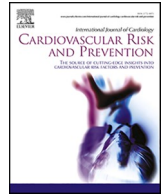




Contents lists available at ScienceDirect

# International Journal of Cardiology Cardiovascular Risk and Prevention

journal homepage: [www.journals.elsevier.com/international-journal-of-cardiology-cardiovascular-risk-and-prevention](http://www.journals.elsevier.com/international-journal-of-cardiology-cardiovascular-risk-and-prevention)



## A next-generation prediction risk model for acute myocardial infarction: Derivation and validation in a multi-centre cohort

Jose David Amorocho-Morales<sup>a</sup>, Sergio Parra Guevara<sup>a,b</sup>, Elias Quintero-Muñoz<sup>a</sup>,  
Guillermo Dimas<sup>c</sup>, Juan Esteban Correa-Morales<sup>d,\*</sup> 

<sup>a</sup> Universidad de La Sabana, Chía, Colombia

<sup>b</sup> Universidad Simón Bolívar, Cúcuta, Colombia

<sup>c</sup> Salud Total, Bogotá DC, Colombia

<sup>d</sup> Faculty of Medicine and Health Sciences, UIC Barcelona – Universitat Internacional de Catalunya, Sant Cugat Del Valles, Barcelona, Spain

### ARTICLE INFO

#### Keywords:

Acute myocardial infarction  
Risk prediction model  
Machine learning in cardiology  
Short-term cardiovascular risk  
Population health stratification

### ABSTRACT

**Aim:** Cardiovascular disease remains the leading global cause of death, and the need for accurate, event-specific risk prediction is particularly critical in regions where long-horizon models perform poorly. We developed and internally validated a probabilistic model to estimate 6- and 12-month risk of acute myocardial infarction, with exploratory 5- and 10-year horizons, using routinely collected electronic health record data from an integrated cardiovascular cohort in Colombia.

**Methods:** The study followed TRIPOD + AI guidance and analysed 382,589 patients contributing 3.9 million encounters. The modelling strategy combined a calibrated gradient-boosting classifier with an interpretable survival ensemble incorporating Cox regression, random survival forests, and discrete-time hazards. Primary outcomes were prediction accuracy, discrimination, calibration, and concordance with legacy score scales.

**Results:** The classifier achieved an AUC of 0.869, while 6- and 12-month survival models reached C-indices of 0.836 and 0.846. Calibration was strong, with predicted vs observed AMI counts nearly identical (O/E = 0.998). Concordance analyses demonstrated only moderate alignment with Framingham and PROCAM, indicating substantial re-ranking at short horizons compared with legacy long-term models. External, label-delayed validation (n = 5602) showed monotonic risk separation across predefined priority bands.

**Conclusion:** This model provides a practical population-health stratification tool for short-term AMI risk, with particular value in resource-constrained settings. Recalibration to local incidence rates is recommended before deployment. Prospective evaluation is warranted to assess real-world clinical and operational impact.

### 1. Introduction

Cardiovascular diseases (CVD) remain the leading cause of morbidity and mortality worldwide, representing one of the greatest global health and economic burdens [1]. In 2023, CVD caused roughly 19.2 million deaths, most of which were attributable to ischemic heart disease, the main cause of heart attacks [2]. Moreover, heart attacks contributed to 437 million disability-adjusted life years in 2023 [2]. Coronary heart disease, including acute myocardial infarction (AMI), is estimated to affect over 250 million people globally, reflecting the growing prevalence and chronicity of cardiovascular conditions [3]. In response to evolving epidemiological patterns and the limitations of traditional cardiovascular risk models, newer and more comprehensive approaches

to risk prediction have emerged [4].

The 2025 ACC/AHA guidelines now recommend the Predicting Risk of Cardiovascular Disease EVENTS (PREVENT) equations as the primary tool for estimating absolute cardiovascular risk [5]. Developed using large, diverse, and contemporary datasets, PREVENT aligns cardiovascular risk assessment with modern concepts of cardiovascular–kidney–metabolic health and enables both 10-year and 30-year risk estimation [6]. In initial external validation cohorts from the United States, PREVENT achieved C-statistics of 0.794 in women and 0.757 in men [7]. However, external validations outside the U.S. have demonstrated variable performance. For example, in a United Arab Emirates cohort, the C-index ranged from approximately 0.81–0.82 in women to 0.70–0.72 in men [8], while in a Swiss cohort of 5064

\* Corresponding author. Carrer de la Immaculada, 22, Sarrià-Sant Gervasi, 08017, Barcelona, Spain

E-mail address: [jecorrea@uic.es](mailto:jecorrea@uic.es) (J.E. Correa-Morales).

<https://doi.org/10.1016/j.ijcrp.2026.200659>

Received 12 February 2026; Received in revised form 6 May 2026; Accepted 30 May 2026

Available online 8 June 2026

2772-4875/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

participants, the C-index ranged from 0.73 to 0.79 (76% concordant pairs), though the model systematically underestimated risk (Observed/Expected ratio  $\approx 1.45$ ) and did not outperform comparator models such as SCORE2 or PCE [9]. These findings underscore that calibration and absolute risk estimates derived from PREVENT may not translate accurately to non-U.S., ethnically diverse, or different baseline-risk populations.

The urgency for modern, more precise prediction models is particularly evident in non-European and non-American regions, where the burden of CVD is disproportionately high [10]. In middle- and low-income countries, CVD accounts for nearly a quarter of all deaths, with mortality rates at least 2.5 times higher than in high-income countries [10]. A systematic review found that widely used Western models—including Framingham, Pooled Cohort Equations, and SCORE—exhibited acceptable discriminative ability but universally poor calibration when predicting 10-year CVD risk in Asian populations [11]. This need for geographical tailoring is globally recognized, exemplified by the World Health Organization's comprehensive effort to revise its CVD risk charts, leading to 21 distinct, recalibrated models designed to better estimate risk in low- and middle-income countries [12]. Despite these large-scale efforts to adapt existing tools, the performance of risk prediction remains a challenge. While the updated SMART2 algorithm was systematically recalibrated, it still demonstrated suboptimal discrimination with a C-statistic of 0.61 (95% CI 0.53–0.68) in the Latin American subcohort [13].

Machine learning (ML) and artificial intelligence have demonstrated promising utility across cardiovascular conditions beyond AMI. In obstructive coronary artery disease, ML approaches applied to exercise-test electrocardiographic features have improved diagnostic accuracy for identifying obstructive lesions [14]. Deep learning models have shown strong performance in predicting short-term mortality in acute pulmonary embolism [15], and AI-based risk stratification has been evaluated in both coronary artery disease and atrial fibrillation [16]. Event-specific risk prediction models have also shown value in high-risk surgical populations [17]. These advances underscore that properly validated ML models can deliver clinically actionable, condition-specific risk estimates that complement and extend traditional scoring frameworks.

Furthermore, existing risk-prediction models focus on long-term horizons and composite cardiovascular outcomes pose important limitations for clinical application [4,6]. By estimating events over 10 years and combining diverse endpoints such as heart attack, stroke, heart failure, or revascularization, these models often produce aggregate risk estimates that overestimate risk and lack specificity for individual outcomes [4,6,18,19]. Emphasizing short-term, event-specific prediction is intended to enhance public health planning by enabling timely, action-oriented decision-making. Although recent models targeting specific outcomes tend to demonstrate higher accuracy [20], many still fall short due to weak methodological rigour and poor adherence to reporting standards, ultimately reducing their reliability [21]. Unlike existing tools that estimate composite cardiovascular outcomes over 10-year horizons, the present study addresses an underserved domain: short-term, event-specific AMI risk stratification using routinely collected electronic health record (EHR) data from a middle-income country setting, with transparent calibration and adherence to TRIPOD + AI guidelines [22]. The purpose of this study is to develop a model capable of independently predicting 6-month and 1-year risk of acute myocardial infarction, with exploratory extensions to 5- and 10-year horizons.

## 2. Methods

### 2.1. Study design and data source

This is a retrospective modelling study using routinely collected data from an integrated cardiovascular cohort in Colombia. The registry

consolidates de-identified records of urban and rural ambulatory encounters, laboratory results, and cardiovascular procedures recorded electronically. The analytic window spanned 1990–2025, with follow-up censored at the last recorded clinical contact. The study was designed and reported in accordance with TRIPOD + AI guidance for clinical prediction models using machine learning and regression [22].

#### 2.1.1. Participants

Adults were eligible if the registry showed at least one cardiovascular risk factor, defined as any recorded diagnosis or structured field indicating hypertension, diabetes mellitus, dyslipidaemia, tobacco exposure, chronic kidney disease, chronic pulmonary disease, or obesity. Patients were not excluded based on treatment status. Although medication orders and procedures were recorded, they did not define the outcome. For time-to-event analyses, only data recorded before the event or censoring were included to prevent label leakage (i.e., to avoid incorporating post-event information into model training). Patient selection is summarized in [Supplementary Fig. S1](#).

### 2.2. Outcomes

The analysis targeted a single endpoint: acute myocardial infarction, defined from structured fields as myocardial infarction with or without revascularization as recorded in the registry. For time-to-event modeling, time zero (baseline) was anchored at each qualifying visit (index), and event times were measured in days to the first occurrence of AMI or censoring. Primary prediction horizons were 6 months (180 days) and 1 year (365 days) to support short-term, actionable decisions. Additional horizons at 5 and 10 years were also analysed to inform operational case management and alignment with long-horizon prevention frameworks. Outcome labels were derived from structured clinical fields consistently captured across sites. No subjective adjudication was performed.

### 2.3. Predictors

Candidate predictors were selected a priori from medical literature and included: age, body mass index (BMI), LDL cholesterol, diabetes (and duration, when available), hypertension, tobacco exposure, chronic kidney disease staging, chronic pulmonary disease severity, functional mobility markers, and indicators of social vulnerability [23–28]. In addition, harmonized “risk signal” from dichotomous clinical flags were synthesized using a Bayesian aggregation layer that yields a latent continuous score reflecting cumulative risk burden [29,30]. This layer takes as inputs a panel of binary clinical flags — including presence of each major risk factor, prior cardiovascular procedures, and markers of clinical surveillance gaps — and models their joint distribution via a latent Gaussian structure with empirical priors estimated on the training data. The output is a single continuous score on a logit-probability scale, static per encounter. This approach was preferred over additive comorbidity indices because it captures covariation among risk flags probabilistically, better reflecting the multiplicative nature of cardiovascular risk. Full details are in [Supplementary Fig. S2](#) and [Supplementary Table S6](#). All predictors were extracted from structured elements of routine care.

### 2.4. Data preparation and quality control

Variable names were harmonized using canonical text mappings to address accent and spelling variants frequent in Colombian documentation. Variables with structural missingness >80% were excluded. Implausible physiologic values were flagged by automated rules and reviewed with clinical leads. For modelling, continuous predictors were centred and scaled using parameters estimated in the training split only. Numeric missingness was imputed with training-median values; the same preprocessing pipeline was applied, unchanged, to validation and test sets to avoid information leakage. Missingness patterns and

distributions were examined across sociodemographic strata, and the same preprocessing steps were applied to all subgroups. Baseline Framingham cardiovascular risk categories (when available in the registry) were retained in their original levels to enable planned concordance analyses.

## 2.5. Sample size and data partitions

The source registry comprised 3,940,059 encounter-level rows for 382,589 unique patients. Modelling was performed on patient/visit-level longitudinal tables with pre-event observations only. The AMI time-to-event model used a stratified 70%/10%/20% split (train/validation/test) on the event indicator, preserving outcome prevalence. The complementary prioritization classifier for AMI employed a 75%/25% split for development and hold-out evaluation. These partitions provide narrow uncertainty around performance and calibration estimates, given the cohort size.

## 2.6. Handling of missing data

Predictors with >80% structural missingness were excluded prior to modelling [31]. Remaining numeric missingness was imputed with training-set medians. Categorical flags were retained as recorded (including explicit “unknown” levels when present). Imputation models and scaling parameters were learned on training data and applied unchanged to validation and test [32]. Missingness indicators were evaluated as candidate features; those not improving validation AUC were excluded. Residual informative missingness is partially captured by surveillance-gap flags in the Bayesian aggregation layer; potential residual bias is acknowledged in the Limitations.

## 2.7. Model development

### 2.7.1. Primary analysis: AMI risk classifier

To mirror clinical prioritization workflows and maximize interpretability, a shallow extreme-gradient-boosting classifier was trained on the Bayesian risk signal plus the core covariates [33]. The estimator produces calibrated probabilities interpreted as the expected 6- and 12-month AMI risk. Calibration and data preprocessing were optimized using the development subset and applied consistently to the hold-out set, with hyperparameters derived from established institutional settings. The 0.50 cut-point follows standard practice for reporting binary discrimination metrics in supervised ML for dichotomous outcomes and is an illustrative reference, not a recommended operational threshold; deployment relies on continuous probability outputs stratified into priority bands (High, Medium, Low) validated externally. Threshold-dependent metrics are reported in [Supplementary Table S3–S4](#); time-dependent AUC and Integrated Brier Score are in [Supplementary Table S5](#).

### 2.7.2. Secondary analysis: complementary AMI time-to-event model

We built an interpretable-first survival ensemble for AMI combining: (i) a Cox proportional hazards model with elastic-net regularization (hazard ratios for clinical review), (ii) a random survival forest (non-linear effects and interactions), and (iii) a discrete-time logistic hazard model emphasizing early risk. Horizon-specific event probabilities (6 months, 1 year, 5 years, and 10 years) were obtained by learning convex blend weights that minimized validation Brier scores, followed by isotonic recalibration on validation data. Residual learners were permitted only if they improved validation Brier without degrading calibration. Covariates matched those of the classifier, and model fitting used pre-event records only.

### 2.7.3. Internal validation, calibration, and leakage control

All preprocessing steps (harmonization, imputation, scaling, calibration) were estimated exclusively on training data and applied

forward to validation and test sets. Blend weights and calibration layers were refit within each training/validation partition to respect uncertainty. Performance was summarized as point estimates on the held-out test sets, and sensitivity analyses computed non-parametric bootstrap intervals to quantify sampling variability for key metrics.

### 2.7.4. Performance metrics

The AMI classifier's performance was evaluated for discrimination and calibration; threshold-dependent metrics at the 0.50 illustrative cut-point are in [Supplementary Table S3](#). For the AMI survival model, the Brier score, C-index, and calibration-in-the-large and calibration slope were reported per horizon; time-dependent AUC and Integrated Brier Score are in [Supplementary Table S5](#). For the AMI survival model, the Brier score, concordance-type index (C-index), and calibration-in-the-large and calibration slope were reported per horizon (6 months, 1 year, 5 years, and 10 years), and visual calibration curves were generated at 6 and 12 months. Decision-curve analyses (net benefit) were conducted on the test set to illustrate potential clinical utility across threshold probabilities pertinent to near-term AMI risk management. Consistent with a probability-as-risk framework, observed versus expected AMI events (sum of predicted probabilities) were summarized by predicted-risk strata over the test set.

### 2.7.5. Legacy comparator and concordance analysis

To contextualize the survival-based prioritization against legacy long-horizon scores, we performed complementary concordance analyses. For each contract with complete information, we formed horizon-specific survival-priority tertiles: a short-term scheme based on the average of the 6- and 12-month survival probabilities, and a long-term scheme based on the average of the 5- and 10-year probabilities. These survival-priority tertiles were cross-tabulated with (i) baseline Framingham risk categories and (ii) tertiles of the continuous AMI-risk score derived from the classifier. Cross-tabulations and Cramér's V from chi-square tests of independence (two-sided,  $\alpha = .05$ ) were used to quantify the strength of association, with emphasis placed on effect size rather than p-values. Analyses used the subset with complete labels for each pair under comparison ( $N = 179,946$  for the main contrasts), and detailed contingency tables are provided in the Supplement. Additional analyses using classifier-based risk tertiles as the reference scheme are also reported in the Supplement.

In a prespecified sensitivity analysis focused on a classical coronary risk score, we also computed the PROCAM 10-year coronary risk according to the standard Münster specification for men aged 35–65 years with complete data, using age, lipid profile (LDL, HDL, triglycerides), systolic blood pressure, smoking, diabetes, and family history of myocardial infarction. The resulting PROCAM\_Risk\_Percent (0–30%) was then compared with the model's AMI risk estimates using Pearson correlation coefficients for short-term and extended-horizon configurations, in order to quantify the degree of linear association in patient ranking between PROCAM and the Corpus probabilities. Framingham and PROCAM were chosen as legacy comparators because both have been validated and are routinely used in clinical practice in Colombia [34].

### 2.7.6. Robustness analysis

Variable directionality (Cox coefficients) and model attribution (forest importance profiles) were inspected to ensure that no single artifact or site-specific variable dominated predictions. As a robustness check, missingness handling was varied (complete-case sensitivity), and primary endpoints were recalculated at alternate index definitions, with results remaining directionally consistent.

### 2.7.7. External, label-delayed validation cohort

An external, label-delayed validation was conducted using a de-identified general-population cohort ( $n = 5602$ ) from an independent insurance provider. Prior to scoring, the insurer's clinical committee

confirmed that the cohort's age distribution was broadly comparable to the derivation population (median age 63 years vs. 64 years in the derivation cohort); detailed patient-level characteristics could not be released under the data-sharing agreement, but aggregate demographic equivalence was verified as a precondition for the validation exercise (see [Supplementary Table S7](#)). AMI probabilities were generated with the locked classifier without re-tuning. Expected events were computed as the sum of predicted probabilities and compared with subsequently observed events (O/E). Pre-specified priority bands (Low <0.25, Medium 0.25–0.75, High ≥0.75) were applied and event rates summarized per band; differences were tested by one-way ANOVA with Tukey HSD. An O/E < 1.0 in the external cohort may reflect lower baseline AMI incidence, delayed administrative coding (label-delayed design), or differential care access; recalibration to local rates is recommended before operational deployment.

2.7.8. Software and reproducibility

Analyses were performed in Python 3.10; key package versions are listed in [Supplementary Table S2](#). Version control preserved an auditable history. Reproducibility packages include scripts that regenerate all tables and figures from curated analytic files (see supplementary materials). To safeguard proprietary components, model families, covariates, and calibration procedures are reported, but full weight vectors and trained artifacts are not disclosed.

2.7.9. Ethics and reporting

The study was approved by the Institutional Review Board of Salud Total with a waiver of informed consent owing to the use of de-identified, retrospective data.

2.7.10. Alignment with clinical use

Focusing on 6- and 12-month AMI risk supports timely triage, targeted diagnostics (e.g., referral optimization), and proactive care management for patients at near-term risk of myocardial infarction. Modelling choices (transparent covariates, Cox-based effect summaries, strict calibration) were selected to maximize clinical interpretability and operational uptake in routine care.

3. Results

3.1. Sample characteristics

The registry comprised 382,589 patients who met eligibility criteria, contributing 3,940,059 encounters (encounter-level records). Overall,

15,511 patients experienced an acute myocardial infarction during follow-up. The patient cohort was 53.1% female and median age was 64 years [IQR 55–72]. At the encounter level, the median body mass index was 27.6 kg/m<sup>2</sup> (IQR 24.9–30.9). The encounter-level median LDL cholesterol was 99 mg/dL (IQR 75–129), and LDL persisted ≥130 mg/dL in 541,065 encounters (13.7%) while triglycerides were ≥150 mg/dL in 1,180,785 encounters (30.0%). Hypertension was documented in 41.9% of encounters, diabetes in 23.7%. Among the 2,912,548 encounters with a registered smoking status, smoking was documented as active in 93,223 (3.2%) and former in 162,222 (5.6%). Chronic kidney disease stage ≥3 was recorded in 0.11% of encounters, COPD in 1.0%, and programme-level social risk indicators were present in 27.4%. Additionally, 1102 encounters (0.03%) noted a prior coronary angiography. Framingham risk classifications were assigned in 857,406 encounters, representing 21.8% of the total. Of these, 7% (about 59,000) were classified as high risk, 44% (about 377,000) as moderate risk, and 49% (about 421,000) as low risk.

3.2. Primary analysis: AMI risk classification

Discrimination and calibration metrics for both the classifier and the survival ensemble are summarized in [Table 1](#). The classifier achieved an AUC of 0.869 on the held-out test set, reflecting robust rank-based discrimination. Threshold-dependent metrics are presented for illustrative purposes in [Supplementary Table S3](#). The model's operational use relies on continuous probability outputs stratified into actionable priority bands (High, Medium, Low) intended to inform outreach scheduling, diagnostic triage, and preventive monitoring. The conventional 0.50 cut-point follows standard reporting practice for binary ML classifiers applied to dichotomous clinical outcomes [\[35,36\]](#) and does not constitute a recommended operational threshold.

3.3. Secondary analysis: AMI time-to-event model

The complementary survival ensemble for AMI preserved low prediction error and good rank ordering at clinically actionable horizons. At 6 months, Brier = 0.013928 and C-index = 0.836; calibration-in-the-large and slope were −0.722 (95% CI −1.035 to −0.410) and 0.801 (95% CI 0.714 to 0.888), respectively. At 12 months, Brier = 0.030350 and C-index = 0.846; calibration-in-the-large and slope were −0.628 (95% CI −0.803 to −0.453) and 0.761 (95% CI 0.698 to 0.825) (see [Table 1](#) and [Fig. 1](#)).

The negative calibration-in-the-large intercepts at short horizons indicate that the raw survival model systematically over-predicted AMI

**Table 1**  
AMI classifier performance and survival prediction accuracy at 6 and 12 months, 5 years, and 10 years.

| Primary Analysis                |         |             |                                |                            |           |          |          |
|---------------------------------|---------|-------------|--------------------------------|----------------------------|-----------|----------|----------|
| Metric                          | AUC     | Sensitivity | Specificity                    | Brier score                | Precision | F1-score | Accuracy |
| Risk stratification performance | 0.869   | 0.812       | 0.757                          | 0.147                      | 0.77      | 0.791    | 0.784    |
| Secondary Analysis              |         |             |                                |                            |           |          |          |
| Survival                        | C-index | Brier score | Calibration intercept (95% CI) | Calibration slope (95% CI) |           |          |          |
| Survival 6 months               | 0.836   | 0.013928    | −0.722 [−1.035, −0.410]        | 0.801 [0.714, 0.888]       |           |          |          |
| Survival 12 months              | 0.846   | 0.03035     | −0.628 [−0.803, −0.453]        | 0.761 [0.698, 0.825]       |           |          |          |
| Survival 5 years [EXPLORATORY]  | 0.983   | 0.039142    | 0.287 [0.214–0.360]            | 0.685 [0.638–0.733]        |           |          |          |
| Survival 10 years [EXPLORATORY] | 0.993   | 0.020913    | 0.557 [0.459–0.655]            | 0.556 [0.525–0.586]        |           |          |          |

**Notes.** (i) Risk stratification performance metrics characterize the model's ability to separate patients who experienced AMI from those who did not, using the predicted probability output as a ranking signal. The 0.50 probability boundary is reported as a conventional reference point — standard in supervised ML evaluation for dichotomous outcomes [\[18\]](#) — and serves to demonstrate that events concentrate in the high-predicted-risk group and are largely absent in the low-predicted-risk group. This is the core property required for population-level prioritization: that patients assigned high probabilities carry substantially higher observed event rates than those assigned low probabilities, as confirmed by the calibration analyses ([Table 2](#)) and external band-level validation ([Table 4](#)). These metrics are not intended as an operational decision threshold; deployment relies on continuous probability outputs stratified into priority bands. (ii) Survival metrics summarise rank discrimination (C-index), prediction error (Brier), and calibration across horizons. (iii) Rows labelled [EXPLORATORY] should not be interpreted as prospective prognostic accuracy; C-indices at 5 and 10 years reflect cumulative risk ordering within a longitudinal cohort and are reported for descriptive purposes only. Time-dependent AUC and Integrated Brier Score at actionable horizons are in [Supplementary Table S5](#).

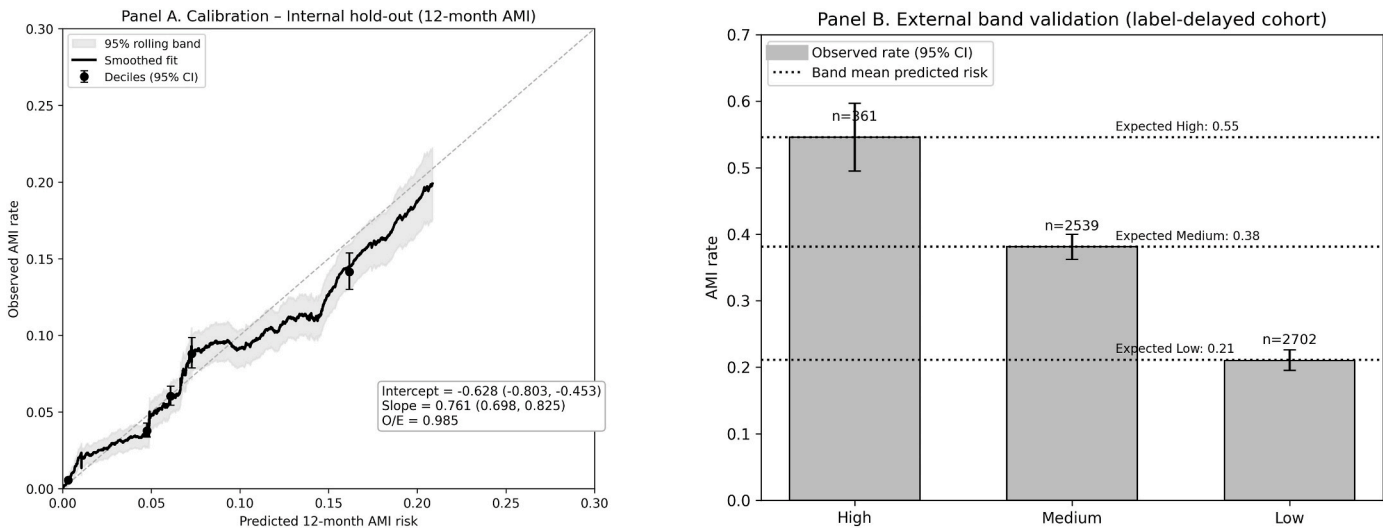


**Fig. 1.** Calibration and external band validation for 6–12-month AMI risk. Panel A shows the internal hold-out calibration plot: the smoothed curve and decile-level confidence intervals closely track the 45° identity line after restricting the axes to clinically relevant risks (0–30%), with annotated intercept, slope, and O/E. Panel B displays the external, label-delayed insurer cohort: observed AMI rates (bars, 95% CIs) and the corresponding band-level mean predicted risks (horizontal dotted lines) for the High, Medium, and Low priority bands.

risk before isotonic recalibration was applied on the validation set. This pattern is expected in high-risk EHR cohorts where baseline hazard is elevated: the Cox component tends to over-estimate absolute risk at short horizons until recalibrated. The aggregate O/E ratio of 0.998 reflects the post-recalibration model. The calibration intercepts in Table 1 are post-isotonic-recalibration values; pre-recalibration intercepts for the survival component were not separately persisted, consistent with standard practice for ensemble models where recalibration is an integral model stage. Supplementary Note 1 provides the full technical account and deployment recommendations. The near-perfect C-indices at 5 and 10 years (0.983 and 0.993) should be interpreted with caution. These values most plausibly reflect cumulative risk ordering within a large longitudinal cohort rather than prospective prognostic accuracy: (a) informative censoring may inflate rank statistics at long horizons; (b) the maximum observed follow-up is 584 days, so 5- and 10-year estimates are extrapolations beyond the observed data range; and (c) visit frequency, correlated with risk, may act as a proxy predictor over long time spans. Time-dependent AUC and Integrated Brier Score — less susceptible to these artifacts — are reported in Supplementary Table S5.

Calibration is the central criterion for evaluating a probabilistic classifier, defined as the agreement between predicted risks and observed event frequencies. Calibration was assessed both in aggregate (the sum of predicted risks approximating observed events) and within subgroups (event rates aligning with average predicted risk after applying a threshold), results are reported in Table 2. On the hold-out test set, aggregate calibration was near perfect: observed AMI events = 4490 vs expected (sum of predicted probabilities) = 4499.3, yielding O/E = 0.998. Class-level calibration at the prespecified 0.50 threshold showed close alignment between observed proportions and average

predicted risks: for Predicted 0 ( $n = 3652$ ), the observed AMI rate was 16.1% (95% CI 14.9–17.3) versus an average predicted risk of 17.4% (Expected = 636.6; Obs–Exp = –49.6); for Predicted 1 ( $n = 5328$ ), the observed AMI rate was 73.3% (95% CI 72.1–74.5) versus an average predicted risk of 72.5% (Expected = 3862.7; Obs–Exp = +40.3). The small, opposite-signed differences and overlap between observed rates and average predicted risks indicate probability coherence at the operational cut-point.

3.4. Concordance with horizon-specific risk strata

Horizon-specific survival priorities demonstrated moderate alignment with legacy risk frameworks. Concordance statistics, cross-tabulations with Framingham categories and continuous AMI-risk tertiles, and correlations with PROCAM estimates are summarized in Table 3. Moderate concordance with Framingham (Cramér’s  $V = 0.326$  at short horizons) is both expected and operationally desirable: the two approaches measure fundamentally different constructs. Framingham estimates 10-year absolute risk based on stable chronic risk-factor burden, whereas the present model captures near-term event probability integrating dynamic clinical signals — biomarker trajectories, comorbidity accrual, and surveillance gaps — invisible to a static scoring equation. A patient classified as moderate 10-year risk by Framingham may carry substantially elevated 6-month AMI probability if glycaemic control has recently deteriorated, renal function has declined, or clinical contact has been interrupted. This is precisely the added operational value of the present model. An illustrative example is in Supplementary Box 1. The two frameworks are therefore complementary: Framingham supports long-horizon prevention planning while the present model enables near-term prioritization of outreach and diagnostics.

As a complementary legacy comparison, we also examined concordance with the PROCAM 10-year coronary risk score in the subset of men aged 35–65 years with complete data. PROCAM\_Risk\_Percent was significantly correlated with the model’s AMI risk estimates, but only with low to moderate magnitude:  $r = 0.24$  ( $n = 44,045$  pairs) for the short-horizon configuration and  $r = 0.34$  ( $n = 53,557$  pairs) for the extended-horizon configuration (both  $p < .001$ ). These results indicate a statistically significant but only partial alignment in patient ranking between PROCAM and the Corpus AMI probabilities.

**Table 2**  
Observed vs. Expected AMI events by predicted class.

| Predicted class            | Patients | Observed AMI, n (%) | Expected AMI, n | Obs – Exp | 95% CI for observed proportion |
|----------------------------|----------|---------------------|-----------------|-----------|--------------------------------|
| Class of low expected AMI  | 3652     | 587 (16.1%)         | 636.6           | –49.6     | 14.9%–17.3%                    |
| Class of high expected AMI | 5328     | 3903 (73.3%)        | 3862.70         | 40.3      | 72.1%–74.5%                    |

**Table 3**

Horizon-specific cross-classification of survival priorities, Framingham risk groups, and continuous AMI-risk tertiles (row-wise %).

| Horizon     | Comparator scheme | Reference stratum         | High  | Medium | Low   |
|-------------|-------------------|---------------------------|-------|--------|-------|
| 6–12 months | Framingham group  | Survival priority: Low    | 75.32 | 19.14  | 5.54  |
| 6–12 months | Framingham group  | Survival priority: Medium | 61.49 | 23.67  | 14.84 |
| 6–12 months | Framingham group  | Survival priority: High   | 26.98 | 26.10  | 46.93 |
| 6–12 months | Survival priority | Risk tertile: Low         | 20.45 | 37.26  | 42.29 |
| 6–12 months | Survival priority | Risk tertile: Medium      | 27.84 | 37.30  | 34.86 |
| 6–12 months | Survival priority | Risk tertile: High        | 41.51 | 35.07  | 23.42 |
| 5–10 years  | Framingham group  | Survival priority: Low    | 71.46 | 19.96  | 8.58  |
| 5–10 years  | Framingham group  | Survival priority: Medium | 59.01 | 23.19  | 17.80 |
| 5–10 years  | Framingham group  | Survival priority: High   | 31.71 | 26.24  | 42.05 |
| 5–10 years  | Survival priority | Risk tertile: Low         | 16.50 | 35.83  | 47.67 |
| 5–10 years  | Survival priority | Risk tertile: Medium      | 34.36 | 33.65  | 31.99 |
| 5–10 years  | Survival priority | Risk tertile: High        | 49.10 | 30.53  | 20.37 |

Notes. Cramér's V: 0.326 (short survival vs. Framingham), 0.148 (short survival vs. continuous-risk tertiles), 0.266 (long survival vs. Framingham), 0.215 (long survival vs. continuous-risk tertiles). N = 179,946 contracts with both measures available. Moderate discordance with Framingham at short horizons is expected given the different risk constructs measured; see text for interpretation and [Supplementary Box 1](#) for an illustrative example.

### 3.5. External, label-delayed validation

The model predicted 1736 AMI events (sum of probabilities), while the observed count was 1,370, yielding O/E = 0.79. This degree of over-prediction may reflect lower baseline AMI incidence in the validation population, delayed administrative coding of events (label-delayed design), or differential access to acute care; it does not necessarily indicate model miscalibration, and recalibration to local event rates is recommended before deployment in any new health system. Priority bands showed monotonic separation of event rates, and ANOVA confirmed highly significant between-group differences ( $F = 147.6$ ;  $p < .001$ ); Tukey HSD indicated that all pairwise contrasts were significant ( $p < .001$ ). Band-level event rates with 95% CIs are reported in [Fig. 1](#) and in [Table 4](#).

## 4. Discussion

Current clinical decision-making remains limited by the imprecision of existing cardiovascular disease CVD risk models, indicating that new

**Table 4**

External label-delayed validation: overall observed vs expected and event rates by priority band.

| Group                | N    | Observed AMI | n (%), 95% CI | p      |
|----------------------|------|--------------|---------------|--------|
| Overall              | 5602 | 1370 (24.5%) | 23.3%–25.7%   | <0.001 |
| High ( $\geq 0.75$ ) | 361  | 197 (54.6%)  | 49.5%–59.7%   | <0.001 |
| Medium (0.25–0.75)   | 2539 | 967 (38.1%)  | 36.2%–40.0%   | <0.001 |
| Low (<0.25)          | 2702 | 567 (21.0%)  | 19.5%–22.6%   | <0.001 |

Notes. (i) Expected AMI for the overall row is the sum of predicted probabilities from the locked classifier; band-level expected counts were not required for the ANOVA/Tukey procedure and are therefore omitted. (ii) 95% CIs are binomial (approximate). (iii) Minor tally differences can arise from timing/definition reconciliation between global and band-specific tallies; conclusions are unchanged (monotonic separation and significant between-group differences).

methods are necessary to facilitate better clinical guidance. For instance, a comparison of the ACC/AHA, Adult Treatment Panel III, and European Society of Cardiology guidelines found that all three models provided poor calibration, underscoring the necessity of improving risk predictions and setting appropriate population-wide thresholds [37]. In this context, the future of CVD prevention is increasingly seen to rely on the adoption of advanced methods like machine learning and digital health technology. Machine learning is capable of generating insights from complex, big data derived from electronic health records, imaging, wearables, and sensors, which can drive a shift toward personalized, precision CVD prevention [38].

In this study, we evaluated a probabilistic model developed to estimate the short-term risk of AMI over 6- and 12-month horizons in a large, real-world population. Our aim was to characterize the model's discriminatory capacity, calibration, and prediction error, and to assess how the resulting prioritization structure aligns with established long-horizon cardiovascular risk classifications. The findings show that near-term AMI risk can be estimated with high accuracy using routinely collected electronic health record data, and that primary 6- and 12-month horizons demonstrate consistent discrimination and calibration. Extended 5- and 10-year horizon analyses are reported as exploratory; the near-perfect C-indices at those horizons most plausibly reflect cumulative risk ordering within a longitudinal cohort — amplified by informative censoring, visit-frequency correlation with risk, and extrapolation beyond the observed 584-day follow-up window — rather than prospective prognostic accuracy. Calendar-time effects and variable completeness of historical look-back diagnoses may contribute additional non-stationarity, reinforcing the need for ongoing monitoring and recalibration. Discrimination was high across components, with an AUC of 0.869 for the classifier and C-indices of 0.836 and 0.846 for the 6- and 12-month survival models, respectively. The calibration slopes (0.76–0.80) confirmed strong risk separability, while the calibration intercepts, after isotonic recalibration, confirmed that aggregate predicted event counts closely matched observed events at the population level (O/E = 0.998). Brier scores indicated low prediction error, and calibration assessments showed close agreement between predicted probabilities and observed event frequencies. Differences between observed and expected events were small in both aggregate and threshold-based subgroup analyses, indicating that the model's output probabilities functioned as reliable estimates of AMI risk.

This model provides clinicians with an event-specific tool capable of estimating AMI risk in a way that is conceptually comparable to traditional long-horizon risk scores. In Colombia, Framingham has demonstrated poor discrimination (AUC 0.65) and consistent overestimation of 10-year risk, while PROCAM—although better calibrated—has shown only modest discrimination (AUC 0.74) and miscalibration in high-risk groups [34]. In contrast, our model delivers materially higher discrimination and stronger calibration, suggesting that it may offer more clinically meaningful estimates of imminent coronary risk. However, because our model was not derived from a general population cohort, further external validation remains necessary. This process is currently underway.

Second, the model may have its greatest utility as a population-health tool. Health systems increasingly need mechanisms to stratify cardiovascular risk within large, heterogeneous populations. By providing calibrated, near-term probabilities rather than broad 10-year composite categories, the model supports more precise prioritization of care pathways, enabling earlier diagnostic evaluation, structured follow-up, and targeted preventive interventions for individuals at highest imminent risk [39]. These advantages are especially relevant in middle- and low-income countries, where diagnostic capacity is often constrained [40]. The probabilistic nature of the model also facilitates clearer risk communication and more informed resource planning, which may strengthen population-level initiatives aimed at reducing avoidable acute events, guiding disease-management outreach, and allocating preventive resources according to anticipated near-term

burden [38].

A growing body of work has explored routinely collected data and advanced analytic methods for cardiovascular risk stratification, with important differences in purpose and time horizon. Population-oriented tools have primarily focused on long-term absolute risk for prevention planning, while post-AMI prognostic models have emphasised longer-term mortality or composite outcomes rather than near-term prioritization [41]. Evaluations of models embedded in clinical guidelines have highlighted that most are designed around long-horizon endpoints and show substantial methodological limitations, including limited calibration assessment across settings [21]. The present study contributes by focusing explicitly on short-term AMI risk as a population-health stratification problem, emphasizing risk separation, calibration, and operational interpretability. Future work should assess the generalizability of these findings in external cohorts and across diverse demographic and socioeconomic settings. Prospective evaluation will be necessary to determine whether integrating short-term AMI predictions into clinical workflows improves care quality, diagnostic yield, or patient outcomes [42,43].

#### 4.1. Operational considerations

Deployment at population scale requires ongoing lifecycle management. Calibration should be monitored using rolling O/E ratios at regular intervals (e.g., quarterly); a drift of  $\pm 20\%$  from the reference O/E should trigger recalibration — updating the isotonic correction layer on recent data without retraining the base model — with full retraining reserved for evidence of distributional shift. Governance responsibilities should be clearly assigned, including an accountable clinical lead and a defined escalation pathway. Probabilistic outputs should be communicated to non-technical stakeholders as priority bands (High, Medium, Low) with observed event rates from validation data, rather than raw probabilities, to support operational uptake without inducing disproportionate alarm.

#### 4.2. Limitations

The study was conducted within a single health system, and performance may differ in populations with different demographic characteristics, care-seeking patterns, or prevalence of cardiovascular disease. As with all EHR-based studies, the model's accuracy depends on the completeness and accuracy of coded data; underdiagnosis, miscoding, and irregular care utilization may introduce misclassification. We did not evaluate potential temporal drift, and model performance may change over time as clinical practices evolve. Missing data in EHR settings may be non-random; median imputation may inadequately capture informative absence, particularly for laboratory values, and could introduce bias. Although missingness indicators were evaluated during model development and partially captured through the Bayesian aggregation layer's surveillance-gap flags, residual bias from informative missingness cannot be excluded. The survival ensemble employs cause-specific censoring, treating non-AMI deaths as censored observations. In older populations with elevated competing mortality, this approach may systematically overestimate AMI risk. Cause-specific modelling was selected to maintain alignment with AMI-specific prioritization objectives, and competing risk extensions using Fine-Gray models are planned for future work. The near-perfect C-indices at 5 and 10 years (0.983 and 0.993) should not be interpreted as prospective prognostic accuracy; they most plausibly reflect artefactual rank separation due to informative censoring and extrapolation beyond the observed follow-up window (maximum 584 days). Recalibration to local event rates is required before deployment in any new health-system context, as demonstrated by the  $O/E = 0.79$  in the external validation cohort.

This study does not differentiate between types of acute myocardial infarction, such as ST-segment elevation myocardial infarction and non-ST-segment elevation myocardial infarction, which may affect risk

stratification in specific populations. Further methodological development may enhance performance. Incorporating longitudinal trajectories, natural-language data from clinical notes, or multimodal inputs such as electrocardiographic features could further refine risk estimation. Finally, research on risk communication is warranted to guide how short-term probabilities can be presented to clinicians and patients in a manner that supports decision-making without inducing alarm or overuse of diagnostic resources.

## 5. Conclusions

In a large outpatient population, a probabilistic model using routinely collected electronic health record data enabled stratification by estimated 6- and 12-month risk of acute myocardial infarction, with consistent discrimination and calibration at the population level. These findings highlight the potential role of short-term, event-specific risk estimation as a complement to traditional long-horizon cardiovascular risk frameworks, particularly for supporting prioritization and planning within health systems. As a population-health stratification tool, this approach may inform the allocation of diagnostic, monitoring, and preventive resources according to anticipated near-term burden. Recalibration to local incidence rates is recommended before deployment in new health-system contexts. Further evaluation in real-world operational settings is needed to assess its potential to reduce avoidable acute events and improve cardiovascular outcomes.

#### Availability of data and materials

Data may be made available upon reasonable request to Juan Esteban Correa-Morales.

#### Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used Calude Sonnet 4.6 in order to check grammar and text readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

#### Funding

This study received no external funding.

#### CRediT authorship contribution statement

**Jose David Amorocho-Morales:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Visualization, Writing – original draft. **Sergio Parra Guevara:** Data curation, Investigation, Methodology, Validation, Writing – review & editing. **Elias Quintero-Muñoz:** Data curation, Investigation, Supervision, Validation, Writing – review & editing. **Guillermo Dimas:** Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing. **Juan Esteban Correa-Morales:** Conceptualization, Formal analysis, Investigation, Methodology, Resources, Validation, Writing – original draft, Writing – review & editing.

#### Declaration of competing interest

José David Amorocho-Morales and Juan Esteban Correa-Morales are employees, co-founders, and shareholders of CorpusAI, the company that developed the algorithm evaluated in this study. The remaining authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

The authors thank Juan Felipe Forero, Santiago Pelaez, Miguel Rojas, Miguel Angel Rojas, Fredy Rodriguez, and Juan Carlos Rodriguez for their invaluable advice and support to this research.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijcrp.2026.200659>.

## References

- [1] B. Chong, J. Jayabaskaran, et al., Global burden of cardiovascular diseases: projections from 2025 to 2050, *Eur. J. Prev. Cardiol.* 32 (11) (2025) 1001–1015, <https://doi.org/10.1093/eurjpc/zwae281>, 25.
- [2] Regional Global, National burden of cardiovascular diseases and risk factors in 204 countries and territories, 1990–2023, *JACC*. Published online Sept. 24 (2025), <https://doi.org/10.1016/j.jacc.2025.08.015>.
- [3] British Heart Foundation, Global Cardiovascular Disease Factsheet, August 2025. Available from: <https://www.bhf.org.uk/-/media/files/for-professionals/research/heart-statistics/bhf-cvd-statistics-global-factsheet.pdf>. (Accessed 5 November 2025).
- [4] J.A. Damen, R. Pajouheshnia, et al., Performance of the framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis, *BMC Med.* 17 (1) (2019) 109, <https://doi.org/10.1186/s12916-019-1340-7>, 13.
- [5] S.V. Rao, M.L. O'Donoghue, M. Ruel, et al., ACC/AHA/ACEP/NAEMSP/SCAI Guideline for the management of patients with acute coronary syndromes: a report of the American college of Cardiology/American heart association joint committee on clinical practice guidelines, *Circulation* 151 (13) (2025) e771–e862, <https://doi.org/10.1161/CIR.0000000000001309>, 2025.
- [6] G.B.J. Mancini, A. Ryomoto, Adoption of the PREVENT (predicting risk of cardiovascular disease EVENTS) risk algorithm: potential international implications, *JACC Adv* 3 (8) (2024) 101122, <https://doi.org/10.1016/j.jaccadv.2024.101122>, 18.
- [7] B. Scheuermann, A. Brown, T. Colburn, H. Hakeem, C.H. Chow, C. Ade, External validation of the American heart association PREVENT cardiovascular disease risk equations, *JAMA Netw. Open* 7 (10) (2024) e2438311, <https://doi.org/10.1001/jamanetworkopen.2024.38311>, 1.
- [8] S. Al-Shamsi, R.D. Govender, External validation of the American heart association's PREVENT equations in predicting atherosclerotic cardiovascular disease risk among Arab women and men: a retrospective cohort study, *BMC Cardiovasc. Disord.* 25 (1) (2025) 732, <https://doi.org/10.1186/s12872-025-05211-8>, 10.
- [9] H. Roxane, M.V. Pedro, V. Julien, External validation of the 2023 American heart association predicting risk of cardiovascular disease EVENTS equations for atherosclerotic cardiovascular disease in primary cardiovascular prevention setting and comparison with 2021 systematic COronary risk evaluation and 2013 pooled Cohort equations, *Eur. J. Prev. Cardiol.* 30 (2025), <https://doi.org/10.1093/eurjpc/zwaf213> zwaf213.
- [10] A. Timmis, V. Aboyans, P. Vardas, et al., ESC national cardiac societies. European society of cardiology: the 2023 atlas of cardiovascular disease statistics, *Eur. Heart J.* 45 (38) (2024) 4019–4062, <https://doi.org/10.1093/eurheartj/ehae466>, 7.
- [11] M. Nomali, D. Khalili, M. Yaseri, M.A. Mansournia, A. Ayati, H. Navid, S. Nedjat, Validity of the models predicting 10-year risk of cardiovascular diseases in Asia: a systematic review and prediction model meta-analysis, *PLoS One* 18 (11) (2023) e0292396, <https://doi.org/10.1371/journal.pone.0292396>, 30.
- [12] WHO CVD Risk Chart Working Group, World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions, *Lancet Glob Health* 7 (10) (2019 Oct) e1332–e1345, [https://doi.org/10.1016/S2214-109X\(19\)30318-3](https://doi.org/10.1016/S2214-109X(19)30318-3). Epub 2019 Sep 2. Erratum in: *Lancet Glob Health*. 2023;11(2): e196. doi: 10.1016/S2214-109X(22)00522-8.
- [13] S.H.J. Hageman, A.J. McKay, et al., Estimation of recurrent atherosclerotic cardiovascular event risk in patients with established cardiovascular disease: the updated SMART2 algorithm, *Eur. Heart J.* 43 (18) (2022) 1715–1727, <https://doi.org/10.1093/eurheartj/ehac056>, 7.
- [14] A. Yilmaz, M.I. Hayiroglu, S. Salturk, L. Pay, A.A. Demircali, C. Coskun, et al., Machine learning approach on high risk treadmill exercise test to predict obstructive coronary artery disease by using P, QRS, and T waves' features, *Curr. Probl. Cardiol.* 48 (2) (2023) 101482, <https://doi.org/10.1016/j.cpcardiol.2022.101482>.
- [15] V. Cicek, A.L. Orhan, F. Saylik, V. Sharma, Y. Tur, A. Erdem, et al., Predicting short-term mortality in patients with acute pulmonary embolism with deep learning, *Circ. J.* 89 (5) (2025) 602–611, <https://doi.org/10.1253/circj.CJ-24-0630>.
- [16] M.I. Hayiroglu, S. Altay, The role of artificial intelligence in coronary artery disease and atrial fibrillation, *Balkan Med. J.* 40 (3) (2023) 151–152, <https://doi.org/10.4274/balkanmedj.galenos.2023.06042023>.
- [17] V. Cicek, M. Babaoglu, F. Saylik, S. Yavuz, A.F. Mazlum, M.S. Genc, et al., A new risk prediction model for the assessment of myocardial injury in elderly patients undergoing non-elective surgery, *J. Cardiovasc. Dev. Dis.* 12 (1) (2024) 6, <https://doi.org/10.3390/jcdd12010006>.
- [18] D.M. Lloyd-Jones, Cardiovascular risk prediction: basic concepts, current status, and future directions, *Circulation* 121 (15) (2010) 1768–1777, <https://doi.org/10.1161/CIRCULATIONAHA.109.849166>, 20.
- [19] A. Pate, R. Emsley, D.M. Ashcroft, B. Brown, T. van Staa, The uncertainty with using risk prediction models for individual decision making: an exemplar cohort study examining the prediction of cardiovascular disease in English primary care, *BMC Med.* 17 (1) (2019) 134, <https://doi.org/10.1186/s12916-019-1368-8>, 17.
- [20] C. Krittanawong, H.U.H. Virk, S. Bangalore, Z. Wang, K.W. Johnson, R. Pinotti, H. Zhang, S. Kaplin, B. Narasimhan, T. Kitai, U. Baber, J.L. Halperin, W.H.W. Tang, Machine learning prediction in cardiovascular diseases: a meta-analysis, *Sci. Rep.* 10 (1) (2020) 16057, <https://doi.org/10.1038/s41598-020-72685-1>, 29.
- [21] C.Y. Jing, L. Zhang, L. Feng, J.C. Li, L.R. Liang, J. Hu, X. Liao, Recommendations for prediction models in clinical practice guidelines for cardiovascular diseases are over-optimistic: a global survey utilizing a systematic literature search, *Front. Cardiovasc. Med.* 17 (11) (2024) 1449058, <https://doi.org/10.3389/fcvm.2024.1449058>.
- [22] G.S. Collins, K.G.M. Moons, et al., TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods, *BMJ* 16 (385) (2024) e078378, <https://doi.org/10.1136/bmj-2023-078378>.
- [23] M.J. Pencina, A.M. Navar, D. Wojdyla, R.J. Sanchez, I. Khan, J. Ellassal, R.B. Sr D'Agostino, E.D. Peterson, A.D. Sniderman, Quantifying importance of major risk factors for coronary heart disease, *Circulation* 139 (13) (2019) 1603–1611, <https://doi.org/10.1161/CIRCULATIONAHA.117.031855>, 26.
- [24] S.H.J. Hageman, C. Petitjaen, et al., Improving 10-year cardiovascular risk prediction in apparently healthy people: flexible addition of risk modifiers on top of SCORE2, *Eur. J. Prev. Cardiol.* 30 (15) (2023) 1705–1714, <https://doi.org/10.1093/eurjpc/zwad187>, 26.
- [25] N. Shakibaie, R. Hassannejad, N. Mohammadifard, H.R. Marateb, M. Mansourian, M.A. Mañanas, N. Sarrafzadegan, Pathways leading to prevention of fatal and non-fatal cardiovascular disease: an interaction model on 15 years population-based cohort study, *Lipids Health Dis.* 19 (1) (2020) 203, <https://doi.org/10.1186/s12944-020-01375-8>, 5.
- [26] J. Zhang, H. Zhang, T. Wei, P. Kang, B. Tang, H. Wang, Predicting angiographic coronary artery disease using machine learning and high-frequency QRS, *BMC Med. Inf. Decis. Making* 24 (1) (2024) 217, <https://doi.org/10.1186/s12911-024-02620-1>, 31.
- [27] E. Bayrakçeken, S. Yarali, U. Ercan, Ö. Alkan, Patterns among factors associated with myocardial infarction: chi-squared automatic interaction detection tree and binary logit model, *BMC Public Health* 25 (1) (2025) 296, <https://doi.org/10.1186/s12889-025-21536-7>, 24.
- [28] P. Lopez-Jaramillo, P. Joseph, J.P. Lopez-Lopez, F. Lanas, A. Avezum, R. Diaz, P. A. Camacho, P. Seron, G. Oliveira, A. Orlandini, S. Rangarajan, S. Islam, S. Yusuf, Risk factors, cardiovascular disease, and mortality in South America: a PURE substudy, *Eur. Heart J.* 43 (30) (2022) 2841–2851, <https://doi.org/10.1093/eurheartj/ehac113>, 7.
- [29] D.B. Dunson, A.H. Herring, Bayesian latent variable models for mixed discrete outcomes, *Biostatistics* 6 (1) (2005) 11–25, <https://doi.org/10.1093/biostatistics/kxh025>.
- [30] K.M. Prenovost, S.D. Fihn, M.L. Maciejewski, K. Nelson, S. Vijan, A.-M. Rosland, Using item response theory with health system data to identify latent groups of patients with multiple health conditions, *PLoS One* 13 (11) (2018) e0206915, <https://doi.org/10.1371/journal.pone.0206915>.
- [31] M.W. Heymans, J.W.R. Twisk, Handling missing data in clinical research, *J. Clin. Epidemiol.* 151 (2022) 185–188, <https://doi.org/10.1016/j.jclinepi.2022.08.016>.
- [32] L. Sasse, E. Nicolaisen-Sobesky, J. Dukart, S.B. Eickhoff, M. Götz, S. Hamdan, et al., Overview of leakage scenarios in supervised machine learning, *J. Big Data* 12 (2025) 135, <https://doi.org/10.1186/s40537-025-01193-8>.
- [33] D.A. Broniatowski, Psychological foundations of explainability and interpretability in artificial intelligence, *Natl. Inst. Stand. Technol. Interag. Intern. Rep.* 8367 (2021) 1–56, <https://doi.org/10.6028/NIST.IR.8367>.
- [34] O.M. Muñoz, N.I. Rodríguez, Á. Ruiz, M. Rondón, Validación de los modelos de predicción de Framingham y PROCAM como estimadores del riesgo cardiovascular en una población colombiana, *Rev. Colomb. Cardiol.* 21 (4) (2014) 202–212.
- [35] E.A. Freeman, G.G. Moisen, A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa, *Ecol. Model.* 217 (1–2) (2008 Aug 24) 48–58, <https://doi.org/10.1016/j.ecolmodel.2008.05.015>.
- [36] O. Rainio, J. Teuhio, R. Klén, Comparison of thresholds for a convolutional neural network classification model used for dental caries detection, *Int. J. Data Sci. Anal.* 19 (2025) 1–12, <https://doi.org/10.1007/s41060-024-00562-6>.
- [37] M. Kavousi, M.J. Leening, et al., Comparison of application of the ACC/AHA guidelines, Adult Treatment Panel III guidelines, and European Society of Cardiology guidelines for cardiovascular disease prevention in a European cohort, *JAMA* 311 (14) (2014) 1416–1423, <https://doi.org/10.1001/jama.2014.2632>, 9.
- [38] A. Javadi, F. Zghyer, C. Kim, E.M. Spaulding, N. Isakadze, J. Ding, D. Kargillis, Y. Gao, F. Rahman, D.E. Brown, S. Saria, S.S. Martin, C.M. Kramer, R. S. Blumenthal, F.A. Marvel, Medicine 2032: the future of cardiovascular disease prevention with machine learning and digital health technology, *Am. J. Prev. Cardiol.* 12 (2022) 100379, <https://doi.org/10.1016/j.ajpc.2022.100379>, 29.
- [39] T. Liu, A. Krentz, L. Lu, V. Curcin, Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis, *Eur Heart J Digit Health* 6 (1) (2024) 7–22, <https://doi.org/10.1093/ehjdh/ztae080>, 27.

- [40] GBD 2017 Cardiovascular Diseases Collaborators, Global, regional, and national burden of cardiovascular diseases, 1990–2017: a systematic analysis for the global Burden of Disease Study 2017, *Lancet*. 396 (10255) (2020) 1257–1280.
- [41] D.G. Manuel, M. Tuna, R. Perez, D. Henry, R.H. Glazier, Development and validation of a cardiovascular disease risk-prediction model using population health surveys, *CMAJ (Can. Med. Assoc. J.)* 190 (29) (2018 Jul 23) E871–E882.
- [42] A. Jeewa, R.M. Muhame, L.L. Guo, A. Wolochacz, K. Jessa, B. Cardiff, et al., Development and prospective evaluation of a machine learning model to predict serious cardiac outcomes among paediatric cardiac inpatients, *BMJ Digit Health* AI 1 (1) (2025 Aug 7) e000114, <https://doi.org/10.1136/bmjdhai-2025-000114>.
- [43] CardioSAVE: a 12-months prospective Longitudinal controlled study for the validation of a novel predictive model and a diagnostic tool, for the risk stratification of AMI patients undergoing PCI [Internet]. ClinicalTrials.gov. Identifier: NCT03038360. [Cited 2025 Nov 15]. Available from: <https://www.clinicaltrials.gov/study/NCT03038360>.