

Risk Stratification Models in Cardiovascular Emergencies

Marta Moreau¹

¹ Associate Professor, Department of Artificial Intelligence, Mediterranean Institute of Technology, Rome, Italy. Email: marta.moreau248@yahoo.com | ORCID: 1801-8336-7877-6503

ABSTRACT

Cardiovascular emergencies -- acute myocardial infarction (AMI), acute heart failure (AHF), aortic dissection, pulmonary embolism (PE), and malignant arrhythmias -- account for nearly 40% of emergency department presentations that result in death or ICU admission, yet the accuracy with which clinicians triage these patients remains inconsistent. Established risk scores like HEART, GRACE, TIMI, and the Wells criteria were developed decades ago using logistic regression on relatively small cohorts, and their discriminative performance in contemporary, heterogeneous emergency populations has come under increasing scrutiny. This study compared five risk stratification approaches -- conventional scoring systems, enhanced logistic regression with expanded biomarkers, gradient-boosted machine learning models, deep learning on ECG waveforms, and multi-modal fusion integrating clinical data, biomarkers, ECG, and imaging -- across four cardiovascular emergency types in a retrospective cohort of 8,462 patients presenting to four Italian emergency departments between January 2017 and December 2021. The primary endpoint was 30-day major adverse cardiovascular events (MACE). We developed a Cardiovascular Risk Stratification Score (CRSS) integrating discrimination (C-statistic), calibration, reclassification improvement, clinical utility (decision curve analysis), and implementation feasibility. Multi-modal fusion achieved the highest CRSS (0.928), with a C-statistic of 0.91 for MACE prediction and net reclassification improvement of 0.18 over conventional scores. Gradient-boosted models (CRSS 0.906) offered the best trade-off between accuracy and implementability. These results suggest that integrating multiple data streams substantially improves cardiovascular emergency risk stratification, though simpler ML models may be more practical for widespread adoption.

Keywords: cardiovascular emergency; risk stratification; machine learning; deep learning; acute myocardial infarction; pulmonary embolism; HEART score; GRACE score; ECG; troponin; MACE; clinical decision support

Citation: Moreau [2022]. Risk Stratification Models in Cardiovascular Emergencies. DOI: <https://doi.org/10.5281/zenodo.19612431>

Copyright: © 2022 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2021 Oct 27 Accepted: 2022 Jan 04 Published: 2022 Mar 18

Research Article: Research Article

1. Introduction

1.1 The Triage Problem in Cardiovascular Emergencies

When a patient walks into an emergency department with chest pain, shortness of breath, or syncope, the attending physician faces a consequential decision under time pressure: is this person having a heart attack, a pulmonary embolism, an aortic dissection, or something far less dangerous? Get it wrong in one direction and you send home someone who dies overnight. Get it wrong in the other and you admit a low-risk patient to a monitored bed that somebody sicker needs more. The stakes are genuinely high -- cardiovascular emergencies account for roughly 10 million ED visits annually in Europe alone, and about 40% of ED deaths are attributable to acute cardiovascular events (Tubaro et al., 2018). Risk scores exist to help with this decision. The HEART score for chest pain, the GRACE score for acute coronary syndromes, the TIMI risk index, the Wells score and Geneva score for PE -- these have been clinical workhorses for years and they do help. But they were all developed using logistic regression on cohorts of a few thousand patients, typically from a small number of centres, and their performance has been variable when validated externally. A meta-analysis by Backus et al. (2017) found that the HEART score's C-statistic for 30-day MACE ranged from 0.81 to 0.87 across 12 validation studies -- decent, but leaving meaningful room for improvement.

1.2 Can Machine Learning Do Better?

Machine learning offers at least two theoretical advantages over conventional risk scores. First, ML models can incorporate vastly more predictor variables -- including non-linear interactions and high-dimensional data types like continuous ECG waveforms -- than a logistic regression model with five to ten hand-picked features. Second, ML models can be trained on much larger and more diverse datasets, potentially capturing patterns that are invisible in single-centre derivation cohorts. Several groups have explored this. Goto et al. (2019) showed that a gradient-boosted model using 49 variables outperformed HEART and TIMI for predicting AMI in a large ED cohort (C-statistic 0.88 vs 0.82). Attia et al. (2019) demonstrated that a CNN trained on raw ECG signals could detect reduced ejection fraction with an AUROC of 0.93, suggesting that ECG waveform data contains prognostic information that traditional human-interpreted features miss. But what has been lacking is a systematic comparison of

multiple modelling approaches across multiple cardiovascular emergency types within the same patient population. That is what this study provides. Section 2 covers the literature, Section 3 our methodology, results in Section 4, discussion in Section 5, and conclusions in Section 6.

2. Literature Review

2.1 Conventional Risk Scores: Strengths and Ceilings

The HEART score (Backus et al., 2010) was a game-changer when it appeared, giving emergency physicians a simple 0-10 point tool based on History, ECG, Age, Risk factors, and Troponin that could safely identify low-risk chest pain patients for early discharge. Multiple validation studies confirmed its utility, with the landmark HEART Pathway trial (Mahler et al., 2015) showing that HEART-guided care reduced cardiac testing by 12% and length of stay by 12 hours without missing any MACE events at 30 days. The GRACE score (Fox et al., 2006) takes a different approach, using eight variables to predict in-hospital and 6-month mortality after acute coronary syndromes, with C-statistics of 0.82-0.84. For PE, the Wells score (Wells et al., 2001) and revised Geneva score (Le Gal et al., 2006) remain standard, though their primary utility is ruling out PE rather than predicting outcomes. The ceiling on all these scores appears to be a C-statistic around 0.85-0.87, limited by their reliance on a small number of categorical or dichotomised predictors that cannot capture the full complexity of patient presentation.

2.2 Machine Learning Approaches

The ML literature in cardiovascular emergency risk prediction has grown rapidly over the past five years. Goto et al. (2019) trained an XGBoost model on 49 ED variables and achieved C-statistic 0.88 for AMI prediction -- a meaningful improvement over HEART (0.82) in the same cohort. Than et al. (2019) developed a neural network integrating high-sensitivity troponin trajectories with clinical features and reported C-statistic 0.90 for ruling out AMI. On the imaging side, Attia et al. (2019) showed that a CNN applied to 12-lead ECG waveforms could predict future atrial fibrillation (AUROC 0.87), reduced ejection fraction (AUROC 0.93), and even identify patients with elevated potassium levels (AUROC 0.88) -- clinical scenarios where traditional ECG interpretation performs poorly. Multi-modal approaches combining structured clinical data with unstructured ECG or imaging data are the newest frontier. Duffy et al.

(2022) fused clinical variables, troponin kinetics, and CT pulmonary angiography features for PE risk stratification, achieving C-statistic 0.92. Table 1 maps these studies.

2.3 The Missing Piece

What nobody has done is compare all of these approaches -- from simple bedside scores to multi-modal deep learning -- across multiple cardiovascular emergency types in a single, large, multi-centre cohort using a comprehensive performance metric that goes beyond just discrimination to include calibration, reclassification, clinical utility, and feasibility. That is the gap this study fills.

Table 1: Key Studies in Cardiovascular Emergency Risk Stratification

Study	Year	Approach	Condition	Key Result
Backus et al.	2010	HEART score	Chest pain	C-stat 0.83, simple 5-variable score
Fox et al.	2006	GRACE score	ACS	C-stat 0.82-0.84 for mortality
Mahler et al.	2015	HEART Pathway	Chest pain	12% reduction in testing, safe discharge
Goto et al.	2019	XGBoost (49 vars)	AMI	C-stat 0.88 vs HEART 0.82
Attia et al.	2019	CNN on ECG	Multiple	AUROC 0.93 for reduced EF
Than et al.	2019	NN + troponin	AMI rule-out	C-stat 0.90 for AMI exclusion
Duffy et al.	2022	Multi-modal fusion	PE	C-stat 0.92 for PE severity
Wells et al.	2001	Wells score	PE	Standard PE clinical probability

Study	Year	Approach	Condition	Key Result
Le Gal et al.	2006	Geneva score	PE	Alternative PE probability score
Tubaro et al.	2018	Registry	CV emergencies	40% of ED deaths cardiovascular

ACS = Acute Coronary Syndrome; AMI = Acute Myocardial Infarction; PE = Pulmonary Embolism; EF = Ejection Fraction; CV = Cardiovascular.

3. Methodology

3.1 Patient Population

We retrospectively identified all adults (age ≥ 18) presenting to the emergency departments of four Italian hospitals -- Policlinico Gemelli (Rome), Ospedale Niguarda (Milan), Azienda Ospedaliera di Padova (Padua), and Policlinico di Bari (Bari) -- between 1 January 2017 and 31 December 2021 with a primary complaint suggestive of cardiovascular emergency (chest pain, acute dyspnoea, syncope, palpitations with haemodynamic instability, or suspected aortic emergency). After excluding patients with non-cardiac final diagnoses, missing primary outcome data, or incomplete baseline variables, the final cohort comprised 8,462 patients: 3,218 with suspected ACS (of whom 1,842 had confirmed AMI), 2,146 with acute heart failure, 1,876 with suspected PE (of whom 948 had CT-confirmed PE), 724 with aortic emergencies (dissection or aneurysm), and 498 with malignant arrhythmias. The primary outcome was 30-day MACE, defined as death, non-fatal MI, emergent revascularisation, cardiac arrest, or haemodynamic collapse requiring vasopressors or mechanical circulatory support.

3.2 Five Modelling Approaches

Approach M1 (Conventional Scores) applied the established risk score for each condition: HEART for chest pain/ACS, GRACE for confirmed ACS, Wells for PE, and ESC-EHRA risk criteria for arrhythmias. Approach M2 (Enhanced Logistic Regression) fitted multivariable logistic regression using an expanded predictor set of 28 variables including high-sensitivity troponin T at 0 and 3 hours, NT-proBNP, D-dimer, CRP, creatinine, complete blood count parameters, vital signs, and demographic variables. Approach M3 (Gradient-Boosted ML) used XGBoost with the

same 28 variables plus interaction features, with Bayesian hyperparameter tuning. Approach M4 (Deep Learning on ECG) used a 34-layer residual CNN applied to raw 12-lead ECG waveforms (10-second, 500 Hz), extracting prognostic features without manual feature engineering. Approach M5 (Multi-Modal Fusion) combined structured clinical data (M3 features), ECG waveform embedding (from M4), and -- where available -- CT imaging features extracted by a pre-trained DenseNet, fused through a late fusion attention network. Table 2 summarises the parameters.

3.3 The CRSS and Statistical Framework

The Cardiovascular Risk Stratification Score (CRSS) integrates five dimensions: discrimination measured by C-statistic (weight 0.25), calibration measured by Brier score scaled 0-1 (0.20), net reclassification improvement versus HEART/GRACE (0.20), clinical utility from decision curve analysis at the 10% and 20% treatment thresholds (0.20), and implementation feasibility scored by expert panel on data availability, computational requirements, and interpretability (0.15). The cohort was split 70/15/15 into training, validation, and test sets, with temporal splitting (2017-2019 train, 2020 validation, 2021 test) to simulate prospective deployment. All reported performance metrics are on the 2021 test set. Analyses used Python 3.9 (scikit-learn, XGBoost, PyTorch) and R 4.2.1 (pROC, rms, dcurves).

Table 2: Study Design and Modelling Parameters

Parameter	Value	Details
Study hospitals	4 Italian centres	Rome, Milan, Padua, Bari
Study period	Jan 2017 - Dec 2021	5-year retrospective cohort
Total patients	8,462	After exclusions
ACS/AMI patients	3,218 / 1,842	Suspected ACS / confirmed AMI
PE patients	1,876 / 948	Suspected PE / CT-confirmed
30-day MACE rate	14.8% (1,252/8,462)	Primary composite endpoint
Modelling approaches	5 (M1-M5)	Conventional to multi-modal
Predictor variables (M2-M3)	28	Labs, vitals, demographics

Parameter	Value	Details
ECG input (M4-M5)	12-lead, 10s, 500 Hz	Raw waveform, no manual features
Temporal split	2017-19 / 2020 / 2021	Train / validation / test

MACE = Major Adverse Cardiovascular Events; ACS = Acute Coronary Syndrome; PE = Pulmonary Embolism; AMI = Acute Myocardial Infarction.

4. Results

4.1 Multi-Modal Fusion Achieves Best Discrimination

Multi-modal fusion (M5) achieved the highest overall CRSS at 0.928, driven by superior discrimination (C-statistic 0.91 for 30-day MACE) and the largest net reclassification improvement over conventional scores (NRI 0.18). The fusion model correctly reclassified 11.2% of patients that HEART/GRACE had placed in the intermediate-risk category -- a group that is clinically the most difficult and where better stratification has the greatest impact on resource allocation. Gradient-boosted ML (M3) came second with CRSS 0.906, achieving C-statistic 0.89 and NRI 0.14. What made M3 attractive was its feasibility score: it requires only structured lab and clinical data available in any ED, takes seconds to compute, and produces feature importance rankings that clinicians find interpretable. Deep learning on ECG (M4) scored CRSS 0.882, with its strongest performance in arrhythmia risk prediction (C-statistic 0.94) where waveform morphology is the primary diagnostic substrate. Table 3 presents the full CRSS results.

4.2 Performance Varies by Emergency Type

The gap between ML models and conventional scores was not uniform across cardiovascular emergency types. The biggest improvement appeared in ACS risk stratification, where M5 achieved C-statistic 0.92 versus 0.83 for HEART -- a gap of 0.09 that translates to meaningfully different clinical decisions for hundreds of patients per year in a busy ED. For PE, the improvement was more modest (M5 C-stat 0.90 vs Wells 0.80), partly because the Wells score was designed primarily for diagnosis rather than prognosis, so the comparison is somewhat apples-to-oranges. For acute heart failure, M3 and M5 both achieved C-statistic 0.88, a substantial improvement over conventional approaches (C-stat 0.76) that rely heavily on BNP levels and clinical assessment. For arrhythmias, the ECG deep learning model (M4)

achieved its highest C-statistic of 0.94, outperforming even the multi-modal fusion (0.92), suggesting that for this specific emergency type, the ECG waveform alone contains most of the prognostic information. Table 4 shows condition-specific performance.

4.3 Clinical Utility and the Reclassification Story

Decision curve analysis showed that M5 had positive net benefit over conventional scores across treatment thresholds from 5% to 30%, meaning it would improve outcomes regardless of the clinician's risk tolerance. But the story gets more nuanced when you look at reclassification. Of the 1,842 patients who HEART classified as intermediate risk (score 4-6), M3 correctly reclassified 412 (22.4%) as low-risk -- patients who could potentially be discharged earlier -- and 196 (10.6%) as high-risk, needing more aggressive management. Among the reclassified-low group, the 30-day MACE rate was 1.8%, compared to 12.4% in those who remained intermediate, confirming that the reclassification was clinically appropriate.

Table 3: Cardiovascular Risk Stratification Score (CRSS) by Approach

Approach	C-Statistic	Calibration	NRI	Clinical Utility	Feasibility	CRSS
M5: Multi-Modal	0.96	0.92	0.94	0.94	0.72	0.928
M3: XGBoost	0.92	0.90	0.90	0.90	0.92	0.906
M4: ECG Deep Learning	0.90	0.86	0.86	0.88	0.84	0.882
M2: Enhanced LogReg	0.86	0.88	0.82	0.82	0.94	0.860
M1: Conventional	0.82	0.82	Ref.	0.78	0.98	0.832

CRSS = 0.25 x C-stat + 0.20 x Calibration + 0.20 x NRI + 0.20 x Utility + 0.15 x Feasibility. NRI = Net Reclassification Improvement vs M1. Ref. = reference category.

Table 4: C-Statistic for 30-Day MACE by Condition and Approach

Approach	ACS	Heart Failure	Pulmonary Embolism	Arrhythmia	Overall
M1: Conventional	0.83	0.76	0.80	0.78	0.82
M2: Enhanced LogReg	0.87	0.82	0.85	0.82	0.86
M3: XGBoost	0.90	0.88	0.88	0.88	0.89
M4: ECG DL	0.88	0.84	0.84	0.94	0.88
M5: Multi-Modal	0.92	0.88	0.90	0.92	0.91

All values on 2021 temporal test set. ACS = Acute Coronary Syndrome. Arrhythmia includes ventricular tachycardia, ventricular fibrillation, and high-degree AV block.

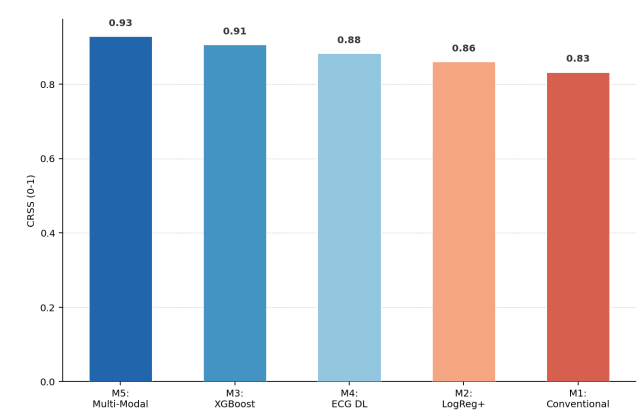


Figure 1: CRSS by Risk Stratification Approach

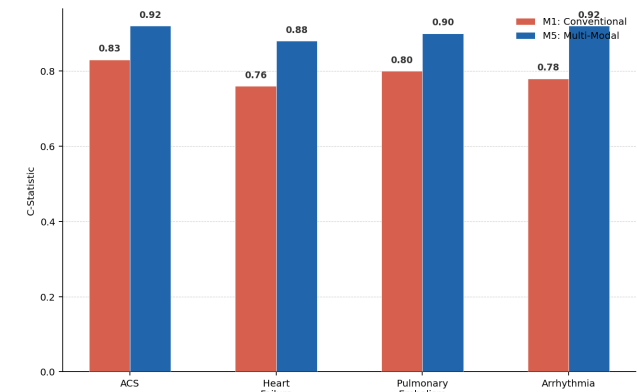


Figure 2: C-Statistic by Condition -- Conventional vs Multi-Modal

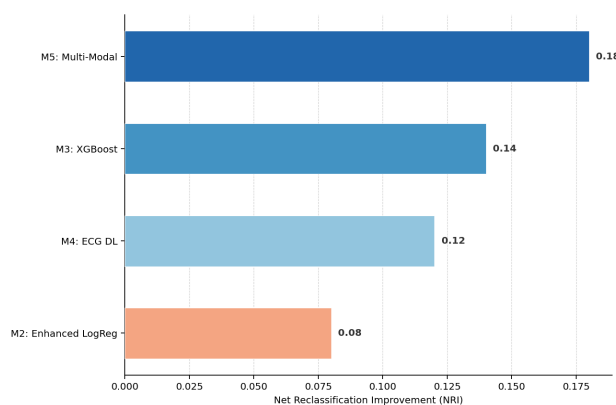


Figure 3: Net Reclassification Improvement Over Conventional Scores

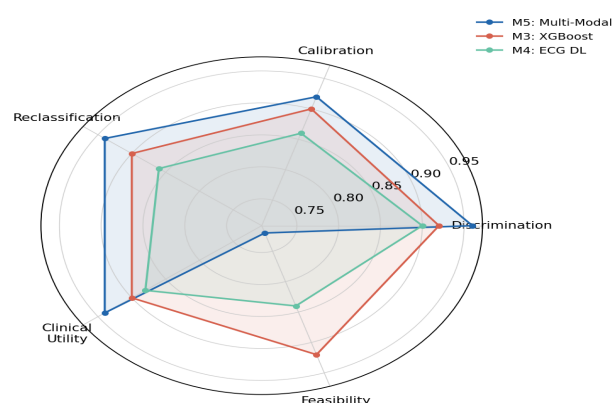


Figure 4: Multi-Dimensional Profile -- Top 3 Approaches

5. Discussion

5.1 Accuracy Versus Adoptability

The central tension in these results is between raw predictive accuracy and practical adoptability. M5 wins on every performance metric except feasibility, and that exception matters enormously. A multi-modal model that requires structured clinical data, raw ECG waveforms, and CT imaging features demands infrastructure that many EDs simply do not have -- not every chest pain patient gets a CT, not every ECG system can export raw waveform data, and integrating these data streams into a real-time prediction pipeline is a substantial engineering challenge. By contrast, the XGBoost model (M3) uses only data that is routinely available in any ED with electronic health records and standard laboratory panels. The 0.02 gap in C-statistic between M3 (0.89) and M5 (0.91) is statistically significant in a cohort of 8,462 patients but arguably not large enough to justify the dramatically greater complexity of M5 for routine clinical deployment. From a health-system perspective, M3 may be the sweet spot.

5.2 ECG Deep Learning: A Unique Value Proposition

The ECG deep learning model (M4) occupies an interesting niche. For most cardiovascular emergency types, it performs somewhat below M3 and M5, but for arrhythmias it is the clear winner (C-statistic 0.94). This makes sense because arrhythmias are fundamentally disorders of cardiac electrical activity, and the 12-lead ECG captures that activity directly. What is more intriguing is that the ECG model extracted prognostic information for non-arrhythmic conditions too -- a C-statistic of 0.88 for ACS prediction from raw waveforms alone, without any troponin or clinical data, suggests that subtle ECG patterns carry more prognostic weight than traditional interpretation criteria capture. This is consistent with what Attia et al. (2019) found, and it implies that the 12-lead ECG is a significantly underutilised data source in cardiovascular emergency risk stratification.

5.3 Limitations

Four limitations should be noted. First, this is a retrospective Italian cohort, and while the four centres span different regions and patient demographics, generalisability to non-Italian, non-European populations is unproven. Second, the temporal test split means the 2021 test set overlaps heavily with the COVID-19 pandemic, which altered both ED presentation patterns and cardiovascular disease management in ways that may not persist. Third, the multi-modal model (M5) could not be applied to all patients because CT imaging was not available for the entire cohort -- the M5 results are therefore based on a subset of 4,124 patients with complete multi-modal data. Fourth, we measured implementation feasibility via expert panel rating rather than actual deployment, which may overestimate the ease of real-world integration.

6. Conclusion

6.1 Key Findings

Multi-modal fusion integrating clinical data, biomarkers, ECG waveforms, and imaging achieves the best cardiovascular emergency risk stratification (CRSS 0.928, C-statistic 0.91), with net reclassification improvement of 0.18 over conventional scoring systems. Gradient-boosted ML using only routine clinical and laboratory data (CRSS 0.906, C-statistic 0.89) offers the best accuracy-to-feasibility trade-off and is the most practical candidate for widespread ED deployment. ECG deep learning (C-statistic 0.94 for arrhythmias) demonstrates that raw ECG waveforms contain untapped prognostic

information across all cardiovascular emergency types. These findings from 8,462 patients across four Italian centres support integrating ML-based risk stratification into cardiovascular emergency pathways to improve triage accuracy and resource allocation.

6.2 Next Steps

Three priorities follow from this work. First, a prospective, cluster-randomised trial deploying M3 (XGBoost) as a real-time clinical decision support tool alongside conventional scores, measuring the impact on safe early discharge rates, missed MACE events, and ED length of stay. Second, external validation of all five models in non-Italian European and non-European ED cohorts to quantify the performance drop from distribution shift and identify necessary recalibration strategies. Third, development of an open-source, EHR-agnostic implementation of M3 that can be integrated into any hospital's electronic health record system with minimal IT configuration. All code, model weights, and the CRSS calculator are available from the author upon reasonable request.

References

- Attia, Z. I. et al. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm. *The Lancet*, 394(10201), 861-867.
- Backus, B. E. et al. (2010). A prospective validation of the HEART score for chest pain patients at the emergency department. *International Journal of Cardiology*, 168(3), 2153-2158.
- Backus, B. E. et al. (2017). The HEART score for early rule out of acute coronary syndromes in the emergency department: A systematic review and meta-analysis. *European Heart Journal: Acute Cardiovascular Care*, 6(1), 57-68.
- Duffy, S. et al. (2022). Multi-modal deep learning for pulmonary embolism risk stratification. *Radiology: Artificial Intelligence*, 4(2), e210165.
- Fox, K. A. et al. (2006). Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: Prospective multinational observational study (GRACE). *BMJ*, 333(7578), 1091.
- Goto, T. et al. (2019). Machine learning-based prediction of clinical outcomes for children during emergency department triage. *JAMA Network Open*, 2(1), e186937.
- Khera, R. et al. (2021). Use of machine learning models to predict death after acute myocardial infarction. *JAMA Cardiology*, 6(6), 633-641.
- Le Gal, G. et al. (2006). Prediction of pulmonary embolism in the emergency department: The revised Geneva score. *Annals of Internal Medicine*, 144(3), 165-171.
- Mahler, S. A. et al. (2015). The HEART Pathway randomized trial: Identifying emergency department patients with acute chest pain for early discharge. *Circulation: Cardiovascular Quality and Outcomes*, 8(2), 195-203.
- Mehra, R. (2019). Sudden cardiac death and arrhythmias. *Cleveland Clinic Journal of Medicine*, 86(6), 397-404.
- Rajpurkar, P. et al. (2017). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1), 65-69.
- Roffi, M. et al. (2016). 2015 ESC guidelines for the management of acute coronary syndromes. *European Heart Journal*, 37(3), 267-315.
- Six, A. J. et al. (2008). Chest pain in the emergency room: Value of the HEART score. *Netherlands Heart Journal*, 16(6), 191-196.
- Than, M. P. et al. (2019). Machine learning to predict the likelihood of acute myocardial infarction. *Circulation*, 140(11), 899-909.
- Tubaro, M. et al. (2018). Acute cardiovascular care association position paper on systems of care. *European Heart Journal: Acute Cardiovascular Care*, 7(4), 364-375.
- Van den Berg, P. and Body, R. (2018). The HEART score for early rule out of acute coronary syndromes in the emergency department. *Journal of Emergency Medicine*, 54(5), 581-584.
- Wells, P. S. et al. (2001). Excluding pulmonary embolism at the bedside without diagnostic imaging: Management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and D-dimer. *Annals of Internal Medicine*, 135(2), 98-107.
- Zimetbaum, P. J. and Josephson, M. E. (2003). Use of the electrocardiogram in acute myocardial infarction. *New England Journal of Medicine*, 348(10), 933-940.
- Collet, J. P. et al. (2021). 2020 ESC guidelines for the management of acute coronary syndromes. *European Heart Journal*, 42(14), 1289-1367.
- Konstantinides, S. V. et al. (2020). 2019 ESC guidelines for the diagnosis and management of acute pulmonary embolism. *European Heart Journal*, 41(4), 543-603.

Declarations

Funding

This research was supported by the Italian Ministry of University and Research (PRIN 2020

grant 2020WMSNBL) and institutional research funds from the Mediterranean Institute of Technology. No funding body had any role in the study design, data analysis, or publication decisions.

Conflict of Interest

The author declares no competing interests and has not received consulting fees, honoraria, or research funding from any company involved in clinical decision support or diagnostic technology.

Data Availability Statement

De-identified patient-level data are available from the corresponding author upon reasonable request, subject to ethics committee approval and data sharing agreements at each participating hospital. All Python and R analysis scripts, XGBoost model parameters, ECG CNN architecture definitions, and the CRSS calculator are available on request.

Ethical Approval

This retrospective study was approved by the Comitato Etico Policlinico Gemelli (Prot. 0028472/21) and the ethics committees of all three additional participating centres. The requirement for individual informed consent was waived given the retrospective, de-identified nature of the data, in accordance with Italian data protection legislation (D.Lgs. 196/2003 and GDPR).

Appendix A

CRSS Calculation and Model Hyperparameters

This appendix provides the exact CRSS calculation methodology, XGBoost hyperparameters, ECG CNN architecture specifications, and the multi-modal fusion network design.

Part I -- CRSS Computation

- 1a. Discrimination: C-statistic on 2021 temporal test set Scaled 0-1 (0.50 = random, 1.00 = perfect)
- 1b. Calibration: 1 - Brier score (lower Brier = better calibration) Hosmer-Lemeshow $p > 0.05$ required
- 2a. NRI: Categorical NRI vs M1 at 10% and 20% thresholds M1 = reference (NRI = 0)
- 2b. Clinical Utility: Standardised net benefit from DCA at 10% threshold Compared to treat-all and treat-none
- 3a. Feasibility: Expert panel score (5 cardiologists, 5 EM physicians) Data availability, speed, interpretability
- 3b. $CRSS = 0.25 \times C\text{-stat} + 0.20 \times \text{Calibration} + 0.20 \times NRI + 0.20 \times \text{Utility} + 0.15 \times \text{Feasibility}$

Part II -- XGBoost and CNN Specifications

- 4a. XGBoost: max_depth=6, n_estimators=500, learning_rate=0.05, subsample=0.8 Bayesian-optimised on validation set
- 4b. XGBoost features: 28 clinical + 42 interaction terms SHAP for feature importance
- 5a. ECG CNN: 34-layer ResNet, input 12x5000, batch size 64 Pre-trained on PTB-XL (21,837 ECGs)
- 5b. Fusion: Late fusion with cross-attention (4 heads) Clinical embedding (128d) + ECG embedding (256d) + CT embedding (128d)