ESC European Society of Cardiology

European Heart Journal - Digital Health (2025) **6**, 989–996
https://doi.org/10.1093/ehjdh/ztaf083

ORIGINAL ARTICLE

Artificial intelligence-driven electrocardiogram analysis for risk stratification in pulmonary embolism

Tanmay A. Gokhale [1,2,*], Nathan T. Riek [2], Brent Medoff [1], Rui Qi Ji[3], Belinda Rivera-Lebron[1,2], Ervin Sejdic [3,4], Murat Akcakaya [2], Samir F. Saba [1,2], Salah Al-Zaiti [2,5], and Catalin Toma[1,2]

[1]Heart and Vascular Institute, University of Pittsburgh Medical Center, 3550 Terrace St, B571.3 Scaife Hall, Pittsburgh, PA 15261, USA; [2]Departments of Medicine, Electrical and Computer Engineering, and Nursing, University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, PA 15260, USA; [3]Departmenf of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada M5S 3G4; [4]Artificial Intelligence for Health Outcomes, North York General Hospital, Toronto, ON, Canada M2K 1E1; and [5]School of Nursing, University of Rochester, Rochester, NY 14642, USA

Received 1 April 2025; revised 10 June 2025; accepted 27 June 2025; online publish-ahead-of-print 23 July 2025

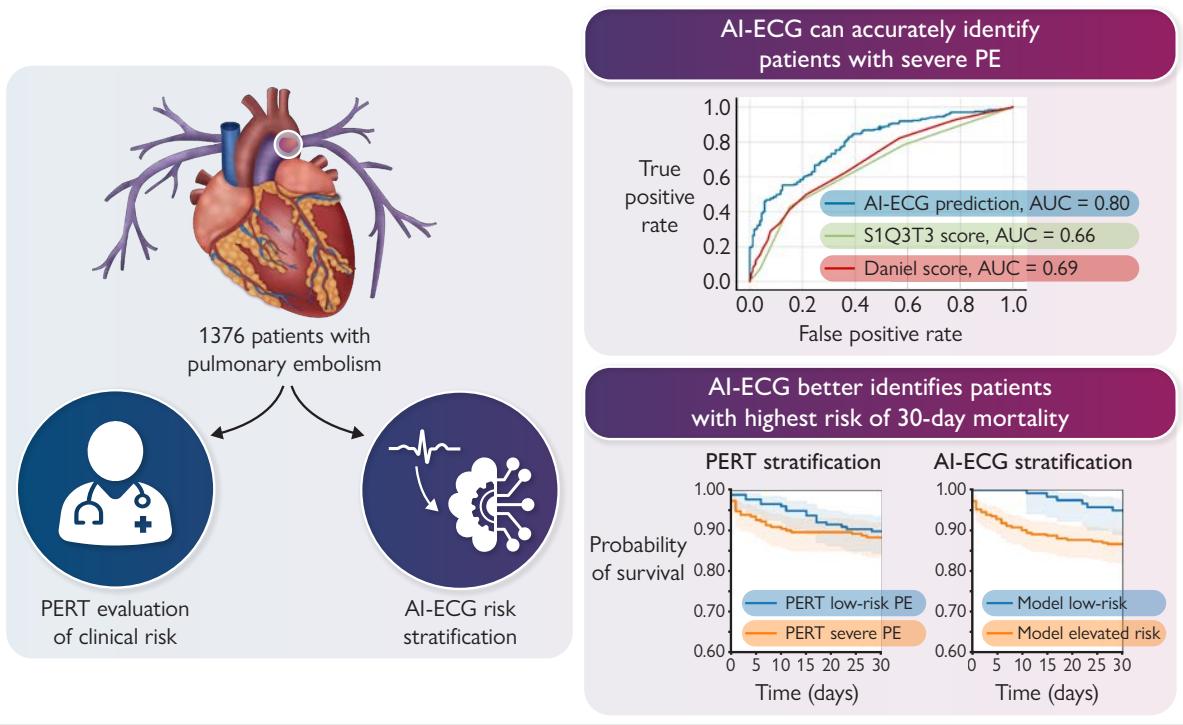
Aims Among patients with acute pulmonary embolism (PE), rapid identification of those with highest clinical risk can help guide life-saving treatment. However, current risk stratification algorithms involve a multistep process requiring physical exam, imaging, and laboratory results. We investigated the utility of electrocardiogram (ECG) alone to rapidly identify patients at elevated clinical risk by developing and validating a feature-based artificial intelligence (AI) model to predict clinical risk.

Methods and results Patients who were diagnosed with PE over a 9-year period, had an ECG within 1 day of presentation, and were evaluated by our PE response team (PERT) were included. A feature-based random forest model was trained to predict the PERT team's risk stratification from the ECG alone. The ability of the model to predict the clinical risk categorization and the accuracy of both risk stratification approaches in predicting mortality were examined on a holdout test set. Of the overall cohort of 1376 patients, 55% had a submassive (intermediate risk) or massive (high risk) PE, which were grouped together as 'severe PE'. The AI-ECG model was able to predict the clinical classification (low-risk vs. severe PE) with an AUC of 0.83 and F1 score of 0.78 in a holdout test set. A 30-day mortality and in-hospital mortality were significantly different between patients classified by the model as low vs. elevated risk.

Conclusion AI-based analysis of 12-lead ECGs may provide a useful tool in the risk stratification of PE, allowing for rapid identification and treatment of those at highest risk of adverse outcomes.

* Corresponding author. Tel: +412 647 3429, Email: gokhaleta@upmc.edu, @tgokhale

Graphical Abstract



Keywords AI • ECG • Pulmonary embolism • Risk stratification

Introduction

Acute pulmonary embolism (PE) carries a substantial risk of morbidity and mortality, and early intervention in those at highest risk of clinical deterioration can improve outcomes. However, many patients with haemodynamically significant PE present with seemingly low acuity symptoms, and identifying those at highest risk remains a challenge. Current guidelines for assessment of PE risk call for assessment by a pulmonary embolism response team (PERT) using a combination of physical exam, laboratory, and imaging tools to best assess clinical risk.[1,2]

As patients with PE frequently present with chest pain or dyspnoea, electrocardiograms (ECGs) are routinely obtained in these patients very early in their clinical course. Electrocardiogram (ECG)-based approach for identifying PE has been previously described,[3,4] including the classical S1Q3T3 (the presence of an S wave in Lead I, and a Q wave and inverted T-wave in Lead III) as well as anterolateral T-wave inversions. However, these semiquantitative features individually have limited clinical utility due to modest to low sensitivity and specificity.

The aim of our study was to build and evaluate a machine learning (ML) model that can predict clinical risk stratification based on the ECG alone. We hypothesized that the RV strain associated with severe PE would lead to a discrete ECG signature that can be characterized with the use of artificial intelligence (AI). To that end, we used data from a cohort of patients seen by the PERT at our institution over a period of 9 years to train and validate a feature-based predictive model and compared the accuracy of our AI-based ECG model with clinical risk stratification in predicting mortality.

Methods

Study population and data sources

This was a retrospective study of patients diagnosed with PE who were seen by the PERT at UPMC Presbyterian Hospital from 1 August 2014 to 15 August 2023. UPMC Presbyterian is a tertiary/quaternary care medical centre in Pittsburgh, PA, USA, that provides comprehensive medical and interventional management of PE including catheter-directed thrombolysis, catheter-based aspiration embolectomy, surgical embolectomy, and extracorporeal membrane oxygenation (ECMO); UPMC is a PERT Consortium Certified PE Center of Excellence. Patients evaluated by the PERT were identified through an internal PERT team database. Included patients were evaluated by the on-call PERT consultants—a pulmonary/critical care medicine fellow and attending physician—via in-person or electronic consultation.

At the time of PERT evaluation, patients were classified as low-risk, intermediate-risk (submassive), or high-risk (massive) based on clinical practice guidelines. Further stratification of intermediate-risk patients into low-intermediate and high-intermediate risk categories was not routinely documented in the PERT database. For the analysis in this study, patients with submassive or massive PE were collectively classified as 'severe PE', in contrast to those felt to have low-risk PE, as those with elevated risk may be the ones benefiting from procedural intervention for PE.

Patients who were diagnosed with PE more than 1 day after hospital admission were excluded. In addition, patients who were transferred to our facility from an outside hospital were excluded unless the primary diagnosis for transfer was acute PE, other venous embolism, or obstetric thromboembolism. All ECGs obtained within 1 day of PE diagnosis were identified with the GE MUSE ECG management system, and if multiple ECGs had been performed, the earliest ECG on the day of diagnosis was utilized.

Patients without an ECG within 1 day of PE diagnosis were excluded. All ECGs were acquired as digital 12-lead, 10 s tracings using GE MAC55 or VU360 machines. Electrocardiogram waveforms for the eight linearly independent leads (I, II, V1–V6) were exported from the GE MUSE system in XML format for analysis.

Patient demographics, clinical comorbidities, and clinical outcomes were obtained from the UPMC Clinical Analytics group's clinical data warehouse, which includes inpatient and outpatient records from 20 UPMC hospitals and associated clinics, including all UPMC sites in southwestern Pennsylvania. Clinical outcomes were censored in March 2024 to ensure a minimum of 6 months of follow-up for all included patients. Mortality data in the clinical data warehouse incorporate both electronic health record documented mortality and data from the social security death index.

Electrocardiogram processing and feature selection

ECGs were filtered at the time of clinical acquisition with a high-pass filter cut-off of 0.05, 0.16, or rarely 0.56 Hz, and a low-pass filter cut-off of 100 or 150 Hz. At the time of analysis, all ECGs were filtered with a low-pass filter at 100 Hz, a high-pass filter at 0.5 Hz, and a 60 Hz powerline notch filter. Noise and ectopic beats were identified and removed. A median beat was then computed for each ECG. Each non-ectopic beat and the median beat were segmented into traditional waveform segments. As previously described,[5] a total of 419 spatio-temporal features were then computed from each ECG. These features include global and lead-specific ECG segment durations, intervals, amplitudes, and areas; ratios of waveform eigenvalues; and vectorcardiographic axes, angles, loops, and gradients.

Included patients were split into training and test sets in a 70/30 split. Further feature selection was performed only on the training set to prevent train–test leakage. Highly collinear feature-pairs, with a Spearman rank correlation coefficient of >0.9, were then identified, and one feature from each pair was removed from the feature set, resulting in 312 remaining features. A Benjamini–Hochberg procedure[6] was then performed with a false discovery rate of 5% and identified 140 features that met criteria for significance. The resulting feature set was considered the 'full' feature set. Forward sequential feature selection was then performed using a random forest estimator to identify the 30 or 60 best features using five-fold cross-validation. Both the 30-feature and 60-feature subsets outperformed the full 140-feature set with respect to AUC. In addition, the 60-feature set had slightly higher AUC, higher average precision, and the highest cross-validation F1 score both before and after model tuning, so the 60-feature set was used for further model development (see Supplementary material online, *Table S1*). All variables were *z*-score normalized.

Model development

Five different ML model types were evaluated on the training set using stratified five-fold cross-validation. The average area under the receiver operator curve (AUC) and the average precision during cross-validation were comparable amongst the random forest classifier, light gradient boosting machine, and support vector machine with radial basis function kernel, and lower with a multi-layer perceptron classifier and logistic regression.

Given comparable performance and our group's prior work with random forests,[7] a random forest classifier was selected for further optimization. Model tuning was performed via random grid search using stratified five-fold cross-validation to optimize AUC. Prediction thresholds were determined by training the model with four folds and selecting the threshold that optimizes F1 score on the holdout fold. The five resulting thresholds were then averaged to determine the threshold to be used for testing. The selected decision threshold was 0.39

Model evaluation

The resulting model was evaluated on the holdout test set. The model prediction was deemed elevated risk when the model output was greater than the threshold determined on the training set. The most informative features in determining the predicted outcome were computed using Shapley additive explanations (SHAP).[8] In addition, the average ECGs for all patients with predicted low risk and with predicted elevated risk were determined by excluding all patients with atrial fibrillation, aligning all ECG median beats to the peak of the QRS complex, and computing the average waveform within each group.

To compare AI-ECG model performance with existing ECG-based tools for assessing PE, the S1Q3T3 score and Daniel score[9] were calculated using the features already computed for each ECG. The S1Q3T3 score was computed by assigning one point for each present feature; the Daniel score was computed as previously described.[9,10] The receiver operating and precision-recall curves were then plotted for the random forest classifier and the S1Q3T3 and Daniel scores.

Statistical methods

Continuous data are presented as mean ± standard deviation and were compared between groups with a Student's *t*-test. Categorical data are reported as frequency with percentage and were compared using the χ^2 test. The AUCs of different models were compared using DeLong's test. Survival curves are presented as Kaplan–Meier survival curves and compared using the log rank test. Unadjusted hazard ratio of an elevated risk prediction was calculated using the univariable Cox proportional hazards model. All analyses were performed in Python using the sci-kit learn, SHAP, PyCaret, and Lifelines packages.

This research study was reviewed by the University of Pittsburgh Institutional Review Board (IRB) and deemed exempt. The IRB waived the need for informed consent, given the observational nature of the study. This study is reported in accordance with the TRIPOD + AI reporting guidelines.

Results

Patient characteristics

A total of 2914 patients were evaluated by the PERT during the 9-year period examined in this study. *Figure 1* shows the flowchart of patient inclusion. Following exclusion criteria, a total of 1376 patients were included in the patient cohort used for model development and validation. Of these, 55% had a severe PE, and 45% had a low-risk PE.

Table 1 shows the baseline characteristics of the study population. Patients with low-risk PE were younger than those with severe PE (57 ± 18 years vs. 62 ± 16 years, $P < 0.01$) and more likely to be male (46.3% female in low-risk PE vs. 52.3% female in severe PE). Patients in both groups had comparable comorbidity profiles, as assessed by the Elixhauser-based van Walraven scores (2.4 ± 5.2 in low-risk PE vs. 2.3 ± 4.7 in severe PE).

Model performance

The final AI-ECG model had a cross-validation AUC of 0.80 and an average precision of 0.82. When tested on the holdout test set, the AI-ECG model demonstrated stable AUC of 0.80 and average precision of 0.85 (*Figure 2*, blue line). At the selected operating point (*Figure 2*, blue diamond), the AI-ECG demonstrated accuracy of 0.72, precision (positive predictive value) of 0.70, negative predictive value of 0.76, and recall (sensitivity) of 0.88, with an F1 score of 0.78. Of note, the specificity of the model was modest at 0.51 indicating a high false-positive rate. An AI-ECG prediction of elevated risk carried an odds ratio of 7.6 (4.6–12.4) of being clinically classified as a severe PE.

The top 20 features driving AI-ECG prediction on the test set are shown in *Figure 3*. These include several traditionally regarded markers of PE and PE severity as well as several novel markers. To better understand the ECG phenotype that the AI-ECG model identified as elevated risk, the average ECG waveform for patients predicted to be of elevated risk and low risk is shown in *Figure 4*. Qualitatively, ECGs from patients with predicted elevated risk showed an S wave in Lead I; a Qs pattern in Lead III; T-wave inversions in Leads III, avF, and V1–V3; and T-wave flattening in Leads II and V4–V6. In addition, predicted elevated risk ECGs had a delayed QRS transition across the precordium.

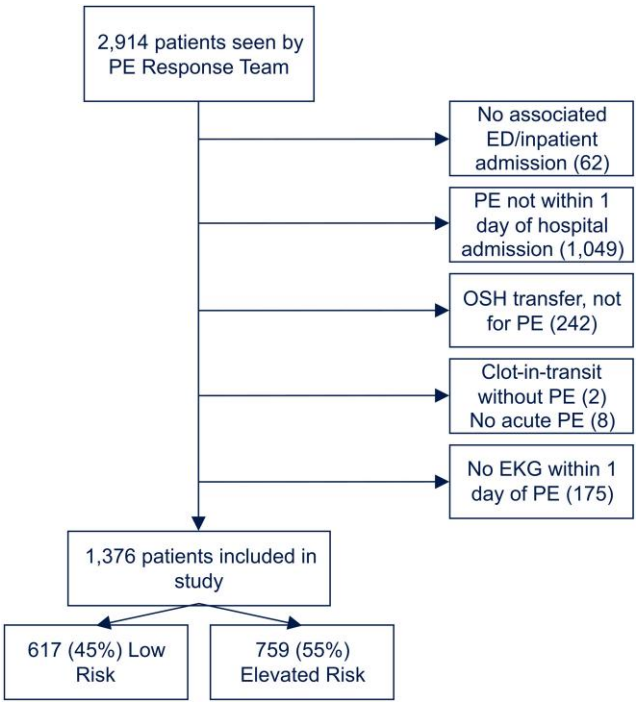


Figure 1 Patient inclusion flowchart showing inclusion/exclusion criteria for subjects in the study. PE, pulmonary embolism; OSH, outside hospital; ECG, electrocardiogram.

Table 1 Baseline characteristics of included patients

	Low-risk PE ($n = 617$)	Severe PE ($n = 759$)	SMD
Demographics			
Age, years [mean (SD)]	57 (18)	62 (16)	0.27
Women	46.30%	52.30%	0.12
Race: Black	19.94%	15.94%	0.10
Medical history			
van Walraven score	2.4 (5.2)	2.3 (4.7)	0.03
Coronary artery disease	10.19%	9.04%	0.04
Diabetes mellitus	14.76%	14.91%	0.01
Hypertension	38.46%	41.21%	0.06
Heart failure	5.82%	4.52%	0.06
Atrial fibrillation	3.53%	5.53%	0.10
Asthma	22.87%	21.44%	0.03
Obstructive lung disease	6.65%	8.04%	0.05
Chronic kidney disease	4.57%	5.19%	0.03
Obesity	14.97%	18.59%	0.10
Malignancy	11.40%	12.40%	0.03
Current/recent tobacco use	15.70%	10.80%	0.14
Clinical course			
ICU stay*	31.76%	77.70%	1.04
Primary PE diagnosis	46.07%	77.68%	0.61

Variables are summarized as percentage or as mean (standard deviation) as appropriate.
Note that ICU stay includes patients in the ICU at the time of their PE diagnosis.
ICU, intensive care unit; SMD, standardized mean difference.

AI-ECG performance on the test set was compared with the S1Q3T3 and Daniel scores. The tuned random forest classifier's AUC of 0.80 outperformed the S1Q3T3 (AUC = 0.66, $P < 0.01$ by DeLong's test) and the Daniel score (AUC 0.69, $P < 0.01$) (*Figure 2*).

Of the 411 patients in the test set for whom clinical outcomes were available (out of 413), 36 (8.8%) died during the index hospitalization and 45 (10.9%) died within 30 days. A 30-day mortality was comparable between patients classified by the PERT team as severe PE and low-risk PE (11.6% vs. 10.1%, $P = 0.34$) (*Figure 5*, top panel). In contrast, 5.1% of patients identified by the AI-ECG model as predicted low risk died within 30 days, compared with 13.3% of those identified as elevated risk ($P < 0.01$; *Figure 5*, bottom panel). AI-ECG elevated risk prediction carried an unadjusted hazard ratio for all-cause mortality of 2.21 [95% confidence interval (CI): 1.33–3.67]. The in-hospital mortality rate was 2.5% in those the model predicted as low risk compared with 11.3% in those the model deemed elevated risk. An elevated risk prediction carries an odds ratio for in-hospital mortality of 4.87 (95% CI: 1.46–16.19), higher than that reported for any single ECG finding in a recent review.[4]

Discussion

In this study, we present a novel AI-ECG model developed to classify the haemodynamic risk of PE, as assessed by a pulmonary disease specialist equipped with comprehensive clinical history, imaging, and laboratory data. We demonstrate that the ECG-based model can accurately identify patients with severe PE (AUC 0.84, 71.9% accuracy). It outperforms both the S1Q3T3 criteria (AUC 0.66) and the Daniel score (AUC 0.69).

The accurate and timely risk stratification of patients with acute PE is crucial for guiding clinical decision-making and optimizing patient outcomes. PE can present with a wide spectrum of clinical severity, from asymptomatic to mild dyspnoea to haemodynamic collapse; however,

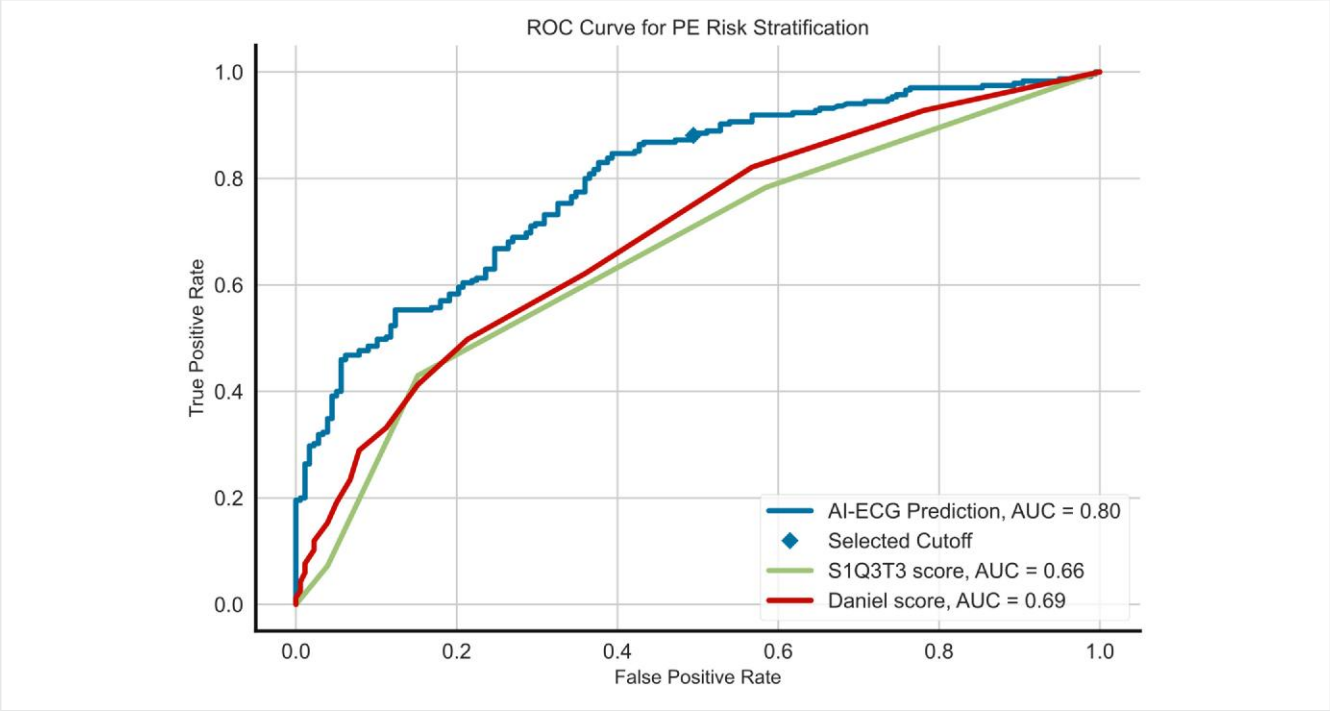


Figure 2 AI-ECG model outperforms existing ECG-based PE scores in identifying patients with severe PE. At the selected operating point (blue diamond), AI-ECG had an accuracy of 72% and F1 score of 78%.

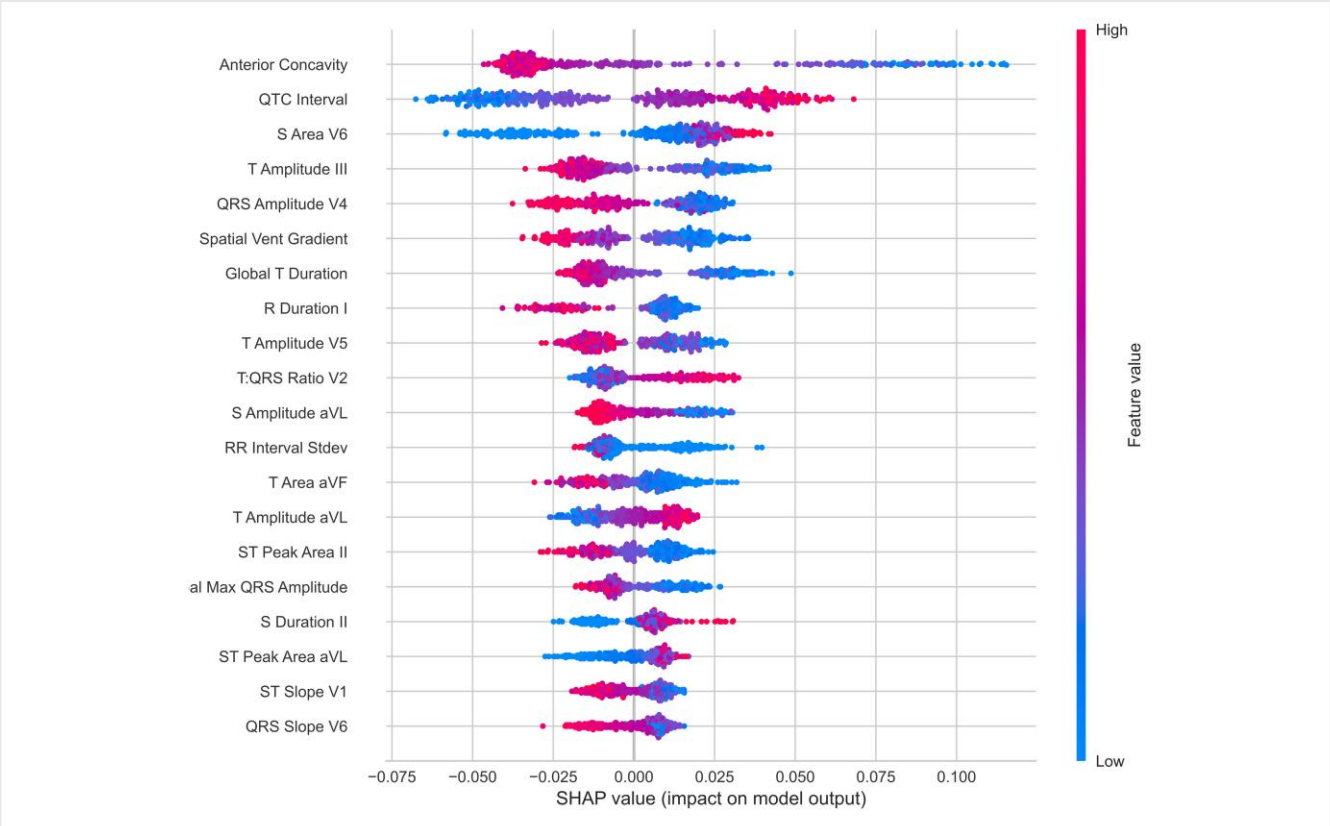


Figure 3 SHAP analysis reveals the ECG features that drove model predictions. The top features include both previously known and novel predictors of severe PE. SHAP, Shapley additive explanations.

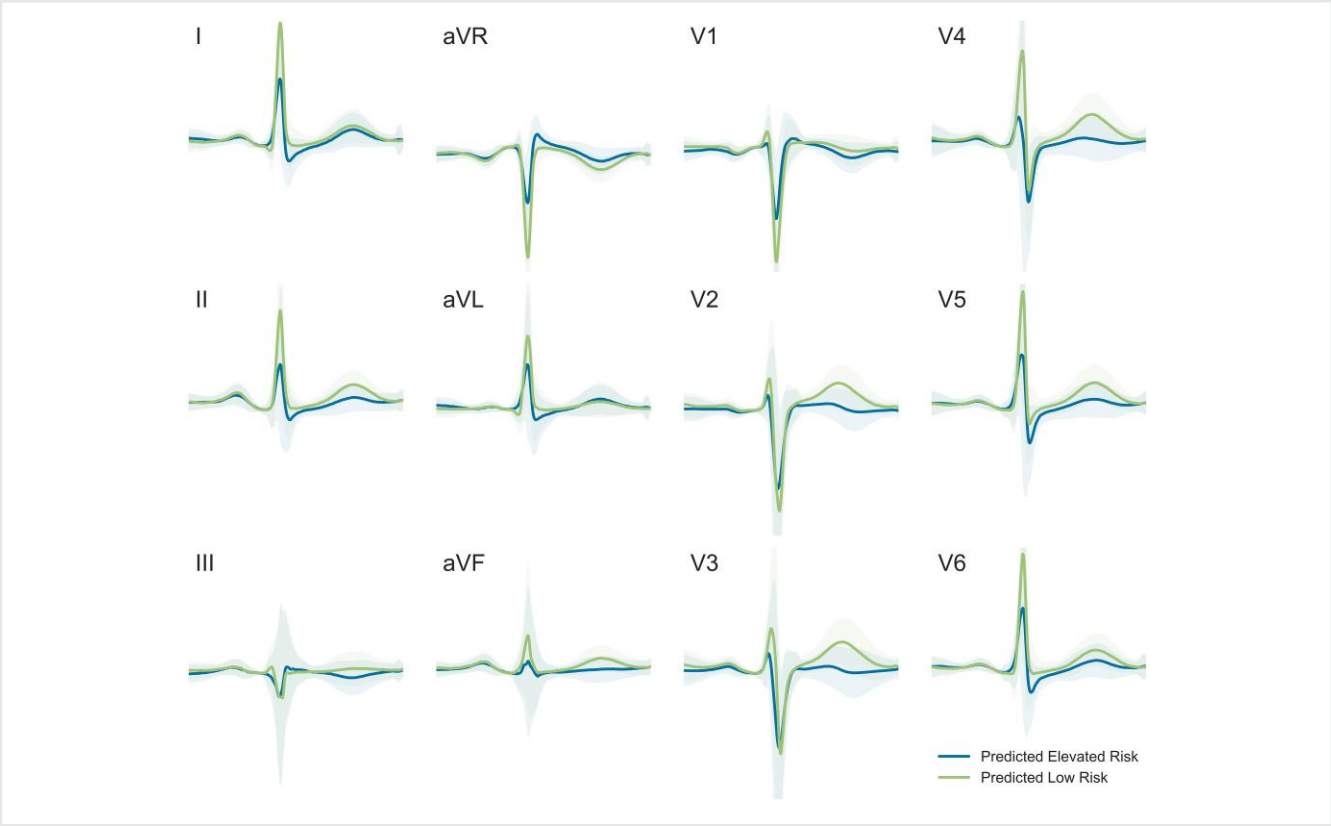


Figure 4 Comparison of average ECG median beat waveforms in patients with AI-ECG predicted elevated risk (blue) vs. AI-ECG predicted low risk (green). ECGs with atrial fibrillation were excluded from averaging.

even patients who present with low acuity symptoms can quickly decompensate if not treated quickly. As such, identifying patients at the highest risk of adverse events allows for targeted interventions such as anticoagulation, thrombolysis, catheter-directed therapies, and embolectomy. Conversely, accurately identifying lower-risk individuals may facilitate shorter hospital stays or outpatient management, reducing healthcare costs and unnecessary utilization.

Several clinical factors are known to be associated with high risk of adverse outcomes. These include physical exam findings like hypotension, tachycardia, hypoxia, and signs of right ventricular failure[11]; laboratory tests including elevated brain natriuretic peptide (BNP) and elevated troponin levels[12,13]; and imaging findings of right ventricular dysfunction by computed tomography (CT) or echocardiography. The pulmonary embolism severity index (PESI) and the later simplified PESI (sPESI) scores[14,15] aggregate several of these factors to stratify patients by risk of 30-day mortality. However, these scores fail to identify patients who have seemingly normal haemodynamics but progressive right ventricular dysfunction. Current clinical practice guidelines recommend a multistep risk stratification algorithm that first identifies those with clinical instability as 'massive' or 'high-risk' and then uses a combination of PESI/sPESI score, biomarkers of myocardial injury or heart failure, and RV dysfunction on cardiac imaging to classify patients as 'low risk' or 'intermediate risk' (previously termed 'submassive' and now subdivided into 'intermediate-low risk' and 'intermediate-high risk').[1,2]

The imaging and laboratory testing required to make this distinction may not always be easily or quickly available, which can delay appropriate diagnosis and care. In addition, the expertise of a PERT is not available in many settings to perform rigorous risk stratification. In contrast, electrocardiography (ECG) is widely available in most care settings, including pre-hospital, and a 12-lead ECG can be obtained within minutes. A number of ECG parameters correlated with PE severity and right ventricular dysfunction have been previously described. The S1Q3T3 pattern has been historically associated with PE but has low sensitivity. The Daniel score combines the S1Q3T3 pattern with tachycardia, right bundle branch block, and anterior T-wave inversion and has been shown to correlate with pulmonary artery pressure,[9] lung perfusion defects,[10] and right ventricular hypokinesis.[16] Further studies have identified additional ECG features with prognostic significance.[3,4] While empirically derived scores like the Daniel score can help with clinical prognostication, ML and AI models provide a more sensitive approach to considering all aspects of the ECG to optimize risk stratification. AI-ECG approaches to identifying patients with any PE have previously been described,[17–19] but these tools provide limited insight into the haemodynamic significance of the identified PE.

One key advantage of our feature-based AI-ECG model is the ability to examine the contribution of specific features towards model prediction. The features that most influenced the model's classification of PE risk included both traditional and novel ECG markers of PE. The feature of highest importance was the concavity of the ST segment in the anterior leads, which reflects both the direction and amplitude of the anterior T waves and the shape of the ST segment. Other model-identified key features that have previously been noted to predict outcomes include lateral S wave (traditionally in Lead I, but in Lead V6 in the model) and T-wave amplitude in Lead III. In addition to these, the model identified QTC prolongation as a predictor of severe PE, which is not captured in existing ECG-based risk scores but has previously been reported.[20,21] Other model-selected parameters included QRS amplitude in lead V4, which is consistent with a delayed

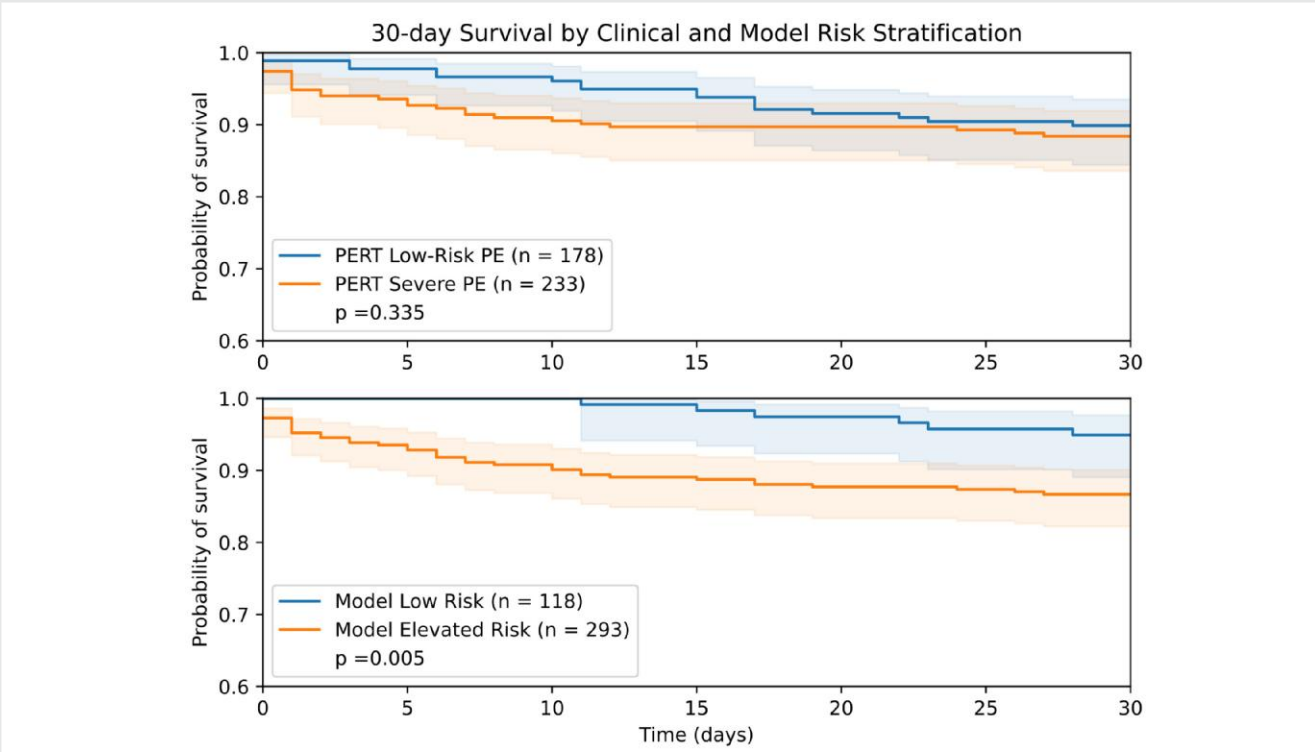


Figure 5 AI-ECG outperforms clinical risk assessment in predicting 30-day mortality. A 30-day mortality was comparable between those identified as low-risk PE and severe PE by the clinical PERT (top panel), but patients with AI-ECG predicted elevated risk had a significantly higher 30-day mortality than those with AI-ECG predicted low risk (bottom panel).

precordial QRS transition and has previously been described as 'clockwise rotation',[22] as well as the spatial ventricular gradient as measured by vectorcardiography, which was recently described in association with adverse PE outcomes.[23]

While our AI-ECG model was trained to predict clinical risk stratification, we also examined its performance on classifying clinical outcomes. Interestingly, the clinical PERT's traditional risk stratification did not predict clinical outcomes, with comparable 30-day mortality between low-risk PE and severe PE patients. This is likely due to referral bias in the low-risk PE group, as the PERT is usually not involved in the care of stable PE patients with few comorbidities who are felt to be at low risk by the treating providers. Indeed, of the low-risk PE patients in the test cohort, only 55% had a primary discharge diagnosis related to PE, compared with 82% of severe PE patients, suggesting that the low-risk PE patients in our cohort had incidental or secondary pulmonary emboli. As such, patients in the low-risk PE group may have been accurately classified by the PERT clinicians as having a lower clinical risk of PE-related adverse outcomes but had overall higher risk of adverse outcomes due to their other illnesses, which most commonly include sepsis, stroke, COVID infection, malignancy, acute myocardial infarction (MI), lower extremity fracture, and intracranial injury. Despite this, the model's prediction successfully identifies a group with higher risk of in-hospital and 30-day mortality. This may be due to the model identifying more subtle evidence of PE related to haemodynamic changes or because the model identifies signals consistent with overall clinical illness. This is consistent with the model's modest specificity as the high false-positive rate is likely due to patients with low PE-related risk who were classified by the model as having a higher-risk phenotype due to visually imperceptible ECG markers of overall illness.

The high sensitivity of our model (0.88) is balanced by its relatively modest specificity (0.51). However, in a population with imaging-confirmed PE, this strikes a clinically reasonable balance in ensuring that patients with severe PE are rapidly identified at the expense of over-estimating the risk of some low-risk PE patients. We propose that this model can be used in conjunction with the traditional risk stratification approach to rapidly deliver care to patients at highest risk. Early identification of patients at the highest risk could allow for escalation of care to an intensive care unit (ICU), coordination of urgent echocardiography, and activation of the PERT team for consideration of interventional therapies. All patients with PE require anticoagulation, so the risks associated with these interventions for a 'false-positive' are relatively minimal.

There are limitations to this work. First, it uses data from a single centre and only includes patients who were evaluated by the PERT. Because patients with low-risk PE are less likely to have the PERT team consulted and because patients with severe PE are more likely to be transferred to a tertiary care centre such as ours, this introduces referral bias into the cohort as noted previously. In addition, the cohort was limited to patients diagnosed with PE within 1 day of hospital admission, or transferred for PE, and exclusion of PEs diagnosed while in-hospital may limit applicability of the resulting risk stratification. Third, we grouped massive/high-risk and submassive/intermediate-risk PE into a single group given the small number of massive PE patients; however, massive PE patients are clinically unstable at presentation and may have more dramatic ECG findings.

While our AI-ECG algorithm demonstrated strong performance in a selected cohort of PE patients evaluated by the PERT team, future work is necessary to assess its generalizability in an unselected, real-world population. In addition, prospective evaluation in this broader clinical

context will allow us to determine how AI-ECG risk stratification can be integrated with traditional markers of high clinical risk to best inform management decisions in patients with acute PE.

Overall, this study demonstrates that a feature-based AI-ECG model can accurately risk stratify patients with PE when compared with the clinical PERT team classification and can outperform the clinical classification in predicting short-term mortality. This model could serve as an important adjunctive tool for clinical risk stratification, by combining traditional and non-traditional ECG markers of PE risk into a single score to augment traditional metrics of PE risk, particularly at sites without PERT expertise.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Author contributions

T.A.G., S.A-Z. and C.T. conceptualized the study. T.A., B.R.L. and C.T. curated the data. T.A.G., N.T.R., E.S., M.A. and S.A-Z. developed the methodology and performed investigations. T.A.G., N.T.R., B.M., R.Q.J., S.A-Z. and C.T. analyzed the data. E.S., M.A., S.F.S., S.A-Z. and C.T. provided supervision. T.A.G., S.F.S. and S.A-Z. acquired funding. T.A.G. wrote the original draft. All authors contributed to reviewing and editing the manuscript.

Funding

This work was supported by the National Heart, Lung, and Blood Institute through grant T32HL129964 (T.A.G.) and R01HL137761 (S.A-Z.) and by the McKamish Family Foundation (T.A.G.). This research was supported in part by the University of Pittsburgh Center for Research Computing and Data, RRID:SCR_022735, through the resources provided. Specifically, this work used the H2P cluster, which is supported by National Science Foundation award number OAC-2117681.

Conflict of interest: none declared.

Data availability

The raw ECG data underlying this article cannot be shared publicly for the privacy of the individuals included in the study. The computed ECG features and the AI model code will be shared on reasonable request.

References

1. Konstantinides SV, Meyer G, Becattini C, Bueno H, Geersing G-J, Harjola V-P, *et al.* 2019 ESC guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS). *Eur Heart J* 2020;**41**:543–603.
2. Rivera-Lebron B, McDaniel M, Ahrar K, Alrifai A, Dudzinski DM, Fanola C, *et al.* Diagnosis, treatment and follow up of acute pulmonary embolism: consensus practice from the PERT consortium. *Clin Appl Thromb* 2019;**25**:1–16.
3. Shopp JD, Stewart LK, Emmett TW, Kline JA. Findings from 12-lead electrocardiography that predict circulatory shock from pulmonary embolism: systematic review and meta-analysis. *Acad Emerg Med* 2015;**22**:1127–1137.
4. Qaddoura A, Digby GC, Kabali C, Kukla P, Zhan ZQ, Baranchuk AM. The value of electrocardiography in prognosticating clinical deterioration and mortality in acute pulmonary embolism: a systematic review and meta-analysis. *Clin Cardiol* 2017;**40**:814–824.
5. Bouzid Z, Faramand Z, Gregg RE, Frisch SO, Martin-Gill C, Saba S, *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* 2021;**10**:1–13.
6. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995;**57**:289–300. http://www.jstor.org/stable/2346101.
7. Al-Zaiti SS, Martin-Gill C, Zègre-Hemsey JK, Bouzid Z, Faramand Z, Alrawashdeh MO, *et al.* Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction. *Nat Med* 2023;**29**:1804–1813.
8. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 2017;**2017**:4766–4775.
9. Daniel KR, Courtney DM, Kline JA. Assessment of cardiac stress from massive pulmonary embolism with 12-lead ECG. *Chest* 2001;**120**:474–481.
10. Iles S, Le Heron CJ, Davies G, Turner JG, Beckert LEL. ECG score predicts those with the greatest percentage of perfusion defects due to acute pulmonary thromboembolic disease. *Chest* 2004;**125**:1651–1656.
11. Kasper W, Konstantinides S, Geibel A, Olschewski M, Heinrich F, Grosser KD, *et al.* Management strategies and determinants of outcome in acute major pulmonary embolism: results of a multicenter registry. *J Am Coll Cardiol* 1997;**30**:1165–1171.
12. Kucher N, Goldhaber SZ. Cardiac biomarkers for risk stratification of patients with acute pulmonary embolism. *Circulation* 2003;**108**:2191–2194.
13. Becattini C, Vedovati MC, Agnelli G. Prognostic value of troponins in acute pulmonary embolism: a meta-analysis. *Circulation* 2007;**116**:427–433.
14. Aujesky D, Obrosky DS, Stone RA, Auble TE, Perrier A, Cornuz J, *et al.* Derivation and validation of a prognostic model for pulmonary embolism. *Am J Respir Crit Care Med* 2005;**172**:1041–1046.
15. Jiménez D, Aujesky D, Moores L, Gómez V, Lobo JL, Uresandi F, *et al.* Simplification of the pulmonary embolism severity index for prognostication in patients with acute symptomatic pulmonary embolism. *Arch Intern Med* 2010;**170**:1383–1389.
16. Ryu HM, Lee JH, Kwon YS, Lee SH, Bae MH, Lee JH, *et al.* Electrocardiography patterns and the role of the electrocardiography score for risk stratification in acute pulmonary embolism. *Korean Circ J* 2010;**40**:499–506.
17. Su XF, Fan N, Yang XM, Song JM, Peng QH, Liu X. A novel electrocardiography model for the diagnosis of acute pulmonary embolism. *Front Cardiovasc Med* 2022;**9**:1–11.
18. Somani SS, Honarvar H, Narula S, Landi I, Lee S, Khachatoorian Y, *et al.* Development of a machine learning model using electrocardiogram signals to improve acute pulmonary embolism screening. *Eur Hear J—Digit Heal* 2022;**3**:56–66.
19. Wysokinski WE, Meverden RA, Lopez-Jimenez F, Harmon DM, Medina Inojosa BJ, Suarez AB, *et al.* Electrocardiogram signal analysis with a machine learning model predicts the presence of pulmonary embolism with accuracy dependent on embolism burden. *Mayo Clin Proc Digit Heal* 2024;**2**:453–462.
20. Gowda RM, Vasavada BC, Khan IA. Prolonged QT interval in acute pulmonary embolism. *Int J Cardiol* 2004;**97**:335–337.
21. Buppajarntham S, Seetha Rammohan HR, Junpaparp P, Figueredo VM. Prognostic value of prolonged QTc interval in patients with acute pulmonary embolism. *Acta Cardiol* 2014;**69**:550–555.
22. Kukla P, Długopolski R, Krupa E, Furtak R, Szełemej R, Mirek-Bryniarska E, *et al.* Electrocardiography and prognosis of patients with acute pulmonary embolism. *Cardiol J* 2011;**18**:648–653.
23. Stabenau HF, Marcus M, Matos JD, McCormick I, Litmanovich D, Manning WJ, *et al.* The spatial ventricular gradient is associated with adverse outcomes in acute pulmonary embolism. *Ann Noninvasive Electrocardiol* 2023;**28**:1–9.