Original Research Article | **Open Access**

The Journal of Cardiovascular Aging

Qin *et al. J Cardiovasc Aging.* 2026;6:53 **DOI:**10.20517/jca.2026.111

Machine-learning risk stratification for 12-month mortality and SGLT2 inhibitor-related absolute risk differences in elderly patients with heart failure and type 2 diabetes

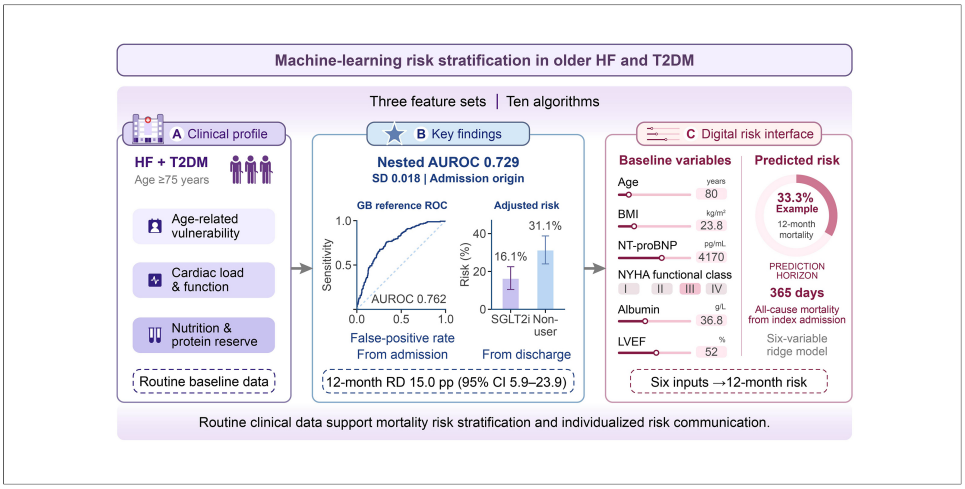
Zhen Qin[1,2,3,4,5,#], Yitong Jin[1,2,3,4,5,#], Xiang Li[1,2,3,4,5,#], Bo Lan[1,2,3,4,5], Kaiwen Wu[1,2,3,4,5], Jieru Zou[6], Tiansheng Liu[7], Yongshun Wang[8], Naijin Zhang[9], Jinying Zhang[1,2,3,4,5], Jie Yuan[10], Junnan Tang[1,2,3,4,5]

Keywords:
Heart failure, type 2 diabetes, machine learning, risk stratification, SGLT2 inhibitor, digital health

Citation: Qin Z, Jin Y, Li X, Lan B, Wu K, Zou J, Liu T, Wang Y, Zhang N, Zhang J, Yuan J, Tang J. Machine-learning risk stratification for 12-month mortality and SGLT2 inhibitor-related absolute risk differences in elderly patients with heart failure and type 2 diabetes. *J Cardiovasc Aging.* 2026;6:53. https://dx.doi.org/10.20517/jca.2026.111

Received: 23 Jul 2026 **First Decision:** 16 Sep 2026 **Revised:** 9 Oct 2026 **Accepted:** 15 Oct 2026 **Published:** 2 Nov 2026

Academic Editor: Xiaoqiang Tang
Copy Editor: Fangling Lan
Production Editor: Fangling Lan



Abstract

Aim: Older adults with heart failure (HF) and type 2 diabetes mellitus (T2DM) face high 12-month mortality. We developed a routine-variable model for mortality prediction and explored SGLT2 inhibitor (SGLT2i)-related absolute risk differences across model-defined risk strata.

Methods: This retrospective study included 500 hospitalized patients aged ≥ 75 years with HF and T2DM. Three feature sets (guideline-informed core, primary analysis, and complete imputed candidate) and ten algorithms were compared using repeated five-fold outer/four-fold inner cross-validation with training-fold preprocessing. Cox and multiple-imputation sensitivity analyses were performed. SGLT2i associations were assessed using propensity-score overlap weighting.

[1]Department of Cardiology, First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, Henan, China.
[2]Key Laboratory of Cardiac Injury and Repair of Henan Province, Zhengzhou 450052, Henan, China.
[3]Henan Province Clinical Research Center for Cardiovascular Diseases, Zhengzhou 450052, Henan, China.
[4]State Key Laboratory of Metabolic Dysregulation & Prevention and Treatment of Esophageal Cancer, Tianjian Laboratory of Advanced Biomedical Sciences, Academy of Medical Sciences, Zhengzhou 450001, Henan, China.
[5]Henan Province Academy of Cardiovascular Disease Prevention and Treatment, Zhengzhou 450052, Henan, China.
[6]Department of Cardiology, Fuwai Shenzhen Hospital, Chinese Academy of Medical Sciences, Shenzhen 518055, Guangdong, China.
[7]Southern University of Science and Technology Hospital, Shenzhen 518055, Guangdong, China.
[8]Department of Cardiology, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University, The First Affiliated

Hospital, Southern University of Science and Technology), Shenzhen 518020, Guangdong, China.
[9]Department of Cardiology, The First Hospital of China Medical University, Shenyang 110001, Liaoning, China.
[10]Department of Cardiology, Shenzhen People's Hospital (First Affiliated Hospital of Southern University of Science and Technology, Second Clinical Medical College of Jinan University), Shenzhen 518020, Guangdong, China.
[#]Authors contributed equally.

Correspondence to: Dr. Jie Yuan, Department of Cardiology, Shenzhen People's Hospital (First Affiliated Hospital of Southern University of Science and Technology, Second Clinical Medical College of Jinan University), Shenzhen 518020, Guangdong, China. E-mail: yuanjie@szhospital.com; Dr. Junnan Tang, Department of Cardiology, First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, Henan, China; Key Laboratory of Cardiac Injury and Repair of Henan Province, Zhengzhou 450052, Henan, China; Henan Province Clinical Research Center for Cardiovascular Diseases, Zhengzhou 450052, Henan, China; State Key Laboratory of Metabolic Dysregulation & Prevention and Treatment of Esophageal Cancer, Tianjian Laboratory of Advanced Biomedical Sciences, Academy of Medical Sciences, Zhengzhou 450001, Henan, China; Henan Province Academy of Cardiovascular Disease Prevention and Treatment, Zhengzhou 450052, Henan, China. E-mail: fcctangjn@zzu.edu.cn

Results: The binary modelling cohort included 434 patients with 167 deaths (38.5%). The nested selection workflow achieved a mean area under the receiver operating characteristic curve (AUROC) of 0.729, area under the precision-recall curve of 0.612, and Brier score of 0.206. Cox sensitivity analysis yielded a censoring-adjusted 365-day AUROC of 0.755; multiple imputation gave an AUROC of 0.758 for the fixed gradient-boosting reference. Among 381 patients, weighted 12-month mortality was 16.1% in SGLT2i users *vs.* 31.1% in non-users, a difference of 15.0 percentage points (95%CI: 5.9-23.9). Stratum-specific differences were 6.5, 18.1, and 19.7 percentage points, with wide confidence intervals. A six-variable interactive interface supported risk communication.

Conclusion: Routine clinical variables supported moderately discriminative mortality risk stratification under nested validation. SGLT2i use was associated with lower adjusted 12-month mortality overall; larger studies are needed to establish whether associations differ across risk strata. External and prospective validation remains necessary.

INTRODUCTION

Heart failure (HF) and type 2 diabetes mellitus (T2DM) frequently coexist in older adults and reinforce each other's adverse prognosis through overlapping cardiac, metabolic, renal and geriatric pathways. In patients aged 75 years or older, the clinical phenotype is rarely driven by one abnormal test alone[1,2]. Symptoms may reflect congestion, impaired systolic or diastolic reserve, renal dysfunction, anaemia, malnutrition, frailty, polypharmacy, social vulnerability, or limited tolerance of guideline-directed therapy[3]. These interacting processes make bedside risk estimation difficult, especially when the clinician must decide which patients need more intensive follow-up, earlier treatment optimization, or careful medication prioritization[4].

Several variables that are routinely recorded during admission can capture different components of this vulnerability. N-terminal pro-B-type natriuretic peptide (NT-proBNP) reflects haemodynamic stress, New York Heart Association (NYHA) grade summarizes functional limitation, left ventricular ejection fraction (LVEF) measures systolic reserve, while body mass index (BMI) and albumin provide accessible signals of nutritional reserve, wasting and systemic illness. However, these signals are usually interpreted one at a time. In elderly patients with HF and T2DM, a single-variable approach can under-recognize compound risk, because moderate abnormalities across several domains may together imply a substantially higher short-term mortality risk than any one variable suggests[4,5].

SGLT2 inhibitors have become central to contemporary HF and diabetes care. Major randomized trials and meta-analyses support their benefits across reduced, mildly reduced, and preserved ejection fraction phenotypes, with additional relevance across the broader cardiometabolic and kidney-disease spectrum[6-10]. These trials established robust relative treatment effects, but clinical decision-making in a frail elderly population also requires attention to absolute risk. A patient with high baseline mortality risk may derive a

larger absolute benefit from an intervention with the same relative effect, whereas a lower-risk patient may have a smaller absolute difference despite similar proportional efficacy[11].

This distinction is particularly relevant in elderly, comorbid patients. Renal instability, hypotension, poor oral intake, recurrent infection, cost, treatment adherence, and clinician concern about tolerability may affect whether SGLT2 inhibitors are prescribed at discharge. Therefore, an analytically useful model should do more than rank patients by mortality probability. It should also help translate predicted risk into a clinically understandable framework for treatment-priority discussions, while keeping the boundary between prognosis and treatment association clear.

Machine-learning approaches can integrate non-linear relations and interactions among routine variables and have been used for mortality prediction and HF classification in older cardiometabolic populations[12]. Nevertheless, prediction studies are often difficult to reuse clinically when they focus mainly on discrimination metrics, omit calibration and decision-curve information, use unstable high-dimensional feature sets, or do not show how predicted risk may change practical decisions. For elderly HF-T2DM patients, the main methodological need is a transparent, internally validated model that remains close to variables already available in routine care[13,14].

We therefore developed and internally validated a routine-variable risk-stratification model for 12-month all-cause mortality in elderly HF-T2DM patients. We benchmarked ten supervised learning algorithms, compared feature-set strategies, audited collinearity in a compact clinical variable set, evaluated calibration and decision utility, and examined whether model-defined risk strata were associated with different observed absolute differences related to SGLT2 inhibitor use. The overall study logic, from patient selection to risk modelling and treatment-oriented interpretation, is summarized in Figure 1.

MATERIAL AND METHODS

Study design and participants

This was a single-center retrospective cohort study conducted at the First Affiliated Hospital of Zhengzhou University. Consecutive patients undergoing their first eligible hospitalization in the Department of Cardiology between 1 January 2020 and 31 December 2024 were screened. Eligible patients were aged ≥ 75 years and had coexisting HF and T2DM. HF was defined by compatible symptoms or signs together with natriuretic peptide elevation and/or cardiac imaging evidence; T2DM was defined by a documented history of T2DM, glucose-lowering therapy, or laboratory criteria consistent with diabetes. Inclusion additionally required sufficient baseline information for core-variable assessment. Patients were excluded if expected survival was < 24 h; heart transplantation or ventricular-assist-device implantation occurred during the hospitalization; active malignant disease, severe hematological disease, severe acute infection, or systemic inflammatory disease was present; or clinical data were insufficient for core-variable assessment.

Index admission was defined as the first hospitalization during the study period that met all eligibility criteria; subsequent readmissions were not treated as separate index records. The cohort-construction process and the subsequent modelling and treatment-stratification workflow are shown in Figure 1, with baseline characteristics summarized in Table 1. Follow-up was obtained from the electronic medical record and structured telephone interviews. Follow-up time was calculated from admission to death or the last confirmed contact. Patients whose 12-month status could not be determined were not entered into the 12-month binary endpoint set, whereas patients who died within 365 days were included as events even when the full interval had not elapsed.

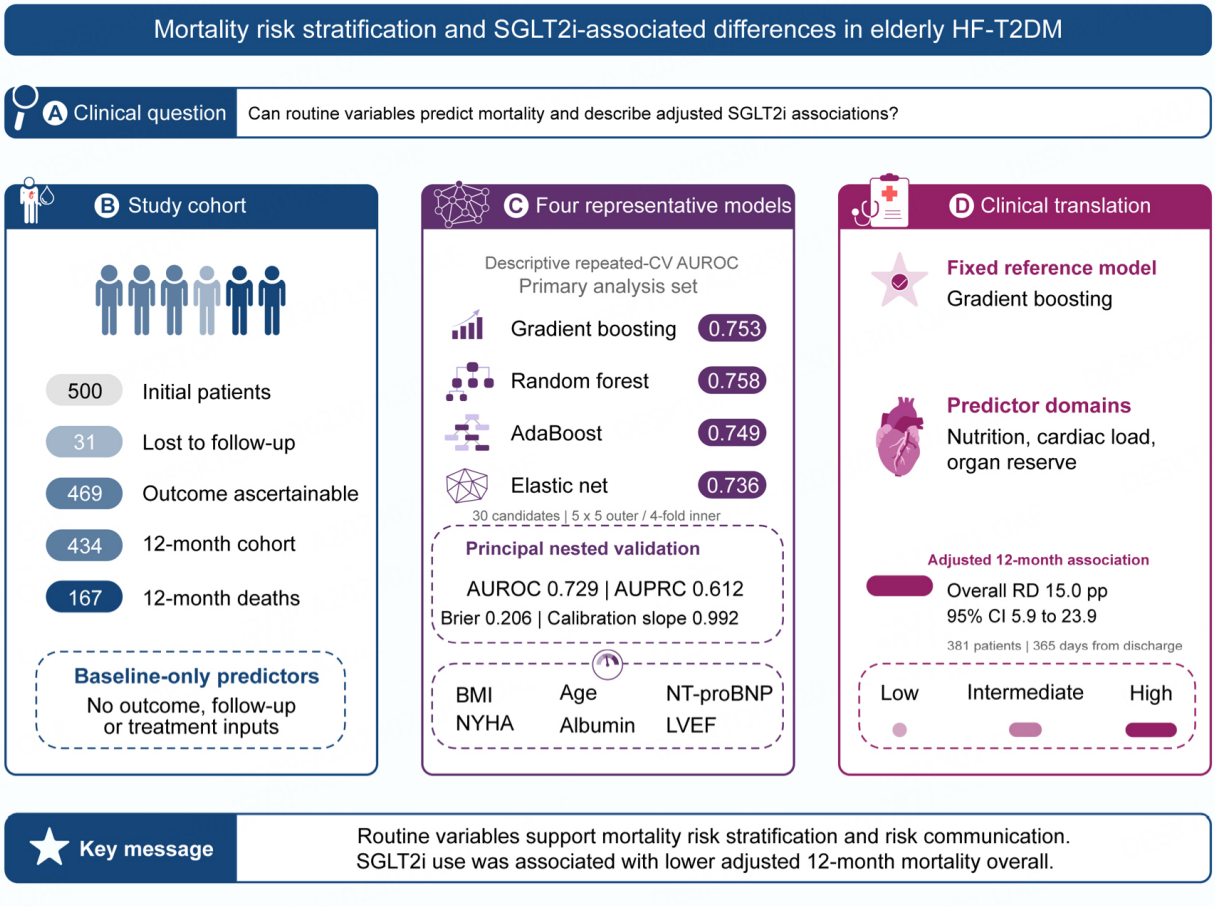


Figure 1. Study overview and principal findings. (A) Clinical question: routine-variable mortality prediction and adjusted SGLT2i-associated differences in older patients with HF and T2DM. (B) Cohort construction: 500 eligible patients, 469 with ascertainable follow-up and 434 with determinable 12-month status, including 167 deaths. (C) Four representative descriptive benchmarks from the ten-algorithm comparison, together with principal mean repeat estimates from nested selection across 30 algorithm-feature-set combinations. These nested estimates evaluate the selection workflow, not one fixed algorithm. The six variables shown below are the display inputs, not the complete predictor roster. (D) Fixed gradient-boosting reference, predictor domains, and the overall overlap-weighted mortality difference in 381 discharge patients through day 365. The low-, intermediate-, and high-risk symbols identify the stratification framework; differential treatment associations are not established. HF: Heart failure; T2DM: type 2 diabetes mellitus; AUROC: area under the receiver operating characteristic curve; AUPRC: area under the precision-recall curve; CV: cross-validation; RD: non-user minus user mortality; CI: confidence interval; pp: percentage points; BMI: body mass index; NYHA: New York Heart Association; LVEF: left ventricular ejection fraction; NT-proBNP: N-terminal pro-B-type natriuretic peptide; SGLT2i: sodium-glucose cotransporter 2 inhibitor.

The study was approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University (approval No. 2025-KY-1217-001). The requirement for informed consent was waived because of the retrospective design. The study was registered with the Chinese Clinical Trial Registry (ChiCTR2500109213).

Candidate predictors and missing data

Candidate predictors comprised routinely collected baseline demographics, vital signs, comorbidities, laboratory measures, and echocardiographic variables available before discharge treatment assignment. Baseline values were defined as the first recorded measurements after admission during the index hospitalization. Endpoint labels, follow-up status, survival time, and discharge medications were excluded from model fitting so that prediction remained anchored to baseline prognosis.

Missing baseline covariates were handled using single imputation within each training fold[15]. Numeric variables were completed using training-fold medians with missingness indicators, and categorical variables using the training-fold mode. Scaling and one-hot encoding were also fitted only on training data and

Table 1. Baseline characteristics according to 12-month all-cause mortality status

Characteristic	12-month death (*n* = 167)	12-month survival (*n* = 267)	Cliff's δ/OR (95%CI)	*P* value
Demographic characteristics				
Age, years	81.0 (78.0-84.0)	79.0 (77.0-82.0)	δ = 0.226 (0.115 to 0.330)	< 0.001
Male sex, n (%)	71 (42.5%)	114 (42.7%)	OR = 0.99 (0.67 to 1.47)	1.000
BMI, kg/m^2	22.7 (19.7-25.4)	24.5 (22.1-27.1)	δ = -0.291 (-0.408 to -0.161)	< 0.001
Clinical status				
Admission SBP, mmHg	130.0 (112.0-145.2)	134.0 (121.0-148.2)	δ = -0.106 (-0.218 to 0.005)	0.066
Admission pulse, beats/min	81.0 (75.0-95.0)	78.0 (68.0-89.0)	δ = 0.196 (0.084 to 0.302)	< 0.001
NYHA class III-IV, n (%)	144 (86.2%)	189 (70.8%)	OR = 2.58 (1.55 to 4.32)	< 0.001
LVEF, %	49.0 (40.0-58.0)	55.0 (43.0-61.0)	δ = -0.223 (-0.329 to -0.113)	< 0.001
Laboratory measurements				
NT-proBNP, pg/mL	7685 (3425-18762)	2950 (1270-7420)	δ = 0.417 (0.316 to 0.515)	< 0.001
eGFR, mL/min/1.73 m^2	49.3 (30.2-73.0)	61.3 (42.6-81.5)	δ = -0.225 (-0.337 to -0.119)	< 0.001
Albumin, g/L	35.4 (32.5-38.1)	37.7 (34.5-40.4)	δ = -0.278 (-0.377 to -0.170)	< 0.001
HbA1c, %	7.4 (6.8-8.5)	7.5 (6.9-8.5)	δ = -0.045 (-0.168 to 0.084)	0.473
Haemoglobin, g/L	111.0 (97.0-129.0)	120.0 (106.0-133.3)	δ = -0.183 (-0.288 to -0.067)	0.001
Comorbidities				
Hypertension, n (%)	139 (83.2%)	237 (88.8%)	OR = 0.63 (0.36 to 1.10)	0.111
Chronic kidney disease, n (%)	65 (38.9%)	83 (31.1%)	OR = 1.41 (0.94 to 2.12)	0.097
Atrial fibrillation, n (%)	47 (28.1%)	71 (26.6%)	OR = 1.08 (0.70 to 1.67)	0.740
Previous myocardial infarction, n (%)	71 (42.5%)	53 (19.9%)	OR = 2.99 (1.94 to 4.59)	< 0.001
Stroke, n (%)	60 (35.9%)	78 (29.2%)	OR = 1.36 (0.90 to 2.05)	0.168

Continuous variables are median (interquartile range) and were compared using the two-sided Mann-Whitney U test; effect estimates are Cliff's δ. Categorical variables are n (%) and were compared using the two-sided Fisher exact test; effect estimates are odds ratios (ORs). Positive δ or OR > 1 indicates a higher value or prevalence in patients who died within 12 months. Values are summarized before imputation. NYHA: New York Heart Association; BMI: body mass index; LVEF: left ventricular ejection fraction; NT-proBNP: N-terminal pro-B-type natriuretic peptide; eGFR: estimated glomerular filtration rate; SBP: systolic blood pressure; HbA1c: glycated haemoglobin.

applied unchanged to validation data, including within the inner validation loop. Endpoint and treatment-exposure fields were not imputed.

Original missingness was calculated before imputation. Variables with at least 30% missingness were excluded from the primary analysis feature set but retained in the complete imputed candidate feature set. This design prioritized more consistently observed predictors in the primary analysis while preserving a broader sensitivity benchmark. The three feature sets contained 7, 51, and 58 variables, respectively. Feature rosters used source-wide missingness, whereas imputation was fitted within training folds.

A multiple-imputation sensitivity analysis evaluated the fixed gradient-boosting model with the primary analysis feature set. Twenty stochastic chained-equations imputations of numeric predictors were generated within each of the 25 outer training samples using Bayesian ridge conditional models with posterior sampling; the fitted models were then applied to the corresponding validation data without using validation outcomes. Complete categorical variables were entered as one-hot auxiliary predictors. Skewed positive laboratory variables were log-transformed for imputation and restored to their recorded scales before model fitting; imputation support bounds were learned from training observations. Observed values were retained, and missingness indicators were included. Twenty held-out probabilities per repeat were averaged for

evaluation. Ten- *vs*. twenty-imputation averages assessed Monte Carlo stability. This analysis tested the fixed model's missing-data sensitivity and did not repeat selection across the 30 candidates.

Model development and validation

The primary outcome was all-cause mortality within 365 days of admission. Three feature sets were compared. The first was a guideline-informed core clinical feature set designed to capture major prognostic domains, including demographic characteristics, ischaemic heart disease history, cardiac function, haemodynamic stress and renal function. It comprised age, sex, albumin, LVEF, NT-proBNP, estimated glomerular filtration rate (eGFR) and previous myocardial infarction. Albumin was additionally included to reflect nutritional reserve and overall systemic disease burden. The second was the primary analysis feature set, which included eligible baseline variables with an original missingness rate of < 30% after excluding patient identifiers, outcome and follow-up fields, derived duplicate variables, and discharge treatment variables. The third was the complete imputed candidate feature set, which retained all eligible baseline predictors after missing-data imputation and was used for sensitivity benchmarking. The seven-variable core set was investigator-defined using guideline-informed prognostic domains.

Ten supervised learning algorithms were systematically compared: L2-penalized logistic regression, elastic net, radial-basis-function support vector machine, k-nearest neighbors, Gaussian naive Bayes, decision tree, random forest, extremely randomized trees, gradient boosting, and AdaBoost. Numeric variables were standardized and categorical variables were one-hot encoded within each training fold. The same predefined model settings were applied across all folds and repeats, without fold-specific hyperparameter searching. The full benchmark is presented in Figure 2. Key hyperparameters are provided in Supplementary Table 1. NT-proBNP entered the primary model in its recorded units.

The principal performance evaluation used five-fold stratified outer cross-validation repeated five times. A four-fold inner loop compared the 30 algorithm-feature-set combinations within each outer training sample; highest inner AUROC, followed by Brier score as a secondary ordering criterion, determined the candidate evaluated on the untouched outer fold. Imputation, encoding, and scaling were fitted within the corresponding training data. Repeat-level metrics were summarized as means and standard deviations. Following the initial exploratory comparison in the same cohort, gradient boosting with the primary analysis feature set was retained as a fixed reference model for interpretation and the digital display. Its repeated-validation and averaged out-of-fold estimates were reported separately. Confidence intervals for fixed-reference averaged predictions used 2,000 patient-level bootstrap samples of saved outcome-prediction pairs without repeating model fitting or selection; these intervals are conditional on the saved predictions.

AUROC assessed rank discrimination, whereas area under the precision-recall curve (AUPRC) was emphasized because mortality was not equally distributed across outcome classes. The Brier score, calibration intercept, and calibration slope assessed probability accuracy. Decision-curve analysis for the fixed gradient-boosting reference compared model-guided intensified review with review-all and review-none strategies across descriptive risk thresholds of 0.05-0.80 [Figure 3]. The thresholds represented hypothetical decision preferences for evaluating net benefit, with p/(1-p) specifying the relative penalty assigned to a false-positive decision[16]. The range was a broad descriptive interval, not an established clinical cutoff. The hypothetical action was intensified clinical review or follow-up rather than initiation or withholding of SGLT2i. An operational threshold would require the benefits and burdens of the intended action to be specified.

Time-to-event sensitivity analysis used the seven-variable core clinical set in Cox regression among all patients with verified observation times, including those censored before 365 days. Admission was time zero; deaths through day 365 were events, and other observations were censored at last confirmed contact or day

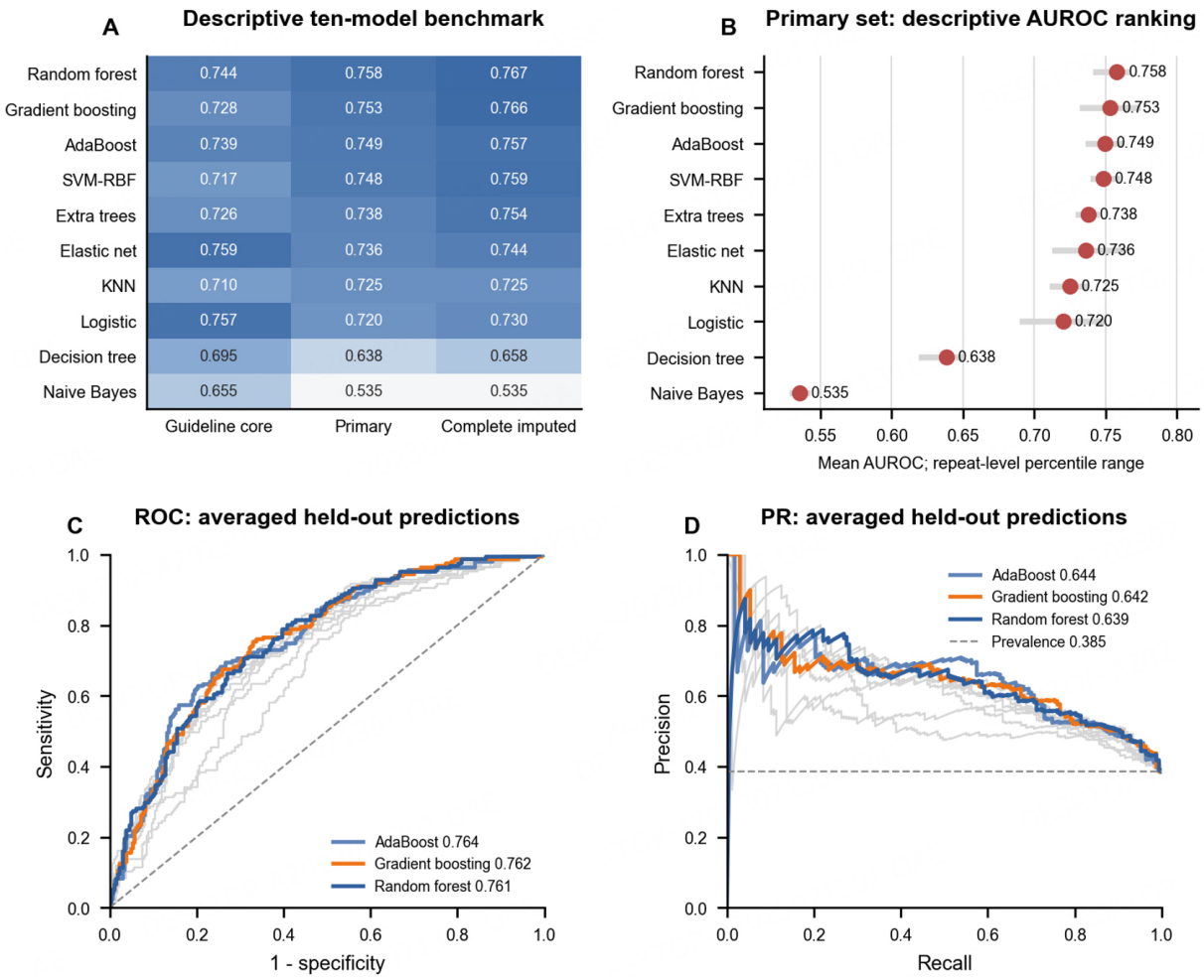


Figure 2. Comparison of ten algorithms across three feature sets. (A) Mean AUROC for each algorithm-feature-set combination across repeated five-fold cross-validation. (B) AUROC ranking within the primary analysis feature set; points show the repeat-level mean and horizontal lines show the 2.5-97.5th percentile range. (C) Receiver operating characteristic curves from stored out-of-fold predictions for the leading models. (D) Precision-recall curves for the same models; the dashed horizontal line indicates the event prevalence. The displayed percentile ranges across five repeats are not confidence intervals. Curves use averaged patient-held-out probabilities. Core: guideline-informed 7-variable comparator; Primary: 51-variable analysis roster; Complete imputed: 58-variable roster, completed within training folds. AUROC: area under the receiver operating characteristic curve; SVM-RBF: radial-basis support vector machine; KNN: k-nearest neighbors.

365. Numeric gaps and missingness indicators were processed within training folds, categorical predictors used reference coding, and NT-proBNP was transformed as ln(1 + x). Repeated five-fold validation provided held-out 365-day probabilities; concordance and cumulative/dynamic AUROC with inverse censoring weights were calculated. Kaplan-Meier estimates used tertiles of averaged held-out Cox probabilities. Rank-based Schoenfeld tests assessed proportional hazards in the final cohort fit. Records without verifiable follow-up time were not treated as known survivors or valid censored observations.

Collinearity and model interpretation

For a separate six-variable interpretability display, variance inflation factors (VIFs) were calculated for age, BMI, NT-proBNP, NYHA grade, albumin, and LVEF. This display was distinct from the guideline-informed core clinical feature set used in model benchmarking. A VIF below 5 was considered acceptable. Permutation importance summarized the decrease in AUROC in five held-out folds of the first repeat, using ten shuffles per variable per fold. NT-proBNP was transformed as ln(1 + x) for the VIF calculation, and missing numeric values were median-imputed for this descriptive audit. VIF estimates are provided in Supplementary Table 2.

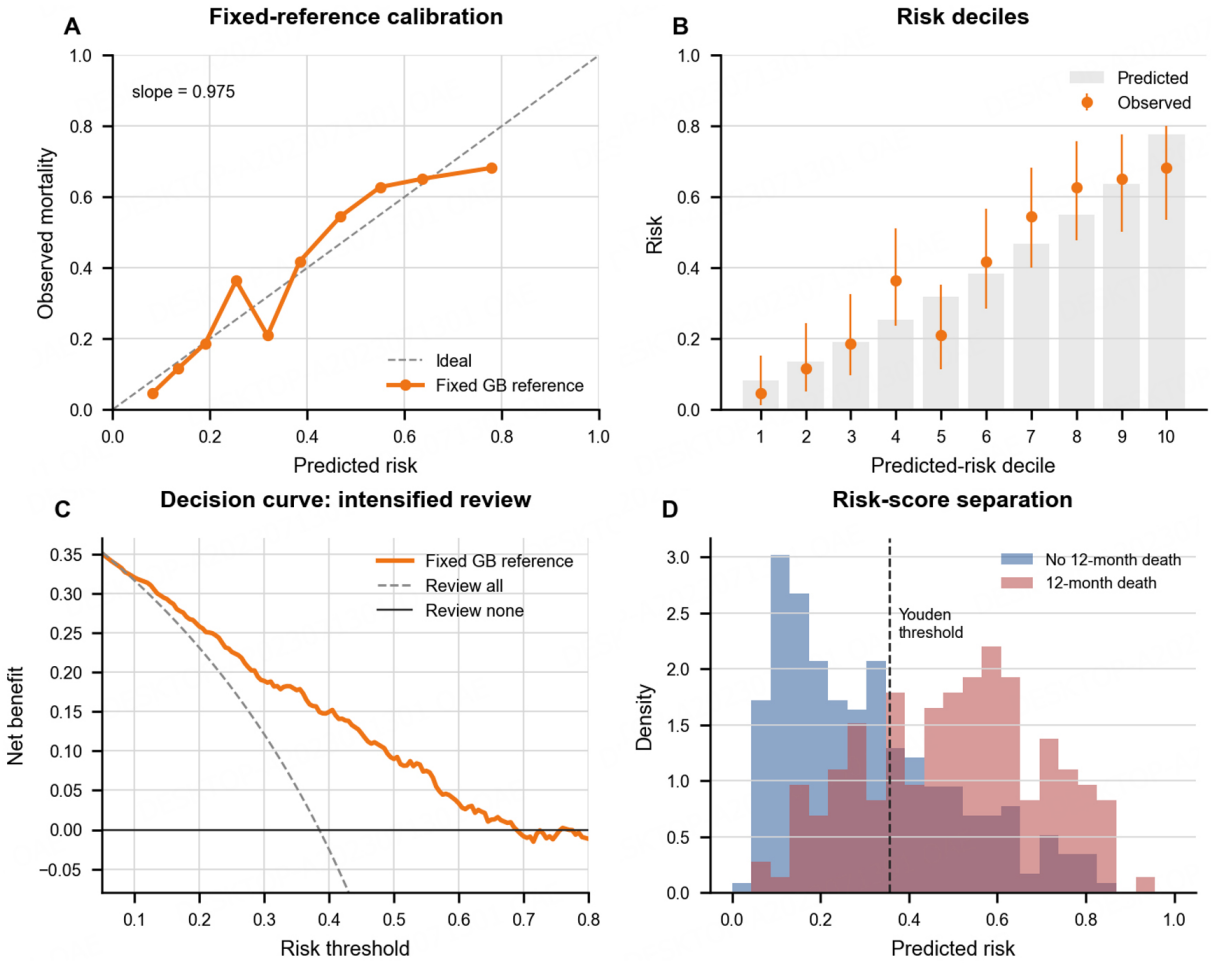


Figure 3. Calibration, decision utility and risk separation of the fixed gradient-boosting reference model. (A) Calibration curve; the dashed line denotes ideal calibration. (B) Mean predicted risk and observed mortality by predicted-risk decile; error bars show Wilson 95% confidence intervals. (C) Decision-curve analysis of hypothetical intensified review or follow-up, comparing the fixed reference with review-all and review-none strategies over thresholds of 0.05-0.80. This is a descriptive display range, not a validated intervention cutoff. (D) Risk distributions among patients with and without death within 365 days of admission; the dashed vertical line marks the descriptive maximum-Youden-index threshold. N = 434, including 167 deaths. Panels use each patient's averaged held-out probability. Density is not a patient count. These fixed-reference plots are separate from the principal nested-workflow estimates in Table 2. GB: Gradient boosting.

The six-variable interface used age, BMI, NT-proBNP, NYHA functional class, albumin, and LVEF, chosen for their clinical interpretability and routine availability. Together, they represented age-related vulnerability, nutritional reserve, cardiac load, symptom severity, and systolic function. The fixed full gradient-boosting reference model supported interpretation, and the six-variable model supported interactive risk communication. The predictor audit and interpretation outputs are presented in Figure 4. The three model-development feature rosters are listed in Supplementary Table 3.

Risk-stratified SGLT2i-related absolute risk differences

The adjusted SGLT2i analysis included 381 patients with documented discharge exposure, ascertainable follow-up and positive post-discharge observation, without requiring survival for 30 days. The endpoint was all-cause mortality within 365 days of discharge; shorter observed follow-up was right-censored. Baseline risk was represented by archived out-of-sample predictions from the fixed gradient-boosting reference model. Tertiles defined three groups of 127 patients; the cohort, score source, grouping rule, and endpoint were fixed before the adjusted comparison. These strata differ from the fitted-score tertiles used in the original 358-patient comparison and were not optimized against treatment differences. Propensity models were fitted

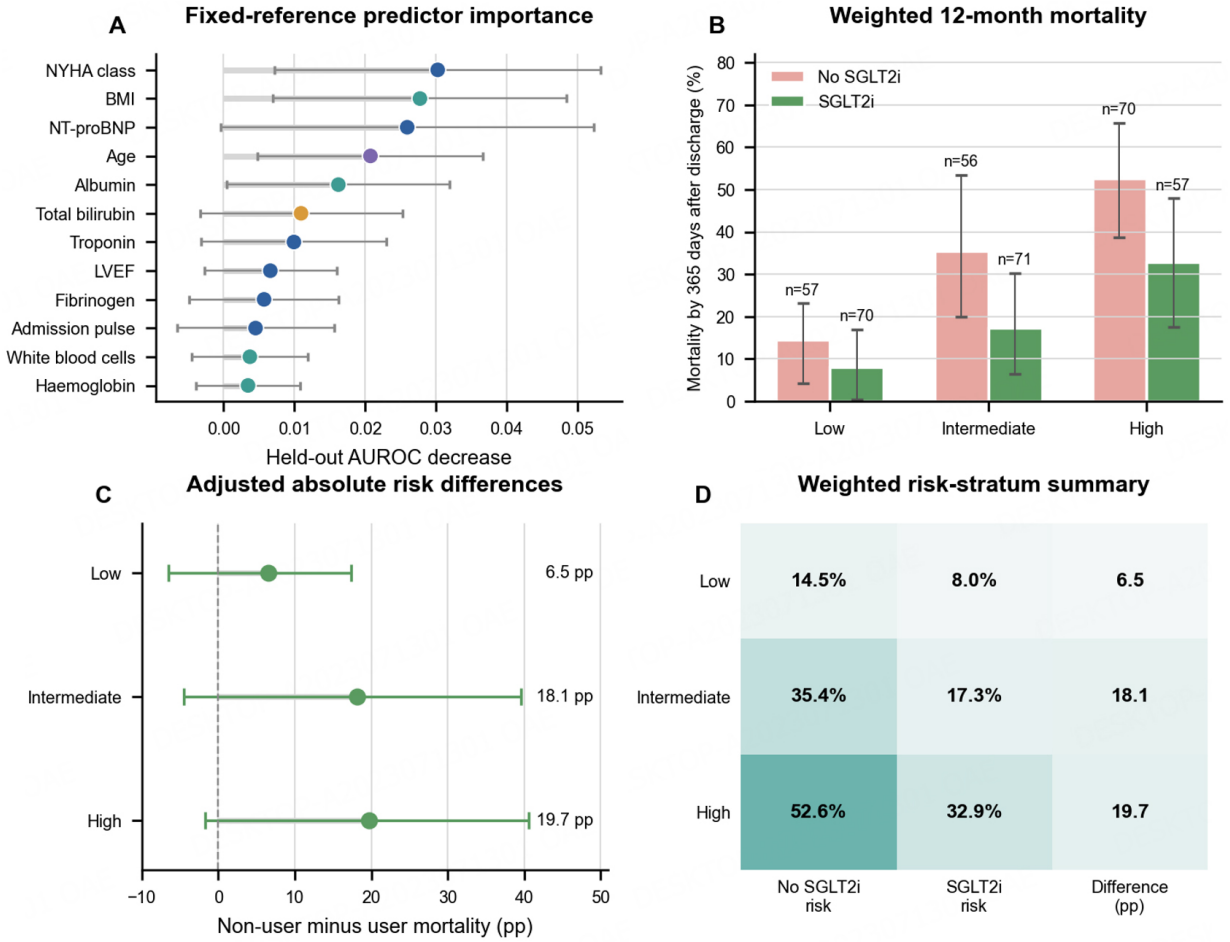


Figure 4. Fixed-reference predictor importance and adjusted SGLT2i-associated 12-month mortality differences. (A) Permutation importance in held-out folds of the fixed gradient-boosting reference: mean AUROC decrease after shuffling each predictor, with one-standard-deviation error bars calculated from 50 fold-by-shuffle values per predictor (five held-out folds, with ten shuffles per fold). (B) Overlap-weighted Kaplan-Meier mortality through 365 days after discharge among SGLT2i users and non-users in the 381-patient treatment cohort. Labels give original group sizes, not effective sample sizes; error bars are bootstrap 95% confidence intervals. Each fixed-reference out-of-sample-score tertile contains 127 patients. (C) Adjusted non-user minus user mortality differences in percentage points, with bootstrap 95% confidence intervals; the dashed line denotes zero. (D) The same weighted mortality and risk-difference point estimates as a numerical summary. Propensity models were fitted overall and within strata; confidence intervals use 1,000 treatment-stratified patient bootstrap samples with propensity refitting and fixed risk scores and strata. The analyses describe observational associations. NYHA: New York Heart Association; pp: percentage points; BMI: body mass index; LVEF: left ventricular ejection fraction; NT-proBNP: N-terminal pro-B-type natriuretic peptide; AUROC: area under the receiver operating characteristic curve; SGLT2i: sodium-glucose cotransporter 2 inhibitor.

overall and within each stratum using admission year, age, sex, BMI, systolic blood pressure, pulse, NYHA class, LVEF, ln(1 + NT-proBNP), eGFR, albumin, glycated haemoglobin, chronic kidney disease, atrial fibrillation and previous myocardial infarction. Missing propensity covariates were median/mode completed. Overlap weights were 1 minus the propensity score for users and the propensity score for non-users. Weighted Kaplan-Meier estimates provided 365-day mortality and non-user minus user risk differences. Covariate balance, propensity overlap, and effective sample size were reported. Confidence intervals used 1,000 treatment-stratified patient bootstrap samples with propensity refitting, conditional on fixed risk scores and strata. Unrecorded prescribing reasons were not imputed.

The original 358-patient, 30-day-survivor comparison summarized deaths over all available follow-up using fixed fitted-model tertiles, Wilson mortality intervals and Newcombe risk-difference intervals. Its absolute risk reduction and number needed to treat summaries are provided separately in Supplementary Table 4.

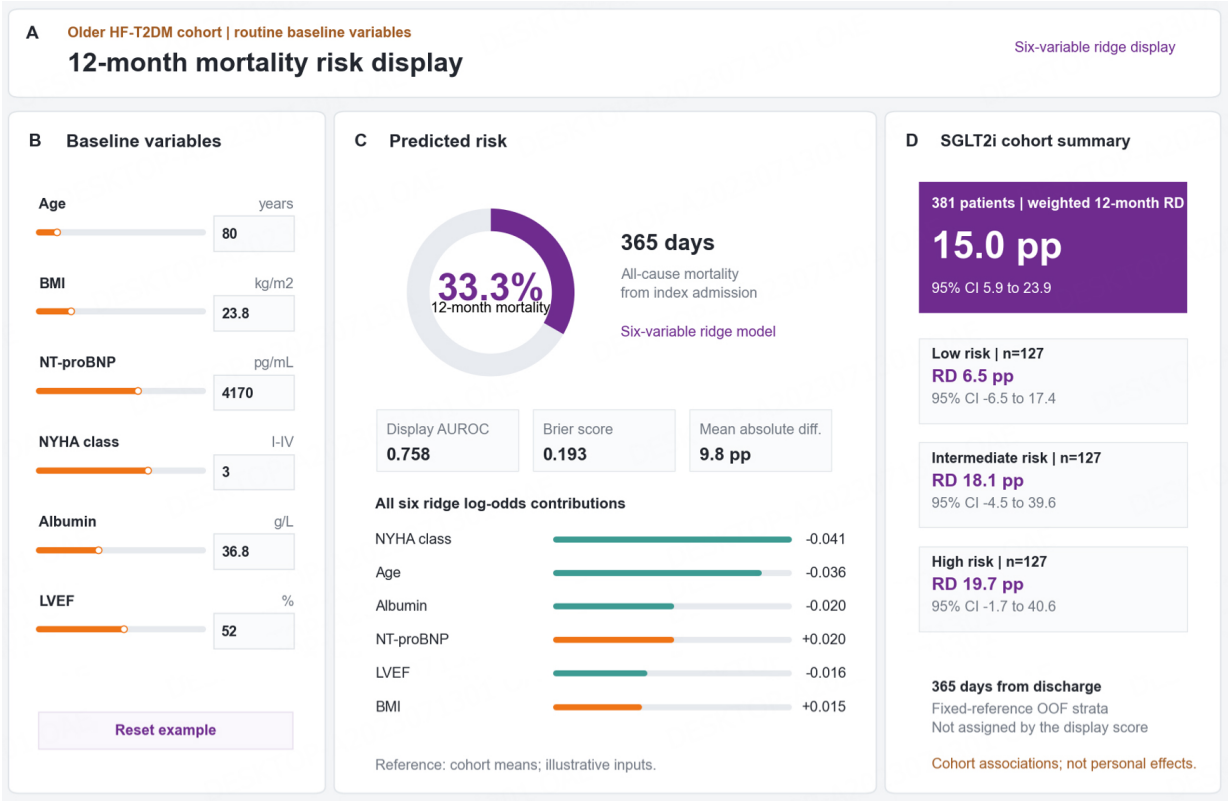


Figure 5. Interactive six-variable mortality-risk display and adjusted cohort summary. Editable schematic rendering of the digital interface. (A) Target population and six-variable ridge-model context. (B) Age, BMI, NT-proBNP, NYHA functional class, albumin and LVEF inputs. (C) Illustrative 365-day admission-origin mortality estimate, held-out performance summaries and all six standardized log-odds contributions to the ridge model. Contributions are relative to cohort means and are distinct from full-reference permutation importance. Mean absolute difference refers to surrogate *vs.* fixed-reference probabilities. (D) Overall and risk-stratified overlap-weighted 12-month post-discharge mortality differences from the 381-patient cohort, with 95% confidence intervals. Those strata use fixed-reference out-of-sample scores and are not assigned by the illustrated six-variable prediction. The inputs are illustrative, not a patient record; cohort treatment differences are not individual treatment-effect predictions. NYHA: New York Heart Association; BMI: body mass index; LVEF: left ventricular ejection fraction; NT-proBNP: N-terminal pro-B-type natriuretic peptide; AUROC: area under the receiver operating characteristic curve; OOF: out-of-fold; RD: non-user minus user mortality; CI: confidence interval; pp: percentage points; SGLT2i: sodium-glucose cotransporter 2 inhibitor.

The corresponding early-death sensitivity analysis retained the original score boundaries [Supplementary Figure 1]. The adjusted analysis used all eligible discharge patients, out-of-sample scores, and a fixed 365-day horizon; the populations, strata, and endpoint definitions are reported separately.

Adjusted treatment estimates are reported in Figure 4B-D and Supplementary Table 5. The Figure 5 interface uses a six-variable ridge surrogate of the fixed gradient-boosting reference model. Within each outer training sample, inner out-of-fold teacher predictions, clipped to 0.01-0.99 and logit-transformed, supplied the targets; the fitted surrogate was evaluated on the outer held-out fold. Ridge alpha was 2.0. Calibration, Brier score, and absolute prediction errors were evaluated together with correlation. Coefficients, intercept, units, ln(1 + NT-proBNP) transformation and standardization are provided in Supplementary Table 6. All analyses were performed in Python.

To support reproducibility, saved output files were used for manuscript tables, figure source data, out-of-fold prediction metrics, risk-stratum summaries and package-level audit reports. This structure allows the reported numbers in the manuscript, figures, supplementary tables, and web display to be traced back to their source files.

Table 2. Principal nested-validation estimates and fixed gradient-boosting reference performance

Quantity	Estimate
Principal evaluation	Nested 30-candidate selection workflow
Nested mean repeat AUROC	0.729 (SD 0.018)
Nested mean repeat AUPRC	0.612 (SD 0.028)
Nested mean repeat Brier score	0.206 (SD 0.008)
Nested mean calibration intercept	-0.254 (SD 0.092)
Nested mean calibration slope	0.992 (SD 0.235)
Fixed reference model	Gradient boosting with the primary analysis feature set
Fixed mean repeat AUROC	0.753 (SD 0.016)
Fixed mean repeat AUPRC	0.632 (SD 0.019)
Fixed averaged-OOF AUROC	0.762 (95%CI: 0.716-0.805)
Fixed averaged-OOF AUPRC	0.642 (95%CI: 0.566-0.720)
Fixed averaged-OOF Brier score	0.191
Fixed averaged-OOF calibration slope	0.975

Nested estimates include candidate selection and summarize five outer validation repeats. Fixed-reference OOF metrics use averaged held-out probabilities; their bootstrap intervals resample saved patient outcome-prediction pairs without refitting. AUROC: Area under the receiver operating characteristic curve; AUPRC: area under the precision-recall curve; OOF: out-of-fold; CI: confidence interval; SD: standard deviation.

RESULTS

Study population and baseline characteristics

The clinical question and cohort construction are summarized in Figure 1A and B. The initial cohort comprised 500 eligible patients. Follow-up sufficient for endpoint ascertainment was available for 469 patients, whereas 31 patients were lost to follow-up. Overall, 198 deaths occurred during the available follow-up. The 12-month modelling cohort comprised 434 patients, of whom 167 died within 12 months (38.5%) [Figure 1B]. Among the 469 patients with ascertainable outcomes, 35 patients without a recorded death had less than 365 days of follow-up and therefore could not be classified for the 12-month binary endpoint; they were excluded from binary modelling. Patients with undetermined 12-month outcomes were not classified as survivors. Baseline characteristics stratified by 12-month all-cause mortality status are shown in Table 1.

Model comparison and predictive performance

Across 30 algorithm-feature-set combinations, descriptive mean AUROC ranged from 0.535 to 0.767 [Figure 2A]. The complete-set random forest had the highest mean AUROC (0.767). Within the primary analysis set, random forest and gradient boosting had mean AUROCs of 0.758 and 0.753, respectively [Figure 2B]. Gradient boosting had a lower mean Brier score (0.195 *vs.* 0.200) and a calibration slope closer to 1 (0.885 *vs.* 1.581). This comparison informed its retention as the fixed reference for probability estimation, interpretation, and the digital display. Candidate selection was evaluated through the principal nested validation workflow. All candidate estimates and outer-fold selection frequencies are in Supplementary Table 1.

The nested candidate-selection workflow achieved mean repeat AUROC 0.729 (SD 0.018), AUPRC 0.612 (SD 0.028), and Brier score 0.206 (SD 0.008). These are the principal performance estimates [Table 2], describing the selection workflow rather than a single fixed algorithm. The fixed gradient-boosting reference had mean repeat AUROC 0.753 and AUPRC 0.632; its averaged out-of-fold predictions gave AUROC 0.762 (95%CI: 0.716-0.805) and AUPRC 0.642 (95%CI: 0.566-0.720), as shown in Figure 2C and D.

Table 3. Overlap-weighted SGLT2i-associated 12-month mortality differences after discharge

Risk stratum	*N*	Users, n	Non-users, n	Weighted user mortality, % (95%CI)	Weighted non-user mortality, % (95%CI)	RD, pp (95%CI)	ESS U/C
Overall	381	198	183	16.1 (10.4 to 22.5)	31.1 (24.0 to 38.8)	15.0 (5.9 to 23.9)	156.6/147.1
Low	127	70	57	8.0 (0.4 to 17.0)	14.5 (4.3 to 23.1)	6.5 (-6.5 to 17.4)	52.5/45.7
Intermediate	127	71	56	17.3 (6.5 to 30.2)	35.4 (20.0 to 53.5)	18.1 (-4.5 to 39.6)	46.5/37.5
High	127	57	70	32.9 (17.5 to 47.9)	52.6 (38.8 to 65.8)	19.7 (-1.7 to 40.6)	39.5/42.5

N = 381. Mortality is the overlap-weighted Kaplan-Meier risk at 365 days after discharge. RD = non-user minus user mortality in percentage points. CIs use 1,000 treatment-stratified patient bootstrap samples with propensity refitting and fixed scores/strata. Each out-of-sample-score tertile contains 127 patients. ESS, effective sample size; U/C, users/non-users. Estimates describe adjusted observational associations.

Calibration and decision-curve analysis

For the fixed gradient-boosting reference model, averaged out-of-fold probabilities yielded a Brier score of 0.191, calibration intercept 0.017, and slope 0.975. The calibration curve and risk-decile plot displayed agreement and deviations across deciles [Figure 3A and B]. Decision-curve analysis compared net benefit across hypothetical thresholds for intensified review or follow-up, rather than treatment eligibility [Figure 3C]. Mortality and non-mortality groups had overlapping predicted-risk distributions, with higher values among those who died [Figure 3D]. These fixed-reference plots are distinct from the principal nested-workflow estimates in Table 2.

Predictor importance and collinearity

Held-out permutation importance ranked NYHA functional class, BMI, NT-proBNP, age, albumin and total bilirubin highest; LVEF ranked eighth [Figure 4A]. These variables represented complementary information on age, nutritional status, haemodynamic burden, symptom severity and cardiac or systemic function. The six variables used in the simplified risk display showed limited collinearity, with a maximum VIF of 1.322 [Supplementary Table 2]. The six-variable set was distinct from both the seven-variable core comparator and the full primary model.

SGLT2i-associated differences across risk strata

The adjusted discharge-treatment cohort included 381 patients, comprising 198 SGLT2i users and 183 non-users; 84 deaths occurred within 365 days of discharge. Overlap-weighted mortality was 16.1% in users and 31.1% in non-users, with a risk difference of 15.0 percentage points (pp) (95%CI: 5.9-23.9). Overall effective sample sizes were 156.6 and 147.1, respectively. Maximum absolute standardized covariate differences were below 0.002 after weighting [Supplementary Table 5].

Each risk stratum contained 127 patients. Adjusted non-user minus user risk differences were 6.5 (-6.5 to 17.4), 18.1 (-4.5 to 39.6) and 19.7 (-1.7 to 40.6) pp in the low-, intermediate- and high-risk strata, respectively [Figure 4B-D and Table 3]. Point estimates were larger in the intermediate- and high-risk strata, but each stratum-specific interval included zero. The high-minus-low contrast was 13.2 pp (95%CI: -10.0 to 36.2).

The original full-follow-up comparison and its early-death sensitivity analysis are presented separately in Supplementary Table 4 and Supplementary Figure 1.

Survival and missing-data sensitivity analyses

Cox sensitivity analysis included all 469 patients with verified observation times, including 35 patients censored before day 365, and recorded 167 deaths through day 365. Mean repeat concordance was 0.724, and censoring-adjusted 365-day AUROC was 0.755. Kaplan-Meier 12-month mortality was 12.9%, 36.8% and

57.7% across tertiles of held-out Cox predictions (log-rank $P < 0.001$; Supplementary Figure 2A and B). No individual rank-based proportional-hazards test had $P < 0.05$ (minimum $P = 0.067$). These estimates assess sensitivity to the observed follow-up structure using the seven-variable core set [Supplementary Table 7].

For the fixed gradient-boosting reference, twenty-imputation mean repeat AUROC was 0.758, AUPRC 0.634, and Brier score 0.193, compared with 0.753, 0.632, and 0.195 using training-fold median/mode completion [Supplementary Figure 2C and D]. Ten- *vs.* twenty-imputation averaged probabilities differed by a mean of 0.37 pp. This sensitivity analysis evaluated the fixed model without repeating selection among the 30 candidates [Supplementary Table 8].

Interactive six-variable risk interface

To enhance the translational accessibility of the risk-stratification framework, we developed an interactive digital interface based on six routinely available clinical variables—age, BMI, NT-proBNP, NYHA functional class, albumin and LVEF—to display admission-origin 12-month mortality estimates [Figure 5]. The ridge surrogate approximated the fixed gradient-boosting reference model's out-of-fold predictions and preserved their overall ordering (held-out Pearson correlation, 0.821). In held-out assessment, the surrogate achieved an AUROC of 0.758, a Brier score of 0.193, and a calibration slope of 1.211; mean absolute and root mean squared differences from fixed-reference risk were 9.8 and 12.5 pp, respectively [Supplementary Table 6]. The interface reduced input burden to six variables and displayed all six standardized log-odds contributions. The treatment panel presented the adjusted cohort-level 12-month mortality differences separately from the individual mortality estimate; treatment-cohort strata were not assigned by the six-variable score.

DISCUSSION

This study evaluated routine-variable mortality risk stratification in an elderly HF-T2DM cohort through three feature sets and ten learning algorithms. Nested validation quantified the performance of the candidate-selection workflow, whereas the fixed gradient-boosting reference supported predictor interpretation and the six-variable display. Cox sensitivity analysis incorporating short-follow-up censored patients supported a graded prognostic pattern. The findings link age-related vulnerability, HF severity, cardiac load and nutritional reserve to multivariable risk assessment without identifying a single algorithm as uniformly superior.

Previous machine-learning studies support routine-variable risk assessment in older cardiometabolic populations. Liu *et al.* modelled in-hospital mortality in older patients with HF and hypertension, and Bai *et al.* examined HF classification in older adults with diabetes and hypertension[12,17]. More directly, Chen *et al.* externally validated a cardiovascular-death/HF-hospitalization model in diabetes with HFpEF; age, NT-proBNP, albumin, eGFR, and LVEF were leading predictors[18]. Their convergence with our findings extends this evidence to admission-origin 12-month all-cause mortality in patients aged ≥ 75 years with HF and T2DM.

Taken together, Figures 1-5 connect model development and validation with clinical interpretation and risk communication.

The predictor profile was clinically coherent. NT-proBNP and NYHA grade indicate haemodynamic stress and symptom burden, LVEF represents systolic reserve, and albumin and BMI reflect nutritional and catabolic status. The high ranking of BMI is particularly relevant in elderly HF, where lower BMI can accompany sarcopenia, malnutrition, chronic inflammation, or advanced disease rather than representing a simple body-size effect[19,20]. Together, these variables span cardiac stress, functional limitation, nutritional

reserve and age-related vulnerability, providing a clinically recognizable basis for their contribution to mortality prediction.

The combination of these variables is clinically important because elderly HF-T2DM risk is not purely cardiac or purely metabolic. A patient with reduced nutritional reserve, high natriuretic peptide level, worse NYHA grade, and lower LVEF may have limited resilience to acute decompensation even if no single value appears extreme. This supports the use of a multivariable model to integrate vulnerability domains that are often reviewed separately during routine care.

After adjustment for measured baseline differences, SGLT2i use remained associated with lower 12-month post-discharge mortality overall. The intermediate- and high-risk strata had larger absolute-difference point estimates than the low-risk stratum, in a direction compatible with trial and age-specific evidence supporting clinically meaningful SGLT2i effects in older HF patients[21]. Stratum-specific intervals were wide, and the high- *vs*. -low contrast did not establish a difference in association across strata. These results support further risk-stratified evaluation rather than a confirmed treatment-priority rule.

The broader evidence base reinforces the clinical relevance of risk-stratified evaluation. Comparative real-world data indicate broadly similar cardiovascular and renal effectiveness of empagliflozin and dapagliflozin[22]. In frail or older cardiometabolic populations, observations relating SGLT2 inhibition to endothelial microRNA profiles and cognitive or physical function suggest complementary vascular and functional effects[23,24], while current understanding of vascular aging provides a mechanistic framework linking cardiorenal protection with preservation of vascular health[25]. Together, these findings support continued evaluation of SGLT2i-related outcomes across baseline-risk strata in older patients with HF and T2DM.

The adjusted treatment comparison complements trial evidence by examining mortality associations in a routinely treated, elderly comorbid cohort. Baseline risk information may support medication review, closer follow-up, renal and volume-status monitoring, and structured reassessment after discharge. Baseline-only inputs, training-fold preprocessing, nested candidate selection, and supplementary survival and missing-data analyses provide complementary assessments of the prediction workflow.

The prediction and treatment analyses use explicitly defined but different clinical time origins. Admission-origin 12-month mortality describes baseline prognosis, including in-hospital deaths; discharge-origin 12-month mortality evaluates outcomes after documented discharge treatment among eligible survivors of hospitalization. The original variable-follow-up proportions are retained as a separate descriptive comparison, not as fixed-horizon adjusted estimates.

Study limitations

Several limitations should be considered. This retrospective, single-center cohort was modest relative to the number of candidate predictors and algorithms, and external validation was not performed. The fully nested estimates therefore provide the principal internal assessment, while the fixed gradient-boosting reference and its digital surrogate require evaluation in independent populations. In addition, 31 patients lacked verifiable follow-up and were excluded from endpoint analyses. The Cox sensitivity analysis retained 35 patients with shorter event-free observation as censored records, but selection related to unavailable follow-up cannot be excluded. Multiple imputation evaluated numeric missing-data uncertainty for the fixed reference model rather than the full candidate-selection workflow or treatment analysis. SGLT2i exposure was observational. Overlap weighting balanced recorded covariates, but unmeasured prescribing factors, including contraindications, clinical judgment, adherence, and treatment costs[26], may influence the associations. The

original comparison also required 30-day post-discharge survival; its separate early-death sensitivity analysis and the revised discharge-origin comparison address the importance of explicit observation windows[27]. Stratum-specific confidence intervals were wide and did not establish heterogeneous treatment associations. Bootstrap intervals for the fixed-reference predictions were conditional on saved outcome-prediction pairs and did not repeat model development; nested assessment included candidate selection. Treatment bootstrap intervals refitted propensity models but remained conditional on fixed risk strata, and overlap weighting targets patients with similar measured treatment-assignment profiles[28]. Finally, the six-variable surrogate had a mean absolute difference of 9.8 pp from the reference predictions. Independent calibration, prospective usability testing and larger multicentre cohorts are needed before routine clinical implementation.

CONCLUSION

In elderly patients with HF and T2DM, routine clinical variables supported moderate internally validated 12-month mortality risk stratification when candidate selection was included within nested validation. Cardiac load, functional severity, age-related vulnerability, and nutritional reserve contributed complementary prognostic information. Cox sensitivity analysis supported risk separation with incomplete follow-up retained as censoring, and multiple imputation assessed the fixed model's missing-data sensitivity. SGLT2i use was associated with lower adjusted 12-month post-discharge mortality overall. Larger cohorts are needed to establish whether treatment associations differ across baseline-risk strata and to validate the prediction workflow and digital risk-communication display externally and prospectively.

DECLARATIONS

Authors' contributions

Conception and design of the study: Qin Z, Jin Y, Li X, Tang J
Data curation, clinical investigation and data verification: Qin Z, Jin Y, Li X, Lan B, Wu K, Zou J, Liu T; Wang Y
Statistical and machine-learning analysis: Qin Z, Jin Y, Li X, Lan B
Interpretation of the results and drafting of the manuscript: Qin Z, Jin Y, Li X, Lan B
Clinical interpretation and critical revision for important intellectual content: Wu K, Zou J, Liu T, Wang Y, Zhang N, Zhang J, Yuan J, Tang J
Supervision and funding acquisition: Zhang J, Yuan J, Tang J
All authors reviewed and approved the final manuscript.

Availability of data and materials

Processed data supporting the main figures and tables, including model-performance summaries, aggregate risk-stratum summaries, variance inflation factor outputs, figure source data, and the six-variable risk-display payload, are provided in the Source Data and Supplementary Material accompanying this manuscript. Patient-level raw clinical data are not publicly available because they contain sensitive health information and are subject to institutional ethics and data-governance restrictions. Qualified researchers may request access to de-identified patient-level data through the corresponding author and the data-governance or ethics review process of the First Affiliated Hospital of Zhengzhou University. Access requires an approved study proposal and a data-use agreement. Individual out-of-fold predictions are not included in the public package. Custom analysis scripts for imputation, model benchmarking, figure generation, the six-variable risk display, and statistical audit are provided in the accompanying reproducibility package.

AI and AI-assisted tools statement

We used ChatGPT (OpenAI, GPT-6; released 2026-09-22) for English-language editing, code refinement and debugging assistance, and statistical-output checking during the preparation of manuscript. No synthetic patient records or AI-clinical images were created. All authors are responsible for verifying the data, methods, analyses, references, interprtation, and final submission content.

Financial support and sponsorship

This work was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (No. 2025ZD0547800), the Henan Province Key R&D Program (No. 241111313300), the Henan Provincial Clinical Medical Scientist Training Project (No. HNCMS202418), the Henan Zhongyuan Medical Science and Technology Innovation and Development Foundation Project (No. ZYYC202305ZD), and the Funding for Scientific Research and Innovation Team of the First Affiliated Hospital of Zhengzhou University (Nos. ZYCXTD2023008 and QNCXTD2023001). The funders had no role in study design, data collection, analysis, interpretation, manuscript preparation, or the decision to submit the work for publication.

Conflicts of interest

Tang J is a Guest Editor of the Special Issue "Integrated Precision Strategies for Cardiac Injury and Heart Failure: From Early Detection to Innovative Therapy" and a member of the Editorial Board of *The Journal of Cardiovascular Aging*, while the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

The study was approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University (approval No. 2025-KY-1217-001). The requirement for informed consent was waived because of the retrospective design. The study was registered with the Chinese Clinical Trial Registry (ChiCTR2500109213).

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

Supplementary Materials

Supplementary Materials

REFERENCES

1. Karwi QG, Ho KL, Pherwani S, Ketema EB, Sun Q, Lopaschuk GD. Concurrent diabetes and heart failure: interplay and novel therapeutic approaches. *Cardiovasc Res.* 2022;118:686-715. DOI PubMed
2. Truby LK, Rogers JG. Advanced heart failure. *JACC Heart Fail.* 2020;8:523-36. DOI PubMed
3. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure. *J Am Coll Cardiol.* 2022;79:e263-421. DOI
4. 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;42:3599-726. DOI
5. Bayes-genis A, Docherty KF, Petrie MC, et al. Practical algorithms for early diagnosis of heart failure and heart stress using NT-proBNP: a clinical consensus statement from the heart failure association of the ESC. *Eur J Heart Fail.* 2023;25:1891-8. DOI
6. Anker SD, Butler J, Filippatos G, et al. Empagliflozin in heart failure with a preserved ejection fraction. *N Engl J Med.* 2021;385:1451-61. DOI
7. Solomon SD, Mcmurray JJ, Claggett B, et al. Dapagliflozin in heart failure with mildly reduced or preserved ejection fraction. *N Engl J Med.* 2022;387:1089-98. DOI PubMed
8. Vaduganathan M, Docherty KF, Claggett BL, et al. SGLT2 inhibitors in patients with heart failure: a comprehensive meta-analysis of five randomised controlled trials. *Lancet.* 2022;400:757-67. DOI
9. Zannad F, Ferreira JP, Pocock SJ, et al. SGLT2 inhibitors in patients with heart failure with reduced ejection fraction: a meta-analysis of the EMPEROR-Reduced and DAPA-HF trials. *Lancet.* 2020;396:819-29. DOI
10. Nassif ME, Windsor SL, Borlaug BA, et al. The SGLT2 inhibitor dapagliflozin in heart failure with preserved ejection fraction: a multicenter randomized trial. *Nat Med.* 2021;27:1954-60. DOI PubMed PMC
11. Kent DM, Paulus JK, Van Klaveren D, et al. The predictive approaches to treatment effect heterogeneity (PATH) statement. *Ann Intern Med.* 2020;172:35-45. DOI PubMed PMC
12. Bai Q, Chen H, Gao Z, et al. Advanced prediction of heart failure risk in elderly diabetic and hypertensive patients using nine machine learning models and novel composite indices: insights from NHANES 2003-2016. *Eur J Prev Cardiol.* 2026;33:53-63. DOI

13. 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:e078378. DOI
14. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ.* 2025;388:e082505. DOI
15. Austin PC, White IR, Lee DS, Van Buuren S. Missing data in clinical research: a tutorial on multiple imputation. *Can J Cardiol.* 2021;37:1322-31. DOI PubMed PMC
16. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26:565-74. DOI PubMed PMC
17. Liu X, Xie Z, Zhang Y, et al. Machine learning for predicting in-hospital mortality in elderly patients with heart failure combined with hypertension: a multicenter retrospective study. *Cardiovasc Diabetol.* 2024;23:407. DOI PubMed PMC
18. Chen Z, Cheng J, Huang C, et al. Machine learning-based prediction of mortality and hospitalization in diabetic patients with heart failure with preserved ejection fraction: the GUARDIAN-P risk score. *Eur Heart J Digit Health.* 2026;7:ztag083. DOI PubMed PMC
19. Vest AR, Didomenico RJ, Lichtenstein L, et al. Malnutrition and Cachexia in Inpatients With Acute Cardiac Conditions: A Scientific Statement From the American Heart Association. *Circulation.* 2026;153:e1078-105. DOI PubMed PMC
20. Von Haehling S, Ebner N, Dos Santos MR, Springer J, Anker SD. Muscle wasting and cachexia in heart failure: mechanisms and therapies. *Nat Rev Cardiol.* 2017;14:323-41. DOI PubMed
21. Martinez FA, Serenelli M, Nicolau JC, et al. Efficacy and safety of dapagliflozin in heart failure with reduced ejection fraction according to age: insights from DAPA-HF. *Circulation.* 2020;141:100-11. DOI
22. Engström A, Söderling J, Hviid A, et al. Comparative cardiovascular and renal effectiveness of empagliflozin and dapagliflozin: Scandinavian cohort study. *Eur Heart J Cardiovasc Pharmacother.* 2024;10:432-43. DOI PubMed PMC
23. Mone P, Guerra G, Lombardi A, et al. Effects of SGLT2 inhibition via empagliflozin on cognitive and physical impairment in frail diabetic elders with chronic kidney disease. *Pharmacol Res.* 2024;200:107055. DOI PubMed PMC
24. Mone P, Lombardi A, Kansakar U, et al. Empagliflozin improves the MicroRNA signature of endothelial dysfunction in patients with heart failure with preserved ejection fraction and diabetes. *J Pharmacol Exp Ther.* 2023;384:116-22. DOI PubMed PMC
25. Fountoulakis N, Miyamoto Y, Pavkov ME, Karalliedde J, Maltese G. Pathophysiology of vascular ageing and the effect of novel cardio-renal protective medications in preventing progression of chronic kidney disease in people living with diabetes. *Diabet Med.* 2024;42:e15464. DOI PubMed PMC
26. Faridi KF, Dayoub EJ, Ross JS, Dhruva SS, Ahmad T, Desai NR. Medicare coverage and out-of-pocket costs of quadruple drug therapy for heart failure. *J Am Coll Cardiol.* 2022;79:2516-25. DOI PubMed PMC
27. Atzema CL, Austin PC, Yu B, et al. Effect of early physician follow-up on mortality and subsequent hospital admissions after emergency care for heart failure: a retrospective cohort study. *Can Med Assoc J.* 2018;190:E1468-77. DOI PubMed PMC
28. Li F, Morgan KL, Zaslavsky AM. Balancing Covariates via propensity score weighting. *J Am Stat Assoc.* 2017;113:390-400. DOI

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.


© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.