were non-diagnostic as per guidelines⁵, suggesting potential value in boosting providers’ confidence when interpreting ‘fuzzy’ ECGs.

Finally, we investigated the potential sources of false negatives in the validation data. Among patients with missed OMI events ($n = 28$, 0.9%), many had high-frequency noise and baseline wander on their initial ECG ($n = 13/28$, 46%) or had low-voltage ECG ($n = 14/28$, 50%), and most patients ($n = 24/28$, 86%) had benign ECGs without any diagnostic ST-T changes (Extended Data Fig. 5). Moreover, we found no significant differences between false negatives and true positives in terms of demographics or clinical characteristics, with the exception that most false negatives had a history of a prior myocardial infarction (93% versus 27%). The latter finding was intriguing given that our

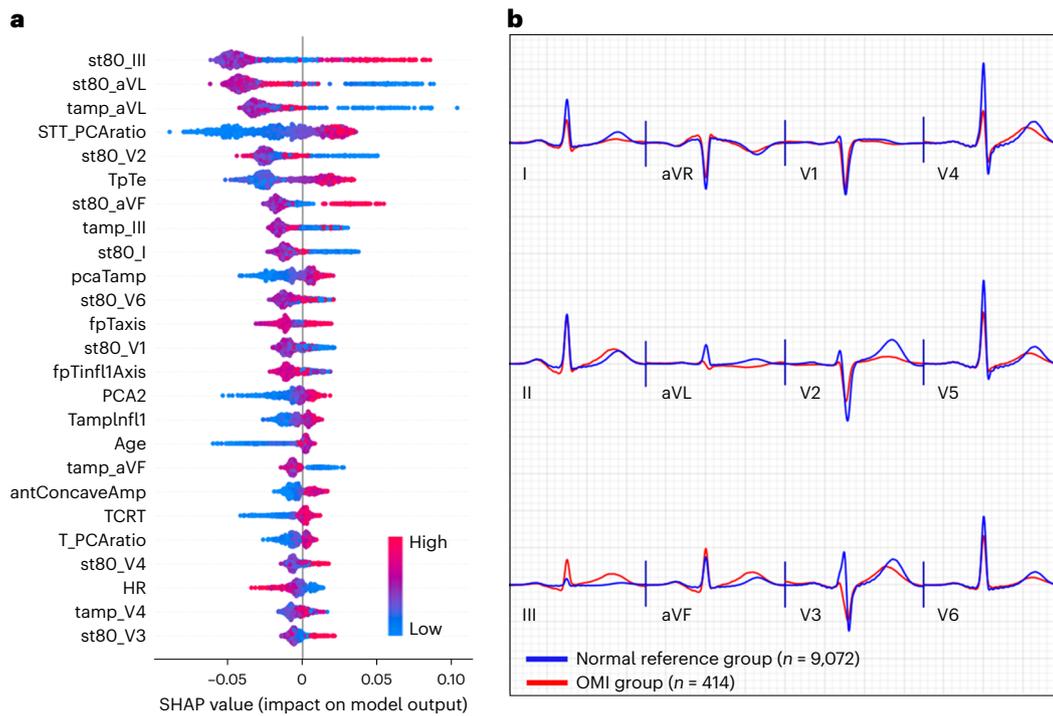


Fig. 3 | Model explainability for OMI detection. This figure shows SHAP values for the 25 most important features driving the predictions of the machine learning classifier in the derivation cohort (a) and the aggregate median beats of ECGs with OMI class (red) and the aggregate median beats of ECGs with normal sinus rhythm and no OMI (blue) (b). antConcaveAmp, the sum of concave

amplitudes in the anterior leads; fpTaxis, T axis in the frontal plane; HR, heart rate; Infl1, the first inflection point before T peak; ST80, ST amplitude at the J point + 80 ms; tamp, T amplitude; TCRT, total cosine R-to-T; TpTe, $T_{peak} - T_{end}$ interval.

OMI model was slightly less specific in patients with known coronary artery disease (CAD) (Fig. 4c). This finding aligns with recent evidence showing diminished NPV in patients with chest pain and known CAD³⁷.

Screening for any ACS event

We further built a model to screen for any potential ACS event at first medical contact. Using the same set of ECG features, we trained and optimized an RF classifier that denoted the likelihood of any ACS event. The model performed well during training (AUROC 0.88 (95% CI 0.87–0.90)) and generalized well during internal testing (AUROC 0.80 (95% CI 0.76–0.84)) (Extended Data Fig. 6). On external validation, the model continued to generalize well (AUROC 0.79 (95% CI 0.76–0.8)), outperforming the commercial system (AUROC 0.68 (95% CI 0.65–0.71), $P < 0.001$) and practicing clinicians (AUROC 0.72 (95% CI 0.69–0.74), $P < 0.001$). Our derived risk score provided a suboptimal rule-out classification for any ACS event (sensitivity 68.2% and NPV 92.5%) but provided superior rule-in accuracy (specificity 98.9% and PPV 82.5%).

Discussion

In this study, we developed and validated a machine learning algorithm for the ECG diagnosis of OMI in consecutive patients with chest pain recruited from multiple clinical sites in the United States. This model outperformed practicing clinicians and other commercial interpretation systems. The derived risk score provided superior rule-in and rule-out accuracy for OMI, boosting the sensitivity by ~28 percentage points and the precision by ~32 percentage points compared to reference standards. When combined with the judgment of experienced emergency personnel, our derived OMI risk score helped correctly reclassify one in three patients with chest pain. To our knowledge, this is the first study using machine learning methods and novel ECG features to optimize OMI detection in patients with acute chest pain and negative STEMI pattern on their presenting ECG.

Mapping myocardial ischemia, a problem of regional metabolic derangement, to coronary occlusion, a problem of diminished blood flow due to an atherosclerotic plaque rupture, is a complex process¹. Essentially, ischemia disproportionately distorts action potentials in different myocardial segments, resulting in tissue-scale currents, often called ‘injury’ currents. Previous studies mapped pronounced ST-elevation to transmural injury currents associated with total coronary occlusion. This has historically driven the current paradigm dichotomy of STEMI versus ‘others’ (any ACS other than STEMI) in determining who might benefit from emergent reperfusion therapy. However, nearly 65% of patients with ACS present with no ST-elevation on their baseline ECG^{35,38}, and, among the latter group, 24–35% have total coronary occlusion requiring emergent catheterization^{9–13}. Thus, determining who would benefit from reperfusion therapy remains an adjudicated diagnosis.

Conceptually, injury currents produced by ischemic cardiac cells are summative in nature, explaining how ST amplitude changes can get attenuated on the surface ECG (Extended Data Fig. 7). These injury currents, however, distort the propagation of both excitation and recovery pathways, altering the configuration of the QRS complex and the ST-T waveform altogether³⁹. Thus, a more comprehensive approach for the ECG detection of ischemia should focus on (1) evaluating temporal characteristics over entire waveform segments rather than the voltage at a given timepoint (for example, J point + 80 ms) and (2) evaluating lead-to-lead spatial characteristics in waveform morphology rather than absolute changes in isolated ECG leads¹.

This study identified several ECG patterns indicative of acute coronary occlusion beyond the criteria recommended by clinical guidelines⁵. Intriguingly, these ECG patterns overlap with those described in the literature. A consensus report in 2012 identified few ECG patterns that should be treated as STEMI equivalent during acute pain episodes: ST-depression in V1–V3; small inverted T waves in V1–V3; deep

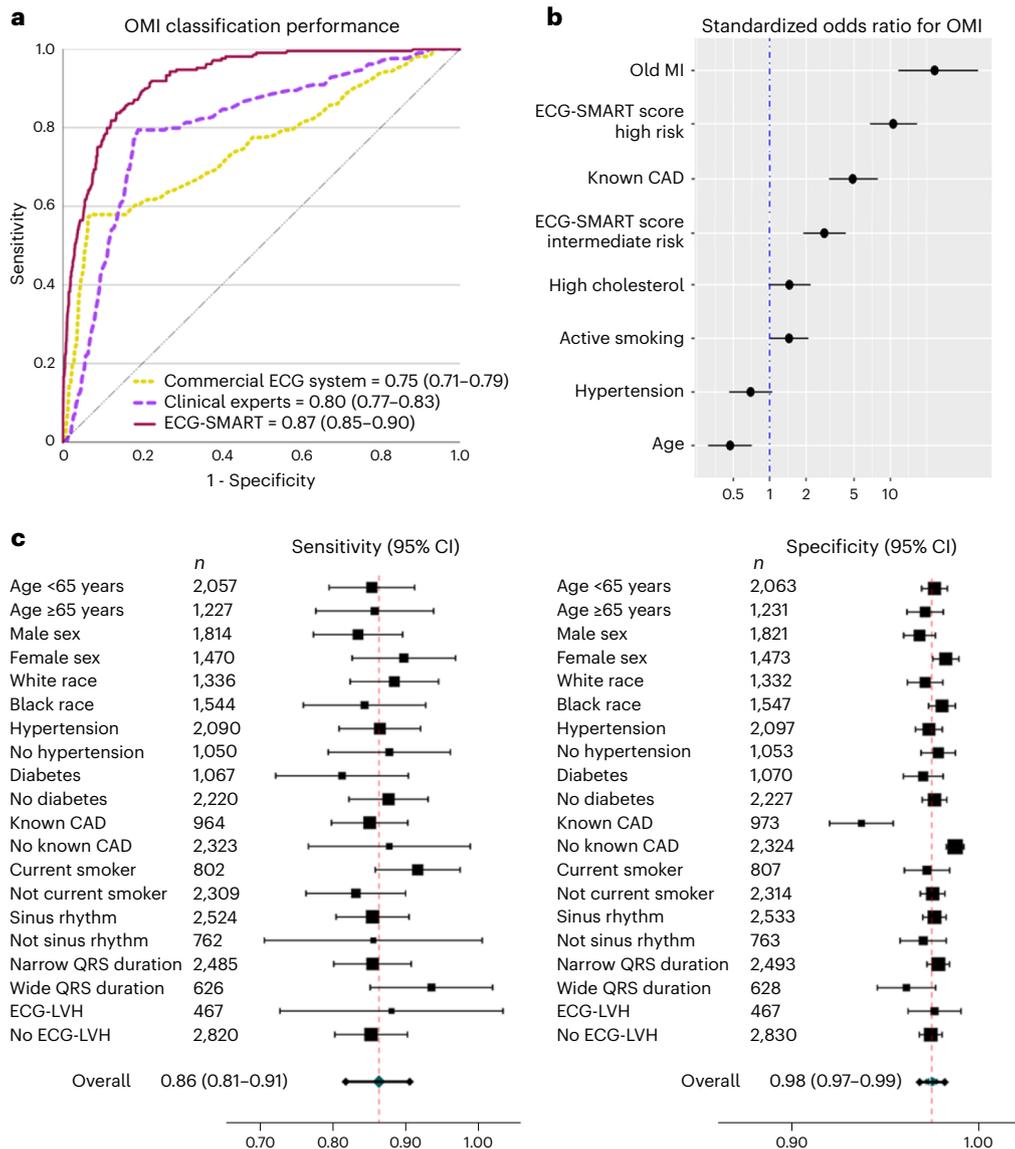


Fig. 4 | External validation of the ECG-SMART algorithm. a–c. This figure shows the classification performance of the machine learning model against other reference standards for detecting OMI on the external validation set ($n = 3,287$) (a), adjusted OR (center) with 95% CI (error bars) for the independent clinical predictors of OMI on the external validation set ($n = 3,287$) (b) and

the overall sensitivity and specificity (center) with 95% CI (error bars) of the derived OMI score, along with breakdown across subgroups based on age, sex, comorbidities and baseline ECG findings (c). The size of the center marker is proportionate to the sample size of the respective subgroup.

negative T waves in pre-cordial leads; widespread ST-depression; and prominent positive T waves²⁰. Similar ECG patterns were also described more recently: ST-depression in V1–V4 (versus V5–V6); reciprocal ST-depression with maximal ST-depression vector toward the apex (leads II and V5, with reciprocal STE in aVR); subtle ST-elevation; acute pathologic Q waves; hyperacute T waves; and loss of terminal S wave²¹. Many of these expert-driven patterns rely on assessing the proportion of repolarization amplitudes or area under the QRS amplitude. They also rely heavily on the visual assessment of waveform morphology and can introduce a high degree of subjectivity and variability among ECG interpreters. We demonstrated that machine learning models not only outperformed practicing clinicians in identifying OMI but also provided an objective, observer-independent approach to quantifying subtle ECG patterns associated with OMI.

Many of the data-driven features identified by our machine learning model are subtle and cannot be easily appreciated by clinical experts. T feature indices were among these most important features,

including $T_{peak}-T_{end}$ interval prolongation, T wave flattening and T wave characteristics at the inflection point preceding T_{peak} (Fig. 3a). Mechanistically, ischemic injury currents interfere with signal propagation, leading to longer activation time⁴⁰. These late activation potentials lead to a loss of terminal S wave and longer recovery time, both manifesting as T wave flattening, shifted T peak and loss of concavity at the initial T wave (Fig. 3b). These STEMI-equivalent patterns were previously described in the literature as small or negative T waves with widespread ST-depression or subtle ST-elevation^{20,21}. Another important subtle feature identified by our model was increased ventricular repolarization dispersion, measured using the ratio between the principal components of the ST-T waveforms (that is, principal component analysis (PCA) metrics), the direction of the T axis and the angle between activation and recovery pathways (for example, total cosine R-to-T). Injury currents disproportionately affect the duration and velocity of repolarization across different myocardial segments⁴¹, resulting in lead-to-lead variability in the morphology of the ST-T waveform^{22–24,39,42}.

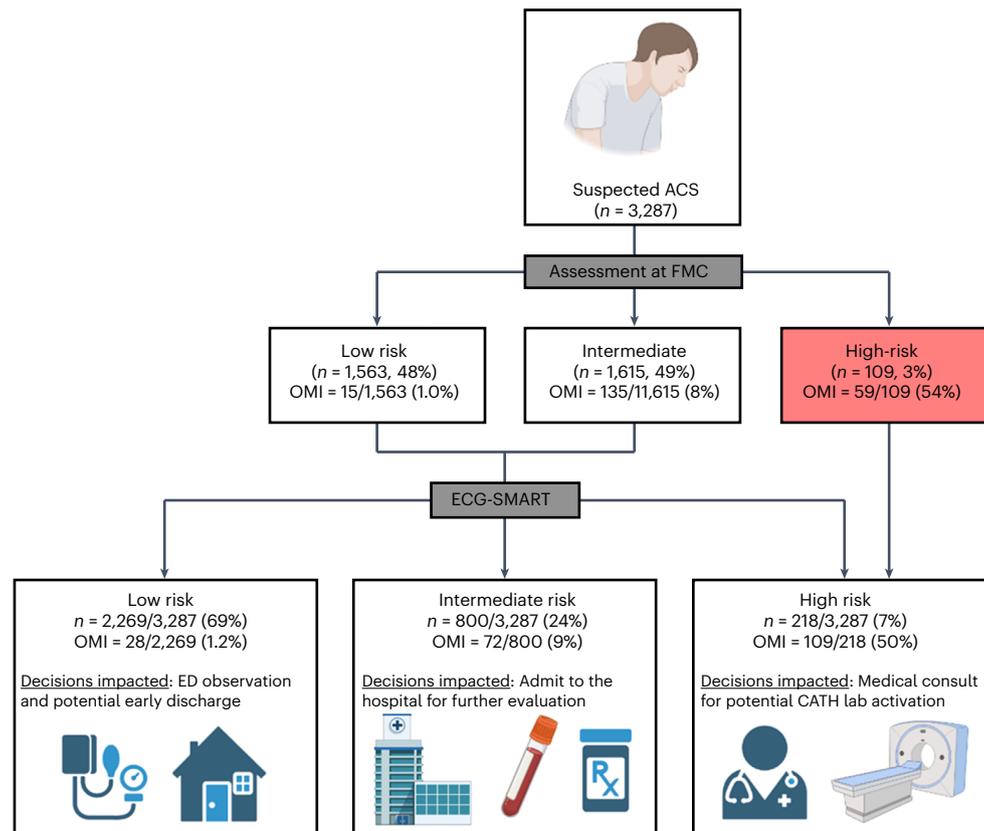


Fig. 5 | NRI of OMI risk score when integrated in the clinical workflow and concept of potential impact on subsequent clinical decisions. This figure describes the incremental gain of the derived risk score in reclassifying the initial triage decisions by emergency personnel at first medical contact and depicts

the concept of potential impact on subsequent clinical decisions. This figure was created with BioRender (credit to S.S.A.-Z.). CATH, catheterization; ED, emergency department; FMC, first medical contact.

These high-risk ECG patterns were previously described as a mixture of deep negative T waves and prominent/hyperacute T waves or reciprocal T wave changes^{20,21}. Our machine learning model provided a more comprehensive, quantitative approach to evaluating this subtle inter-lead variability in repolarization morphology.

Machine learning is well suited to address many challenges in 12-lead ECG interpretation. Myocardial ischemia distorts the duration and amplitude of the Q wave, R peak, R', QRS complex, ST segment and T wave as well as the morphology and configuration of these waveforms (for example, upsloping, downsloping, concavity, symmetry and notching). These distortions are lead specific yet come with dynamic inter-lead correlations. Thus, ECG interpretation involves many complex aspects and parameters, making it a highly dimensional, decision space problem¹. Few experienced clinicians excel in such pattern recognition,²¹ which explains why so many patients with OMI are not reperfused in a timely way; this is also why simple, rule-based commercial systems that use simple regression models are suboptimal for OMI detection. Machine learning algorithms can provide powerful tools to solve such highly dimensional, nonlinear mathematical representations found in 12-lead ECG data.

Although the literature on machine learning for the ECG diagnosis of coronary disease is ubiquitous, it comes with many serious limitations. First, many studies focused on detecting the known STEMI group^{34,35,43,44} rather than the critical OMI group without ST-elevation. Second, most previous work used open-source ECG datasets, such as PTB and PTB-XL⁴⁵, which are highly selected datasets that focus on ECG-adjudicated diagnoses. Our unique cohorts included unselected, consecutive patients with clinical profiles and disease prevalence like that seen in real-world settings. Third, many studies used a full range

of input features based on both ECG data and clinical data elements (for example, patient history, physical examination abnormalities, laboratory values and diagnostic tests)^{46–49}, which limits the applicability to real-world settings. Fourth, to our knowledge, most studies used a single derivation cohort for training and testing⁵⁰, without the use of an independent validation cohort. Finally, previous studies paid little attention to model explainability⁵¹, shedding little light on novel markers and pathways of ischemia than what is already known. Without explanation aids of clinical meaningfulness, machine learning models for ECG interpretation would have limited clinical utility⁵².

This study has important clinical implications. Our model can be integrated into systems of care for real-time deployment where risk score assignments can be made readily available to clinicians at the time of ECG acquisition. Enhanced decision support can help emergency personnel identify 85% more patients with critical coronary occlusion despite the absence of a STEMI pattern and without any loss in precision. Our models can also help inform care in more than 50% of patients in whom the initial assessment is indeterminate, placing 45% more patients in the low-risk group for OMI without any loss in NPV. This incremental gain in rule-in and rule-out accuracy can help re-allocate critical emergency resources to those in utmost need while optimizing the clinical workflow. This can impact numerous decisions at first medical contact, including targeted pre-hospital interventions, catheterization laboratory activation, administration of anti-ischemic therapies, hospital destination decisions, the need for medical consults, referrals for expedited diagnostic testing (for example, ECG and imaging scans) and early discharge decisions. Furthermore, until now, clinicians never had sensitive or highly specific tools that would allow the ultra-early identification of OMI in the absence of a STEMI pattern.

Such enhanced diagnostics can allow the design and implementation of prospective interventional trials to assess the therapeutic effectiveness of targeted interventions in this vulnerable group (for example, early upstream P2Y₁₂ inhibitor administration⁵³, emergent versus delayed reperfusion therapy⁵⁴ and glucose–insulin–potassium infusion⁵⁵).

Several limitations merit consideration. First, the features that we used for model building were based on manufacturer-specific software. There are known discrepancies between manufacturers in ECG pre-processing, which means that our models would need retraining when using different software for signal processing. Alternatively, deep neural networks can be used to analyze raw ECG signal without explicit feature engineering. However, these techniques require much training samples (for example, >10,000) and might not yield a meaningful improvement over feature engineering-based machine learning for traditional 12-lead ECG-based diagnosis⁵⁶. Second, we found slight differences between the derivation and validation cohorts in terms of disease prevalence and practicing clinicians' accuracy in ECG interpretation. These cohorts came from two different regions in the United States, and emergency medical systems (EMSs) follow state-specific protocols. It is possible that discrepancies in EMS protocols and in-hospital practices resulted in slight differences in types and proportions of patients who receive pre-hospital 12-lead ECGs and the corresponding outcome adjudications. However, it is reassuring that our models continued to generalize well among the study sites. Third, it is worth noting that our model for 'any ACS event' boosted the performance of only the rule-in arm. This means that a low-risk determination suggests that a given patient would unlikely have OMI, but they might have a less subtle phenotype of NSTEMI-ACS that does not require reperfusion therapy. It is likely that serial ECG testing might improve the detection of missed events where patients might switch to a higher-risk category in the following hours³⁴, but this remains to be confirmed. Coronary occlusion is a dynamic process that evolves over time, so an initial low-risk class by our models should not lead to a lower level of active monitoring. Finally, although this study used prospective patients, all analyses were completed offline. Prospective validation where OMI probabilities and decision support is provided in real time is warranted.

In conclusion, we developed and externally validated models for the ECG diagnosis of OMI in 7,313 patients with chest pain from multiple sites in the United States. The results demonstrated the superiority of machine learning in detecting subtle ischemic ECG changes indicative of OMI in an observer-independent approach. These models outperformed practicing clinicians and commercial ECG interpretation software, significantly boosting precision and recall. ECG features driving our models were evaluated, providing plausible mechanistic links to myocardial injury. Our derived OMI risk score provided enhanced rule-in and rule-out accuracy when compared to HEAR score, and, when combined with the clinical judgment of trained emergency personnel, this score helped correctly reclassify one in three patients with chest pain. The benefits of this new clinical pathway in terms of clinical outcomes should be evaluated in prospective trials. Future work should also focus on the prospective deployment where OMI probabilities and decision support is provided in real time.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-023-02396-3>.

References

- Al-Zaiti, S., Macleod, M. R., Van Dam, P. M., Smith, S. W. & Birnbaum, Y. Emerging ECG methods for acute coronary syndrome detection: recommendations & future opportunities. *J. Electrocardiol.* **74**, 65–72 (2022).
- Birnbaum, Y. et al. ECG diagnosis and classification of acute coronary syndromes. *Ann. Noninvasive Electrocardiol.* **19**, 4–14 (2014).
- Goodacre, S. et al. Clinical diagnosis of acute coronary syndrome in patients with chest pain and a normal or non-diagnostic electrocardiogram. *Emerg. Med. J.* **26**, 866–870 (2009).
- Ioannidis, J. P., Salem, D., Chew, P. W. & Lau, J. Accuracy and clinical effect of out-of-hospital electrocardiography in the diagnosis of acute cardiac ischemia: a meta-analysis. *Ann. Emerg. Med.* **37**, 461–470 (2001).
- Thygesen, K. et al. What's new in the Fourth Universal Definition of Myocardial Infarction?. *Eur. Heart J.* **39**, 3757–3758 (2018).
- Gulati, M. et al. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR guideline for the evaluation and diagnosis of chest pain. *J. Am. Coll. Cardiol.* **78**, e187–e285 (2021).
- Levine, G. N. et al. 2015 ACC/AHA/SCAI focused update on primary percutaneous coronary intervention for patients with ST-elevation myocardial infarction: an update of the 2011 ACCF/AHA/SCAI guideline for percutaneous coronary intervention and the 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction. *J. Am. Coll. Cardiol.* **67**, 1235–1250 (2016).
- Amsterdam, E. A. et al. 2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes: executive summary. *Circulation* **130**, 2354–2394 (2014).
- Dixon, W. C. et al. Anatomic distribution of the culprit lesion in patients with non-ST-segment elevation myocardial infarction undergoing percutaneous coronary intervention: findings from the National Cardiovascular Data Registry. *J. Am. Coll. Cardiol.* **52**, 1347–1348 (2008).
- Wang, T. Y. et al. Multivessel vs culprit-only percutaneous coronary intervention among patients 65 years or older with acute myocardial infarction. *Am. Heart J.* **172**, 9–18 (2016).
- Karwowski, J. et al. Relationship between infarct artery location, acute total coronary occlusion, and mortality in STEMI and NSTEMI patients. *Pol. Arch. Intern. Med.* **127**, 401–411 (2017).
- Figueras, J. et al. Area at risk and collateral circulation in a first acute myocardial infarction with occluded culprit artery. STEMI vs non-STEMI patients. *Int. J. Cardiol.* **259**, 14–19 (2018).
- Tanaka, T. et al. Comparison of coronary atherosclerotic disease burden between ST-elevation myocardial infarction and non-ST-elevation myocardial infarction: non-culprit Gensini score and non-culprit SYNTAX score. *Clin. Cardiol.* **44**, 238–243 (2021).
- Aslanger, E. K., Meyers, H. P., Bracey, A. & Smith, S. W. The STEMI/nonSTEMI dichotomy needs to be replaced by occlusion MI vs. non-occlusion MI. *Int. J. Cardiol.* **330**, 15 (2021).
- Avdikos, G., Michas, G. & Smith, S. W. From Q/non-Q myocardial infarction to STEMI/NSTEMI: why it's time to consider another simplified dichotomy; a narrative literature review. *Arch. Acad. Emerg. Med.* **10**, e78 (2022).
- Aslanger, E. K., Meyers, H. P. & Smith, S. W. STEMI: a transitional fossil in MI classification? *J. Electrocardiol.* **65**, 163–169 (2021).
- Meyers, H. P. et al. Comparison of the ST-elevation myocardial infarction (STEMI) vs. NSTEMI and occlusion MI (OMI) vs. NOMI paradigms of acute MI. *J. Emerg. Med.* **60**, 273–284 (2021).
- Bouزيد, Z. et al. In search of an optimal subset of ECG features to augment the diagnosis of acute coronary syndrome at the emergency department. *J. Am. Heart Assoc.* **10**, e017871 (2021).
- Meyers, H. P. et al. Ischemic ST-segment depression maximal in V1–V4 (versus V5–V6) of any amplitude is specific for occlusion myocardial infarction (versus nonocclusive ischemia). *J. Am. Heart Assoc.* **10**, e022866 (2021).

20. Birnbaum, Y. et al. Common pitfalls in the interpretation of electrocardiograms from patients with acute coronary syndromes with narrow QRS: a consensus report. *J. Electrocardiol.* **45**, 463–475 (2012).
21. Meyers, H. P. et al. Accuracy of OMI ECG findings versus STEMI criteria for diagnosis of acute coronary occlusion myocardial infarction. *Int. J. Cardiol. Heart Vasc.* **33**, 100767 (2021).
22. Al-Zaiti, S., Callaway, C. W., Kozik, T. M., Carey, M. & Pelter, M. Clinical utility of ventricular repolarization dispersion for real-time detection of non-ST elevation myocardial infarction in emergency departments. *J. Am. Heart Assoc.* **4**, e002057 (2015).
23. Al-Zaiti, S. et al. Evaluation of beat-to-beat ventricular repolarization lability from standard 12-lead ECG during acute myocardial ischemia. *J. Electrocardiol.* **50**, 717–724 (2017).
24. Al-Zaiti, S. et al. Spatial indices of repolarization correlate with non-ST elevation myocardial ischemia in patients with chest pain. *Med. Biol. Eng. Comput.* **56**, 1–12 (2018).
25. Sharma, A. et al. Interobserver variability among experienced electrocardiogram readers to diagnose acute thrombotic coronary occlusion in patients with out of hospital cardiac arrest: impact of metabolic milieu and angiographic culprit. *Resuscitation* **172**, 24–31 (2022).
26. Gregg, R. E., Yang, T., Smith, S. W. & Babaeizadeh, S. ECG reading differences demonstrated on two databases. *J. Electrocardiol.* **69**, 75–78 (2021).
27. Cook, D. A., Oh, S.-Y. & Pusic, M. V. Accuracy of physicians' electrocardiogram interpretations: a systematic review and meta-analysis. *JAMA Intern. Med.* **180**, 1461–1471 (2020).
28. McRae, A. D. et al. Undetectable concentrations of an FDA-approved high-sensitivity cardiac troponin T assay to rule out acute myocardial infarction at emergency department arrival. *Acad. Emerg. Med.* **24**, 1267–1277 (2017).
29. Body, R. & Mahler, S. Welcome to the real world: do the conditions of FDA approval devalue high sensitivity troponin? *Acad. Emerg. Med.* **24**, 1278–1280 (2017).
30. Wereski, R. et al. High-sensitivity cardiac troponin concentrations at presentation in patients with ST-segment elevation myocardial infarction. *JAMA Cardiol.* **5**, 1302–1304 (2020).
31. Cotterill, P. G., Deb, P., Shrank, W. H. & Pines, J. M. Variation in chest pain emergency department admission rates and acute myocardial infarction and death within 30 days in the Medicare population. *Acad. Emerg. Med.* **22**, 955–964 (2015).
32. Kang, M. G. et al. Cardiac mortality benefit of direct admission to percutaneous coronary intervention-capable hospital in acute myocardial infarction: community registry-based study. *Medicine (Baltimore)* **100**, e25058 (2021).
33. Quinn, T. et al. Effects of prehospital 12-lead ECG on processes of care and mortality in acute coronary syndrome: a linked cohort study from the Myocardial Ischaemia National Audit Project. *Heart* **100**, 944–950 (2014).
34. Bouzid, Z. et al. Incorporation of serial 12-lead electrocardiogram with machine learning to augment the out-of-hospital diagnosis of non-ST elevation acute coronary syndrome. *Ann. Emerg. Med.* **81**, 57–69 (2023).
35. Al-Zaiti, S. et al. Machine learning-based prediction of acute coronary syndrome using only the pre-hospital 12-lead electrocardiogram. *Nat. Commun.* **11**, 3966 (2020).
36. Stopyra, J. P. et al. Prehospital modified HEART score predictive of 30-day adverse cardiac events. *Prehosp. Disaster Med.* **33**, 58–62 (2018).
37. Ashburn, N. P. et al. Performance of the European Society of Cardiology 0/1-hour algorithm with high-sensitivity cardiac troponin T among patients with known coronary artery disease. *JAMA Cardiol.* **8**, 347–356 (2023).
38. Sabatine, M. S. et al. Combination of quantitative ST deviation and troponin elevation provides independent prognostic and therapeutic information in unstable angina and non-ST-elevation myocardial infarction. *Am. Heart J.* **151**, 25–31 (2006).
39. Lux, R. L. Non-ST-segment elevation myocardial infarction: a novel and robust approach for early detection of patients at risk. *J. Am. Heart Assoc.* **4**, e002279 (2015).
40. Marrusa, S., Zhangc, M. & Arthurb, M. Identification of acute coronary syndrome via activation and recovery times in body-surface mapping and inverse electrocardiography. *Int. J. Bioelectromagnetism* **21**, 1–6 (2019).
41. Lux, R. L. Basis and ECG measurement of global ventricular repolarization. *J. Electrocardiol.* **50**, 792–797 (2017).
42. Al-Zaiti, S., Runco, K. & Carey, M. Increased T-wave complexity can indicate subclinical myocardial ischemia in asymptomatic adults. *J. Electrocardiol.* **44**, 684–688 (2011).
43. Forberg, J. L. et al. In search of the best method to predict acute coronary syndrome using only the electrocardiogram from the emergency department. *J. Electrocardiol.* **42**, 58–63 (2009).
44. Green, M. et al. Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room. *Artif. Intell. Med.* **38**, 305–318 (2006).
45. Hong, S., Zhou, Y., Shang, J., Xiao, C. & Sun, J. Opportunities and challenges of deep learning methods for electrocardiogram data: a systematic review. *Comput. Biol. Med.* **122**, 103801 (2020).
46. Baxt, W. G. & Skora, J. Prospective validation of artificial neural network trained to identify acute myocardial infarction. *Lancet* **347**, 12–15 (1996).
47. Tsien, C. L., Fraser, H. S., Long, W. J. & Kennedy, R. L. Using classification tree and logistic regression methods to diagnose myocardial infarction. *Stud. Health Technol. Inform.* **52**, 493–497 (1998).
48. Berikol, G. B., Yildiz, O. & Özcan, İT. Diagnosis of acute coronary syndrome with a support vector machine. *J. Med. Syst.* **40**, 84 (2016).
49. Wu, C.-C. et al. An artificial intelligence approach to early predict non-ST-elevation myocardial infarction patients with chest pain. *Comput. Methods Prog. Biomed.* **173**, 109–117 (2019).
50. Brisk, R. et al. Neural networks for ischaemia detection: revolution or red herring? A systematic review and meta-analysis. *J. Electrocardiol.* **69**, 79 (2021).
51. Bond, R., Finlay, D., Al-Zaiti, S. S. & Macfarlane, P. Machine learning with electrocardiograms: a call for guidelines and best practices for 'stress testing' algorithms. *J. Electrocardiol.* **69S**, 1–6 (2021).
52. Elul, Y., Rosenberg, A. A., Schuster, A., Bronstein, A. M. & Yaniv, Y. Meeting the unmet needs of clinicians from AI systems showcased for cardiology with deep-learning-based ECG analysis. *Proc. Natl Acad. Sci. USA* **118**, e2020620118 (2021).
53. Cohen, M. V. & Downey, J. M. What are optimal P2Y12 inhibitor and schedule of administration in patients with acute coronary syndrome? *J. Cardiovasc. Pharmacol. Ther.* **25**, 121–130 (2020).
54. Tziakas, D., Chalikias, G., Al-Lamee, R. & Kaski, J. C. Total coronary occlusion in non ST elevation myocardial infarction: time to change our practice? *Int. J. Cardiol.* **329**, 1–8 (2021).
55. Udelson, J. E., Selker, H. P. & Braunwald, E. Glucose-insulin-potassium therapy for acute myocardial infarction: 50 years on and time for a relook. *Circulation* **146**, 503–505 (2022).
56. Zvuloni, E., Read, J., Ribeiro, A. H., Ribeiro, A. L. P. & Behar, J. A. On merging feature engineering and deep learning for diagnosis, risk-prediction and age estimation based on the 12-lead ECG. *IEEE Trans. Biomed. Eng.* **70**, 2227–2236 (2022).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless

indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Methods

Ethics statement

The derivation cohort included pre-hospital data from the City of Pittsburgh Bureau of Emergency Medical Services and in-hospital data from three tertiary care hospitals from the University of Pittsburgh Medical Center (UPMC) healthcare system: UPMC Presbyterian Hospital, UPMC Shadyside Hospital and UPMC Mercy Hospital (Pittsburgh, Pennsylvania). All consecutive eligible patients were recruited under a waiver of informed consent. This observational trial was approved by the institutional review board of the University of Pittsburgh and was registered at <https://www.clinicaltrials.gov/> (identifier NCT04237688). The analyses described in this paper were pre-specified by the trial protocol that was funded by the National Institutes of Health. The first external validation cohort included data from Orange County Emergency Medical Services (Chapel Hill, North Carolina). This study actively consented eligible patients and was approved by the institutional review board of the University of North Carolina at Chapel Hill. The second external validation cohort included data from Mecklenburg County Emergency Medical Services and Atrium Health (Charlotte, North Carolina). Data were collected through a healthcare registry, and all consecutive eligible patients were enrolled under a waiver of informed consent. This study was also approved by the institutional review board of the University of North Carolina at Chapel Hill. These two external datasets were collected by the same local investigative team and were similar in terms of age, sex and disease prevalence. Thus, we combined these two datasets into one cohort for external validation purposes.

Study design and data collection

This was a prospective, observational cohort study. The methods for each study cohort were described in detail elsewhere^{57,58}. All study cohorts enrolled adult patients with an emergency call for non-traumatic chest pain or anginal equivalent symptoms (arm, shoulder or jaw pain, shortness of breath, diaphoresis or syncope). Eligible patients were transported by an ambulance and had at least one recorded pre-hospital 12-lead ECG. There were no selective exclusion criteria based on sex, race, comorbidities or acuity of illness. For this pre-specified analysis, we included only non-duplicate ECGs from unique patient encounters, and we removed patients with pre-hospital ECGs showing ventricular tachycardia or ventricular fibrillation (that is, these patients are managed by ACLS algorithms). We also removed patients with confirmed pre-hospital STEMI, which included machine-generated ***ACUTE MI*** warning, EMS documentation of STEMI and medical consult for potential catheterization laboratory activation.

Independent reviewers extracted data elements from hospital systems on all patients meeting eligibility criteria. If a pre-hospital ECG had no patient identifiers, we used a probabilistic matching approach to link each encounter with the correct hospital record. This previously validated data linkage protocol was based on the ECG-stamped birth date, sex and date/time logs as well as based on EMS dispatch logs and receiving hospital records. All probabilistic matches were manually reviewed by research specialists for accuracy. The match success rate ranged from 98.6% to 99.8%.

Clinical outcomes

Adjudications were made by independent reviewers at each local site after reviewing all available medical records within 30 d of the indexed encounter. Reviewers were blinded from all ECG analyses and models' predictions. OMI was defined as coronary angiographic evidence of an acute culprit lesion in at least one of the three main coronary arteries (left anterior descending (LAD), left circumflex (LCX) and right coronary artery (RCA)) or their primary branches with TIMI flow grade of 0–1. TIMI flow grade of 2 with severe coronary narrowing >70% and peak troponin of 5–10.0 ng ml⁻¹ was also considered indicative of OMI^{17,21}. These adjudications were made by two independent reviewers.

The kappa coefficient statistic between the two reviewers was 0.771 (that is, substantial agreement). All disagreements were resolved by a third reviewer.

ACS was defined per the Fourth Universal Definition of Myocardial Infarction as the presence of symptoms of ischemia (that is, diffuse discomfort in the chest, upper extremity, jaw or epigastric area for more than 20 min) and at least one of the following criteria: (1) subsequent development of labile, ischemic ECG changes (for example, ST changes and T inversion) during hospitalization; (2) elevation of cardiac troponin (that is, >99th percentile) during the hospital stay with rise and/or drop on serial testing; (3) coronary angiography demonstrating greater than 70% stenosis, with or without treatment; and/or (4) functional cardiac evaluation (stress testing) that demonstrates ECG, echocardiographic or radionuclide evidence of focal cardiac ischemia⁵. Patients with type 2 myocardial infarction and pre-existing subacute coronary occlusion were labeled as negative for ACS and OMI. This included around 10% of patients with positive troponin but with no rise and/or drop in concentration on serial testing (that is, chronic leak) or with troponin leak attributed to non-coronary occlusive conditions, such as pericarditis. On a randomly selected small subset of patients ($n = 1,209$), the kappa coefficient statistic for ACS adjudication ranged from 0.846 to 0.916 (that is, substantial to perfect agreement).

ECG methods

Pre-hospital ECGs were obtained in the field by paramedics as part of routine care. ECGs were acquired using either Heart Start MRX (Philips Healthcare) or LIFEPAK-15 (Physio-Control) monitor–defibrillator devices. All digital 12-lead ECGs were acquired at a sampling rate of 500 samples per second (0.05–150 Hz) and transmitted to the respective EMS agency and receiving hospital. Digital ECG files were exported in .xml format and stored in a secondary server at each local site. ECG images were de-identified and manually annotated by independent reviewers or research specialists; ECGs with poor quality or missing leads were removed from the study. Next, digital .xml files were transmitted to the Philips Advanced Algorithm Research Center (Cambridge, Massachusetts) for offline analysis.

ECG featurization was described in detail elsewhere¹⁸. In brief, ECG signal pre-processing and feature extraction were performed using manufacturer-specific software (Philips DXL diagnostic 12/16 lead ECG analysis program). ECG signals were first pre-processed to remove noise, artifacts and baseline wander. Ectopic beats were removed, and representative median beats were calculated for each lead. Median beats refer to the representative average (or median) of the sequential beats in a given ECG lead after temporal alignment of R peaks. Next, we used the root mean square (RMS) signal to identify global waveform fiducials, including the onset, offset and peak of the P wave, QRS complex and T wave. Lead-specific fiducials were then identified to further segment individual waveforms into Q, R, R', S, S' and J point.

We then computed a total of 554 ECG features based on (1) the amplitude, duration, area, slope and/or concavity of global and lead-specific waveforms; (2) the QRS and T axes and angles in the frontal, horizontal, spatial, x - y , x - z and y - z planes, including directions at peak, inflection point and initial/terminal loops; (3) eigenvalues of the principal components of orthogonal ECG leads (I, II and V1–V6), including PCA ratios for individual ECG waveform segments; and (4) T loop morphology descriptors. Features with zero distribution were removed to prevent representation bias.

Next, we previously identified an optimal parsimonious list of the most important ECG features that are mechanistically linked to cardiac ischemia as described in detail elsewhere¹⁸. In brief, to prevent omitted feature bias, we used a hybrid approach that combines domain knowledge with a data-driven strategy. First, clinical scientists identified 24 classical features that are known to correlate with cardiac ischemia (that is, lead-specific ST and T wave amplitudes). Next, starting with a comprehensive list of 554 candidate features, we used data-driven

algorithms (for example, recursive feature elimination and LASSO) to identify 198 supplemental features potentially related to ischemia. LASSO selects features with non-zero coefficients after L1 norm regularization, and recursive feature elimination uses repeated regression iterations to identify the features that have significant impact on model predictions. We then examined the feature pairs in this expanded list of 222 features and removed features with very high collinearity scores that contains redundant information (for example, we kept QTc if both QT and QTc were selected by the model). Finally, we used feature importance ranking to identify the most parsimonious subset of features that are complementary and can boost the classification performance. This hybrid approach eventually yielded a subset of 73 features that can serve as plausible markers of ischemia¹⁸.

Machine learning methods

We followed best practices recommended by 'ROBUST-ML' and 'ECG-AI stress test' checklists to design and benchmark our machine learning algorithms^{51,59}. To prevent measurement bias, ECG features were manually reviewed to identify erroneous calculations. Physiologically plausible outliers were replaced with ± 3 s.d. On average, each feature had a 0.34% missingness rate (range, 0.1–1.6%). Thus, we imputed missing values with the mean, median or mode of that feature after consultation with clinical experts. ECG metrics were then z-score normalized and used as input features in machine learning models. The derivation and validation datasets were cleaned independently to prevent data leakage. Both cohorts were recruited over the same time window, suggesting the lack of temporal bias. To prevent potential mismatch with intended use, input features for model development included only ECG data plus the machine-stamped age. No other clinical data were used for model building.

We randomly split the derivation cohort into an 80% training set and a 20% internal testing set. On the training set, we fit 10 machine learning classifiers: regularized logistic regression, linear discriminant analysis, support vector machine (SVM), Gaussian naive Bayes, RF, gradient boosting machine, extreme gradient boosting, stochastic gradient descent logistic regression, *k*-nearest neighbors and artificial neural networks. Each classifier was optimized over 10-fold cross validation to fine-tune hyperparameters. After selecting optimal hyperparameters, models were retrained on the entire training subset to derive final weights and create a lockout model to evaluate on the hold-out test set. We calibrated our classifiers to produce a probabilistic output that can be interpreted as a confidence level (probability risk score). Trained models were compared using the AUROC curve with Wilcoxon signed-rank test for pairwise comparisons. ROC-optimized cutoffs were chosen using the Youden index, and classifications on confusion matrix were compared using McNemar's test.

The RF classifier achieved high accuracy on the training set (low bias) with a relatively small drop in performance on the test set (low variance), indicating an acceptable bias–variance tradeoff and low risk of overfitting (Extended Data Fig. 8). Although the SVM model had lower variance on the test set, when compared to the RF model there were no significant differences in AUROC (DeLong's test) or their binary classifications (McNemar's test). Moreover, there were no differences between the RF and SVM models in terms of Kolmogorov–Smirnov goodness of fit (0.716 versus 0.715) or the Gini purity index (0.82 versus 0.85). Due to its scalability and intuitive architecture, we chose the probability output of the RF model to build our derived OMI score. We generated density plots of these probability scores for positive and negative classes and selected classification thresholds for low-risk, intermediate-risk and high-risk groups based on pre-specified NPV > 0.99 and true-positive rate > 0.50. Finally, we used the lock-out RF classifier to generate probability scores and risk classes on the completely unseen external validation cohort. The code to generate probability scores is included with the supplementary materials of this manuscript.

Reference standard

To reduce the risk of evaluation bias, we benchmarked our machine learning models against multiple reference standards used during routine care in clinical practice. First, we used a commercial, FDA-approved ECG interpretation software (Philips DXL diagnostic algorithm) to denote the likelihood of ischemic myocardial injury. This likelihood (yes/no) was based on a composite of the following: (1) diagnostic codes for '>>Acute MI<<', including descriptive statements that denote 'acute', 'recent', 'age indeterminate', 'possible' or 'probable'; and (2) diagnostic codes for '>>Acute Ischemia<<', including descriptive statements that denote 'possible', 'probable' or 'consider'. Diagnostic statements that denoted 'old' [infarct], 'nonspecific' [ST depression] or 'secondary to' [LVH or high heart rate] were excluded from this composite reference standard.

We also used practicing clinicians' overread of ECGs to denote the likelihood of ischemic myocardial injury on a given ECG (yes/no) when a STEMI pattern does not exist, which is congruent with how emergency department physicians evaluate these patients in clinical practice. Independent physician reviewers annotated each 12-lead ECG image as per the Fourth Universal Definition of Myocardial Infarction criteria⁵, including two contiguous leads with ST-elevation (≥ 0.2 mV for V2–V3 in men ≥ 40 years of age and ≥ 2.5 mm in men <40 years of age; ≥ 0.15 mV for V2–V3 in women; or ≥ 0.1 mV in other leads) or ST-depression (new horizontal or downsloping depression ≥ 0.05 mV), with or without T wave inversion (> 0.1 mV in leads with prominent R wave or R/S ratio > 1). Reviewers were also prompted to use their clinical judgment to identify highly suspicious ischemic changes (for example, reciprocal changes and hyperacute T waves) as well as to account for potential confounders (for example, BBBs and early repolarization). On a randomly selected subset of patients in the derivation cohort ($n = 1,646$), the kappa coefficient statistic between two emergency physicians who interpreted the ECGs was 0.568 (that is, moderate agreement). A third reviewer was used to adjudicate discrepancies on this randomly selected subset. Similarly, on a randomly selected subset of patients in the external validation cohort ($n = 375$), the kappa coefficient statistic between the two board-certified cardiologists who interpreted the ECGs was 0.690 (that is, substantial agreement).

Finally, given that clinicians largely depend on risk scores to triage patients in the absence of STEMI, which would greatly affect how patients with OMI are diagnosed and treated in clinical practice, we compared our derived OMI risk score against the HEART score. This score is commonly used in US hospitals, and it has been well validated for triaging patients in the emergency department⁶⁰. The HEART score is based on the patient's history at presentation, ECG interpretation, age, risk factors and initial troponin values (range, 0–10). This score places patients in low-risk (0–3), intermediate-risk (4–6) and high-risk (7–10) groups. Given that troponin results are not usually available at first medical contact, we used a modified HEAR score after dropping the troponin values, which has also been previously validated for use by paramedics before hospital arrival³⁶. The comparison against the HEART score herein focused on establishing the incremental gain of using the derived OMI score over routine care at initial triage. We compared how the new risk classes assigned by our derived OMI score agree with or differ from the risk classes assigned by the HEART score, which could inform potential incremental gain over routine care.

Statistical analysis

Descriptive statistics were reported as mean \pm s.d. or n (%). Missing data were assessed for randomness and handled during ECG feature selection (see 'Machine learning methods' subsection above). Normality of distribution was assessed before hypothesis testing where deemed necessary. ECG features were z-score normalized as part of standard input architectures for machine learning models. Comparisons between cohorts were performed using the chi-square test (for discrete variables) and independent samples *t*-test or the Mann–Whitney *U*-test (for continuous variables). The level of significance was set at an alpha of 0.05 for two-tailed hypothesis testing where applicable.

All diagnostic accuracy values were reported as per Standards for Reporting Diagnostic Accuracy Studies (STARD) recommendations. We reported classification performance using AUROC curve, sensitivity (recall), specificity, PPV (precision) and NPV, along with 95% CI where applicable. For 10-fold cross validation, we compared the multiple classifiers using the Wilcoxon signed-rank test (for AUROC curves) and McNemar's test (for confusion matrices). We derived low-risk, intermediate-risk and high-risk categories for the final classifier using kernel density plot estimates between classes. The adequacy of these risk classes was evaluated using log-rank chi-square of accumulative risk for clinically important outcomes over the length of stay during the indexed admission.

For assessing the incremental gain in classification performance, we compared the AUROC of the final model against reference standards using DeLong's test. For ease of comparison, the confidence bounds for AUROC of the reference standards (commercial system and practicing clinicians) were generated using 1,000 bootstrap samples. To place the incremental gain value in a broader context of the clinical workflow, we also computed the NRI index of our model against the HEAR score during the initial assessment at first medical contact. Risk scores are an integral part of clinical workflow in patients with suspected ACS who do not meet STEMI criteria. As per STARD recommendations, the NRI index evaluates the net gain between up-triage and down-triage when correctly reclassifying risk class assignments of an 'old' test (HEART score) using a 'new' test (the derived OMI score).

We used logistic regression to identify the independent predictive value of OMI risk classes. We used variables significant in univariate analysis and then built multivariate models with the stepwise backward selection method using Wald chi-square criteria. We reported ORs with 95% CI for all significant predictors. All analyses were completed using Python version 3.8.5 and SPSS version 24.

Reporting Summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The ECG-SMART trial makes use of extracted ECG features to train and evaluate an RF classifier to denote the probability of OMI. The ECG features used in the derivation and external validation datasets, along with linked clinical outcomes, are publicly available through GitHub (<https://github.com/zeineb-bouid/sharing-github-nature-medicine.git>). Researchers wanting the source binary files to compute their own features should contact the corresponding author to arrange for proper approvals and institutional data use agreements. Interested researchers from non-commercial entities can submit a request by emailing the corresponding author at ssa33@pitt.edu. Requests will be processed within a 2-week timeframe.

Code availability

The Python codes to evaluate these models, along with the derivation and external validation datasets, are available through GitHub (<https://github.com/zeineb-bouid/sharing-github-nature-medicine.git>).

References

- Al-Zaiti, S. S., Martin-Gill, C., Sejdic, E., Alrawashdeh, M. & Callaway, C. Rationale, development, and implementation of the Electrocardiographic Methods for the Prehospital Identification of Non-ST Elevation Myocardial Infarction Events (EMPIRE). *J. Electrocardiol.* **48**, 921–926 (2015).
- Zègre-Hemsey, J. K. Prehospital ECG with ST-depression and T-wave inversion are associated with new onset heart failure in individuals transported by ambulance for suspected acute coronary syndrome. *J. Electrocardiol.* **69S**, 23–28 (2021).
- Al-Zaiti, S. S. et al. A clinician's guide to understanding and critically appraising machine learning studies: a checklist for Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML). *Eur. Heart J. Digit. Health* **3**, 125–140 (2022).
- Al-Zaiti, S. S. et al. Comparison of clinical risk scores for triaging high-risk chest pain patients at the emergency department. *Am. J. Emerg. Med.* **37**, 461–467 (2019).

Acknowledgements

This study was funded by grants from the National Institutes of Health (NIH), the National Heart, Lung, and Blood Institute (NHLBI), the National Center for Advancing Translational Sciences (NCATS) and the National Institute for Nursing Research (NINR) through grants R01HL137761 (S.S.A.-Z.), UL1TR001857 (S.S.A.-Z.), K23NR017896 (J.K.Z.-H.) and KL2TR002490 (J.K.Z.-H.).

Author contributions

S.S.A.-Z., C.M.-G., J.K.Z.-H., S.S., E.S. and C.W.C. conceived the study, secured funding and supervised the research. Y.B. and S.W.S. advised on the scientific direction of the study. S.S.A.-Z., J.K.Z.-H., Z.F., M.O.A., K.K.-P. and S.H. supervised dataset creation and annotation. S.S.A.-Z., C.M.-G., J.K.Z.-H., Z.F., M.O.A. and S.H. supervised clinical outcomes adjudication. R.E.G., S.S.A.-Z., M.A., Z.B., P.V.D. and N.R. supervised ECG signal processing and feature extraction. S.S.A.-Z., Z.B., N.R., S.M.S. and E.S. performed feature engineering, machine learning modeling, statistical analysis and results interpretation. S.S.A.-Z. drafted the manuscript. All authors critically revised the manuscript for important intellectual content. All authors provided their final approval of the version to be published. All authors are accountable for the work.

Competing interests

US Patent 10820822; owner: University of Pittsburgh; inventors: S.S.A.-Z., E.S. and C.W.C. This patent describes methods and systems for identifying increased likelihood of non-ST elevation myocardial infarction (NSTEMI) in a patient based on ECG data. This patent is not under any licensing or commercial agreement whatsoever. The remaining authors declare no competing interests.

Additional information

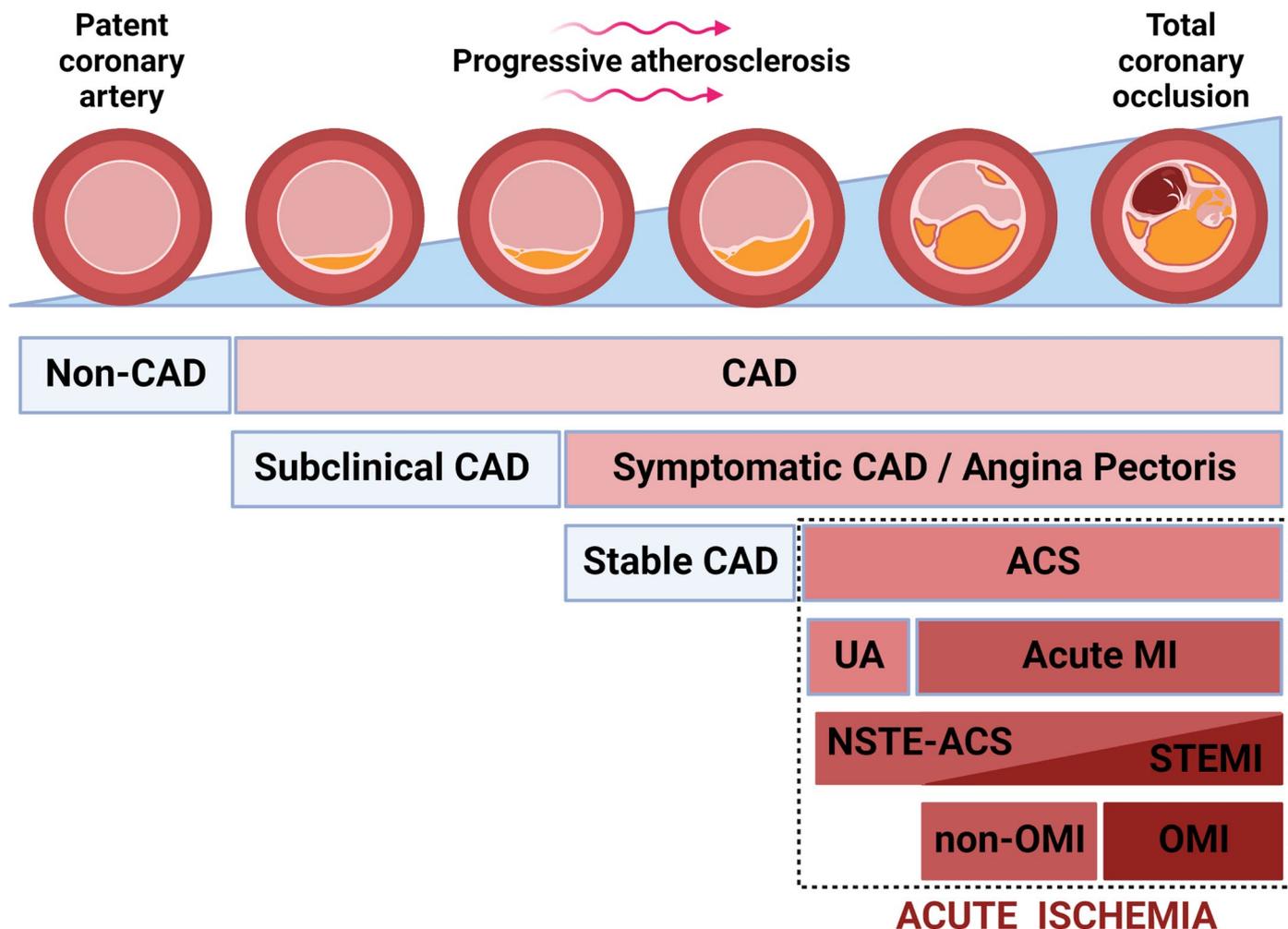
Extended data is available for this paper at <https://doi.org/10.1038/s41591-023-02396-3>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41591-023-02396-3>.

Correspondence and requests for materials should be addressed to Salah S. Al-Zaiti.

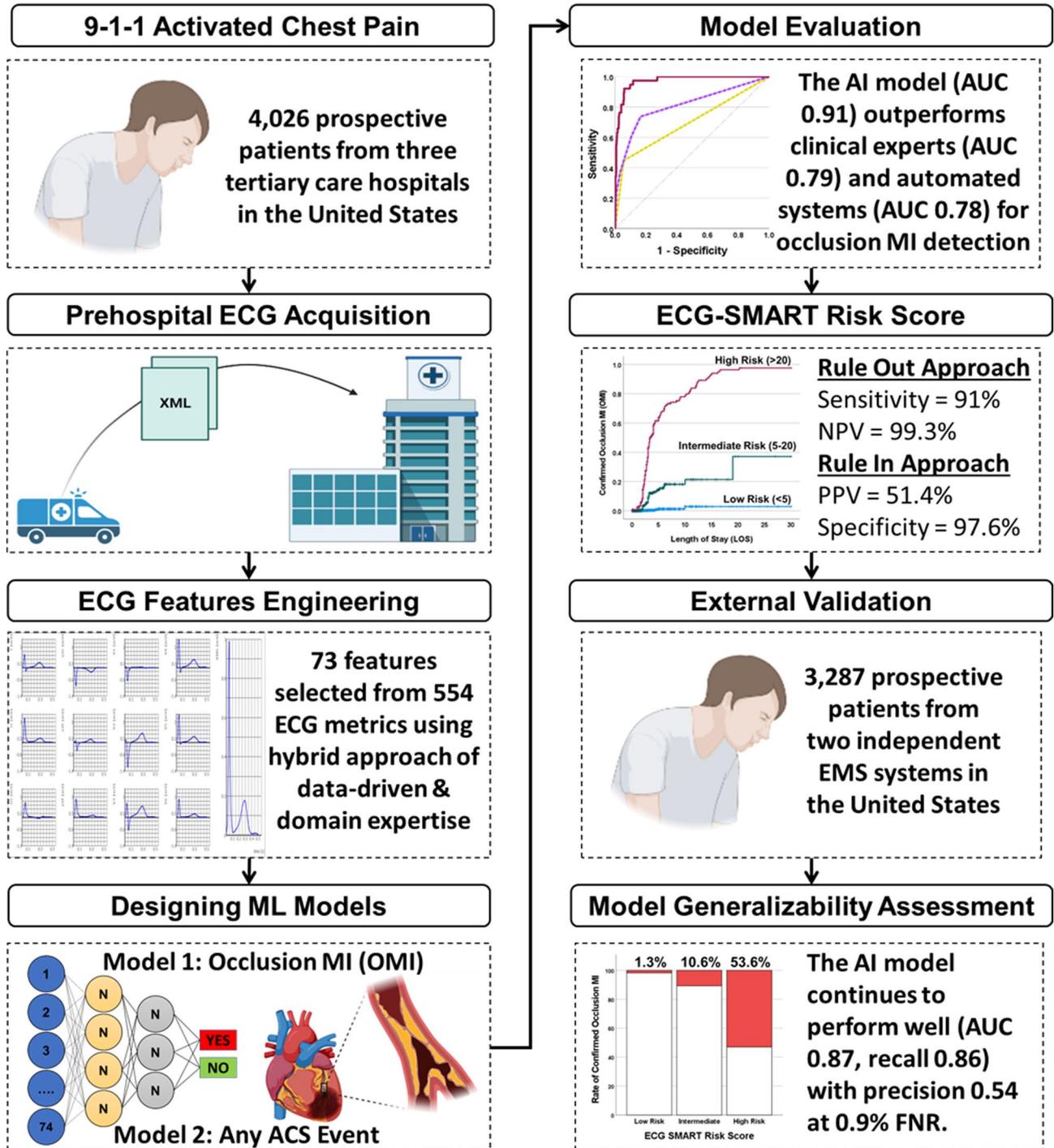
Peer review information *Nature Medicine* thanks Chengyu Liu, Antonio Luiz Ribeiro and Giulio Guagliumi for their contribution to the peer review of this work. Primary handling editor: Lorenzo Righetto, in collaboration with the *Nature Medicine* team.

Reprints and permissions information is available at www.nature.com/reprints.

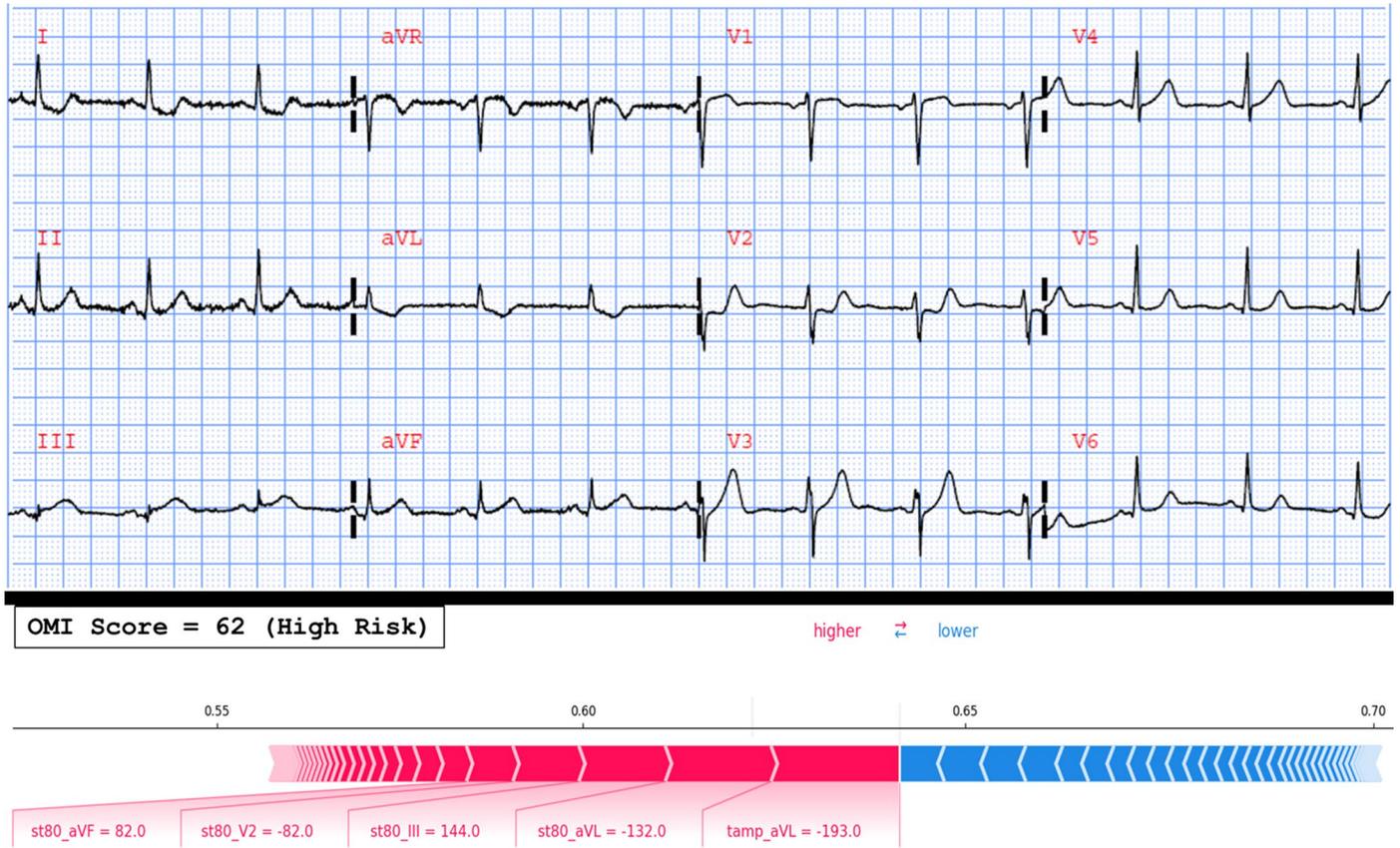


Extended Data Fig. 1 | The relationship between the magnitude of vessel occlusion and the classification of acute coronary events. This figure shows the spectrum of coronary artery disease (CAD) as a function of severity and extent of atherosclerosis plaque progression, ranging from patent coronary artery (far left) to total coronary occlusion (far right). Among patients who develop symptomatic CAD, including those evaluated for chest pain or angina-like symptoms, a subset is diagnosed with acute coronary syndrome (ACS). This group is subclassified as either acute myocardial infarction (MI) or unstable angina (UA). Those with acute MI can be further subclassified, based on the presence of ST-elevation on the ECG, as either ST-elevation myocardial

infarction (STEMI) or without ST-elevation (NSTEMI). The STEMI and NSTEMI patients overlap in terms of the presence or absence of total occlusion (depicted as triangles across the continuum in the figure). Alternatively, the same group with acute MI can be subclassified, based on angiographic TIMI flow criteria, as either occlusion (OMI) or non-occlusion (non-OMI) myocardial infarction. Unlike STEMI, OMI classification better aligns with focal angiographic findings since this group exclusively contains patients with total coronary occlusion. This Figure was created with BioRender.com. Reproduced with permission from Al-Zaiti et. al.1 (permission number 5471421247333, Licensed content publisher: Elsevier).

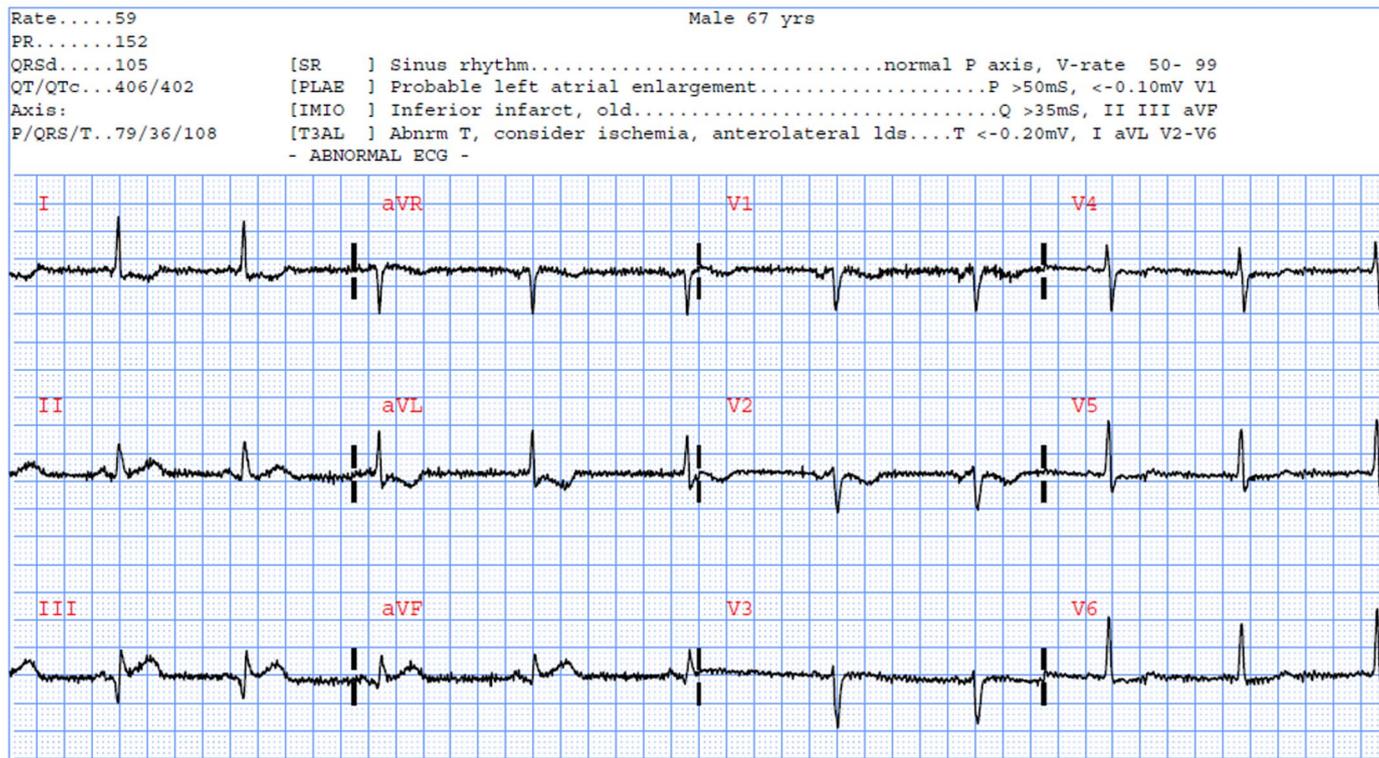


Extended Data Fig. 2 | Graphical abstract summarizing the flow of study and main finding. This figure provides a graphical summary of the study flow and main findings. This Figure was created with BioRender.com (Credit to Salah Al-Zaiti).



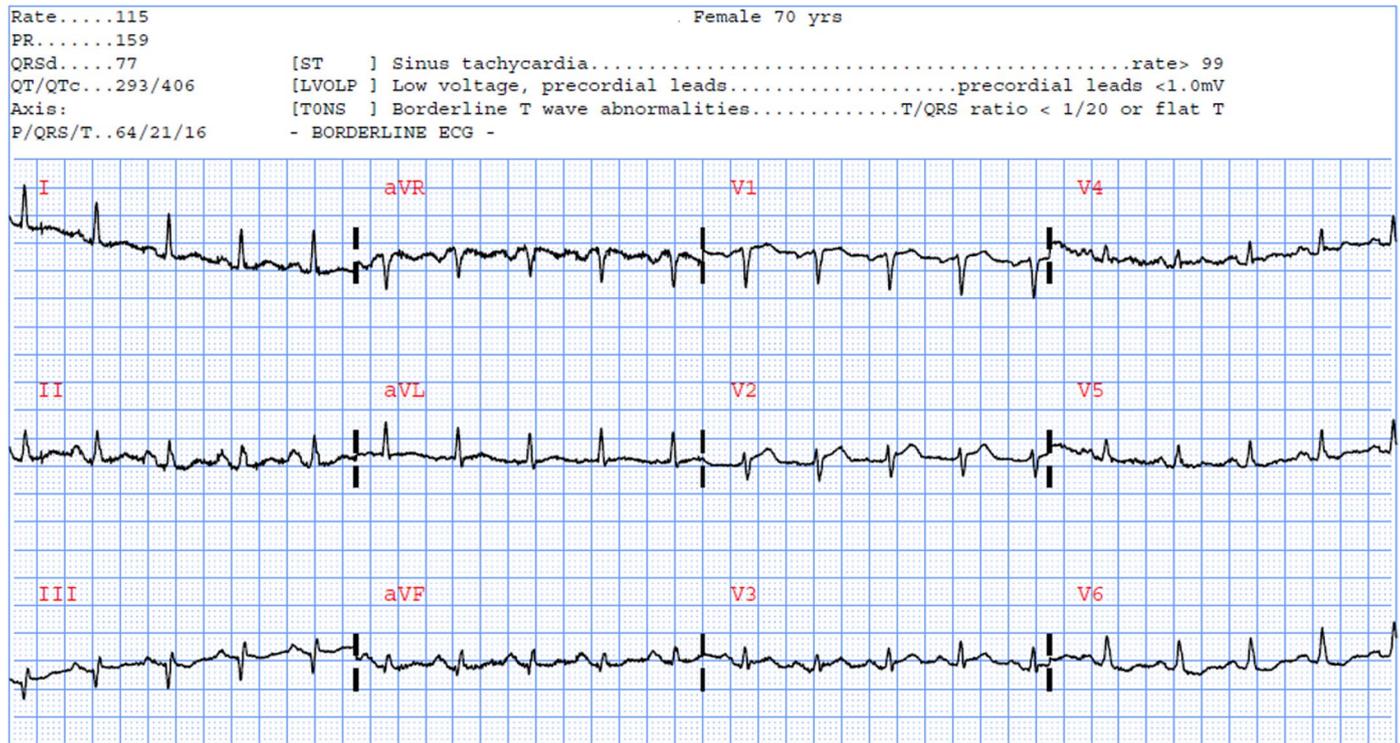
Extended Data Fig. 3 | Local explainability of feature importance on a selected example. This figure shows the baseline ECG of a 50-year-old female with a past medical history of hypertension, high cholesterol, prior myocardial infarction, and current smoking. The ECG was documented as benign with isolated non-specific T wave changes, and the patient was triaged as intermediate risk. The patient was later sent to the catheterization lab where she had complete

occlusion of the right coronary artery. The OMI score on this baseline ECG was 62 indicating high risk designation. The force plot identified the five most important ECG features that met the contribution threshold of the random forest model: negative T wave in aVL, slight ST depression in aVL and V2, and slight ST elevation in aVF and III.



Extended Data Fig. 4 | Selected example of a patient correctly reclassified as OMI. This figure shows an ECG that was correctly reclassified as occlusion myocardial infarction by the machine learning model. This baseline ECG was for a 67-year-old male with a past medical history of high cholesterol and a prior

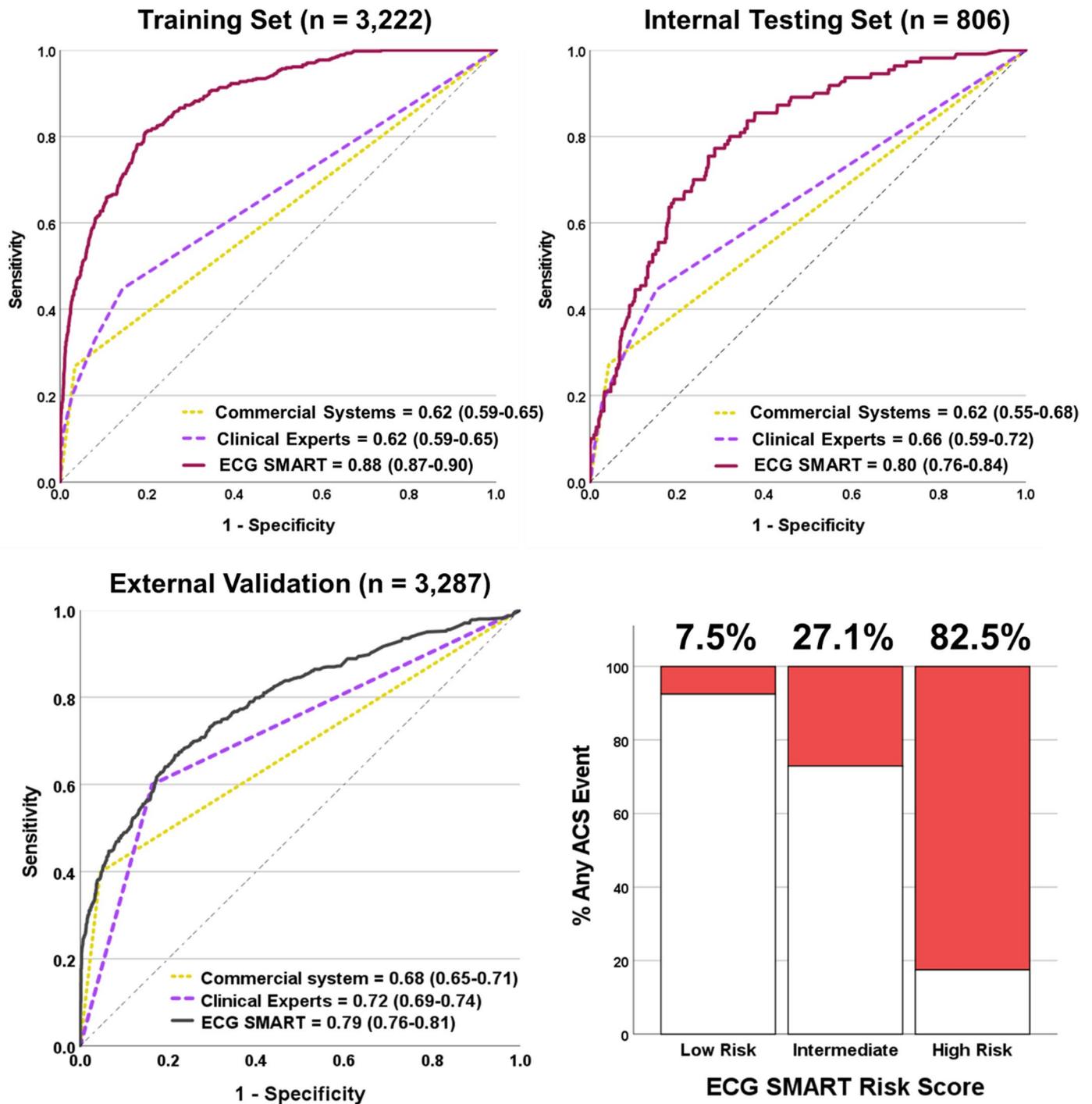
myocardial infarction. The ST-depression in anterior-lateral leads were noted, and the patient was triaged as intermediate risk. The OMI score was 49 indicating the need to up-triage. The patient was later sent to the catheterization lab where he had complete occlusion of the right coronary artery.



Extended Data Fig. 5 | Selected example of a missed OMI by our model.

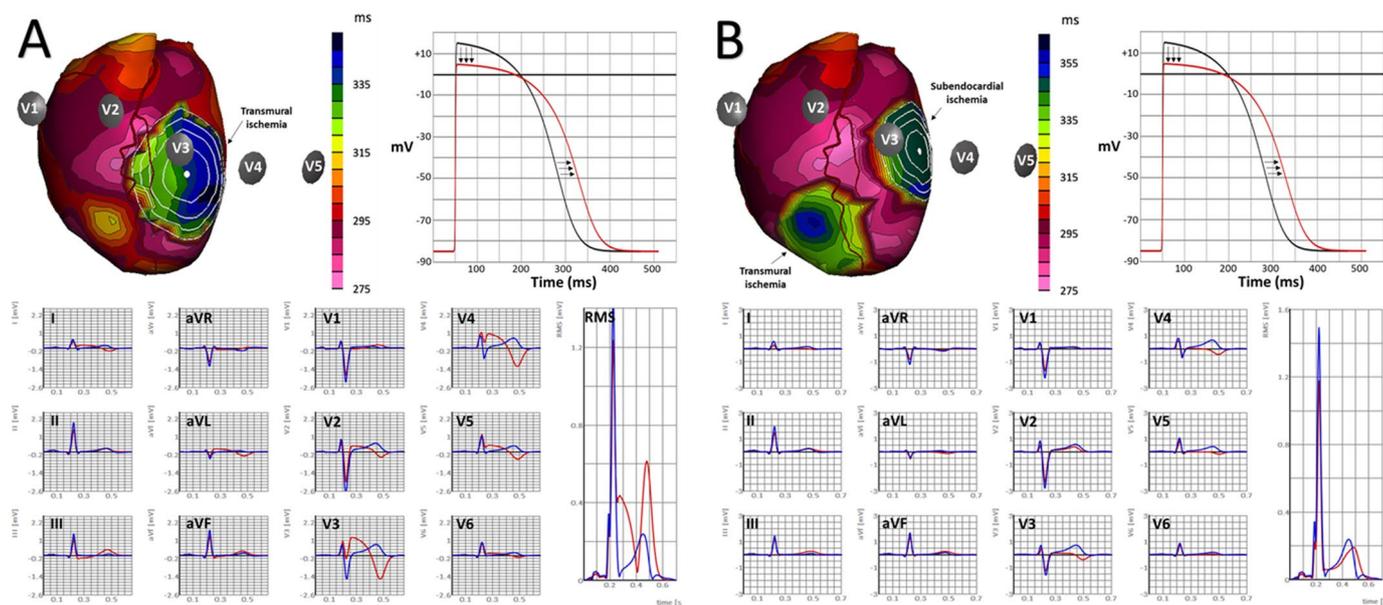
This figure provides a selected example of a patient with occlusion myocardial infarction that was missed by the machine learning model and other reference standards. This ECG was obtained on a 70-year-old female with a past medical history of hypertension, high cholesterol, prior myocardial infarction, and current smoking. The baseline clinical interpretation suggests normal sinus rhythm with benign findings. There are isolated Q waves in inferior leads, low ECG voltage, and some baseline wander and high frequency noise in few leads.

The OMI risk score was 2 indicating a low risk. The patient was later sent to the catheterization lab, which showed severe left main occlusion and had many stents placed. The patient developed new-onset HF during hospitalization. A closer look at this ECG by experienced ECG readers suggests that this ECG could resemble the 'precordial swirl pattern', a rightward ST-elevation vector, with STE in V1 and aVR and reciprocal ST-depression in V5 and V6. This pattern was found to correlate with LAD occlusion.



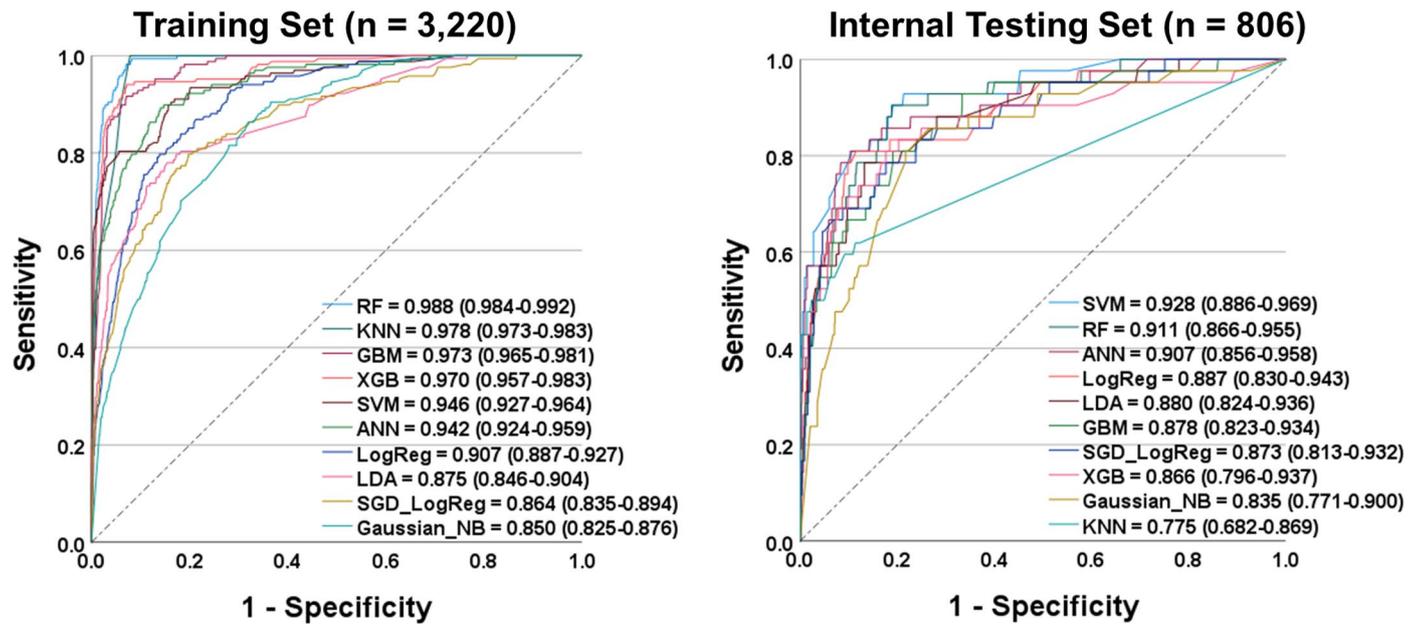
Extended Data Fig. 6 | Development and validation of an algorithm to screen for any ACS event. This figure shows the classification performance of the machine learning model against other reference standards for detecting any acute coronary syndrome event (ACS). The figure also shows the distribution of

patients in low-risk, intermediate risk, and high-risk groups as per our derived risk score. There is a notable gain in precision (rule-in) but a significant loss in recall (rule-out).



Extended Data Fig. 7 | Limitations of ST amplitude on surface ECG as a sole marker of myocardial ischemia. This figure shows: (a) cardiac model of anterior wall epicardial ischemia with corresponding ST-elevation on V3 to V5 of the 12-lead ECG. (b) cardiac model of anterolateral and inferior-apical epicardial

ischemia with corresponding attenuation of ST changes on the 12-lead ECG. This figure was generated using ECGSIM (www.ecgsim.org). Reproduced with permission from Al-Zaiti et. al.1 (permission number 5471421247333, Licensed content publisher: Elsevier).



Extended Data Fig. 8 | Comparison between 10 algorithms trained on the derivation cohort to classify OMI. This figure compares the area under the receiver operator characteristics curves (95% confidence interval) of 10 classifiers during training (left) and testing (right) on the derivation cohort. RF: random forest; KNN: K-nearest neighbors; GBM: gradient boosting

machine; XGB: extreme gradient boosting; SVM: support vector machine; ANN: artificial neural networks; LogReg: regularized logistic regression; LDA: linear discriminant analysis; SGD_LogReg: stochastic gradient descent logistic regression; G_NB: Gaussian Naïve Bayes.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Prehospital 12-lead ECGs were obtained in the field by paramedics as part of routine care. These ECGs were acquired using either Heart Start MRX (Philips Healthcare) or LIFEPAK-15 (Physio-Control Inc.) monitor-defibrillator devices. Demographic and clinical data elements were abstracted from hospital systems on all patients meeting eligibility criteria. Adjudications of outcomes were made by independent reviewers at each local site after reviewing all available medical records within 30 days of the indexed encounter.

Data analysis Each ECG was stored as a digital XML file, and all files were transmitted to the Philips Advanced Algorithm Research Center (Cambridge, Massachusetts, USA). These source binary files were preprocessed using the commercially available Philips DXL diagnostic 12/16 lead ECG analysis program. The extracted ECG features were then used for model derivation and testing. These ECG features are made available in CSV format along with the code used to derive the machine learning classifier. All diagnostic accuracy values were reported as per STARD recommendations (Reporting Guidelines for Diagnostic Accuracy Studies). We reported classification performance using AUROC curve, sensitivity (recall), specificity, PPV (precision), and NPV, along with 95% confidence interval (CI) where applicable. For 10-fold cross validation, we compared the multiple classifiers using the Wilcoxon signed-rank test (for AUROC curves) and McNemar's test (for confusion matrices). We derived low-, intermediate-, and high-risk categories for the final classifier using Kernel density plot estimates between classes. The adequacy of these risk classes was evaluated using Log-rank chi-square of accumulative risk for clinically important outcomes over the length of stay during the indexed admission. All analyses were completed using Python v3.8.5 and SPSS v24.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The ECG SMART trial makes use of extracted ECG features to train and evaluate a random forest classifier to denote the probability of OMI. The ECG features used in the derivation and external validation datasets along with linked clinical outcomes are publicly available through GitHub (<https://github.com/zeineb-bouzid/sharing-github-nature-medicine.git>). Researchers wishing the source binary files to compute their own features should contact the corresponding author to arrange for proper approvals and institutional data use agreements. Interested researchers from non-commercial entities can submit a requested by emailing the corresponding author at ssa33@pitt.edu. Requests will be processed within 2-week time frame.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Our manuscript is compliant with the Journal's policy on sex and gender reporting. As per the National Institute of Health, sex was treated as a biological variable in this study. We collected biological sex on all participants as documented in the electronic health records. We enrolled all consecutive patients with suspected acute coronary syndrome, so all eligible patients from both sexes were recruited. Given that the rate of outcomes is typically higher in males than females and to avoid algorithmic bias, we decided a priori not to include sex as an input feature to estimate probability of OMI in our models. Instead, we show post-hoc evaluation of the diagnostic accuracy of the OMI score in each sex separately (Fig. 4C). There were no significant differences in sensitivity and specificity of the OMI score between males and females.

Population characteristics

All study cohorts enrolled consecutive adult patients > 18 years of age with an emergency call for non-traumatic chest pain or anginal equivalent symptoms (arm, shoulder, jaw pain, shortness of breath, diaphoresis, syncope). Eligible patients were transported by an ambulance and had at least one recorded prehospital 12-lead ECG. There were no selective exclusion criteria based on sex, race, comorbidities, or acuity of illness. After excluding patients with cardiac arrest, ventricular tachyarrhythmias, confirmed prehospital STEMI, and duplicate ECGs, our derivation cohort included 4,026 consecutive patients with chest pain (age 59±16 years, 47% females, 5.2% OMI). The two external validation cohorts together included 3,287 patients (age 60±15 years, 45% females, 6.4% OMI). The derivation and validation cohorts were comparable in terms of age, sex, baseline clinical characteristics, and 30-day cardiovascular mortality.

Recruitment

This was an observational cohort study that enrolled all consecutive eligible patients. There were no selective exclusion criteria based on sex, race, comorbidities, or acuity of illness. All eligible, consecutive patients were recruited under a waiver of informed consent to minimize any potential sources of selection bias. We collected prehospital ECGs obtained during routine care. Independent reviewers extracted data elements from hospital systems on all patients meeting eligibility criteria. Study cohorts altogether enrolled patients that were transported by a participating EMS agency between May 1, 2013, and June 29, 2022.

Ethics oversight

The derivation cohort included prehospital data from the City of Pittsburgh Bureau of Emergency Medical Services (EMS) and in-hospital data from the University of Pittsburgh Medical Center (UPMC). This cohort study was approved by the institutional review board of the University of Pittsburgh. The external validation cohorts included data from Orange County EMS and Mecklenburg County EMS and Atrium Health. These two cohort studies were approved by the institutional review board of the University of North Carolina at Chapel Hill.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No minimum sample size was specified. All eligible consecutive patients over the duration of the study were enrolled. All the available data for this analysis were used.

Data exclusions	We removed patients with confirmed prehospital ST-elevation myocardial infarction (STEMI) because these patients undergo emergent catheterization as per clinical guidelines without the need for additional diagnostic testing.
Replication	We used rigorous 10-fold cross-validation to estimate models' hyper-parameters on the derivation dataset and assessed bias-variance trade-off between training and internal testing splits. We also used an external validation cohort from two independent sites and demonstrated model generalizability on unseen patients. The results were reproducible across folds and datasets. We have made our Python code and the derivation and external validation datasets available to ensure reproducibility.
Randomization	Randomization was not relevant to this analysis. This was an observational cohort study and there was no active assignment of participants to groups.
Blinding	Blinding was not relevant to this analysis. This was an observational cohort study and there was no active assignment of participants to groups that required blinding..

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern		

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	This study was registered in www.ClinicalTrials.gov (identifier # NCT04237688).
Study protocol	The analyses described in this paper were pre-specified by the trial protocol that was funded by the National Institute of Health.
Data collection	This was an observational cohort study that enrolled all consecutive eligible patients. We collected prehospital ECGs obtained during routine care. Independent reviewers extracted data elements from hospital systems on all patients meeting eligibility criteria. The derivation cohort included prehospital data from the City of Pittsburgh Bureau of Emergency Medical Services (EMS) and in-hospital data from the University of Pittsburgh Medical Center (UPMC). The external validation cohorts included data from Orange County EMS and Mecklenburg County EMS and Atrium Health. Study cohorts altogether enrolled patients that were transported by a participating EMS agency between May 1, 2013, and June 29, 2022. A brief synopsis of the study protocol can be found on the public domain of the National Institute of Health through www.NIHreporter.gov
Outcomes	<p>The primary study outcomes were ACS and OMI, which were predefined on study protocol and ClinicalTrials.gov prior to start of the study. ACS was defined per the fourth universal definition of myocardial infarction as the presence of symptoms of ischemia (i.e. diffuse discomfort in the chest, upper extremity, jaw, or epigastric area for more than 20 minutes) and at least one of the following criteria: (1) subsequent development of labile, ischemic ECG changes (e.g., ST changes, T inversion) during hospitalization; (2) elevation of cardiac troponin (i.e., > 99th percentile) during the hospital stay with rise and/or drop on serial testing; (3) coronary angiography demonstrating greater than 70% stenosis, with or without treatment; and/or (4) functional cardiac evaluation (stress testing) that demonstrates ECG, echocardiographic, or radionuclide evidence of focal cardiac ischemia.</p> <p>OMI was defined as coronary angiographic evidence of an acute culprit lesion in at least one of the three main coronary arteries (LAD, LCX, RCA) or their primary branches with TIMI flow grade of 0-1. TIMI flow grade of 2 with significant coronary narrowing > 70% and peak troponin of 5-10.0 ng/mL was also considered indicative of OMI.</p> <p>Adjudications were made by independent reviewers at each local site after reviewing all available medical records within 30 days of the indexed encounter. Reviewers were blinded from all ECG analyses and models' predictions. Each outcome was adjudicated by two independent reviewers. The Kappa coefficient statistic between the two reviewers for OMI was 0.771 (i.e., substantial agreement) and for ACS was 0.846 to 0.916 (i.e., substantial to perfect agreement). All disagreements were resolved by a third reviewer.</p>