

Supplementary Table S1. Baseline characteristics of the training set after SMOTENC resampling

Variable	Levels	N	Overall	Sepsis	ICU-AW	P-value
			N = 3,799	N = 2,533	N = 1,266	
Age		3,799	64.91 (14.84)	66.50 (15.27)	61.74 (13.38)	< 0.001
Glucose		3,799	8.93 (5.44)	8.78 (5.40)	9.23 (5.50)	0.015
Sofa		3,799	8.36 (3.79)	8.53 (3.88)	8.04 (3.60)	< 0.001
Charlson		3,799	5.54 (2.96)	5.94 (3.06)	4.74 (2.56)	< 0.001
Antibiotics		3,799				< 0.001
CRRT	0	3,799	3,108.00 (81.81%)	2,432.00 (96.01%)	676.00 (53.40%)	0.178
			691.00 (18.19%)	101.00 (3.99%)	590.00 (46.60%)	
	1	3,799	3,371.00 (88.73%)	2,260.00 (89.22%)	1,111.00 (87.76%)	
			428.00 (11.27%)	273.00 (10.78%)	155.00 (12.24%)	
	0	3,799	1,746.00 (45.96%)	1,010.00 (39.87%)	736.00 (58.14%)	
			2,053.00 (54.04%)	1,523.00 (60.13%)	530.00 (41.86%)	
VENTILATION						< 0.001

Note. Continuous variables are presented as mean $\pm$ standard deviation (SD), and categorical variables are expressed as frequencies and percentages [n (%)]. P-values were calculated using the Student's t-test or Wilcoxon rank-sum test for continuous variables, and the chi-square (χ^2) test for categorical variables. SMOTENC was applied to address class imbalance in the training dataset. Abbreviations: ICU-AW = intensive care unit-acquired weakness; CRRT = continuous renal replacement therapy; SOFA = Sequential Organ Failure Assessment score; Charlson = Charlson Comorbidity Index; VENTILATION = mechanical ventilation.

Supplementary Table S2. Quantitative performance indicators of different models on the training dataset

Model	Accuracy	Sensitivity	Specificity	Precision	F1-Score
Logistic	0.792	0.505	0.936	0.799	0.619
Bayes	0.789	0.461	0.954	0.834	0.593
Xgboost	0.951	0.907	0.974	0.944	0.926
Lightgbm	0.952	0.909	0.974	0.945	0.927
Random forest	0.813	0.533	0.953	0.851	0.656
Catboost	0.954	0.899	0.981	0.960	0.929
SVM	0.884	0.792	0.930	0.900	0.820

Note. Model performance was evaluated using accuracy, sensitivity, specificity, precision, and F1-score in the training dataset. Sensitivity represents the proportion of true positive cases correctly identified, while specificity represents the proportion of true negative cases correctly identified. Precision indicates the proportion of predicted positive cases that are true positives. The F1-score is the harmonic mean of precision and sensitivity, reflecting the balance between false positives and false negatives. Abbreviations: SVM = support vector machine.

Supplementary Table S3. Quantitative performance metrics of models in the external validation set

Model	Accuracy	Sensitivity	Specificity	Precision	F1-Score
Logistic	0.807	0.967	0.644	0.734	0.835
Bayes	0.681	0.683	0.678	0.683	0.683
Xgboost	0.555	0.233	0.881	0.667	0.346
Lightgbm	0.521	0.200	0.847	0.571	0.296
Random forest	0.832	1.000	0.661	0.750	0.857
Catboost	0.580	0.267	0.898	0.727	0.390
SVM	0.437	0.333	0.542	0.426	0.374

Note. Model performance in the external validation set was evaluated using accuracy, sensitivity, specificity, precision, and F1-score. The external cohort was independently collected and used to assess model generalizability across different clinical settings. Variations in model performance may reflect differences in patient characteristics, outcome definitions, and data distribution between cohorts. Abbreviations: SVM = support vector machine.