Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning Prediction of In-Hospital Mortality in Elderly ICU Patients with COPD and Heart Failure: A Dual-Database Study

Jiayi Lin
Suining Central Hospital
Shuaixian Han
Suining Central Hospital
Ruhao Wu
First Affiliated Hospital of Zhengzhou University
Jianping Liu
Suining Central Hospital
Nan Xie
1051144840@qq.com

Suining Central Hospital

Research Article

Keywords: Machine learning, chronic obstructive pulmonary disease, heart failure, in-hospital mortality, intensive care unit, MIMIC-IV, eICU, SHAP

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11158956/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Machine Learning Prediction of In-Hospital Mortality in Elderly ICU Patients with COPD and Heart Failure: A Dual-Database Study

Running title: ML prediction of mortality in elderly COPD with heart failure

Jiayi Lin1†, Shuaixian Han2†, Ruhao Wu3, Jianping Liu2#, Nan Xie1*

1 Department of Nursing, Suining Central Hospital, Suining, Sichuan 629000, China.

2 Department of Cardiovascular Surgery, Suining Central Hospital, Suining, Sichuan 629000, China.

3 Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, Henan, China.

* Corresponding author: Nan Xie, Department of Nursing, Suining Central Hospital, Suining, Sichuan 629000, China. Email: 1051144840@qq.com

Co-corresponding author: Jianping Liu, Department of Cardiovascular Surgery, Suining Central Hospital, Suining, Sichuan 629000, China. Email: 18928939@qq.com

† Jiayi Lin and Shuaixian Han contributed equally to this work as co-first authors.

Abstract

Background. Elderly patients with coexisting chronic obstructive pulmonary disease (COPD) and heart failure (HF) represent a high-risk ICU population with substantial in-hospital mortality. Traditional scoring systems rely on static physiologic thresholds and may not fully capture the complex, nonlinear interactions among demographic, laboratory, inflammatory, nutritional, and comorbidity factors. We aimed to develop and externally validate a machine learning (ML) model for predicting in-hospital mortality in this population.

Methods. This retrospective cohort study used the Medical Information Mart for Intensive Care IV (MIMIC-IV, v3.1) as the training cohort and the eICU Collaborative Research Database (eICU v2.0) as the external validation cohort. Elderly patients (≥65 years) with concurrent COPD and HF admitted to the ICU were included (MIMIC-IV: n = 2,569, 18.4% mortality; eICU: n = 2,132, 13.8% mortality). Thirty-three candidate variables encompassing demographics, laboratory values, inflammatory and nutritional indices, comorbidities, and illness severity were harmonized across databases. Missing values were imputed using Multiple Imputation by Chained Equations (MICE). Feature selection comprised variance inflation factor (VIF) filtering (threshold = 5), followed by Least Absolute Shrinkage and Selection Operator (LASSO) regression and Boruta algorithm. Eight ML models were trained with hyperparameter optimization and compared. The best-performing model was evaluated on the internal test set and the external eICU cohort. Model performance was assessed by area under the receiver operating characteristic curve (AUC), calibration, and decision curve analysis (DCA). SHapley Additive exPlanations (SHAP) values were used for model interpretability.

Results. VIF filtering removed 8 collinear variables (retaining 25), and LASSO selected 11 features: sepsis, age, red cell distribution width, albumin, lactate, monocyte count, neutrophil-to-

lymphocyte ratio, platelet-to-lymphocyte ratio, monocyte-to-lymphocyte ratio, neutrophil-to-albumin ratio, and illness severity score. The Random Forest model achieved the highest performance with an internal test AUC of 0.769 (95% confidence interval [CI]: 0.729–0.811) and an external validation AUC of 0.723 (95% CI: 0.694–0.754). Brier scores were 0.129 and 0.114, respectively. Calibration curves showed good agreement between predicted and observed probabilities. DCA demonstrated positive net benefit across clinically relevant threshold probabilities. SHAP analysis identified sepsis, illness severity, lactate, and neutrophil-to-lymphocyte ratio as the most influential predictors.

Conclusions. A Random Forest model integrating routinely available clinical variables demonstrated good discrimination and calibration for predicting in-hospital mortality in elderly ICU patients with COPD and HF, with robust external validation. The combination of inflammatory, nutritional, and physiologic markers provides complementary prognostic information. This tool may assist clinicians in early risk stratification and clinical decision-making for this high-risk population.

Keywords: Machine learning; chronic obstructive pulmonary disease; heart failure; in-hospital mortality; intensive care unit; MIMIC-IV; eICU; SHAP

Background

Chronic obstructive pulmonary disease (COPD) and heart failure (HF) frequently coexist, particularly among elderly patients, owing to shared risk factors including smoking, systemic inflammation, and age-related cardiovascular remodeling [1–3]. Patients with concurrent COPD and HF are at heightened risk for acute decompensation, respiratory failure, and circulatory collapse, often necessitating intensive care unit (ICU) admission [2–4]. Despite advances in guideline-directed medical therapy and supportive care, in-hospital mortality among critically ill elderly patients with this dual burden remains high, ranging from 14% to 22% in recent cohorts [3–5]. Early and accurate identification of patients at greatest risk of death is essential for clinical decision-making, resource allocation, and timely escalation of care.

Traditional prognostic scoring systems, such as the Sequential Organ Failure Assessment (SOFA) and the Acute Physiology and Chronic Health Evaluation (APACHE), are widely used in the ICU [6–7]. However, these scores were developed for general critically ill populations, rely on fixed linear combinations of physiologic variables, and do not account for the complex, nonlinear interactions among demographic, laboratory, inflammatory, nutritional, and comorbidity factors that characterize elderly patients with cardiopulmonary comorbidity [8–9]. Moreover, these systems do not incorporate composite inflammatory and nutritional biomarkers that have emerged as important predictors in critical illness [10–12].

Inflammatory and nutritional biomarkers have gained attention as accessible prognostic indicators in critical illness. The neutrophil-to-lymphocyte ratio (NLR), systemic immune-inflammation index (SII), and platelet-to-lymphocyte ratio (PLR) reflect systemic inflammatory burden and immune dysregulation [10–11]. Nutritional indices, such as the geriatric nutritional risk index (GNRI), prognostic nutritional index (PNI), and albumin-based composite ratios (e.g.,

neutrophil-to-albumin ratio [NAR], red cell distribution width-to-albumin ratio [RAR]), integrate nutritional reserve with inflammatory stress [12–14]. These indices have been associated with mortality in COPD, HF, and sepsis individually, yet their combined contribution within an ML framework for elderly patients with both COPD and HF has not been explored.

Machine learning (ML) offers advantages over traditional scoring by capturing nonlinear relationships, high-order interactions, and complex patterns in high-dimensional clinical data [15–17]. Several studies have applied ML for mortality prediction in COPD or HF separately [18–20], but few have specifically addressed the overlap population of elderly patients with both conditions. Furthermore, external validation in independent cohorts is frequently lacking, limiting the generalizability of existing models [21–22].

Therefore, we conducted a retrospective dual-database study to develop and externally validate an ML model for predicting in-hospital mortality in elderly ICU patients with coexisting COPD and HF. We constructed a comprehensive feature set of 33 candidate variables spanning demographics, laboratory values, inflammatory and nutritional indices, comorbidities, and illness severity, and applied a systematic feature selection pipeline (VIF → LASSO + Boruta). We compared eight ML models and evaluated the best-performing model on an independent external cohort for discrimination, calibration, and clinical utility. We also employed SHAP analysis to provide model interpretability.

Methods

Data sources

This study used two publicly available, de-identified critical care databases. The Medical Information Mart for Intensive Care IV (MIMIC-IV, v3.1) contains comprehensive clinical data

for ICU admissions at the Beth Israel Deaconess Medical Center (Boston, MA, USA) between 2008 and 2022 [23]. The eICU Collaborative Research Database (eICU v2.0) contains data from over 200,000 ICU admissions across 208 hospitals in the United States [24]. One author (RW) completed the required training and was granted access to both databases (PhysioNet credential: 9810312). As both databases contain fully de-identified data, the requirement for informed consent was waived. The study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [25].

Study population

In the MIMIC-IV cohort, we identified elderly patients (≥65 years) with concurrent diagnoses of COPD (ICD-9 codes 491, 492, 496; ICD-10 codes J41–J44) and heart failure (ICD-9 code 428; ICD-10 codes I50) who were admitted to the ICU for the first time. In the eICU cohort, patients with COPD and HF were identified by diagnosis string matching and ICD-9 codes. Patients were excluded if they lacked outcome data. The final analytic cohorts comprised 2,569 patients in MIMIC-IV and 2,132 patients in eICU.

Outcome

The primary endpoint was in-hospital mortality, defined as death during the index hospital admission. In MIMIC-IV, mortality was ascertained from hospital discharge records. In eICU, mortality was defined as hospital death (discharge status "Expired").

Candidate variables

33 candidate variables were harmonized across both databases, including:

- **Demographics and laboratory values (7 continuous):** age, body mass index (BMI), white blood cell count, platelet count, red cell distribution width (RDW), serum albumin, and blood lactate.

- **Inflammatory indices (9 continuous):** neutrophil count, lymphocyte count, monocyte count, systemic immune-inflammation index (SII = neutrophil × platelet / lymphocyte), systemic inflammation response index (SIRI = neutrophil × monocyte / lymphocyte), neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), monocyte-to-lymphocyte ratio (MLR), and neutrophil-to-platelet ratio (NPR).

- **Nutritional indices (7 continuous):** GNRI, PNI, NAR, lymphocyte-to-albumin ratio (LAR), monocyte-to-albumin ratio (MAR), platelet-to-albumin ratio (PAR), and RAR.

- **Comorbidities and treatments (9 binary):** male sex, hypertension, diabetes, coronary artery disease, chronic kidney disease (CKD), cancer, atrial fibrillation, sepsis, and vasopressor use.

- **Illness severity (1 continuous):** the severity score was derived from the SOFA score in MIMIC-IV and the APACHE Acute Physiology Score (APS) in eICU. Because these two scores use different scales and variable sets, each was min-max normalized within its respective cohort to produce a cross-database-compatible severity score ranging from 0 to 1. This approach allows the model to learn the relative severity gradient within each cohort while maintaining a common feature representation across databases.

All laboratory values were taken from the first measurement within 24 hours of ICU admission. Binary comorbidity variables were ascertained from ICD-9/ICD-10 diagnosis codes (MIMIC-IV) or diagnosis string matching (eICU).

Missing data and imputation

Missing values were handled using Multiple Imputation by Chained Equations (MICE) via the IterativeImputer (scikit-learn), with a RandomForestRegressor as the imputation estimator (n_estimators = 20, max_iter = 10, random_state = 20220905). The imputer was fitted on the MIMIC-IV training data and then applied to transform the eICU cohort, ensuring that imputation did not leak information from the external validation set. After imputation, binary variables were rounded to 0 or 1 and continuous variables were clipped to non-negative values.

Feature selection

The feature selection pipeline comprised three sequential steps (Figure 1):

Step 1: Variance Inflation Factor (VIF) filtering. VIF was computed for all 33 variables on the training set using ordinary least squares regression. Variables with VIF > 5 were iteratively removed (highest VIF first) until all remaining variables had VIF < 5. This step removed 8 variables—prognostic nutritional index (PNI, VIF = 63.75), neutrophil count (VIF = 56.12), platelet-to-albumin ratio (PAR, VIF = 33.35), lymphocyte-to-albumin ratio (LAR, VIF = 29.20), white blood cell count (VIF = 24.35), red cell distribution width-to-albumin ratio (RAR, VIF = 16.55), systemic immune-inflammation index (SII, VIF = 7.73), and monocyte-to-albumin ratio (MAR, VIF = 6.74)—leaving 25 variables (Table S1, Figure S1).

Step 2: LASSO regression. L1-penalized logistic regression was applied to the 25 VIF-retained variables. The regularization parameter (lambda) was selected by 10-fold stratified cross-validation using the one-standard-error rule (lambda.1se), which favors sparsity. LASSO selected 11 variables: sepsis, age, RDW, albumin, lactate, monocyte count, NLR, PLR, MLR, NAR, and illness severity score.

Step 3: Boruta algorithm. The Boruta algorithm, a wrapper method based on random forest, was applied to the 25 VIF-retained variables. Boruta selected 4 variables: sepsis, lactate, NLR, and illness severity score.

Step 4: Intersection and fallback. The intersection of LASSO and Boruta comprised 4 variables (sepsis, lactate, NLR, illness severity score). Because the intersection contained fewer than 5 variables (deemed insufficient for a robust model), the LASSO-selected set of 11 variables was used as the final feature set (Table S2). This decision was made a priori to prioritize the more conservative and clinically interpretable LASSO selection over the smaller Boruta intersection.

Machine learning models

Eight ML models were trained and compared:

1. **Random Forest (RF):** an ensemble of decision trees with hyperparameters optimized over n_estimators (500, 1000, 2000), max_depth (3, 5, 10, None), and min_samples_leaf (1, 5, 10).

2. **Bagged Classification and Regression Tree (Bagged CART):** bootstrap-aggregated decision trees with n_estimators (50, 100, 200, 500) and max_samples (0.5, 0.7, 1.0).

3. **Naive Bayes:** Gaussian Naive Bayes classifier (no hyperparameters).

4. **Neural Network (NN):** multi-layer perceptron with hidden_layer_sizes [(5,), (10,), (15,), (10,5)], alpha [1e-3, 1e-2, 0.1], and learning_rate_init [1e-3, 1e-2], using early stopping.

5. **K-Nearest Neighbors (KNN):** with n_neighbors [3, 5, 7, 11, 15] and weights [uniform, distance].

6. **Classification and Regression Tree (CART):** single decision tree with max_depth [2, 3, 5, 7, 10], min_samples_split [2, 5, 10], and min_samples_leaf [1, 5, 10].

7. **Gradient Boosting Machine (GBM):** with n_estimators [100, 200, 300], learning_rate [0.01, 0.05, 0.1], max_depth [3, 5, 7], and subsample [0.7, 0.8, 1.0].

8. **Boosted Generalized Linear Model (Boosted GLM):** L1-penalized logistic regression with C values from 1e-3 to 1e1 (8 values on log scale).

Hyperparameter optimization was performed using RandomizedSearchCV (20 iterations, or fewer if the grid was smaller) with repeated stratified 5-fold cross-validation (3 repeats), optimized for AUC. All models were implemented in Python using scikit-learn. The random state was fixed at 123456 for reproducibility.

Training and validation strategy

The MIMIC-IV cohort was split into a training set (70%) and an internal test set (30%) using stratified random sampling to preserve the mortality rate. The eICU cohort served as a completely independent external validation set. Continuous variables were standardized using StandardScaler fitted on the training set and applied to both the internal test and external validation sets. The optimal classification threshold was determined by maximizing the Youden index on the training set.

Model evaluation

Model performance was assessed using the following metrics on both the internal test set and the external eICU validation set:

- **Discrimination:** area under the receiver operating characteristic curve (AUC) with 95% confidence intervals computed by bootstrapping (1000 resamples).

- **Calibration:** Brier score and calibration curves (10 quantile bins) comparing predicted probabilities with observed mortality.

- **Clinical utility:** decision curve analysis (DCA) quantifying net benefit across threshold probabilities, compared with "treat all" and "treat none" strategies.

- **Classification metrics:** accuracy, F1 score, precision, and recall at the optimal threshold.

Model interpretability

SHapley Additive exPlanations (SHAP) values were computed for the best-performing model using TreeExplainer on the internal test set (500 samples). SHAP bar chart (mean absolute SHAP values) and SHAP beeswarm plot (distribution of SHAP values by feature value) were generated to rank feature importance and visualize the direction and magnitude of each feature's contribution to the prediction.

Statistical analysis

Continuous variables are presented as median (interquartile range [IQR]) and compared using the Mann–Whitney U test between survivors and non-survivors. Categorical variables are presented as counts (percentages) and compared using the chi-square test. All statistical analyses were performed in Python (version 3.11) using scikit-learn, statsmodels, and scipy. Two-sided $P < 0.05$ was considered statistically significant.

Results

Baseline characteristics

The study included 2,569 elderly patients (≥65 years) with concurrent COPD and HF in the MIMIC-IV cohort and 2,132 in the eICU cohort. The in-hospital mortality rate was 18.4% (472/2569) in MIMIC-IV and 13.8% (295/2132) in eICU.

In the eICU cohort (Table 1), non-survivors were older (median 78 vs 75 years, P = 0.002), had higher white blood cell counts (11.7 vs 10.4 × 10^3/μL, P < 0.001), higher RDW (16.25% vs 15.50%, P < 0.001), lower albumin (3.00 vs 3.09 g/dL, P < 0.001), and higher lactate (2.10 vs 1.91 mmol/L, P < 0.001). Among inflammatory indices, non-survivors had significantly higher SII (3186 vs 2411, P < 0.001), SIRI (9.17 vs 6.48, P < 0.001), NLR (15.98 vs 13.34, P < 0.001), MLR (1.00 vs 0.82, P < 0.001), and NPR (0.05 vs 0.04, P < 0.001). Among nutritional indices, PNI was lower (33.63 vs 35.06, P < 0.001), while NAR (3.31 vs 2.86, P < 0.001) and RAR (5.41 vs 5.24, P < 0.001) were higher in non-survivors. The severity score was significantly higher in non-survivors (0.36 vs 0.27, P < 0.001). Comorbidities significantly associated with mortality included CKD (54.9% vs 40.3%, P < 0.001), cancer (13.2% vs 6.2%, P < 0.001), atrial fibrillation (45.1% vs 32.4%, P < 0.001), sepsis (36.3% vs 18.7%, P < 0.001), and vasopressor use (23.7% vs 8.9%, P < 0.001).

In the MIMIC-IV cohort (Table 2), a similar pattern was observed. Non-survivors were older (median 79 vs 76 years, P < 0.001) and had higher white blood cell counts (12.55 vs 10.50 × 10^3/μL, P < 0.001), higher RDW (15.90% vs 15.10%, P < 0.001), and higher lactate (2.00 vs 1.60 mmol/L, P < 0.001). Among inflammatory indices, SII (3722 vs 3077, P < 0.001), SIRI (11.00 vs 7.55, P < 0.001), NLR (17.01 vs 16.13, P < 0.001), MLR (0.99 vs 0.87, P < 0.001), and NPR (0.06 vs 0.05, P < 0.001) were higher in non-survivors. NAR was significantly higher (3.63 vs 3.09, P < 0.001), while albumin showed no significant difference (P = 0.733). The severity score was higher in non-survivors (0.29 vs 0.14, P < 0.001). CKD (47.7% vs 39.9%, P = 0.002), cancer (14.8% vs 9.2%, P < 0.001), sepsis (44.7% vs 14.9%, P < 0.001), and vasopressor use (49.6% vs 28.3%, P < 0.001) were significantly more prevalent among non-survivors.

Feature selection

The feature selection pipeline is summarized in Figure 1 and detailed in Tables S1 and S2, with VIF values visualized in Figure S1.

Starting from 33 candidate variables, VIF filtering (threshold = 5) removed 8 variables due to multicollinearity, retaining 25 variables with a maximum VIF of 3.98 after filtering. The removed variables included indices that are mathematically derived from shared component variables (e.g., PNI and albumin, neutrophil count and NLR, PAR and platelet count), reflecting expected structural collinearity.

LASSO regression (lambda.1se = 10.0) selected 11 of 25 variables. The Boruta algorithm confirmed 4 of these 11 as strongly important (sepsis, lactate, NLR, and illness severity score, all with Boruta rank 1). Because the LASSO–Boruta intersection contained only 4 variables (fewer than the pre-specified threshold of 5), the LASSO-selected set of 11 variables was retained as the final feature set.

The final 11 predictors were: sepsis, age, red cell distribution width, albumin, lactate, monocyte count, neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio, monocyte-to-lymphocyte ratio, neutrophil-to-albumin ratio, and illness severity score.

Model performance

Eight ML models were trained on the 11 selected features and compared on both the internal MIMIC-IV test set and the external eICU validation set (Table 3, Table S3, Figure 2).

The Random Forest model achieved the best overall performance, with an internal test AUC of 0.769 (95% CI: 0.729–0.811) and an external validation AUC of 0.723 (95% CI: 0.694–0.754). The Boosted GLM and GBM achieved comparable performance (Boosted GLM: AUC 0.768/0.723; GBM: AUC 0.764/0.719). The Bagged CART showed slightly lower discrimination

(AUC 0.751/0.712), followed by Naive Bayes (AUC 0.728/0.699), KNN (AUC 0.721/0.681), CART (AUC 0.674/0.658), and Neural Network (AUC 0.608/0.562).

For the Random Forest model, Brier scores were 0.129 (internal test) and 0.114 (external validation), indicating good calibration. At the optimal threshold (0.186 for internal test, 0.177 for external validation), the model achieved an F1 score of 0.477 (internal) and 0.344 (external), recall of 0.683 (internal) and 0.698 (external), and accuracy of 0.724 (internal) and 0.632 (external). The relatively high recall indicates that the model correctly identified approximately 68–70% of non-survivors.

Calibration

Calibration curves (Figure 4) demonstrated good agreement between predicted and observed mortality probabilities for the Random Forest model in both the internal test set and the external eICU validation set, with Brier scores of 0.129 and 0.114, respectively. The calibration curves closely followed the diagonal reference line, indicating that predicted probabilities were well-calibrated and not systematically overestimating or underestimating risk.

Decision curve analysis

Decision curve analysis (Figure 3) showed that the Random Forest model provided positive net benefit across a wide range of threshold probabilities (approximately 5–35%) in both the internal test set and the external validation set, compared with the "treat all" and "treat none" strategies. The model's net benefit exceeded that of "treat all" at threshold probabilities above approximately 15%, indicating that using the model to guide clinical decisions would result in a more favorable benefit–harm trade-off than treating all patients.

Model interpretability (SHAP analysis)

SHAP analysis (Figure 5) revealed the relative importance and direction of each feature's contribution to the Random Forest model's predictions. The SHAP bar chart ranked features by mean absolute SHAP value, while the beeswarm plot showed the distribution and direction of SHAP values across individual predictions.

The four Boruta-confirmed features—sepsis, illness severity score, lactate, and NLR—were among the top-ranked features by SHAP importance. Higher values of sepsis (presence), illness severity, lactate, and NLR were associated with higher SHAP values, indicating increased predicted mortality risk. Conversely, higher albumin values were associated with lower SHAP values (protective effect). The SHAP beeswarm plot further demonstrated the nonlinear and interactive nature of these relationships, with the direction and magnitude of feature effects varying across individual patients.

Discussion

In this dual-database study, we developed and externally validated a machine learning model for predicting in-hospital mortality in elderly ICU patients with coexisting COPD and HF. Using a systematic feature selection pipeline (VIF → LASSO + Boruta), we identified 11 routinely available clinical variables—spanning inflammatory, nutritional, physiologic, and comorbidity domains—and compared eight ML models. The Random Forest model achieved the best discrimination, with an internal AUC of 0.769 and an external validation AUC of 0.723, demonstrating robust generalizability across independent cohorts. Calibration and decision curve analyses confirmed that the model's predictions were well-calibrated and clinically useful.

Key findings

Several features of this study merit discussion. First, the feature selection pipeline effectively reduced 33 candidate variables to a parsimonious set of 11 predictors. VIF filtering removed variables with structural collinearity (e.g., PNI and albumin share the same component; NLR and neutrophil count share the same numerator), while LASSO further eliminated less informative variables. The final set retained a clinically coherent combination of inflammation markers (NLR, PLR, MLR, monocyte count), a nutrition-inflammation composite (NAR), a tissue perfusion marker (lactate), a routine hematologic marker (RDW), a demographic factor (age), a physiologic severity measure, and a key clinical diagnosis (sepsis). This combination reflects the multidimensional pathophysiology of elderly patients with COPD and HF, where systemic inflammation, nutritional depletion, impaired tissue perfusion, and acute illness severity converge [1–5].

Second, the Random Forest model outperformed traditional single-tree methods (CART) and simpler classifiers (Naive Bayes, KNN), consistent with the advantage of ensemble methods in capturing nonlinear interactions [15–17]. The Neural Network performed poorly (AUC 0.608/0.562), possibly due to the relatively small sample size for deep learning and the absence of complex feature representations. The Boosted GLM and GBM, which also use ensemble or boosting strategies, achieved performance comparable to Random Forest, reinforcing the value of ensemble approaches for this prediction task.

Third, SHAP analysis provided clinically interpretable insights. The top features—sepsis, illness severity, lactate, and NLR—are well-established prognostic factors in critical illness. Sepsis represents acute systemic infection and inflammation; illness severity (SOFA/APACHE) captures multi-organ dysfunction; lactate reflects tissue hypoperfusion and metabolic stress; and NLR quantifies the neutrophil-driven inflammatory response [6–14]. The finding that these four

features were also the only variables confirmed by the Boruta algorithm (all rank 1) underscores their robust importance. Albumin, although not selected by Boruta, was retained by LASSO and demonstrated a protective direction in SHAP (higher albumin → lower mortality), consistent with its role as a nutritional and inflammatory marker [14].

Comparison with previous studies

Our AUC of 0.769 (internal) and 0.723 (external) is comparable to previously reported ML models for ICU mortality in COPD and HF populations. A systematic review of ML models for COPD exacerbation outcomes reported AUCs ranging from 0.70 to 0.85, with external validation frequently lacking [18]. For HF readmission and mortality prediction, AUCs of 0.70–0.80 are typical [19–20]. Our study extends these findings by focusing specifically on the elderly overlap population and by incorporating inflammatory and nutritional biomarkers not commonly used in existing models.

Clinical implications

The practical implications of this model are twofold. First, the 11 selected variables are routinely available from standard laboratory tests and clinical assessments within the first 24 hours of ICU admission, making the model feasible for real-time implementation. Second, the DCA demonstrated that the model provides positive net benefit across clinically relevant threshold probabilities (5–35%), suggesting that clinicians could use the model's risk estimates to guide decisions regarding intensity of monitoring, early goal-directed therapy, and goals-of-care discussions. The high recall (68–70%) indicates that the model is sensitive in identifying non-survivors, which is particularly important in a high-risk population where missing a patient at elevated risk could have serious consequences.

Strengths and limitations

This study has several strengths. First, the dual-database design with an independent external validation cohort (eICU, 2,132 patients from 208 hospitals) provides strong evidence of generalizability. Second, the systematic feature selection pipeline (VIF → LASSO + Boruta) reduces overfitting and multicollinearity while maintaining clinical interpretability. Third, the comprehensive evaluation (discrimination, calibration, DCA, and SHAP) provides a thorough assessment of model performance and interpretability. Fourth, the inclusion of 13 inflammatory and nutritional indices as candidate variables reflects the growing recognition of these biomarkers in critical care prognosis.

Several limitations should be acknowledged. First, this is a retrospective study using electronic health records, which are subject to measurement bias, coding errors, and selection bias. Second, the severity score was derived from different scoring systems (SOFA in MIMIC-IV vs. APACHE APS in eICU) and harmonized by min-max normalization within each cohort; while this approach allows cross-database compatibility, it may not fully capture differences in the underlying physiologic constructs. Third, mechanical ventilation was excluded because no patient in the analytic cohort had this variable recorded, which may reflect the specific extraction criteria or database limitations and may limit the model's applicability to populations where mechanical ventilation is prevalent. Fourth, the model was developed and validated in US ICU populations; its performance in other healthcare settings or populations may differ. Fifth, the relatively modest AUC (0.723 in external validation), while clinically useful, indicates that a substantial proportion of mortality risk remains unexplained, and additional variables (e.g., echocardiographic parameters, biomarkers such as natriuretic peptides, medication data) may further improve performance.

Conclusions

In this dual-database study, a Random Forest model integrating 11 routinely available clinical variables—including inflammatory, nutritional, and physiologic markers—demonstrated good discrimination and calibration for predicting in-hospital mortality in elderly ICU patients with COPD and HF. The model was robust to external validation in an independent multicenter cohort. Sepsis, illness severity, lactate, and NLR emerged as the most influential predictors. This model may serve as a practical tool for early risk stratification and clinical decision-making in this high-risk population. Future studies should validate the model in prospective settings and explore the incorporation of additional clinical variables and longitudinal trajectories.

Abbreviations

APACHE: Acute Physiology and Chronic Health Evaluation; AUC: area under the receiver operating characteristic curve; BMI: body mass index; CART: Classification and Regression Tree; CI: confidence interval; CKD: chronic kidney disease; COPD: chronic obstructive pulmonary disease; DCA: decision curve analysis; GBM: Gradient Boosting Machine; GLM: generalized linear model; GNRI: geriatric nutritional risk index; HF: heart failure; ICU: intensive care unit; IQR: interquartile range; LAR: lymphocyte-to-albumin ratio; LASSO: Least Absolute Shrinkage and Selection Operator; MAR: monocyte-to-albumin ratio; MICE: Multiple Imputation by Chained Equations; MLR: monocyte-to-lymphocyte ratio; NAR: neutrophil-to-albumin ratio; NLR: neutrophil-to-lymphocyte ratio; NPR: neutrophil-to-platelet ratio; PAR: platelet-to-albumin ratio; PNI: prognostic nutritional index; PLR: platelet-to-lymphocyte ratio; RAR: red cell distribution width-to-albumin ratio; RDW: red cell distribution width; RF: Random Forest; SHAP: SHapley Additive exPlanations; SII: systemic immune-inflammation index; SIRI: systemic inflammation response index; SOFA: Sequential Organ Failure Assessment; VIF: variance inflation factor.

Declarations

Ethics approval and consent to participate

Both MIMIC-IV and eICU databases contain fully de-identified data. The requirement for informed consent was waived.

Consent for publication

Not applicable.

Clinical trial number

Not applicable.

Availability of data and materials

The MIMIC-IV and eICU databases are publicly available at https://physionet.org/ after completion of required training and data use agreement. The analytic code is available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Acknowledgments

Not applicable.

Authors' contributions

JL and SH contributed equally to this work. JL, SH, and RW: Conceptualization, Formal analysis, Methodology, Software, Writing – original draft. JPL and NX: Conceptualization,

Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. All authors read and approved the final manuscript.

References

1. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. 2024 Report. Available at: https://goldcopd.org

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

3. Müllerova D, Agusti A, Erqou S, Mapel DW. Cardiovascular comorbidity in COPD: systematic literature review. Chest. 2013;144(4):1163–1198.

4. Rutten FH, Cramer MJ, Grobbee DE, Hoes AW. Unrecognized heart failure in elderly patients with stable chronic obstructive pulmonary disease. Eur Heart J. 2005;26(18):1887–1894.

5. Hannink JD, Dekhuijzen PN, van Helvoort HA, et al. Heart failure and COPD: two comorbid diseases with a shared burden. Respir Med. 2021;186:106534.

6. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Med. 1996;22(7):707–710.

7. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818–829.

8. Brinkman S, Baken JJ, Riet G, et al. Validation of the SOFA score in the ICU: a prospective cohort study. Intensive Care Med. 2017;43(6):856–867.

9. Keegan MT, Gajic O, Afessa B. Severity of illness scoring systems in the intensive care unit. Crit Care Med. 2011;39(1):163–169.

10. Zahorec R. Ratio of neutrophil to lymphocyte counts—rapid and simple parameter of systemic inflammation and stress in critically ill. Bratisl Lek Listy. 2001;102(1):5–13.

11. Hu B, Yang XR, Xu Y, et al. Systemic immune-inflammation index predicts prognosis of patients after curative resection for hepatocellular carcinoma. Clin Cancer Res. 2014;20(23):6212–6222.

12. Bouillanne O, Morineau G, Dupont C, et al. Geriatric nutritional risk index: a new index for evaluating at-risk elderly medical patients. Am J Clin Nutr. 2005;82(4):777–782.

13. Buzby GP, Williford WD, Peterson OL, et al. A randomized clinical trial of total parenteral nutrition in malnourished surgical patients: the rationale and impact of clinical trials. Am J Clin Nutr. 1988;47(2 Suppl):357–365.

14. Vincent JL, Dubois MJ, Navickis RJ, Wilkes MM. Hypoalbuminemia in acute illness: is there a rationale for intervention? Ann Surg. 2003;237(3):319–321.

15. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347–1358.

16. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.

17. Johnson AEW, Ghassemi MM, Nemati S, et al. Machine learning and medical informatics in the intensive care unit. Lancet Digit Health. 2019;1(3):e136–e145.

18. Pinto JR, Sampaio L, Cardoso R, et al. Machine learning models for predicting in-hospital mortality in COPD: a systematic review. Respir Med. 2022;201:106940.

19. Frizzell JD, Liang L, Schulte PJ, et al. Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure. Circ Heart Fail. 2017;10(6):e003816.

20. Desai RJ, Wang SV, Vaduganathan M, et al. Comparison of machine learning methods with traditional models for clinical prediction. Circulation. 2020;141(19):1462–1467.

21. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). Ann Intern Med. 2015;162(1):55–63.

22. Luo Y, Szolovits P. Predicting in-hospital mortality using ensemble learning. AMIA Jt Summits Transl Sci Proc. 2017;2017:215–219.

23. Johnson AEW, Bulgarelli L, Pollard T, et al. MIMIC-IV (version 3.1). PhysioNet. 2023. https://doi.org/10.13026/6v2e-fc69

24. Pollard TJ, Johnson AEW, Raffa JD, et al. eICU Collaborative Research Database (version 2.0). PhysioNet. 2019. https://doi.org/10.13026/0c3w-jp48

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

Figure legends

Figure 1. Feature selection pipeline. (A) Variance inflation factor (VIF) bar chart for all 33 candidate variables, colored by retained (green) or removed (red) status. The dashed line indicates the VIF = 5 threshold. (B) LASSO coefficient paths across regularization parameters (lambda), showing the shrinkage of coefficients as lambda increases. The vertical dashed line

indicates the selected lambda (lambda.1se). (C) Cross-validation AUC across lambda values, with error bars representing standard deviation. (D) Boruta feature importance, with confirmed (green), tentative (yellow), and rejected (red) variables.

Figure 2. Model performance comparison across eight machine learning models. Bar charts show (A) AUC, (B) accuracy, (C) F1 score, (D) precision, and (E) recall for the internal MIMIC-IV test set and the external eICU validation set.

Figure 3. Decision curve analysis for the Random Forest model. Net benefit is plotted against threshold probability for the internal test set (left) and external eICU validation set (right). Solid lines represent the model; dashed lines represent "treat all"; dotted lines represent "treat none."

Figure 4. Calibration curves for the Random Forest model. Predicted probability is plotted against observed mortality for the internal test set (left) and external eICU validation set (right). The diagonal dashed line represents perfect calibration. Brier scores are shown in the legend.

Figure 5. SHAP (SHapley Additive exPlanations) analysis for the Random Forest model. (A) SHAP bar chart showing mean absolute SHAP values ranked by feature importance. (B) SHAP beeswarm plot showing the distribution of SHAP values for each feature, with color indicating feature value (blue = low, red = high).

Figure S1. Variance inflation factor (VIF) values for all 33 candidate variables, displayed as a horizontal bar chart. Variables with VIF > 5 (red) were removed; variables with VIF ≤ 5 (green) were retained for subsequent feature selection.

Tables

Table 1. Baseline characteristics of the eICU cohort by in-hospital mortality status (n = 2,132).

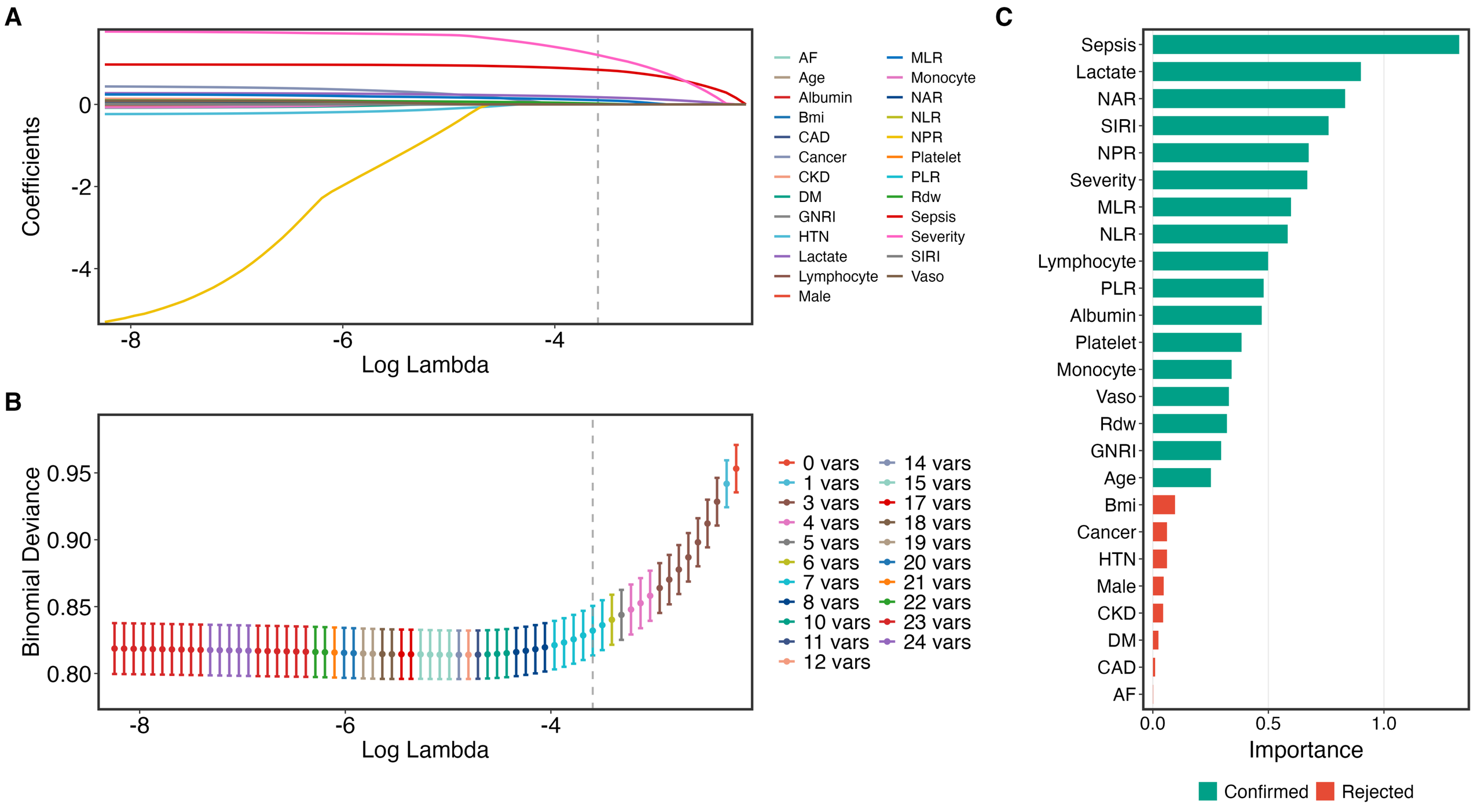
Table 2. Baseline characteristics of the MIMIC-IV cohort by in-hospital mortality status (n = 2,569).

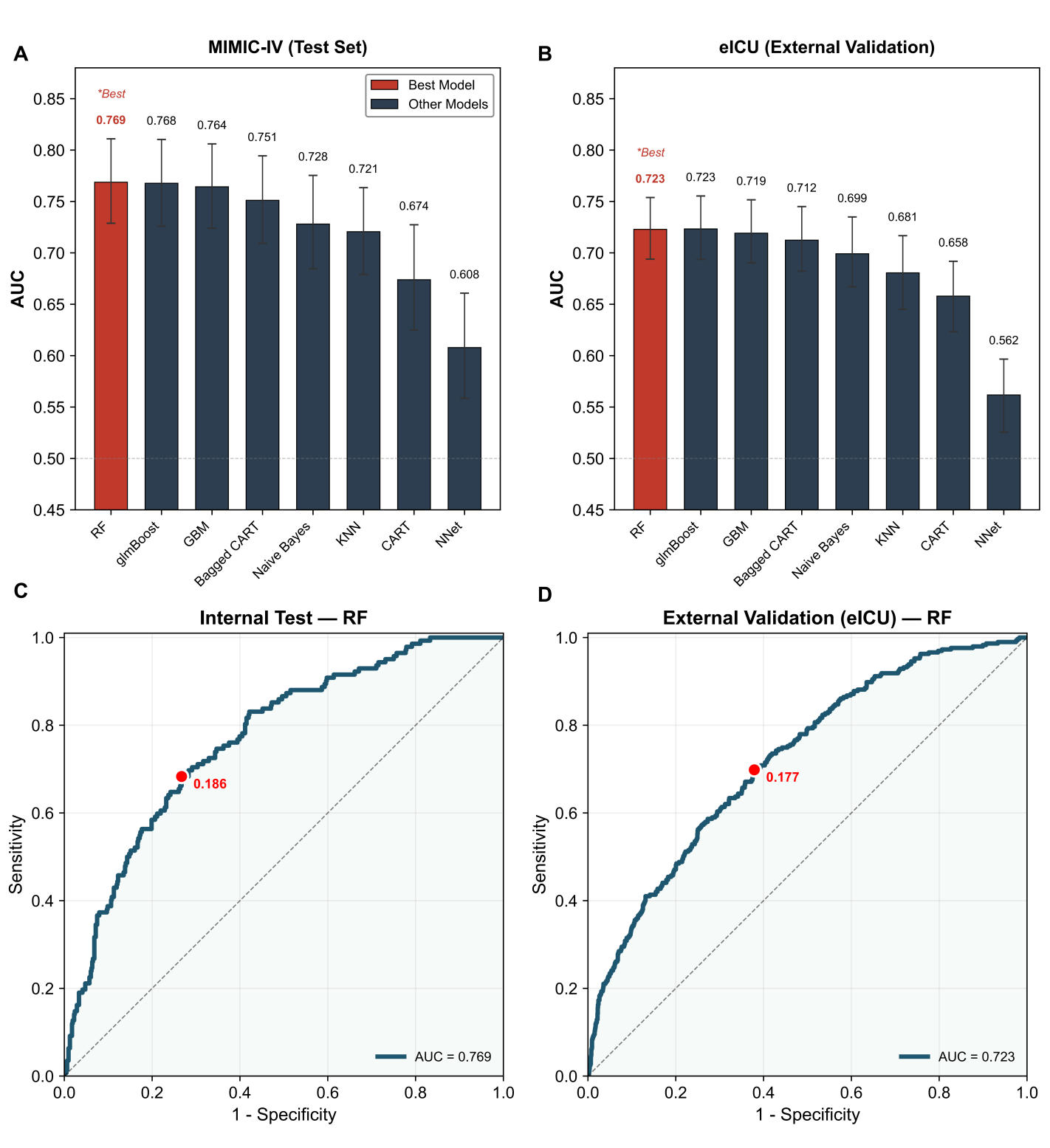
Table 3. Performance comparison of eight machine learning models on the internal MIMIC-IV test set and external eICU validation set.

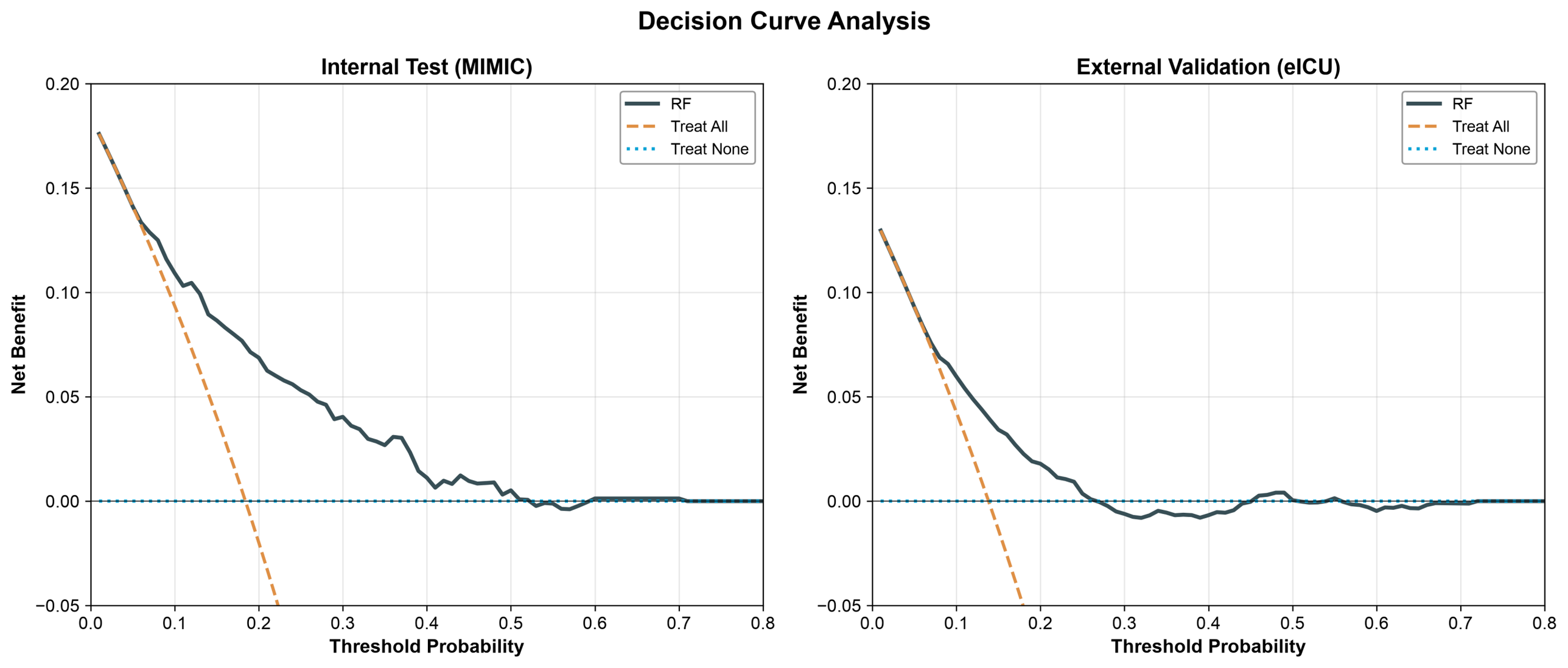
Table S1. Variance inflation factor (VIF) summary for all candidate variables.

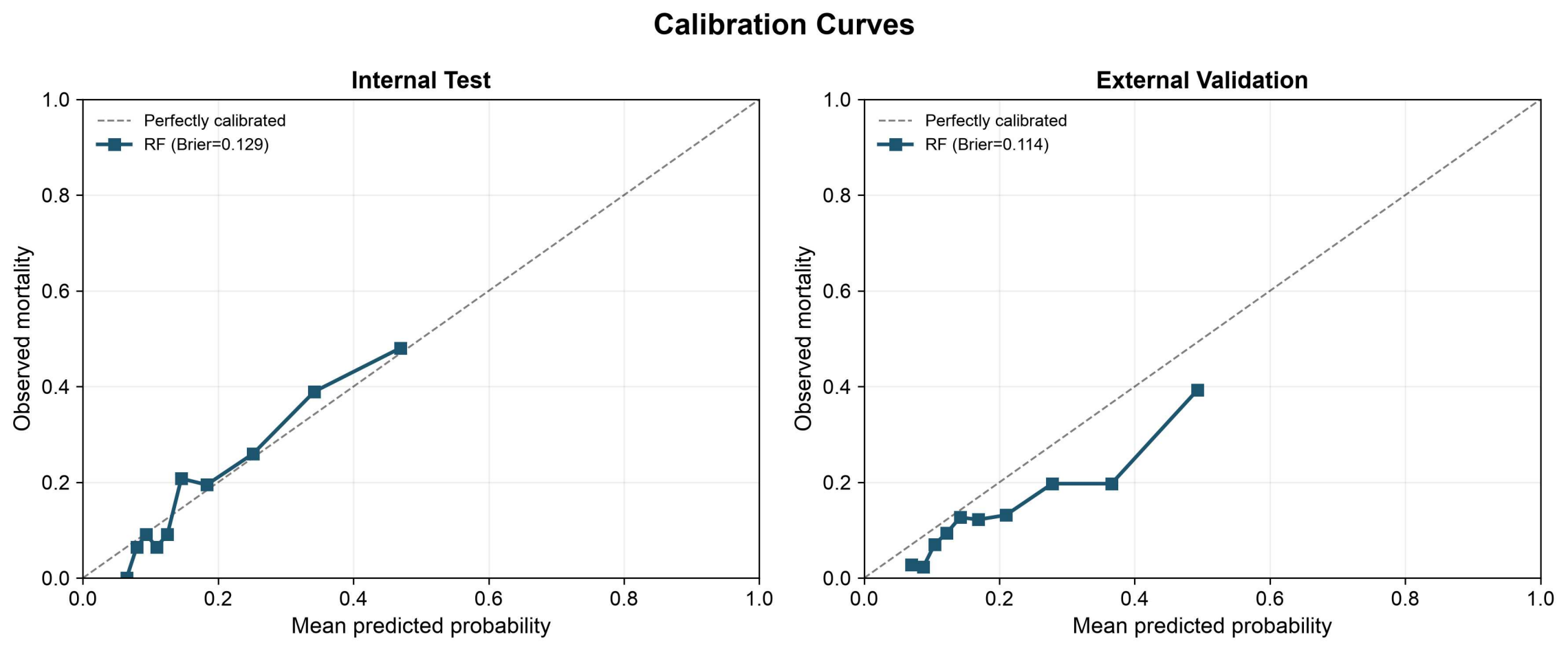
Table S2. Feature selection summary: VIF retention, LASSO selection, Boruta selection, and final selected variables.

Table S3. Area under the receiver operating characteristic curve (AUC) with 95% confidence intervals for eight machine learning models.


A
Coefficients
0
-2
-4
-8
-6
-4
Log Lambda
AF
Age
Albumin
Bmi
CAD
Cancer
CKD
DM
GNRI
HTN
Lactate
Lymphocyte
Male
MLR
Monocyte
NAR
NLR
NPR
Platelet
PLR
Rdw
Sepsis
Severity
SIRI
Vaso
B
Binomial Deviance
0.95
0.90
0.85
0.80
-8
-6
-4
Log Lambda
0 vars
1 vars
3 vars
4 vars
5 vars
6 vars
7 vars
8 vars
10 vars
11 vars
12 vars
14 vars
15 vars
17 vars
18 vars
19 vars
20 vars
21 vars
22 vars
23 vars
24 vars
C
Sepsis
Lactate
NAR
SIRI
NPR
Severity
MLR
NLR
Lymphocyte
PLR
Albumin
Platelet
Monocyte
Vaso
Rdw
GNRI
Age
Bmi
Cancer
HTN
Male
CKD
DM
CAD
AF
0.0
0.5
1.0
Importance
Confirmed
Rejected


A
MIMIC-IV (Test Set)
Best Model
Other Models
*Best
0.769
0.768
0.764
0.751
0.728
0.721
0.674
0.608
AUC
RF
glmBoost
GBM
Bagged CART
Naive Bayes
KNN
CART
NNet
B
eICU (External Validation)
*Best
0.723
0.723
0.719
0.712
0.699
0.681
0.658
0.562
C
Internal Test — RF
0.186
Sensitivity
1 - Specificity
AUC = 0.769
D
External Validation (eICU) — RF
0.177
AUC = 0.723


Decision Curve Analysis
Internal Test (MIMIC)
External Validation (eICU)
RF
Treat All
Treat None
Net Benefit
Threshold Probability
0.20
0.15
0.10
0.05
0.00
−0.05
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8


Calibration Curves
Internal Test
Perfectly calibrated
RF (Brier=0.129)
Observed mortality
Mean predicted probability
0.0
0.2
0.4
0.6
0.8
1.0
External Validation
Perfectly calibrated
RF (Brier=0.114)
Observed mortality
Mean predicted probability

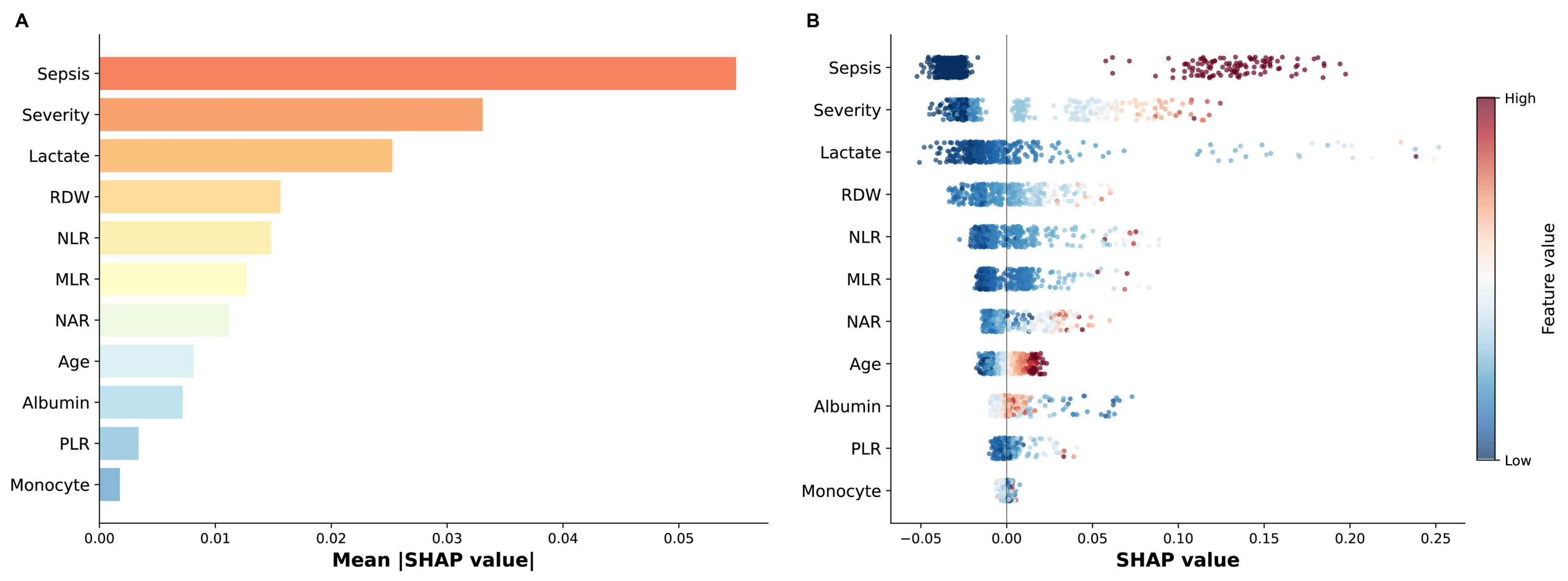

A
Sepsis
Severity
Lactate
RDW
NLR
MLR
NAR
Age
Albumin
PLR
Monocyte
0.00
0.01
0.02
0.03
0.04
0.05
Mean |SHAP value|
B
Sepsis
Severity
Lactate
RDW
NLR
MLR
NAR
Age
Albumin
PLR
Monocyte
−0.05
0.00
0.05
0.10
0.15
0.20
0.25
SHAP value
High
Low
Feature value

Table 1. Baseline Characteristics of Elderly Patients with COPD and Heart Failure (eICU Cohort)

Variable	Total (n=2132)	Survived (n=1837)	Died (n=295)	P value
Demographics & Laboratory				
Male, n (%)	1075 (50.4%)	920 (50.1%)	155 (52.5%)	0.470
Age (years), median [IQR]	76.00 [70.00, 82.00]	75.00 [70.00, 82.00]	78.00 [71.50, 83.00]	0.002
Body Mass Index (kg/m²), median [IQR]	28.63 [23.86, 34.47]	28.69 [23.81, 34.71]	28.31 [24.04, 33.88]	0.502
White Blood Cell Count (10³/μL), median [IQR]	10.60 [7.70, 14.30]	10.40 [7.64, 13.90]	11.70 [8.30, 16.90]	<0.001
Platelets (10³/μL), median [IQR]	202.00 [158.00, 259.00]	202.00 [158.00, 259.00]	210.00 [148.00, 256.50]	0.928
Red Cell Distribution Width (%), median [IQR]	15.60 [14.50, 17.15]	15.50 [14.40, 17.00]	16.25 [15.00, 17.90]	<0.001
Albumin (g/dL), median [IQR]	3.01 [2.72, 3.40]	3.09 [2.74, 3.40]	3.00 [2.60, 3.30]	<0.001
Lactate (mmol/L), median [IQR]	1.95 [1.51, 2.46]	1.91 [1.50, 2.40]	2.10 [1.60, 3.22]	<0.001
Neutrophils (10³/μL), median [IQR]	8.60 [6.24, 11.98]	8.50 [6.22, 11.56]	10.06 [6.53, 14.42]	<0.001
Lymphocytes (10³/μL), median [IQR]	0.74 [0.52, 1.25]	0.76 [0.52, 1.27]	0.66 [0.46, 1.12]	0.004
Monocytes (10³/μL), median [IQR]	0.70 [0.51, 0.92]	0.70 [0.51, 0.91]	0.72 [0.49, 1.00]	0.310
Systemic Immune-Inflammation Index, median [IQR]	2469.27 [1232.54, 3941.20]	2411.13 [1208.09, 3876.64]	3186.43 [1501.69, 4608.78]	<0.001
Systemic Inflammation Response Index, median [IQR]	6.95 [3.55, 12.86]	6.48 [3.42, 12.30]	9.17 [5.03, 17.38]	<0.001
Neutrophil-to-Lymphocyte Ratio, median [IQR]	13.80 [6.50, 18.15]	13.34 [6.27, 17.70]	15.98 [8.76, 22.04]	<0.001
Platelet-to-Lymphocyte Ratio, median [IQR]	273.78 [158.10, 383.44]	272.19 [156.56, 376.59]	292.66 [170.66, 431.00]	0.021
Monocyte-to-Lymphocyte Ratio, median [IQR]	0.84 [0.54, 1.16]	0.82 [0.52, 1.14]	1.00 [0.67, 1.42]	<0.001
Neutrophil-to-Platelet Ratio, median [IQR]	0.04 [0.03, 0.06]	0.04 [0.03, 0.06]	0.05 [0.03, 0.07]	<0.001
Geriatric Nutritional Risk Index, median [IQR]	46.01 [45.44, 46.61]	46.01 [45.57, 46.61]	46.01 [45.17, 46.46]	0.004

Prognostic Nutritional Index, median [IQR]	34.85 [31.02, 39.41]	35.06 [31.12, 39.58]	33.63 [29.52, 38.19]	<0.001
Neutrophil-to-Albumin Ratio, median [IQR]	2.90 [2.05, 4.00]	2.86 [2.02, 3.91]	3.31 [2.31, 4.89]	<0.001
Lymphocyte-to-Albumin Ratio, median [IQR]	0.61 [0.51, 0.80]	0.61 [0.50, 0.75]	0.70 [0.53, 1.16]	<0.001
Monocyte-to-Albumin Ratio, median [IQR]	0.23 [0.17, 0.32]	0.23 [0.17, 0.32]	0.24 [0.17, 0.36]	0.058
Platelet-to-Albumin Ratio, median [IQR]	67.27 [50.57, 86.67]	66.76 [50.57, 86.21]	70.00 [50.17, 89.81]	0.155
Red Cell Distribution Width-to-Albumin Ratio, median [IQR]	5.26 [4.75, 5.82]	5.24 [4.69, 5.71]	5.41 [5.08, 6.50]	<0.001
Severity score (normalized), median [IQR]	0.28 [0.20, 0.37]	0.27 [0.19, 0.35]	0.36 [0.26, 0.51]	<0.001
Comorbidities & Treatments				
Hypertension, n (%)	809 (37.9%)	703 (38.3%)	106 (35.9%)	0.482
Diabetes, n (%)	721 (33.8%)	622 (33.9%)	99 (33.6%)	0.972
Coronary artery disease, n (%)	617 (28.9%)	531 (28.9%)	86 (29.2%)	0.986
Chronic Kidney Disease, n (%)	902 (42.3%)	740 (40.3%)	162 (54.9%)	<0.001
Cancer, n (%)	153 (7.2%)	114 (6.2%)	39 (13.2%)	<0.001
Atrial fibrillation, n (%)	729 (34.2%)	596 (32.4%)	133 (45.1%)	<0.001
Sepsis, n (%)	450 (21.1%)	343 (18.7%)	107 (36.3%)	<0.001
Vasopressor use, n (%)	234 (11.0%)	164 (8.9%)	70 (23.7%)	<0.001

Data are presented as median [interquartile range] for continuous variables and n (%) for categorical variables. P values were calculated using the Mann-Whitney U test for continuous variables and the Chi-square test for categorical variables. COPD = Chronic Obstructive Pulmonary Disease; WBC = White Blood Cell Count; RDW = Red Cell Distribution Width; SII = Systemic Immune-Inflammation Index; SIRI = Systemic Inflammation Response Index; NLR = Neutrophil-to-Lymphocyte Ratio; PLR = Platelet-to-Lymphocyte Ratio; MLR = Monocyte-to-Lymphocyte Ratio; NPR = Neutrophil-to-Platelet Ratio; GNRI = Geriatric Nutritional Risk Index; PNI = Prognostic Nutritional Index; NAR = Neutrophil-to-Albumin Ratio; LAR = Lymphocyte-to-Albumin Ratio; MAR = Monocyte-to-Albumin Ratio; PAR = Platelet-to-Albumin Ratio; RAR = RDW-to-Albumin Ratio; CKD = Chronic Kidney Disease.

Table 2. Baseline Characteristics of Elderly Patients with COPD and Heart Failure (MIMIC-IV Cohort)

Variable	Total (n=2569)	Survived (n=2097)	Died (n=472)	P value
Demographics & Laboratory				
Male, n (%)	1338 (52.1%)	1090 (52.0%)	248 (52.5%)	0.865
Age (years), median [IQR]	77.00 [71.00, 83.00]	76.00 [71.00, 83.00]	79.00 [73.00, 85.00]	<0.001
Body Mass Index (kg/m²), median [IQR]	29.40 [26.36, 32.51]	29.48 [26.36, 32.55]	29.03 [26.36, 32.43]	0.310
White Blood Cell Count (10³/μL), median [IQR]	10.80 [7.80, 15.20]	10.50 [7.70, 14.60]	12.55 [8.70, 18.40]	<0.001
Platelets (10³/μL), median [IQR]	199.00 [150.00, 260.00]	199.00 [152.00, 260.00]	193.95 [140.00, 255.25]	0.155
Red Cell Distribution Width (%), median [IQR]	15.20 [14.10, 16.80]	15.10 [14.10, 16.60]	15.90 [14.60, 17.50]	<0.001
Albumin (g/dL), median [IQR]	2.90 [2.75, 3.30]	2.90 [2.75, 3.30]	3.00 [2.70, 3.30]	0.733
Lactate (mmol/L), median [IQR]	1.63 [1.30, 2.20]	1.60 [1.20, 2.10]	2.00 [1.50, 3.00]	<0.001
Neutrophils (10³/μL), median [IQR]	9.15 [6.73, 13.43]	8.83 [6.58, 12.90]	10.81 [7.51, 16.35]	<0.001
Lymphocytes (10³/μL), median [IQR]	0.62 [0.49, 1.01]	0.62 [0.49, 1.02]	0.63 [0.47, 0.96]	0.153
Monocytes (10³/μL), median [IQR]	0.68 [0.51, 0.87]	0.68 [0.52, 0.86]	0.66 [0.50, 0.87]	0.697
Systemic Immune-Inflammation Index, median [IQR]	3221.65 [1949.98, 4356.29]	3076.55 [1886.57, 4223.45]	3722.05 [2233.04, 4838.46]	<0.001
Systemic Inflammation Response Index, median [IQR]	7.85 [5.04, 14.23]	7.55 [4.94, 12.96]	11.00 [5.47, 18.52]	<0.001
Neutrophil-to-Lymphocyte Ratio, median [IQR]	16.25 [11.87, 19.42]	16.13 [11.24, 18.73]	17.01 [13.72, 22.50]	<0.001
Platelet-to-Lymphocyte Ratio, median [IQR]	313.48 [212.04, 410.63]	315.40 [212.09, 407.88]	306.53 [211.17, 413.85]	0.911
Monocyte-to-Lymphocyte Ratio, median [IQR]	0.89 [0.68, 1.15]	0.87 [0.66, 1.13]	0.99 [0.75, 1.34]	<0.001
Neutrophil-to-Platelet Ratio, median [IQR]	0.05 [0.03, 0.07]	0.05 [0.03, 0.07]	0.06 [0.04, 0.08]	<0.001
Geriatric Nutritional Risk Index, median [IQR]	46.01 [45.71, 46.61]	46.01 [45.71, 46.61]	46.16 [45.63, 46.61]	0.510

Prognostic Nutritional Index, median [IQR]	33.83 [30.85, 37.59]	33.79 [30.92, 37.53]	34.05 [30.09, 37.70]	0.465
Neutrophil-to-Albumin Ratio, median [IQR]	3.18 [2.21, 4.41]	3.09 [2.20, 4.21]	3.63 [2.47, 5.83]	<0.001
Lymphocyte-to-Albumin Ratio, median [IQR]	0.55 [0.43, 0.74]	0.53 [0.42, 0.69]	0.67 [0.50, 1.08]	<0.001
Monocyte-to-Albumin Ratio, median [IQR]	0.23 [0.17, 0.30]	0.23 [0.17, 0.30]	0.23 [0.17, 0.31]	0.477
Platelet-to-Albumin Ratio, median [IQR]	67.57 [49.95, 88.67]	67.63 [50.36, 88.26]	67.46 [46.95, 90.70]	0.559
Red Cell Distribution Width-to-Albumin Ratio, median [IQR]	5.23 [4.93, 5.50]	5.21 [4.91, 5.42]	5.31 [5.08, 5.86]	<0.001
Severity score (normalized), median [IQR]	0.14 [0.00, 0.29]	0.14 [0.00, 0.29]	0.29 [0.14, 0.43]	<0.001
Comorbidities & Treatments				
Hypertension, n (%)	2040 (79.4%)	1687 (80.4%)	353 (74.8%)	0.007
Diabetes, n (%)	1029 (40.1%)	846 (40.3%)	183 (38.8%)	0.563
Coronary artery disease, n (%)	1577 (61.4%)	1290 (61.5%)	287 (60.8%)	0.815
Chronic Kidney Disease, n (%)	1061 (41.3%)	836 (39.9%)	225 (47.7%)	0.002
Cancer, n (%)	263 (10.2%)	193 (9.2%)	70 (14.8%)	<0.001
Atrial fibrillation, n (%)	1479 (57.6%)	1190 (56.7%)	289 (61.2%)	0.084
Sepsis, n (%)	523 (20.4%)	312 (14.9%)	211 (44.7%)	<0.001
Vasopressor use, n (%)	827 (32.2%)	593 (28.3%)	234 (49.6%)	<0.001

Data are presented as median [interquartile range] for continuous variables and n (%) for categorical variables. P values were calculated using the Mann-Whitney U test for continuous variables and the Chi-square test for categorical variables. COPD = Chronic Obstructive Pulmonary Disease; WBC = White Blood Cell Count; RDW = Red Cell Distribution Width; SII = Systemic Immune-Inflammation Index; SIRI = Systemic Inflammation Response Index; NLR = Neutrophil-to-Lymphocyte Ratio; PLR = Platelet-to-Lymphocyte Ratio; MLR = Monocyte-to-Lymphocyte Ratio; NPR = Neutrophil-to-Platelet Ratio; GNRI = Geriatric Nutritional Risk Index; PNI = Prognostic Nutritional Index; NAR = Neutrophil-to-Albumin Ratio; LAR = Lymphocyte-to-Albumin Ratio; MAR = Monocyte-to-Albumin Ratio; PAR = Platelet-to-Albumin Ratio; RAR = RDW-to-Albumin Ratio; CKD = Chronic Kidney Disease.

Table 3. Performance Comparison of Machine Learning Models

Model	Cohort	AUC	Brier Score	Accuracy	F1 Score	Precision	Recall	Threshold
Random Forest	MIMIC-IV Test	0.769	0.129	0.724	0.477	0.366	0.683	0.186
Random Forest	eICU External	0.723	0.114	0.632	0.344	0.228	0.698	0.177
Bagged Classification and Regression Tree	MIMIC-IV Test	0.751	0.132	0.698	0.452	0.339	0.676	0.218
Bagged Classification and Regression Tree	eICU External	0.712	0.127	0.671	0.355	0.244	0.654	0.258
Naive Bayes	MIMIC-IV Test	0.728	0.176	0.685	0.446	0.330	0.690	0.010
Naive Bayes	eICU External	0.699	0.163	0.711	0.370	0.265	0.614	0.035
Neural Network	MIMIC-IV Test	0.608	0.148	0.642	0.337	0.255	0.493	0.201
Neural Network	eICU External	0.562	0.136	0.679	0.253	0.186	0.393	0.268
K-Nearest Neighbors	MIMIC-IV Test	0.721	0.138	0.728	0.429	0.350	0.556	0.208
K-Nearest Neighbors	eICU External	0.681	0.117	0.669	0.336	0.233	0.607	0.194
Classification and Regression Tree	MIMIC-IV Test	0.674	0.143	0.742	0.416	0.357	0.500	0.222
Classification and Regression Tree	eICU External	0.658	0.138	0.689	0.320	0.229	0.529	0.222
Gradient Boosting Machine	MIMIC-IV Test	0.764	0.131	0.685	0.461	0.337	0.732	0.147
Gradient Boosting Machine	eICU External	0.719	0.117	0.670	0.351	0.241	0.644	0.191
Boosted Generalized Linear Model	MIMIC-IV Test	0.768	0.132	0.733	0.488	0.377	0.690	0.175
Boosted Generalized Linear Model	eICU External	0.723	0.115	0.653	0.349	0.235	0.671	0.177

AUC, Area Under the Receiver Operating Characteristic Curve. MIMIC-IV Test refers to internal test set; eICU External refers to external validation cohort. Brier Score measures prediction calibration (lower is better). F1 Score is the harmonic mean of precision and recall.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.png
- TableS1.docx
- TableS2.docx
- TableS3.docx