

A Machine Learning Approach for Predicting 28-Day ICU Mortality in ARDS Patients Based on the New Global Definition

Shuo Yuan

the Affiliated Hospital of Xuzhou Medical University

Lanxin Lü

the Affiliated Hospital of Xuzhou Medical University

Heng Dong

Xuzhou Cancer Hospital

Weichao Ding

the Affiliated Hospital of Xuzhou Medical University

Shuzhan Zhang

the Affiliated Hospital of Xuzhou Medical University

Hui Peng

Feng Xian People's Hospital

Fei Wang

Xuzhou Cancer Hospital

Xianliang Yan

the Affiliated Hospital of Xuzhou Medical University

Ningjun Zhao

zhaonjdoc@163.com

the Affiliated Hospital of Xuzhou Medical University

Research Article

Keywords: ARDS, Machine learning, Support vector classifier (SVC), Interpretable model, Intensive care unit, Mortality prediction

Posted Date: July 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10349847/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Acute respiratory distress syndrome (ARDS) is a life-threatening form of acute diffuse lung injury that can be triggered by various factors, including pneumonia, sepsis, trauma, and other etiologies. The definition of ARDS has undergone multiple revisions to date, with the new global definition (2023) introducing significant updates to the clinical diagnosis and treatment of this syndrome. This study aimed to develop and validate an interpretable prognostic prediction model for ARDS patients via machine learning techniques based on a new global definition.

Methods

We extracted patient data from the Medical Information Mart for Intensive Care-IV database (MIMIC-IV, version 2.2) for model training. For external validation, we used clinical data from patients who met the new global definition of ARDS and were admitted to the Affiliated Hospital of Xuzhou Medical University. LASSO regression with cross-validation was employed to select key prognostic variables. We subsequently constructed predictive models for 28-day mortality in ARDS patients via eight machine learning algorithms: logistic regression, random forest, decision tree, support vector classifier (SVC), LightGBM, XGBoost, AdaBoost, and multilayer perceptron (MLP). Model performance was assessed via receiver operating characteristic (ROC) curves, decision curve analysis (DCA), and calibration curves. SHAP values were used to interpret the machine learning models.

Results

A total of 3382 ARDS patients were included in our analysis, from which 15 key variables were selected for model development. On the basis of the area under the ROC curve (AUC), DCA, and calibration curves, the SVC model demonstrated strong performance, achieving an average AUC of 0.819 (95% CI: 0.786–0.855) in the internal validation set and 0.871 in the external validation set.

Conclusion

Our machine learning-based prognostic prediction model for ARDS patients, developed in accordance with the new global definition, demonstrates reliable performance and can support clinical decision-making by helping clinicians formulate personalized treatment plans.

Introduction

Acute respiratory distress syndrome (ARDS) is a clinical syndrome primarily characterized by refractory hypoxemia and respiratory distress. Its fundamental pathophysiological alterations involve increased permeability of the alveolar–capillary membrane resulting from various direct or indirect insults, leading to alveolar and interstitial edema, intra-alveolar hemorrhage, and hyaline membrane formation, which collectively cause hypoxemia and respiratory compromise [1]. ARDS represents one of the most common critical illnesses in intensive care units (ICUs), complicating approximately 10% of all ICU admissions and 23% of mechanically ventilated patients [2]. Despite its high incidence, studies have reported an alarming in-hospital mortality rate of 35–45% among ICU patients with ARDS [2, 3].

The diagnostic criteria for ARDS are crucial for early identification and severity stratification. The clinical definition of ARDS has undergone multiple revisions, with the 2012 Berlin definition playing a pivotal role in guiding clinical diagnosis and therapeutic management. The Berlin definition establishes four key diagnostic parameters: mechanical ventilation, oxygenation index ($\text{PaO}_2/\text{FiO}_2$ ratio), positive end-expiratory pressure (PEEP), and pulmonary imaging findings [3]. However, during its decade-long clinical application, several limitations have become apparent. Most notably, the Berlin definition fails to adequately diagnose ARDS in critically ill patients who are not receiving mechanical ventilation or who lack arterial blood gas analysis results, particularly in resource-constrained settings [4].

To address these limitations, an international panel of 32 critical care experts jointly published the New Global Definition of ARDS in May 2023. This updated framework incorporates diagnostic criteria for patients receiving noninvasive ventilation or high-flow nasal oxygen (HFNO) therapy. Importantly, it introduces pulse oximetry-derived oxygenation ($\text{SpO}_2/\text{FiO}_2$ ratio) as a key diagnostic parameter, replacing the traditional arterial blood gas-dependent oxygenation index [5]. The new global definition thus significantly expands clinical applicability while providing important updates for ARDS diagnosis, severity assessment, and therapeutic management.

Machine learning (ML), a critical branch of artificial intelligence, empowers computer systems to perform decision-making and predictive tasks through advanced data analytics. In the medical field, ML techniques are being rapidly implemented for comprehensive database analyses, with applications spanning disease prediction, diagnostic support, personalized treatment strategies, pharmaceutical development, epidemiological investigations, and medical image interpretation [6]. Recent studies have increasingly employed ML algorithms to predict clinical outcomes in ARDS patients. For example, Maddali *et al.* developed ML models to identify and validate ARDS subphenotypes [7], and Tran *et al.* constructed a mortality prediction model for ARDS patients via gradient boosting algorithms [8]. Notably, these investigations focused primarily on patient cohorts diagnosed under the Berlin criteria. To date, research exploring the clinical characteristics and outcome prediction in ARDS patients defined by the new global definition remains notably limited.

The Medical Information Mart for Intensive Care-IV (MIMIC-IV) database is a comprehensive critical care database rich in anonymized clinical information that is designed to facilitate medical information sharing and advance critical care research. This database aggregates diverse data sources from hospital information systems, including demographic data, laboratory test results, medication records, clinical notes, vital signs, imaging reports, and physiological waveforms, covering numerous ICU patients over a period of more than ten years. As such, the MIMIC-IV database provides a solid foundation for developing and validating clinical prognostic prediction models for ARDS patients diagnosed according to the new global definition.

Therefore, this study aims to identify ARDS patients meeting the new global definition criteria from the MIMIC database, systematically collect their clinical characteristics, identify prognostic risk factors, and ultimately develop a clinical outcome prediction model to address this critical research gap.

Methods

Study design and data sources

For this investigation, we used the MIMIC-IV, version 2.2, as the primary data repository. The MIMIC-IV is a publicly accessible, comprehensive database that includes clinical data from over 190,000 patients treated at Beth Israel Deaconess Medical Center between 2008 and 2019. It documents more than 450,000 hospital admissions and approximately 70,000 ICU stays. The database integrates a wealth of information, including patient demographics, International Classification of Diseases, Ninth and Tenth Revision (ICD-9 and ICD-10) codes, vital signs, laboratory

reports, imaging data, real-time physiological monitoring data from the ICU, and clinical outcomes. Rigorous anonymization procedures and stringent privacy protection measures are in place to safeguard patient information. Access to and use of the database are governed by the Massachusetts Institute of Technology's Committee on the Use of Human Subjects, ensuring ethical and regulatory compliance. One author (Shuo Yuan) approved the extraction of data from the database for research purposes (Certification No. 66326379).

We extracted data from sepsis patients with ARDS who met the new global definition from the database. The new global definition updates the Berlin definition, with key diagnostic criteria including the following:

1. Acute onset or worsening of hypoxemic respiratory failure within one week of a known risk factor or the appearance of new or worsening respiratory symptoms.
2. Chest imaging revealed bilateral infiltrates or edema, excluding cardiogenic pulmonary edema.
3. Classification based on ventilation status:
4. Noninvasive ARDS: high-flow nasal oxygen (HFNO) with an oxygen flow rate of ≥ 30 L/min or noninvasive ventilation/continuous positive airway pressure (NIV/CPAP) with a PEEP of ≥ 5 cm H₂O.
5. Intubated ARDS: similar to the Berlin criteria.
6. ARDS in resource-limited settings: Diagnosis can be made under conventional oxygen therapy alone, without requiring specific respiratory support (e.g., PEEP or high-flow oxygen).
7. Under the above conditions, when $\text{SpO}_2 \leq 97\%$, a $\text{SpO}_2/\text{FiO}_2$ ratio ≤ 315 is required for ARDS diagnosis [5].

Furthermore, we employed machine learning methods to predict 28-day ICU mortality in ARDS patients diagnosed under the new global definition while also analyzing potential risk factors that may influence clinical outcomes.

Data extraction

Training cohort

We first used structured query language (SQL) in Navicat Premium software (version 16.3.8) to retrieve and extract raw data from the MIMIC-IV database, including patients' basic clinical information, laboratory test results, imaging examinations, clinical comorbidities, intensive care records, advanced life support treatments, and clinical outcome information.

Patients meeting the following criteria and their corresponding data were included in this study: (1) first hospital admission and first ICU stay; (2) ICU length of stay > 24 hours; (3) age at admission ≥ 18 years; and (4) diagnosis of ARDS within 24 hours of admission on the basis of the new global definition criteria (detailed diagnostic basis as described above; the data extraction process is shown in Fig. 1). For extraction of ARDS patients under the new global definition, we referenced the open-source code from Qian *et al.*[9], which included the following: (i) initial ventilation treatment status after ICU admission; (ii) pulmonary imaging (chest X-ray or CT) reports containing text information such as "bilateral infiltrates" or "edema" suggesting bilateral pulmonary edema; and (iii) oxygenation index ($\text{PaO}_2/\text{FiO}_2$ ratio) and $\text{SpO}_2/\text{FiO}_2$ ratio. We made reasonable improvements to some of the codes: for example, we defined the oxygenation index and $\text{SpO}_2/\text{FiO}_2$ under initial ventilation status as the worst values within 24 hours of ICU admission (for cases where the initial ventilation duration was < 24 h, the worst values during the ventilation period were used), and we used the ICD codes "428" and "I50" and their subcodes to identify and exclude acute cardiogenic pulmonary edema.

The specific data we extracted included the following: (1) basic clinical information, such as age at admission, sex, weight, and height; (2) intensive care records, such as ICU length of stay and vital signs within 24 hours of ICU admission (including blood pressure, heart rate, respiratory rate, body temperature, oxygen saturation, urine output, blood glucose, etc.); (3) laboratory test results within 24 hours of ICU admission, such as complete blood count, liver and kidney function tests, coagulation function, and arterial blood gas; (4) advanced life support treatments, such as renal replacement therapy, mechanical ventilation, and vasoactive drug use; (5) imaging examinations, such as text information from pulmonary imaging (chest X-rays and chest CT scans); and (6) patient death records in the database. Death within 28 days of intensive care unit (ICU) admission was considered a positive outcome. The vital signs and laboratory test results were included as independent features according to their maximum, minimum, and/or average values.

External validation cohort

We included patients with ARDS who met the new global definition criteria and were admitted to the Affiliated Hospital of Xuzhou Medical University from March 2022 to October 2024. The exclusion criteria were consistent with those used in the training cohort. Clinical data for patients in the external validation cohort were collected on the basis of the 15 features selected from the training cohort after model training. These features included age at admission, mean SpO₂, mean body temperature, mean respiratory rate, body weight, red blood cell distribution width (RDW), presence of metastatic solid tumors, lactate level, serum creatinine level, serum calcium level, chloride level, international normalized ratio (INR), mean corpuscular volume, anion gap, and logistic organ dysfunction score (LODS).

Statistical analysis

In the preliminary phase of data analysis, the normality of the data was assessed via the Shapiro–Wilk test. Continuous variables with a normal distribution are reported as the means (standard deviations) and were compared between groups via t tests. Variables with nonnormal distributions are presented as medians (interquartile ranges) and were analyzed via the Wilcoxon test. Categorical variables are expressed as frequencies (percentages) and were examined for intergroup differences with the chi-square test, with a significance level set at $p < 0.05$. Patients were classified into survival and mortality groups on the basis of their status 28 days after ICU admission and further divided into the Berlin definition and new global definition groups according to the ARDS diagnostic criteria.

The machine learning models were developed via Python 3.11.7 within a Jupyter Notebook environment. During the data preprocessing phase, as depicted in Fig. 2, the missing package was employed to visualize the missing data. Variables with more than 30% missing values (e.g., bicarbonate and albumin) were excluded. For the remaining missing values, we applied mouse forest multiple imputation, which effectively captures complex relationships among variables via a random forest model [10]. Continuous variables were normalized with MinMaxScaler, and categorical variables were encoded via OneHotEncoder.

In the training and validation phase of the machine learning model, we evaluated a spectrum of established and validated algorithms following LASSO-CV feature selection: logistic regression (LR), random forest (RF), decision tree (DT), support vector classification (SVC), light gradient boosting machine (LightGBM), eXtreme gradient boosting (XGBoost), adaptive boosting (AdaBoost), and multilayer perceptron (MLP).

For model development and validation, we employed a fivefold nested cross-validation approach, which was repeated five times, resulting in a total of 25 independent models. The inner loop was dedicated to hyperparameter

optimization, whereas the outer loop assessed model performance. In the outer loop, the dataset was partitioned into five segments, with four allocated for training and one allocated for testing, to prevent data leakage. The inner loop further divides the training subset into ten parts: nine for hyperparameter tuning and one for performance evaluation. The optimal model was validated on the outer test set, and this procedure was iterated five times to determine the best-performing model. To address dataset imbalance, the synthetic minority oversampling technique (SMOTE) and the Tomek link technique were applied, with hyperparameters refined through Bayesian optimization. Model performance was compared via the receiver operating characteristic (ROC) curve, area under the curve (AUC), accuracy, sensitivity, specificity, recall, and F1 score.

SHapley Additive exPlanations (SHAP) is an interpretive tool in machine learning that employs Shapley values to elucidate the impact of model features on predictions, thereby enhancing model transparency. In this investigation, we performed SHAP analysis on the MLP model with the highest AUC to facilitate interpretability.

Results

The study cohort comprised 3,382 ARDS patients fulfilling the criteria of the new global definition. Patients were assigned to survivor or nonsurvivor groups on the basis of their vital status 28 days after ICU admission.

Baseline characteristics

Table 1 presents an overview of the baseline characteristics of the patient population reclassified under the new global definition, including critical medical history, physiological parameters, laboratory findings, complications, and advanced life support interventions. The overall mortality rate of the cohort was 23.48%. Univariate analysis revealed significant differences between survivors and nonsurvivors with respect to age, weight, urine output, mean pulse oxygen saturation, mean arterial blood pressure, body temperature, pH, arterial oxygen partial pressure, lactate levels, oxygenation index ($\text{PaO}_2/\text{FiO}_2$), and the $\text{SpO}_2/\text{FiO}_2$ ratio ($p < 0.001$).

Table 1
Baseline characteristics of the total population

Variable	Survival (n = 2588)	Nonsurvival (n = 794)	p value ²
Age, Median (IQR)	64.06 (52.28–74.39)	67.78 (57.53–79.34)	< 0.001
Gender Male, n (%)	1,538 (59)	442 (56)	0.060
Heart rate_mean, Median (IQR)	88.53 (78.20–100.48)	93.36 (79.84–104.42)	< 0.001
Respiratory_mean, Median (IQR)	19.96 (17.37–22.86)	22.01 (19.00–25.70)	< 0.001
spo2_mean, Median (IQR)	96.68 (95.19–97.99)	96.34 (94.42–98.05)	< 0.001
temperature_mean, Median (IQR)	36.98 (36.68–37.36)	36.80 (36.42–37.20)	< 0.001
mbp_mean, Median (IQR)	76.48 (71.08–83.71)	73.92 (68.52–81.56)	< 0.001
hematocrit_min, Median (IQR)	30.40 (25.90–35.50)	29.50 (24.20–35.20)	0.006
hemoglobin_min, Median (IQR)	10.10 (8.60–11.80)	9.60 (8.00–11.50)	< 0.001
wbc_max, Median (IQR)	13.90 (10.30–19.10)	15.60 (10.60–21.60)	< 0.001
platelets_min, Median (IQR)	172.50 (119.00–233.00)	150.00 (78.00–226.00)	< 0.001
RDW_max, Median (IQR)	14.50 (13.60–15.90)	15.60 (14.20–17.60)	< 0.001
MCH_min, Median (IQR)	30.00 (28.60–31.40)	30.10 (28.70–31.70)	0.006
MCHC_min, Median (IQR)	32.60 (31.60–33.70)	32.20 (30.90–33.30)	< 0.001
MCV_min, Median (IQR)	90.00 (86.00–94.00)	91.00 (87.00–97.00)	< 0.001
pH_min, Median (IQR)	7.31 (7.23–7.37)	7.27 (7.16–7.37)	< 0.001
po2_min, Median (IQR)	70.00 (49.00–91.00)	55.00 (40.00–76.00)	< 0.001
pco2_max, Median (IQR)	48.00 (42.00–56.00)	48.00 (40.00–58.00)	0.83
spo2fio2ratio_min, Median (IQR)	158.33 (96.00–194.00)	118.75 (92.00–188.00)	< 0.001
pao2fio2ratio_min, Median (IQR)	135.00 (88.00–204.56)	102.25 (68.00–166.00)	< 0.001
baseexcess_min, Median (IQR)	-2.00 (-6.00–0.00)	-4.00 (-10.00–0.00)	< 0.001
aniongap_max, Median (IQR)	16.00 (13.00–19.00)	18.00 (15.00–23.00)	< 0.001
lactate_max, Median (IQR)	2.00 (1.40–3.20)	2.90 (1.60–5.50)	< 0.001
INR_max, Median (IQR)	1.30 (1.20–1.50)	1.50 (1.20–2.10)	< 0.001
PT_max, Median (IQR)	14.40 (12.80–16.90)	16.10 (13.50–22.20)	< 0.001
APTT_max, Median (IQR)	32.10 (28.10–40.40)	38.15 (29.90–57.90)	< 0.001
bun_max, Median (IQR)	20.00 (14.00–31.00)	30.00 (20.00–47.00)	< 0.001
creatinine_max, Median (IQR)	1.00 (0.80–1.60)	1.30 (0.90–2.30)	< 0.001
glucose_mean, Median (IQR)	133.50 (114.52–160.49)	140.88 (115.83–180.11)	< 0.001
calcium_min, Median (IQR)	7.90 (7.30–8.40)	7.80 (7.10–8.40)	0.030

Variable	Survival (n = 2588)	Nonsurvival (n = 794)	p value ²
calcium_max, Median (IQR)	8.40 (7.90–8.90)	8.50 (8.00–9.00)	< 0.001
sodium_min, Median (IQR)	137.00 (135.00–140.00)	137.00 (133.00–140.00)	0.021
sodium_max, Median (IQR)	140.00 (138.00–143.00)	140.00 (137.00–144.00)	0.21
potassium_min, Median (IQR)	3.90 (3.50–4.20)	3.80 (3.40–4.30)	0.011
potassium_max, Median (IQR)	4.50 (4.10–4.90)	4.60 (4.20–5.30)	< 0.001
chloride_max, Median (IQR)	107.00 (103.00–110.00)	106.00 (101.00–112.00)	0.13
chloride_min, Median (IQR)	103.00 (99.00–106.00)	101.00 (97.00–106.00)	< 0.001
urineoutput, Median (IQR)	1 586.00 (1,000.00–2,340.00)	1 097.00 (550.00–1,815.00)	< 0.001
sapsii, Median (IQR)	37.00 (29.00–47.00)	50.00 (40.00–61.00)	< 0.001
sofa, Median (IQR)	6.00 (4.00–10.00)	10.00 (7.00–14.00)	< 0.001
lods, Median (IQR)	5.00 (3.00–8.00)	9.00 (6.00–12.00)	< 0.001
myocardial_infarct, n (%)	272 (11)	116 (15)	0.002
peripheral_vascular_disease, n (%)	250 (9.7)	80 (10)	0.73
cerebrovascular_disease, n (%)	319 (12)	136 (17)	< 0.001
chronic_pulmonTary_disease, n (%)	743 (29)	221 (28)	0.63
rheumatic_disease, n (%)	95 (3.7)	25 (3.1)	0.49
diabetes_with_cc, n (%)	156 (6.0)	35 (4.4)	0.084
diabetes_without_cc, n (%)	579 (22)	148 (19)	0.025
renTal_disease, n (%)	352 (14)	113 (14)	0.65
malignTant_cancer, n (%)	402 (16)	220 (28)	< 0.001
metastatic_solid_tumor, n (%)	154 (6.0)	148 (19)	< 0.001
mild_liver_disease, n (%)	374 (14)	210 (26)	< 0.001
severe_liver_disease, n (%)	179 (6.9)	119 (15)	< 0.001
vasoactive_drugs_use, n (%)	1 108 (43)	459 (58)	< 0.001
rrt_use, n (%)	83 (3.2)	57 (7.2)	< 0.001
aki_2day, n (%)	1 880 (73)	656 (83)	< 0.001
vent_status, n (%)			< 0.001
Oxygen	1 358 (52)	309 (39)	
Highflow	2 (< 0.1)	2 (0.3)	
Noninvasive	6 (0.2)	2 (0.3)	
Invasive	1 222 (47)	481 (61)	

Feature selection

In the feature selection phase, we employed LASSO regression with cross-validation (LASSO-CV) to screen the variables. LASSO regression mitigates collinearity by examining feature coefficients and applying regularization. LASSO-CV automatically determines the optimal lambda value to strike a balance between model complexity and fit, as depicted in Fig. 3. In this investigation, we incorporated 64 preprocessed features from the new global definition group patient data into the LASSO-CV algorithm via a 10-fold cross-validation approach. The final optimal lambda value was determined to be 0.0003252151161302468, leading to the selection of 45 significant features. To streamline the model and avoid overfitting, we selected the top 15 features with the highest absolute coefficient values for model construction. These features included the presence of metastatic solid tumors, logistic organ dysfunction score (LODS), mean respiratory rate, age, anion gap, body temperature, mean corpuscular volume (MCV), red cell distribution width (RDW), creatinine, serum calcium, body weight, serum chloride, mean SpO₂, international normalized ratio (INR), and lactate dehydrogenase.

(A) Determination of the optimal lambda value; (B) The variation of variable coefficients with the lambda value, the black dashed line indicates the coefficients of each variable at the optimal lambda value; (C) The ranking of variable coefficients.

Model performance comparison

The 15 selected features were used to construct machine learning models, resulting in 25 independent models corresponding to 8 distinct algorithms. Hyperparameter optimization was performed via five iterations of fivefold nested cross-validation to maximize generalizability. For comprehensive performance evaluation, we calculated the mean and 95% confidence intervals for the AUC, F1 score, recall, precision, accuracy, sensitivity, and specificity across the 8 models, with detailed results provided in Table 2.

Table 2
Prediction performance of the eight types of machine learning algorithms

Model	AUC	F1-score	Recall	Accuracy	Sensitivity	Specificity
LightGBBoost	0.793(0.756–0.826)	0.513(0.462–0.561)	0.478(0.428–0.536)	0.786(0.759–0.814)	0.478(0.428–0.536)	0.881(0.854–0.914)
Logistic Regression	0.819(0.786–0.855)	0.583(0.553–0.625)	0.748(0.692–0.811)	0.749(0.731–782)	0.748(0.692–0.811)	0.750(0.721–0.790)
Decision Tree	0.741(0.711–0.776)	0.509(0.479–0.542)	0.678(0.605–0.804)	0.693(0.652–0.74)	0.678(0.678–0.804)	0.698(0.601–0.767)
AdaBoost	0.795(0.766–0.831)	0.553(0.526–0.596)	0.685(0.63–0.751)	0.740(0.712–0.767)	0.685(0.63–0.751)	0.757(0.719–0.796)
Random Forest	0.804(0.774–0.838)	0.553(0.511–0.598)	0.643(0.562–0.709)	0.757(0.733–0.78)	0.643(0.562–0.709)	0.792(0.761–0.828)
XGBoost	0.810(0.779–0.843)	0.556(0.506–0.607)	0.595(0.518–0.669)	0.777(0.745–0.807)	0.595(0.518–0.669)	0.685(0.78–0.869)
MLP	0.817(0.782–0.851)	0.575(0.532–0.614)	0.751(0.686–0.818)	0.739(0.709–0.763)	0.751(0.686–0.818)	0.735(0.689–0.773)
SVC	0.819(0.789–0.859)	0.568(0.534–0.0.614)	0.756(0.694–0.826)	0.730(0.708–0.756)	0.756(0.694–0.826)	0.723(0.686–0.749)

To further analyze model performance, we employed receiver operating characteristic (ROC) curves, decision curve analysis (DCA), and calibration curves (Fig. 4). The ROC curve was used to evaluate classification efficacy, while the calibration curve was used to assess prediction accuracy.

As shown in Table 2, the logistic regression and SVC models achieved identical AUC values of 0.819 (logistic regression: 95% CI 0.786–0.855; SVC: 95% CI 0.789–0.859). Furthermore, the two algorithms exhibited comparable performance across metrics, including accuracy, sensitivity, specificity, F1 score, and recall. Moreover, as illustrated in Fig. 4, the two models exhibited similar performance in terms of classification ability, predictive accuracy, and clinical decision support. Accordingly, we validated the performance of both models in an external validation cohort.

External validation

We performed external validation on 100 patients with acute respiratory distress syndrome (ARDS) who met the new global definition and were admitted to the Affiliated Hospital of Xuzhou Medical University between March 2022 and October 2024. Notably, there was no overlap between the external validation cohort and the training cohort, which enabled a more reliable assessment of the model's generalizability and predictive performance in real-world clinical settings.

The performance of the logistic regression and support vector classification (SVC) models in the external validation cohort is summarized in Table 3 and Fig. 5. Specifically, the SVC model achieved a higher AUC value, a lower Brier score, and a greater net benefit (area under the curve) at clinically relevant threshold probabilities in the decision curve analysis (DCA) than did the logistic regression model. The overall performance of the SVC model was superior to that of the logistic regression model.

Table 3 Prediction performance of the SVC model and logistic regression model in the external validation cohort						
Model	AUC	F1-score	Recall	Accuracy	Sensitivity	Specificity
Logistic Regression	0.864(0.858–0.87)	0.75(0.728–0.764)	0.797(0.767–0.81)	0.708(0.689–0.723)	0.776(0.76–0.79)	0.797(0.767–0.81)
SVC	0.871(0.863–0.881)	0.747(0.722–0.764)	0.792(0.762–0.81)	0.706(0.682–0.723)	0.706(0.682–0.723)	0.792(0.762–0.81)

Interpretability analysis

Given the superior predictive performance of the SVC model in both the training cohort and the external validation cohort, we performed SHAP value calculations and visualizations to elucidate the impact of individual variables on the model's predictions. First, we computed the mean SHAP values for each feature in the SVC model and ranked them by importance (Fig. 6A), which reflects the average influence of each feature on the model's overall output [11]. In the beeswarm plot (Fig. 6B), we present the distribution of features via horizontally aligned data points. The x-axis represents the SHAP values of the features, while the color indicates the magnitude of the feature values (red: higher values; blue: lower values). Each data point represents a sample's feature value and its corresponding SHAP value, and its horizontal position reflects the strength of its impact on the prediction; the density of points reveals the pattern of data distribution. Moreover, the relationship between the color (i.e., feature value magnitude) and the SHAP values exposes the directional effect of the features. For example, the age feature has a more positive predictive effect on positive outcomes with higher feature values, whereas the urine output feature has the opposite effect. Through this analysis, we identified features such as LODS, mean respiratory rate, age, anion gap, MCV, RDW, INR, and lactate

dehydrogenase as positively correlated with the risk of 28-day mortality. Conversely, features such as the mean body temperature, creatinine, and body weight were negatively correlated with the risk of 28-day mortality.

To enhance the local interpretability of the model, we examined the influence of various features on prognostic outcomes and the complex linear and nonlinear relationships among them. We generated scatter plots of SHAP values against feature values for 14 of the 15 quantitative features (excluding metastatic solid tumors). Locally weighted scatterplot smoothing (LOWESS) fitting curves were employed to visually represent the data distribution trends through locally weighted regression. Figure 7 shows the fitting curve with the red line, and the intersection points where the SHAP value equals zero are marked with blue dashed lines, indicating the corresponding feature values. For example, the relationship between age and the risk of 28-day ICU mortality is nonlinear. An increase in age shifts the contribution of this feature from negative to positive, with the intersection point at 71.28 years, indicating that patients older than 71.28 years at admission are at a higher risk of 28-day mortality. Similarly, a respiratory rate greater than 21 breaths/min, a LODS greater than 7.93, and other specific threshold values were identified as risk factors for increased 28-day mortality.

The influence of key features on prognosis was examined through SHAP values and corresponding visualizations. The ROC and calibration curves suggested that the multilayer perceptron (MLP) model exhibited the highest prediction accuracy. However, to evaluate its clinical applicability, decision curve analysis (DCA) was also conducted, revealing the superior performance of the SVC model in terms of clinical utility. Ultimately, the SVC model-based predictive model distinguished itself with robust performance and significant clinical application value. Additionally, SHAP values were used to elucidate the machine learning model. The feature importance plot delineates the relative impact of each feature, whereas the swarm plot reveals the distribution of features and their directional influence on prediction outcomes. This approach not only enhances the understanding of the model's decision-making process but also highlights the heterogeneity of ARDS, facilitating the advancement of personalized and precision treatment strategies and accelerating the clinical translation and application of our findings.

Discussion

In this study, we explored factors influencing 28-day mortality in patients with ARDS who met the new global definition. The feature importance plot revealed that the top five features affecting patient prognosis were the logistic organ dysfunction score (LODS), the presence of metastatic solid tumors, the mean respiratory rate, age, and the maximum anion gap.

The LODS is a scoring system used to assess the severity of organ dysfunction in ICU patients. It evaluates the occurrence and severity of organ dysfunction on the basis of heart rate, mean arterial pressure, urine output, and Glasgow Coma Scale (GCS) score. The LOD reflects the overall severity of a patient's condition in the ICU. Although we also included the SAPS II and SOFA scores in our study, the LODS score proved to be more meaningful in predicting the prognosis of ARDS patients. This aligns with the findings of Kallet [12] and Leclerc [13] in their studies on ARDS patients diagnosed under the Berlin definition.

Compared with our previous study on sepsis-associated ARDS under the new global definition [14], which focused on a single etiology, the present study extends the scope to all ARDS patients regardless of cause, thereby providing a more generalized prognostic model for the broader ARDS population. This broader inclusion may have contributed to the improved discriminative performance observed in the current model, which achieved higher AUC values (0.819 vs. 0.792 in internal validation; 0.871 vs. 0.816 in external validation), suggesting that a more heterogeneous ARDS population may offer greater discriminative power for mortality prediction.

Additionally, cancer patients may be at increased risk for 28-day mortality in ARDS patients. Studies have shown that individuals with cancer-induced immunosuppression have a relatively high incidence and mortality rate of ARDS [15]. This may be related to immune dysfunction, coagulation abnormalities, and a greater likelihood of opportunistic infections caused by the tumor itself and/or tumor-related treatments [16, 17]. However, whether the significant influence of metastatic tumors in our model is due to more severe immunosuppression, aggressive antitumor therapy, or poorer nutritional status in these patients requires further investigation.

Basic vital signs, such as body temperature, respiratory rate, and oxygen saturation, were also reflected in the predictive model. Hypothermia was identified as a risk factor for mortality in our study. A low body temperature within the first 24 hours of ICU admission may be associated with 28-day mortality, whereas higher temperatures are correlated with reduced mortality, which is consistent with findings from Schell-Chaple HM and Kushimoto S, among others [18, 19]. An elevated respiratory rate may reflect the severity of respiratory distress, and ventilator-induced lung injury caused by high respiratory rates could contribute to 28-day mortality in ARDS patients [20]. Furthermore, in our model, oxygen saturation (SpO_2) demonstrated greater predictive value for 28-day mortality outcomes in ARDS patients, with lower SpO_2 levels indicating a greater likelihood of ICU death. SpO_2 serves as a reliable indicator of patients' oxygenation status and is highly consistent with arterial oxygen saturation (SaO_2) and the partial pressure of arterial oxygen (PaO_2). Although current ARDS diagnosis and severity assessment rely heavily on the oxygenation index ($\text{PaO}_2/\text{FiO}_2$), SpO_2 offers distinct advantages because of its continuous, cost-effective, and noninvasive measurement capability [21]. In contrast, arterial blood gas analysis and its derived oxygenation index can capture only a patient's oxygenation status at a single time point. Moreover, the percentage of oxyhemoglobin better reflects the pathophysiological status of ARDS patients, such as intrapulmonary shunting [21, 22]. Our findings align with those of numerous other studies demonstrating that SpO_2 is more strongly correlated with ARDS patient outcomes than the oxygenation index ($\text{PaO}_2/\text{FiO}_2$).

A greater red cell distribution width (RDW) is associated with poor prognosis in ARDS patients, possibly because increased oxidative stress leads to greater erythrocyte fragility and destruction, as well as hypoxia-induced erythropoietin production and the release of immature red blood cells [23]. Furthermore, the effects of other indicators in our model, such as chloride ion concentrations, calcium ion concentrations, the anion gap, and creatinine, on the prognosis of ARDS patients remain understudied. These biochemical markers may influence the pathophysiological processes and clinical outcomes of ARDS through various mechanisms, but current research evidence remains relatively limited. Specifically, chloride ions, as key electrolytes for maintaining osmotic pressure and acid–base balance, may exacerbate pulmonary edema by impairing alveolar–capillary barrier function when their levels become abnormal [24]. Calcium homeostasis not only plays a role in coagulation and inflammatory regulation but also may exacerbate lung injury by impairing alveolar epithelial cell function. However, studies focusing on calcium metabolism in ARDS patients are scarce [25]. An elevated anion gap typically indicates lactate accumulation or renal dysfunction, serving as a potential marker of tissue hypoperfusion; however, its specific prognostic value in ARDS remains unclear [26]. Serum creatinine, a marker of renal function, is typically elevated in patients with multiorgan dysfunction. Higher creatinine levels often necessitate the implementation of renal replacement therapy, which may explain why initially elevated creatinine values contribute to a lower risk of mortality. However, the unique mechanisms underlying kidney–lung cross talk in ARDS require further investigation. These parameters are routinely measured in clinical practice. If their association with ARDS prognosis can be confirmed, they may offer a cost-effective and efficient dimension for risk stratification. Future large-scale clinical studies combined with basic research are needed to systematically elucidate the mechanistic roles of these electrolytes and metabolic markers in ARDS progression, thereby providing a theoretical foundation for refining prognostic prediction models.

Nevertheless, this study has several limitations. First, the patient data used for model training were derived primarily from Western populations, which differ substantially from the demographic characteristics of our external validation cohort. Second, our predictive models were constructed using only routinely available clinical variables, without performing a more comprehensive analysis of the database, potentially leading to the omission of certain critical details. Third, as an observational and retrospective study, our findings may be subject to inherent biases or inaccuracies. Fourth, the small size ($n = 100$) of our external validation cohort may limit the generalizability of our conclusions. Despite these limitations, our model continued to demonstrate robust predictive performance in the external validation cohort.

Conclusion

In conclusion, this study validates the utility of machine learning as a reliable tool for predicting the clinical course of ARDS. The clinical prediction model optimized according to the latest ARDS diagnostic criteria provides in-depth insights into the potential factors influencing the clinical outcomes of ARDS patients. This approach holds promise for significantly improving clinical practice by assisting clinicians in developing precise, personalized treatment strategies, thereby maximizing survival rates among ARDS patients.

Abbreviations

ARDS

Acute respiratory distress syndrome

ML

Machine Learning

ICU

Intensive care unit

LODS

Logistic Organ Dysfunction Score

SVC

Support Vector Classifier

SHAP

SHapley Additive exPlanations

AUC

Area under the curve

ROC

Receiver operating characteristic

DCA

Decision Curve Analysis

LASSO

Least absolute shrinkage and selection operator

Declarations

Ethics approval and consent to participate: This study was conducted in accordance with the Declaration of Helsinki. The institutional review boards of the Affiliated Hospital of Xuzhou Medical University granted a waiver of informed consent and approved the sharing of the research resources. The external validation study was approved by the Institutional Review Board of the Affiliated Hospital of Xuzhou Medical University (approval number: [XYFY2025-KL103-01]).

Consent for publication: Not applicable. The individual data presented in this article were anonymized and not identifiable.

Availability of data and materials: The data used for this study are publicly available from the MIMIC-IV database.

Competing interests: The authors declare that they have no competing interests.

Funding: This project was funded by the Construction Project of the High-Level Hospital of Jiangsu Province (GSPJS202419; GSPJS202519; GSPJS202522; GSPJS202530); the Xuzhou National Clinical Key Specialty Cultivation Project (2018ZK004); the excellent young and middle-aged talent project of the affiliated hospital of Xuzhou Medical University (2019128009); the Xuzhou Medical University Affiliated Hospital "Pairing Assistance" Scientific Research Project (FXJDBF2024215); the Concept Validation Project of Xuzhou Medical University (GNYZ2024010); and the Innovation and Entrepreneurship Education Practice Center Project of Xuzhou Medical University Science Park (2024).

Author contributions: All the authors contributed substantially to the conception and design of the study, the acquisition of data, or the analysis and interpretation of the data. They also drafted (SY and LL) or provided (NZ, XY, FW) critical revision of the article. All authors provided final approval of the version submitted for publication.

Acknowledgments: We would like to thank the owners of the MIMIC database for allowing the scientific community access to these data.

Supplementary Information: The online version contains supplementary material available at

References

1. Matthay MA, Zemans RL, Zimmerman GA, Arabi YM, Beitler JR, Mercat A, Herridge M, Randolph AG, Calfee CS. Acute respiratory distress syndrome. *Nat Rev Dis Primers*. 2019;5(1):18.
2. Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, Gattinoni L, van Haren F, Larsson A, McAuley DF, et al. Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. *JAMA*. 2016;315(8):788–800.
3. Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, Camporota L, Slutsky AS. Acute respiratory distress syndrome: the Berlin Definition. *JAMA*. 2012;307(23):2526–33.
4. Gershengorn HB, Hu Y, Chen JT, Hsieh SJ, Dong J, Gong MN, Chan CW. The Impact of High-Flow Nasal Cannula Use on Patient Mortality and the Availability of Mechanical Ventilators in COVID-19. *Ann Am Thorac Soc*. 2021;18(4):623–31.
5. Matthay MA, Arabi Y, Arroliga AC, Bernard G, Bersten AD, Brochard LJ, Calfee CS, Combes A, Daniel BM, Ferguson ND, et al. A New Global Definition of Acute Respiratory Distress Syndrome. *Am J Respir Crit Care Med*. 2024;209(1):37–47.
6. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol*. 2022;23(1):40–55.
7. Maddali MV, Churpek M, Pham T, Rezoagli E, Zhuo H, Zhao W, He J, Delucchi KL, Wang C, Wickersham N, et al. Validation and utility of ARDS subphenotypes identified by machine-learning models using clinical data: an observational, multicohort, retrospective analysis. *Lancet Respir Med*. 2022;10(4):367–77.
8. Tran TK, Tran MC, Joseph A, Phan PA, Grau V, Farmery AD. A systematic review of machine learning models for management, prediction and classification of ARDS. *Respir Res*. 2024;25(1):232.

9. Qian F, van den Boom W, See KC. The new global definition of acute respiratory distress syndrome: insights from the MIMIC-IV database. *Intensive Care Med.* 2024;50(4):608–9.
10. Chen Y, Zong C, Zou L, Zhang Z, Yang T, Zong J, Wan X. A novel clinical prediction model for in-hospital mortality in sepsis patients complicated by ARDS: A MIMIC IV database and external validation study. *Heliyon.* 2024;10(13):e33337.
11. Zuo D, Yang L, Jin Y, Qi H, Liu Y, Ren L. Machine learning-based models for the prediction of breast cancer recurrence risk. *BMC Med Inf Decis Mak.* 2023;23(1):276.
12. Kallet RH, Lipnick MS, Zhuo H, Pangilinan LP, Gomez A. Characteristics of Nonpulmonary Organ Dysfunction at Onset of ARDS Based on the Berlin Definition. *Respir Care.* 2019;64(5):493–501.
13. Leclerc F, Duhamel A, Deken V, Le Reun C, Lacroix J, Leteurtre S. Nonrespiratory pediatric logistic organ dysfunction-2 score is a good predictor of mortality in children with acute respiratory failure. *Pediatr Crit Care Med.* 2014;15(7):590–3.
14. Zhang P, Yuan S, Zhang S, Yuan Z, Ye Z, Lv L, Yang H, Peng H, Li H, Zhao N. Under the background of the new global definition of ARDS: an interpretable machine learning approach for predicting 28-day ICU mortality in patients with sepsis complicated by ARDS. *Front Physiol.* 2025;16:1617196.
15. Soubani AO, Shehada E, Chen W, Smith D. The outcome of cancer patients with acute respiratory distress syndrome. *J Crit Care* 2014, 29(1):183.e187-183.e112.
16. Azoulay E, Lemiale V, Mokart D, Pène F, Kouatchet A, Perez P, Vincent F, Mayaux J, Benoit D, Bruneel F, et al. Acute respiratory distress syndrome in patients with malignancies. *Intensive Care Med.* 2014;40(8):1106–14.
17. Inokuchi R, Manabe H, Ohta F, Nakamura K, Nakajima S, Yahagi N. Granulocyte colony-stimulating factor-producing lung cancer and acute respiratory distress syndrome. *Clin Respir J.* 2015;9(2):250–2.
18. Schuliga M, Read J, Knight DA. Aging mechanisms that contribute to tissue remodeling in lung disease. *Aging Res Rev.* 2021;70:101405.
19. Kushimoto S, Abe T, Ogura H, Shiraishi A, Saitoh D, Fujishima S, Mayumi T, Hifumi T, Shiino Y, Nakada TA, et al. Impact of Body Temperature Abnormalities on the Implementation of Sepsis Bundles and Outcomes in Patients With Severe Sepsis: A Retrospective Sub-Analysis of the Focused Outcome Research on Emergency Care for Acute Respiratory Distress Syndrome, Sepsis and Trauma Study. *Crit Care Med.* 2019;47(5):691–9.
20. Gattinoni L, Quintel M. How ARDS should be treated. *Crit Care.* 2016;20:86.
21. Wick KD, Matthay MA, Ware LB. Pulse oximetry for the diagnosis and management of acute respiratory distress syndrome. *Lancet Respir Med.* 2022;10(11):1086–98.
22. Patel S, Singh G, Zarbiv S, Ghiassi K, Rachoin JS. Mortality Prediction Using SaO₂/FiO₂ Ratio Based on eICU Database Analysis. *Crit Care Res Pract.* 2021;2021:6672603.
23. Dai J, Guo Y, Zhou Q, Duan XJ, Shen J, Zhang X. The relationship between red cell distribution width, serum calcium ratio, and in-hospital mortality among patients with acute respiratory failure: A retrospective cohort study of the MIMIC-IV database. *Med (Baltim).* 2024;103(15):e37804.
24. Shukla A, Hashiguchi N, Chen Y, Coimbra R, Hoyt DB, Junger WG. Osmotic regulation of cell function and possible clinical applications. *Shock.* 2004;21(5):391–400.
25. Islam MN, Gusarova GA, Das SR, Li L, Monma E, Anjaneyulu M, Owusu-Ansah E, Bhattacharya S, Bhattacharya J. The Mitochondrial Calcium Uniporter of Pulmonary Type 2 Cells Determines Severity of ARDS. *bioRxiv* 2021.
26. Fennes AZ, Emmett M. Approach to Patients With High Anion Gap Metabolic Acidosis: Core Curriculum 2021. *Am J Kidney Dis.* 2021;78(4):590–600.

Figures

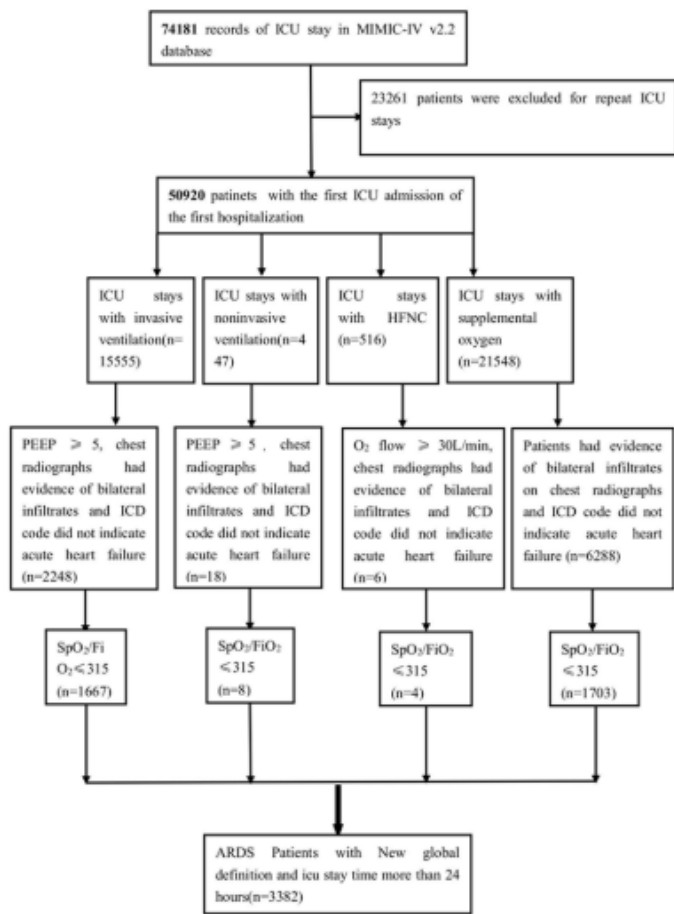


Figure 1

Flowchart of screening

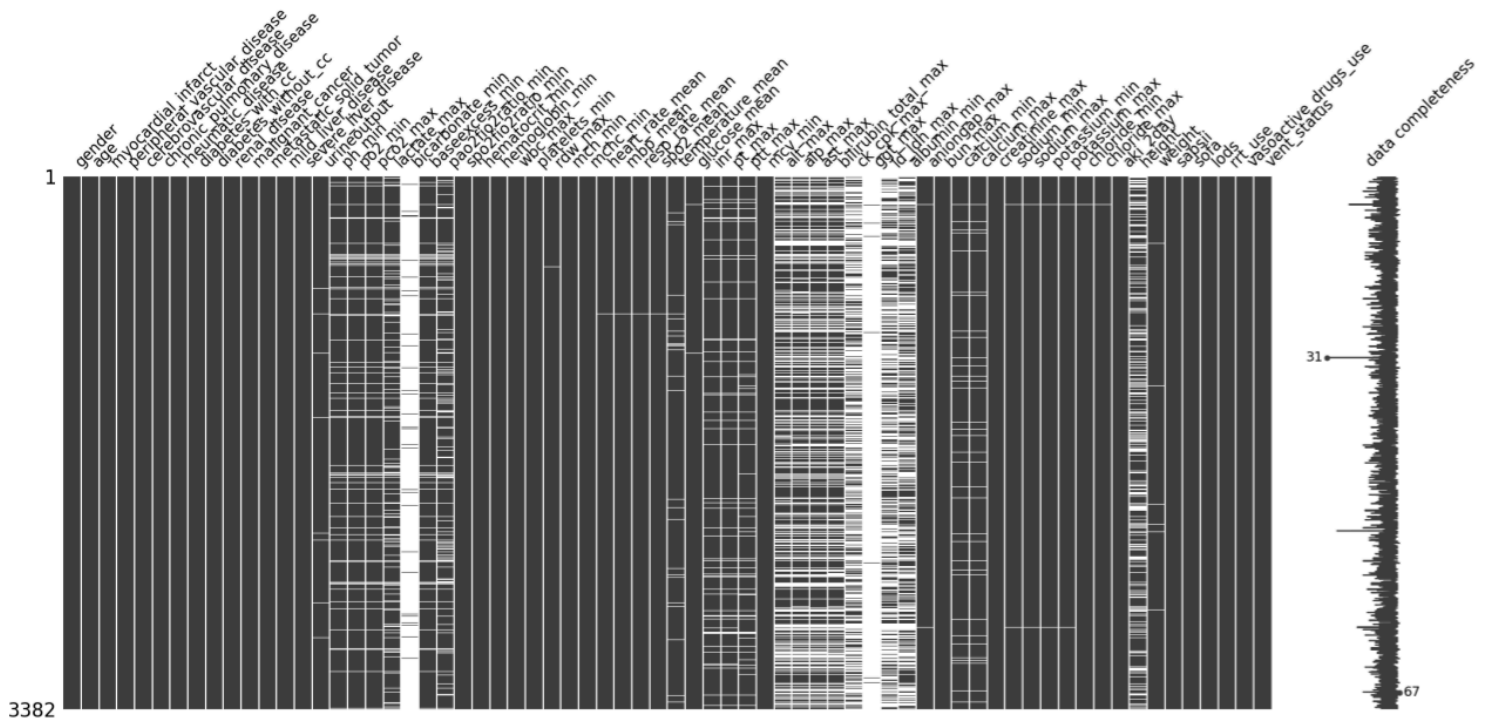


Figure 2

Missing value visualization: Each column represents a clinical variable, and the white lines represent missing values.

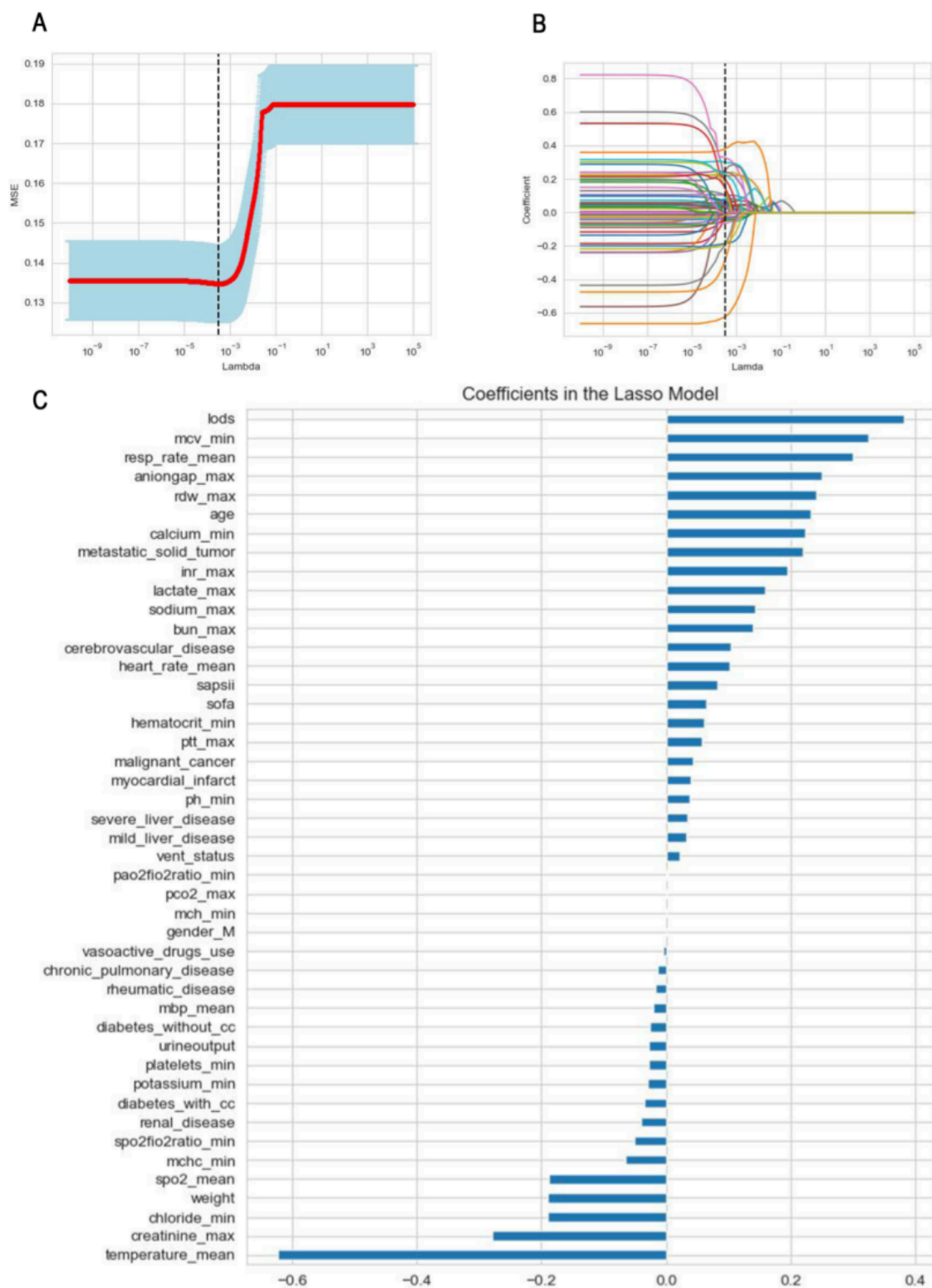


Figure 3

Feature selection using LASSO regression with cross-validation.

(A) Determination of the optimal lambda value; (B) The variation of variable coefficients with the lambda value, the black dashed line indicates the coefficients of each variable at the optimal lambda value; (C) The ranking of variable coefficients.

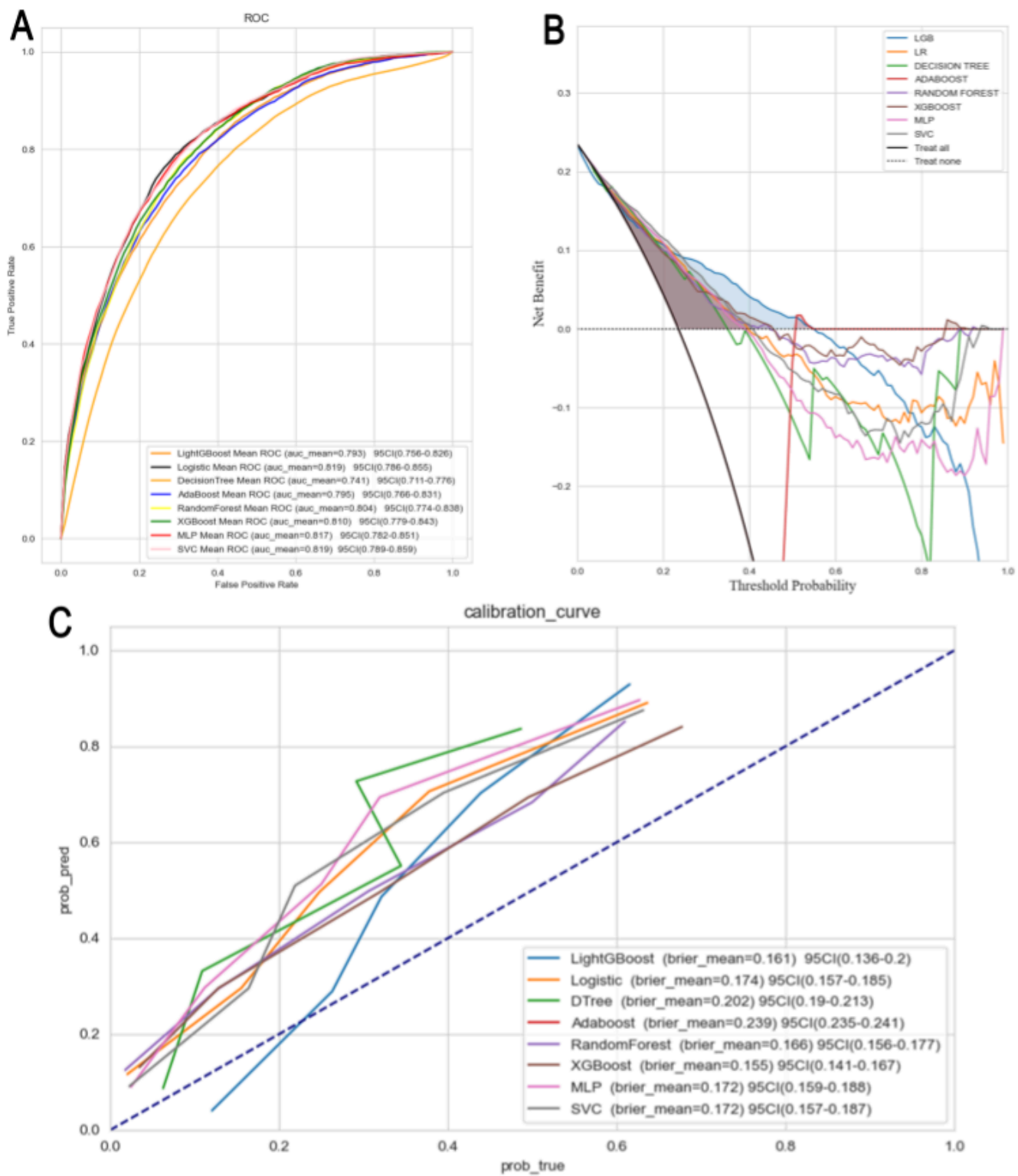


Figure 4

ROC curve DCA curve and Calibration curve comparison of eight models

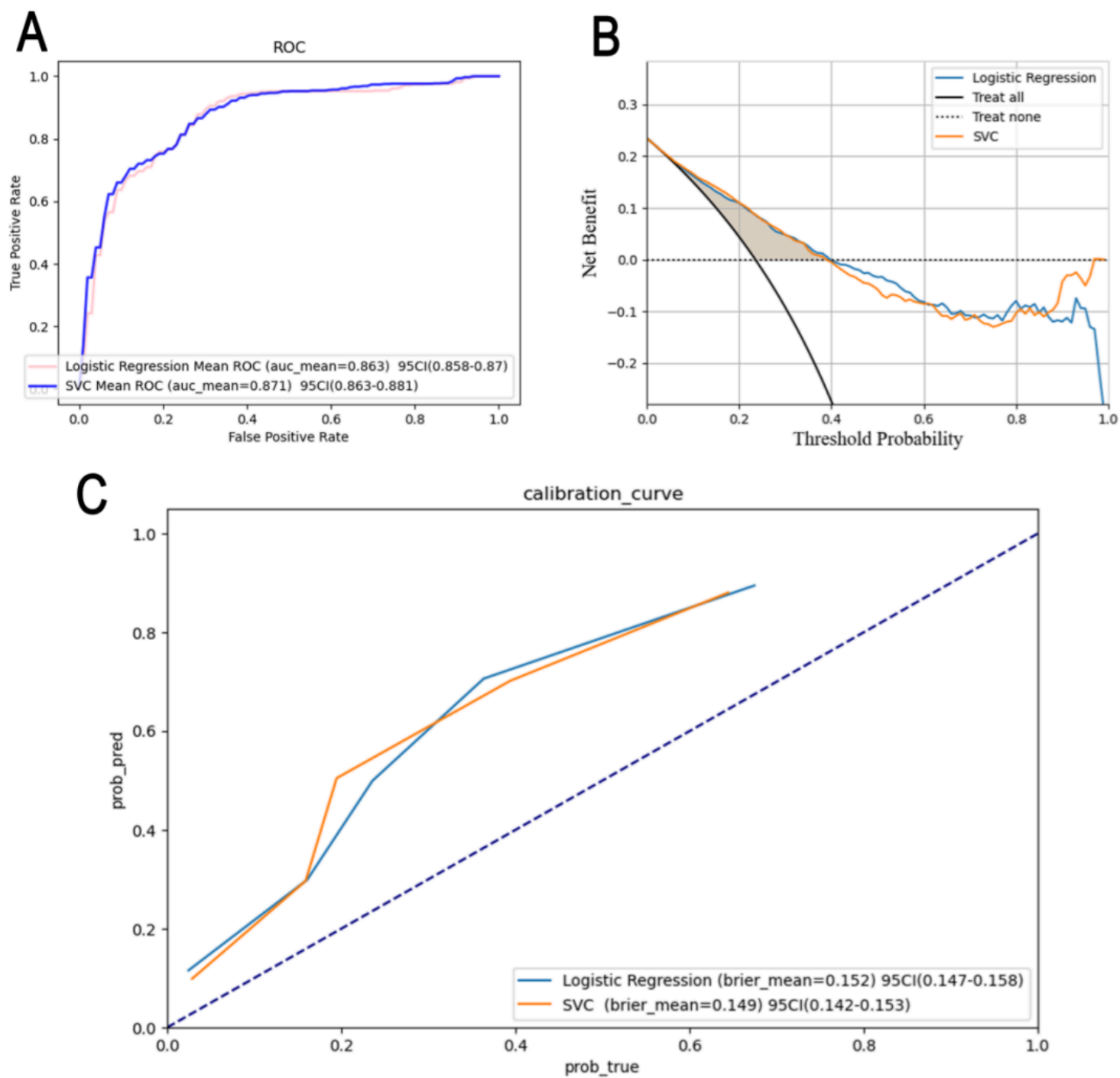


Figure 5

ROC curve, DCA curve and Calibration curve comparison of SVC model and Logistic Regression model

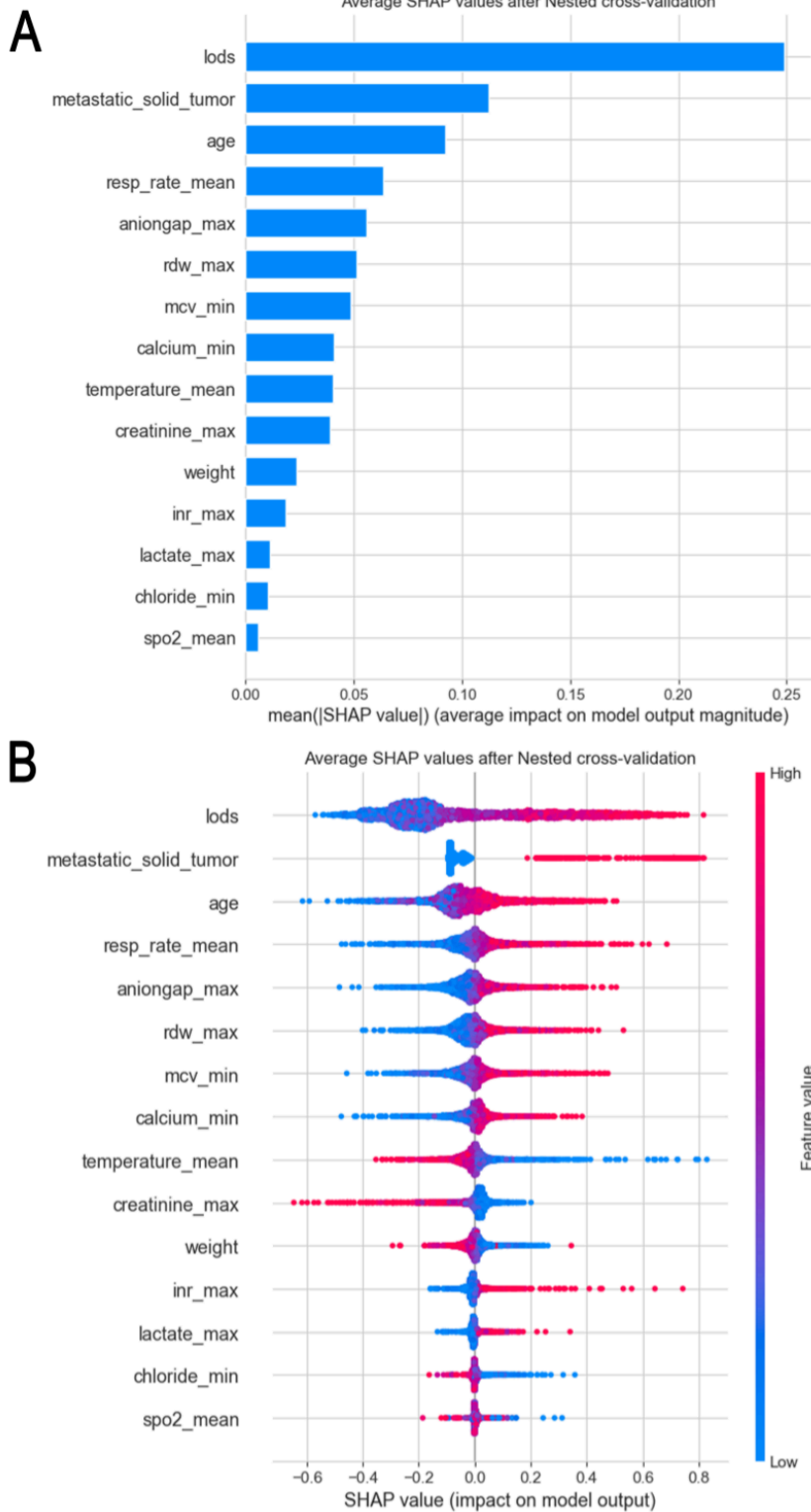


Figure 6

From a global perspective, we calculated the average SHAP value for each feature and used a swarm plot to display the distribution of features and SHAP values. (A) Plot of Features Importance; (B) Swarm Plot.

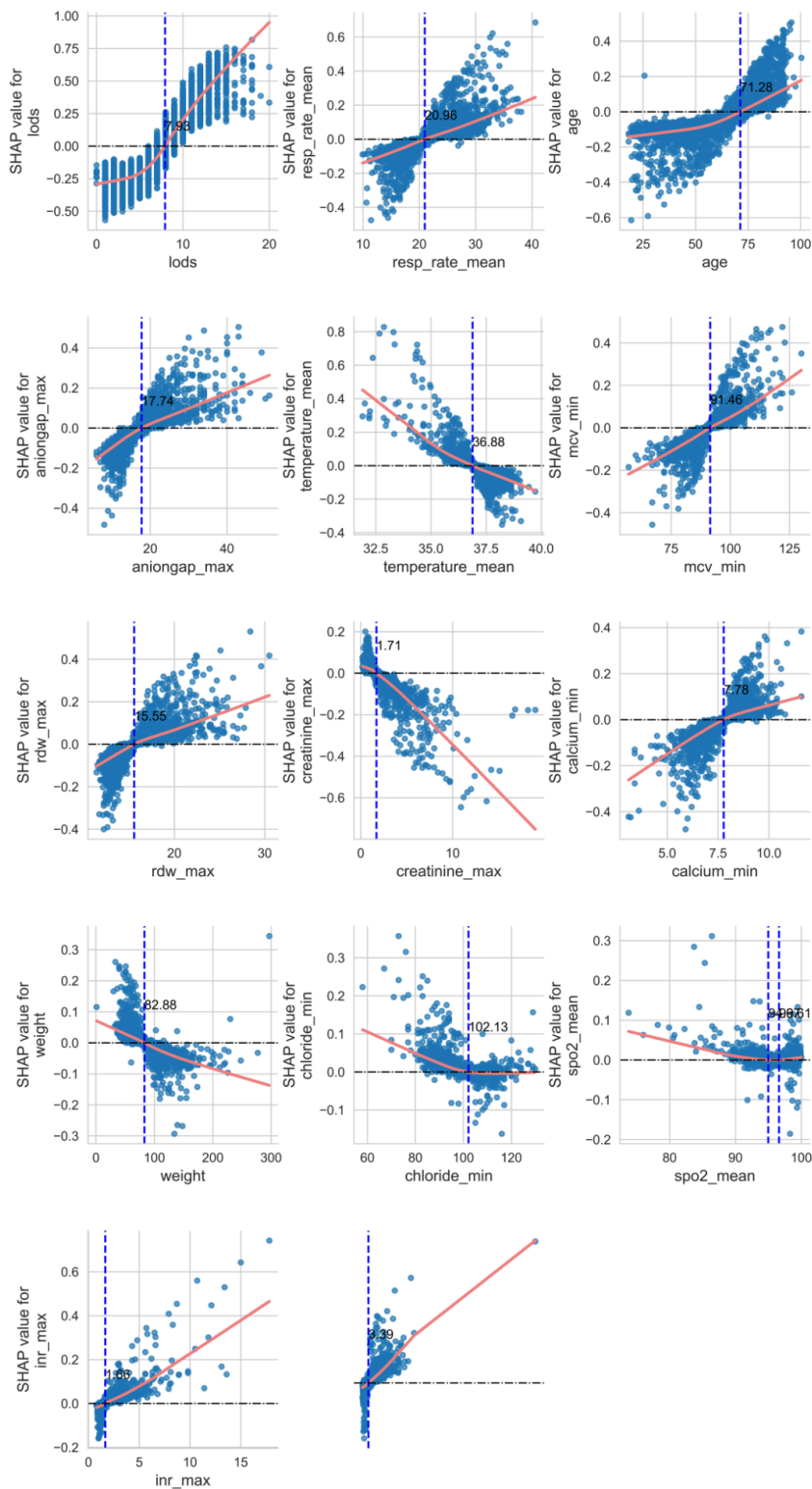


Figure 7

Linear or nonlinear effects of quantitative type data on prognostic outcomes.

Supplementary Files

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

- [ARDSnewglobalmimic.xlsx](#)
- [XuzhouMedicalUniversityAffiliatedHospitalexternalvalidation.xlsx](#)