

Prediction at Hospital Admission of In-Hospital Mortality and High-Cost Hospitalization Using Machine Learning in Patients With Severe Community-Acquired Pneumonia: An Exploratory Dual-Risk Stratification Study

Shiting Chen

China Pharmaceutical University

Jiaxin Ren

China Pharmaceutical University

Hui Jiang

Nanjing First Hospital, Nanjing Medical University

Zhou Zhou

Nanjing First Hospital, Nanjing Medical University

Ziye Huang

China Pharmaceutical University

Weixuan Shi

Hebei University

Zhengyang Wu

China Pharmaceutical University

Jianjun Zou

zoujianjun100@126.com

Nanjing First Hospital, Nanjing Medical University

Hao Liu

China Pharmaceutical University

Meng Wan

Nanjing First Hospital, Nanjing Medical University

Research Article

Keywords: Severe community-acquired pneumonia, Machine learning, In-hospital mortality, High-cost hospitalization, Risk stratification, Healthcare resource management.

Posted Date: September 17th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10632723/v1>

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

Additional Declarations: No competing interests reported.

1 **Prediction at Hospital Admission of In-Hospital Mortality and**
2 **High-Cost Hospitalization Using Machine Learning in Patients With**
3 **Severe Community-Acquired Pneumonia: An Exploratory Dual-Risk**
4 **Stratification Study**

5

6 Shiting Chen^{a,b,1}, Jiaxin Ren^{a,b,1}, Hui Jiang^{c,1}, Zhou Zhou^{b,1}, Ziye Huang^{a,b}, Weixuan
7 Shi^d, Zhengyang Wu^{a,b}, Jianjun Zou^{b,*}, Hao Liu^{e,*}, Meng Wan^{b,f,*}

8

9 ^a School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University,
10 Nanjing 211198, China

11 ^b Department of Pharmacy, Nanjing First Hospital, Nanjing Medical University,
12 Nanjing 210006, China

13 ^c Department of Neurosurgery, Nanjing First Hospital, Nanjing Medical University,
14 Nanjing 210006, China

15 ^d School of Nursing, Hebei University, Baoding 071002, China

16 ^e State Key Laboratory of Natural Medicines, Key Laboratory of Drug Metabolism,
17 China Pharmaceutical University, Nanjing 210009, China

18 ^f School of International Pharmaceutical Business, China Pharmaceutical University,
19 Nanjing, 211198, China

20 ^{*}Corresponding authors

21 ¹ These authors have contributed equally to this work as first authors.

22

23 **Corresponding authors:**

24 Jianjun Zou, Tel: +86 15380998951, E-mail: zoujianjun100@126.com

25 Hao Liu, Tel: +86 13905622859, E-mail: lenhartkoo@foxmail.com

26 Meng Wan, Tel: +86 13584017093, E-mail: wanmeng1985@sina.cn

27

Abstract

Background: Patients with severe community-acquired pneumonia (sCAP) have high in-hospital mortality and healthcare resource use. Existing risk assessment mainly focuses on clinical outcomes, whereas early prediction of high-cost hospitalization remains limited. We aimed to develop and validate machine learning models at hospital admission for predicting in-hospital mortality and high-cost hospitalization in patients with sCAP and to explore the joint distribution of the two predicted risks.

Methods: This retrospective cohort study included adults with sCAP from January 2020 to January 2025. Predictors were limited to information available at admission. Common predictors were selected using dual-outcome Elastic Net with 500 bootstrap resamples. Feature selection, hyperparameter optimization, and model selection were performed in the training cohort, whereas the temporal validation cohort was reserved for final evaluation. L2-regularized logistic regression, random forest, XGBoost, LightGBM, and CatBoost models were evaluated, optimal models were selected by internal cross-validation using AUROC as the prespecified primary criterion. Performance was assessed using discrimination, calibration, and decision curve analysis. The optimal models were compared with CURB-65 and an age-plus-comorbidity logistic regression baseline, respectively. SHAP and exploratory four-quadrant analyses examined feature contributions and joint risk patterns.

Results: Among 1,543 patients, 16 admission predictors were retained. In temporal validation, the LightGBM mortality model achieved an AUROC of 0.734, and the

CatBoost high-cost model achieved an AUROC of 0.700. Both outperformed their respective comparators. Exploratory four-quadrant analysis revealed differences in clinical characteristics and resource use across predicted risk combinations. Group II had an in-hospital mortality rate of 84.2%, whereas Group IV had a high-cost hospitalization rate of 37.3%.

Conclusions: Using the same information available at hospital admission, we developed and temporally validated separate models for in-hospital mortality and high-cost hospitalization in patients with sCAP. Exploratory analysis of the joint predicted risks showed differences in clinical characteristics and resource use across risk combinations.

Keywords: Severe community-acquired pneumonia; Machine learning; In-hospital mortality; High-cost hospitalization; Risk stratification; Healthcare resource management.

1. Introduction

Community-acquired pneumonia (CAP) is a major cause of hospitalization and infection-related mortality in adults. Severe community-acquired pneumonia (sCAP) progresses rapidly and is difficult to manage, and patients admitted to intensive care units (ICU) remain at high risk of death [1, 2]. Beyond mortality, CAP imposes a substantial burden on healthcare resources. Previous studies estimated its annual treatment costs at approximately US\$8.4 billion, while the cost of a single CAP hospitalization may increase to US\$11,148–51,219 with advancing age and higher

71 risk levels [3]. Therefore, early management of sCAP should assess both in-hospital
72 mortality and high-cost hospitalization to support the prioritization of critically ill
73 patients and the appropriate allocation of healthcare resources.

74 Currently, risk stratification of CAP mainly relies on traditional scoring tools
75 including PSI and CURB-65 to evaluate in-hospital mortality and disease severity
76 [4-6]. In recent years, machine learning methods have also been applied to predict
77 in-hospital mortality, ICU admission, and length of stay in patients with CAP,
78 showing promising predictive value [7-9]. By contrast, early prediction of high-cost
79 hospitalization remains relatively underexplored [10, 11]. Previous studies have
80 shown that CAP hospitalization costs are associated with initial disease severity,
81 comorbidities, complications, ICU treatment, and length of stay [12-14]. However,
82 some of these factors become available only during hospitalization. These factors are
83 therefore better suited to explaining costs already incurred than to supporting
84 management at admission. Accordingly, in the early management of sCAP, it is
85 necessary to develop separate predictive models for in-hospital mortality and
86 high-cost hospitalization based on hospital admission data, then stratify patients for
87 management according to different combinations of these two risk categories.

88 Based on this, we used variables available at hospital admission to construct
89 separate prediction models for in-hospital mortality and high-cost hospitalization
90 among patients with sCAP. We further compared feature contributions between the
91 two models and performed an exploratory four-quadrant analysis based on the joint
92 distribution of the two predicted risks. This analysis aimed to characterize differences

in hospital course and resource use across risk combinations and to support future evaluation of risk-informed clinical and resource planning.

2. Methods

2.1. Study population

We conducted a single-center retrospective cohort study at Nanjing First Hospital. The study screened 9,958 adult patients hospitalized with community-acquired pneumonia (CAP) between January 2020 and January 2025.

The inclusion criteria were: (1) age ≥ 18 years; and (2) a first hospitalization for sCAP diagnosed according to the American Thoracic Society and Infectious Diseases Society of America guideline [6]. The exclusion criteria were: (1) hospital-acquired pneumonia; and (2) evident recording errors. After screening, 1,543 unique patients with index admissions for sCAP were included. Patients with index admissions between January 1, 2020, and December 31, 2023, formed the training cohort, whereas those with index admissions between January 1, 2024, and January 1, 2025, formed the temporal validation cohort.

The study complied with the Declaration of Helsinki. The Ethics Committee of Nanjing First Hospital approved the study (approval No. KY20260108-KS-09) and waived the requirement for informed consent because we used de-identified retrospective clinical data. We reported the study in accordance with the TRIPOD+AI guideline. Figure 1 presents the study flowchart.

2.2. Data collection

We obtained data from the electronic medical record system of Nanjing First

Hospital. The prediction time was hospital admission. Candidate predictors were restricted to information documented and available at this time, including demographic characteristics, pre-existing comorbidities, admitting department, initial vital signs, and initial laboratory results available during the admission assessment. Measurements first obtained after admission were not eligible as predictors. The primary outcomes included in-hospital mortality and high-cost hospitalization. In-hospital mortality was defined as all-cause death occurring after the prediction time during the index hospitalization. High-cost hospitalization was defined based on the 75th percentile of total hospitalization costs in the training cohort [15, 16], corresponding to a total hospitalization cost $\geq$ CNY 38,969. The main models included only predictors available at hospital admission. Variables related to subsequent treatment, ICU admission, mechanical ventilation, resource utilization, hospitalization costs, and outcomes were excluded from model development and were used only in secondary explanatory analyses.

2.3. Data processing

We excluded patients with missing primary outcome data or missing total hospitalization costs from the corresponding outcome analysis. Missing continuous predictors were imputed using median values estimated from the corresponding training data, whereas missing categorical values were coded as “Unknown.” For L2-regularized logistic regression, random forest, XGBoost, and LightGBM, a signed log1p transformation was applied to C-reactive protein, procalcitonin, alanine aminotransferase, aspartate aminotransferase, serum creatinine, blood urea nitrogen,

D-dimer, and lactate. Categorical variables were one-hot encoded for these models. CatBoost retained categorical variables and handled them natively, while missing continuous values were replaced with training-data medians. Continuous variables were standardized using Z-scores when required by the model. Within nested cross-validation, all preprocessing procedures were fitted separately using each training fold and applied to the corresponding validation fold. For temporal validation, preprocessing was fitted using the full training cohort and applied unchanged to the temporal validation cohort.

2.4. Feature selection

To ensure comparability between the two outcome models, we used the same set of candidate predictors available at hospital admission. We excluded all variables unavailable at the prediction time, including post-admission laboratory measurements, treatment-related variables, ICU admission, mechanical ventilation, resource-use variables, cost variables, and outcome-related variables. Microbiological results were eligible only when they were available at hospital admission. We also excluded the CURB-65 score from the main machine learning models because it structurally overlapped with its component variables.

Feature selection was fully nested within the outer five-fold cross-validation. In each outer training fold, preprocessing and variable screening were repeated independently. Variables with more than 50% missing data, constant values, or a single value accounting for at least 99.5% of observations were removed. Highly correlated variables were excluded using an absolute Spearman correlation threshold

of $|r| \geq 0.90$. The common feature set was determined using dual-outcome Elastic Net bootstrap stability selection. Separate Elastic Net logistic regression models were fitted for in-hospital mortality and high-cost hospitalization, with 500 stratified bootstrap resamples for each outcome. A categorical variable was considered selected when at least one of its one-hot-encoded dummy variables entered the model. Variables with a selection frequency of ≥ 0.60 for either outcome were prioritized, and a joint stability score was used to determine the common feature set. The fold-specific feature sets were used for inner-loop model tuning and out-of-fold prediction, and feature retention frequencies across the five outer folds were recorded. After internal validation, the entire selection procedure was repeated in the full training cohort to determine the final feature set for temporal validation.

2.5. Model development and performance evaluation

We developed separate models for in-hospital mortality and high-cost hospitalization using the same set of predictors available at hospital admission. We compared five prespecified algorithms: L2-regularized logistic regression, random forest, XGBoost, LightGBM, and CatBoost. For each algorithm, we calculated the Spearman correlation between the predicted probabilities of the two outcomes, with bootstrap 95% confidence intervals (CIs). We also assessed agreement in predicted probabilities for the same outcome across algorithms as a measure of algorithmic robustness.

We used stratified nested cross-validation for model development. The outer five-fold cross-validation estimated internal generalizability and generated out-of-fold

(OOF) predicted probabilities. The inner cross-validation optimized hyperparameters using Optuna. The AUROC was prespecified as the primary criterion for model selection, and for each outcome, the algorithm with the highest mean AUROC across the outer cross-validation folds was selected as the final model. To address class imbalance, we included class-weight parameters for L2-regularized logistic regression, random forest, and LightGBM. We used `scale_pos_weight` for XGBoost and included `auto_class_weights` among the candidate settings for CatBoost.

We assessed discrimination using the area under the receiver operating characteristic curve (AUROC) and the area under the precision–recall curve (AUPRC). Bootstrap resampling provided 95% CIs for the primary performance metrics. We evaluated calibration using the Brier score, calibration curves, calibration intercept, and calibration slope. Decision curve analysis (DCA) assessed clinical net benefit. Threshold-dependent metrics included accuracy, recall, specificity, precision, and the F1 score. We selected classification thresholds from the OOF predicted probabilities in the training cohort and applied them unchanged to the temporal validation cohort. The CURB-65 score served as the conventional comparator for in-hospital mortality prediction. For high-cost hospitalization, we developed a simple logistic regression baseline model based on age and the number of comorbidities. Paired DeLong tests were used to compare the AUROCs of the optimal machine learning models with those of their respective reference models in the same patients. Differences in AUROC and their 95% confidence intervals were also calculated. Comparisons were performed using the out-of-fold predictions in the training cohort

and the predictions in the temporal validation cohort.

2.6. Feature contribution and predictor inconsistency analysis

To compare feature contributions between the two outcome models, we grouped early admission variables into seven clinical domains: demographics and admission characteristics, chronic comorbidities, acute physiological dysfunction, inflammatory and infection markers, hepatic, renal, and nutritional function, hematologic and coagulation profiles, and microbiological testing status. These domains were used only for SHAP analysis.

Predictor discordance was assessed in four aspects. Outcome-level analyses included Spearman correlation, the Mann–Whitney U test, the chi-square test, and logistic or linear regression adjusted for age, CURB-65, and major comorbidities. Model-output discordance was evaluated using the Spearman correlation between predicted probabilities of in-hospital mortality and high-cost hospitalization. Feature-level analyses compared SHAP rank correlations, top-K overlap, and SHAP contributions across seven clinical domains, with bootstrap 95% CIs for rank correlations. For exploratory joint-risk grouping, we used the 75th percentiles of the out-of-fold predicted probabilities in the training cohort as thresholds. Patients were classified into four quadrants according to high-cost hospitalization and in-hospital mortality, respectively: Group I, high–high; Group II, low–high; Group III, low–low; and Group IV, high–low. Clinical characteristics, costs, and outcomes were compared across groups.

2.7. Exploratory analysis

We conducted an exploratory analysis of Groups II and IV from the four-quadrant stratification. Variables included the hospital course, number of medications, and cost structure. Effect sizes were calculated as Group IV minus Group II, using Cliff's delta for continuous variables and differences in proportions for categorical variables. Positive values indicated higher levels in the cost-dominant discordant group, whereas negative values indicated higher levels in the mortality-dominant discordant group. Key features were defined by an adjusted q value <0.10 and an absolute effect size ≥ 0.147 for continuous variables or an absolute proportion difference ≥ 0.10 for categorical variables. For binary variables, only the positive category was included in ranking and visualization.

2.8. Sensitivity analysis

Sensitivity analyses included redefining high-cost hospitalization using the 80th and 90th percentiles of total cost in the training cohort; excluding patients in the highest 1% of total costs; using median imputation with missingness indicators; excluding records with a hospital stay ≥ 20 days and total cost $<10,000$ yuan; For four-quadrant stratification, we repeatedly retrained both models using bootstrap resampling and reassigned patients to quadrants in each iteration. Stability was defined as the proportion of iterations in which each patient remained in the original quadrant.

To examine whether the relatively low observed hospitalization costs in Group II were attributable to early death and shorter hospital stays, we compared length of stay, mortality within 3 and 7 days after admission, total hospitalization costs, and

hospitalization costs per day across the four groups. We further repeated the four-quadrant analysis after excluding patients who died within 3 and 7 days after admission. The original predicted probabilities and risk thresholds derived from the training cohort were retained, and the models were not retrained.

2.9. Statistical analysis

A two-sided p value <0.05 was considered statistically significant. All analyses and model development were performed in Python 3.12.4. Continuous variables were compared using the Mann–Whitney U test for two groups and the Kruskal–Wallis test for multiple groups. Categorical variables were compared using the chi-square test or Fisher’s exact test. Spearman rank correlation was used for correlation analyses. The Benjamini–Hochberg method controlled the false discovery rate in multiple comparisons.

3. Results

3.1. Study population

The final cohort included 1,543 patients with sCAP, comprising 1,262 in the training cohort and 281 in the temporal validation cohort. In-hospital death occurred in 345 of 1,262 patients (27.3%) in the training cohort and 93 of 281 patients (33.1%) in the temporal validation cohort. Among patients who died, the median length of hospital stay was 11 days (IQR, 8–15) in the training cohort and 11 days (IQR, 8–14) in the temporal validation cohort. The mean age was 75.90 ± 11.84 years, and 981 patients (63.6%) were male. Common comorbidities included hypertension in 961 patients (62.3%), heart disease in 760 (49.3%), and respiratory disease in 920 (59.6%).

In-hospital mortality occurred in 438 patients (28.4%). The median total hospitalization cost was CNY 21,948.56 (IQR, 14,105.00–37,816.58). Baseline characteristics are shown in Table 1.

3.2. Feature selection

After collinearity assessment, 41 candidate variables entered dual-outcome Elastic Net bootstrap stability selection (Table S1). Within the fully nested outer five-fold cross-validation, 11-18 features were retained in each fold, and all 500 bootstrap resamples were successfully completed. Ten predictors were retained in all five outer folds, including age, admitting department, Glasgow Coma Scale score, arterial partial pressure of oxygen, lactate, respiratory rate, albumin, hemoglobin, procalcitonin, and respiratory disease, indicating consistent retention of a core predictor set.

When the feature-selection procedure was repeated in the full training cohort, 26 variables met the stability criterion of a selection frequency ≥ 0.60 for either outcome. After ranking by the joint stability score and limiting the design matrix to fewer than 20 degrees of freedom, 16 shared features were retained: age, admitting department, Glasgow Coma Scale score, arterial partial pressure of oxygen, lactate, systolic blood pressure, diastolic blood pressure, respiratory rate, albumin, hemoglobin, procalcitonin, serum potassium, respiratory disease, heart disease, viral test results, and Mycoplasma pneumoniae test results.

3.3. Model development and performance evaluation

Among the five candidate models, LightGBM performed best for in-hospital

mortality, whereas CatBoost performed best for high-cost hospitalization. In temporal validation, the LightGBM model achieved an AUROC of 0.734 (95% CI, 0.663–0.794) and an AUPRC of 0.586 (95% CI, 0.486–0.701) for in-hospital mortality. The CatBoost model achieved an AUROC of 0.700 (95% CI, 0.620–0.773) and an AUPRC of 0.334 (95% CI, 0.246–0.464) for high-cost hospitalization.

In temporal validation, the mortality model significantly outperformed CURB-65, with AUROCs of 0.734 and 0.538, respectively (difference, 0.196; 95% CI, 0.099–0.293; DeLong $P < 0.001$). The high-cost hospitalization model also significantly outperformed the baseline model based on age and comorbidity count, with AUROCs of 0.700 and 0.534, respectively (difference, 0.166; 95% CI, 0.064–0.267; DeLong $P = 0.001$). Temporal validation performance is shown in Table 2, and the model curves are presented in Figure 2.

3.4. Feature contribution and predictor inconsistency analysis

After adjustment for age, CURB-65, and major comorbidities, in-hospital mortality was not significantly associated with high-cost hospitalization ($P = 0.468$) or log10-transformed total hospitalization cost ($\beta = 0.032$, $P = 0.150$). Across the five candidate algorithms, predictions for the same outcome were highly consistent. Cross-outcome Spearman ρ values ranged from 0.240 to 0.389 in the training cohort and from 0.153 to 0.340 in temporal validation, whereas median within-outcome correlations across algorithms were substantially higher. These findings indicate that the differences between the two outcome models were robust to algorithm choice (Table S2).

Predicted probabilities of in-hospital mortality and high-cost hospitalization were weakly correlated, with a Spearman ρ of 0.359 in temporal validation ($P < 0.001$). SHAP feature rankings differed between the two models ($P = 0.018$). Four of the top 10 features overlapped: admitting department, GCS, systolic blood pressure, and PaO_2 . In the four-quadrant analysis, Group II had an in-hospital mortality rate of 84.2%, whereas Group IV had a high-cost hospitalization rate of 37.3% in temporal validation. Details are shown in Figure 3 and Table S3.

3.5. Exploratory analysis

Exploratory quadrant analysis showed that Group II was characterized by initial lactate, initial procalcitonin, age, admitting department, respiratory rate, ICU admission, and mechanical ventilation. Group IV showed higher observed expenditures across several cost categories and differences in medication profiles and selected clinical variables, including initial albumin, respiratory disease, initial hemoglobin, admitting department, medical consumable costs, other costs, nursing costs, number of reimbursed medications, examination costs, initial PaO_2 , proportion of generic drugs, and medication costs. The observed cost difference between Groups II and IV was distributed across medical consumable, other, nursing, examination, and medication cost categories (Figure 4).

3.6. Sensitivity analysis

When high-cost hospitalization was redefined using the 80th and 90th percentiles, the model achieved AUROCs of 0.663 and 0.669, respectively. After excluding the highest 1% of total costs, the AUROC was 0.659. With median imputation and

missingness indicators, the AUROCs were 0.688 for in-hospital mortality and 0.694 for high-cost hospitalization. Excluding records with hospital stays ≥ 20 days and total costs $< \text{CNY } 10,000$ did not change the results. Bootstrap analysis showed that the mean proportions of patients reassigned to their original quadrants were 0.625 for training-cohort OOF predictions and 0.756 for temporal validation; the corresponding medians were 0.715 and 0.845. Mean stability was 0.653 for Group IV and 0.543 for Group II.

After excluding patients who died during hospital stays of ≤ 3 or ≤ 7 days, Group II remained identifiable, and the contrasting mortality and high-cost hospitalization profiles of Groups II and IV persisted. After the exclusion of deaths within 7 days, 197 of the original 220 patients in Group II were retained in the training cohort (89.5%), and 16 of the original 19 patients were retained in the temporal validation cohort (84.2%). In temporal validation, Group II continued to have a higher in-hospital mortality rate than Group IV (81.2% vs. 21.3%) but a lower rate of high-cost hospitalization (12.5% vs. 41.0%). Median length of stay was identical between the two groups (11.0 vs. 11.0 days; $P=0.786$), whereas median hospitalization cost per day remained lower in Group II (CNY 2,103.19 vs. CNY 3,629.58; Benjamini–Hochberg-adjusted $P=0.044$). Table S4.

4. Discussion

We developed separate models for in-hospital mortality and high-cost hospitalization in patients with sCAP, establishing a dual-risk assessment framework based on the same information available at hospital admission. In temporal validation,

the LightGBM model for in-hospital mortality and the CatBoost model for high-cost hospitalization achieved AUROCs of 0.734 and 0.700, respectively. Based on the joint distribution of the two predicted risks, patients were further classified into Groups I, II,III, and IV using four-quadrant stratification, and Groups II and IV were identified as two patient subgroups that were particularly challenging to manage in clinical practice. This framework may support risk-specific management while jointly considering clinical outcomes and hospital costs.

A key contribution of this study was the use of a unified set of information available at hospital admission to assess both in-hospital mortality and high-cost hospitalization. Previous CAP cost studies often incorporated in-hospital factors, including length of stay, complications, and ICU treatment [12, 13]. Such information was useful for explaining costs retrospectively but offered limited support for decisions at admission. By restricting model inputs to hospital admission, our framework shifted cost assessment from retrospective explanation toward earlier risk prediction. A common feature set was then identified using dual-outcome Elastic Net with 500 bootstrap resamples. Under this unified input structure, both models retained acceptable discrimination in temporal validation and outperformed their respective reference tools. These findings indicated that the same at-admission information could support prediction of both outcomes. This design avoided reliance on data generated after hospitalization had begun and provided a consistent basis for subsequent four-quadrant stratification.

378 Compared with conventional reference tools, these variables available at hospital
379 admission provided additional discrimination for both risk assessments. In temporal
380 validation, the model for in-hospital mortality outperformed CURB-65, while the
381 model for high-cost hospitalization outperformed the simple baseline model based on
382 age and comorbidity count. Both differences were statistically significant. CURB-65
383 incorporates confusion, urea, respiratory rate, blood pressure, and age [5, 6], whereas
384 the simple baseline model for high-cost hospitalization included only age and
385 comorbidity count. In contrast, our common feature set covered multiple clinical
386 domains assessed at hospital admission, including oxygenation, tissue perfusion,
387 inflammatory burden, physiological reserve, and underlying cardiopulmonary disease.
388 A lower PaO_2 reflected impaired alveolar capillary gas exchange, which could lead to
389 hypoxemia and reduced tissue oxygen delivery [17]. Elevated lactate indicated tissue
390 hypoperfusion, microcirculatory dysfunction, and metabolic stress and was an
391 important sign of circulatory decompensation [18, 19]. These variables captured key
392 pathophysiological dimensions not fully represented in the conventional reference
393 tools, which may partly explain the additional discrimination achieved by both models.

394 More importantly, a clinical challenge arises when predicted mortality risk and
395 predicted risk of high-cost hospitalization do not align. Two clinically relevant
396 discordant profiles may therefore be observed: patients with a high predicted risk of
397 in-hospital mortality but a low predicted risk of high-cost hospitalization, and patients
398 with a high predicted risk of high-cost hospitalization but a relatively low predicted
399 risk of death. Conventional single-outcome assessment may not readily reveal these

discordant profiles or their different clinical implications. By jointly examining the two predicted probabilities, the exploratory four-quadrant analysis illustrated these risk discrepancies at the patient level. Patients in Group II were generally older and had higher respiratory rates, lactate levels, and procalcitonin levels at admission. They were also more likely to require ICU admission and mechanical ventilation. An increased respiratory rate reflected greater respiratory workload and could indicate worsening respiratory failure, whereas elevated procalcitonin was associated with a greater bacterial burden and a more pronounced systemic inflammatory response [5, 20-22]. Collectively, the higher respiratory rate, lactate and procalcitonin levels, together with the more frequent need for ICU admission and mechanical ventilation, suggested greater physiological instability and a higher likelihood of clinical deterioration. In contrast, Group IV was characterized by higher expenditures on medical consumables, nursing care, examinations, and medications. The increase across multiple cost categories suggested cumulative resource intensity rather than a single dominant cost driver. These descriptive findings may generate hypotheses regarding the appropriateness and efficiency of clinical processes and resource use, but they do not establish a specific resource-management strategy. The four-quadrant analysis therefore provided exploratory information on differences across predicted risk combinations, and its value for guiding clinical monitoring or resource planning requires prospective validation.

Our study also has certain limitations. First, it was a single-center retrospective study and was therefore subject to selection and information bias. Although temporal

validation was performed, the models had not been independently validated across multiple centers. Their applicability and generalizability to other institutions therefore remained uncertain. Second, high-cost hospitalization was influenced by local pricing standards, healthcare policies, and care pathways, which could vary substantially among hospitals. The models therefore needed to be revalidated and recalibrated before use at other centers. Finally, SHAP analysis and four-quadrant stratification were used to interpret model outputs and describe patient characteristics. They could not establish causal relationships between predictors and outcomes. The proposed risk stratification and its value for clinical management required further evaluation in prospective multicenter studies.

5. Conclusion

In conclusion, we developed and temporally validated separate machine learning models for in-hospital mortality and high-cost hospitalization in patients with sCAP using the same set of information available at hospital admission. Exploratory four-quadrant analysis illustrated differences among patients with different combinations of predicted risks. These models may support risk assessment at admission, whereas their clinical utility and generalizability require further validation.

Abbreviations

CAP, community-acquired pneumonia; sCAP, severe community-acquired pneumonia; PSI, Pneumonia Severity Index; CNY, Chinese yuan; CURB-65, confusion, urea,

respiratory rate, blood pressure, and age ≥ 65 years; GCS, Glasgow Coma Scale; ICU, intensive care unit; WBC, white blood cell count; ALT, alanine aminotransferase; APTT, activated partial thromboplastin time; AST, aspartate aminotransferase; BUN, blood urea nitrogen; CKD, chronic kidney disease; CRP, C-reactive protein; PaCO₂, arterial partial pressure of carbon dioxide; PaO₂, arterial partial pressure of oxygen; PCT, procalcitonin; PT, prothrombin time; AUPRC, area under the precision–recall curve; AUROC, area under the receiver operating characteristic curve; CI, confidence interval; DCA, decision curve analysis; IQR, interquartile range; OOF, out-of-fold; SD, standard deviation; L2-LR, L2-regularized logistic regression; LightGBM, Light Gradient Boosting Machine; LR, logistic regression; RF, random forest; XGBoost, eXtreme Gradient Boosting; CatBoost, Categorical Boosting; SHAP, SHapley Additive exPlanations; TRIPOD+AI, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis–Artificial Intelligence.

Declarations

Ethics approval and consent to participate: The Ethics Committee of Nanjing First Hospital approved our study protocols. Moreover, informed permission was exempted due to the retrospective study design and anonymous data collection.

Author statement: We certify that this article is not under consideration for publication elsewhere, and all authors approve its publication. If accepted, it will not be published elsewhere in the same form in any other language.

Availability of data and materials: The datasets used and analyzed during the current study cannot be shared publicly due to ethical reasons but are available from the corresponding author on reasonable request.

Disclosure: The authors report no conflicts of interest in this work.

Funding: This work was supported by National Natural Science Foundation of China (82173899); Jiangsu Pharmaceutical Association (Q202202, JY202207, Z04JKM2023E040, A202309); Jiangsu Research Hospital Association for Precision Medication (SYHKJ-JY-2025-15).

Authors' contributions: **Shiting Chen:** Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft. **Jiaxin Ren:** Methodology, Investigation, Formal analysis, Writing – review & editing. **Hui Jiang:** Methodology, Resources, Data curation, Writing – review & editing. **Zhou Zhou:** Conceptualization, Software, Data curation, Writing – review & editing. **Ziye Huang:** Methodology, Resources, Data curation, Writing – review & editing. **Weixuan Shi:** Data curation, Resources, Writing – review & editing. **Zhengyang Wu:** Data curation, Resources, Writing – review & editing. **Jianjun Zou:** Conceptualization, Formal analysis, Visualization, Validation, Writing – review & editing. **Hao Liu:** Conceptualization, Formal analysis, Writing – review & editing. **Meng Wan:** Conceptualization, Methodology, Project administration, Writing – review & editing. All authors read and approved the final manuscript.

Acknowledgment: We are deeply grateful to all patients who participated in this

study. To all the people who made efforts for this study, we would also like to express our gratitude for their collaboration.

E-mail address of each author:

Shiting Chen: chenst.0916@foxmail.com

Jiaxin Ren: 15735264968@163.com

Hui Jiang:jianghui2000@foxmail.com

Zhou Zhou: evanzhow_0826@163.com

Ziye Huang: huangziye2020@163.com

Weixuan Shi: yxswx77@163.com

Zhengyang Wu: 13765008574@163.com

References

1. Restrepo MI, Anzueto A: Severe community-acquired pneumonia. *Infect Dis Clin North Am* 2009, 23(3):503–520.
2. Niederman MS, Torres A: Severe community-acquired pneumonia. *Eur Respir Rev* 2022, 31(166).
3. Niederman MS, McCombs JS, Unger AN, Kumar A, Popovian R: The cost of treating community-acquired pneumonia. *Clin Ther* 1998, 20(4):820–837.
4. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, Coley CM, Marrie TJ, Kapoor WN: A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* 1997, 336(4):243–250.

- 509 5. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI,
510 Lewis SA, Macfarlane JT: Defining community acquired pneumonia severity
511 on presentation to hospital: an international derivation and validation study.
512 *Thorax* 2003, 58(5):377–382.
- 513 6. Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, Cooley
514 LA, Dean NC, Fine MJ, Flanders SA *et al*: Diagnosis and Treatment of Adults
515 with Community-acquired Pneumonia. An Official Clinical Practice Guideline
516 of the American Thoracic Society and Infectious Diseases Society of America.
517 *Am J Respir Crit Care Med* 2019, 200(7):e45–e67.
- 518 7. Cilloniz C, Ward L, Mogensen ML, Pericàs JM, Méndez R, Gabarrús A,
519 Ferrer M, Garcia-Vidal C, Menendez R, Torres A: Machine-Learning Model
520 for Mortality Prediction in Patients With Community-Acquired Pneumonia:
521 Development and Validation Study. *Chest* 2023, 163(1):77–88.
- 522 8. Chen H, Zhang S, Matsumoto H, Tsuchiya N, Yamada C, Okasaki S, Miyasaka
523 A, Yumoto K, Kanou D, Kashizaki F *et al*: Employing a low-code machine
524 learning approach to predict in-hospital mortality and length of stay in patients
525 with community-acquired pneumonia. *Sci Rep* 2025, 15(1):309.
- 526 9. Pan J, Guo T, Kong H, Bu W, Shao M, Geng Z: Prediction of mortality risk in
527 patients with severe community-acquired pneumonia in the intensive care unit
528 using machine learning. *Sci Rep* 2025, 15(1):1566.

- 529 10. Uematsu H, Yamashita K, Kunisawa S, Imanaka Y: Prediction model for
530 prolonged length of stay in patients with community-acquired pneumonia
531 based on Japanese administrative data. *Respir Investig* 2021, 59(2):194–203.
- 532 11. Wu C, Zha D, Gao H: Prediction of Bronchopneumonia Inpatients' Total
533 Hospitalization Expenses Based on BP Neural Network and Support Vector
534 Machine Models. *Comput Math Methods Med* 2022, 2022:9275801.
- 535 12. Reyes S, Martinez R, Vallés JM, Cases E, Menendez R: Determinants of
536 hospital costs in community-acquired pneumonia. *Eur Respir J* 2008,
537 31(5):1061–1067.
- 538 13. Gümüş A, Çilli A, Çakın Ö, Karakurt Z, Ergan B, Aksoy E, Cengiz M: Factors
539 Affecting Cost of Patients with Severe Community-Acquired Pneumonia in
540 Intensive Care Unit. *Turk Thorac J* 2019, 20(4):216–223.
- 541 14. Fine MJ, Pratt HM, Obrosky DS, Lave JR, McIntosh LJ, Singer DE, Coley
542 CM, Kapoor WN: Relation between length of hospital stay and costs of care
543 for patients with community-acquired pneumonia. *Am J Med* 2000,
544 109(5):378–385.
- 545 15. Bhanvadia RR, Baky FJ, Ashbrook CQ, Lotan Y, Woldu SL: Pathologic
546 fracture in metastatic kidney cancer: Identifying widening disparities and
547 opportunity for quality improvement. *Urol Oncol* 2022, 40(8):384.e381–
548 384.e388.

- 549 16. Kelly NE, An A, Nash KA, Abramson EL: Admissions for Pediatric Somatic
550 Symptom and Related Disorders: A National Database Study. *Hosp Pediatr*
551 2026, 16(1):77–84.
- 552 17. Bahçecioğlu SN, Köktürk N, Baha A, Yapar D, Aksakal FNB, Gunduz C,
553 Tasbakan S, Sayiner A, Coskun AS, Yaman F *et al*: A new scoring system to
554 predict mortality in community-acquired pneumonia: CURB (S)-65. *Eur Rev*
555 *Med Pharmacol Sci* 2023, 27(13):6293–6300.
- 556 18. Gwak MH, Jo S, Jeong T, Lee JB, Jin YH, Yoon J, Park B: Initial serum
557 lactate level is associated with inpatient mortality in patients with
558 community-acquired pneumonia. *Am J Emerg Med* 2015, 33(5):685–690.
- 559 19. Frenzen FS, Kutschan U, Meiswinkel N, Schulte-Hubbert B, Ewig S, Kolditz
560 M: Admission lactate predicts poor prognosis independently of the
561 CRB/CURB-65 scores in community-acquired pneumonia. *Clin Microbiol*
562 *Infect* 2018, 24(3):306.e301–306.e306.
- 563 20. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC,
564 Dowell SF, File TM, Jr., Musher DM, Niederman MS *et al*: Infectious
565 Diseases Society of America/American Thoracic Society consensus guidelines
566 on the management of community-acquired pneumonia in adults. *Clin Infect*
567 *Dis* 2007, 44 Suppl 2(Suppl 2):S27–72.
- 568 21. Krüger S, Ewig S, Marre R, Papassotiriou J, Richter K, von Baum H, Suttorp
569 N, Welte T: Procalcitonin predicts patients at low risk of death from

570 community-acquired pneumonia across all CRB-65 classes. *Eur Respir J* 2008,
571 31(2):349–355.

572 22. Schuetz P, Suter-Widmer I, Chaudri A, Christ-Crain M, Zimmerli W, Mueller
573 B: Prognostic value of procalcitonin in community-acquired pneumonia. *Eur*
574 *Respir J* 2011, 37(2):384–392.

575

576

577 **Figure legends**

578 **Figure 1. Overall analytical workflow for dual risk prediction of sCAP.**

579 The figure outlines the overall analytical workflow, including retrospective sCAP
580 cohort construction, extraction of predictors available at hospital admission, data
581 cleaning, redundancy removal and elastic net bootstrap feature selection. Dual
582 prediction models targeting in-hospital mortality and high-cost hospitalization were
583 developed with multiple machine learning algorithms. Multi-level discordance
584 analyses were used to assess the mismatch between clinical and economic risk,
585 followed by exploratory sub-analysis to identify major drivers of increased hospital
586 costs.

587 **Abbreviations:** sCAP, severe community-acquired pneumonia; Z-score, standard
588 score; LR, L2 regularized logistic regression; RF, random forest; XGBoost, extreme
589 gradient boosting; LightGBM, light gradient boosting machine; CatBoost, categorical
590 boosting; SHAP, SHapley Additive exPlanations; ICU, intensive care unit; OOF,
591 out-of-fold.

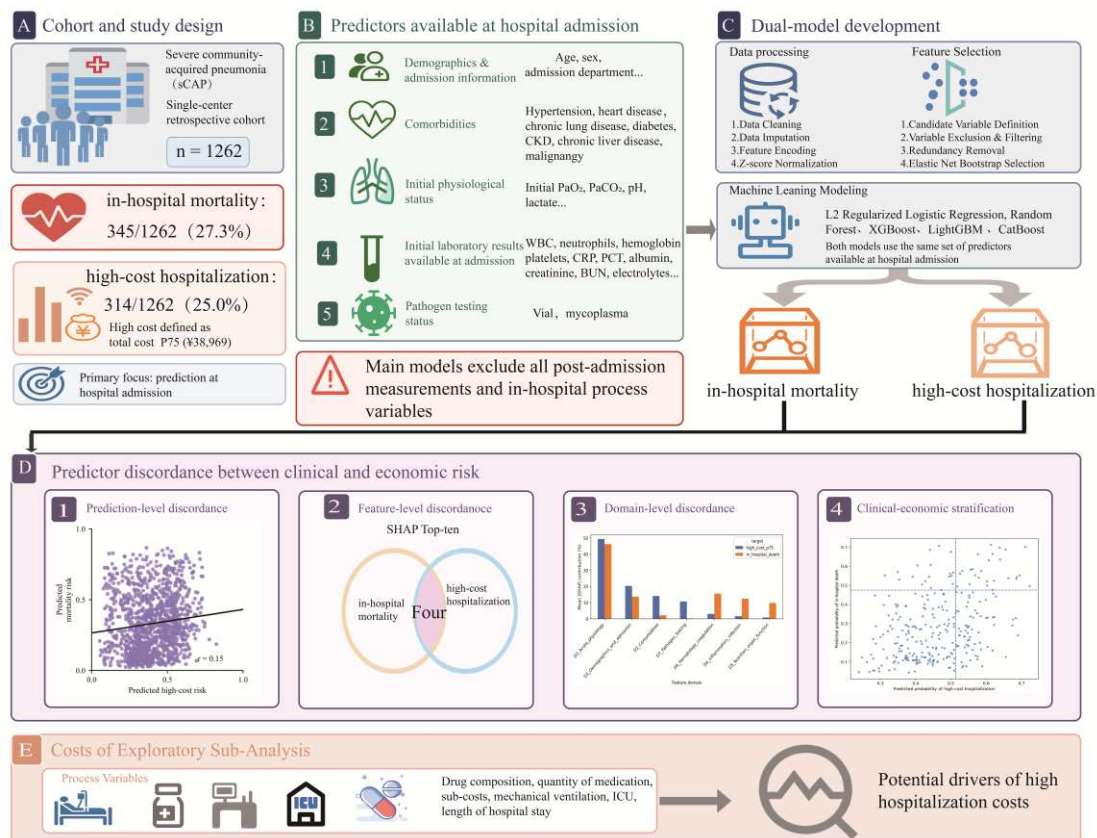


Figure 2. Performance comparison of machine learning models.

(A-D) ROC , PR , calibration, and DCA curves of five machine learning models for predicting in-hospital mortality.

(E-H) ROC, PR, calibration, and DCA curves of five machine learning models for predicting high-cost hospitalization.

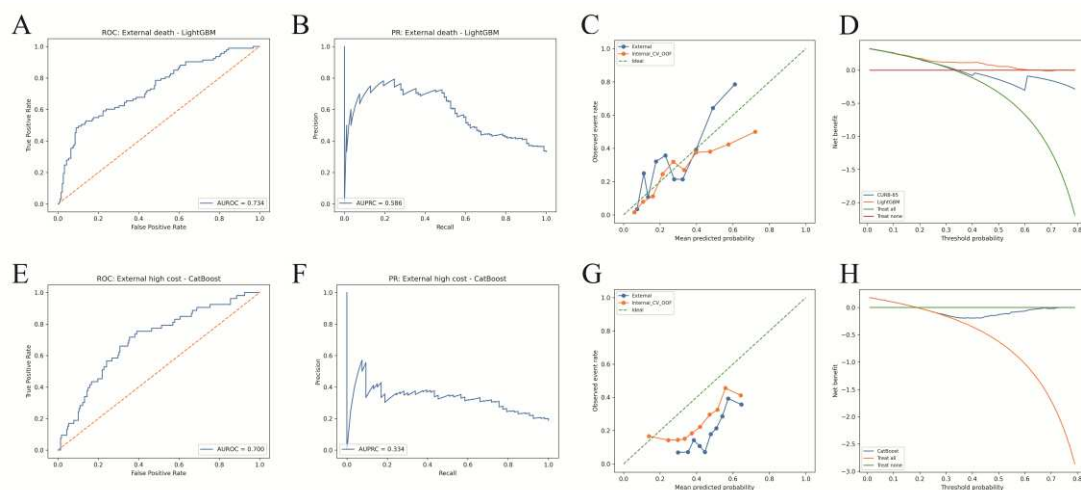


Figure 3. Feature analysis of in-hospital mortality and high-cost hospitalization.

(A) SHAP summary plot for the high-cost hospitalization prediction model.

(B) SHAP summary plot for the in-hospital mortality prediction model.

(C) Bar chart of mean SHAP contributions of distinct feature domains for the two outcomes.

(D) Four-quadrant scatter plot of predicted in-hospital mortality versus high-cost hospitalization.

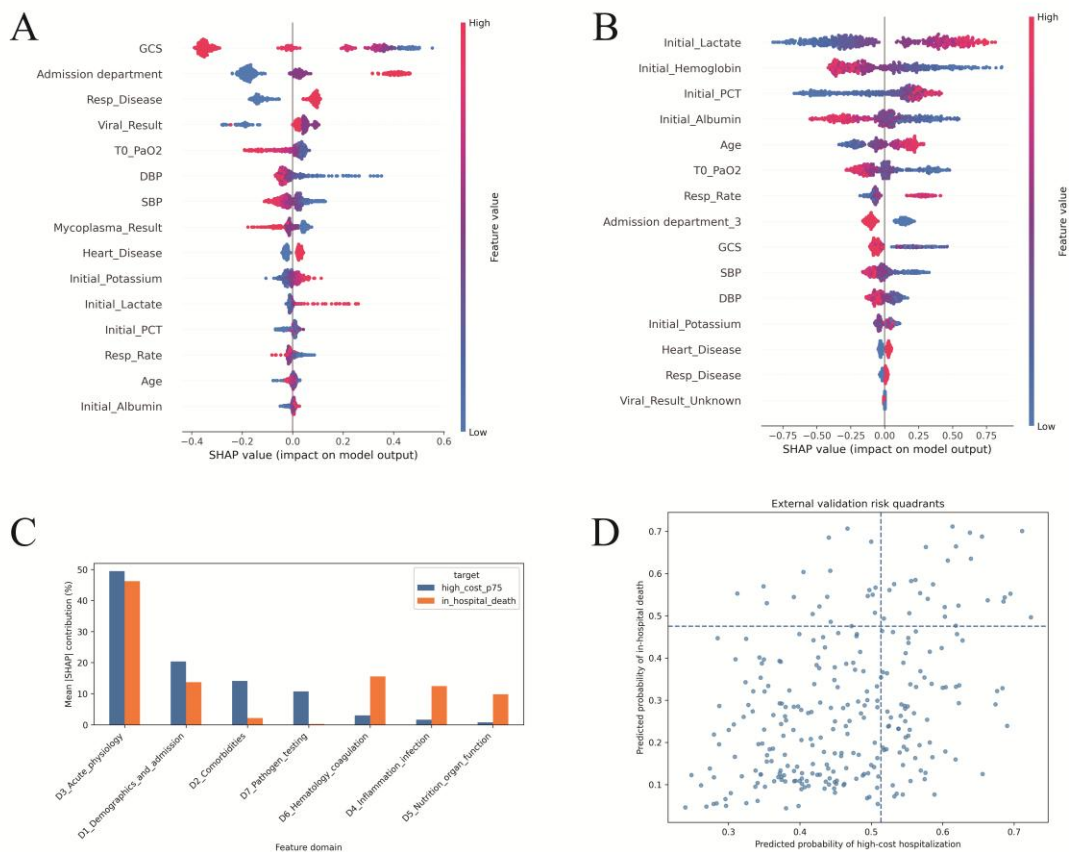
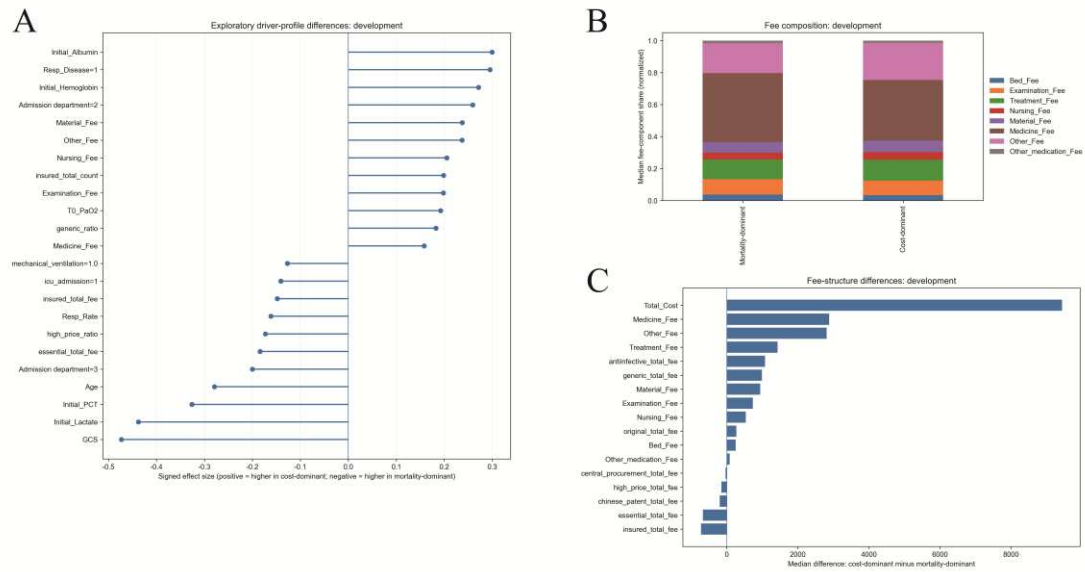


Figure 4. Exploratory analysis

(A) Differences in variable characteristics between Group II and Group IV.

(B) Composition of hospitalization fees for Group II and Group IV.

(C) Cost-component differences between discordant groups.



611

612 **Abbreviations:** Group II, High in-hospital mortality with low high-cost
 613 hospitalization; Group IV, Low in-hospital mortality with high high-cost
 614 hospitalization

615

616 Table 1 The demographic and clinical characteristics in the whole cohort.

Characteristics	Overall (N=1543)	Training Cohort (N=1262)	Temporal validation cohort (N=281)	P Value
Demographic characteristics				
Age, years	75.90 ± 11.84	75.58 ± 11.74	77.30 ± 12.23	0.033
Gender, n (%)				<0.001
Female	562 (36.4)	414 (32.8)	148 (52.7)	
Male	981 (63.6)	848 (67.2)	133 (47.3)	
Receipt of surgery, n (%)				1.000
No	1070 (69.3)	875 (69.3)	195 (69.4)	
Yes	473 (30.7)	387 (30.7)	86 (30.6)	
Disease severity and vital signs at admission				
Critical illness, n (%)				0.694
No	110 (7.1)	92 (7.3)	18 (6.4)	
Yes	1433 (92.9)	1170 (92.7)	263 (93.6)	
ICU admission, n (%)				0.802
No	628 (40.7)	516 (40.9)	112 (39.9)	
Yes	915 (59.3)	746 (59.1)	169 (60.1)	
Mechanical ventilation, n (%)				<0.001
No	884 (59.8)	749 (62.0)	135 (49.8)	
Yes	595 (40.2)	459 (38.0)	136 (50.2)	
Missing	64 (4.1%)	54 (4.3%)	10 (3.6%)	
Respiratory rate, breaths/min	25.59 ± 5.42	25.59 ± 5.41	25.60 ± 5.44	0.962
Systolic blood pressure, mmHg	114.69 ± 19.59	114.97 ± 19.64	113.40 ± 19.34	0.219
Diastolic blood pressure, mmHg	69.26 ± 15.50	69.32 ± 15.46	68.97 ± 15.70	0.730
Initial laboratory parameters				
WBC, ×10 ⁹ /L	11.62 ± 4.89	11.57 ± 4.91	11.86 ± 4.84	0.365
Hemoglobin, g/L	114.94 ± 19.51	115.11 ± 19.62	114.21 ± 19.02	0.474
Creatinine, μmol/L	97.00 [69.00, 131.00]	97.00 [68.00, 131.00]	97.00 [71.00, 131.00]	0.674
C-reactive protein, mg/L	142.30 [84.46, 206.00]	140.75 [83.40, 201.67]	153.30 [94.90, 215.00]	0.068

Characteristics	Overall (N=1543)	Training Cohort (N=1262)	Temporal validation cohort (N=281)	P Value
PaO ₂ , mmHg	71.52 ± 20.80	71.54 ± 20.78	71.41 ± 20.94	0.926
Microbiological testing				
Viral test, n (%)				0.530
No	462 (29.9)	373 (29.6)	89 (31.7)	
Yes	1081 (70.1)	889 (70.4)	192 (68.3)	
Not tested	708 (45.9%)	586 (46.4%)	122 (43.4%)	
Mycoplasma test, n (%)				0.246
No	528 (34.2)	423 (33.5)	105 (37.4)	
Yes	1015 (65.8)	839 (66.5)	176 (62.6)	
Chlamydia result, n (%)				0.212
No	723 (86.6)	580 (85.8)	143 (89.9)	
Yes	112 (13.4)	96 (14.2)	16 (10.1)	
Comorbidities				
Hypertension, n (%)				0.945
No	582 (37.7)	475 (37.6)	107 (38.1)	
Yes	961 (62.3)	787 (62.4)	174 (61.9)	
Heart disease, n (%)				0.990
No	783 (50.7)	641 (50.8)	142 (50.5)	
Yes	760 (49.3)	621 (49.2)	139 (49.5)	
Respiratory disease, n (%)				0.587
No	623 (40.4)	505 (40.0)	118 (42.0)	
Yes	920 (59.6)	757 (60.0)	163 (58.0)	
Outcomes and resource use				
In-hospital death, n (%)				0.062
No	1105 (71.6)	917 (72.7)	188 (66.9)	
Yes	438 (28.4)	345 (27.3)	93 (33.1)	
Length of stay, days	11.00 [8.00, 15.00]	11.00 [8.00, 15.00]	11.00 [8.00, 14.00]	0.896
Total hospitalization cost, CNY	21948.56 [14105.00, 37816.58]	20842.50 [13420.23, 38969.00]	25520.49 [19050.10, 34857.81]	<0.001

Notes: Continuous variables are presented as mean ± SD or median [IQR] according to distribution. Categorical variables are presented as n (%). P values were calculated using Welch t test or Mann–Whitney U test for continuous variables and chi-square test or Fisher exact test for categorical variables, as appropriate.

622 Table 2 Performance of the various models in time-segmented validation

Name	Model	AUROC	AUPRC	Accuracy	F1	Brier
in-hospital mortality						
model	L2-LR	0.733	0.610	0.679	0.558	0.188
in-hospital mortality						
model	RF	0.731	0.596	0.715	0.560	0.194
in-hospital mortality						
model	XGBoost	0.714	0.563	0.622	0.561	0.194
in-hospital mortality						
model	LightGBM	0.734	0.586	0.590	0.541	0.188
in-hospital mortality						
model	CatBoost	0.733	0.574	0.565	0.567	0.213
VS ^a	CURB-65	0.538	0.348			0.278
high-cost						
hospitalization model	L2-LR	0.663	0.286	0.654	0.374	0.246
high-cost						
hospitalization model	RF	0.692	0.351	0.665	0.453	0.170
high-cost						
hospitalization model	XGBoost	0.696	0.330	0.804	0.179	0.218
high-cost						
hospitalization model	LightGBM	0.680	0.307	0.804	0.126	0.216
high-cost						
hospitalization model	CatBoost	0.700	0.334	0.469	0.376	0.224
VS ^b	LR	0.534	0.281			0.218

623 Notes: VS^a: A Comparison of CURB-65 and the In-Hospital Mortality Model Using
624 LightGBM; VS^b: A comparison of logistic regression models for age and the number of
625 comorbidities with the CatBoost model for predicting the risk of high-cost hospitalisation

Supplementary Files

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

- [TableS1.docx](#)
- [TableS2.docx](#)
- [TableS3.docx](#)
- [TableS4.docx](#)