ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Machine learning risk stratification for major adverse cardiovascular events after acute myocardial infarction in patients receiving early empagliflozin therapy

Raisa Trigulova[1†], Alisher Ikramov[2†*], Dilafruz Akhmedova[3], Shokhista Akhmedova[3], and Shakhnoza Mukhtarova[4]

[1]Research Department of Preventive Cardiology, State Institution "Republican Specialized Scientific and Practical Medical Center of Cardiology", Tashkent, Uzbekistan

[2]Department of Mathematics, School of Humanities, Natural and Social Sciences, New Uzbekistan University, Tashkent, Uzbekistan

[3]Laboratory of Preventive Cardiology, State Institution "Republican Specialized Scientific and Practical Medical Center of Cardiology", Tashkent, Uzbekistan

[4]Department of Endocrinology and Pediatric Endocrinology, Faculty of Medical Education and General Medicine, Tashkent State Medical University, Tashkent, Uzbekistan

[†]*These authors contributed equally to this work.*

***Corresponding author:**
Alisher Ikramov
(a.ikramov@newuu.uz)

Citation: Trigulova R, Ikramov A, Akhmedova D, Akhmedova S, Mukhtarova S. Machine learning risk stratification for major adverse cardiovascular events after acute myocardial infarction in patients receiving early empagliflozin therapy. *Artif Intell Health*. doi: 10.36922/AIH026350130

Received: August 27, 2026

Revised: September 11, 2026

Accepted: September 18, 2026

Published online: September 30, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract

Sodium-glucose cotransporter-2 (SGLT2) inhibitors are increasingly prescribed after acute myocardial infarction (AMI), yet residual risk on this therapy is rarely examined alongside the drug's association with outcomes in one cohort. We investigated 311 patients with AMI at one center (January 2024–May 2025): 251 received early empagliflozin 10 mg/day with standard therapy and were followed prospectively for six months; 60 contemporaneous patients without an SGLT2 inhibitor formed a non-randomized control group. The endpoint was six-month major adverse cardiovascular events (MACEs): all-cause death, non-fatal reinfarction, urgent percutaneous coronary intervention, or unplanned cardiovascular hospitalization. Classical and machine learning classifiers were compared under repeated nested cross-validation, with imputation, scaling, and tuning inside each fold; the empagliflozin–MACE association was estimated by propensity-score matching on 20 covariates. MACE occurred in 83 patients (26.7%; 24.7% with empagliflozin vs. 35.0% in controls), including 17 deaths (5.5%). Cross-validated backward elimination reduced 24 candidate predictors to eight: body mass index, prior infarction, current smoking, mineralocorticoid-receptor antagonist use, peak A velocity, left-ventricular end-diastolic volume, empagliflozin exposure, and discretized total cholesterol. On these features, L2-regularized logistic regression reached a pooled out-of-fold area under the curve of 0.673 (95% confidence interval 0.601–0.741) with good calibration (slope 0.88), matching random forest, gradient boosting, XGBoost, support-vector machine, k-nearest neighbors, and a multilayer perceptron. After matching (44 pairs), the odds ratio for MACE with empagliflozin was 0.79 (95% CI 0.32–1.98). Discrimination was moderate and needs external validation; the association between empagliflozin and fewer events was small and hypothesis-generating.

Keywords: Acute myocardial infarction; Post-infarction remodeling; Empagliflozin; Machine learning; Propensity score; Heart-failure phenotype; Major adverse cardiovascular events

1. Introduction

Acute myocardial infarction (AMI) remains the leading precursor of incident heart failure worldwide and a major determinant of long-term prognosis and healthcare resource use, despite substantial progress in reperfusion strategies and secondary prevention.[1,2] The early post-infarction period is increasingly recognized as a window of opportunity during which timely intervention can influence the trajectory of myocardial recovery.

Sodium-glucose cotransporter-2 (SGLT2) inhibitors have become central to the treatment of chronic heart failure and chronic kidney disease, with landmark trials demonstrating reproducible reductions in heart-failure hospitalization and cardiovascular death irrespective of diabetes status: DAPA-HF reduced the composite of cardiovascular death or worsening heart failure by 26% in patients with reduced ejection fraction[3], EMPEROR-Reduced showed a closely concordant 25% reduction with empagliflozin in the same population[4], and DAPA-CKD extended this benefit to progression of chronic kidney disease.[5] This naturally raised the question of whether early initiation after AMI—a setting of acute hemodynamic and neurohormonal stress rather than established chronic heart failure—would confer similar benefit.

Three trials have since addressed this question directly, with results that only partially confirm this expectation. EMMY (n = 476), the first to report, randomized patients within 72 hours of primary percutaneous coronary intervention for large infarcts to empagliflozin or placebo and found a significant 15% relative reduction in N-terminal pro-B-type natriuretic peptide (NT-proBNP) over 26 weeks, together with improvements in left-ventricular ejection fraction (LVEF), diastolic function, and echocardiographic remodeling parameters[6]; the trial, however, was not powered for clinical events. DAPA-MI (n = 4,017), powered for clinical endpoints, subsequently found no significant benefit of early dapagliflozin on the composite of cardiovascular death or heart-failure hospitalization in a lower-risk, largely nondiabetic population without known heart failure, despite confirming an acceptable safety profile.[7] EMPACT-MI (n = 6,522), enrolling a higher-risk population with LVEF < 45% and/or signs of congestion within 14 days of infarction, similarly found no significant difference in all-cause mortality between empagliflozin and placebo, although a reduction was observed in the secondary composite of first hospitalization for heart failure.[8] Mechanistic and imaging sub-studies have added further detail: EMPRESS-MI found no incremental effect of empagliflozin on indexed left-ventricular end-systolic volume (LVESV) beyond that achieved by background therapy alone[9], whereas EMPA-TROPISM and SUGAR-DM-HF, conducted in patients with established reduced ejection fraction rather than the acute post-infarction setting, did demonstrate favorable effects of SGLT2 inhibition on ventricular volumes.[10,11] Taken together, this body of evidence has shaped the current view of early post-infarction SGLT2 inhibition as a selective, phenotype-dependent strategy rather than a routine one for all comers.[12]

Randomized trials, by design, answer the question of average treatment effect; they are not designed to characterize what determines residual risk in an individual patient who is already receiving such therapy. Yet this is precisely the question a clinician faces at hospital discharge, and machine learning (ML) has increasingly been applied across cardiovascular medicine.

The performance of ML models for cardiovascular risk prediction, however, varies substantially by clinical outcome, algorithm, and dataset, and no single algorithm is uniformly best. In an empirical comparison of six algorithms on the University of California, Irvine cardiovascular disease dataset (76 candidate attributes), a hybrid artificial neural network–genetic algorithm model achieved the highest classification accuracy (86.4%), ahead of a support-vector machine (84.0%), k-nearest neighbors (79.0%), decision tree (77.8%), and K-means clustering (59.6%), illustrating both the potential gains from hybrid architectures and the wide performance spread across otherwise standard algorithms applied to the same data.[13] A network meta-analysis pooling 17 studies and 285,213 patients similarly found that the best-performing algorithm differed by outcome: Deep-learning models achieved an area under the receiver-operating-characteristic curve (AUC) of 0.843 (95% confidence interval [CI] 0.840–0.845) for heart-failure prediction, gradient boosting machines reached 91.1% average accuracy for the same outcome, support-vector machines performed best for stroke prediction, and random forests performed best for hypertension —showing that algorithm selection should be empirical and outcome-specific rather than assumed a priori.[14] For outcomes recorded as time-to-event rather than a simple binary label, a systematic review of 33 studies found that random survival forests, survival gradient-boosting models, and penalized Cox regression were the most frequently and successfully applied approaches, reflecting the additional structure such outcomes require relative to standard classification.[15] Reviews of artificial intelligence (AI) across cardiovascular medicine reach a similar conclusion: performance is strongest for image- and signal-based diagnosis and far less settled for prognostic and therapeutic decision support, where small samples and heterogeneous outcomes dominate.[16]

Closer to the present clinical question, a 2025 systematic review and meta-analysis of ML models specifically for predicting major adverse cardiovascular events (MACEs) after myocardial infarction, pooling 28 studies and 59,392 patients, reported a pooled validation-set C-index of 0.77 (95% CI 0.74–0.81)[17]—a useful external benchmark against which the discrimination achieved in the present cohort (**Section 3.4**) can be judged. Beyond structured tabular prediction, AI is also being applied to arrhythmia and electrocardiogram-based diagnostics in this population[18], and, more broadly, to prediction tasks spanning general cardiovascular disease risk[13]; across this literature, data on the phenotypic structure of post-infarction remodeling and on individualized prognosis specifically in patients already receiving early SGLT2 inhibitor therapy remain limited. Rigorous, transparent reporting of prognostic models of this kind—as codified in the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis and Artificial Intelligence (TRIPOD+AI) statement—is essential if this class of model is to move from exploratory analysis toward clinical utility.[19]

This study therefore has two separate aims, which are analyzed and reported separately throughout. The first is to develop and internally validate models for six-month MACE in patients with AMI treated with early empagliflozin, comparing regularized logistic regression with ML classifiers under one cross-validated protocol and quantifying discrimination, calibration, and optimism. The secondary aim was to estimate, in a non-randomized comparison with contemporaneous patients treated without an SGLT2 inhibitor, the association between early empagliflozin and six-month MACE after propensity-score adjustment, and to describe how six-month remodeling and outcome differed by LVEF phenotype.

2. Data and methods

2.1. Design and participants

This was a single-center cohort study conducted at the Republican Specialized Scientific and Practical Medical Center of Cardiology (Tashkent, Uzbekistan) between January 2024 and May 2025 (Figure 1). Consecutive patients aged 35–75 years with confirmed AMI treated by primary percutaneous coronary intervention were eligible. Exclusion criteria were chronic heart failure pre-dating the index infarction, an estimated glomerular filtration rate below 30 mL/min/1.73 m^2, insulin therapy before admission, chronic rheumatic heart disease, acute inflammatory or infectious disease, and a history of malignancy. Of 324 patients assessed, 13 were excluded because of severely reduced LVEF, leaving 311 in the analysis. In 251 patients, an SGLT2 inhibitor (empagliflozin 10 mg/day in 249, dapagliflozin 10 mg/day in 2) was added to standard therapy on days 3–5 of the index admission; these patients were enrolled prospectively and assessed at baseline (visit 1, days 1–3) and at six months (visit 2). Prescription was at the discretion of the treating physician and was neither randomized nor governed by an institutional protocol; physicians relied on the EMMY trial[6] and on the 2023 European Society of Cardiology (ESC) guidelines for acute coronary syndromes.[2] The control group comprised 60 contemporaneous patients from the same center who met the same eligibility criteria, were not prescribed an SGLT2 inhibitor during the index admission or follow-up, and for whom baseline and six-month data (echocardiography, Holter monitoring, laboratory tests) were available in the hospital records; their data were extracted retrospectively, and outcomes were verified by telephone contact. Because controls were selected on the availability of six-month data, and because the decision to prescribe empagliflozin was clinical, the two groups differ in several baseline characteristics and in background therapy; these differences are addressed by propensity-score methods (**Section 2.5**) and discussed as limitations (**Section 4.5**). Throughout, "empagliflozin exposure" denotes the 251 patients treated with an early SGLT2 inhibitor; the two patients treated with dapagliflozin were retained in this group without a separate analysis.

2.2. Clinical and instrumental assessment

Echocardiography was performed at both visits according to a standard protocol. LVEF, LVESV, and left-ventricular end-diastolic volume (LVEDV) were assessed using the biplane method of disks (modified Simpson's rule) in the apical four- and two-chamber views, as recommended by the American Society of Echocardiography/European Association of Cardiovascular Imaging (ASE/EACVI)[20]; linear dimensions (left-ventricular [LV] end-diastolic dimension, LV end-systolic dimension) and LV mass were also recorded. LV mass and left-atrial volume were also indexed to body-surface area. Diastolic function was assessed according to the 2016 ASE/EACVI algorithm[21]: Peak early (E) and late (A) transmitral velocities and their ratio, septal and lateral mitral annular e', E/e', tricuspid regurgitation velocity and gradient, and pulmonary artery systolic pressure. Heart-failure phenotype was defined from baseline LVEF according to the 2021 ESC classification[22] as reduced (<40%), mildly reduced (40%–49%), or preserved (≥50%) LVEF. Killip class and other clinical markers of heart failure at admission were not recorded in the electronic dataset, so heart-failure phenotype was characterized by LVEF category together with NT-proBNP. We performed 24-hour Holter electrocardiogram monitoring and calculated time- and frequency-domain heart-rate

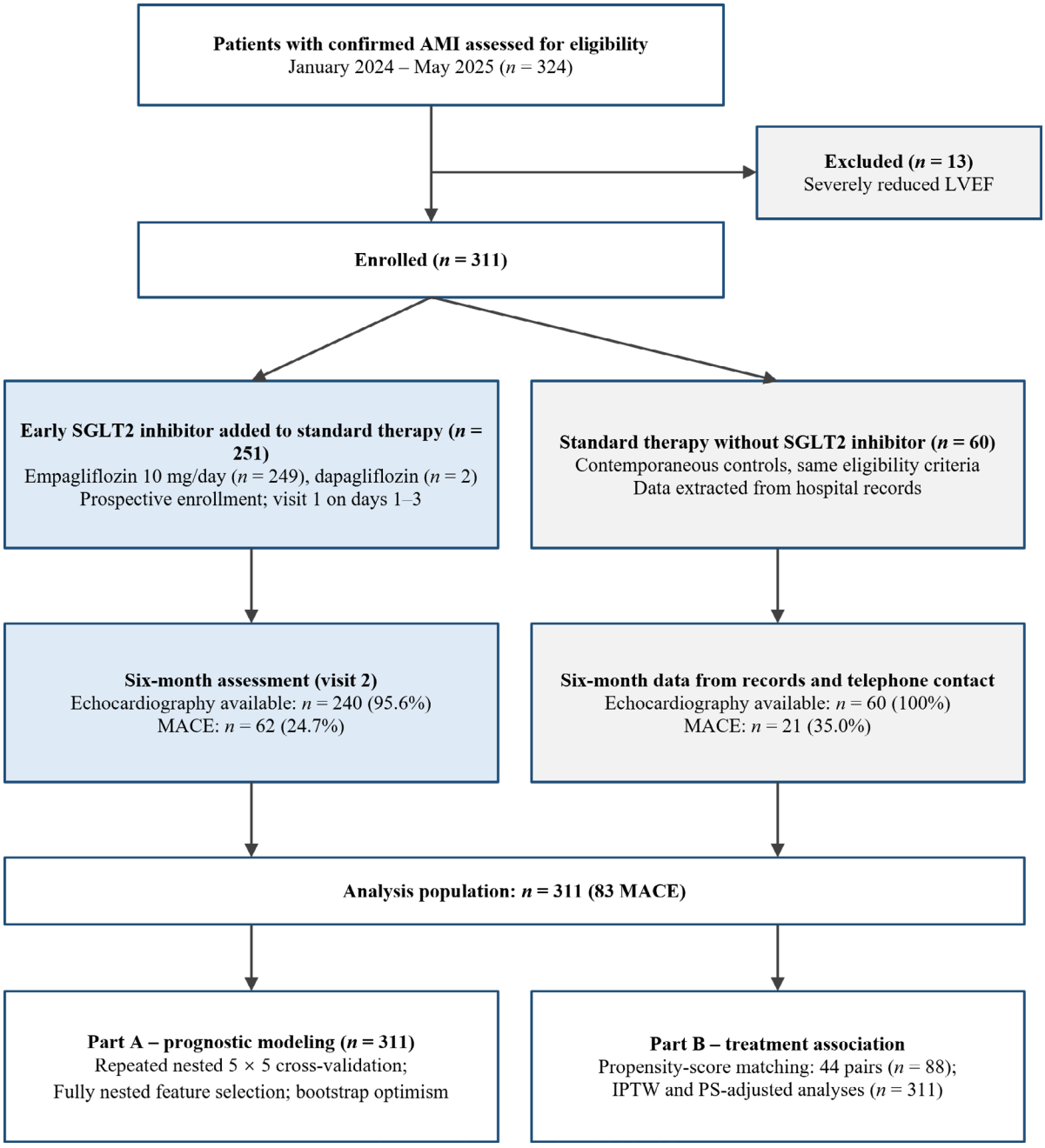


Figure 1. Study flow chart

AMI: Acute myocardial infarction; IPTW: Inverse-probability-of-treatment weighting; LVEF: Left-ventricular ejection fraction; MACE: Major adverse cardiovascular events; PS: Propensity score; SGLT2: Sodium-glucose cotransporter-2.

variability indices. Laboratory assessments included creatinine with estimated glomerular filtration rate (eGFR; Chronic Kidney Disease Epidemiology Collaboration equation), NT-proBNP, glycated hemoglobin (HbA1c), lipid profile, fibrinogen, and C-reactive protein (CRP).

2.3. Endpoints

The primary endpoint was a composite of MACEs at six months, defined as the first occurrence of death from any cause, non-fatal reinfarction, urgent (unplanned) percutaneous coronary intervention, or unplanned hospitalization for unstable angina or heart failure. This

composite corresponds to the extended ("MACE-plus") definition used in regulatory guidance on cardiovascular outcome assessment, in which the core components (death, non-fatal infarction, stroke) are supplemented by hospitalization for acute coronary syndrome, urgent revascularization, and heart failure.[23] Two deliberate deviations from the classical three-point definition were made. All-cause rather than cardiovascular death was used because, over the first six months after infarction, the cause of death could not be adjudicated reliably from the available documentation; this choice is conservative. Unplanned hospitalization was included because, in a contemporary reperfused population, early residual risk is expressed mainly through recurrent ischemia and decompensation leading to readmission rather than through death. Planned, staged admissions for elective percutaneous coronary intervention were part of the revascularization strategy and were not counted as events, and no stroke occurred during follow-up. For patients with more than one event, the first event was used; event timing was recorded to the nearest month. Component events were recorded prospectively in the empagliflozin group. For the retrospectively assembled control group, death and rehospitalization could be verified from records and telephone contact, whereas reinfarction and urgent revascularization could not be separated reliably from the composite; component counts are therefore reported in full for the empagliflozin group only (**Section 3.8**).

2.4. Statistical analysis—Part A: Prognostic model development and internal validation

Continuous variables are presented as mean ± standard deviation with interquartile range; between-group comparisons used the Kruskal–Wallis test. Categorical variables are presented as percentages and compared with Fisher's exact test. All statistical tests were two-sided; $p < 0.05$ was considered significant. Baseline imbalance between treatment groups was also quantified by standardized mean differences (SMD), with $|\text{SMD}| > 0.10$ taken as meaningful.[24]

Univariable logistic regression was used to screen 84 candidate predictors of MACE (51 raw clinical, laboratory, echocardiographic, and treatment variables, 32 discretized deviation scores described below, and empagliflozin exposure; retention threshold $p < 0.20$). Because many continuous measurements in this cohort lack a population-specific reference range—every participant, by definition, had just had a myocardial infarction—each continuous variable with an established clinical normal range was additionally discretized into a three-level ordinal deviation score: 0 if the value fell within the normal interval; 1 if it fell outside the interval but within three interval-widths of the nearest boundary; and 2 if the deviation exceeded three interval-widths. Sex-specific reference ranges were applied where clinically appropriate (creatinine, LVEF, LV volumes and dimensions, LV mass). Both raw and discretized versions of each variable were offered to the feature-selection procedure.

Starting from the predictors retained by univariable screening, a greedy backward-elimination procedure was applied with L2-regularized logistic regression as the base learner: at each step, every remaining feature was tentatively removed, the AUC of the reduced model was estimated by repeated stratified five-fold cross-validation (five repeats), and the feature whose removal produced the highest cross-validated AUC was dropped; the procedure stopped when no removal improved cross-validated AUC. Cross-validated rather than in-sample AUC was used to limit the optimism that in-sample selection produces in datasets of this size.[25]

Classification algorithms ($n = 10$) were compared under an identical protocol: Logistic regression with L2 and with L1 penalties, random forest[26], extra trees, gradient boosting, XGBoost[27], Gaussian naïve Bayes, support-vector machine with radial basis kernel, k-nearest neighbors, and a multilayer perceptron. The whole pipeline—imputation (median for continuous, mode for binary and ordinal variables), standardization, and the classifier—was fitted only on the training portion of each fold, so that no information from validation folds entered model fitting. Hyperparameters were tuned by grid search in an inner stratified five-fold loop; performance was estimated in an outer repeated stratified five-fold loop with five repeats (25 outer folds of 62–63 patients and 16–17 events each). For every algorithm, we reported the mean ± standard deviation of AUC, area under the precision–recall curve (AUPRC), and Brier score across the 25 outer folds, the 2.5th–97.5th percentiles of the fold AUCs and, as an indication of the best realization of each algorithm, the AUC of the best model, defined as the fitted model from the outer fold with the highest test AUC. The best-model value is an optimistic upper bound and is not used as the estimate of expected performance. Model comparison was performed on three feature sets: (i) all 84 candidates, (ii) the 24 predictors retained by univariable screening, and (iii) the eight predictors retained by backward elimination. Because screening and backward elimination were performed on the full cohort before this comparison, the comparisons carry the optimism of feature selection[25]; to quantify it, the entire pipeline (screening, backward elimination, and tuning) was also repeated inside each of the 25 outer folds, and the resulting fully nested estimate is reported alongside (**Section 3.5**). No separate hold-out set

was used; all estimates come from cross-validation.

Calibration of the final model was assessed on out-of-fold predictions averaged over the five repeats: a calibration plot with octiles of predicted risk and a loess smoother (Figure 2), the calibration slope (logistic recalibration), and the calibration-in-the-large intercept.[28] The 95% CIs for the pooled out-of-fold AUC, AUPRC, Brier score, slope, and intercept were obtained from 2,000 bootstrap resamples of patients. Optimism of the final eight-feature model was also estimated by Harrell's bootstrap (1,000 resamples; apparent minus test performance).[29] Sample-size adequacy was judged against the events-per-variable ratio and the criteria of Riley *et al.*[30]

Feature contributions to the final model were characterized using SHapley Additive exPlanations (SHAP) (Figure 3).[31] Spearman rank correlations were calculated between each candidate feature and the MACE indicator, with Benjamini–Hochberg correction for multiple comparisons across the full feature set.

Analyses were performed in Python (version 3.x, Python Software Foundation, Netherlands; scikit-learn, XGBoost, statsmodels, SHAP) and R (version 4.x, R Foundation for Statistical Computing, Austria; tableone, MatchIt, glmnet, rms). Reporting follows the TRIPOD+AI statement[19], its predecessor TRIPOD[32], and the development-and-validation framework of Steyerberg and Vergouwe[33]; a condensed TRIPOD+AI checklist with the location of each item is given in Table 1. The analysis code (notebooks and scripts) and the de-identified dataset are available from the corresponding author on reasonable request.

2.5. Statistical analysis—Part B: Association between empagliflozin and major adverse cardiovascular event

The association between empagliflozin exposure and six-month MACE was analyzed separately from the prognostic models; outcomes by treatment group and LVEF phenotype are tabulated in **Section 3.8**. A propensity score for receiving empagliflozin was estimated by logistic regression on 20 baseline covariates chosen a priori as predictors of treatment assignment and/or outcome: Age, sex, body mass index (BMI), Q-wave infarction, prior myocardial infarction, prior percutaneous coronary intervention, type 2 diabetes, hypertension, chronic kidney disease, chronic obstructive pulmonary disease, current smoking, heart rate, systolic blood pressure, eGFR, HbA1c, low-density lipoprotein (LDL)-cholesterol, CRP, log-transformed NT-proBNP, and LVEF phenotype (reduced and mildly reduced vs. preserved). Overlap was examined from the distribution

Table 1. Condensed TRIPOD+AI checklist and location of each item in the paper

TRIPOD+AI item (condensed)	Where reported
Title and abstract: Development and internal validation of a prediction model; population, outcome, and summary of performance	Title; Abstract
Background, objectives, and intended use; the two research questions stated separately	**Section 1**
Source of data, setting, enrollment and follow-up dates, eligibility, participant flow	**Sections 2.1, 3.1; Figure 1**
Outcome definition, timing, and assessment	**Section 2.3**
Predictors: Definition, measurement, transformations (deviation scores)	**Sections 2.2, 2.4; Table 3**
Sample size and events per variable	**Sections 2.4, 4.4, 4.5**
Handling of missing data (imputation inside each fold)	**Section 2.4**
Model building: Candidates, screening, selection, algorithms, hyperparameter tuning	**Section 2.4**
Internal validation: Repeated nested cross-validation, fully nested selection, bootstrap optimism	**Sections 2.4, 3.4, 3.5**
Performance with confidence intervals: Discrimination (AUC, AUPRC), calibration (slope, intercept, plot), overall (Brier)	**Sections 3.4, 3.5; Tables 4–6; Figure 2**
Model specification and explainability	**Section 3.6; Figure 3**
Participant characteristics and outcome frequency	**Sections 3.1, 3.2, 3.8; Tables 1, 8**
Fairness and subgroup performance	Not assessed (**Section 4.5**)
Limitations, interpretation, implications for practice and validation	**Sections 4.4–4.6**
Ethics, funding, competing interests	Declarations
Availability of data and code	Declarations (Availability of data)

Abbreviations: AUC: Area under the receiver-operating-characteristic curve; AUPRC: Area under the precision–recall curve; TRIPOD+AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis and Artificial Intelligence.

of the propensity score in the two groups and from the region of common support. Controls were matched 1:1 to treated patients by nearest-neighbor matching on the logit of the propensity score without replacement, with a caliper of 0.2 pooled standard deviations[24]; balance after matching was re-assessed using SMD (Figure 4). The treatment effect in the matched sample was estimated as an odds ratio from logistic regression with cluster-robust standard errors for matched pairs and confirmed by McNemar's test on discordant pairs. Two sensitivity analyses used the whole cohort: inverse-probability-of-treatment weighting with average-treatment-effect-in-the-treated weights and robust standard errors, and logistic regression of MACE on treatment adjusted for the logit propensity score. Effect modification by LVEF phenotype was explored by adding a treatment × phenotype interaction term to the propensity-adjusted model. Finally, the propensity model was re-estimated with six baseline drug classes added (angiotensin-converting enzyme [ACE] inhibitor, mineralocorticoid-receptor antagonist, metformin, beta-blocker, clopidogrel, loop diuretic) to examine whether background therapy could be balanced.

2.6. Ethical considerations

The study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of the Republican Specialized Scientific and Practical Medical Center of Cardiology. All participants provided written informed consent.

3. Results

3.1. Participant flow and baseline characteristics

Of 324 patients assessed, 311 were included (Figure 1): 251 treated with early empagliflozin and 60 contemporaneous controls. Six-month echocardiography was available in 240 (95.6%) and 60 (100%) patients, respectively, and vital status and MACE were ascertained in all 311. The mean age of the cohort was 62.20 ± 10.50 years (median 62, interquartile range 55–70), consistent with the typical age profile of patients with AMI. The two groups were comparable in age (62.41 ± 10.62 years on empagliflozin vs. 61.28 ± 9.90 years in controls; $p = 0.390$). Women comprised 27.3% ($n = 85$) of the cohort, with no significant between-group difference (29.1% vs. 20.0%; $p = 0.197$), although the modest size of the control group ($n = 60$) limits power to detect such a difference. Mean BMI for the cohort was 29.16 ± 5.03 kg/m^2, placing the sample at the upper margin of the overweight range, close to the threshold for obesity. BMI was the only demographic/anthropometric variable that differed significantly between groups: patients receiving empagliflozin had higher BMI (29.56 ± 4.79 vs. 27.46 ± 5.61 kg/m^2; $p = 0.004$), an absolute difference of 2.1 kg/m^2. This pattern is consistent with confounding by indication, as SGLT2 inhibitors are preferentially prescribed to patients with a more pronounced cardiometabolic phenotype in routine practice. BMI was therefore included in the propensity model (**Section 2.5**).

Half of the cohort had sustained a Q-wave infarction (51.8%, $n = 161$), with no between-group difference (52.2% vs. 50.0%; $p = 0.776$). A substantial proportion of patients had an adverse coronary history: prior myocardial infarction was present in 26.4% ($n = 82$) and prior percutaneous coronary intervention in 18.3% ($n = 57$), both comparable between groups ($p = 0.627$ and $p = 0.853$, respectively)—approximately one in four patients presented with a recurrent, rather than first, infarction, characterizing this as a high-baseline-risk population. Type 2 diabetes was present in 42.1% (43.8% vs. 35.0%; $p = 0.245$) and hypertension in 49.8% (50.6% vs. 46.7%; $p = 0.667$), without significant between-group differences; chronic kidney disease was diagnosed in 14.5% (13.1% vs. 20.0%; $p = 0.219$). Two variables differed significantly: current smoking was more prevalent in the control group (33.3% vs. 19.5%; odds ratio 2.06; $p = 0.025$), as was chronic obstructive pulmonary disease (10.0% vs. 3.6%; odds ratio 2.99; $p = 0.048$)—a biologically consistent pair of findings, given smoking as the principal risk factor for the latter. Both differences point toward a less favorable respiratory profile in the control group and warrant consideration when interpreting variables sensitive to pulmonary pathology, particularly pulmonary artery pressure and right-heart parameters.

Mean heart rate was 80.09 ± 15.78 bpm, systolic blood pressure 131.97 ± 20.08 mmHg, and diastolic blood pressure 81.82 ± 11.18 mmHg, consistent with assessment during the acute post-infarction period after hemodynamic stabilization. All three variables were significantly higher in the control group (heart rate 83.28 vs. 79.32 bpm, $p = 0.023$; systolic blood pressure 136.67 vs. 130.85 mmHg, $p = 0.024$; diastolic blood pressure 85.00 vs. 81.06 mmHg, $p = 0.025$). Although each absolute difference is modest, the consistent direction across all three parameters suggests a systematic rather than incidental pattern, most plausibly explained by differences in background therapy: ACE inhibitors were used far more frequently in the control group (86.7% vs. 41.8%), suggesting these agents were preferentially prescribed in response to higher blood pressure rather than as part of a uniform protocol; the higher heart rate in controls is not explained by beta-blocker use, which was, if anything, more frequent in that group (100% vs. 91.5%).

Fasting glucose (7.25 ± 2.85 mmol/L), postprandial glucose (9.34 ± 3.95 mmol/L), and HbA1c (6.82 ± 1.86%)

reflected moderate dysglycemia consistent with the 42.1% prevalence of diabetes. Groups were comparable for fasting glucose (p = 0.183) and HbA1c (p = 0.717); only postprandial glucose differed significantly (9.19 vs. 9.97 mmol/L; p = 0.005), which, given comparable HbA1c and fasting glucose, more likely reflects distributional asymmetry than a genuine difference in overall glycemic control. This comparability of integrated glycemic markers makes baseline glycemic differences an unlikely explanation for the remodeling patterns described in **Section 3.2**; it does not, by itself, establish a non-glycemic mechanism of empagliflozin's effect.

The lipid profile combined hypercholesterolemia, low high-density lipoprotein (HDL)-C, and hypertriglyceridemia—a pattern of atherogenic dyslipidemia typical of the cardiometabolic phenotype (total cholesterol 196.91 ± 51.65 mg/dL [5.09 mmol/L]; LDL-C 117.46 ± 43.67 mg/dL [3.04 mmol/L]; HDL-C 41.59 ± 10.82 mg/dL [1.08 mmol/L]; triglycerides 189.19 ± 131.44 mg/dL [2.14 mmol/L]). Because every patient had sustained an infarction, all fall into the very-high-cardiovascular-risk category with an LDL-C target < 55 mg/dL, unchanged in the 2025 focused update of the ESC/European Atherosclerosis Society (EAS) dyslipidemia guidelines[34]; the cohort mean exceeds this target more than two-fold, and the lower quartile (84 mg/dL) indicates that at least three-quarters of patients remained above even the 70 mg/dL threshold, pointing to substantial unrealized potential for lipid-lowering intensification, particularly given the 2025 recommendation to initiate combination therapy already during the index hospitalization. Groups were comparable for total cholesterol (p = 0.596), LDL-C (117.06 vs. 119.12 mg/dL; p = 0.675)—an important condition for unbiased interpretation of subsequent outcome comparisons—and triglycerides (p = 0.845). Only HDL-C differed significantly (43.70 vs. 41.09 mg/dL; p = 0.019), an absolute difference of 2.6 mg/dL within the range of biological variability, likely reflecting the previously noted BMI difference between groups.

Mean creatinine was 106.16 ± 31.55 µmol/L and eGFR 69.13 ± 21.28 mL/min/1.73m^2; both differed significantly between groups, with higher creatinine (112.35 vs. 104.68 µmol/L; p = 0.002) and lower eGFR (60.93 vs. 71.09 mL/min/1.73m^2; p = 0.001) in controls. This 10.2 mL/min difference is clinically meaningful—mean eGFR in the control group sits at the boundary of chronic kidney disease stage 3, versus stage 2 in the empagliflozin group—and most plausibly reflects prescribing practice: physicians appear to have been more reluctant to initiate empagliflozin in patients with reduced eGFR, even though current labeling does not require this caution. Renal function was therefore included in the propensity model.

Mean LVEF was 49.23 ± 9.04%, placing the cohort within the moderately reduced range of systolic function overall. LVEF was significantly higher in the empagliflozin group (49.75% vs. 47.04%; p = 0.013), though the 2.7-point absolute difference does not exceed the reproducibility of serial echocardiographic measurement. Volumetric and linear LV parameters were fully comparable between groups (LVEDV 108.14 vs. 103.01 mL, p = 0.677; LVESV 56.56 vs. 54.81 mL, p = 0.432; LV end-diastolic dimension 5.09 vs. 5.06 cm, p = 0.959; LV end-systolic dimension 3.71 cm in both groups, p = 0.898). LV mass differed at trend level (206.05 vs. 216.09 g; p = 0.073), but after indexing to body-surface area the difference reached significance (LV mass index 104.25 vs. 115.86 g/m^2; p = 0.003)—left-ventricular hypertrophy was more pronounced in the control group, with the indexed value exceeding the upper limit of normal for both sexes; this divergence between absolute and indexed mass is explained by the between-group BMI difference (a smaller body-surface area yields a higher index for the same absolute mass). By LVEF phenotype, 37 (14.7%), 74 (29.5%), and 140 (55.8%) patients in the empagliflozin group and 13 (21.7%), 23 (38.3%), and 24 (40.0%) controls had reduced, mildly reduced, and preserved LVEF, respectively (p = 0.084), and NT-proBNP was five-fold higher in the empagliflozin group (mean 3,056 vs. 605 pg/mL; p < 0.001), so that the treated group combined a somewhat better LVEF distribution with markedly stronger neurohormonal activation.

Mean transmitral flow velocities were E 68.71 ± 20.28 cm/s and A 75.01 ± 21.12 cm/s, with an E/A ratio of 1.01 ± 0.51—a near-unity ratio combined with globally reduced tissue velocities, consistent with impaired relaxation masked by pseudonormalization in a subset of patients, a typical post-infarction pattern. Tissue velocities were the most informative parameters and were significantly higher in the control group across all three sites (septal e' 7.34 vs. 6.40 cm/s; lateral e' 10.08 vs. 8.47 cm/s; average e' 8.72 vs. 7.43 cm/s; all p < 0.001, with the largest test statistics in the entire baseline table). In the empagliflozin group, mean septal and lateral e' fell below the lower limit of normal (7 and 10 cm/s, respectively), indicating impaired LV relaxation. E/e' did not differ significantly (9.60 vs. 8.30; p = 0.173), nor did diastolic dysfunction grade (1.20 vs. 1.02; p = 0.160). The E/A difference reached nominal significance (1.00 vs. 1.06; p = 0.004) but is of negligible clinical magnitude. Lower tissue velocities and higher NT-proBNP in the empagliflozin group point in the same direction—toward more pronounced diastolic dysfunction and neurohormonal activation—reinforcing the impression that the empagliflozin group was, if anything, the sicker of the two at baseline.

Mean left-atrial volume was 54.16 ± 15.88 mL and left-atrial volume index (LAVI) 27.93 ± 8.80 mL/m^2; the cohort-wide LAVI did not reach the threshold for dilation (34 mL/m^2) but was significantly higher in the control group (29.94 vs. 27.45 mL/m^2; p = 0.002); non-indexed volume differed at trend level (p = 0.054), while antero–posterior diameter did not differ (p = 0.638)—again attributable to the BMI difference between groups. Tricuspid regurgitation velocity was similar between groups (2.47 vs. 2.45 m/s; p = 0.617), while its gradient was higher in the empagliflozin group (27.07 vs. 24.60 mmHg; p = 0.011). Full baseline characteristics are presented in Table 2.

3.2. Six-month outcomes and remodeling

Over six months of follow-up, 17 patients (5.5%) died, 74 (23.8%) were rehospitalized, and 83 (26.7%) experienced the composite MACE endpoint. MACE occurred in 62 of 251 patients (24.7%) in the empagliflozin group and in 21 of 60 controls (35.0%) (unadjusted odds ratio 0.61, 95% CI 0.33–1.11; p = 0.108); outcomes by treatment group and LVEF phenotype are analyzed in Part B (**Section 3.8**). In the prospectively followed empagliflozin group, LVEF improved in 37.5% of patients with mildly reduced (40%–49%) and 27.8% of patients with reduced (< 40%) baseline LVEF, and the improvement rate did not differ by glycemic control: 16.3% vs. 17.0% for HbA1c ≤ 8% vs. > 8% (p = 0.929). These remodeling data are observational and do not by themselves identify the mechanism of SGLT2 inhibitor action (**Section 4.2**).

Ventricular reverse remodeling occurred substantially faster than atrial remodeling. Large Spearman effect sizes were observed for the six-month change in LV end-systolic and end-diastolic volumes (r = 0.82–0.83), while indexed left-atrial volume showed essentially no change in any phenotype over the same interval (r = 0.12–0.19).

3.3. Prognostic modeling of six-month MACE: Univariable predictor screening

A univariable logistic regression screen for six-month MACE was performed across all 84 candidate predictors, including the discretized deviation-from-normal versions of continuous laboratory and echocardiographic variables. At the pre-specified retention threshold of p < 0.20, 24 predictors were carried forward to feature selection (Table 3). The strongest univariable associations were with lower body mass index, lower peak A velocity, larger LVEDV (both as raw volume and as deviation score), current smoking, larger LVESV, prior myocardial infarction, higher E/A ratio, and lower LVEF. Empagliflozin exposure was associated with lower odds of MACE (odds ratio 0.61, 95% CI 0.33–1.11; p = 0.107) and mineralocorticoid-receptor antagonist use with higher odds (odds ratio 1.57, 95% CI 0.89–2.75; p = 0.120). NT-proBNP (log-transformed), eGFR, age, sex, and diabetes did not pass the screening threshold.

3.4. Feature selection and model comparison

Cross-validated backward elimination with logistic regression reduced the 24 screened predictors to eight: body mass index, prior myocardial infarction, current smoking, mineralocorticoid-receptor antagonist use, peak A velocity, LVEDV, empagliflozin exposure, and discretized total cholesterol. Discretized total cholesterol was retained although the raw value had not passed univariable screening (p = 0.25), the one instance in which the deviation-score representation added information beyond the raw measurement.

Tables 4–6 show the cross-validated performance of the ten algorithms on the full candidate set, the 24 screened predictors, and the eight backward-eliminated predictors. Mean discrimination improved with more targeted feature selection: the best mean AUC rose from 0.580 (naïve Bayes) on the full set to 0.670 (naïve Bayes) on the screened set and to 0.682 (naïve Bayes) and 0.674 (L2 logistic regression) on the eight-feature set, while the best Brier score improved from 0.194 to 0.181. On the eight-feature set, L2-regularized logistic regression had a mean AUC of 0.674 ± 0.079 (fold range 0.43–0.85; 2.5th–97.5th percentiles 0.51–0.80), AUPRC 0.489 ± 0.089, and Brier score 0.181 ± 0.015. Naïve Bayes had a marginally higher mean AUC (0.682 ± 0.080) but a worse Brier score (0.193 ± 0.028) and lower AUPRC (0.465 ± 0.103); the tree-based ensembles, support-vector machine, k-nearest neighbors, and multilayer perceptron were inferior in discrimination. L2 logistic regression was therefore selected as the final model on the combination of discrimination, precision–recall, calibration, and interpretability. Its best model—the fit from the outer fold with the highest test AUC—reached an AUC of 0.850 (AUPRC 0.664, Brier score 0.153) on 63 patients with 17 events (Figure 2A); this value marks the upper end of fold-to-fold variability and is not an estimate of expected performance. When the full pipeline, including feature selection, was repeated inside each outer fold, the mean AUC was 0.587 ± 0.074 (**Section 3.5**).

3.5. Calibration, optimism, and fully nested validation

Out-of-fold predictions of the final model, averaged over the five cross-validation repeats, gave a pooled AUC of 0.673 (95% bootstrap CI 0.601–0.741), AUPRC 0.465 (0.371–0.574), and Brier score 0.179 (0.157–0.204). Calibration was acceptable (Figure 2B): the calibration slope was 0.88 (95% CI 0.49–1.33) and the calibration-in-the-large intercept 0.00 (−0.27 to 0.25); mean predicted

Table 2. Baseline characteristics of the overall cohort and by treatment group

Variable	Overall (*n* = 311)	Empagliflozin (*n* = 251)	Control (*n* = 60)	*p*
Age, years	62.20 ± 10.50 (55.00–70.00)	62.41 ± 10.62 (55.00–70.00)	61.28 ± 9.90 (55.75–67.25)	0.390
Female sex, %	27.3	29.1	20.0	0.197
BMI, kg/m^2	29.16 ± 5.03 (25.63–31.98)	29.56 ± 4.79 (26.07–32.05)	27.46 ± 5.61 (23.86–31.28)	0.004
Q-wave MI, %	51.8	52.2	50.0	0.776
Prior MI, %	26.4	27.1	23.3	0.627
Prior PCI, %	18.3	18.7	16.7	0.853
Type 2 diabetes, %	42.1	43.8	35.0	0.245
Hypertension, %	49.8	50.6	46.7	0.667
COPD, %	4.8	3.6	10.0	0.048
CKD, %	14.5	13.1	20.0	0.219
Current smoking, %	22.2	19.5	33.3	0.025
Heart rate, bpm	80.09 ± 15.78 (68–89.5)	79.32 ± 15.96 (68–88)	83.28 ± 14.55 (72–94)	0.023
Systolic BP, mmHg	131.97 ± 20.08 (120–145)	130.85 ± 20.26 (120–140)	136.67 ± 18.61 (120–150)	0.024
Diastolic BP, mmHg	81.82 ± 11.18 (75–90)	81.06 ± 11.30 (70–90)	85.00 ± 10.04 (80–90)	0.025
Fasting glucose, mmol/L	7.25 ± 2.85 (5.20–8.20)	7.18 ± 2.83 (5.20–8.20)	7.51 ± 2.90 (5.60–8.40)	0.183
HbA1c, %	6.82 ± 1.86 (5.60–7.80)	6.86 ± 1.95 (5.60–7.90)	6.68 ± 1.44 (5.78–7.43)	0.717
Postprandial glucose, mmol/L	9.34 ± 3.95 (6.30–11.15)	9.19 ± 4.04 (6.23–11.15)	9.97 ± 3.5 (7.6–10.8)	0.005
Creatinine, μmol/L	106.16 ± 31.55 (85–120)	104.68 ± 32.79 (84–118)	112.35 ± 24.79 (98–130)	0.002
eGFR, mL/min/1.73m^2	69.13 ± 21.28 (51.44–85.08)	71.09 ± 21.44 (53.44–88.32)	60.93 ± 18.51 (46.72–71.72)	0.001
Total cholesterol, mg/dL	196.91 ± 51.7 (161–229)	196.73 ± 53.4 (160–229)	197.68 ± 43.6 (175–228)	0.596
LDL-C, mg/dL	117.46 ± 43.67 (84–145)	117.06 ± 43.86 (84–145)	119.12 ± 42.82 (84–149)	0.675
HDL-C, mg/dL	41.59 ± 10.82 (35–48)	41.09 ± 10.96 (34–47)	43.70 ± 9.92 (38–51)	0.019
Triglycerides, mg/dL	189.19 ± 131.44 (114.5–217.5)	193.29 ± 141.04 (112–216.5)	172.07 ± 77.22 (122.25–219)	0.845
NT-proBNP, pg/mL	2583.24 ± 4578.6 (405–2721.5)	3056.11 ± 4974.0 (450–3289.5)	605.08 ± 558.9 (292.5–774.5)	<0.001
CRP, mg/L	33.24 ± 64.08 (4.65–24.53)	37.42 ± 70.53 (3.8–25.52)	15.72 ± 9.72 (9.18–18.23)	0.027
LVEF, %	49.23 ± 9.04 (44.1–55.83)	49.75 ± 9.07 (45.16–56.58)	47.04 ± 8.55 (42–52.38)	0.013
LA volume, mL	54.16 ± 15.88 (43–62.95)	53.71 ± 16.31 (42–61.9)	56 ± 13.8 (46.18–64)	0.054
LAVI, mL/m^2	27.93 ± 8.80 (22.1–32.25)	27.45 ± 9.02 (21.55–30.23)	29.94 ± 7.48 (25.03–33.78)	0.002
E, cm/s	68.71 ± 20.28 (55–79.5)	68.55 ± 21.19 (54–80)	69.38 ± 15.87 (61–78)	0.252
A, cm/s	75.01 ± 21.12 (60–88)	75.83 ± 19.78 (65–88)	71.60 ± 25.73 (51.5–99)	0.075
E/A ratio	1.01 ± 0.51 (0.69–1.20)	1.00 ± 0.54 (0.68–1.17)	1.06 ± 0.34 (0.82–1.25)	0.004
e' septal, cm/s	6.58 ± 1.47 (5.7–7.3)	6.40 ± 1.21 (5.7–7)	7.34 ± 2.08 (5.97–8.72)	<0.001
e' lateral, cm/s	8.80 ± 1.66 (7.89–9.8)	8.47 ± 1.34 (7.81–9.23)	10.08 ± 2.15 (8.1–11.47)	<0.001
e' average, cm/s	7.69 ± 1.49 (6.95–8.35)	7.43 ± 1.17 (6.86–7.90)	8.72 ± 2.08 (7.07–10.25)	<0.001
E/e' ratio	9.34 ± 4.12 (7.03–10.43)	9.60 ± 4.43 (7.10–10.53)	8.30 ± 2.26 (6.57–9.57)	0.173
TR velocity, m/s	2.47 ± 0.55 (2.16–2.70)	2.47 ± 0.59 (2.11–2.70)	2.45 ± 0.38 (2.22–2.64)	0.617
TR gradient, mmHg	26.58 ± 8.05 (22–28)	27.07 ± 8.04 (22.2–28)	24.60 ± 7.79 (19.6–28)	0.011
LVEDV, mL	107.12 ± 34.75 (83.83–127.62)	108.14 ± 36.00 (84.15–128)	103.01 ± 28.86 (83.00–124.72)	0.677

(*Cont'd...*)

Table 2. (*Continued*)

Variable	Overall (*n* = 311)	Empagliflozin (*n* = 251)	Control (*n* = 60)	*p*
LVESV, mL	56.21 ± 26.20 (38–67.15)	56.56 ± 27.84 (37–67)	54.81 ± 18.18 (41–67.6)	0.432
LVEDD, cm	5.09 ± 0.64 (4.6–5.5)	5.09 ± 0.64 (4.6–5.5)	5.06 ± 0.62 (4.7–5.5)	0.959
LVESD, cm	3.71 ± 0.75 (3.2–4.19)	3.71 ± 0.75 (3.2–4.2)	3.71 ± 0.73 (3.19–4.11)	0.898
LV mass, g	208.07 ± 56.43 (171.35–237.73)	206.05 ± 57.30 (171.01–232.28)	216.09 ± 52.06 (173.75–244.43)	0.073
LV mass index, g/m²	106.58 ± 27.59 (87.45–121.64)	104.25 ± 26.56 (86.57–117.16)	115.86 ± 29.62 (94.33–137.10)	0.003
LA antero–posterior diameter, cm	3.92 ± 0.57 (3.5–4.3)	3.91 ± 0.57 (3.5–4.28)	3.94 ± 0.60 (3.46–4.39)	0.638
Aspirin, %	93.2	91.6	100.0	0.018
Clopidogrel, %	49.2	46.2	61.7	0.043
ACE inhibitor, %	50.5	41.8	86.7	<0.001
Mineralocorticoid-receptor antagonist, %	67.6	78.3	23.3	<0.001
Loop diuretic, %	24.1	24.7	21.7	0.737
Metformin, %	8.7	3.6	30.0	<0.001
Beta-blocker, %	96.3	91.5	100.0	<0.001

Notes: Continuous variables are mean ± standard deviation (interquartile range); categorical variables are percentages; *p*-values are from Fisher's exact test (categorical) or the Kruskal–Wallis test (continuous). Abbreviations: A: Peak late transmitral velocities; ACE: Angiotensin-converting enzyme; BMI: Body mass index; BP: Blood pressure; CKD: Chronic kidney disease; COPD: Chronic obstructive pulmonary disease; CRP: C-reactive protein; E: Peak early transmitral velocities; eGFR: Estimated glomerular filtration rate; HbA1c: Glycated hemoglobin; HDL: High-density lipoprotein; LA: Left-atrial; LAVI: Left-atrial volume index; LDL: Low-density lipoprotein; LV: Left-ventricular; LVEDD: Left-ventricular end-diastolic dimension; LVEDV: Left-ventricular end-diastolic volume; LVEF: Left-ventricular ejection fraction; LVESD: Left-ventricular end-systolic dimension; LVESV: Left-ventricular end-systolic volume; MI: Myocardial infarction; NT-proBNP: N-terminal pro-B-type natriuretic peptide; PCI: Percutaneous coronary intervention; TR: Tricuspid regurgitation.

Table 3. Univariable predictors of six-month MACE retained at $p < 0.20$ (logistic regression; odds ratio per unit of the raw variable or per level of the deviation score)

Predictor	OR	95% CI	*p*
Body mass index, per kg/m²	0.920	0.870–0.970	0.002
Peak A velocity, per cm/s	0.981	0.969–0.994	0.004
LVEDV, deviation score	2.200	1.240–3.890	0.007
Current smoking	2.140	1.210–3.770	0.009
LVESV, per mL	1.012	1.003–1.022	0.010
LVEDV, per mL	1.009	1.002–1.016	0.017
Prior myocardial infarction	1.920	1.110–3.300	0.019
E/A ratio, per unit	1.680	1.060–2.680	0.029
LVESD, per cm	1.460	1.039–2.050	0.029
LVEF, per %	0.970	0.940–0.998	0.033
E/A ratio, deviation score	0.580	0.350–0.965	0.036
Triglycerides, per mg/dL	0.997	0.995–1.000	0.042
Body mass index, deviation score	0.580	0.320–1.041	0.068
LVESV, deviation score	1.580	0.965–2.570	0.069
Triglycerides, deviation score	0.700	0.480–1.029	0.070

(*Cont'd...*)

Table 3. (*Continued*)

Predictor	OR	95% CI	*p*
TR velocity, per m/s	1.460	0.930–2.290	0.097
CRP, per mg/L	1.003	0.999–1.007	0.103
Empagliflozin exposure	0.610	0.330–1.110	0.107
Mineralocorticoid-receptor antagonist use	1.570	0.890–2.750	0.120
HbA1c, deviation score	0.700	0.440–1.120	0.138
Peak A velocity, deviation score	0.690	0.420–1.150	0.154
Average e', deviation score	0.650	0.350–1.230	0.191
Aspirin use	2.290	0.660–7.970	0.195
LA volume, deviation score	1.400	0.840–2.350	0.200

Notes: Deviation score: 0, within the reference range; 1, outside the range by up to three range-widths; 2, beyond three range-widths. Abbreviations: CI: Confidence interval; CRP: C-reactive protein; E/A ratio: Peak early/late transmitral velocity ratio; HbA1c: Glycated hemoglobin; LA: Left-atrial; LVEDV: Left-ventricular end-diastolic volume; LVEF: Left-ventricular ejection fraction; LVESD: Left-ventricular end-systolic dimension; LVESV: Left-ventricular end-systolic volume; MACE: Major adverse cardiovascular event; OR: Odds ratio; TR: Tricuspid regurgitation.

Table 4. Cross-validated performance of 10 algorithms on the full candidate feature set (84 predictors)

Model	AUC, mean ± SD	AUC, 95% range	AUC, best model	AUPRC, mean ± SD	Brier score, mean ± SD
Naïve Bayes	0.580 ± 0.071	0.431–0.667	0.667	0.381 ± 0.078	0.322 ± 0.053
Random forest	0.577 ± 0.071	0.453–0.687	0.711	0.370 ± 0.072	0.194 ± 0.007
Logistic regression (L2)	0.556 ± 0.075	0.416–0.688	0.710	0.361 ± 0.062	0.201 ± 0.017
Extra trees	0.552 ± 0.072	0.414–0.682	0.735	0.353 ± 0.065	0.195 ± 0.005
Logistic regression (L1)	0.545 ± 0.069	0.418–0.656	0.693	0.337 ± 0.057	0.215 ± 0.024
Gradient boosting	0.539 ± 0.076	0.392–0.649	0.664	0.343 ± 0.067	0.226 ± 0.023
XGBoost	0.537 ± 0.077	0.403–0.676	0.698	0.353 ± 0.073	0.209 ± 0.017
k-nearest neighbors	0.516 ± 0.081	0.363–0.625	0.625	0.314 ± 0.052	0.231 ± 0.029
SVM (RBF)	0.487 ± 0.067	0.375–0.603	0.616	0.328 ± 0.050	0.198 ± 0.005
MLP	0.485 ± 0.056	0.382–0.573	0.594	0.290 ± 0.055	0.306 ± 0.062

Notes: Values are mean ± SD across the 25 outer folds of repeated stratified five-fold cross-validation (5 repeats) after inner-loop hyperparameter tuning; 95% range is the 2.5th–97.5th percentile of the fold AUCs; best model is the AUC of the fitted model from the single outer fold with the highest test AUC (an optimistic upper bound, not an estimate of expected performance). Abbreviations: AUC: Area under the receiver-operating-characteristic curve; AUPRC: Area under the precision–recall curve; MLP: Multilayer perceptron; RBF: Radial basis function; SD: Standard deviation; SVM: Support vector machine.

risk (26.7%) equaled the observed incidence, and observed and predicted risk agreed across octiles, with wide CIs in the highest-risk octile. Harrell's bootstrap gave an apparent AUC of 0.720, an optimism of 0.034, and an optimism-corrected AUC of 0.686 for the fixed eight-feature model. When univariable screening, backward elimination, and tuning were all repeated inside each of the 25 outer folds, the fully nested AUC was 0.587 ± 0.074 (fold range 0.40–0.75; pooled out-of-fold AUC 0.54–0.61 across repeats), with AUPRC 0.400 ± 0.082 and Brier score 0.201 ± 0.018. Feature selection was moderately stable: body mass index and current smoking were selected in all 25 folds, peak A velocity in 80%, prior myocardial infarction in 76%, the LVEDV deviation score in 72%, and triglycerides in 68%, whereas empagliflozin exposure (48%) and mineralocorticoid-receptor antagonist use (36%) were selected in fewer than half. The difference between 0.674 and 0.587 is the optimism attributable to selecting predictors on the same 311 patients on which the model was then evaluated; it, rather than the best-model value, sets the realistic expectation for a new sample.

Table 5. Cross-validated performance of 10 algorithms on the 24 predictors retained by univariable screening (Table 3)

Model	AUC, mean ± SD	AUC, 95% range	AUC, best model	AUPRC, mean ± SD	Brier score, mean ± SD
Naïve Bayes	0.670 ± 0.084	0.527–0.822	0.845	0.476 ± 0.102	0.258 ± 0.046
Logistic regression (L1)	0.631 ± 0.082	0.469–0.754	0.795	0.442 ± 0.090	0.193 ± 0.018
Logistic regression (L2)	0.629 ± 0.086	0.467–0.769	0.834	0.449 ± 0.093	0.191 ± 0.017
Random forest	0.613 ± 0.080	0.460–0.740	0.789	0.426 ± 0.079	0.189 ± 0.010
Extra trees	0.605 ± 0.077	0.460–0.738	0.776	0.402 ± 0.079	0.191 ± 0.007
Gradient boosting	0.584 ± 0.081	0.420–0.721	0.738	0.389 ± 0.087	0.213 ± 0.026
XGBoost	0.570 ± 0.081	0.421–0.713	0.728	0.385 ± 0.084	0.211 ± 0.024
SVM (RBF)	0.565 ± 0.079	0.444–0.700	0.730	0.400 ± 0.072	0.194 ± 0.005
k-nearest neighbors	0.519 ± 0.066	0.390–0.656	0.669	0.327 ± 0.048	0.211 ± 0.020
MLP	0.473 ± 0.074	0.280–0.560	0.569	0.296 ± 0.054	0.261 ± 0.028

Notes: Values are mean ± SD across the 25 outer folds of repeated stratified five-fold cross-validation (5 repeats) after inner-loop hyperparameter tuning; 95% range is the 2.5th–97.5th percentile of the fold AUCs; best model is the AUC of the fitted model from the single outer fold with the highest test AUC (an optimistic upper bound, not an estimate of expected performance). Abbreviations: AUC: Area under the receiver-operating-characteristic curve; AUPRC: Area under the precision–recall curve; MLP: Multilayer perceptron; RBF: Radial basis function; SD: Standard deviation; SVM: Support vector machine.

Table 6. Cross-validated performance of 10 algorithms on the eight predictors retained by cross-validated backward elimination

Model	AUC, mean ± SD	AUC, 95% range	AUC, best model	AUPRC, mean ± SD	Brier score, mean ± SD
Naïve Bayes	0.682 ± 0.080	0.540–0.806	0.832	0.465 ± 0.103	0.193 ± 0.028
Logistic regression (L2)	0.674 ± 0.079	0.505–0.801	0.850	0.489 ± 0.089	0.181 ± 0.015
Logistic regression (L1)	0.672 ± 0.085	0.497–0.799	0.844	0.488 ± 0.100	0.181 ± 0.015
Random forest	0.639 ± 0.086	0.436–0.765	0.793	0.457 ± 0.089	0.186 ± 0.010
MLP	0.639 ± 0.068	0.520–0.781	0.798	0.488 ± 0.085	0.413 ± 0.024
Extra trees	0.636 ± 0.075	0.514–0.769	0.772	0.406 ± 0.087	0.189 ± 0.007
SVM (RBF)	0.624 ± 0.081	0.471–0.776	0.802	0.466 ± 0.085	0.184 ± 0.011
k-nearest neighbors	0.622 ± 0.065	0.489–0.731	0.738	0.404 ± 0.061	0.195 ± 0.015
XGBoost	0.588 ± 0.083	0.410–0.724	0.730	0.404 ± 0.076	0.200 ± 0.017
Gradient boosting	0.580 ± 0.069	0.428–0.676	0.678	0.397 ± 0.076	0.209 ± 0.025

Notes: L2-regularized logistic regression was selected as the final prognostic model; values are mean ± SD across the 25 outer folds of repeated stratified five-fold cross-validation (5 repeats) after inner-loop hyperparameter tuning; 95% range is the 2.5th–97.5th percentile of the fold AUCs; best model is the AUC of the fitted model from the single outer fold with the highest test AUC (an optimistic upper bound, not an estimate of expected performance). Abbreviations: AUC: Area under the receiver-operating-characteristic curve; AUPRC: Area under the precision–recall curve; MLP: Multilayer perceptron; RBF: Radial basis function; SD: Standard deviation; SVM: Support vector machine.

3.6. Model interpretability

The SHAP analysis of the final logistic-regression model (Figure 3) was used to characterize the direction and relative magnitude of each retained feature's contribution to predicted MACE risk, providing case-level as well as cohort-level interpretability in line with the explainability expectations for clinical prediction models discussed in **Section 4.6**.

3.7. Spearman correlation analysis

Spearman rank correlations between individual candidate features and the MACE indicator (Table 7) were of small-to-moderate magnitude throughout the feature set ($|\rho| \leq 0.17$) and, after Benjamini–Hochberg correction for multiple comparisons across the full candidate set, none remained significant at the conventional $\alpha = 0.05$ threshold, although BMI, current smoking, and peak A velocity showed the strongest unadjusted associations (*p*

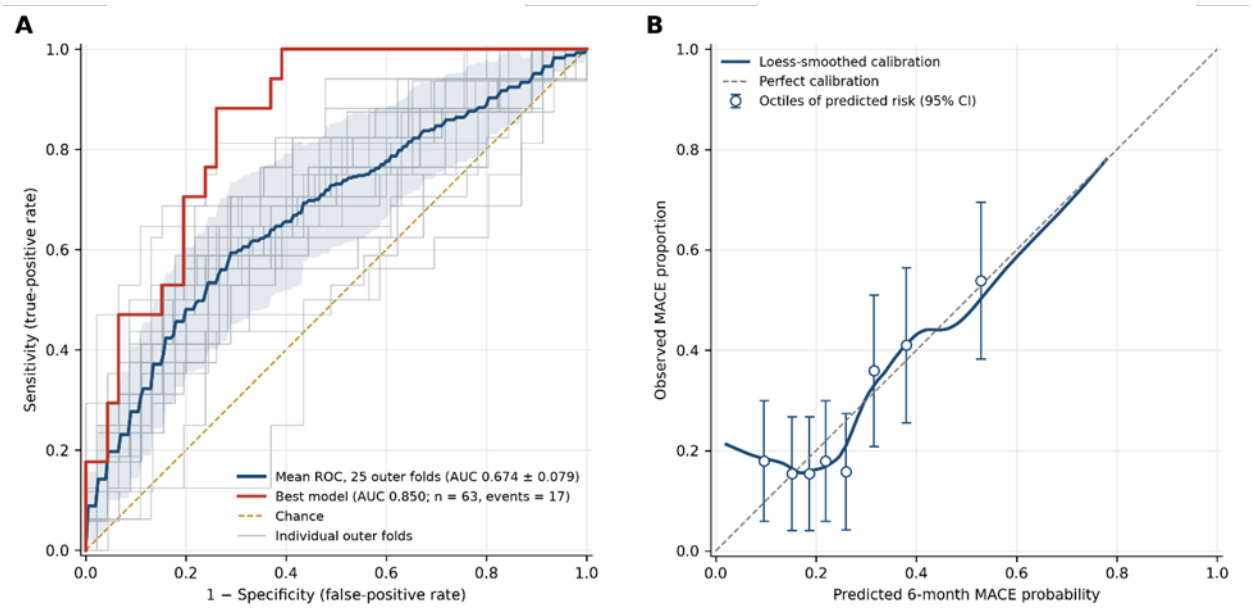


Figure 2. Discrimination and calibration of the final model (L2-regularized logistic regression, eight predictors). (A) ROC curves: Mean curve over the 25 outer folds with ± 1 SD band (blue), the best model (red; AUC 0.850 on 63 patients with 17 events), and individual outer folds (grey). (B) Calibration of out-of-fold predictions averaged over five cross-validation repeats: octiles of predicted risk with 95% CIs, a loess smoother, and the identity line; calibration slope 0.88, intercept 0.00.
Abbreviations: AUC: Area under the receiver operating characteristic curve; CI: Confidence interval; ROC: Receiver operating characteristic; SD: Standard deviation.

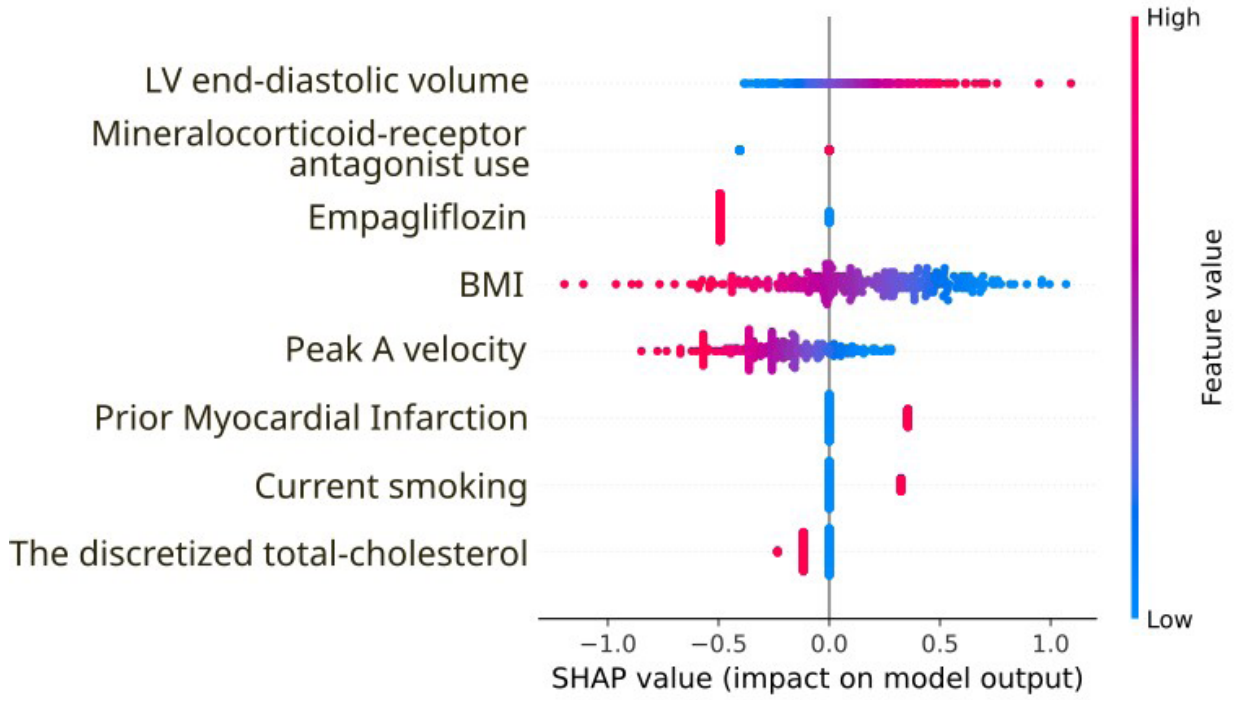


Figure 3. SHAP summary plot for the final logistic-regression model (eight backward-eliminated features), showing the direction and magnitude of each feature's contribution to predicted six-month MACE probability
Abbreviations: BMI: Body mass index; LV: Left-ventricular; MACE: Major adverse cardiovascular event; SHAP: SHapley Additive exPlanations.

< 0.01). This pattern—modest univariable correlations that nonetheless combine into a moderately discriminating multivariable model—indicates that individual-marker signal is weak in this cohort and that risk stratification depends on the joint, rather than marginal, information carried by these variables.

Table 7. Spearman correlations between selected candidate features and six-month MACE (top ranked by |ρ|; full results available from the corresponding author)

Feature	ρ	*p* (raw)	*p* (BH-adjusted)
BMI	−0.165	0.004	0.144
Current smoking	0.150	0.008	0.144
Peak A velocity	−0.149	0.008	0.144
Prior MI	0.134	0.018	0.236
LVESV	0.125	0.030	0.241
TR velocity	0.124	0.032	0.241
LVEDV	0.124	0.032	0.241
HbA1c	−0.108	0.057	0.318
TR gradient	0.107	0.064	0.318
CRP	0.104	0.066	0.318
LVEF	−0.103	0.070	0.318
E/A ratio	0.102	0.073	0.318
Empagliflozin exposure	−0.092	0.106	0.392

Abbreviations: BH: Benjamini–Hochberg; BMI: Body mass index; CRP: C-reactive protein; E/A ratio: Peak early/late transmitral velocity ratio; HbA1c: Glycated hemoglobin; LVEDV: Left-ventricular end-diastolic volume; LVEF: Left-ventricular ejection fraction; LVESV: Left-ventricular end-systolic volume; MACE: Major adverse cardiovascular event; MI: Myocardial infarction; TR: Tricuspid regurgitation.

3.8. Association between early empagliflozin and six-month MACE: Outcomes by treatment group and LVEF phenotype

Major adverse cardiovascular events occurred in 62 of 251 patients (24.7%) in the empagliflozin group and in 21 of 60 controls (35.0%) (Table 8). In the empagliflozin group, in which component events were recorded prospectively, 12 patients (4.8%) died, 62 (24.7%) had an unplanned cardiovascular hospitalization, 9 (3.6%) had a non-fatal reinfarction, and 7 (2.8%) underwent urgent percutaneous coronary intervention. In the control group, 5 patients (8.3%) died, and 12 (20.0%) were rehospitalized; reinfarction and urgent revascularization could not be separated reliably from the composite in the retrospective records. Baseline LVEF phenotype differed between groups at trend level ($p = 0.084$): reduced LVEF in 14.7% vs. 21.7%, mildly reduced in 29.5% vs. 38.3%, and preserved in 55.8% vs. 40.0%. In the empagliflozin group, MACE incidence rose with worsening phenotype (20.0%, 25.7%, and 40.5% for preserved, mildly reduced, and reduced LVEF), whereas in controls it was highest in the preserved-LVEF stratum (41.7%, 30.4%, and 30.8%). The between-group difference was thus concentrated in patients with preserved LVEF (20.0% vs. 41.7%) and absent, or reversed, in those with reduced LVEF (40.5% vs. 30.8%). The treatment × phenotype interaction did not reach significance (odds ratio 2.12 per phenotype category; $p = 0.067$) and, with 13 controls in the reduced-LVEF stratum, this pattern is exploratory.

3.9. Propensity-score analysis

The 20-covariate propensity model separated the groups strongly (c-statistic 0.89): the median propensity score was 0.94 (interquartile range 0.83–0.98) in treated patients and 0.53 (0.36–0.77) in controls, and 102 treated patients lay outside the region of common support (0.06–0.96), whereas all controls lay within it (Figure 4A). Before matching, 16 of 20 covariates had |SMD| > 0.10, the largest for NT-proBNP (0.90), eGFR (0.51), CRP (0.43), body mass index (0.40), current smoking (−0.32), and systolic blood pressure (−0.30) (Table 9). Nearest-neighbor matching within the caliper produced 44 pairs (88 patients); 16 controls, mostly with low propensity scores, could not be matched. Matching reduced the SMD of the logit propensity score from 1.70 to 0.04 and brought 12 of 20 covariates below 0.10, but residual imbalance (|SMD| 0.11–0.24) remained for body mass index, diabetes, chronic kidney disease, systolic blood pressure, eGFR, NT-proBNP, sex, and reduced LVEF (Figure 4B)—a consequence of the small control group rather than of the matching algorithm. In the matched sample, MACE occurred in 25.0% of treated and 29.5% of control patients (odds ratio 0.79, 95% CI 0.32–1.98; $p = 0.62$; 7 vs. 9 discordant pairs, McNemar $p = 0.80$). Sensitivity analyses on the whole cohort gave odds ratios of 0.88 (95% CI 0.56–1.40) with inverse-probability weighting and 0.68 (0.33–1.41) with adjustment for the logit propensity score; a simpler six-covariate model (age, sex, Q-wave infarction, diabetes, hypertension, eGFR) gave 0.67 (0.30–1.47). When the six baseline drug classes were added to the propensity model, its *c*-statistic rose to 0.98, only 19 pairs could be formed, and the matched odds ratio was 0.58 (0.15–2.26): background therapy in this cohort is so closely tied to the decision to prescribe empagliflozin that it cannot be balanced by design, and all estimates above should be read with this in mind.

Table 8. Six-month outcomes and LVEF phenotype by treatment group

Variable	Empagliflozin (n = 251)	Control (n = 60)	p
Outcomes			
MACE (composite)	62 (24.7)	21 (35.0)	0.108
Death from any cause	12 (4.8)	5 (8.3)	0.338
Unplanned cardiovascular hospitalization	62 (24.7)	12 (20.0)	0.503
Non-fatal reinfarction	9 (3.6)	n/a	–
Urgent percutaneous coronary intervention	7 (2.8)	n/a	–
Stroke	0	0	–
Baseline LVEF phenotype			0.084
Reduced (<40%)	37 (14.7)	13 (21.7)	
Mildly reduced (40%–49%)	74 (29.5)	23 (38.3)	
Preserved (≥50%)	140 (55.8)	24 (40.0)	
MACE within LVEF phenotype			
Reduced (<40%)	15/37 (40.5)	4/13 (30.8)	0.742
Mildly reduced (40%–49%)	19/74 (25.7)	7/23 (30.4)	0.788
Preserved (≥50%)	28/140 (20.0)	10/24 (41.7)	0.033

Notes: Values are n (%); p-values are from Fisher's exact test (outcomes) or the χ^2 test (phenotype distribution). Abbreviations: LVEF: Left-ventricular ejection fraction; MACE: Major adverse cardiovascular event; n/a: Not available.

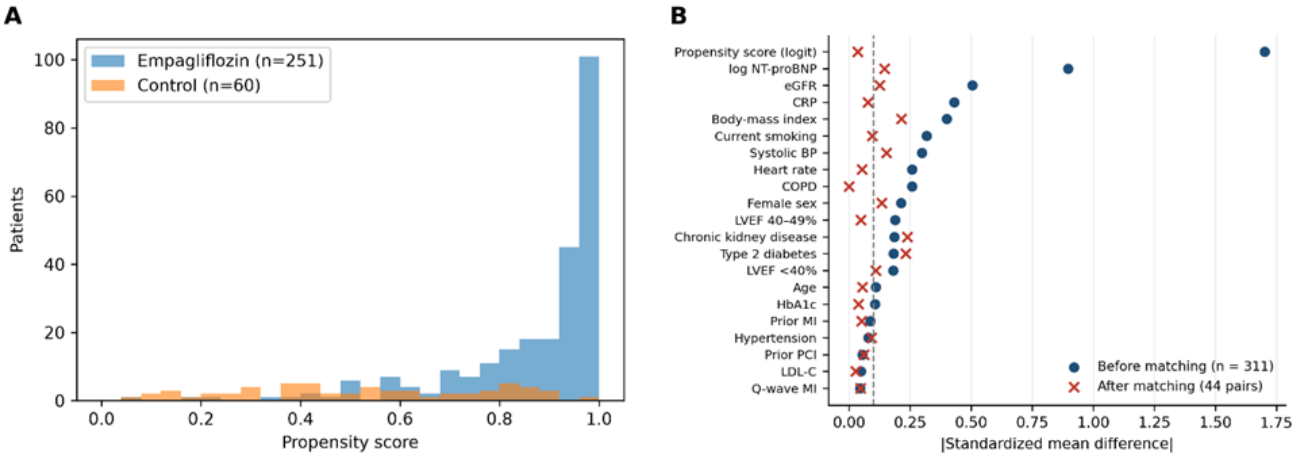


Figure 4. Propensity-score diagnostics. (A) Distribution of the estimated propensity score for empagliflozin in treated patients and controls before matching. (B) Absolute standardized mean differences for the 20 covariates and the logit propensity score before (circles) and after (crosses) 1:1 caliper matching; the dashed line marks |SMD| = 0.10. Abbreviations: BP: Blood pressure; COPD: Chronic obstructive pulmonary disease; CRP: C-reactive protein; eGFR: Estimated glomerular filtration rate; HbA1c: Glycated hemoglobin; LDL: Low-density lipoprotein; LVEF: Left-ventricular ejection fraction; MI: Myocardial infarction; NT-proBNP: N-terminal pro-B-type natriuretic peptide; PCI: Percutaneous coronary intervention; SMD: Standardized mean difference.

Table 9. Covariate balance before and after 1:1 propensity-score matching (44 pairs)

Covariate	SMD before matching (n = 311)	SMD after matching (n = 88)
Age	0.11	−0.05
Female sex	0.21	−0.13
Body mass index	0.40	0.21
Q-wave MI	0.04	0.05
Prior MI	0.09	0.05
Prior PCI	0.05	0.06
Type 2 diabetes	0.18	0.23
Hypertension	0.08	0.09
Chronic kidney disease	−0.19	0.24
COPD	−0.26	0.00
Current smoking	−0.32	0.09
Heart rate	−0.26	0.05
Systolic blood pressure	−0.30	0.15
eGFR	0.51	−0.13
HbA1c	0.11	0.04
LDL-cholesterol	−0.05	−0.03
CRP	0.43	0.08
log NT-proBNP	0.90	0.14
LVEF < 40%	−0.18	0.11
LVEF 40%–49%	−0.19	−0.05
Propensity score (logit)	1.70	0.04

Notes: SMD = Empagliflozin minus control; |SMD| > 0.10 indicates meaningful imbalance. Abbreviations: COPD: Chronic obstructive pulmonary disease; CRP: C-reactive protein; eGFR: Estimated glomerular filtration rate; HbA1c: Glycated hemoglobin; LDL: Low-density lipoprotein; LVEF: Left-ventricular ejection fraction; MI: Myocardial infarction; NT-proBNP: N-terminal pro-B-type natriuretic peptide; PCI: Percutaneous coronary intervention; SMD: Standardized mean difference.

4. Discussion

This single-center cohort of 311 patients with AMI addressed two separate questions. On the first—which patients remain at risk during the six months after infarction despite early empagliflozin—the answer is clinical rather than technical: prior infarction, current smoking, lower body mass index, lower peak A velocity, a larger left-ventricular end-diastolic volume, and treatment with a mineralocorticoid-receptor antagonist identified the patients who went on to die, to be readmitted, or to need urgent revascularization, and they did so about as well in a regularized logistic regression as in any of nine ML classifiers (mean cross-validated AUC 0.67, with good calibration). On the second—whether early empagliflozin was associated with fewer events than contemporaneous standard care—the propensity-matched estimate (odds ratio 0.79, 95% CI 0.32–1.98) is compatible with anything from a substantial benefit to a modest harm, and neither adds to nor contradicts the randomized evidence. Two descriptive findings frame both answers: LVEF recovery was related to baseline systolic phenotype and not to glycemic control, and ventricular reverse remodeling far outpaced atrial remodeling over six months.

4.1. Early sodium-glucose cotransporter-2 inhibition after myocardial infarction: The trials and this cohort

The randomized evidence on SGLT2 inhibition started early after infarction is now consistent. EMMY showed that empagliflozin lowers NT-proBNP and improves LVEF, diastolic indices, and ventricular volumes over 26 weeks[6], but it was a surrogate-endpoint trial. DAPA-MI, the first outcome trial, randomized 4,017 patients without diabetes or established heart failure; its hierarchical composite favored dapagliflozin, but the benefit came entirely from the cardiometabolic components (new-onset diabetes and weight), with no difference in death, heart-failure hospitalization, or non-fatal infarction – an absence of efficacy on hard outcomes in a low-risk population.[7,35] EMPACT-MI, which enrolled higher-risk patients with LVEF < 45% or congestion, found no reduction in all-cause death or heart-failure hospitalization with empagliflozin, although first heart-failure hospitalizations were fewer[8] and the benefit concentrated in patients with the lowest LVEF and highest natriuretic peptides.[36] EMPRESS-MI found no incremental effect on ventricular volumes beyond background therapy[9], and the timing of initiation within the first 72 hours did not modify the NT-proBNP response in EMMY.[37] The view that has emerged treats early SGLT2 inhibition as a selective strategy for patients with impaired systolic function or overt congestion rather than as routine post-infarction therapy[12,38], and the drug is not a guideline-mandated component of post-infarction care.[2]

In our center, empagliflozin was prescribed at the discretion of the treating cardiologist on the strength of EMMY and the 2023 ESC guidance, which is why the treated group was large (251 of 311) and why it differed from controls in the way that clinical discretion produces: higher body mass index, NT-proBNP, and CRP, lower mitral annular velocities, more frequent mineralocorticoid-receptor antagonist use, and less frequent ACE inhibitor and metformin use. In other words, physicians tended to give empagliflozin to patients they judged to be at higher risk of heart failure, and the control group is enriched for patients in whom that concern was absent. Against

this background, a neutral matched estimate with wide confidence limits is what 60 controls can deliver, and our data neither support nor refute a treatment effect. The one signal worth carrying forward is the phenotype pattern: the lower event rate in treated patients was confined to those with preserved LVEF (20.0% vs. 41.7%) and absent in those with reduced LVEF (40.5% vs. 30.8%), the opposite of the EMPACT-MI subgroup finding. With 13 controls in the reduced-LVEF stratum and an interaction *p*-value of 0.067, this is exploratory, but it argues for stratifying by LVEF phenotype in any future non-randomized comparison rather than treating post-infarction patients as one population.

4.2. Heart-failure phenotype, remodeling, and glycemia

Left-ventricular ejection fraction improved in roughly a third of treated patients with reduced or mildly reduced baseline LVEF, and the proportion improving was the same whether HbA1c was above or below 8%. Because fasting glucose and HbA1c were balanced at baseline, glycemia is an unlikely explanation for this pattern. The observation is compatible with, but does not prove, the non-glycemic mechanisms proposed for SGLT2 inhibitors – a shift of myocardial substrate use toward ketone bodies and fatty acids[39], natriuresis and ventricular unloading, and an anti-inflammatory effect documented in EMMY[40]—and with the favorable effects on volumes seen in nondiabetic patients with reduced LVEF in EMPA-TROPISM and SUGAR-DM-HF.[10,11] Our remodeling data are observational and were obtained in the treated group alone, so they cannot attribute recovery to the drug rather than to reperfusion, guideline-directed therapy, and time.

The dissociation between ventricular and atrial remodeling has a more direct clinical use. Ventricular volumes changed substantially within six months (r = 0.82–0.83), while indexed left-atrial volume barely moved (r = 0.12–0.19). Left-atrial volume integrates filling pressure over months to years and is regarded by the ASE/EACVI as a marker of chronic rather than instantaneous diastolic load[21]; the mean LAVI in this cohort (27.9 mL/m^2) was below the dilation threshold, although mitral annular velocities in the treated group were already below normal (septal e' 6.4, lateral e' 8.5 cm/s), that is, relaxation was impaired ahead of structural atrial change. For following treatment response over a six-month horizon, ventricular volumes and NT-proBNP are informative; LAVI is not, and its stability should not be read as absence of effect. Heart-failure phenotype defined by baseline LVEF[22] behaved as expected in the treated group, with MACE rising from 20% in preserved to 41% in reduced LVEF, but it did not add discrimination to the prognostic model once LVEDV and peak A velocity were included (mean AUC 0.672 vs. 0.674). We could not characterize clinical heart failure at admission because Killip class was not recorded, which limits the phenotype description to LVEF and NT-proBNP.

4.3. Who remains at risk on early empagliflozin: A clinical reading of the model

The eight predictors retained by cross-validated selection form a coherent clinical picture. Prior infarction and current smoking are established markers of recurrent ischemic events and were the two most stably selected predictors (retained in 76% and 100% of nested folds). A larger LVEDV reflects the extent of infarct-related dilatation and the substrate for adverse remodeling. Lower peak A velocity in a population with impaired relaxation indicates either loss of atrial contractile reserve or progression toward pseudonormal and restrictive filling, both of which carry prognostic weight after infarction.[21] Mineralocorticoid-receptor antagonist use is not plausibly harmful; in this cohort it identifies patients whose physicians judged them to have post-infarction heart failure or reduced LVEF—confounding by indication that, in a prognostic model, works as a marker of severity. The inverse association of body mass index with events (odds ratio 0.92 per kg/m^2) is the pattern usually called the "obesity paradox" and most likely reflects the frailty, malnutrition, and comorbidity of lower-weight patients rather than a protective effect of adiposity; the parallel inverse associations of triglycerides and of the total-cholesterol deviation score point the same way. Two expected predictors were absent: Neither NT-proBNP nor eGFR passed univariable screening, even after log transformation of NT-proBNP. The most likely explanation is the strong dependence of both on treatment allocation in this cohort (NT-proBNP was five-fold higher in treated patients), which makes their marginal association with outcome uninformative once empagliflozin exposure—itself retained in the model—is among the candidates. The baseline lipid profile deserves a clinical comment independent of the model: mean LDL-C (117 mg/dL) was more than twice the < 55 mg/dL target that the 2025 ESC/EAS update retains for patients after acute coronary syndrome, and three-quarters of patients remained above 70 mg/dL, although the same update recommends starting combination lipid-lowering therapy during the index admission.[34]

4.4. Performance of the prognostic model in context

The expected discrimination of the model is moderate. A mean cross-validated AUC of 0.67 (95% CI 0.60–0.74) on the fixed eight-feature set, and 0.59 when the cost of selecting those features is included, lies below the pooled validation C-index of 0.77 reported by the 2025 meta-

analysis of ML models for MACE after infarction.[17] The gap is informative: Most models in that meta-analysis were developed on registries an order of magnitude larger, and several were evaluated without nested feature selection. The best model in our comparison reached an AUC of 0.85 on one outer fold of 63 patients; had that value been reported alone, the model would have appeared to outperform the literature, which illustrates why fold-level maxima should be shown but not quoted as performance.[25,29] Three features of the result are nevertheless encouraging for a cohort of this size. Calibration was good (slope 0.88, intercept 0.00), so predicted risks can be read at face value within this population.[28] Logistic regression matched or exceeded every ML classifier, in line with the pattern across cardiovascular prediction tasks in which the best algorithm is outcome- and dataset-specific and flexible learners rarely beat regression on small tabular datasets.[14–16] Additionally, the events-per-variable ratio of the final model (83 events, eight predictors) meets the classical rule of thumb, although it remains below the sample size that current criteria would require for a model intended for deployment.[30] The model should therefore be regarded as a ranked set of clinically interpretable risk markers with a calibrated probability scale, not as a tool ready for individual decision-making. Because the final model is a logistic regression, it can be refitted across centers without pooling patient-level data through a federated iteratively reweighted least-squares scheme in which each center transmits only its Hessian and gradient contributions; this is the approach planned for the multicenter extension of this cohort.

4.5. Study limitations

Several limitations of the present study should be acknowledged and considered when interpreting the results:

(i) Non-randomized treatment allocation: Empagliflozin was prescribed at the physician's discretion, and the treated group had higher body mass index and NT-proBNP, lower mitral annular velocities, better renal function, less smoking, and a markedly different background regimen (ACE inhibitors 41.8% vs. 86.7%, mineralocorticoid-receptor antagonists 78.3% vs. 23.3%, metformin 3.6% vs. 30.0%; all $p < 0.001$). Propensity matching on 20 covariates balanced the propensity score but left residual imbalance in eight covariates, and background therapy could not be balanced without discarding most of the cohort. The treatment estimates are therefore hypothesis-generating.

(ii) Retrospective, selected control group: Controls were assembled from records of the same period and center and were included only if six-month data were available. Patients who died early or were lost to follow-up before six months may therefore be under-represented among controls; this would bias the comparison against empagliflozin if anything, but the direction cannot be verified. Component events other than death and rehospitalization could not be separated reliably for controls.

(iii) Sample size and events: With 83 events, 60 controls, and 13 controls with reduced LVEF, CIs for the treatment effect, for the phenotype interaction, and for fold-level model performance are wide, and the fully nested AUC shows that the feature-selection procedure itself is unstable at this sample size.

(iv) No external validation: All performance estimates are internal; transportability to other centers, and to populations in which SGLT2 inhibitors are prescribed differently, is unknown. Fairness and subgroup performance of the model were not assessed.

(v) Phenotype characterization: Killip class and clinical heart-failure status at admission were not recorded; heart-failure phenotype rests on LVEF category and NT-proBNP.

(vi) Single center and single drug: The study was performed in one center, and 249 of 251 treated patients received empagliflozin, so no inference about other SGLT2 inhibitors is possible.

(vii) Follow-up horizon: Six months captures ventricular, but not atrial remodeling, and late events are not covered.

4.6. Clinical implications and next steps

For the clinician discharging a patient after infarction on early empagliflozin, the practical message is that the familiar markers—prior infarction, continued smoking, a dilated ventricle, advanced diastolic filling abnormalities, low body weight, and the need for a mineralocorticoid-receptor antagonist—still identify the patients most likely to return within six months, and that the drug does not remove this risk. Explainable model output (SHAP; Figure 3) makes these contributions visible at the individual level[31], but transparency does not substitute for validation: the barriers to adoption of AI-based decision support—data quality, workflow integration, and clinician trust—are well described[13,16,18,41,42], and the present model must first be shown to hold in an external cohort and, ideally, in a randomized or multicenter setting that also settles the treatment question.[19]

5. Conclusion

In 311 patients with AMI, 251 of whom received early empagliflozin, six-month MACE occurred in 26.7%. First, eight predictors selected by cross-validated backward

elimination—body mass index, prior infarction, current smoking, mineralocorticoid-receptor antagonist use, peak A velocity, LVEDV, empagliflozin exposure, and discretized total cholesterol—predicted MACE with a mean cross-validated AUC of 0.674 (95% CI 0.60–0.74) and good calibration in an L2-regularized logistic regression, which performed at least as well as nine ML classifiers; the fully nested AUC of 0.587 shows that the estimate depends on how predictors are chosen, and the model is not ready for clinical use without external validation. Second, after propensity-score matching on 20 covariates, early empagliflozin was associated with a non-significant reduction in MACE (odds ratio 0.79, 95% CI 0.32–1.98), consistent across sensitivity analyses and concentrated in patients with preserved LVEF; this association is hypothesis-generating and does not alter the neutral randomized evidence. Third, LVEF recovery was related to baseline systolic phenotype rather than to glycemic control, and ventricular reverse remodeling preceded atrial remodeling, so ventricular volumes and NT-proBNP, not left-atrial volume, are the appropriate markers of early treatment response.

Acknowledgments

None.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Ministry of Innovative Development of the Republic of Uzbekistan under grant No. AL-9925025470-R3.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

Conceptualization: Raisa Trigulova
Formal analysis: Raisa Trigulova, Shokhista Akhmedova
Investigation: Dilafruz Akhmedova, Alisher Ikramov
Methodology: Shakhnoza Mukhtarova, Raisa Trigulova, Alisher Ikramov, Dilafruz Akhmedova
Writing–original draft: Dilafruz Akhmedova
Writing–review & editing: Raisa Trigulova, Alisher Ikramov

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Republican Specialized Scientific and Practical Medical Center of Cardiology. Written informed consent was obtained from all participants prior to enrollment.

Consent for publication

Consent for publication was not applicable because this study used anonymized data, and no identifiable individual information is included in this manuscript.

Availability of data

The datasets generated and/or analyzed during the current study, together with the full analysis code (R and Python), are available from the corresponding author upon reasonable request, subject to institutional data-sharing and patient-privacy requirements.

References

1. Timmis A, Vardas P, Townsend N, *et al.* European Society of Cardiology: cardiovascular disease statistics 2021. *Eur Heart J*. 2022;43(8):716-799.

   doi: 10.1093/eurheartj/ehab892

2. Byrne RA, Rossello X, Coughlan JJ, *et al.* 2023 ESC guidelines for the management of acute coronary syndromes. *Eur Heart J*. 2023;44(38):3720-3826.

   doi: 10.1093/eurheartj/ehad191

3. McMurray JJV, Solomon SD, Inzucchi SE, *et al.* Dapagliflozin in patients with heart failure and reduced ejection fraction. *N Engl J Med*. 2019;381(21):1995-2008.

   doi: 10.1056/NEJMoa1911303

4. Packer M, Anker SD, Butler J, *et al.* Cardiovascular and renal outcomes with empagliflozin in heart failure. *N Engl J Med*. 2020;383(15):1413-1424.

   doi: 10.1056/NEJMoa2022190

5. Heerspink HJL, Stefánsson BV, Correa-Rotter R, *et al.* Dapagliflozin in patients with chronic kidney disease. *N Engl J Med*. 2020;383(15):1436-1446.

   doi: 10.1056/NEJMoa2024816

6. von Lewinski D, Kolesnik E, Tripolt NJ, *et al.* Empagliflozin in acute myocardial infarction: the EMMY trial. *Eur Heart J*. 2022;43(41):4421-4432.

   doi: 10.1093/eurheartj/ehac494

7. James S, Erlinge D, Storey RF, *et al.* Dapagliflozin in myocardial infarction without diabetes or heart failure. *NEJM Evid*. 2024;3(2).

   doi: 10.1056/EVIDoa2300286

8. Butler J, Jones WS, Udell JA, *et al.* Empagliflozin after acute myocardial infarction. *N Engl J Med*. 2024;390(16):1455-1466.

   doi: 10.1056/NEJMoa2314051

9. Carberry J, Petrie MC, Lee MMY, *et al.* Empagliflozin to prevent worsening of left ventricular volumes and systolic function after myocardial infarction (EMPRESS-MI). *Eur J Heart Fail*. 2024;27(3):566-576.

   doi: 10.1002/ejhf.3560

10. Santos-Gallego CG, Vargas-Delgado AP, Requena-Ibanez JA, *et al.* Randomized trial of empagliflozin in nondiabetic patients with heart failure and reduced ejection fraction (EMPA-TROPISM). *J Am Coll Cardiol*. 2021;77(3):243-255.

    doi: 10.1016/j.jacc.2020.11.008

11. Lee MMY, Brooksbank KJM, Wetherall K, *et al.* Effect of empagliflozin on left ventricular volumes in patients with type 2 diabetes, or prediabetes, and heart failure with reduced ejection fraction (SUGAR-DM-HF). *Circulation*. 2021;143(6):516-525.

    doi: 10.1161/circulationaha.120.052186

12. Hammer A, Sossalla S, Sulzgruber P. Early initiation of SGLT2 inhibitors after acute myocardial infarction. *Eur Heart J Cardiovasc Pharmacother*. 2025;11(4):365-366.

    doi: 10.1093/ehjcvp/pvaf020

13. Umar BU, Ajao LA, Dogo EM, Ajao FJ, Atama M. Artificial intelligence model for prediction of cardiovascular disease: an empirical study. *Artif Intell Health*. 2023;1(1):42-56.

    doi: 10.36922/aih.1746

14. Baashar Y, Alkawsi G, Alhussian H, *et al.* Effectiveness of artificial intelligence models for cardiovascular disease prediction: network meta-analysis. *Comput Intell Neurosci*. 2022;2022:5849995.

    doi: 10.1155/2022/5849995

15. Teshale AB, Htun HL, Vered M, Owen AJ, Freak-Poli R. A systematic review of artificial intelligence models for time-to-event outcome applied in cardiovascular disease risk prediction. *J Med Syst*. 2024;48(1):68.

    doi: 10.1007/s10916-024-02087-7

16. Aşkın L, Polat E, Hoşoğlu Y, Tanrıverdi O. The application of artificial intelligence in the field of cardiovascular diseases focuses on both diagnostic and therapeutic aspects. *Exp Appl Med Sci*. 2024;5(2):22-35.

    doi: 10.46871/eams.1438927

17. Xiang Y, Liu D, Guo L, *et al.* Accuracy of machine learning models for early prediction of major cardiovascular events post myocardial infarction: a systematic review and meta-analysis. *Rev Cardiovasc Med*. 2025;26(6):37224.

    doi: 10.31083/RCM37224

18. Splinter R. Artificial intelligence in cardiac rhythm diagnostics and management: challenges and opportunities. *Artif Intell Health*. 2025;2(3):107-124.

    doi: 10.36922/aih.8468

19. Collins GS, Moons KGM, Dhiman P, *et al.* TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385.

    doi: 10.1136/bmj-2023-078378

20. Lang RM, Badano LP, Mor-Avi V, *et al.* Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the ASE and the EACVI. *J Am Soc Echocardiogr*. 2015;28(1):1-39.

    doi: 10.1016/j.echo.2014.10.003

21. Nagueh SF, Smiseth OA, Appleton CP, *et al.* Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the ASE and the EACVI. *J Am Soc Echocardiogr*. 2016;29(4):277-314.

    doi: 10.1016/j.echo.2016.01.011

22. McDonagh TA, Metra M, Adamo M, *et al.* 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J*. 2021;42(36):3599-3726.

    doi: 10.1093/eurheartj/ehab368

23. Hirshberg B, Katz A. Cardiovascular outcome studies with novel antidiabetes agents: scientific and operational considerations. *Diabetes Care*. 2013;36(suppl 2):S253-S258.

    doi: 10.2337/dcS13-2041

24. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011;46(3):399-424.

    doi: 10.1080/00273171.2011.568786

25. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*. 2006;7(1):91.

    doi: 10.1186/1471-2105-7-91

26. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32.

    doi: 10.1023/A:1010933404324

27. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016:785-794.

    doi: 10.1145/2939672.2939785

28. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol*. 2016;74:167-176.

    doi: 10.1016/j.jclinepi.2015.12.005

29. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247.

    doi: 10.1016/j.jclinepi.2015.04.005

30. Riley RD, Snell KIE, Ensor J, *et al.* Minimum sample size for developing a multivariable prediction model: part II – binary and time-to-event outcomes. *Stat Med*. 2019;38(7):1276-1296.

doi: 10.1002/sim.7992

31. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst (NeurIPS 2017)*. 2017;30:4765-4774.

doi: 10.48550/arXiv.1705.07874

32. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). *BMJ*. 2015;350:g7594.

doi: 10.1136/bmj.g7594

33. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925-1931.

doi: 10.1093/eurheartj/ehu207

34. Mach F, Koskinas KC, Roeters van Lennep JE, *et al.* 2025 focused update of the 2019 ESC/EAS guidelines for the management of dyslipidaemias. *Atherosclerosis*. 2025;409:120479.

doi: 10.1016/j.atherosclerosis.2025.120479

35. Kato ET, Ono K, Lewis BS. SGLT2 inhibitors in acute myocardial infarction: what can we learn from the DAPA-MI trial? *Eur Heart J Cardiovasc Pharmacother*. 2024;10(2):95-97.

doi: 10.1093/ehjcvp/pvad099

36. Hernandez AF, Udell JA, Jones WS, *et al.* Effect of empagliflozin on heart failure outcomes after acute myocardial infarction: insights from the EMPACT-MI trial. *Circulation*. 2024;149(21):1627-1638.

doi: 10.1161/circulationaha.124.069217

37. Tripolt NJ, Kolesnik E, Benedikt M, *et al.* Timing of SGLT2 inhibitor initiation after acute myocardial infarction. *Cardiovasc Diabetol*. 2023;22(1):269.

doi: 10.1186/s12933-023-02000-5

38. Trigulova R, Akhmedova D, Akhmedova S, Mukhtarova S, Ikramov A, Alimova D. Early use of sodium-glucose cotransporter 2 inhibitors after acute myocardial infarction: 2023-2025 data and clinical implications. *Cardiol Belarus*. 2026;18(2):312-320.

doi: 10.34883/PI.2026.18.2.011

39. Ferrannini E, Baldi S, Frascerra S, *et al.* Shift to fatty substrate utilization in response to sodium-glucose cotransporter 2 inhibition. *Diabetes*. 2016;65(5):1190-1195.

doi: 10.2337/db15-1356

40. Benedikt M, Mangge H, Aziz F, *et al.* Impact of the SGLT2 inhibitor empagliflozin on inflammatory biomarkers after acute myocardial infarction: a post-hoc analysis of the EMMY trial. *Cardiovasc Diabetol*. 2023;22(1):166.

doi: 10.1186/s12933-023-01904-6

41. Hovenga E, Atalag K. Optimizing electronic health records to support artificial intelligence. *Artif Intell Health*. 2024;1(3):10-25.

doi: 10.36922/aih.3056

42. Zribi A, Ayaad O. Challenges in incorporating artificial intelligence into daily healthcare practice. *Artif Intell Health*. 2025;3(2):025350070.

doi: 10.36922/AIH025350070