BMJ Open Respiratory Research

Machine learning-based model for predicting all-cause mortality in severe pneumonia

Weichao Zhao ,[1,2] Xuyan Li,[1] Lianjun Gao,[3] Zhuang Ai,[4] Yaping Lu,[4] Jiachen Li ,[5] Dong Wang,[1] Xinlou Li,[6] Nan Song,[7] Xuan Huang,[7] Zhao-hui Tong[1,7]

To cite: Zhao W, Li X, Gao L, *et al*. Machine learning-based model for predicting all-cause mortality in severe pneumonia. *BMJ Open Respir Res* 2025;**12**:e001983. doi:10.1136/bmjresp-2023-001983

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjresp-2023-001983).

WZ, XL and LG contributed equally.

Received 27 July 2023
Accepted 15 October 2024

For numbered affiliations see end of article.

Correspondence to
Zhao-hui Tong; tongzhaohuicy@sina.com and Dr Xuan Huang; huangxuan03@163.com

ABSTRACT

Background Severe pneumonia has a poor prognosis and high mortality. Current severity scores such as Acute Physiology and Chronic Health Evaluation (APACHE-II) and Sequential Organ Failure Assessment (SOFA), have limited ability to help clinicians in classification and management decisions. The goal of this study was to analyse the clinical characteristics of severe pneumonia and develop a machine learning-based mortality-prediction model for patients with severe pneumonia.

Methods Consecutive patients with severe pneumonia between 2013 and 2022 admitted to Beijing Chaoyang Hospital affiliated with Capital Medical University were included. In-hospital all-cause mortality was the outcome of this study. We performed a retrospective analysis of the cohort, stratifying patients into survival and non-survival groups, using mainstream machine learning algorithms (light gradient boosting machine, support vector classifier and random forest). We aimed to construct a mortality-prediction model for patients with severe pneumonia based on their accessible clinical and laboratory data. The discriminative ability was evaluated using the area under the receiver operating characteristic curve (AUC). The calibration curve was used to assess the fit goodness of the model, and decision curve analysis was performed to quantify clinical utility. By means of logistic regression, independent risk factors for death in severe pneumonia were figured out to provide an important basis for clinical decision-making.

Results A total of 875 patients were included in the development and validation cohorts, with the in-hospital mortality rate of 14.6%. The AUC of the model in the internal validation set was 0.8779 (95% CI, 0.738 to 0.974), showing a competitive discrimination ability that outperformed those of traditional clinical scoring systems, that is, APACHE-II, SOFA, CURB-65 (confusion, urea, respiratory rate, blood pressure, age ≥65 years), Pneumonia Severity Index. The calibration curve showed that the in-hospital mortality in severe pneumonia predicted by the model fit reasonably with the actual hospital mortality. In addition, the decision curve showed that the net clinical benefit was positive in both training and validation sets of hospitalised patients with severe pneumonia. Based on ensemble machine learning algorithms and logistic regression technique, the level of ferritin, lactic acid, blood urea nitrogen, creatine kinase, eosinophil and the requirement of vasopressors were identified as top independent predictors of in-hospital mortality with severe pneumonia.

Conclusion A robust clinical model for predicting the risk of in-hospital mortality after severe pneumonia was successfully developed using machine learning techniques. The performance of this model demonstrates the effectiveness of these techniques in creating accurate predictive models, and the use of this model has the potential to greatly assist patients and clinical doctors in making well-informed decisions regarding patient care.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Risk-assessment systems for severe pneumonia have been well-applied in the complex and increasing intensive care unit (ICU) population. However, these tools have limited accuracy and applicability for mortality in severe pneumonia. Further research is needed to identify independent risk factors and build predictive models, enabling us to differentiate patients at risk of death early and decide whether to provide preventive or aggressive treatment.

WHAT THIS STUDY ADDS

⇒ By focusing on ensemble machine learning algorithms, we successfully constructed a mortality-prediction model for patients with severe pneumonia using comprehensive clinical and laboratory data, which outperformed conventional scoring systems. Additionally, we identified ferritin, lactic acid, blood urea nitrogen, creatine kinase, eosinophil and the requirement of vasopressors as predictive factors for overall in-hospital mortality in severe pneumonia.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The use of this model has the potential to greatly assist patients and clinical doctors in making well-informed decisions regarding patient care. With the aid of this model, there is a potential for more efficient treatment of patients with severe pneumonia, leading to a rational allocation of ICU resources and alleviating the burden on society. And the identified risk indicators will provide guiding significance for the mechanistic study of disease progression in severe pneumonia.

INTRODUCTION

Pneumonia is a major health problem associated with high incidence rates and both short- and long-term mortality. Pneumonia in critically ill patients can be characterised as community-acquired pneumonia (CAP), hospital-acquired pneumonia or ventilator-associated pneumonia.[1]

Severe CAP (SCAP) is the most critical type of CAP with high mortality, accounting for 17%–21% of hospitalised patients with CAP.[2 3] The number of patients admitted to the intensive care unit (ICU) globally due to SCAP is increasing every year, especially among the elderly, those with underlying diseases and those who are immuno-suppressed.[4] Previous studies have found that the mortality rate of CAP is still high, ranging from 17% in the hospital to 50% in 1 year.[5–7] Up to 35% of ICU patients with SCAP may die during their hospital stay, and >50% of patients die within 1 year after acute onset.[8] Delayed admission to the ICU can further increase the mortality rate of SCAP.[9] Therefore, exploring early diagnosis, predicting prognosis and uncovering-related molecular markers are particularly important for timely appropriate treatment and prevention measures to improve patient outcomes.

In recent years, risk-assessment systems for the severity of pneumonia have been well-applied in the complex and increasing ICU population. However, these tools have limited accuracy and applicability for mortality, and often require a collection of cumbersome information that is difficult to obtain within 24 hours of admission.

The traditional Acute Physiology and Chronic Health Evaluation (APACHE-II), Sequential Organ Failure Assessment (SOFA), CURB-65 score (confusion, urea, respiratory rate, blood pressure, age ≥65 years), Pneumonia Severity Index (PSI) and Simplified Acute Physiology Score (SAPS 3) are widely used scoring systems for risk stratification and clinical decision-making support in critically ill patients with pneumonia. However, these scoring systems have limitations that affect their ability in the prediction of SCAP mortality to serve as clinical decision-making support tools. SAPS 3 lacks sufficient evidence for use in ICU-admitted patients with pneumonia, with lower positive and negative predictive values for predicting ICU demand, which can underestimate severity in certain age groups. CURB-65 and PSI often provide severity scores that contradict death outcomes when predicting the efficacy of corticosteroid use.[10] APACHE-II and SOFA require more clinical and laboratory indicators, making them intricate and time-consuming to use,[11] also lacking clinical evidence for predicting mortality in Asian patients with SCAP. Overall, previous specialised evaluations of pneumonia and ICU severity scores have performed poorly in predicting mortality rates in critically ill patients with pneumonia.[12 13] Therefore, further research is needed to identify new risk factors and build models that enable us to differentiate patients at risk of death early and decide whether to provide preventive or aggressive treatment.

In light of the growing understanding and appreciation for the potential of machine learning (ML) and artificial intelligence in enhancing the prognostic capabilities related to infectious events such as sepsis, antibiotic resistance and pandemic coronavirus disease 2019 (COVID-19), scientists have embarked on additional investigations into respiratory diseases. Specifically, comprehensive analyses and interpretations of multiple modalities, such as chest imaging, lung pathology and physiological data, have been undertaken, aiming to deepen our understanding of these intricate conditions within the respiratory system.[14 15] Recent studies have shown that ML outperforms existing single-purpose scores which are based on common regression methods in predicting 30-day mortality in CAP.[16] Whereas there is limited research evaluating the effectiveness of ML methods in predicting all-cause mortality during hospitalisation for severe pneumonia.

The purpose of this study was to explore the risk factors and to construct prognostic models for fatal outcomes in severe pneumonia. With a focus on mainstream ML algorithms, that is, light gradient boosting machine (LightGBM), support vector classifier (SVC) and random forest (RF), we sought to construct a rapid risk-stratification and mortality-prediction model for patients with severe pneumonia based on their clinical and laboratory data. This ensemble model was subsequently compared with traditional scoring systems to prove the superiority of ML analysis. Based on the variables sorted out by the ML model, we further identified independent risk factors for severe pneumonia mortality by logistic regression, providing important guidance for research into the disease progression. The implementation of an efficient and robust rapid risk-stratification system for severe pneumonia, using ML algorithms, is anticipated to improve clinical tools for optimising healthcare resource distribution and timely intervention. This progress could be critical for securing an early diagnosis and making prompt treatments, which, as a result, may enhance patient prognosis and alleviate the consequential strain on healthcare infrastructures.

METHODS

Patient selection with inclusion and exclusion criteria

This is a retrospective, single-centre observational study that enrolled patients with severe pneumonia between July 2013 and March 2022 admitted to Beijing Chao-Yang Hospital, where the Beijing Institute of Respiratory Medicine is located. According to the consensus guidelines from the Infectious Diseases Society of America and the American Thoracic Society,[17] patients with severe pneumonia were included if they met one major criterion or three minor criteria; among them, major criteria included (1) requiring mechanical ventilation via tracheal intubation and (2) requiring vasopressor therapy despite adequate fluid resuscitation for septic shock, while minor criteria included (1) respiratory rate ≥30 breaths/min,

(2) ratio of arterial oxygen pressure to fractional inspired oxygen (PaO2/FiO2)≤250, (3) multi-lobar infiltrates, (4) confusion or disorientation, (5) blood urea nitrogen ≥7.14 mmol/L and (6) systolic blood pressure <90 mm Hg requiring aggressive fluid resuscitation. Patients were excluded if they were any (1) <18 years of age, (2) underwent cardiopulmonary resuscitation within 24 hours before admission, (3) had a hospital stay of <24 hours, (4) had a prolonged stay of >90 days due to multiple hospital admissions, (5) had more than 20% of their variables missing. This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis reporting guidelines,[18] which aim to improve the accuracy and transparency of reporting for prediction models in medicine.

Clinical data collection

We extracted clinical data from the electronic medical records system, including medical history, demographic data, symptoms and vital signs, comorbidities, laboratory test results within 24 hours of admission, clinical management and clinical outcomes. The ICU severity scores (APACHE-II, SOFA, etc) were calculated based on the related parameters. Symptoms of interest included fever, cough, sputum production, haemoptysis, chest pain, shortness of breath and respiratory distress. Vital signs included body temperature, mean blood pressure, heart rate and respiratory rate. Here, laboratory test results included complete blood count, coagulation test, liver and kidney function, electrolytes, cardiac markers, N-terminal pro-B-type natriuretic peptide, arterial oxygen pressure, PaO2/FiO2, procalcitonin and lactic acid, etc. Clinical management included the need for invasive mechanical ventilation, the occurrence of acute respiratory distress syndrome and the requirement for vasopressor drugs. The primary clinical outcome was in-hospital mortality. We compared the clinical features of the survival group and non-survival group at admission and developed an ML model to predict the in-hospital mortality in severe pneumonia. All data were cross-checked by at least two clinical physicians or researchers.

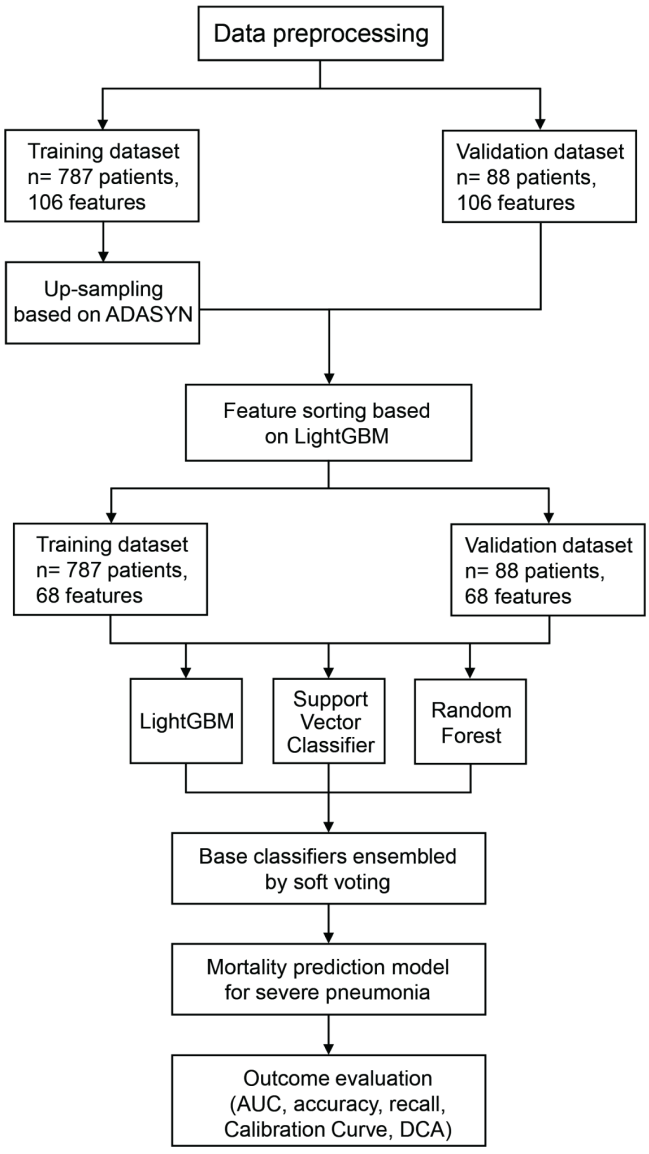

Figure 1 The workflow of the model construction. ADASYN, adaptive synthetic sampling; AUC, area under receiver operating characteristic curve; DCA, decision curve analysis; LightGBM, light gradient boosting machine.

Model building

This study included 106 clinical features as candidate features for analysis, including age, sex, smoking history, alcohol consumption history, hypertension, type 2 diabetes, chronic heart disease, chronic kidney disease, fever, cough, chest pain, vital signs and laboratory tests. All these clinical features were easily obtainable from clinical work. The basic process of model construction is illustrated in figure 1. First, samples with all the features were collected and normalised. For features that contained missing values and were non-continuous or categorical, we chose to fill them with −1. For missing values in continuous features, we set them to 0.5, while non-missing values were subjected to min-max normalisation. Patients were randomly divided into a training set and a validation set in a ratio of 9:1. To address the potential issue of an imbalanced classification, adaptive synthetic sampling (ADASYN) was applied for model training by oversampling the training set. In this study, feature sorting was conducted using LightGBM to identify the optimal subset of features for variable screening. Random subsets of features were selected from all features for model fitting, and the model's performance was evaluated using accuracy (ACC), recall and area under the receiver operator characteristic curve (AUC). This process was repeated for all possible feature combinations until the best feature subset was obtained. Based on the selected optimal feature subset and the training and validation set data, an ensemble learning model was constructed using three basic classifiers: RF, SVC and LightGBM. The performance of this model was evaluated

Table 1 Optimisation parameter-adjustment process of machine learning

Results of different processing methods	ACC	Recall	AUC
Raw data set	0.8295	0.3077	0.7713
Raw data set+normalisation	0.8295	0.3846	0.8379
Raw data set+normalisation+ADASYN upsampling	0.8750	0.7692	0.8246
Raw data set+normalisation+RandomOverSampler upsampling	0.7841	0.3846	0.8144
Raw data set+normalisation+upsampling+ensemble learning	0.8750	0.5384	0.8758
Raw data set+normalisation+upsampling+ensemble learning+reducing feature number	0.8977	0.8462	0.8779

ACC, accuracy; ADASYN, adaptive synthetic sampling; AUC, area under receiver operating characteristic curve.

from multiple dimensions including discriminative ability, calibration goodness and clinical utility. The optimisation process of the ML model is shown in table 1, and the parameter settings for the ensemble learning algorithm can be found in online supplemental table S1. The ML algorithms were operated by Scikit-learn V.1.1.3 in Python V.3.8.8.

Validation and evaluation of the prognostic model

We conducted internal validation of the model, which was then compared with conventional clinical scoring systems, that is, APACHE-II, SOFA, CURB-65 and PSI. The discriminative ability of the model was evaluated by AUC and the concordance index (C-index). A calibration curve was employed to assess the goodness of fit from the consistency between the predicted values from the model and the actual values. Decision curve analysis (DCA) was then performed to evaluate the clinical utility from the net benefits of this prognostic model.

Visualisation by nomogram

The top 20 variables with the most significant influence on the clinical outcomes, as determined by ML feature sorting and SHapley Additive exPlanations (SHAP) analysis, were included in a multivariable logistic regression analysis to identify independent risk factors. Based on the results of the multivariable logistic regression analysis, a predictive model was constructed using the 'rms' R package and it was visualised as an alignment diagram known as a nomogram. The nomogram displayed the prediction lines, which allowed for the identification of points received from each factor. The sum of these points was located on the 'Total Points' axis. Finally, a line was drawn vertically to project onto the underlying probability scale, thereby determining the risk factors and prediction of mortality in severe pneumonia.

Statistical analysis

The sample size was determined based on the available data, which is greater than the sample size estimation by software PASS V.15 software (NCSS, Kaysville, Utah, USA). Baseline features were compared between the survival and non-survival groups. Age, respiratory rate, heart rate, white blood cell count, neutrophil count and haemoglobin value were considered as numerical variables, while the requirement for vasopressors, requirement for invasive mechanical ventilation, immunocompromise status, cough, fever, pleural effusion, diabetes and hypertension were listed as categorical variables. Candidate variables with missing data of >20% were excluded.

Descriptive statistics were presented as median (IQR) values or mean with SD values for numerical variables and as frequency and percentage (%) for categorical variables. The distribution normality of numerical variables was assessed using the Shapiro-Wilk test. If the data followed a normal distribution, an independent two-sample t-test was used; otherwise, the Mann-Whitney U test was employed. For categorical variables, the χ^2 test or Fisher's exact test was used. Differences between the survival and non-survival groups were evaluated by the relevant statistical analysis, and $p<0.05$ was considered statistically significant, conducted by GraphPad Prism V.9.0 (Dotmatics, San Diego, California, USA).

Patient and public involvement

Patients were not involved in the planning and conduct of this research.

RESULTS

Patient selection

In our study, we extracted a cohort of 1041 patients with severe pneumonia from electronic medical records on admission. However, 47 patients were excluded from the analysis due to being <18 years of age. Among the remaining 994 patients, we further excluded 82 cases with repeated hospitalisations for medical billing purposes. Additionally, 27 patients who had a hospital stay of <24 hours or >90 days were excluded, and 10 patients with a significant proportion of missing variables exceeding 20% were similarly excluded. Ultimately, our analysis included a total of 875 patients, as depicted in figure 2.

Clinical features

Clinical features at admission for the survival and non-survival cohorts are presented in online supplemental table S2. Notably, no significant correlation was observed

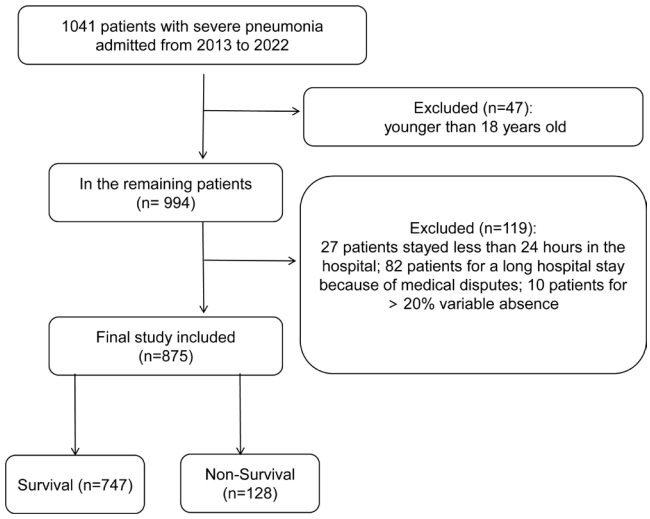

Figure 2 Flowchart describing the patient-selection procedure.

in sex between survival and non-survival (32.1% vs 28.1%, p=0.368). The median age of patients with severe pneumonia was 62 (IQR=47–72) years, with a median age of 62.5 (IQR=51–72) years for those who did not survive and a median age of 62 (IQR=47–72) years for those who did (p=0.422). The top three predominant symptoms in both groups were fever, cough and respiratory distress. Fever (63.2% vs 49.2%) and cough (54.2% vs 37.5%) were more commonly observed in the survival group compared with the non-survival group (p<0.05) with significance, while respiratory distress (26.8% vs 11.8%) was more prevalent in the non-survival group, although the difference was not statistically significant (p=0.067).

Regarding comorbidities, the three most common conditions in the survival group were hypertension (41.5%), immunosuppression (37.1%) and diabetes (22.8%). In the non-survival group, the leading comorbidities differed slightly from the survival group, with immunosuppression (53.1%), hypertension (32%) and chronic heart disease (17.1%) being the most prevalent. Notably, the patients with immunosuppression of the non-survival (53.1%) accounted for a larger percentage than those with survival (37.1%) significantly (p<0.001).

Univariate analysis results revealed no significant differences between the survival and non-survival groups in terms of vital signs, including body temperature, respiratory rate, heart rate and mean arterial pressure. In addition, this study also investigated clinical management, including the requirement for endotracheal intubation (invasive mechanical ventilation, IMV) and vasopressors, the duration of IMV and the incidence of acute respiratory distress syndrome (ARDS). Our findings showed that patients in the non-survival group included a greater proportion of those receiving IMV (66.4% vs 36.8%, p<0.001) and a greater requirement for vasopressor medications (71.9% vs 32.1%, p<0.001). Nevertheless, the duration time of IMV had no significant effect on mortality (p=0.185). Furthermore, although the non-survival group exhibited a slightly greater incidence of ARDS, the difference was not statistically significant (22.7% vs 18.7%, p=0.3).

The laboratory test results demonstrated that, compared with the survival group, the non-survival group exhibited significantly elevated white blood cell counts (10.08 vs 8.04×10^9/L, p=0.03), neutrophil counts (8.42 vs 7.42×10^9/L, p=0.01), neutrophil-to-lymphocyte ratios (11.24 vs 8.79, p=0.02), creatinine levels (72.30 vs 64.20 µmol/L, p=0.04), blood urea nitrogen levels (7.56 vs 6.72 µmol/L, p<0.001), serum lactate dehydrogenase levels (424 vs 328 U/L, p=0.002), blood potassium

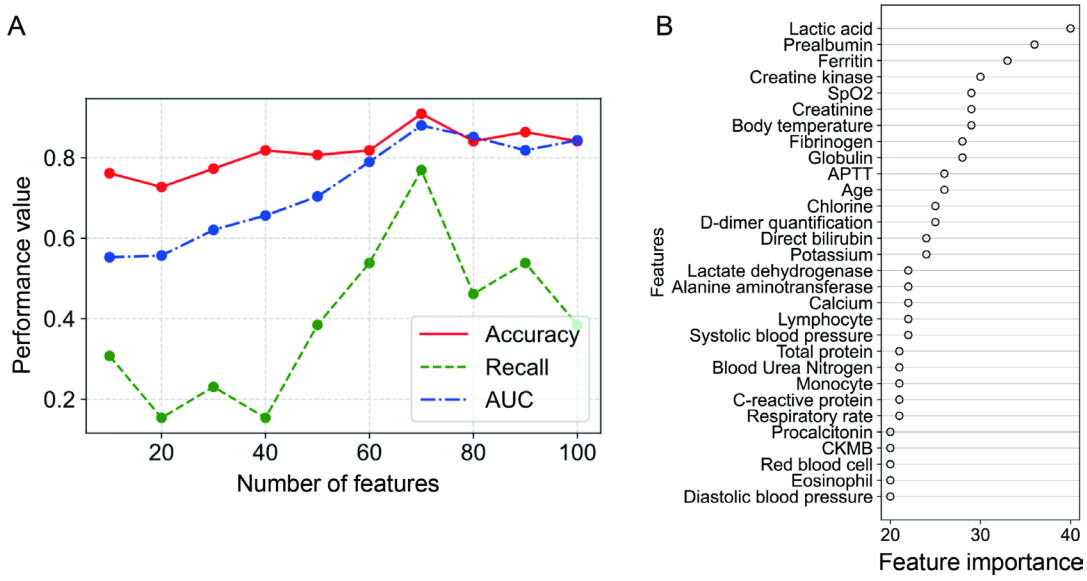

Figure 3 Candidate predictors screening by light gradient boosting machine. APTT, activated partial thromboplastin clotting time; AUC, area under receiver operating characteristic curve; CKMB, creatine kinase myocardial band; SpO2, percutaneous oxygen saturation.

levels (4.14 vs 3.9 mmol/L, p=0.03), blood sodium levels (138.35 vs 137.7 mmol/L, p=0.03), prothrombin times (13.1 vs 12.8 s, p=0.048), D-dimer values (3.4 vs 2.71 mg/L, p=0.04), lactic acid levels (1.7 vs 1.3 mmol/L, p<0.001) and FiO2 (0.5 vs 0.4, p=0.002). The non-survival group also had significantly lower values of percutaneous oxygen saturation (94% vs 95%, p=0.018) and oxygenation indices (PaO2/FiO2) (153.35 vs 200 mm Hg, p<0.001) compared with the survival group. However, there were no statistically significant differences observed between the two groups in terms of infection-related indicators or other biochemical laboratory test results at admission.

Candidate predictor selection

On the basis of 875 patients being in the training set and validation sets, 68 potential predictors were sorted out from 106 features by LightGBM. The 68 characteristic variable combinations showed the highest value for recall with relatively high values for ACC and AUC (figure 3A). As the number of variables exceeded 68, recall experienced a significant decrease. Recall here means non-survivors would be predicted correctly in advance, which makes the most difference to clinical application, so the number of 68 variables were selected. The importance of each feature was quantified by the number of times a feature was used to split between the model, and a higher value of feature importance was associated with a greater contribution to the mortality prediction of the model. As a result, the top 30 feature variables were ranked in figure 3B. Notably, this process of feature screening could facilitate model building by employing ensemble learning algorithms in the next step.

Next, we constructed an ensemble model, composed of the RF, SVC and LightGBM classifiers, to rank the 68 feature variables by SHAP analysis. From this analysis, the top 20 feature variables were sorted as shown in figure 4, in order to exhibit the relationship between the SHAP value of the feature and its exact impact. The absolute SHAP value provides relative importance in the training data set, and the SHAP summary plots reflect the range or distribution of the feature's impact on the output. Among these variables, the partial SHAP dependence plots for the top six variables (cough, hypertension, mechanical ventilation time, ferritin, diabetes, lactic acid) with the mean absolute SHAP values are illustrated in figure 5 to increase the interpretability of the model. The regression curves show the varying contribution of numerical variables at different ranges to the model. The prioritisation of these important variables not only revealed the biomarkers for mortality but also led to the optimal predictive accuracy of the ensemble model.

Evaluation of the model performance

To evaluate the discriminatory ability of the ensemble model, we calculated the AUC in both the combination of training and validation namely all set (0.9922; 95% CI, 0.981 to 0.999) (figure 6A), and the validation set (0.8779; 95% CI, 0.738 to 0.974) (figure 6B). In terms of predictive accuracy, our model is superior to other

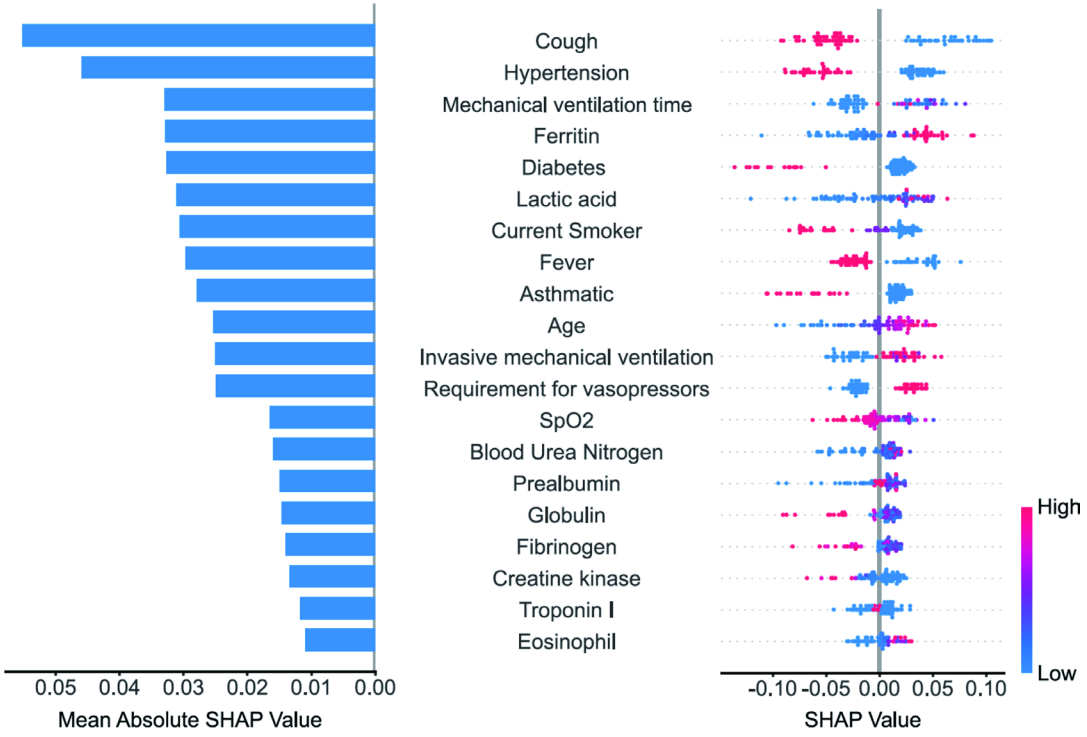

Figure 4 Importance of the selected variables by ensemble model of machine learning. SHAP, SHapley Additive exPlanations; SpO2, percutaneous oxygen saturation.

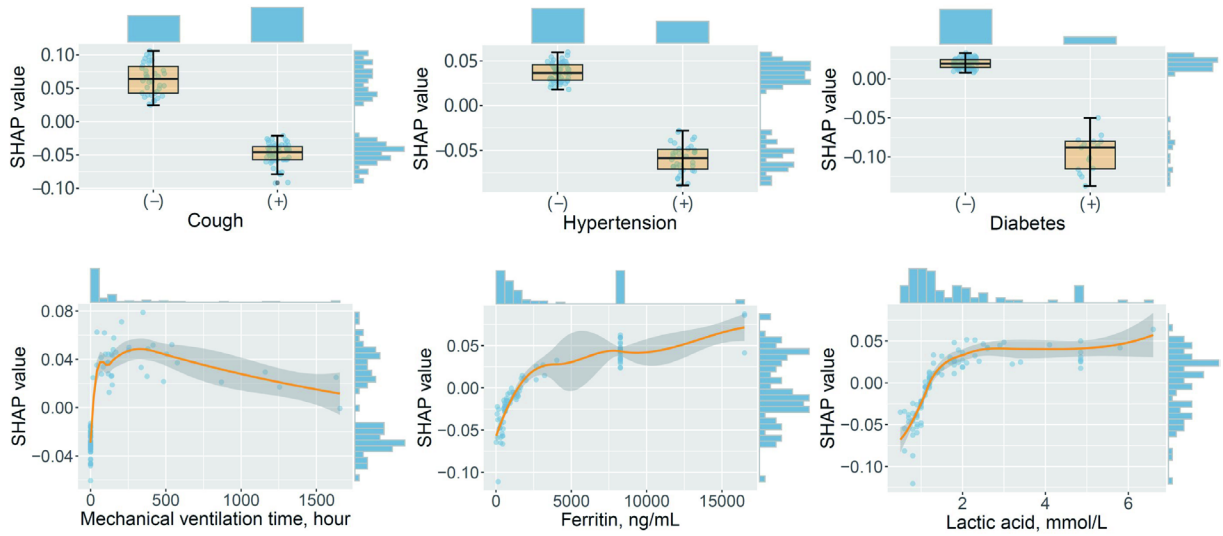

Figure 5 The partial SHAP dependence plots for the top six variables with the mean absolute SHAP values. SHAP, SHapley Additive exPlanations.

scoring systems (APACHE-II, SOFA, CURB-65, PSI), in the aspect of AUC, optimal specificity and recall. To evaluate the goodness of fit of the model, we applied calibration curve analysis by plotting the fitted values against the actual average values. Calibration plots demonstrated that the model has a good agreement between the predicted probability and observed incidence of death (figure 7A). In addition, to assess the clinical utility of the model, DCA was employed to provide further insights (figure 7B). The clinical net benefit exhibits a positive trend across all threshold probabilities and surpasses that of the two extreme strategies, which means the value of clinical application presented by the ensemble model.

Comparison with basic machine learning algorithms

The results show that the ensemble model proposed in this study outperforms the basic algorithms in terms of ACC, recall and AUC metrics, as depicted in table 2. Specifically, our model achieved impressive scores of 0.8977, 0.8462 and 0.8779 for ACC, recall and AUC in the validation data set, respectively. In contrast, the basic ML algorithms that is, MLPClassifier, DecisionTreeClassifier and KNeighborsClassifier exhibited relatively mediocre performance across these metrics: MLPClassifier had lower recall; KNeighborsClassifier showed lower ACC and AUC; DecisionTreeClassifier was

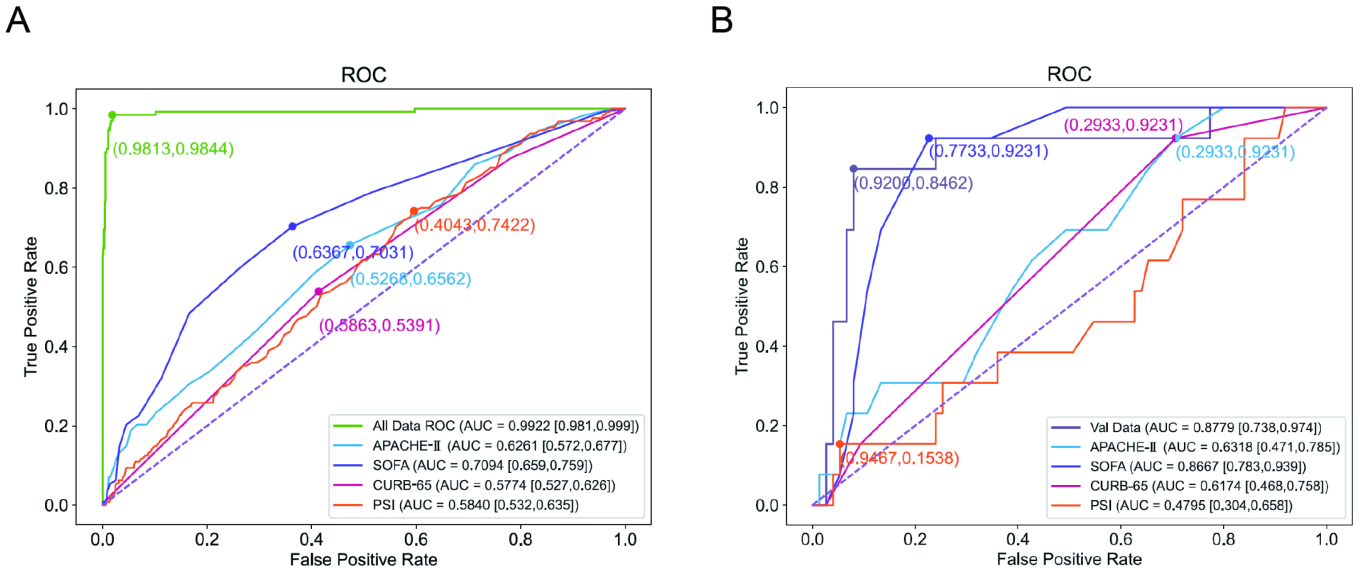

Figure 6 Evaluation of the ensemble model discriminative ability. APACHE-II, Acute Physiology and Chronic Health Evaluation; AUC, area under receiver operating characteristic curve; CURB-65, confusion, urea, respiratory rate, blood pressure, age ≥65 years; PSI, Pneumonia Severity Index; SOFA, Sequential Organ Failure Assessment; ROC, receiver operator characteristic curve.

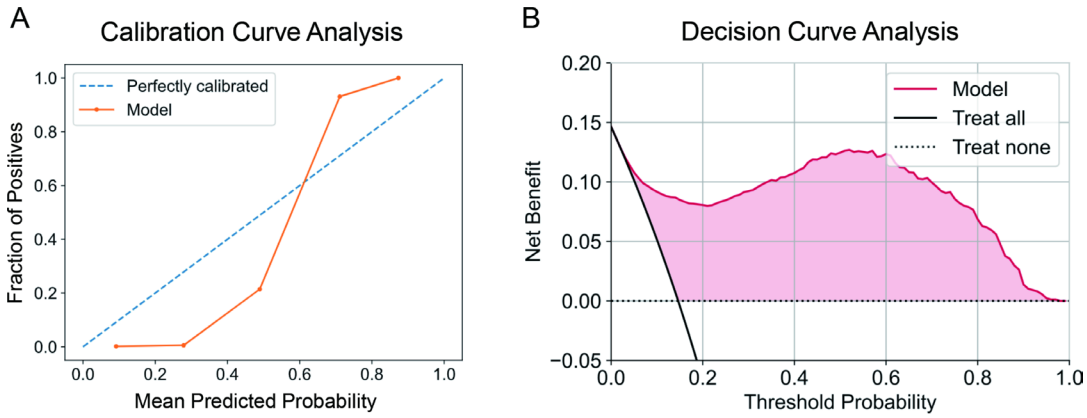

Figure 7 Calibration curve and decision curve analysis.

inferior to the ensemble model in all evaluation dimensions.

Based on these findings, our ensemble model for predicting all-cause mortality in hospital for severe pneumonia demonstrates superior performance in predictive accuracy, effectively detecting positive incidences. The ensemble model was established as a consistently reliable and efficacious solution that adeptly confronts the task of predicting all-cause mortality in hospital associated with severe pneumonia.

Identification of independent risk factor

In order to identify the independent risk factor for fatal pneumonia, multivariate logistic analysis was applied in forward conditional stepwise regression to filter from the top 20 variables figured out by SHAP analysis before. As a result, eight risk factors were identified to be independently associated with in-hospital mortality through logistic regression, including cough, ferritin, diabetes, lactic acid, the requirement for vasopressors, blood urea nitrogen, creatine kinase and eosinophil, reported with OR (95% CI) and p values in table 3.

To visualise the results based on the logistic regression model, a nomogram was generated to predict the risk of death in a simple way (figure 8A). In the nomogram, each predictive factor was assigned a weighted score, displayed at the top of the column chart. The total score for each sample was the sum of individual scores. At the bottom of the column chart, the probability of in-hospital mortality was calculated based on the total score. According to the column chart, the patients with severe pneumonia who exhibit higher levels of ferritin, lactic acid, blood urea nitrogen and eosinophil, along with lower levels of cough and creatine kinase, acquired higher total scores, indicating a poorer prognosis of mortality. Although diabetes here seemed like a protective factor contradicting the previous studies, it is possibly confounding bias from a limited sample size of non-survival. Notably, lactic acid and blood urea nitrogen showed more pronounced elevations in the non-survival group with statistical significance, as depicted in figure 8B. Furthermore, ferritin, eosinophil and the requirement of vasopressors were also considered to be potential independent risk factors for fatal pneumonia after a rigorous process of successive selection.

Table 2 Performance comparison of ensemble learning and basic algorithms in validation data set

Algorithms	ACC	Recall	AUC
MLPClassifier	0.8409	0.5385	0.8656
DecisionTreeClassifier	0.75	0.3846	0.599
KNeighborsClassifier	0.6477	0.8462	0.7482
Ensemble learning	0.8977	0.8462	0.8779

ACC, accuracy; AUC, area under receiver operating characteristic curve.

DISCUSSION

In this study, we employed ML techniques to develop and validate a clinical prediction model for in-hospital mortality of patients with severe pneumonia. Our research yielded two major findings: first, the ensemble model integrated by ML basic algorithms demonstrated potential in predicting all-cause mortality for severe

Table 3 Risk factors selected by multivariate logistic analysis of all data set

Risk factors	OR	95% CI for OR	P value
Cough	0.580	(0.383 to 0.879)	0.010
Ferritin	3.274	(1.400 to 7.657)	0.006
Diabetes	0.482	(0.275 to 0.845)	0.011
Lactic acid	9.430	(2.239 to 39.723)	0.002
Requirement for vasopressors	5.282	(3.427 to 8.142)	< 0.001
Blood urea nitrogen	11.078	(1.710 to 71.789)	0.012
Creatine kinase	0.003	(0.000 to 0.301)	0.013
Eosinophil	7.840	(1.022 to 60.166)	0.048

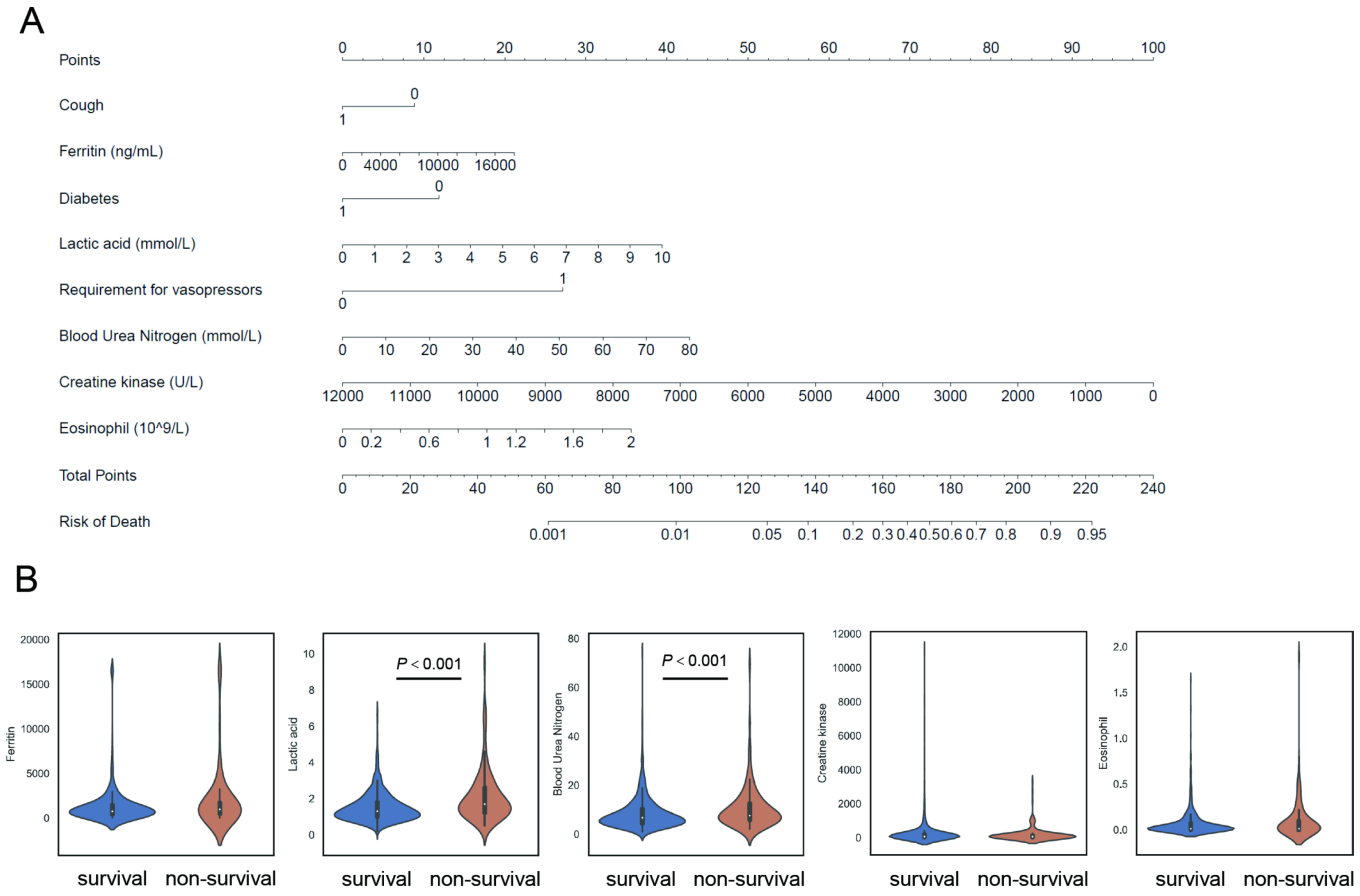

Figure 8 Risk factors for mortality of severe pneumonia.

pneumonia, superior to any single ML algorithm or conventional scoring systems; second, risk factors for fatal pneumonia were identified by means of the ML ensemble model and logistic regression analysis with forward conditional selection, possibly serving as pivotal indicators in clinical decision-making.

In clinical practice, several traditional tools, such as PSI, the CURB-65 score and APACHE-II, have been approved for assessing the severity of CAP and for predicting mortality. Among the widely used severity assessment tools for CAP, the PSI and CURB-65 scores have shown good performance in predicting 30-day mortality but limited utility in identifying the need for ICU admission in severe CAP.[19] Previous studies have demonstrated the limitations of the PSI score in identifying severe patients with CAP.[20] ML has been extensively used for risk stratification and predicting disease outcomes or mortality in prognostic tools.[21] However, only a few studies to date have reported the application of ML in specific pneumonia-related research.[21 22] For instance, Xu *et al* investigated the effectiveness of ML models in predicting adverse outcomes in CAP. Their study included 2302 patients with pneumonia, and the data were divided into training and testing sets using a 7:3 ratio. The results demonstrated the feasibility and effectiveness of ML algorithms in predicting adverse outcomes in patients with CAP.[23] Another study used ML techniques to develop a simple yet accurate model for predicting mortality in CAP based on readily available and common clinical variables.[24] Additionally, there have been recent studies performed using ML for predicting the occurrence of severe pneumonia in kidney transplant recipients.[25] Ensemble ML has been demonstrated to reduce generalisation errors and improve model performance in multiple clinical prediction scenarios.[26–29] Nevertheless, the existing researches examining the utilisation of ML techniques to forecast in-hospital mortality aimed at severe pneumonia remains limited in scope.

In the previous study, logistic regression analysis was employed to develop a predictive model for in-hospital mortality in patients with SCAP admitted to the ICU. The study found that lymphocytes, PaO2/FiO2, shock and the APACHE-II score were independent risk factors for in-hospital mortality.[30] Moreover, Li *et al* evaluated predictive factors using multivariable logistic regression and constructed a prediction model for the in-hospital mortality of elderly patients with CAP admitted to the ICU.[31] ML methods were also applied to develop novel pneumonia scores to predict 1 year and in-hospital mortality in patients with pneumonia on ICU admission. But most of them extracted data from the public databases such as MIMIC-IV and eICU which involve the western ethnic populations. The data of our study are mainly derived from patients in North China which provides insightful comparison.[32] Jeon *et al* previously

developed and validated an ML model for mortality prediction using data from patients with severe pneumonia admitted to the ICU.[33] In contrast, our study enrolled a broader population of patients with severe pneumonia, including both ICU patients and patients with severe pneumonia in general wards.

To the best of our knowledge, this study represents a pioneering approach in employing an ensemble ML model that was developed specifically for hospitalised patients with severe pneumonia and subsequently validated using an internal cohort. The primary objective of this model was to accurately predict in-hospital mortality among patients with severe pneumonia. Our evaluation of the performance of the ML model yielded compelling results, demonstrating a high degree of predictive accuracy with an AUC and a C-index exceeding 0.85 superior to conventional scoring systems in both the training and validation data sets. Furthermore, the calibration analysis indicated an acceptable goodness of fit between the predicted in-hospital mortality derived from the ML model and the actual observed one. Compared with the commonly used basic ML algorithm, the ensemble algorithm applied in this study exhibited robust discriminatory and calibration capabilities, bolstering its efficacy and reliability. Through DCA, we found conclusive evidence of positive net clinical benefits in all threshold probability for patients with severe pneumonia admitted to the hospital, thereby reinforcing the practical value and clinical utility of this model.

We incorporated eight independent risk factors into the nomogram to predict the in-hospital mortality of patients with severe pneumonia, including cough, diabetes, requirement for vasopressors, ferritin, lactic acid, blood urea nitrogen, creatine kinase and eosinophil. These predictive indicators can be obtained in the shortest possible time. The nomogram has the potential to facilitate the identification of risk factors for mortality in severe pneumonia by healthcare professionals, aiding them in making rapid preliminary diagnoses. However we did not conduct an assessment of the nomogram predicting performance from the logistic regression model, since the primary purpose of this section was to identify the independent risk factor on the foundation of ML results. According to clinical expertise, as a common symptom of respiratory diseases, cough is not a specific indicator of the severity of pneumonia.[34] Considering that the average age of the cases included in this study was 62 years old, the elderly usually developed severe pneumonia with insidious onset and often atypical symptoms.[35] Diabetes as a prevalent comorbidity has been confirmed as a significant risk factor for increased mortality in severe pneumonia.[36 37] The contradiction may come from the confounding bias of the limited sample size of non-survival. Creatine kinase, as an important component of myocardial enzyme spectrum, usually significantly increases during cardiac injury. It is reported that heart damage is one of the fatal complications in patients with COVID-19 pneumonia. The elevation of creatine kinase was observed in 13.7% of patients and in 19% of severe patients.[38 39] Although we did not observe thsimilar difference in our data set because of targeted medical intervention or clinical management, creatine kinase was considered as an undeniable risk factor for mortality among patients with severe pneumonia.

Also, consensus between several of the risk factors in the nomogram and those reported in the previous studies has been observed. The requirement for vasopressors was considered an important factor with increased mortality in SCAP, which could be attributed to the severity of the underlying infection leading to a systemic inflammatory response, resulting in haemodynamic instability and the need for vasopressor support. The elevated lactic acid level reflects increased anaerobic metabolism in the body during tissue hypoxia and reduced perfusion. Consistent with previous studies, the level of lactic acid was positively associated with prolonged hospitalisation and higher mortality in severe pneumonia.[40 41] Furthermore, serum ferritin was discovered serving as an indispensable predictive marker for the severity and prognosis. Typically associated with infectious diseases, the release of serum ferritin is triggered by inflammatory responses, tissue damage, hypoxia and stress.[42] Eosinophil was also figured out as a key risk factor for mortality in the nomogram. Eosinophils are circulating and tissue resident leucocytes that have potent pro-inflammatory effects in many diseases.[43] Eosinophils are able to act as immune regulatory cells and also directly against parasites and bacteria.[44] Unlike virus infection, for example, COVID-19, where increased eosinophil count was associated with better prognosis, including lower complication incidence and mortality, the nomogram demonstrated that increased eosinophil count in SCAP implied higher mortality. The underlying cause for this observation is that the patients included in our data set predominantly exhibited bacterial aetiology of severe pneumonia, without the presence of COVID-19.[45] Another identified risk factor blood urea nitrogen was also included as an indicator in the PSI and CURB-65 scoring system, meantime with a significant difference between the survival and non-survival group.

This study has certain limitations that should be acknowledged. First, our model analysis was based on single-centre, retrospective data, which may introduce some degree of bias. To facilitate further investigation, a prospective external validation cohort is needed for confirmation in order to enhance the generalisability and mitigate overfitting. Second, it may not be straightforward to apply predictive models with a large number of explanatory features in clinical situation. Thereby the implementation of highly digitised healthcare facilities would be essential for the practical application of this ensemble model. Third, although a specialised team for the diagnosis and treatment of severe pneumonia ensured the representativeness and credibility of the data, the results of cross-validation were not good enough to prove robustness of the ensemble model. A balanced data set with a larger scale or multimodality such as CT

images would be necessary to enhance the predictive performance of the model.

CONCLUSION

In this study, we successfully developed an ensemble ML model based on clinical features of severe pneumonia for early prediction of in-hospital mortality. The internal validation confirmed the effectiveness of our model. The overall performance of the ML model in predicting in-hospital mortality was significantly better than that of existing severity-specific scores for severe pneumonia. Additionally, we identified ferritin, lactic acid, blood urea nitrogen, creatine kinase, eosinophil and the requirement of vasopressors as predictive factors for overall in-hospital mortality in severe pneumonia. The visualisation of predictive factors by nomogram enables clinicians to identify and recognise the prognosis of severe pneumonia at an early stage in a simple and intuitive way. These risk indicators will also provide guiding significance for the mechanistic study of disease progression in severe pneumonia.

Importantly, our study demonstrates that the ML model outperforms other widely used severity scoring systems, such as the APACHE-II score, in predicting mortality among patients with severe pneumonia. By harnessing the power of ML techniques, we have exploited the potential to analyse vast amounts of data in a manner transcending the capabilities of conventional approaches. Furthermore, this approach facilitates the direct collection of data from electronic medical records by investigators, thereby offering unprecedented opportunities to advance research endeavours in this field.

Author affiliations
[1]Department of Respiratory and Critical Care Medicine, Capital Medical University, Beijing, China
[2]Department of Respiratory Medicine, the Ninth Medical Center, Chinese PLA General Hospital, Beijing, China
[3]Beijing Boai hospital, Department of Respiratory and Critical Care Medicine, Beijing, China
[4]Sinopharm Genomics Technology Co Ltd, Changzhou, Jiangsu, China
[5]Department of Clinical Epidemiology, Capital Medical University, Beijing, China
[6]Department of Medical Research, the Ninth Medical Center, Chinese PLA General Hospital, Beijing, China
[7]Medical Research Center, Beijing Institute of Respiratory Medicine and Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China

Contributors Z-hT, XH, WZ, NS, LG and XuL conceived, designed and supervised the study. WZ, XuL, ZA, YL, JL, DW and XiL collected, cleaned and analysed the data. Z-hT, XH and WZ wrote the draft of the manuscript and interpreted the findings. Z-hT is responsible for the overall content as the guarantor. All authors read and approved the final report.

Funding This work was supported by the National Natural Science Foundation of China (Grant No. 82200007), the Ministry of Science and Technology of the People's Republic of China (Grant No. 2021YFC0863600, 2023YFC0872500), the Capital's Funds for Health Improvement and Research (Grant No.2022-1-1061), the Reform and Development Program of Beijing Institute of Respiratory Medicine to XH (Ggyfz202421), Beijing Hospitals Authority Youth Programme to XH (code: QML20230302).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The research was approved by the Ethics Committee of Beijing Chao-yang Hospital (No. 2023–KE–439). This study was retrospective and approved by the human ethics committee and deemed of low risk and hence informed consent was waived.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The datasets produced and analysed in the present study can be obtained from the corresponding author upon a reasonable request. All machine learning techniques were executed using scikit-learn (https://scikit-learn.org/stable/index.html). The custom script developed for this research will be accessible at https://github.com/YHHAZ/severe-pneumonia-machine-learning.

ORCID iDs
Weichao Zhao http://orcid.org/0000-0001-8785-1702
Jiachen Li http://orcid.org/0000-0001-9054-1975

REFERENCES

1 Cillóniz C, Torres A, Niederman MS. Management of pneumonia in critically ill patients. *BMJ* 2021;375:e065871.
2 Jain S, Self WH, Wunderink RG, *et al*. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. *N Engl J Med* 2015;373:415–27.
3 Qu J, Zhang J, Chen Y, *et al*. Aetiology of severe community acquired pneumonia in adults identified by combined detection methods: a multi-centre prospective study in China. *Emerg Microbes Infect* 2022;11:556–66.
4 Laporte L, Hermetet C, Jouan Y, *et al*. Ten-year trends in intensive care admissions for respiratory infections in the elderly. *Ann Intensive Care* 2018;8:84.
5 Cillóniz C, Dominedò C, Garcia-Vidal C, *et al*. Community-acquired pneumonia as an emergency condition. *Curr Opin Crit Care* 2018;24:531–9.
6 Montull B, Menéndez R, Torres A, *et al*. Predictors of Severe Sepsis among Patients Hospitalized for Community-Acquired Pneumonia. *PLoS ONE* 2016;11:e0145929.
7 Cavallazzi R, Furmanek S, Arnold FW, *et al*. The Burden of Community-Acquired Pneumonia Requiring Admission to ICU in the United States. *Chest* 2020;158:1008–16.
8 Metlay JP, Waterer GW, Long AC, *et al*. Diagnosis and Treatment of Adults with Community-acquired Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am J Respir Crit Care Med* 2019;200:e45–67.
9 Woodhead M, Welch CA, Harrison DA, *et al*. Community-acquired pneumonia on the intensive care unit: secondary analysis of 17,869 cases in the ICNARC Case Mix Programme Database. *Crit Care* 2006;10 Suppl 2:S1.
10 Carmo TA, Ferreira IB, Menezes RC, *et al*. Derivation and Validation of a Novel Severity Scoring System for Pneumonia at Intensive Care Unit Admission. *Clin Infect Dis* 2021;72:942–9.
11 Wang X, Jiao J, Wei R, *et al*. A new method to predict hospital mortality in severe community acquired pneumonia. *Eur J Intern Med* 2017;40:56–63.
12 Vicco MH, Ferini F, Rodeles L, *et al*. In-hospital mortality risk factors in community acquired pneumonia: evaluation of immunocompetent

adult patients without comorbidities. *Rev Assoc Med Bras* 2015;61:144–9.
13 Reyes LF, Garcia-Gallo E, Pinedo J, *et al*. Scores to Predict Long-term Mortality in Patients With Severe Pneumonia Still Lacking. *Clin Infect Dis* 2021;72:e442–3.
14 Peiffer-Smadja N, Rawson TM, Ahmad R, *et al*. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin Microbiol Infect* 2020;26:584–95.
15 Patel D, Kher V, Desai B, *et al*. Machine learning based predictors for COVID-19 disease severity. *Sci Rep* 2021;11:4673.
16 Cilloniz C, Ward L, Mogensen ML, *et al*. Machine-Learning Model for Mortality Prediction in Patients With Community-Acquired Pneumonia: Development and Validation Study. *Chest* 2023;163:77–88.
17 Mandell LA, Wunderink RG, Anzueto A, *et al*. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis* 2007;44 Suppl 2:S27–72.
18 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55–63.
19 Yang L, He D, Huang D, *et al*. Development and Validation of Nomogram for Hospital Mortality in Immunocompromised Patients with Severe Pneumonia in Intensive Care Units: A Single-Center, Retrospective Cohort Study. *Int J Gen Med* 2022;15:451–63.
20 Chalmers JD, Mandal P, Singanayagam A, *et al*. Severity assessment tools to guide ICU admission in community-acquired pneumonia: systematic review and meta-analysis. *Intensive Care Med* 2011;37:1409–20.
21 Kang SY, Cha WC, Yoo J, *et al*. Predicting 30-day mortality of patients with pneumonia in an emergency department setting using machine-learning models. *Clin Exp Emerg Med* 2020;7:197–205.
22 Cooper GF, Aliferis CF, Ambrosino R, *et al*. An evaluation of machine-learning methods for predicting pneumonia mortality. *Artif Intell Med* 1997;9:107–38.
23 Xu Z, Guo K, Chu W, *et al*. Performance of Machine Learning Algorithms for Predicting Adverse Outcomes in Community-Acquired Pneumonia. *Front Bioeng Biotechnol* 2022;10:903426.
24 Feng D-Y, Ren Y, Zhou M, *et al*. Deep Learning-Based Available and Common Clinical-Related Feature Variables Robustly Predict Survival in Community-Acquired Pneumonia. *Risk Manag Healthc Policy* 2021;14:3701–9.
25 Luo Y, Tang Z, Hu X, *et al*. Machine learning for the prediction of severe pneumonia during posttransplant hospitalization in recipients of a deceased-donor kidney transplant. *Ann Transl Med* 2020;8:82.
26 Shokhirev MN, Johnson AA. An integrative machine-learning meta-analysis of high-throughput omics data identifies age-specific hallmarks of Alzheimer's disease. *Ageing Res Rev* 2022;81:101721.
27 Callender T, Imrie F, Cebere B, *et al*. Assessing eligibility for lung cancer screening using parsimonious ensemble machine learning models: A development and validation study. *PLoS Med* 2023;20:e1004287.
28 Yoo D, Divard G, Raynaud M, *et al*. A Machine Learning-Driven Virtual Biopsy System For Kidney Transplant Patients. *Nat Commun* 2024;15:554.
29 Narula S, Shameer K, Salem Omar AM, *et al*. Machine-Learning Algorithms to Automate Morphological and Functional Assessments in 2D Echocardiography. *J Am Coll Cardiol* 2016;68:2287–95.
30 Pan J, Bu W, Guo T, *et al*. Development and validation of an in-hospital mortality risk prediction model for patients with severe community-acquired pneumonia in the intensive care unit. *BMC Pulm Med* 2023;23:303.
31 Li N, Chu W. Development and validation of a survival prediction model in elder patients with community-acquired pneumonia: a MIMIC-population-based study. *BMC Pulm Med* 2023;23:23.
32 Wang B, Li Y, Tian Y, *et al*. Novel pneumonia score based on a machine learning model for predicting mortality in pneumonia patients on admission to the intensive care unit. *Respir Med* 2023;217:107363.
33 Jeon ET, Lee HJ, Park TY, *et al*. Machine learning-based prediction of in-ICU mortality in pneumonia patients. *Sci Rep* 2023;13:11527.
34 Argenziano MG, Bruce SL, Slater CL, *et al*. Characterization and clinical course of 1000 patients with coronavirus disease 2019 in New York: retrospective case series. *BMJ* 2020;369:m1996.
35 Liu Y-N, Zhang Y-F, Xu Q, *et al*. Infection and co-infection patterns of community-acquired pneumonia in patients of different ages in China from 2009 to 2020: a national surveillance study. *Lancet Microbe* 2023;4:e330–9.
36 Huang D, He D, Gong L, *et al*. Clinical characteristics and risk factors associated with mortality in patients with severe community-acquired pneumonia and type 2 diabetes mellitus. *Crit Care* 2021;25:419.
37 Edqvist J, Lundberg C, Andreasson K, *et al*. Severe COVID-19 Infection in Type 1 and Type 2 Diabetes During the First Three Waves in Sweden. *Diabetes Care* 2023;46:570–8.
38 Meyer EJ, Nenke MA, Rankin W, *et al*. Corticosteroid-Binding Globulin: A Review of Basic and Clinical Advances. *Horm Metab Res* 2016;48:359–71.
39 Lv GY, An L, Sun XD, *et al*. Pretreatment albumin to globulin ratio can serve as a prognostic marker in human cancers: a meta-analysis. *Clin Chim Acta* 2018;476:81–91.
40 Trzeciak S, Dellinger RP, Chansky ME, *et al*. Serum lactate as a predictor of mortality in patients with infection. *Intensive Care Med* 2007;33:970–7.
41 Lindholm MG, Hongisto M, Lassus J, *et al*. Serum Lactate and A Relative Change in Lactate as Predictors of Mortality in Patients With Cardiogenic Shock - Results from the Cardshock Study. *Shock* 2020;53:43–9.
42 Ghosh S, Baranwal AK, Bhatia P, *et al*. Suspecting Hyperferritinemic Sepsis in Iron-Deficient Population: Do We Need a Lower Plasma Ferritin Threshold?*. *Pediatr Crit Care Med* 2018;19:e367–73.
43 Lindsley AW, Schwartz JT, Rothenberg ME. Eosinophil responses during COVID-19 infections and coronavirus vaccination. *J Allergy Clin Immunol* 2020;146:1–7.
44 Rodrigo-Muñoz JM, Gil-Martínez M, Sastre B, *et al*. Emerging Evidence for Pleiotropism of Eosinophils. *Int J Mol Sci* 2021;22:7075.
45 Macchia I, La Sorsa V, Urbani F, *et al*. Eosinophils as potential biomarkers in respiratory viral infections. *Front Immunol* 2023;14:1170035.