

RESEARCH

Open Access



A clinical data-driven machine learning approach for predicting the effectiveness of piperacillin-tazobactam in treating lower respiratory tract infections

Yemeng Yang^{1†}, Kun Han^{2†}, Jiatao Li^{3†}, Tao Zhang⁴, Zhijing Zhu⁵, Ling Su¹, Zhaoyong Han¹, Chunyan Xu¹, Yi Lu^{4*}, Likun Pan^{2,6*} and Tao Yang^{4*}

Abstract

Background In hospitalized patients, inadequate antibiotic dosage leading to bacterial resistance and increased antimicrobial use intensity due to overexposure to antibiotics are common problems. In the present study, we constructed a machine learning model based on patients' clinical information to predict the clinical effectiveness of Piperacillin-tazobactam (TZP) (4:1) in treating bacterial lower respiratory tract infections (LRTIs), to assist clinicians in making better clinical decisions.

Methods We collected data from patients diagnosed with LRTIs or equivalent diagnoses admitted to the Department of Pulmonary and Critical Care Medicine at Shanghai Pudong Hospital, Shanghai, between January 1, 2021, and July 31, 2023. A total of 26 relevant clinical features were extracted from this cohort. Following data preprocessing, we trained four models: Logistic Regression, Random Forest, Support Vector Machine, and Gaussian Naïve Bayes. The dataset was split into training and test sets using a 7:3 ratio. The top-performing models, as determined by Receiver Operating Characteristic (ROC)-Area Under the Curve (AUC) on the independent test set, were subsequently ensemble. Ensemble model (EL) performance was evaluated using bootstrap resampling on the training set and ROC-AUC, recall, accuracy, precision, F1-score, and log loss on an independent test set. The optimal model was then deployed as a web application for clinical outcome prediction.

Results A total of 1,314 patients primarily treated with TZP as initial empiric antibiotic therapy were enrolled in the analysis. The success group comprised 995 patients (75.7%), while the failure group consisted of 319 patients

[†]Yemeng Yang, Kun Han and Jiatao Li contributed equally to this work and shared the first authorship.

*Correspondence:

Yi Lu

liulixiyang@163.com

Likun Pan

lkpan@phy.ecnu.edu.cn

Tao Yang

yangtao12@fudan.edu.cn

Full list of author information is available at the end of the article



(24.3%). We constructed an ensemble learning model based on the Logistic Regression, Support Vector Machine and Random Forest models, which showed better overall performance. The EL model demonstrated robust performance on an independent test set, exhibiting a ROC-AUC of 0.69, a recall of 0.69, an accuracy of 0.64, a precision of 0.40, a F1-score of 0.50, and a log loss of 0.66. A corresponding web application was then developed and made available at <http://106.12.146.54:1020/>.

Conclusions In this study, we successfully developed and validated an EL model that effectively predicts the clinical effectiveness of TZP (4:1) in treating bacterial LRTIs. The model achieved a balanced performance across key evaluation metrics, demonstrating the model's potential utility in clinical decision-making. The web-based application makes this model readily accessible to clinicians, potentially helping optimize antibiotic dosing decisions and reduce both inadequate treatment and overexposure. While promising, future studies with larger datasets and prospective validation are needed to further improve the model's performance and validate its clinical utility. This work represents a step forward in using machine learning to support antimicrobial stewardship and personalized antibiotic therapy.

Keywords Machine learning, Piperacillin-tazobactam, Lower respiratory tract infections, Clinical data, Web app

Introduction

Lower respiratory tract infections (LRTIs), clinically diagnosed as pneumonia or bronchiolitis in the Global Burden of Diseases (GBD) Study, stand as a significant global cause of mortality [1]. In 2019, around 2.49 million deaths were attributed to LRTIs globally, positioning it as the fourth foremost cause of death across all age groups [2]. It has been proved that LRTIs are also the main cause of death from infectious diseases worldwide [3]. In China, the major pathogens responsible for LRTIs are bacteria including *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Acinetobacter baumannii*, *Klebsiella pneumoniae*, *Escherichia coli*, and *Pseudomonas aeruginosa*, etc [4–6].

Due to the limitations of traditional pathogen detection methods, which cannot meet clinical needs and lack sufficient head-to-head clinical trial support, clinical treatment of LRTIs is mostly empirical [7]. In the empirical treatment of bacterial infections, the combination of penicillin and β -lactamase inhibitor is one of the preferred antimicrobial treatments for respiratory tract infections requiring hospitalization but not intensive care unit (ICU) admission [5, 6]. Piperacillin-tazobactam (TZP) is a powerful antibiotic combining a β -lactam antibiotic (piperacillin) with a β -lactamase inhibitor (tazobactam) and has been extensively used in the empirical treatment of hospitalized patients with LRTIs due to its broad spectrum of antibacterial activity, higher safety and efficacy, effectively targeting a majority of the bacteria associated with these infections including gram-positive, anaerobic, and gram-negative types [8, 9]. Notably, it remains effective against many multidrug-resistant strains of *Pseudomonas aeruginosa* and Enterobacteriaceae species [10, 11]. However, TZP still has a certain failure rate in clinical practice, and some patients have to upgrade to higher-level antibiotics, such as carbapenems, tigecycline. Therefore, if the effective response of TZP can be predicted before empirical treatment of LRTIs, it will be of great help to clinicians in clinical decision-making,

and patients who do not respond to TZP treatment can be identified, and early intensive treatment can be carried out for these patients.

Previous studies have shown that the treatment failure is undoubtedly associated with factors such as bacterial spectrum beyond that of TZP, drug-resistant strains, and undetected non-bacterial pathogens (such as mycoplasma pneumoniae, viruses) [12] and limited dosage of piperacillin due to the fixed ratio with tazobactam (8:1 or 4:1) (4:1 or 8:1 refers to the fixed ratio of piperacillin to tazobactam) [13]. Additionally, other factors such as gender, age, weight, concurrent medication information, and infection related characteristics of patients can also affect the treatment effectiveness of LRTIs [14–17]. However, on the one hand, due to the large amount of existing patient data, it is difficult to clarify effective relationships between them. On the other hand, the accuracy of these factors as predictive indicators for TZP is not yet high, and they cannot provide clinicians with enough information to accurately identify non-responders to TZP. Therefore, it is necessary to find effective strategies to predict the efficacy of TZP so as to help clinicians make clinical decisions.

At present, Artificial Intelligence (AI) has been widely used in medical image analysis, auxiliary diagnosis, clinical trials and other fields, which has significantly improved the quality and efficiency of medical care [18, 19]. Machine Learning (ML), as one of the typical representatives of data-driven AI methods, can process extensive, complex, and heterogeneous datasets, revealing subtle patterns and associations that may be overlooked by conventional statistical approaches, which facilitates the identification of non-linear relationships and interactions among variables, leading to more detailed and complex insights [20, 21]. As modern hospital information systems gain popularity and computing power continues to improve, the integration of ML and medicine has become increasingly seamless [22].

For instance, based on clinical data, an antibiotic combination recommendation model was constructed to predict antibiotic combination recommendations by Qin et al. [23]. Kim et al. developed prediction models to estimate the likelihood of antibiotics being ineffective against urinary tract infections suspected to be hospital-acquired [24]. Wang et al. built ML models to provide an automated identification for patients at risk of chronic obstructive pulmonary disease [25]. However, within the field of anti-infectives, the researches mainly focus on predicting antibiotic resistance, discovering new drugs, and drug design [26–28]. There is still a lack of relevant research on the development of ML algorithms specifically for predicting the effectiveness of antimicrobial drugs based on the clinical characteristics of patients with bacterial LRTIs, especially for TZP that are widely used in clinical practice.

In this study, a web-based ensemble learning (EL) model, integrating Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF), was developed to predict the clinical effectiveness of TZP (4:1) as the primary therapeutic strategy for treating bacterial LRTIs using patient clinical data. The model demonstrated balanced performance across key evaluation metrics. This tool has the potential to aid clinicians

in optimizing antibiotic prescribing and mitigating the development of antibiotic resistance.

Methods

Study design and setting

The entire experimental procedure is shown in Fig. 1. Our prediction framework was based on four common ML models: LR, RF, SVM, and Gaussian Naive Bayes (GNB). To investigate the feature importance, we employed the SHapley Additive exPlanations (SHAP) algorithm to conduct a global analysis after data imputation and pre-processing [29, 30]. SelectKBest and RFE were utilized to assess the sensitivity of different models to varying feature numbers. Subsequently, a comprehensive evaluation was conducted using grid search, cross-validation, and an independent test set, with performance metrics including Receiver Operating Characteristic (ROC)-Area Under the Curve (AUC), recall, precision, accuracy, F1-score, and log loss. The final model's input features and hyperparameters were determined through iterative optimization.

Study participants

The inclusion criteria for the study were as follows: patients were admitted to the Department of Pulmonary and Critical Care Medicine at Shanghai Pudong Hospital,

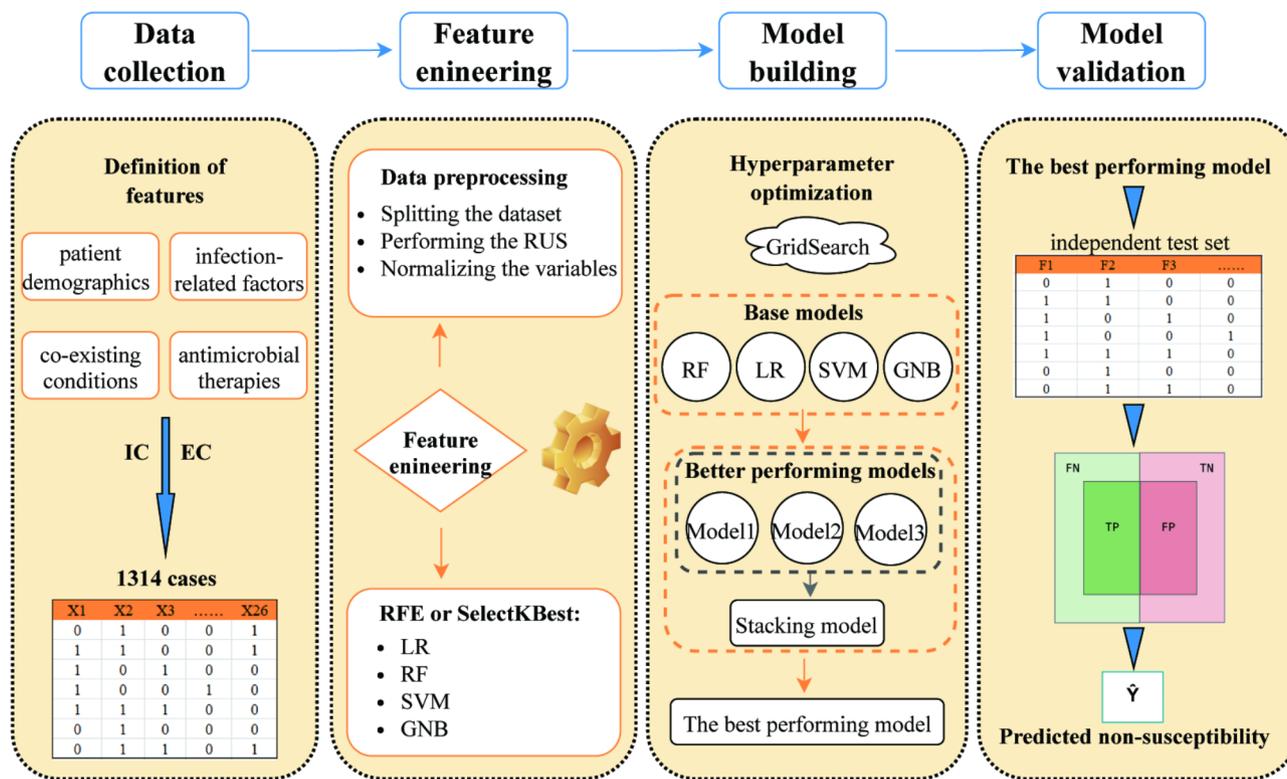


Fig. 1 Workflow of the research. IC, inclusion criteria; EC, exclusion criteria; RUS, the Random Under-Sampling; RFE, Recursive Feature Elimination; LR, logistic regression; RF, random forest; SVM, Support Vector Machine; GNB, Gaussian Naive Bayes; FN, false negative; TN, true negative; TP, true positive; FP, false positive

Shanghai, China, between January 1, 2021, and July 31, 2023; there were no gender restrictions; at least one clinical symptom or laboratory finding was associated with bacterial pneumonia; computed tomography (CT) confirmed bacterial pneumonia; patients were diagnosed with bacterial LRTIs; TZP (4:1 ratio) (it will be abbreviated as ‘TZP’ unless otherwise specified) was used as the main first-line treatment regimen after admission; patients were aged 18 years or older;

The exclusion criteria for the study were as follows: incomplete clinical data; concurrent infection at other sites; TZP treatment for less than 3 days; those with severe hepatic or renal dysfunction (Child-Pugh class B or higher; estimated glomerular filtration rate (EGFR) below 20 ml/min) were excluded as well.

Endpoints and assessments

Based on the comprehensive assessment of patient symptoms and signs (such as temperature, lung auscultation, mental status, etc.), laboratory indicators, and lung imaging, clinicians determine whether piperacillin is effective or treatment has failed.

Criteria for clinical ineffectiveness: after 3 days or more for treatment, if none of the three indicators mentioned above show significant improvement, clinicians escalate

the antibiotic or consider transferring the patient. Criteria for clinical effectiveness: after 3 days or more for treatment, if the overall assessment of the mentioned criteria shows improvement, clinicians discontinue or reduce the use of the antibiotic.

Definition of features

In this study, we initially selected 26 factors as features for predicting TZP treatment of LRTIs. These features include aspects of patient demographics (such as gender, age, and weight), infection-related factors (such as temperature, white blood cell count, and neutrophil ratio), coexisting conditions (such as diabetes, chronic obstructive pulmonary disease, and bronchiectasis), and antimicrobial therapy details (such as TZP dosage and concurrent antimicrobial usage). Some features, initially continuous variables, were converted into binary variables, with reference cutoffs primarily based on relevant guidelines and indicators modified by our hospital’s laboratory according to the guidelines. Specific details can be referred to Table 1.

Data preprocessing

The dataset was initially randomly split into training and test sets using a 7:3 ratio. For the training set, missing

Table 1 Variables and their assignments

Factors	Abbreviations	Assignments/units
Age	Age(X ₁)	continuous variable, year
Weight	Weight(X ₂)	continuous variable, Kg
Estimated glomerular filtration rate	EGFR(X ₃)	continuous variable, ml/min
Blood urea nitrogen	BUN(X ₄)	continuous variable, mmol/L
Serum Albumin	ALB(X ₅)	continuous variable, g/L
White Blood Cell Count	WBC(X ₆)	continuous variable, ×10 ⁹ /L
D-dimer	DD(X ₇)	continuous variable, mg/L
C-Reactive Protein	CRP(X ₈)	continuous variable, mg/L
Procalcitonin	PCT(X ₉)	continuous variable, ng/ml
Neutrophil-to-Lymphocyte Ratio	NLR(X ₁₀)	continuous variable, dimensionless
Lymphocyte count	LYM(X ₁₁)	continuous variable, ×10 ⁹ /L
Body Temperature	Temp(X ₁₂)	≥37.3 °C = 1; <37.3 °C = 0
Sex	Sex(X ₁₃)	Female = 0; Male = 1
Respiratory Failure	RF(X ₁₄)	YES = 1; NO = 0
Recent hospitalization history(in 1 month)	RH(X ₁₅)	YES = 1; NO = 0
Diabetes Mellitus	DM(X ₁₆)	YES = 1; NO = 0
Active tumor	CA(X ₁₇)	YES = 1; NO = 0
Chronic Obstructive Pulmonary Disease	COPD(X ₁₈)	YES = 1; NO = 1
Bronchiectasis	BE(X ₁₉)	YES = 1; NO = 0
Interstitial lung disease	ILD(X ₂₀)	YES = 1; NO = 0
Cerebrovascular disease	CVD(X ₂₁)	YES = 1; NO = 0
Heart Failure	HF(X ₂₂)	YES = 1; NO = 0
Outpatient treatment history before admission	OPTH (X ₂₃)	YES = 1; NO = 0
Dosage of Piperacillin/Tazobactam	TZPD(X ₂₄)	2.5 g, Q8H = 1; 3.125 g, Q8H = 0
Combining erythromycin	Ery(X ₂₅)	YES = 1; NO = 0
Combining Nemonoxacin	NXC(X ₂₆)	YES = 1; NO = 0
Clinical outcomes (Result)		Failure = 1; Success = 0

values in continuous variables were imputed using the K-Nearest Neighbors (KNN) method, while categorical binary variables were imputed using mode imputation (SimpleImputer). Subsequently, continuous variables were standardized (StandardScaler) after imputation. The imputers and standardizers derived from the training set were then applied to process the test set data. Prior to model fitting, the processed training data underwent random under-sampling to ensure balanced classes in the outcome variable [31, 32]. This process was executed using the “imblearn” package in Python.

Feature selection

To ensure model robustness, we employed a two-pronged approach to feature selection. Initially, Recursive Feature Elimination (RFE) was used to identify a subset of potentially relevant features [33]. Subsequently, SHAP were used to assess the importance of these features and provide further insights into their impact on model predictions. The final feature set was then determined by evaluating the performance of models trained with different combinations of these features on both the test set and using cross-validation.

Hyperparameter optimization

After the feature selection process, we proceeded with hyperparameter optimization for the selected ML models. We GridSearch to maximize the AUC, a crucial metric for evaluating model performance.

Evaluation parameters

Considering the imbalanced classes within the dataset, we utilized several metrics, including AUC, Recall, Accuracy, Precision, and F1 Score, to evaluate the models. These metrics were computed using the following formulas.

$$Recall = \frac{TP}{TP + FN} \tag{1}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

$$Log\ loss = -\frac{1}{N} \sum_{i=1}^N [y_i * \log(pi) + (1 - y_i) * \log(1 - pi)] \tag{5}$$

True positive (TP) refers to the number of cases accurately predicted by the model as treatment failures, which were indeed treatment failures in reality. False positive (FP) represents the number of cases erroneously classified by the model as treatment failures, whereas they were actually successful treatments in reality. True negative (TN) signifies the number of cases accurately classified by the model as treatment successes, which indeed were successful treatments in reality. False negative (FN) indicates the number of cases incorrectly identified by the model as treatment successes, whereas they were actually treatment failures in reality. N: The number of samples. y_i : the true label for the i -th sample (either 0 or 1 in binary classification). p_i : the predicted probability of the positive class for the i -th sample. \log : The natural logarithm.

Model ensembling and interpretability

An EL model was constructed utilizing a VotingClassifier with a soft voting strategy and equal weighting of the constituent base learners. The performance of the EL model was rigorously evaluated via a 1000-iteration bootstrap resampling methodology applied to the training dataset. Subsequent performance assessment was conducted using an independent test dataset.

The relative importance of input features for the final predictive model was characterized using both SHAP and Boruta algorithms. SHAP force plot analysis was further conducted to elucidate the complex interplay of individual features in influencing clinical outcomes.

Online deployment of the model

For online deployment, the developed EL model was integrated into a web application hosted on a low-resource server platform running Windows Server 2012 R2. Flask (v2.2) was used to implement the web application’s front-end interface.

Results

Baseline characteristics

A total of 1,314 patients primarily treated with TZP as initial empiric antibiotic therapy were included in the analysis, with 995 (75.7%) in the success group and 319 (24.3%) in the failure group. The median age of all patients was 74.00 years (IQR: 66.00–82.00). The data were divided into training and test sets using stratified random sampling in a 7:3 ratio. As shown in the table, the distributions of most baseline characteristics, including Age, EGFR, Weight, and so on, were similar between the training and test sets (all $p > 0.05$). This suggests that the test set is representative of the training set and the overall population. The outcome variable, ‘Result’, also showed comparable distributions between the training and test sets ($p = 0.227$), indicating no significant bias in outcome

distribution across the datasets. Specific details can be referred to Table 2.

Feature selection

Based on the combined results of Recursive Feature Elimination (RFE) (Fig. 2a), SHapley Additive exPlanations (SHAP) analysis (Fig. 2b), and an assessment of model performance across various feature combinations (Fig. 3), we identified 13 key features (ALB, CRP, DD, LYM, NLR, EGFR, PCT, CA, HF, BUN, RH, RE, and CVD) as the common input features for all subsequent base models.

Model training

Feature selection was performed iteratively using RFE or SelectKBest methods, complemented by SHAP analysis on the training dataset. The predictive performance of the four base models was subsequently evaluated on the independent test set, with their respective ROC-AUC curves shown in Fig. 3.

Table 3 presents the AUC, recall, accuracy, precision, F1 score, and log loss obtained through independent test set of the four models. The results presented in Table 3 indicate that Logistic LR, RF, and SVM exhibited the highest overall performance among the models assessed.

These models achieved ROC-AUC values of 0.68, 0.69, and 0.68, respectively, and were therefore chosen as the base models for the subsequent ensemble modeling.

Model ensembling

To further enhance predictive performance, an EL model was developed using a VotingClassifier (voting='soft', weights= [1]), with the RF, LR, and SVM models as base learners. The EL model's performance was rigorously evaluated using 1000-iteration bootstrap cross-validation on the training set and subsequently assessed on an independent test set. This evaluation yielded a cross-validated ROC-AUC of 0.71 (Fig. 4a) and a test set ROC-AUC of 0.69 (Fig. 4b). On the independent test set, the EL model achieved a recall of 0.69 (Fig. 4c-d), accuracy of 0.64, precision of 0.40, F1-score of 0.50, and log loss of 0.66 (Fig. 4d).

Model interpretability

To enhance model interpretability, we employed both SHAP and Boruta algorithms to analyze feature importance in the ensemble model. The SHAP analysis revealed that ALB, CRP, and NLR were the most influential features in model predictions (Fig. 5a). This finding was

Table 2 Comparison of baseline characteristics between train set and test set*

Variable	Overall (n= 1314)	Train set (n= 919)	Test set(n= 395)	Pvalue
Age	74.00 (66.00–82.00)	74.00 (66.00–81.00)	75.00 (65.00–82.00)	0.600
EGFR	87.83 (72.15–97.14)	87.82 (71.23–97.12)	87.90 (72.80–97.21)	0.914
Weight	60.00 (53.00–70.00)	60.00 (52.75–70.00)	60.00 (53.30–70.00)	0.851
BUN	6.10 (4.80–7.90)	6.00 (4.80–7.80)	6.30 (4.90–8.20)	0.129
ALB	38.00 (35.30–40.60)	37.90 (35.15–40.65)	38.20 (35.50–40.60)	0.307
LYM	1.15 (0.74–1.61)	1.14 (0.73–1.62)	1.17 (0.76–1.58)	0.734
PCT	0.10 (0.07–0.19)	0.10 (0.07–0.19)	0.10 (0.08–0.19)	0.506
WBC	7.51 (5.75–10.20)	7.59 (5.87–10.38)	7.39 (5.60–9.99)	0.126
CRP	10.27 (1.03–51.67)	10.36 (0.91–51.27)	10.19 (1.45–52.16)	0.457
DD	0.60 (0.32–1.23)	0.61 (0.32–1.25)	0.58 (0.32–1.21)	0.650
NLR	0.75 (0.66–0.83)	0.76 (0.66–0.83)	0.74 (0.65–0.83)	0.348
Sex	863 (65.7%)	603 (65.6%)	260 (65.8%)	0.992
Temp	302 (23.0%)	211 (23.0%)	91 (23.0%)	1.000
RF	286 (21.8%)	208 (22.6%)	78 (19.7%)	0.276
RH	215 (16.4%)	147 (16.0%)	68 (17.2%)	0.641
DM	226 (17.2%)	164 (17.8%)	62 (15.7%)	0.386
CA	213 (16.2%)	156 (17.0%)	57 (14.4%)	0.286
COPD	735 (55.9%)	515 (56.0%)	220 (55.7%)	0.957
BE	242 (18.4%)	170 (18.5%)	72 (18.2%)	0.969
ILD	104 (7.9%)	74 (8.1%)	30 (7.6%)	0.865
CVD	508 (38.7%)	367 (39.9%)	141 (35.7%)	0.166
HF	576 (43.8%)	403 (43.9%)	173 (43.8%)	1.000
OPTH	961 (73.1%)	663 (72.1%)	298 (75.4%)	0.242
PIPD	974 (74.1%)	690 (75.1%)	284 (71.9%)	0.255
Ery	351 (26.7%)	246 (26.8%)	105 (26.6%)	0.999
NXC	192 (14.6%)	142 (15.5%)	50 (12.7%)	0.219
Result	319(24.3%)	214 (23.3%)	105 (26.6%)	0.227

*See Table 1 for full feature names

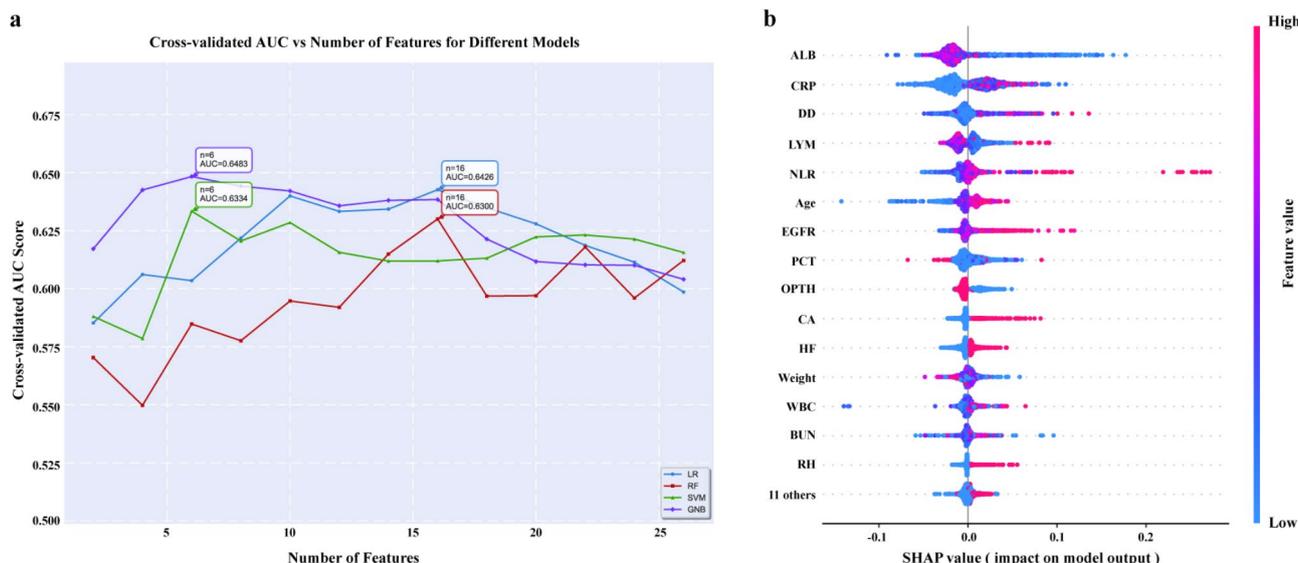


Fig. 2 Feature selection curves for each model. **(a)** Cross-validated AUC scores versus the number of selected features for different models. The optimal number of features and corresponding AUC scores are annotated for each model. **(b)** SHAP values showing the impact of each feature on model predictions. Features are ranked by their absolute SHAP values, with blue indicating lower values and red indicating higher values. ALB and CRP showed the strongest impact on model predictions

further supported by the Boruta algorithm, which identified ALB, CRP, and NLR as the top-ranked important features (Fig. 5b). To demonstrate the model’s decision-making process, we presented two representative cases using SHAP force plots. In a correctly predicted treatment failure case (Fig. 5c), elevated NLR (0.93) and DD (1.72) values strongly contributed to the prediction of treatment failure, while lower PCT (0.1) had an opposing effect. Conversely, in a successfully predicted treatment success case (Fig. 5d), normal ALB (42.9) and LYM (1.2) values were key factors driving the prediction toward treatment success, while elevated BUN (9.2) suggested potential treatment failure.

Web-based model deployment

To facilitate clinical application, we developed a web-based prediction system using the Flask framework (Fig. 6). This system implements our EL to predict TZP treatment outcomes in LRTIs patients. The interface allows clinicians to input 13 key clinical parameters identified through our feature importance analysis. The system provides real-time predictions categorized as either “Success” (Fig. 6a) or “Failure” (Fig. 6b), accompanied by corresponding treatment recommendations. The web application is freely accessible at <http://106.12.146.54:1020/>.

Discussion

Antibiotic resistance is a growing global threat, resulting in an estimated 1.3 million deaths in 2022. If left unchecked, this issue could lead to a staggering 10 million

deaths annually by 2050 due to infections caused by antibiotic-resistant bacteria [34]. One contributing factor is the common issue of inadequate antibiotic dosage in clinical practice [35–37]. This can happen because of a phenomenon called the mutation selection window [38]. When the dosage is too low, drug concentrations in the blood may fall within this window, allowing the growth of resistant bacteria and ultimately treatment failure.

Although the use of TZP (4:1) can reduce the intensity of antibiotic use to meet the requirements of hospital management to some extent, the dosage of piperacillin in TZP (4:1) is obviously insufficient for many patients, due to the extreme limitation of tazobactam [39]. In the absence of therapeutic drug monitoring (TDM), making preliminary predictions about the clinical efficacy of TZP (4:1) for patients based on their clinical information may hold positive significance in enhancing its clinical effectiveness and decreasing the emergence of resistant bacteria. This approach could contribute to a beneficial balance between the intensity of antimicrobial drug use and their therapeutic effectiveness.

In ML model evaluation metrics, accuracy and recall are two conflicting indicators [40]. Which side to lean towards often depends on the nature of the research question. In this research, we place a higher emphasis on the model’s recall. This is because we prefer to identify as many cases as possible that may not respond well to treatment, allowing for the early consideration of alternative medications such as TZP (8:1), cefoperazone-sulbactam, and others. Additionally, if the model predicts cases that could potentially fail the treatment as cases likely

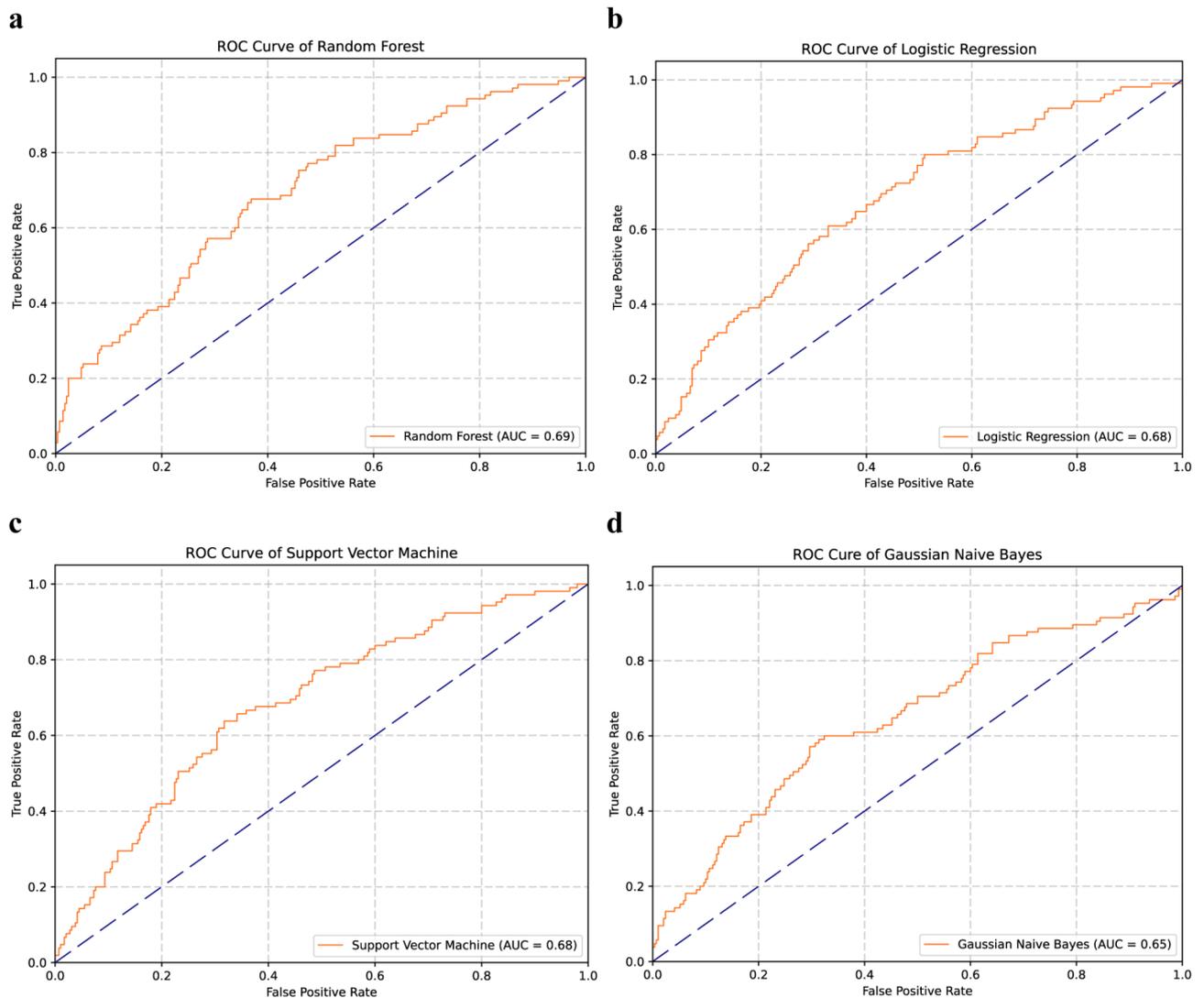


Fig. 3 ROC curves of four machine learning models on the independent test set. **(a)** Receiver Operating Characteristic (ROC) curves for RF (AUC=0.69), **(b)** LR (AUC=0.68), **(c)** SVM (AUC=0.68), and **(d)** GNB (AUC=0.65) models, respectively. The dashed diagonal line represents random prediction (AUC=0.5). The x-axis shows the False Positive Rate, and the y-axis shows the True Positive Rate

Table 3 Performance metrics of the models based on independent test set

Model	Accuracy	Precision	Recall	F1 score	AUC	Log loss
LR	0.63	0.38	0.63	0.47	0.68	0.68
RF	0.64	0.40	0.68	0.50	0.69	0.65
SVM	0.61	0.37	0.68	0.48	0.68	0.66
GNB	0.52	0.32	0.71	0.44	0.65	1.17

to succeed, it does not seem to be a significant issue in clinical practice. In such situations, the success rate of alternative antimicrobial treatments chosen by clinicians may be higher. In summary, a model with higher recall can filter out more cases where TZP (4:1) treatment is unsuccessful, and a slightly lower precision does not have a substantial impact on clinical practice.

In this study, we also found many factors that contribute to the failure of TZP (4:1) anti-infection, such as

hypoproteinemia, higher estimated glomerular filtration rate (EGFR) (high EGFR levels/activity may contribute to reduced drug concentrations), increased neutrophil ratio (NLR), increased C-reactive protein (CRP), and decreased serum albumin (ALB), and active tumors (CA) (Fig. 4). Individually, these features have limited impact on the outcome of infection, but when combined, they exhibit a strong predictive effect [41, 42]. Additionally, these features hold relevance for future related research.

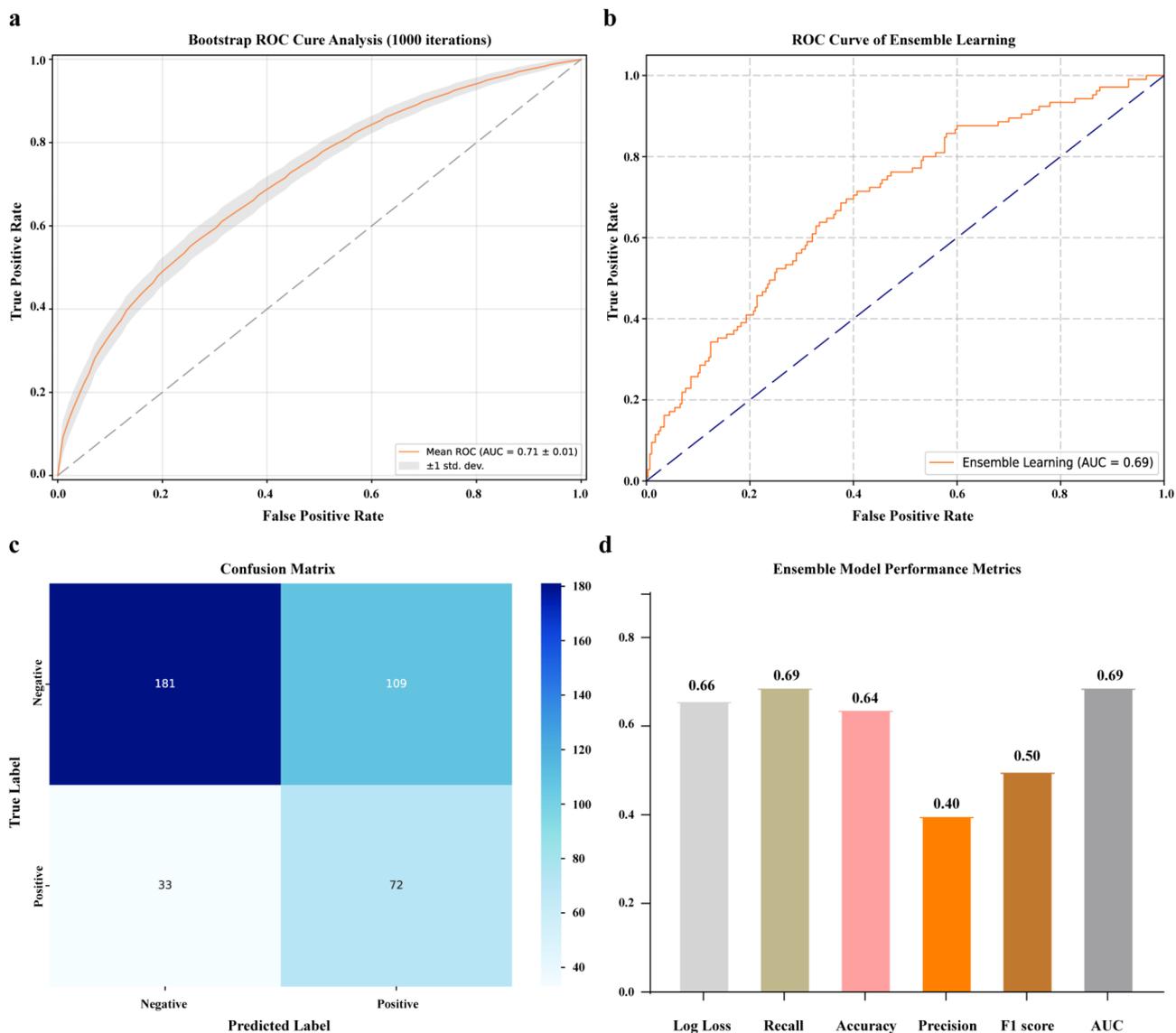


Fig. 4 Performance of the EL model. **(a)** Mean ROC curve (with standard deviation shown as shaded area) derived from 1000 bootstrap iterations on the training set (AUC = 0.71 ± 0.01). **(b)** ROC curve on the independent test set (AUC = 0.69). **(c)** Confusion matrix on the test set. **(d)** Performance metrics on the test set: log loss = 0.66, recall = 0.69, accuracy = 0.64, precision = 0.40, F1-score = 0.50, ROC-AUC = 0.69

Additionally, there is a tendency towards overuse or misuse of antibiotics, including combinations like PIPT with other agents such as fluoroquinolones or macrolides. It is well-recognized that antibiotic misuse and overuse are primary drivers of antimicrobial resistance (AMR) [43], which has led to a significant increase in bacterial resistance rates in recent years. Initially, we included concomitant erythromycin or nemonoxacin administration as baseline characteristics in the TZP analysis. However, feature importance ranking in both initial and final iterations consistently placed these factors low among all features. This suggests that combined antibiotic therapy may not be a primary determinant of anti-infective treatment success, and other factors such

as inflammatory markers, nutritional status, and comorbidities should be considered.

The AUC value reflects the model's robust classification capability, furnishing dependable predictions in discerning between patients who have undergone successful and failed treatments [44, 45]. Although the AUC value on the test set in the present study did not exceed 0.7 (Fig. 4), lower AUC values seem to be a common problem for models that use clinical information as features [25, 46]. This seems to be related to the complexity of clinical data and the imbalance of data. In some cases, the situation may indeed exceed the prediction ability of the model, such as when the bacteria infecting the patient exceed the antibacterial spectrum of TZP [12]. This problem may

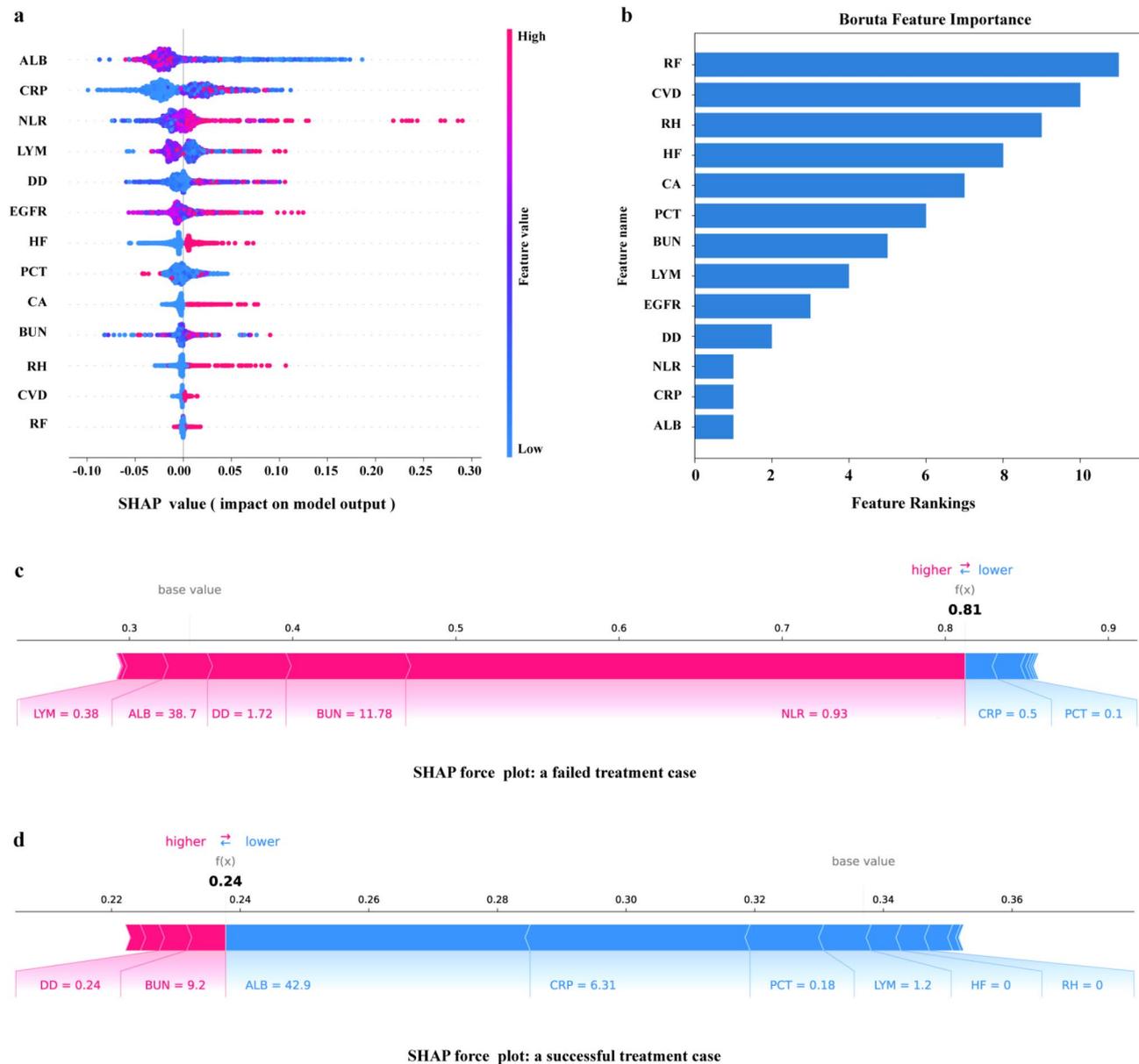


Fig. 5 Feature importance analysis and interpretation of the ensemble model predictions. **(a)** SHAP summary plot showing the impact and distribution of each feature on model output. Red indicates higher feature values, while blue represents lower values. The x-axis represents the SHAP value impact on model predictions. **(b)** Boruta feature importance rankings displaying the relative importance of selected features in the final model. **(c)** SHAP force plot demonstrating the prediction process for a correctly identified treatment failure case. Features in red pushed the prediction toward treatment failure, while blue features opposed this prediction. **(d)** SHAP force plot illustrating the prediction process for a correctly predicted treatment success case. Features in red suggested treatment failure, while blue features supported treatment success

be improved in the future by increasing the amount of research data and using more advanced models.

The feature importance analysis provided valuable insights into the model’s decision-making process. Both SHAP and Boruta algorithms consistently identified key clinical parameters that significantly influenced treatment outcomes. Laboratory indicators such as ALB, CRP, and NLR demonstrated substantial impact on model predictions, which aligns with their established clinical relevance in infection and inflammation. The SHAP force

plots further validated our feature selection by illustrating how these parameters interact to predict treatment outcomes. For instance, in treatment failure cases, elevated inflammatory markers (NLR, DD) strongly contributed to the prediction, while in successful cases, normal albumin levels and lymphocyte counts were key positive predictors. This interpretability not only validates our model’s clinical reliability but also reflects established pathophysiological mechanisms of respiratory infections. The consistency between our model’s feature importance

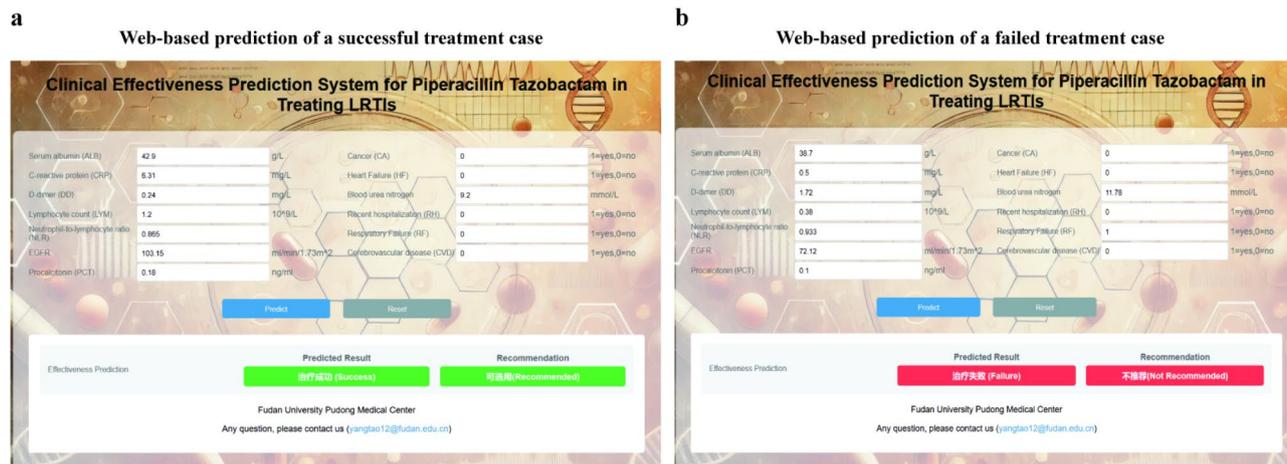


Fig. 6 Web-based implementation of the prediction model for TZP treatment effectiveness. (a) Screenshot of the web interface showing a successful treatment prediction case. The system displays input fields for 13 clinical parameters and provides a “Success” prediction with corresponding treatment recommendation. (b) Screenshot of the web interface showing a failed treatment prediction case. The same interface predicts treatment failure based on different clinical parameter values, with appropriate recommendations against PIPT treatment

rankings and clinical knowledge enhances its credibility as a decision support tool. Moreover, the selected features are readily available in routine clinical practice, making our model both interpretable and practical for real-world applications.

This study has several limitations that should be addressed to validate the proposed model’s effectiveness further. First, while we carefully selected 26 clinical features for inclusion, expanding the initial feature set might uncover additional variables that significantly impact treatment outcomes. This expansion could enhance the model’s predictive capability. Second, the sample size, though sufficient for this analysis, remains limited in scope, particularly in representing the diverse patient populations and pathogen profiles observed in broader clinical settings. Increasing the dataset size and diversity would likely improve the model’s generalizability and robustness. Third, the complex and multifactorial etiology of LRTIs, including infections caused by pathogens outside the antibacterial spectrum of TZP or influenced by factors unmeasured in this study, may challenge the model’s predictive accuracy. Addressing this would require incorporating data on potential microbial resistance profiles and using advanced modeling approaches capable of capturing such complexities. Despite these limitations, the study provides a robust foundation for applying ML in antimicrobial decision-making, supporting future research to refine and validate the model’s clinical utility.

Conclusions

In this study, we successfully developed and validated an ensemble machine learning model that effectively predicts the clinical effectiveness of TZP (4:1) in treating

LRTIs. The model achieved robust performance metrics including an AUC of 0.69 and a recall of 0.69 on the independent test set. Through feature importance analysis using both SHAP and Boruta algorithms, we identified key clinical parameters that significantly influence treatment outcomes, with laboratory indicators such as ALB, CRP, and NLR demonstrating substantial predictive value. The model’s interpretability was enhanced through SHAP force plots, which revealed how different clinical parameters interact to influence predictions. To facilitate clinical application, we deployed the model as a web-based tool that provides real-time predictions and treatment recommendations (<http://106.12.146.54:1020/>). This work represents a significant step forward in applying machine learning to support clinical decision-making in antibiotic therapy, potentially helping optimize treatment selection and reduce both inadequate treatment and antibiotic overexposure. While the current model shows promise, future studies with larger datasets and prospective validation are needed to further enhance its clinical utility and reliability.

Abbreviations

- LRTIs Lower respiratory tract infections
- ML Machine Learning
- EL Ensemble Learning
- LR Logistic Regression
- RF Random Forest
- AUC Area Under the Curve
- ROC Receiver Operating Characteristic
- TZP Piperacillin-tazobactam
- SVM Support Vector Machine
- GNB Gaussian Naive Bayes
- SHAP the SHapley Additive exPlanations
- RFE the Recursive Feature Elimination
- TP True positive
- FP False positive
- TN True negative

FN False negative

Acknowledgements

Not applicable.

Author contributions

Conceptualization: K.H., T.Y.; Methodology: Y.M., K.H. and J.T.L.; Investigation and Supervision: Y.L., L.K.P.; Formal analysis: L.S.; Validation: T.Z.; Visualization: Z.J.Z. Resources: Z.Y. H.; Software: C.Y. X.; Writing - original draft: Y.M., Y.T.Y.; Writing - review & editing: L.K.P. All authors reviewed the manuscript.

Funding

This research is supported by grants from the Health Commission of Pudong New Area Health and Technology Project (Grant No: PW2023A-31), Clinical Pharmacy Key Specialized Subject Construction Project of Pudong Hospital affiliated to Fudan University (Grant No: Tszk2020-05), Pulmonary and Critical Care Medicine Key Subject Construction Project of Pudong Hospital affiliated to Fudan University (Grant No: Zdxk2020-08), and Clinical Diagnosis and Treatment Innovation Research Project of Shanghai Pudong Hospital (Grant No: YJLC202409). The authors would also like to acknowledge all interviewers for survey data collection work.

Data availability

Data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The study is a retrospective analysis based on historical data and falls under the waiver of informed consent. Before the study's commencement, the Ethics Committee of Pudong Hospital, Shanghai confirmed the waiver of informed consent. The research protocol, which strictly adhered to the principles of the Declaration of Helsinki and its subsequent amendments, also received approval from the committee (Approval NO: 2024-KY-001-E01).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author details

¹Department of Respiratory and Critical Care Medicine, Shanghai Pudong Hospital, Fudan University Pudong Medical Center, Shanghai, China

²Shanghai Key Laboratory of Magnetic Resonance, School of Physics and Electronic Science, East China Normal University, Shanghai, China

³School of Pharmacy, Shanghai University of Medicine & Health Sciences, Shanghai, China

⁴Department of Pharmacy, Shanghai Pudong Hospital, Fudan University Pudong Medical Center, Shanghai, China

⁵School of Materials and Chemistry, University of Shanghai for Science and Technology, Shanghai, China

⁶Institute of Magnetic Resonance and Molecular Imaging in Medicine, East China Normal University, Shanghai 200241, China

Received: 6 May 2024 / Accepted: 5 March 2025

Published online: 17 March 2025

References

- GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *Lancet*. 2020;396:1204–22.
- GBD 2019 Diseases and Injuries Collaborators. Lower respiratory infections-Level 3 cause. 2019. <https://www.thelancet.com/pb-assets/Lancet/gbd/summaries/diseases/lower-respiratory-infections.pdf>. Accessed 22 Feb 2024.
- Langelier C, Kalantar KL, Moazed F, Wilson MR, Crawford ED, Deiss T et al. Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. *Proc Natl Acad Sci USA*. 2018;115:E12353–e62.369.
- Cai BQ, Cai SX, Chen RC, Cui LY, Feng YL, Gu YT, et al. Expert consensus on acute exacerbation of chronic obstructive pulmonary disease in the People's Republic of China. *Int J Chron Obstruct Pulmon Dis*. 2014;9:381–95.
- Cao B, Huang Y, She DY, Cheng QJ, Fan H, Tian XL, et al. Diagnosis and treatment of community-acquired pneumonia in adults: 2016 clinical practice guidelines by the Chinese thoracic society, Chinese medical association. *Clin Respir J*. 2018;12:1320–60.
- Shi Y, Huang Y, Zhang TT, Cao B, Wang H, Zhuo C, et al. Chinese guidelines for the diagnosis and treatment of hospital-acquired pneumonia and ventilator-associated pneumonia in adults (2018 Edition). *J Thorac Dis*. 2019;11:2581–616.
- Zhu YG, Tang XD, Lu YT, Zhang J, Qu JM. Contemporary situation of Community-acquired pneumonia in China: A systematic review. *J Transl Int Med*. 2018;6:26–31.
- Arain S, Khalawi F, Parakkal SA, AlHamad HS, Thorakkattil SA, Alghashmari FFJ et al. Drug utilization evaluation and impact of pharmacist interventions on optimization of Piperacillin/Tazobactam use: A retrospective analysis and prospective audit. *Antibiot (Basel)*. 2023;12.
- Qian ET, Casey JD, Wright A, Wang L, Shotwell MS, Siemann JK, et al. Cefepime vs Piperacillin-Tazobactam in adults hospitalized with acute infection: the ACORN randomized clinical trial. *JAMA*. 2023;330:1557–67.
- Bryson HM, Brogden RN. Piperacillin/tazobactam. A review of its antibacterial activity, Pharmacokinetic properties and therapeutic potential. *Drugs*. 1994;47:506–35.
- Lister PD, Prevan AM, Sanders CC. Importance of beta-lactamase inhibitor pharmacokinetics in the pharmacodynamics of inhibitor-drug combinations: studies with piperacillin-tazobactam and piperacillin-sulbactam. *Antimicrob Agents Chemother*. 1997;41:721–7.
- Havey TC, Hull MW, Romney MG, Leung V. Retrospective cohort study of inappropriate piperacillin-tazobactam use for lower respiratory tract and skin and soft tissue infections: opportunities for antimicrobial stewardship. *Am J Infect Control*. 2015;43:946–50.
- Cojutti PG, Pai MP, Tonetti T, Siniscalchi A, Viale P, Pea F. Balancing the scales: achieving the optimal beta-lactam to beta-lactamase inhibitor ratio with continuous infusion Piperacillin/tazobactam against extended spectrum beta-lactamase producing enterobacterales. *Antimicrob Agents Chemother*. 2024:e0140423.
- Fujii M, Karumai T, Yamamoto R, Kobayashi E, Ogawa K, Tounai M, et al. Pharmacokinetic and pharmacodynamic considerations in antimicrobial therapy for sepsis. *Expert Opin Drug Metab Toxicol*. 2020;16:415–30.
- Novy E, Roger C, Roberts JA, Cotta MO. Pharmacokinetic and pharmacodynamic considerations for antifungal therapy optimisation in the treatment of intra-abdominal candidiasis. *Crit Care*. 2023;27:449.
- Sanders CV, Jr. Piperacillin/tazobactam in the treatment of community-acquired and nosocomial respiratory tract infections: a review. *Intensive Care Med*. 1994;20(Suppl 3):S21–6.
- Shlaes DM, Baughman R, Boylen CT, Chan JC, Charan NB, Cormier YC, et al. Piperacillin/tazobactam compared with ticarcillin/clavulanate in community-acquired bacterial lower respiratory tract infection. *J Antimicrob Chemother*. 1994;34:565–77.
- Alowais SA, Alghamdi SS, Alsuhbany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ*. 2023;23:689.
- Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of health care. *Am J Med*. 2019;132:795–801.
- Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inf*. 2007;2:59–77.
- Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8–17.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–44.
- Qin J, Yang Y, Ai C, Ji Z, Chen W, Song Y, et al. Antibiotic combinations prediction based on machine learning to multicentre clinical data and drug interaction correlation. *Int J Antimicrob Agents*. 2024;63:107122.
- Kim C, Choi YH, Choi JY, Choi HJ, Park RW, Rhie SJ. Translation of machine Learning-Based prediction algorithms to personalised empiric antibiotic selection: A Population-Based cohort study. *Int J Antimicrob Agents*. 2023;62:106966.

25. Wang X, Ren H, Ren J, Song W, Qiao Y, Ren Z, et al. Machine learning-enabled risk prediction of chronic obstructive pulmonary disease with unbalanced data. *Comput Methods Programs Biomed.* 2023;230:107340.
26. Diéguez-Santana K, González-Díaz H. Machine learning in antibacterial discovery and development: A bibliometric and network analysis of research hotspots and trends. *Comput Biol Med.* 2023;155:106638.
27. Yelin I, Snitser O, Novich G, Katz R, Tal O, Parizade M, et al. Personal clinical history predicts antibiotic resistance of urinary tract infections. *Nat Med.* 2019;25:1143–52.
28. Yurtseven A, Buyanova S, Agrawal AA, Bochkareva OO, Kalinina OV. Machine learning and phylogenetic analysis allow for predicting antibiotic resistance in *M. tuberculosis*. *BMC Microbiol.* 2023;23:404.
29. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global Understanding with explainable AI for trees. *Nat Mach Intell.* 2020;2:56–67.
30. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng.* 2018;2:749–60.
31. Feng H, Qin W, Wang H, Li Y, Hu G. A combination of resampling and ensemble method for text classification on imbalanced data. In: Wei J, Zhang L-J, editors. *Big Data – BigData 2021*. Cham: Springer International Publishing; 2022. pp. 3–16.
32. Japkowicz N, Stephen S. The class imbalance problem: A systematic study. *Intell Data Anal.* 2002;6:429–49.
33. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for Cancer classification using support vector machines. *Mach Learn.* 2002;46:389–422.
34. O'Neill J. Tackling drug-resistant infections globally: final report and recommendations. 2016.
35. Donadello K, Antonucci E, Cristallini S, Roberts JA, Beumier M, Scolletta S, et al. β -Lactam pharmacokinetics during extracorporeal membrane oxygenation therapy: A case-control study. *Int J Antimicrob Agents.* 2015;45:278–82.
36. Póvoa P, Moniz P, Pereira JG, Coelho L. Optimizing antimicrobial drug dosing in critically ill patients. *Microorganisms.* 2021;9.
37. Yang H, Cui X, Ma Z, Liu L. Evaluation outcomes associated with alternative dosing strategies for Piperacillin/Tazobactam: A systematic review and Meta-Analysis. *J Pharm Pharm Sci.* 2016;19:274–89.
38. Drlica K, Zhao X. Mutant selection window hypothesis updated. *Clin Infect Dis.* 2007;44:681–8.
39. Filius PM, Liem TB, van der Linden PD, Janknegt R, Natsch S, Vulto AG, et al. An additional measure for quantifying antibiotic use in hospitals. *J Antimicrob Chemother.* 2005;55:805–8.
40. Nabipour M, Nayyeri P, Jabani H, Shahab S, Mosavi A. Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. *IEEE Access.* 2020;8:150199–212.
41. de Jager CPC, van Wijk PTL, Mathoera RB, de Jongh-Leuvenink J, van der Poll T, Wever PC. Lymphocytopenia and neutrophil-lymphocyte count ratio predict bacteremia better than conventional infection markers in an emergency care unit. *Crit Care.* 2010;14:R192.
42. Hatherill M, Tibby SM, Sykes K, Turner C, Murdoch IA. Diagnostic markers of infection: comparison of procalcitonin with C reactive protein and leucocyte count. *Arch Dis Child.* 1999;81:417–21.
43. Bassetti M, Russo A, Carnelutti A, La Rosa A, Righi E. Antimicrobial resistance and treatment: an unmet clinical safety need. *Expert Opin Drug Saf.* 2018;17:669–80.
44. Kim T, Lee J-S. Maximizing AUC to learn weighted Naive Bayes for imbalanced data classification. *Expert Syst Appl.* 2023;217:119564.
45. Ling CX, Huang J, Zhang H, AUC: A better measure than accuracy in comparing learning algorithms. In: Xiang Y, Chaib-draa B, editors. *Advances in artificial intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2003. pp. 329–41.
46. Ouyang Y, Cheng M, He B, Zhang F, Ouyang W, Zhao J, et al. Interpretable machine learning models for predicting in-hospital death in patients in the intensive care unit with cerebral infarction. *Comput Methods Programs Biomed.* 2023;231:107431.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.