Journal of Biosciences and Medicines, 2026, 14(9), 209-232
https://www.scirp.org/journal/jbm
ISSN Online: 2327-509X
ISSN Print: 2327-5081

Development and Temporal Validation of Machine Learning Models for Predicting High Cardiac Biomarker Burden in Patients Undergoing Hemodialysis

Tong Wang[1,2], Jingchun Yang[1,2], Xiaoxiao Zhang[1], Jing Wang[1], Lin Li[1,3*]

[1]Blood Purification Center, The First Affiliated Hospital of Yangtze University, Jingzhou, China
[2]Department of Medicine, Yangtze University, Jingzhou, China
[3]Department of Obstetrics, The First Affiliated Hospital of Yangtze University, Jingzhou, China
Email: *15671611676@163.com

How to cite this paper: Wang, T., Yang, J.C., Zhang, X.X., Wang, J. and Li, L. (2026) Development and Temporal Validation of Machine Learning Models for Predicting High Cardiac Biomarker Burden in Patients Undergoing Hemodialysis. *Journal of Biosciences and Medicines*, **14**, 209-232.
https://doi.org/10.4236/jbm.2026.149012

Received: July 28, 2026
Accepted: September 7, 2026
Published: September 10, 2026

Abstract

Background: Cardiovascular assessment in patients undergoing maintenance hemodialysis is complicated by elevations and variability in cardiac biomarkers. We aimed to develop and temporally validate an interpretable machine learning model for classifying a study-defined composite of elevated B-type natriuretic peptide (BNP) or high-sensitivity cardiac troponin I (hs-cTnI) at the same clinical assessment episode. **Methods:** This study included 960 adults receiving maintenance hemodialysis in China. Patients treated from January 2021 to December 2023 comprised the development-period cohort, and those treated from January 2024 to December 2025 comprised the temporal validation cohort. The development-period cohort was divided into training and same-period holdout sets. Least absolute shrinkage and selection operator (LASSO) regression selected predictors, and six models were developed. Three top-performing models were combined in a stacking ensemble. Discrimination, grouped calibration, Brier score, decision-curve analysis, and Shapley additive explanations (SHAP) were evaluated. **Results:** LASSO retained 13 predictors. The stacking ensemble achieved ROC-AUC point estimates of 0.848, 0.863, and 0.835 in training out-of-fold evaluation, same-period holdout evaluation, and temporal validation, respectively. Corresponding precision-recall AUC point estimates were 0.77, 0.76, and 0.70, and Brier scores were 0.15, 0.15, and 0.16. In temporal validation, sensitivity, specificity, and accuracy point estimates were 0.75, 0.76, and 0.75. SHAP identified prognostic nutritional index (PNI), log-transformed D-dimer, age, red cell distribution width (RDW), and beta-2 microglobulin as leading predictors. Small

between-model differences should be interpreted cautiously because confidence intervals were not estimated. **Conclusions:** The stacking ensemble showed comparable point-estimate performance across same-period holdout and temporal validation data and may support further investigation of contemporaneous biomarker-burden classification. Multicenter external validation, standardized sampling, uncertainty quantification, and prospective evaluation are required before clinical implementation.

Keywords

Maintenance Hemodialysis, Cardiac Biomarkers, Machine Learning, Stacking Ensemble, Temporal Validation, Contemporaneous Classification, Explainable Artificial Intelligence

1. Introduction

Cardiovascular disease is a major contributor to morbidity and mortality among people with kidney failure treated with dialysis. Conventional in-center hemodialysis adds recurrent hemodynamic fluctuations and rapid changes in volume and solute concentrations to an already complex cardiovascular risk profile [1]. These features make cardiovascular assessment difficult, particularly when symptoms are nonspecific and conventional risk factors do not fully characterize subclinical myocardial stress or injury.

B-type natriuretic peptide (BNP) and high-sensitivity cardiac troponin I (hs-cTnI) represent different aspects of cardiovascular stress. BNP is released in response to myocardial wall stress, whereas hs-cTnI reflects myocardial injury; both have been investigated for cardiovascular risk assessment in chronic kidney disease [2]. Their interpretation in patients receiving dialysis is not straightforward. High-sensitivity troponin concentrations are frequently elevated in clinically stable dialysis patients and show substantial within-person variability, so a single measurement may not reliably distinguish chronic elevation from acute ischemic injury [3]. In a systematic review and meta-analysis of 61 studies in patients with end-stage kidney disease, higher BNP or N-terminal pro-BNP concentrations were associated with greater risks of cardiovascular and all-cause mortality, although the estimated risks varied across thresholds and some estimates were imprecise [4]. Classifying a study-defined state of biomarker elevation is therefore distinct from using either marker alone to diagnose acute myocardial infarction or heart failure.

Machine learning offers a way to model heterogeneous clinical data and potentially nonlinear relations among demographic characteristics, comorbidities, vital signs, and laboratory measurements. Previous studies have applied random forests and gradient-boosting methods to mortality prediction in hemodialysis, including an incident cohort of 1571 patients and the international Monitoring Dialysis Outcomes data set of 95,142 patients [5] [6]. These studies demonstrate fea-

sibility while also showing that model performance and selected predictors depend on the algorithm, population, and assessment target. Evaluation of a new classification target should therefore extend beyond a single discrimination statistic to include calibration, temporal validation, and transparent reporting. Explainable artificial intelligence methods, including Shapley additive explanations (SHAP), can describe the direction and magnitude of variable contributions at the population and individual levels [7]. The TRIPOD + AI statement further emphasizes complete reporting of model development and evaluation regardless of whether regression or machine-learning methods are used [8].

Accordingly, an interpretable classification framework may offer a structured way to integrate multidimensional clinical information when assessing cardiac biomarker elevation in hemodialysis. In this retrospective, single-center model-development and temporal-validation study, we aimed to classify a study-defined composite of elevated cardiac biomarkers at the index assessment episode and to evaluate model performance in a later-period cohort. The intended use is research-stage identification of patients whose routinely recorded clinical profile is consistent with contemporaneous high biomarker burden when biomarker results are unavailable or pending; the model does not predict a future event and is not intended to diagnose heart failure or acute myocardial infarction. We further sought to determine whether explainable classification could support transparent characterization of model output at both the population and individual levels.

2. Methods

2.1. Ethics and Study Design

This study was reviewed and approved by the Ethics Committee of the participating hospital (Approval No. YJ202570) and was conducted in accordance with the Declaration of Helsinki and relevant institutional requirements. Because this retrospective observational study used routinely collected and de-identified clinical data, the requirement for written informed consent was waived in accordance with the decision of the ethics committee. All data were accessed only by authorized investigators, and strict measures were implemented to protect patient privacy and confidentiality.

This was a single-center model-development and temporal-validation study conducted at a tertiary hospital in Jingzhou, Hubei Province, China. Patients treated from January 2021 through December 2023 constituted the development-period cohort, whereas patients treated from January 2024 through December 2025 constituted the temporal validation cohort. The development-period cohort was used for model training and same-period holdout evaluation, and all fitted models were subsequently applied without refitting or retuning to the later-period cohort. Because the later-period data originated from the same hospital, this cohort represents temporal validation rather than external validation in an independent setting. All observations represented actual clinical records; no synthetic data generation, bootstrap-based sample expansion, artificial perturbation of clin-

ical variables, model-guided outcome assignment, or performance-targeted cohort construction was performed.

2.2. Data Sources and Study Population

Demographic characteristics, comorbidities, vital signs, anthropometric measurements, and laboratory findings were obtained from the hospital electronic medical record system, laboratory information system, and hemodialysis management database. Consecutive adult patients receiving maintenance hemodialysis during the study periods were screened. Patients were eligible if they were aged 18 years or older, had received maintenance hemodialysis for at least 3 months, had available BNP and hs-cTnI measurements during the index clinical assessment, and had sufficient information for outcome definition, feature selection, and model development. Patients receiving temporary dialysis for acute kidney injury, those without a definable primary outcome, those lacking variables required for the final model, and those with duplicate or internally inconsistent records that could not be resolved were excluded. When more than one eligible assessment was available for the same patient during a study period, only the first qualifying assessment was retained. Patient identifiers were cross-checked between study periods to prevent the same patient from contributing observations to both cohorts. After application of the eligibility criteria, 600 patients treated from January 2021 through December 2023 were included in the development-period cohort, and 360 patients treated from January 2024 through December 2025 were included in the temporal validation cohort. Each patient contributed one patient-level analytical record.

2.3. Data Collection and Outcome Definition

Clinical information was extracted according to a prespecified data dictionary, and identical variable definitions, coding procedures, measurement units, and extraction rules were applied to the development and temporal validation cohorts. Predictor values and index cardiac-biomarker measurements were obtained from the same clinical assessment episode. When multiple predictor measurements were available, the result obtained closest to and before, or on the same day as, the index BNP and hs-cTnI assessment was retained. Blood samples were obtained as part of routine clinical care, but the retrospective source systems did not consistently encode whether each specimen was collected before dialysis, during dialysis, or after dialysis, nor the immediately preceding ultrafiltration volume. Sampling time relative to dialysis therefore could not be standardized retrospectively and may have influenced BNP and hs-cTnI concentrations. Body mass index (BMI) was calculated as weight in kilograms divided by height in meters squared. The prognostic nutritional index (PNI) was calculated as serum albumin in g/L plus five times the absolute lymphocyte count in 10^9/L, and the neutrophil-to-lymphocyte ratio (NLR) was calculated as the absolute neutrophil count divided by the absolute lymphocyte count. D-dimer concentrations were natural-log transformed because of their right-skewed distribution. The primary outcome was a

study-defined contemporaneous composite elevation of cardiac biomarkers. A patient was classified as having the outcome when BNP was at least 1045.13 pg/mL or hs-cTnI was at least 47.62 ng/L. These thresholds corresponded to the 75th percentiles of the observed BNP and hs-cTnI distributions in the complete development-period cohort and were calculated using the type 8 quantile estimator before the cohort was randomly partitioned. The outcome definition was subsequently applied unchanged to the training set, same-period holdout set, and temporal validation cohort. BNP and hs-cTnI were used exclusively to define the outcome and were excluded from feature selection and model development. Because this was a study-specific contemporaneous biomarker endpoint, these thresholds should not be interpreted independently as diagnostic cutoffs for heart failure or acute myocardial infarction.

2.4. Data Preprocessing, Partitioning, and Feature Selection

The completeness and consistency of the data were assessed before model development. Because no missing values were present in the variables included in the analytical datasets, neither single nor multiple imputation was performed. Data-quality checks were conducted to identify duplicate observations, impossible values, inconsistent measurement units, and data-entry errors. No patient was excluded solely because of an extreme but clinically plausible value, and no Winsorization or model-driven outlier removal was performed. After the outcome had been defined, the development-period cohort was divided once into a training set and a same-period holdout set at a ratio of 7:3 using outcome-stratified random sampling and random seed 123. Because the biomarker percentile thresholds had been estimated from the complete development-period cohort before this split, the holdout set was not fully independent of outcome-definition development and is described as same-period holdout evaluation rather than strict internal validation. Least absolute shrinkage and selection operator (LASSO) logistic regression was conducted exclusively in the training set to select predictors for model development. The penalty parameter lambda was determined through 10-fold cross-validation using the one-standard-error rule (lambda.1se), and predictors with nonzero coefficients at the selected lambda value were retained. The selected predictor set was subsequently fixed and used in all base models and the stacking ensemble, with no further feature selection performed in the holdout or temporal validation cohorts. Spearman correlation coefficients were calculated among the retained continuous predictors, and predictor pairs with an absolute correlation coefficient greater than 0.50 were flagged for review rather than automatically excluded.

2.5. Model Development and Stacking Ensemble

Six base models were developed using the predictors retained by LASSO: logistic regression, classification and regression tree (CART), elastic net, random forest, radial support vector machine, and gradient boosting machine. Outcome-strati-

fied 10-fold cross-validation was conducted within the training set, and identical fold assignments were used for all base models. Hyperparameters were selected by maximizing the mean cross-validated area under the receiver operating characteristic curve (ROC-AUC), and out-of-fold predicted probabilities were retained for every patient in the training set. Centering and scaling were incorporated into the modeling pipelines for logistic regression, elastic net, and the radial support vector machine, whereas tree-based models used predictors on their retained analytical scales. Logistic regression was fitted without a tuning grid. For elastic net, alpha values of 0, 0.5, and 1 and 10 logarithmically spaced lambda values ranging from 0.001 to 1 were evaluated. CART evaluated 10 complexity-parameter values ranging from 0.001 to 0.05. Random forest used 600 trees and evaluated mtry values of 3, 5, 8, and 11 and minimum node sizes of 5 and 12 under Gini splitting. For the radial support vector machine, automatic tuning was performed through the caret training interface using a tuning length of six. Gradient boosting evaluated 75, 125, and 200 trees; interaction depths of 1, 2, and 3; shrinkage values of 0.03 and 0.06; and a minimum of 10 observations per terminal node.

The three base models with the highest mean cross-validated ROC-AUC values in the training set were selected for inclusion in the stacking ensemble. Their out-of-fold predicted probabilities were used as inputs to an L1-penalized logistic meta-model. To obtain training-set stacking predictions, each outer fold was predicted using a meta-model fitted on the remaining folds, with the penalty parameter selected through an additional five-fold cross-validation procedure. A final meta-model was subsequently fitted using the complete out-of-fold prediction matrix, with its penalty parameter selected through 10-fold cross-validation. The final meta-model was used to combine the corresponding base-model predictions in the same-period holdout and temporal validation cohorts. No feature selection, preprocessing parameter, model coefficient, hyperparameter, stacking weight, or classification threshold was re-estimated using the temporal validation cohort.

2.6. Model Performance Evaluation

ROC-AUC was the primary measure of discrimination and was reported separately for the training out-of-fold, same-period holdout, and temporal validation data sets. Precision-recall area under the curve was calculated by trapezoidal integration of precision over recall. Additional classification measures included accuracy, balanced accuracy, sensitivity, specificity, positive predictive value, negative predictive value, F1 score, Matthews correlation coefficient, Jaccard index, and detection prevalence. For each model, the classification threshold was selected by maximizing the Youden index using its training-set out-of-fold predictions and was then applied unchanged to the same-period holdout and temporal validation cohorts. Calibration was described by Brier score and by grouping predicted probabilities into 10 approximately equal-sized groups and comparing mean predicted probability with observed event proportion. Calibration-in-the-large/intercepts, calibration slopes, and confidence intervals were not estimated in the original

analysis; therefore, the reported calibration results should be interpreted as descriptive rather than as a complete assessment of systematic miscalibration. Likewise, ROC-AUC, precision-recall AUC, Brier score, and threshold-based metrics are reported as point estimates without confidence intervals, so small differences between models should not be interpreted as evidence of superiority. Decision curve analysis was conducted over threshold probabilities from 0.01 to 0.80 in increments of 0.01. Net benefit was calculated as the true-positive proportion minus the false-positive proportion weighted by the odds of the threshold probability and was compared with the treat-all and treat-none strategies. Because no clinical intervention was assigned according to model predictions, decision curve analysis was interpreted as an assessment of potential decision-analytic value rather than direct evidence of improved clinical outcomes.

2.7. SHAP Analysis Methods

SHapley Additive exPlanations (SHAP) were used to interpret the stacking ensemble. SHAP values were estimated on the predicted-probability scale using a Monte Carlo approximation with 60 repetitions and adjustment enabled. The training predictor matrix was used as the background dataset, and patients in the same-period holdout set were used as the cases to be explained. Global feature importance was defined as the mean absolute SHAP value for each predictor. SHAP summary plots displayed the magnitude and direction of predictor contributions, whereas dependence plots illustrated the relationships between predictor values and their corresponding SHAP contributions.

2.8. Statistical Analysis

Continuous variables were summarized as medians and interquartile ranges (IQRs) and were compared between the development-period and temporal validation cohorts using the two-sided Wilcoxon rank-sum test. Categorical variables were summarized as frequencies and percentages and were compared using Pearson's chi-square test or Fisher's exact test, as appropriate. Baseline comparisons were descriptive and were not used for feature selection, model tuning, or modification of the fitted models. All statistical analyses were performed using R version 4.4.3. All tests were two-sided, and a P value < 0.05 was considered statistically significant.

3. Results

3.1. Participants

A total of 960 patients undergoing maintenance hemodialysis were included: 600 in the January 2021-December 2023 development-period cohort and 360 in the January 2024-December 2025 temporal validation cohort. The development-period cohort was divided into a training set (n = 420) and a same-period holdout set (n = 180). Overall, the median age was 64.0 years (IQR 54.0 - 72.0), and 590 of 960 patients (61.5%) were male. High cardiac biomarker burden was present in

340 of 960 patients (35.4%), including 213 of 600 (35.5%) in the development-period cohort and 127 of 360 (35.3%) in the temporal validation cohort (P = 0.944; Table 1). Compared with the development-period cohort, the temporal validation cohort had higher red cell distribution width (median 14.59%, IQR 13.66 - 15.52 vs 14.30%, IQR 13.50 - 15.27; P = 0.006), neutrophil count (median 5.37, IQR 3.60 - 7.83 vs 4.80, IQR 3.36 - 6.62 × 10^9/L; P = 0.008), beta-2 microglobulin concentration (median 21.25, IQR 17.21 - 26.18 vs 19.25, IQR 15.84 - 23.95 mg/L; P < 0.001), and neutrophil-to-lymphocyte ratio (median 6.46, IQR 4.27 - 11.27 vs 5.77, IQR 3.79 - 9.67; P = 0.033). BNP and high-sensitivity cardiac troponin I distributions did not differ between cohorts (P = 0.971 and P = 0.322, respectively; Table 1).

Table 1. Baseline characteristics of the development-period and temporal validation cohorts.

Characteristics	Overall (N = 960)	Development (N = 600)	Temporal validation (N = 360)	P value
Demographics and lifestyle				
Gender, n (%)				0.706
Male	590 (61.5%)	366 (61.0%)	224 (62.2%)	
Female	370 (38.5%)	234 (39.0%)	136 (37.8%)	
Age (years)	64.0 [54.0; 72.0]	64.0 [53.0; 71.0]	65.0 [54.0; 73.0]	0.188
Marital status, n (%)				0.213
Married	853 (88.9%)	539 (89.8%)	314 (87.2%)	
Unmarried/other	107 (11.1%)	61 (10.2%)	46 (12.8%)	
Education, n (%)				0.282
Illiterate	84 (8.8%)	45 (7.5%)	39 (10.8%)	
Primary school	350 (36.5%)	227 (37.8%)	123 (34.2%)	
Junior middle school	324 (33.8%)	196 (32.7%)	128 (35.6%)	
Secondary/high school	133 (13.9%)	88 (14.7%)	45 (12.5%)	
College or above	69 (7.2%)	44 (7.3%)	25 (6.9%)	
Smoking, n (%)				0.353
No	785 (81.8%)	496 (82.7%)	289 (80.3%)	
Yes	175 (18.2%)	104 (17.3%)	71 (19.7%)	
Alcohol use, n (%)				0.680
No	900 (93.8%)	561 (93.5%)	339 (94.2%)	
Yes	60 (6.3%)	39 (6.5%)	21 (5.8%)	
Vital signs and anthropometry				
Height (cm)	165.00 [158.00; 170.00]	165.00 [158.00; 169.84]	165.00 [158.00; 170.00]	0.589
Weight (kg)	61.65 [54.83; 70.00]	61.00 [54.61; 69.16]	62.65 [55.03; 70.98]	0.098
BMI (kg/m^2)	22.79 [20.68; 25.47]	22.65 [20.67; 25.34]	23.04 [20.71; 25.88]	0.236
SBP (mmHg)	148.00 [130.11; 167.00]	148.11 [131.09; 166.73]	146.50 [129.00; 167.58]	0.501
DBP (mmHg)	83.54 [73.00; 94.00]	83.67 [73.00; 94.33]	83.51 [73.27; 93.26]	0.672

Continued

Comorbidity disease, n (%)				
Hypertension				0.693
No	193 (20.1%)	123 (20.5%)	70 (19.4%)	
Yes	767 (79.9%)	477 (79.5%)	290 (80.6%)	
Diabetes mellitus				0.767
No	619 (64.5%)	389 (64.8%)	230 (63.9%)	
Yes	341 (35.5%)	211 (35.2%)	130 (36.1%)	
Stroke				0.828
No	827 (86.1%)	518 (86.3%)	309 (85.8%)	
Yes	133 (13.9%)	82 (13.7%)	51 (14.2%)	
Coronary heart disease				0.126
No	839 (87.4%)	532 (88.7%)	307 (85.3%)	
Yes	121 (12.6%)	68 (11.3%)	53 (14.7%)	
Glucose and hematological indicators				
FBG (mmol/L)	6.29 [5.30; 7.60]	6.24 [5.29; 7.57]	6.40 [5.31; 7.66]	0.239
WBC (10^9/L)	6.25 [4.96; 8.63]	6.24 [4.87; 8.50]	6.34 [5.20; 8.80]	0.171
PLT (10^9/L)	174 [131; 221]	174 [132; 218]	173 [129; 226]	0.953
HB (g/L)	77.0 [64.0; 92.2]	77.0 [64.0; 92.0]	77.1 [64.2; 92.9]	0.904
MCV (fL)	92.30 [88.57; 96.30]	92.30 [88.52; 96.26]	92.27 [88.62; 96.35]	0.996
RDW (%)	14.40 [13.56; 15.40]	14.30 [13.50; 15.27]	14.59 [13.66; 15.52]	0.006
MPV (fL)	10.53 [9.80; 11.40]	10.51 [9.80; 11.40]	10.60 [9.86; 11.50]	0.551
NEU (10^9/L)	4.99 [3.47; 7.14]	4.80 [3.36; 6.62]	5.37 [3.60; 7.83]	0.008
LYM (10^9/L)	0.77 [0.53; 1.10]	0.77 [0.54; 1.11]	0.76 [0.52; 1.09]	0.740
MONO (10^9/L)	0.40 [0.27; 0.54]	0.40 [0.27; 0.54]	0.40 [0.27; 0.59]	0.455
Nutritional and liver function indicators				
TP (g/L)	65.32 [60.02; 70.16]	65.32 [59.75; 70.23]	65.31 [60.08; 70.06]	0.830
ALB (g/L)	35.60 [32.20; 39.58]	35.90 [32.49; 39.58]	35.25 [31.66; 39.57]	0.218
GLB (g/L)	29.02 [25.51; 32.70]	29.00 [25.45; 32.52]	29.22 [25.72; 32.95]	0.434
A/G ratio	1.24 [1.03; 1.47]	1.26 [1.04; 1.48]	1.22 [1.01; 1.45]	0.137
ALT (U/L)	12.00 [8.00; 18.00]	12.00 [8.00; 18.00]	12.00 [8.00; 18.00]	0.710
AST (U/L)	13.00 [10.00; 18.00]	13.00 [10.00; 18.00]	13.00 [10.00; 18.00]	0.870
GGT (U/L)	22.00 [15.00; 36.00]	22.00 [15.00; 37.00]	21.50 [16.00; 35.00]	0.949
ALP (U/L)	82.00 [64.00; 107.00]	81.00 [62.00; 107.58]	84.00 [65.42; 106.58]	0.248
PNI	40.19 [36.44; 43.94]	40.38 [36.81; 43.81]	39.64 [35.98; 44.17]	0.211
Renal function, electrolytes, and metabolism				
Scr (μmol/L)	722.06 [527.10; 956.83]	726.53 [531.45; 963.11]	708.53 [520.20; 946.86]	0.623
UREA (mmol/L)	31.91 [25.14; 42.17]	31.78 [24.71; 42.14]	31.95 [25.83; 42.21]	0.550
UA (μmol/L)	458.50 [395.83; 533.38]	456.14 [391.95; 530.43]	467.77 [398.79; 536.90]	0.335
K (mmol/L)	4.95 [4.42; 5.62]	4.97 [4.45; 5.61]	4.94 [4.37; 5.64]	0.723
Na (mmol/L)	138.21 [135.90; 140.87]	138.20 [135.93; 140.85]	138.26 [135.61; 140.90]	0.901

Continued

Cl (mmol/L)	107.95 [104.22; 112.06]	108.12 [104.27; 112.03]	107.70 [104.14; 112.26]	0.746
Ca (mmol/L)	1.95 [1.78; 2.11]	1.95 [1.78; 2.12]	1.93 [1.75; 2.10]	0.575
Phosphate (mmol/L)	1.82 [1.50; 2.35]	1.80 [1.51; 2.35]	1.86 [1.48; 2.36]	0.958
TCO_2 (mmol/L)	19.68 [17.51; 21.69]	19.58 [17.51; 21.57]	19.80 [17.51; 21.97]	0.353
B2M (mg/L)	19.76 [16.23; 24.85]	19.25 [15.84; 23.95]	21.25 [17.21; 26.18]	<0.001
PTH (pg/mL)	352.00 [209.00; 487.00]	349.00 [203.83; 483.17]	356.50 [221.42; 488.00]	0.419
Lipid profile				
TC (mmol/L)	4.02 [3.37; 4.85]	4.02 [3.38; 4.84]	4.01 [3.37; 4.86]	0.839
TG (mmol/L)	1.39 [1.09; 1.88]	1.40 [1.08; 1.96]	1.37 [1.11; 1.85]	0.715
LDL-C (mmol/L)	2.01 [1.65; 2.62]	2.01 [1.64; 2.59]	2.01 [1.66; 2.69]	0.620
HDL-C (mmol/L)	0.96 [0.80; 1.12]	0.97 [0.80; 1.13]	0.96 [0.79; 1.09]	0.390
Cardiac biomarkers				
hs-cTnI (ng/L)	25.83 [11.07; 46.95]	26.34 [11.19; 47.60]	24.83 [10.26; 45.99]	0.322
BNP (pg/mL)	396.66 [150.72; 1054.83]	397.03 [154.09; 1046.14]	396.66 [141.22; 1058.79]	0.971
Coagulation indicators				
PT (sec)	12.50 [11.90; 13.21]	12.43 [11.90; 13.19]	12.57 [11.95; 13.30]	0.121
PT (%)	89.90 [77.20; 100.00]	90.30 [78.37; 100.00]	88.40 [75.80; 98.24]	0.152
INR	1.06 [1.02; 1.13]	1.05 [1.01; 1.13]	1.07 [1.02; 1.14]	0.195
FIB (g/L)	4.24 [3.49; 5.42]	4.19 [3.46; 5.41]	4.34 [3.68; 5.42]	0.207
APTT (sec)	30.00 [26.60; 33.20]	30.03 [26.60; 33.11]	29.76 [26.62; 33.45]	0.887
TT (sec)	16.11 [15.40; 16.80]	16.13 [15.44; 16.82]	16.10 [15.20; 16.79]	0.380
D-dimer (mg/L FEU)	1.24 [0.72; 2.48]	1.21 [0.70; 2.29]	1.34 [0.74; 2.77]	0.123
FDP (mg/L)	4.97 [1.73; 6.56]	4.94 [1.77; 6.52]	5.00 [1.66; 6.70]	0.853
Inflammatory index				
NLR	5.96 [3.95; 10.34]	5.77 [3.79; 9.67]	6.46 [4.27; 11.27]	0.033
Outcome, n (%)				
High cardiac biomarker burden				0.944
No	620 (64.6%)	387 (64.5%)	233 (64.7%)	
Yes	340 (35.4%)	213 (35.5%)	127 (35.3%)	

Values are median [25th percentile; 75th percentile] or n (%). Continuous variables were compared using the two-sided Wilcoxon rank-sum test; categorical variables were compared using Pearson's chi-square test. P values compare the January 2021-December 2023 development-period cohort with the January 2024-December 2025 same-hospital temporal validation cohort; this later-period cohort is not an external validation cohort. BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; FBG, fasting blood glucose; WBC, white blood cell; PLT, platelet; HB, hemoglobin; MCV, mean corpuscular volume; RDW, red cell distribution width; MPV, mean platelet volume; NEU, neutrophil; LYM, lymphocyte; MONO, monocyte. TP, total protein; ALB, albumin; GLB, globulin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; ALP, alkaline phosphatase; Scr, serum creatinine; UA, uric acid; PTH, parathyroid hormone; B2M, beta-2 microglobulin; PNI, prognostic nutritional index; NLR, neutrophil-to-lymphocyte ratio. TC, total cholesterol; TG, triglyceride; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; hs-cTnI, high-sensitivity cardiac troponin I; BNP, B-type natriuretic peptide; PT, prothrombin time; INR, international normalized ratio; FIB, fibrinogen; APTT, activated partial thromboplastin time; TT, thrombin time; FDP, fibrin degradation products.

3.2. Feature Selection and Correlation Assessment

Ten-fold cross-validated least absolute shrinkage and selection operator logistic regression performed in the training set retained 13 predictors with nonzero coefficients at the lambda.1se value: age, male sex, body mass index, systolic blood pressure, diabetes mellitus, coronary heart disease, hemoglobin, red cell distribution width, prognostic nutritional index, neutrophil-to-lymphocyte ratio, log-transformed D-dimer, beta-2 microglobulin, and phosphate (**Figure S1**). These 13 predictors constituted the fixed predictor set for all subsequent model development and evaluation analyses. Spearman correlation analysis was subsequently performed among the 10 continuous predictors. No predictor pair exceeded the prespecified absolute correlation threshold of 0.50. The strongest inverse correlation was observed between the prognostic nutritional index and log-transformed D-dimer (rho = −0.42), whereas the strongest positive correlation was observed between beta-2 microglobulin and phosphate (rho = 0.34). Therefore, no predictor was removed after the correlation assessment, and all 13 LASSO-selected predictors were carried forward into the six base models and the stacking analysis (**Figure S2**).

3.3. Model Development

Using the same 13 predictors retained by LASSO, 6 base models were developed: logistic regression, classification and regression tree, elastic net, random forest, radial support vector machine, and gradient boosting machine. Training-set out-of-fold evaluation identified elastic net and logistic regression as the strongest individual models for discrimination, with ROC-AUCs of 0.852 and 0.851 and precision-recall AUCs of 0.777 and 0.756, respectively. The radial support vector machine achieved a ROC-AUC of 0.835 and a precision-recall AUC of 0.744, followed by gradient boosting (0.828 and 0.749), random forest (0.821 and 0.732), and CART (0.744 and 0.646; **Figures 1(a)-(e)**). The unweighted overall performance scores ranged from 0.68 for CART to 0.77 for logistic regression (**Figure 1(c)**). Decision-curve analysis showed positive net benefit relative to the treat-none strategy across a broad range of low-to-moderate threshold probabilities, although the magnitude and persistence of benefit varied among models (**Figure 1(f)**). Grouped calibration plots provided a descriptive visual assessment of agreement between predicted and observed risks (**Figure 1(g)**). Cross-validation results were consistent with the out-of-fold estimates. Mean ROC-AUCs across the 10 folds were 0.86 (SD 0.06) for logistic regression, 0.86 (SD 0.07) for elastic net, 0.85 (SD 0.07) for the support vector machine, 0.84 (SD 0.06) for gradient boosting, 0.83 (SD 0.07) for random forest, and 0.75 (SD 0.05) for CART (**Figure 1(h)**). The corresponding precision-recall AUC results across the 10 cross-validation folds are shown in **Figure 1(i)**. Logistic regression, elastic net, and the support vector machine consequently provided the 3 out-of-fold probability inputs for the stacking ensemble.

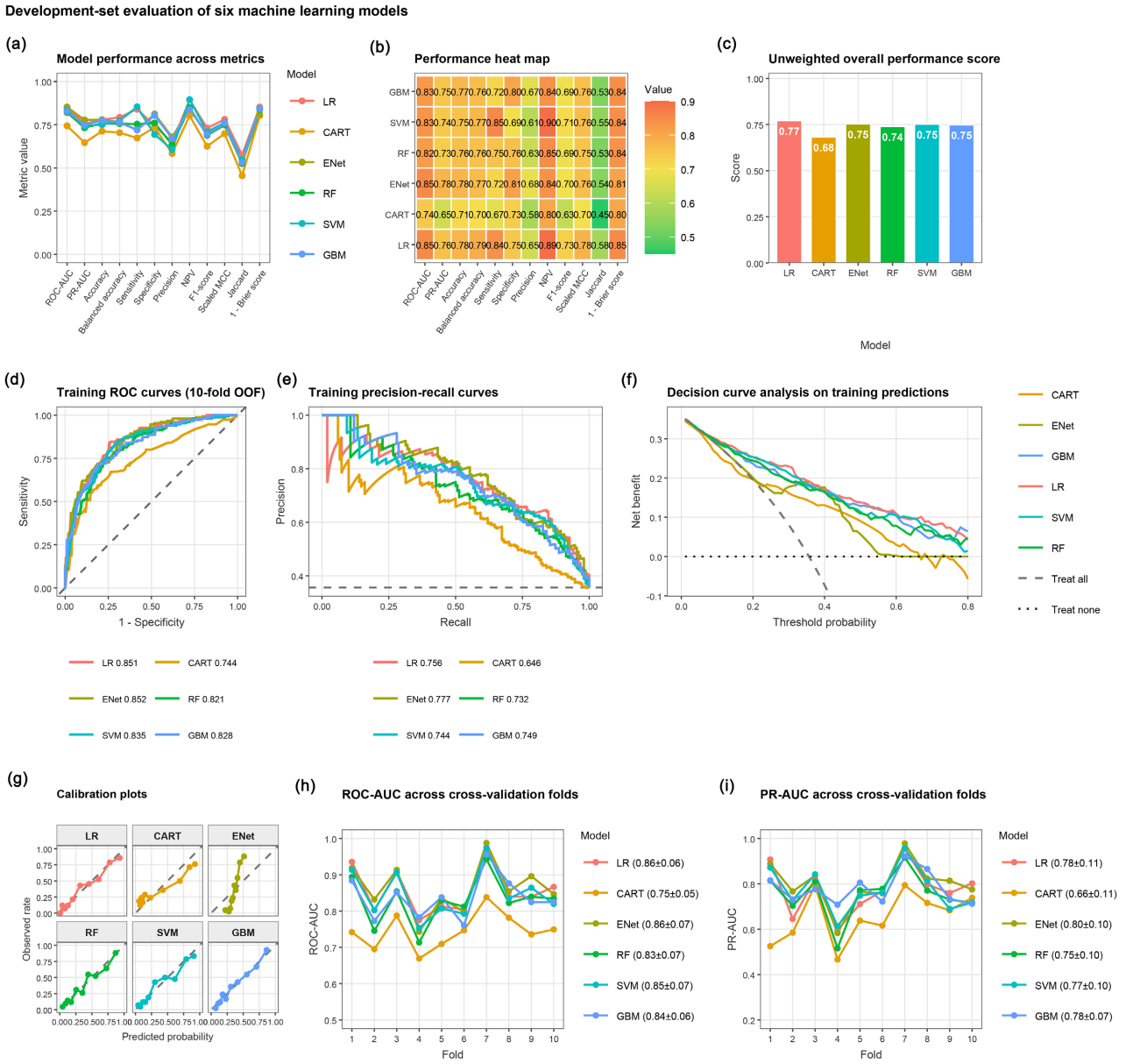

Figure 1. Development-period evaluation of six machine learning models. (a) Performance profiles across evaluation metrics; (b) Heat map of metric values; (c) Unweighted overall performance scores; (d) Training out-of-fold ROC curves; (e) Training out-of-fold precision-recall curves; (f) Decision-curve analysis; (g) Grouped calibration plots; (h) ROC-AUC across 10 cross-validation folds; (I) PR-AUC across 10 cross-validation folds.

3.4. Same-Period Holdout and Temporal Validation of Individual Machine Learning Models

In the same-period holdout set, logistic regression and elastic net had the highest discrimination point estimates among the individual models, with ROC-AUCs of 0.87 and 0.86, respectively. The support vector machine and gradient boosting each achieved a ROC-AUC of 0.84, compared with 0.80 for random forest and 0.77 for CART. In the temporal validation cohort, ROC-AUC point estimates were 0.83 for logistic regression, elastic net, and gradient boosting; 0.82 for the

support vector machine; 0.81 for random forest; and 0.74 for CART (Figure 2(a)). The corresponding precision-recall curves for the same-period holdout and temporal validation cohorts are shown in Figure 2(b). Decision curves indicated that most models provided positive net benefit over treat-none across low-to-moderate thresholds in both evaluation cohorts, with net benefit generally diminishing at higher thresholds (Figure 2(c)). Descriptive calibration performance varied by model and validation period. Logistic regression had the lowest reported Brier score in same-period holdout evaluation (0.149), whereas gradient boosting had the lowest Brier score in temporal validation (0.162). The corresponding holdout and temporal validation Brier scores were 0.185 and 0.221 for CART, 0.189 and 0.190 for elastic net, 0.177 and 0.170 for random forest, and 0.161 and 0.173 for the support vector machine (Figure 2(d)). Because uncertainty intervals were not estimated, the small between-model differences in these point estimates should not be interpreted as evidence of superiority.

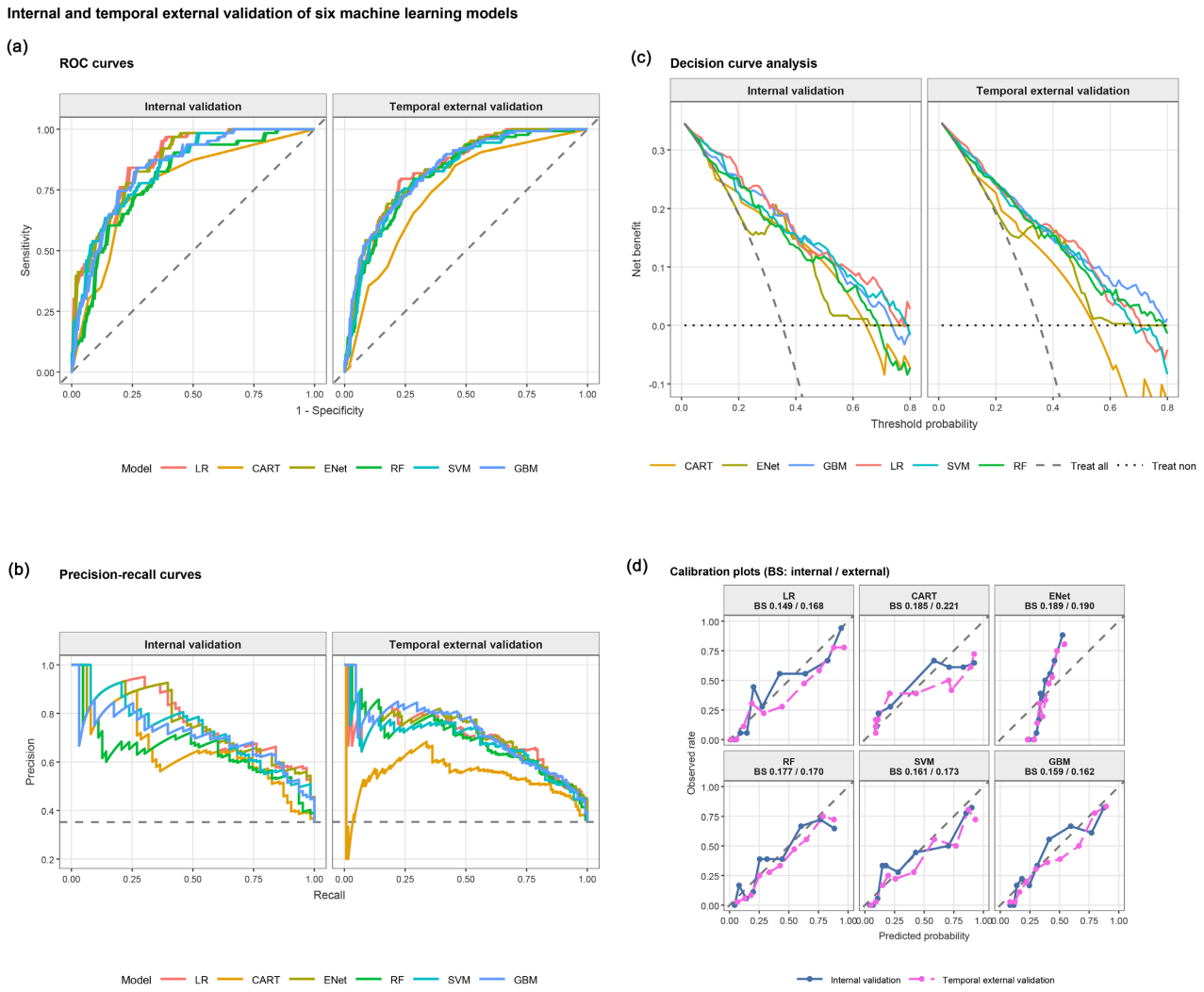

Figure 2. Performance of six machine learning models in the same-period holdout and temporal validation cohorts. (a) ROC curves; (b) Precision-recall curves; (c) Decision-curve analysis; (d) Grouped calibration plots with Brier scores.

3.5. Stacking Ensemble Development and Performance

The three base models with the highest mean cross-validated ROC-AUCs were logistic regression, elastic net, and the radial support vector machine; these models were selected as components of the stacking ensemble. Each component model was developed using the same 13 LASSO-selected predictors. Their out-of-fold predicted probabilities, rather than the original predictor values, were used as inputs to an L1-penalized logistic meta-model. The ensemble achieved ROC-AUC point estimates of 0.848 in training out-of-fold evaluation, 0.863 in same-period holdout evaluation, and 0.835 in temporal validation. The corresponding precision-recall AUC point estimates were 0.77, 0.76, and 0.70, and the Brier scores were 0.15, 0.15, and 0.16, respectively (Figure 3(a)). Thus, the ensemble remained comparable by point estimate with the strongest individual models; the small observed differences cannot establish algorithmic superiority without uncertainty intervals.

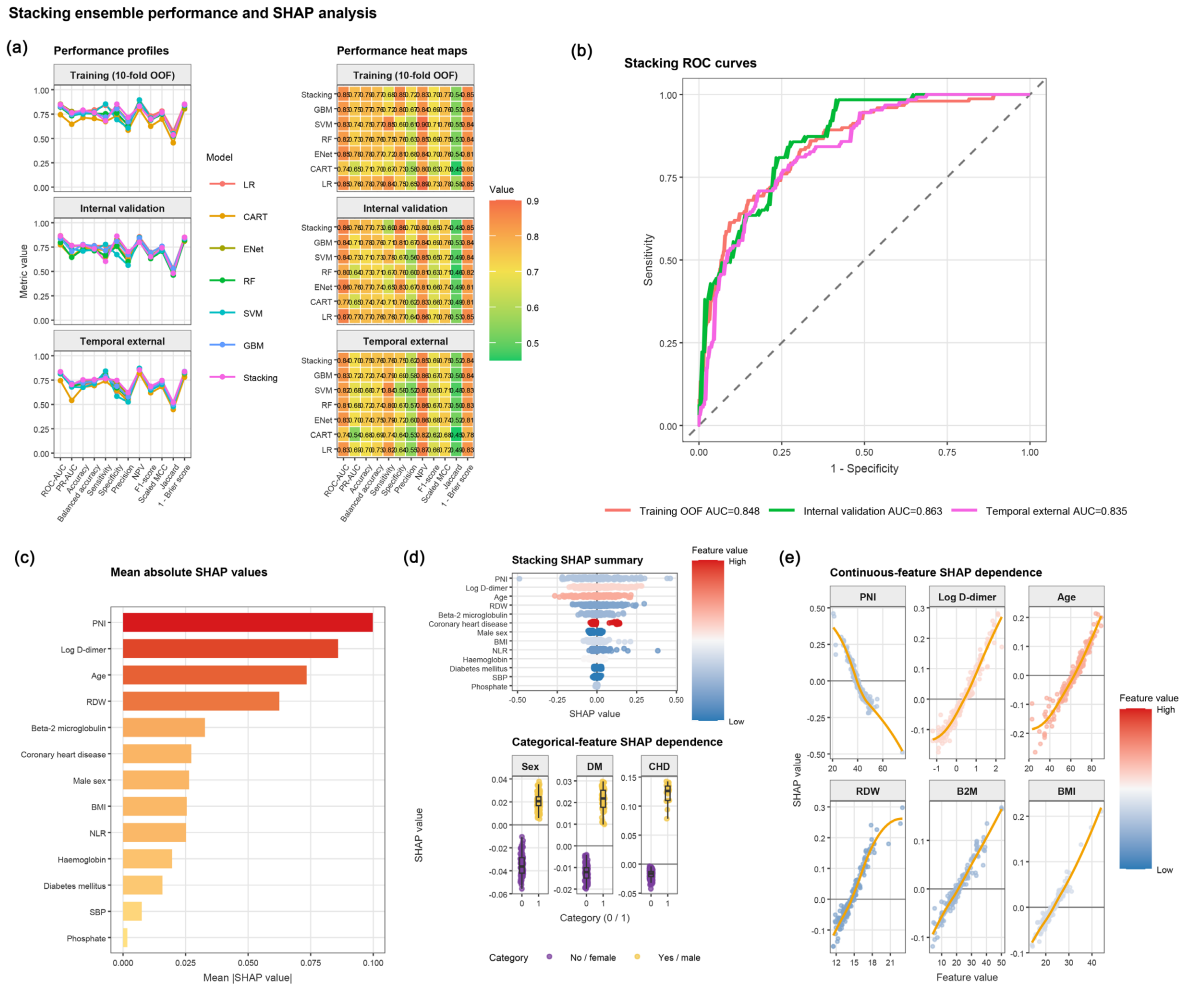

Figure 3. Performance and interpretation of the stacking ensemble. (a) Performance profiles and heat maps for the base models and stacking ensemble across training out-of-fold evaluation, same-period holdout evaluation, and temporal validation; (b) ROC curves for the stacking ensemble across the three data sets; (c) Mean absolute SHAP values; (d) SHAP summary and categorical-feature dependence plots; (e) Continuous-feature SHAP dependence plots.

At the prespecified classification threshold, the stacking model had sensitivity and specificity point estimates of 0.68 and 0.85 in training out-of-fold evaluation, 0.60 and 0.86 in same-period holdout evaluation, and 0.75 and 0.76 in temporal validation. Accuracy point estimates were 0.79, 0.77, and 0.75 across these data sets, respectively (**Figure 3(a)**). The ROC curves for the stacking ensemble across the three data sets are shown in **Figure 3(b)**.

3.6. Global and Individual SHAP Explanations

In the same-period holdout set, global SHAP analysis quantified the contributions of the 13 LASSO-selected predictors to the stacking ensemble. In descending order of mean absolute SHAP value, the predictors were prognostic nutritional index, log-transformed D-dimer, age, red cell distribution width, beta-2 microglobulin, coronary heart disease, male sex, body mass index, neutrophil-to-lymphocyte ratio, hemoglobin, diabetes mellitus, systolic blood pressure, and phosphate (**Figure 3(c)** and **Figure 3(d)**). Lower prognostic nutritional index values and higher values of log-transformed D-dimer, age, red cell distribution width, beta-2 microglobulin, and body mass index were associated with larger positive contributions to predicted risk across their observed ranges (**Figure 3(e)**). Coronary heart disease, male sex, and diabetes mellitus also contributed positively to model predictions, whereas systolic blood pressure and phosphate had comparatively small mean absolute SHAP values.

The single-patient explanation illustrated how these population-level patterns combined within an individual prediction. For the example patient, the predicted probability was 0.450 compared with an adjusted background expectation of 0.342. Red cell distribution width of 17.90%, log-transformed D-dimer of 1.92, male sex, body mass index of 24.01 kg/m^2, and a neutrophil-to-lymphocyte ratio of 8.19 increased the prediction, whereas a prognostic nutritional index of 44.65, age of 43 years, absence of coronary heart disease, beta-2 microglobulin of 16.86 mg/L, and hemoglobin of 83.00 g/L decreased it (**Figure S3**).

4. Discussion

In this single-center retrospective study of 960 adults receiving maintenance hemodialysis, we developed and temporally evaluated an interpretable model for a study-defined contemporaneous composite of high cardiac biomarker burden. Training-set LASSO retained 13 routinely available demographic, clinical, and laboratory predictors spanning nutritional, inflammatory, hematologic, coagulation-related, and uremic domains. The stacking ensemble had ROC-AUC point estimates of 0.848 in training out-of-fold evaluation and 0.835 in the later-period cohort; its temporal-validation precision-recall AUC was 0.70 and its Brier score was 0.16. SHAP analysis identified the prognostic nutritional index (PNI), log-transformed D-dimer, age, red cell distribution width (RDW), and beta-2 microglobulin as the leading contributors to model predictions. These findings support the feasibility of integrating routine data to characterize contemporaneous bi-

omarker burden, but they do not establish future-risk prediction, a diagnostic test for heart failure or acute myocardial infarction, a causal model of cardiovascular injury, or an intervention that improves outcomes.

The ensemble should be interpreted as a relatively stable aggregation of complementary models rather than as evidence that stacking was uniformly superior. Logistic regression achieved a slightly higher ROC-AUC point estimate than stacking in same-period holdout evaluation (0.87 vs 0.863), and several individual models had temporal-validation ROC-AUCs between 0.81 and 0.83. The observed differences in point estimates are too small to support a claim of decisive algorithmic superiority, particularly because confidence intervals were not estimated. The strong performance of logistic regression and elastic net suggests that much of the available signal was captured by additive or regularized relations, whereas stacking may have helped preserve performance by combining their probability estimates with those from a radial support vector machine. Recent hemodialysis studies have likewise reported useful mortality prediction with random forests, deep learning, and other machine-learning approaches, but their outcomes, prediction horizons, feature sets, and validation designs differ substantially from those used here; their reported ROC-AUCs therefore should not be compared directly with ours [9]-[11].

Temporal validation was informative because the later-period cohort differed from the development-period cohort in RDW, neutrophil count, beta-2 microglobulin, and neutrophil-to-lymphocyte ratio (NLR), yet the outcome prevalence remained similar and discrimination declined only modestly by point estimate. This provides a limited test of robustness to calendar-time change within the same hospital, not external validation across independent settings. Methodological guidance regards temporal validation as weaker evidence of transportability than evaluation in an independent setting and recommends examining calibration and clinical utility in addition to discrimination [12]. The relevance of this distinction is illustrated by a recent eight-center external validation of a Bayesian-network mortality tool in hemodialysis, in which performance and sensitivity differed between incident and prevalent patients [13].

The outcome in this study was deliberately defined as biomarker burden rather than a clinical diagnosis or future event. This distinction is particularly important in dialysis. In 200 asymptomatic patients receiving hemodialysis, elevated hs-cTnI and hs-cTnT were common and the two troponins showed different changes after dialysis [14]. In a prospective cohort of 178 patients, ultrafiltration volume and dialysis treatment time influenced intradialytic hs-cTnI concentrations [15], whereas monthly measurements in another cohort demonstrated marked between-patient variation in chronic troponin baselines [16]. A randomized crossover study further showed that the trajectories of hs-cTnI, hs-cTnT, and N-terminal pro-B-type natriuretic peptide differed by dialysis modality and sampling time [17], and a smaller crossover trial found different pre-post patterns of hs-cTnI and natriuretic peptide concentrations with routine versus cooled dialysate [18]. These

studies support cautious interpretation of a single absolute biomarker value. Because dialysis-relative sampling time and immediately preceding ultrafiltration volume were not consistently encoded in our retrospective source systems, residual measurement heterogeneity may have affected outcome classification and model performance. The present endpoint should not be used to rule in or rule out an acute cardiovascular condition.

BNP and hs-cTnI nevertheless provide complementary information about myocardial wall stress and injury. A review focused on dialysis emphasized that natriuretic peptide concentrations reflect a combination of cardiovascular disease, volume status, reduced renal clearance, and dialysis-related factors, while suggesting that serial measurements may be more informative than isolated values [19]. In 3182 participants with nondialysis chronic kidney disease, BNP and hs-cTnT contributed to a multibiomarker score associated with incident heart failure, although adding the biomarker score to clinical factors produced only modest gains in discrimination [20]. That study used a different population and a longitudinal clinical outcome, so it does not validate our percentile-based thresholds. It does, however, support the biological rationale for considering markers of stretch and injury together while maintaining clinical context.

PNI had the largest mean absolute SHAP value, and lower PNI contributed to higher predicted probability. Because PNI combines serum albumin and lymphocyte count, it summarizes nutritional and inflammatory information rather than a single pathway. In a cohort of 101,616 patients initiating hemodialysis, higher PNI was associated with progressively lower mortality and provided better one-year mortality prediction than albumin or lymphocyte count alone [21]. A 2024 meta-analysis of 29 dialysis cohorts also found that malnutrition was associated with higher mortality, with a similar direction in the hemodialysis subgroup [22]. These mortality data do not establish that low PNI causes cardiac biomarker elevation, but they make the prominence of PNI in our model clinically coherent and identify nutritional assessment as an appropriate component of subsequent clinical review. The contributions of D-dimer, NLR, and RDW should also be interpreted as associations with a broader systemic state rather than as disease-specific signals. In a direct study of 167 maintenance hemodialysis patients, D-dimer exceeded 500 micrograms/L in 75% overall and remained positive in 52% of patients without an additional acute illness or predisposing chronic disease, limiting its specificity for thromboembolic exclusion in this population [23]. A recent meta-analysis of 19 chronic hemodialysis studies associated elevated NLR with both all-cause and cardiovascular mortality [24], and a cohort study of 181 maintenance hemodialysis patients associated higher RDW with all-cause mortality while finding no significant relation with cardiovascular mortality [25]. These prior outcomes are different from the present composite, and the SHAP findings neither diagnose thrombosis nor justify anticoagulation. They instead suggest that coagulation-related, inflammatory, and erythropoietic abnormalities jointly inform the model's estimate of biomarker burden. Beta-2 microglobulin ranked fifth in

the global SHAP analysis. In 5332 participants from the Dialysis Outcomes and Practice Patterns Study, beta-2 microglobulin remained positively associated with mortality in the high-flux hemodialysis era and was also related to dialysis vintage, residual urine volume, and inflammation [26]. Its contribution here may therefore represent accumulated uremic and dialysis-related burden rather than a specific cardiac mechanism. Phosphate and systolic blood pressure had relatively small mean absolute SHAP values. Low importance in this model does not imply low clinical importance because feature importance is conditional on the selected predictors, their distributions, the fitted model, and the contemporaneous endpoint. Longer-term phosphorus exposure above 4.5 mg/dL was associated with cardiovascular mortality in 17,414 hemodialysis patients [27]; phosphate variability was associated with mortality in a cohort of 302,613 incident hemodialysis patients [28]; and a contemporary Australian and New Zealand registry analysis found a U-shaped relation between serum phosphate and mortality [29]. A 2026 systematic review and meta-analysis of dialysis cohorts similarly found that high phosphate was associated with higher cardiovascular mortality [30]. These longitudinal findings also show why a single phosphate measurement may have limited predictive prominence for the cross-sectional endpoint used here.

SHAP adds transparency by showing how each recorded feature shifted a prediction relative to a background expectation. This can help clinicians audit whether a prediction is driven by plausible recorded information and can expose unexpected dependence on a feature. However, explainable artificial intelligence frameworks distinguish population-associative explanations from mechanistic explanations. The observed SHAP direction and magnitude therefore should not be interpreted as an intervention target, a causal effect, or proof that changing a predictor will change BNP, hs-cTnI, or clinical outcomes.

The next step is prospective, multicenter evaluation with standardized biomarker sampling, an explicitly defined clinical action, and assessment of workflow and patient consequences. The size of an external validation cohort should be chosen to estimate calibration, discrimination, and net benefit with adequate precision rather than by an arbitrary event count [31]. If the model is embedded in clinical decision support, early-stage evaluation should also assess safety, human factors, and actual clinical performance before any large-scale implementation or impact trial [32].

This study has several strengths but also important limitations. First, feature selection and model fitting were confined to the training data, the 13-predictor set was fixed before evaluation, and the stacking model used out-of-fold predictions. Performance was evaluated in a later-period cohort without refitting using multiple point-estimate measures, together with global and individual SHAP analyses. Second, the retrospective single-center design and same-hospital temporal cohort limit generalizability and do not constitute independent external validation. Third, the percentile-based composite was study-specific and contemporaneous rather than diagnostic or prognostic. Its thresholds were estimated from the com-

plete development-period cohort before the training/holdout split; this allowed holdout information to influence outcome-definition development, so the same-period holdout evaluation was not fully independent. Re-estimating thresholds from the training set would change some outcome labels and would require rebuilding and reevaluating the complete modeling pipeline, which was not performed in the present revision. The later-period temporal cohort remained uninvolved in threshold estimation. Fourth, dialysis-relative sampling time and ultrafiltration immediately before sampling were not consistently available, and longitudinal dialysis-related variables were unavailable; these factors may influence biomarker concentrations. Fifth, calibration intercepts and slopes and confidence intervals for performance measures were not estimated, so grouped calibration plots, Brier scores, and small between-model differences should be interpreted descriptively. The complete-case dataset also did not permit evaluation of model performance under missingness. Finally, retrospective decision-curve analysis and the absence of subgroup fairness or prospective impact evaluation preclude claims of clinical superiority or implementation readiness.

5. Conclusion

A stacking ensemble based on 13 LASSO-selected routine clinical variables showed comparable point-estimate discrimination in same-period holdout and temporal validation for a study-defined contemporaneous composite of high BNP or hs-cTnI burden in patients receiving maintenance hemodialysis. PNI, D-dimer, age, RDW, and beta-2 microglobulin contributed most strongly to predictions, linking the model to nutritional, inflammatory, coagulation-related, and uremic domains. The findings support further investigation of an interpretable classification approach, but not future-risk prediction or immediate diagnostic or therapeutic use. Multicenter external validation, standardized longitudinal sampling, clinically anchored thresholds, complete calibration assessment, uncertainty quantification, and prospective impact evaluation are required before the model can inform care.

Ethics Approval and Consent to Participate

This retrospective observational study was conducted in accordance with the Declaration of Helsinki and relevant institutional requirements. The study protocol was reviewed and approved by the Ethics Committee of The First Affiliated Hospital of Yangtze University (Approval No. YJ202570). Because the study used routinely collected, de-identified clinical data and involved no prospective intervention or direct participant contact, the requirement for written informed consent was waived by the ethics committee. All data were accessible only to authorized investigators and were handled in accordance with institutional requirements for patient privacy and confidentiality.

Availability of Data and Materials

The de-identified patient-level data supporting the findings of this study are not

publicly available because they contain sensitive clinical information and are subject to ethical and institutional data-protection restrictions. Reasonable requests for access may be directed to the corresponding author and will be considered subject to approval by The First Affiliated Hospital of Yangtze University and its ethics committee, and completion of an appropriate data-use agreement.

Code Availability

The analysis code used to develop and evaluate the models is available from the corresponding author upon reasonable request.

Author Contributions

All authors contributed to the study conception and design. Writing-original draft preparation: Tong Wang; Writing-review and editing: Jingchun Yang; Conceptualization: Xiaoxiao Zhang; Methodology: Tong Wang; Formal analysis and investigation: Tong Wang; Resources: Jing Wang; Supervision: Lin Li. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Reaves, A.C., Weiner, D.E. and Sarnak, M.J. (2024) Home Dialysis in Patients with Cardiovascular Diseases. *Clinical Journal of the American Society of Nephrology*, **19**, 1646-1655. https://doi.org/10.2215/cjn.0000000000000410

[2] Kula, A. and Bansal, N. (2022) Applications of Cardiac Biomarkers in Chronic Kidney Disease. *Current Opinion in Nephrology & Hypertension*, **31**, 534-540. https://doi.org/10.1097/mnh.0000000000000829

[3] Snaedal, S., Bárány, P., Lund, S.H., Qureshi, A.R., Heimbürger, O., Stenvinkel, P., *et al.* (2021) High-Sensitivity Troponins in Dialysis Patients: Variation and Prognostic Value. *Clinical Kidney Journal*, **14**, 1789-1797. https://doi.org/10.1093/ckj/sfaa215

[4] Harrison, T.G., Shukalek, C.B., Hemmelgarn, B.R., Zarnke, K.B., Ronksley, P.E., Iragorri, N., *et al.* (2020) Association of NT-proBNP and BNP with Future Clinical Outcomes in Patients with ESKD: A Systematic Review and Meta-Analysis. *American Journal of Kidney Diseases*, **76**, 233-247. https://doi.org/10.1053/j.ajkd.2019.12.017

[5] Garcia-Montemayor, V., Martin-Malo, A., Barbieri, C., Bellocchio, F., Soriano, S., Pendon-Ruiz de Mier, V., *et al.* (2021) Predicting Mortality in Hemodialysis Patients Using Machine Learning Analysis. *Clinical Kidney Journal*, **14**, 1388-1395. https://doi.org/10.1093/ckj/sfaa126

[6] Chaudhuri, S., Larkin, J., Guedes, M., Jiao, Y., Kotanko, P., Wang, Y., *et al.* (2023) Predicting Mortality Risk in Dialysis: Assessment of Risk Factors Using Traditional and Advanced Modeling Techniques within the Monitoring Dialysis Outcomes Initiative. *Hemodialysis International*, **27**, 62-73. https://doi.org/10.1111/hdi.13053

[7] Loh, H.W., Ooi, C.P., Seoni, S., Barua, P.D., Molinari, F. and Acharya, U.R. (2022) Application of Explainable Artificial Intelligence for Healthcare: A Systematic Review of the Last Decade (2011–2022). *Computer Methods and Programs in Biomedicine*,

226, Article 107161. https://doi.org/10.1016/j.cmpb.2022.107161

[8] Collins, G.S., Moons, K.G.M., Dhiman, P., *et al.* (2024) TRIPOD + AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods. *British Medical Journal*, **385**, e078378.

[9] Noh, J., Park, S.Y., Bae, W., Kim, K., Cho, J., Lee, J.S., *et al.* (2024) Predicting Early Mortality in Hemodialysis Patients: A Deep Learning Approach Using a Nationwide Prospective Cohort in South Korea. *Scientific Reports*, **14**, Article No. 29658. https://doi.org/10.1038/s41598-024-80900-6

[10] Chen, M., Zeng, Y., Liu, M., Li, Z., Wu, J., Tian, X., *et al.* (2024) Interpretable Machine Learning Models for the Prediction of All-Cause Mortality and Time to Death in Hemodialysis Patients. *Therapeutic Apheresis and Dialysis*, **29**, 220-232. https://doi.org/10.1111/1744-9987.14212

[11] Huang, J., Chen, L., Luo, H., Song, J., Bi, Z., Chen, K., *et al.* (2025) Use of Machine Learning Models to Predict Mortality in Dialysis Patients. *Frontiers in Public Health*, **13**, Article 1683285. https://doi.org/10.3389/fpubh.2025.1683285

[12] Binuya, M.A.E., Engelhardt, E.G., Schats, W., Schmidt, M.K. and Steyerberg, E.W. (2022) Methodological Guidance for the Evaluation and Updating of Clinical Prediction Models: A Systematic Review. *BMC Medical Research Methodology*, **22**, Article No. 316. https://doi.org/10.1186/s12874-022-01801-8

[13] Granal, M., Brokhes-Le Calvez, S., Dimitrov, Y., Chantrel, F., Borni-Duval, C., Muller, C., *et al.* (2024) External Validation of the 2-Year Mortality Prediction Tool in Hemodialysis Patients Developed Using a Bayesian Network. *Clinical Kidney Journal*, **17**, sfae095. https://doi.org/10.1093/ckj/sfae095

[14] Wongcharoen, W., Chombandit, T., Phrommintikul, A. and Noppakun, K. (2021) Variability of High-Sensitivity Cardiac Troponin T and I in Asymptomatic Patients Receiving Hemodialysis. *Scientific Reports*, **11**, Article No. 17334. https://doi.org/10.1038/s41598-021-96658-0

[15] Collister, D., Mazzetti, A., Bhalerao, A., Tyrwhitt, J., Kavsak, P., Brimble, K.S., *et al.* (2021) Variability in Cardiac Biomarkers during Hemodialysis: A Prospective Cohort Study. *Clinical Chemistry*, **67**, 308-316. https://doi.org/10.1093/clinchem/hvaa299

[16] Gremaud, S., Fellay, B., Hemett, O.M., Magnin, J. and Descombes, E. (2022) Monthly Measurement of High-Sensitivity Cardiac Troponins T and Creatine Kinase in Asymptomatic Chronic Hemodialysis Patients: A One-Year Prospective Study. *Hemodialysis International*, **26**, 166-175. https://doi.org/10.1111/hdi.12985

[17] Drivsholm, C.L., Nygaard, L., Dam-Dalgeir, G., Glerup, R.I., Birnum, A., Jensen, J.D., *et al.* (2026) Intradialytic Kinetics of Cardiac Biomarkers during High-Flux Hemodialysis and Post-Dilution Hemodiafiltration: A Randomized Crossover Trial. *Kidney* 360. https://doi.org/10.34067/kid.0000001284

[18] Bathish, Y., Beiruti, K., Safadi, H., Nov, A.S., Bukovetzky, E., Edelstein, M., *et al.* (2022) Effects of Hemodialysis with Cooled Dialysate on High-Sensitivity Cardiac Troponin I and Brain Natriuretic Peptide. *Hemodialysis International*, **26**, 533-539. https://doi.org/10.1111/hdi.13039

[19] Chaikijurajai, T., Choles, H.R. and Tang, W.H.W. (2020) Do Natriuretic Peptide Measurements Provide Insights into Management of End-Stage Renal Disease Patients Undergoing Dialysis? *Current Heart Failure Reports*, **17**, 449-456. https://doi.org/10.1007/s11897-020-00488-6

[20] Janus, S.E., Hajjari, J., Chami, T., Mously, H., Badhwar, A.K., Karnib, M., *et al.* (2022) Multi-Variable Biomarker Approach in Identifying Incident Heart Failure in Chronic Kidney Disease: Results from the Chronic Renal Insufficiency Cohort Study. *European Journal of Heart Failure*, **24**, 988-995. https://doi.org/10.1002/ejhf.2543

[21] Miyasato, Y., Hanna, R.M., Morinaga, J., Mukoyama, M. and Kalantar-Zadeh, K. (2023) Prognostic Nutritional Index as a Predictor of Mortality in 101,616 Patients Undergoing Hemodialysis. *Nutrients*, **15**, Article 311. https://doi.org/10.3390/nu15020311

[22] Rashid, I., Sahu, G., Tiwari, P., Willis, C., Asche, C.V., Bagga, T.K., *et al.* (2024) Malnutrition as a Potential Predictor of Mortality in Chronic Kidney Disease Patients on Dialysis: A Systematic Review and Meta-Analysis. *Clinical Nutrition*, **43**, 1760-1769. https://doi.org/10.1016/j.clnu.2024.05.037

[23] Gubensek, J., Lolic, M., Ponikvar, R. and Buturovic-Ponikvar, J. (2016) D-Dimer Levels in Maintenance Hemodialysis Patients: High Prevalence of Positive Values Also in the Group without Predisposing Diseases. *Hemodialysis International*, **20**, 198-203. https://doi.org/10.1111/hdi.12371

[24] Vempati, R., Damarlapally, N., Vasudevan, S.S., Patel, V., Banda, P., Mourad, D., *et al.* (2025) Association of Neutrophil-Lymphocyte Ratio with Cardiovascular and All-Cause Mortality in Patients Receiving Chronic Hemodialysis: Systematic Review and Meta-Analysis. *World Journal of Methodology*, **15**, Article 107468. https://doi.org/10.5662/wjm.v15.i4.107468

[25] Zhang, J., Lu, X., Feng, J., Wang, S. and Li, H. (2023) Prognostic Value of Red Cell Distribution Width and Mean Corpuscular Volume on Mortality in Hemodialysis Patients. *Seminars in Dialysis*, **36**, 18-23. https://doi.org/10.1111/sdi.13109

[26] Kanda, E., Muenz, D., Bieber, B., Cases, A., Locatelli, F., Port, F.K., *et al.* (2021) Beta-2 Microglobulin and All-Cause Mortality in the Era of High-Flux Hemodialysis: Results from the Dialysis Outcomes and Practice Patterns Study. *Clinical Kidney Journal*, **14**, 1436-1442. https://doi.org/10.1093/ckj/sfaa155

[27] Lopes, M.B., Karaboyas, A., Bieber, B., Pisoni, R.L., Walpen, S., Fukagawa, M., *et al.* (2020) Impact of Longer Term Phosphorus Control on Cardiovascular Mortality in Hemodialysis Patients Using an Area under the Curve Approach: Results from the Dopps. *Nephrology Dialysis Transplantation*, **35**, 1794-1801. https://doi.org/10.1093/ndt/gfaa054

[28] ter Meulen, K.J., Ye, X., Wang, Y., Usvyat, L.A., van der Sande, F.M., Konings, C.J., *et al.* (2023) Variability of Serum Phosphate in Incident Hemodialysis Patients: Association with All-Cause Mortality. *Kidney* 360, **4**, 374-380. https://doi.org/10.34067/kid.0000000000000062

[29] Tiong, M.K., Ullah, S., McDonald, S.P., Tan, S., Lioufas, N.M., Roberts, M.A., *et al.* (2021) Serum Phosphate and Mortality in Incident Dialysis Patients in Australia and New Zealand. *Nephrology*, **26**, 814-823. https://doi.org/10.1111/nep.13904

[30] Franczyk, B., Rysz, J. and Gluba-Sagr, A. (2026) Impact of Serum Phosphate, Potassium and Other Electrolyte Levels on Sudden Cardiac Death and Cardiovascular Mortality in Haemodialysis and Peritoneal Dialysis: A Systematic Review and Meta-Analysis. *Biomedicines*, **14**, Article 605. https://doi.org/10.3390/biomedicines14030605

[31] Riley, R.D., Debray, T.P.A., Collins, G.S., Archer, L., Ensor, J., van Smeden, M., *et al.* (2021) Minimum Sample Size for External Validation of a Clinical Prediction Model with a Binary Outcome. *Statistics in Medicine*, **40**, 4230-4251. https://doi.org/10.1002/sim.9025

[32] Vasey, B., Nagendran, M., Campbell, B., Clifton, D.A., Collins, G.S., Denaxas, S., *et al.* (2022) Reporting Guideline for the Early-Stage Clinical Evaluation of Decision Support Systems Driven by Artificial Intelligence: DECIDE-AI. *Nature Medicine*, **28**, 924-933. https://doi.org/10.1038/s41591-022-01772-9

Supplement

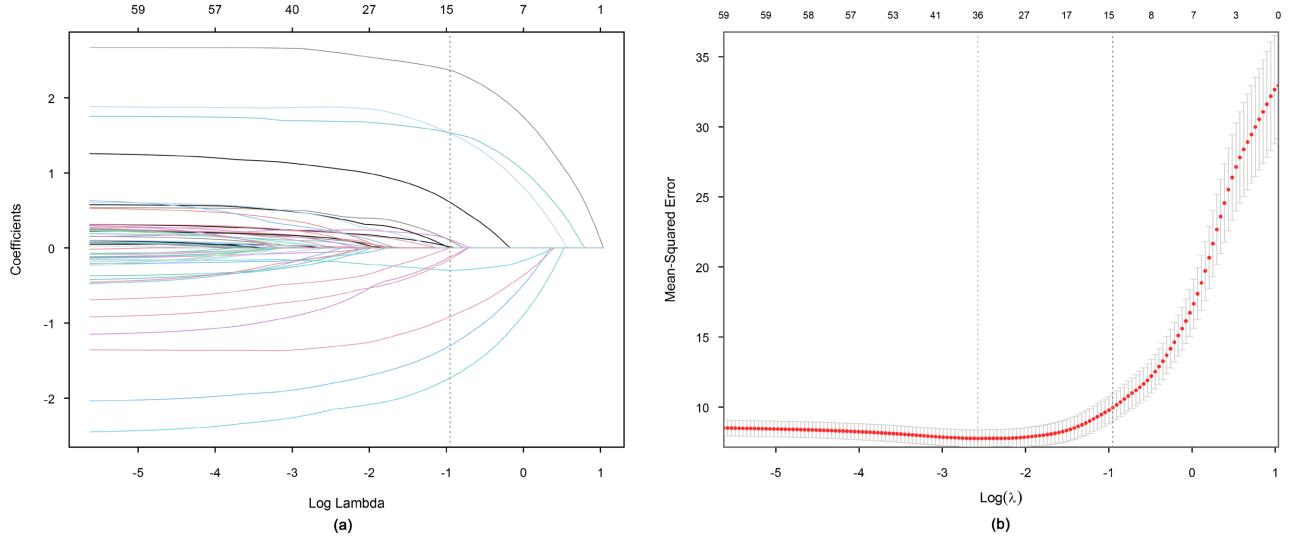

Figure S1. LASSO selection of model predictors. (a) Predictor coefficient trajectories across log-transformed λ values. (b) Ten-fold cross-validated mean squared error used to select the penalty parameter.

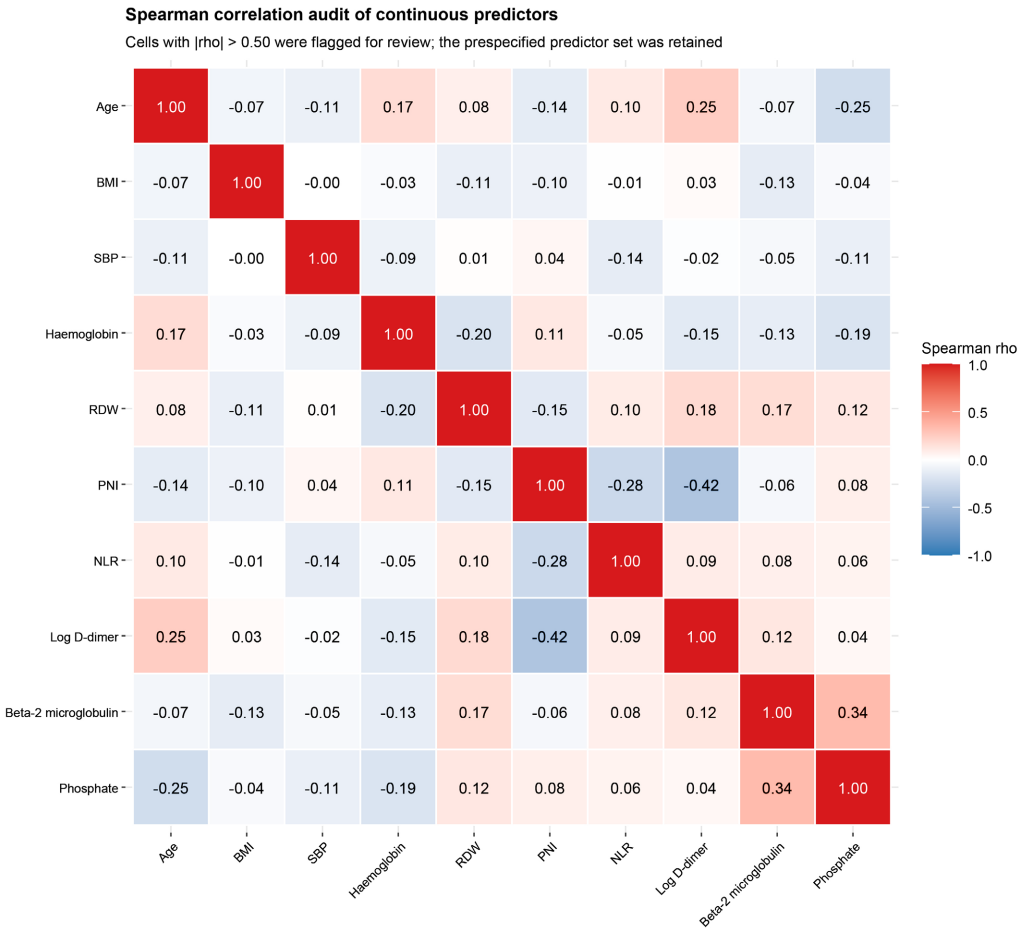

Figure S2. Spearman correlation matrix of the continuous LASSO-selected predictors. Cell values represent Spearman correlation coefficients, and no predictor pair exceeded the prespecified absolute correlation threshold of 0.50.

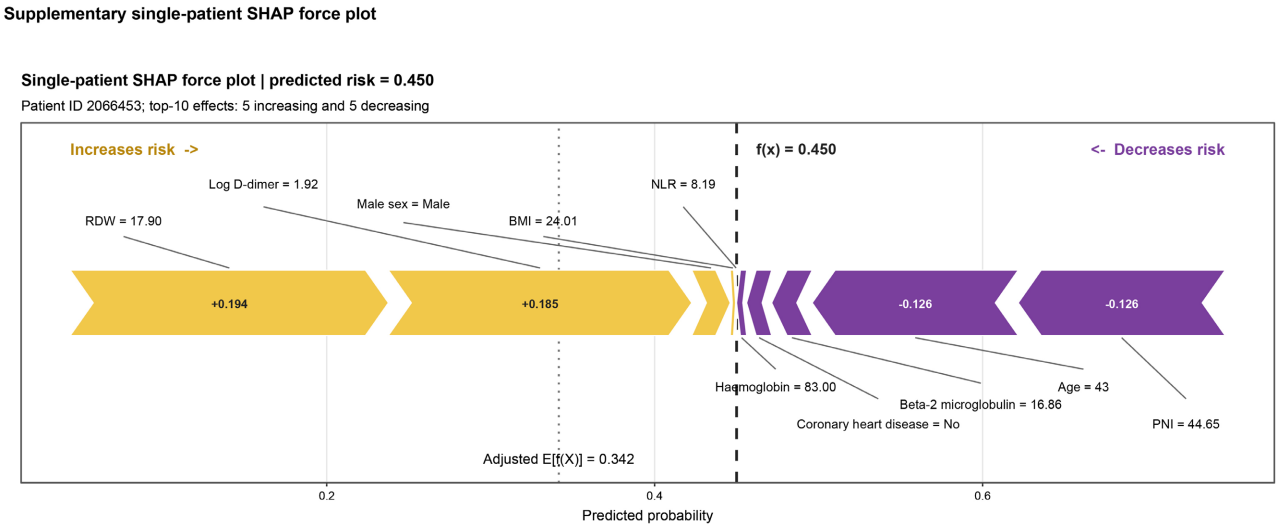

Figure S3. Individual SHAP explanation for an example patient. Contributions of individual features that increased or decreased the predicted probability relative to the adjusted background expectation.