



OPEN ACCESS

EDITED BY
Pei-Hui Wang,
Shandong University, China

REVIEWED BY
Yuanchia Chu,
Taipei Veterans General Hospital, Taiwan
Mehmet Huyut,
Erzincan University, Türkiye
Fida Hussain,
Iqra University, Pakistan

*CORRESPONDENCE
Yi-Shan Zheng
✉ 284159264@qq.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 09 January 2025
ACCEPTED 05 March 2025
PUBLISHED 19 March 2025

CITATION
Xu J-S, Yang K, Quan B, Xie J and Zheng Y-S (2025) A multicenter study on developing a prognostic model for severe fever with thrombocytopenia syndrome using machine learning. *Front. Microbiol.* 16:1557922. doi: 10.3389/fmicb.2025.1557922

COPYRIGHT
© 2025 Xu, Yang, Quan, Xie and Zheng. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A multicenter study on developing a prognostic model for severe fever with thrombocytopenia syndrome using machine learning

Jian-She Xu^{1†}, Kai Yang^{2†}, Bin Quan³, Jing Xie² and Yi-Shan Zheng^{1,2*}

¹School of Public Health, Nanjing Medical University, Nanjing, China, ²Department of Intensive Care Unit, The Second Hospital of Nanjing, Affiliated of Nanjing University of Chinese Medicine, Nanjing, China, ³Department of Infectious Disease, The First Affiliated Hospital of Wannan Medical College, Wuhu, China

Background: Severe Fever with Thrombocytopenia Syndrome (SFTS) is a disease caused by infection with the Severe Fever with Thrombocytopenia Syndrome virus (SFTSV), a novel Bunyavirus. Accurate prognostic assessment is crucial for developing individualized prevention and treatment strategies. However, machine learning prognostic models for SFTS are rare and need further improvement and clinical validation.

Objective: This study aims to develop and validate an interpretable prognostic model based on machine learning (ML) methods to enhance the understanding of SFTS progression.

Methods: This multicenter retrospective study analyzed patient data from two provinces in China. The derivation cohort included 292 patients treated at The Second Hospital of Nanjing from January 2022 to December 2023, with a 7:3 split for model training and internal validation. The external validation cohort consisted of 104 patients from The First Affiliated Hospital of Wannan Medical College during the same period. Twenty-four commonly available clinical features were selected, and the Boruta algorithm identified 12 candidate predictors, ranked by Z-scores, which were progressively incorporated into 10 machine learning models to develop prognostic models. Model performance was assessed using the area under the receiver-operating-characteristic curve (AUC), accuracy, recall, and F1 score. The clinical utility of the best-performing model was evaluated through decision curve analysis (DCA) based on net benefit. Robustness was tested with 10-fold cross-validation, and feature importance was explained using SHapley Additive exPlanation (SHAP) both globally and locally.

Results: Among the 10 machine learning models, the XGBoost model demonstrated the best overall discriminatory ability. Considering both AUC index and feature simplicity, a final interpretable XGBoost model with 7 key features was constructed. The model showed high predictive accuracy for patient outcomes in both internal (AUC = 0.911, 95% CI: 0.842–0.967) and external validations (AUC = 0.891, 95% CI: 0.786–0.977). A clinical tool based on this model has been developed and implemented using the Streamlit framework.

Conclusion: The interpretable XGBoost-based prognostic model for SFTS shows high predictive accuracy and has been translated into a clinical tool. The model's

7 key features serve as valuable indicators for early prognosis of SFTS, warranting close attention from healthcare professionals in clinical practice.

KEYWORDS

severe fever with thrombocytopenia syndrome, machine learning, Boruta algorithm, prognostic model, clinical validation

Introduction

Severe Fever with Thrombocytopenia Syndrome (SFTS) is an acute infectious disease caused by the Severe Fever with Thrombocytopenia Syndrome virus (SFTSV), a member of the Bunyaviridae family, and is primarily transmitted through tick bites (Yu et al., 2011). SFTS has a high incidence in East Asia, with a fatality rate of up to 30%, particularly in China, Japan, and South Korea (Kim et al., 2013; Liu et al., 2014). The clinical characteristics of the disease include high fever, thrombocytopenia, leukopenia, and multi-organ dysfunction, with severe cases often resulting in death (Gai et al., 2012; Liu et al., 2014). Currently, there is no specific treatment available, making early identification of critically ill patients crucial for improving outcomes (Wang et al., 2021). Clinically, there is a lack of accurate and reliable prognostic models to predict disease progression and patient outcomes (Jia et al., 2017; Liu et al., 2023). As a heterogeneous disease, the assessment of disease severity and prognosis in SFTS patients relies on various clinical and laboratory indicators.

Routine blood tests (RBT) play a critical role in the diagnosis, prognosis, and follow-up of many diseases (Huyut, 2023; Huyut and Kocaturk, 2022; Tahir Huyut et al., 2022). Due to their accessibility and cost-effectiveness, RBT data have been widely used in artificial intelligence studies for disease diagnosis and prognosis (Huyut and Huyut, 2021, 2023; Santos-Silva et al., 2024; Üstündag et al., 2023). Recent clinical research has demonstrated the utility of RBT in the early detection and prognosis of COVID-19 and other diseases (Huyut and Ilkbahar, 2021; Huyut and Velichko, 2023; Mertoglu et al., 2022). For instance, Huyut et al. developed successful AI models using RBT data for the diagnosis

and prognosis of COVID-19 (Huyut and Ustundag, 2022; Huyut and Velichko, 2022; Huyut et al., 2022; Velichko et al., 2022). Similarly, Yan et al. (2021) applied the Gradient Boosting Decision Tree method to RBT data for the early diagnosis of multiple myeloma (Yan et al., 2021). Other studies have also highlighted the potential of RBT-based AI models in various clinical contexts. For example, Soerensen et al. (2022) developed an AI model for cancer risk prognosis in primary care using RBT data (Soerensen et al., 2022), while Wu C. C. et al. (2019) achieved 95.7% accuracy in predicting lung cancer using the Random Forest algorithm based on 19 routine blood test parameters (Wu J. et al., 2019). Additionally, Wu J. et al. (2019) utilized the Random Forest method to predict high-risk liver diseases (Wu C. C. et al., 2019), and Kawakami et al. (2019) applied AI models such as Naive Bayes, Artificial Neural Networks, and Elastic Net for the preoperative diagnosis and prognosis of epithelial ovarian cancer (Kawakami et al., 2019). More recently, Haider et al. (2022) used artificial neural networks to analyze RBT data for early differentiation among leukemias (Haider et al., 2022). These studies underscore the significant potential of RBT data in enhancing disease prognosis and management through AI techniques.

Machine learning (ML) techniques, which excel at handling complex and high-dimensional data, have garnered widespread attention for their ability to improve predictive accuracy when used to build prognostic models (Fan et al., 2023; Tian et al., 2023). However, despite the superior performance of ML models, their “black box” (Azodi et al., 2020) nature makes it difficult to interpret the basis of their prognoses, thus limiting their clinical applicability. To address the challenge of model interpretability, the SHapley Additive exPlanation (SHAP) method offers an effective solution. SHAP, based on Shapley values from game theory, quantifies the contribution of each input feature to the model’s prognosis, providing interpretability for complex ML models (Mosca et al., 2022; Zhou et al., 2024). SHAP helps visualize how each feature influences the prognosis, enabling clinicians to better understand the decision-making process of ML models, and thus increasing the credibility and usability of these models in clinical practice (Feretzakis et al., 2024). For a disease like SFTS, which is influenced by multiple factors, SHAP can reveal the role of various clinical features in prognosis, facilitating the development of individualized treatment plans.

Despite many previous studies identifying risk factors for death from Thrombocytopenia Syndrome among adults, there is still a need to clarify the characteristics affecting Thrombocytopenia Syndrome. Our study, modeling “comorbidities and laboratory tests” together, provides low-cost, rapid and reliable results on Thrombocytopenia Syndrome. This study aims to develop a machine learning-based prognostic model for SFTS and improve

Abbreviations: SFTS, Severe Fever with Thrombocytopenia Syndrome; SFTSV, Severe Fever with Thrombocytopenia Syndrome virus; RBT, Routine blood tests; ML, Machine learning; AUC, Area under the receiver-operating-characteristic curve; DCA, Decision curve analysis; SHAP, SHapley Additive explanation; SOFA, Sequential Organ Failure Assessment; APACHE II, Acute Physiology and Chronic Health Evaluation II; PNI, Prognostic Nutritional Index; MV, Mechanical ventilation; HF, High flow oxygen therapy; CRRT, Continuous renal replacement therapy; T, Temperature; HR, Heart rate; MAP, Mean arterial pressure; WBC, White blood cell count; N, Neutrophil count; L, Lymphocyte count; PLT, Platelet count; U, Urea; TBil, Total bilirubin; ALT, Alanine aminotransferase; AST, Aspartate aminotransferase; ALB, Albumin; LDH, Lactate dehydrogenase; ALP, Alkaline phosphatase; Cr, Creatinine; PNI, Prognostic Nutritional Index; LR, Logistic Regression; SVM, Support Vector Machine; KNN, K-Nearest Neighbors; NB, Naive Bayes; DT, Decision Tree; RF, Random Forest; GBDT, Gradient Boosting Decision Tree; XGBoost, Extreme Gradient Boosting.

its interpretability using SHAP, ultimately resulting in a practical clinical tool based on the model.

Methods

Study population

This study is a retrospective multicenter cohort study conducted in China, aiming to derive and validate a prognostic model for patients with SFTS. The derivation cohort consisted of 292 patients treated at The Second Hospital of Nanjing between January 2022 and December 2023, while the external validation cohort included 104 patients from The First Affiliated Hospital of Wannan Medical College during the same period. Inclusion criteria were: SFTS patients were diagnosed based on the presence of acute fever (with a temperature of 38°C or higher) and platelet count $<100 \times 10^9/L$, with lab-confirmed SFTS virus (SFTSV) infection by qRT-PCR (Huang et al., 2019; Yu and Morita, 2019), hospitalized adults aged 18 years or older, and those with complete demographic information, laboratory test results, and final outcome data. Exclusion criteria: pregnant patients and cancer patients were excluded from the study.

Data collection and processing

We collected demographic characteristics, vital signs, and laboratory data at admission from the electronic medical record system. Missing values for categorical variables were imputed using the mode, while missing continuous variables were imputed using the median method (Berkelmans et al., 2022). To minimize the impact of different measurement scales on the model, all continuous variables were standardized. The Sequential Organ Failure Assessment (SOFA), Acute Physiology and Chronic Health Evaluation II (APACHE II) scores, and Prognostic Nutritional Index (PNI) were calculated based on the first examination at admission.

To avoid the impact of missing data on model construction, features with more than 20% missing values were excluded. Twenty-four clinical features were selected as candidate features, including: sex, mechanical ventilation (MV), high flow oxygen therapy (HF), continuous renal replacement therapy (CRRT), underlying disease, consciousness, age, temperature (T), heart rate (HR), mean arterial pressure (MAP), white blood cell count (WBC), neutrophil count (N), lymphocyte count (L), platelet count (PLT), urea (U), total bilirubin (TBil), alanine aminotransferase (ALT), aspartate aminotransferase (AST), albumin (ALB), lactate dehydrogenase (LDH), alkaline phosphatase (ALP), creatinine (Cr), D-dimer, and Prognostic Nutritional Index (PNI).

Model development and comparison

During the feature selection stage, the Boruta algorithm was used to identify the most predictive features from the 24

clinical features, ranking their importance based on Z-scores. The derivation cohort data was split into a 7:3 ratio for training and internal validation sets to avoid overfitting. We progressively incorporated the selected clinical features into 10 different machine learning models to construct prognostic models, including Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Gradient Boosting Decision Tree (GBDT), AdaBoost, Voting Classifier, and Extreme Gradient Boosting (XGBoost).

To optimize model performance, we used GridSearchCV for hyperparameter tuning. Each model underwent 5-fold and 10-fold cross-validation to identify the best combination of hyperparameters. Model stability was further validated using bootstrapping, through which we computed confidence intervals for the performance metrics, thereby enhancing the reliability of the model.

Additionally, to evaluate the prognostic performance of APACHE II and SOFA scores, we constructed three groups of models: (1) a multivariable model based on the 7 selected features, (2) a univariable model based solely on the Acute Physiology and Chronic Health Evaluation score (APACHEII score), and (3) a univariable model based solely on the Sequential Organ Failure Assessment Score (SOFA score). All models were trained and validated using the XGBoost algorithm, and their performance was compared using AUC, accuracy, precision, recall, and F1 score. To further assess model performance based on disease severity, we conducted a stratified analysis using the SOFA score. The patients were divided into two groups: Low-risk group: Patients with a SOFA score ≤ 6 ; High-risk group: Patients with a SOFA score ≥ 7 .

External validation

To assess the external performance of the model, we conducted external validation using data from 104 SFTS patients treated at The First Affiliated Hospital of Wannan Medical College between January 2022 and December 2023. The inclusion and exclusion criteria for the external validation cohort were consistent with those of the derivation cohort. As the primary goal of the external validation was to evaluate the effectiveness of the prognostic model, only the outcome variables and the clinical features included in the final model were collected for these patients.

Feature selection and model interpretation

In the feature selection phase, we applied the Boruta algorithm to identify the most predictive features from the initial set of 24 clinical features. The algorithm ranked the importance of each feature based on Z-scores. To address the “black box” issue associated with machine learning models, we employed the SHAP method for model interpretation. SHAP provided both global and local explanations, detailing the contribution of each selected feature to the prognostic prognoses. This approach enabled a clearer understanding of how the model arrived at its prognoses, offering transparency for clinical application.

Web tool deployment based on the Streamlit framework

To facilitate the clinical application of the final prognostic model, we deployed it in a web application built on the Streamlit framework. Users can input the required feature values into the model and receive real-time prognostic probabilities for individual patients, along with relevant visualizations. This tool aids clinicians in conducting personalized risk assessments and making informed decisions.

Statistical analysis

All data analyses and model construction were conducted in the Python 3.9.13 environment. Feature selection was performed using the Boruta algorithm, while the Delong test was completed in R 4.3.3. Continuous variables with skewed distributions were presented as medians with interquartile ranges (IQRs) and compared using the Mann-Whitney *U* test or the Kruskal-Wallis *H* test. Categorical variables were expressed as frequencies and percentages and compared using the chi-square test or Fisher's exact test. The area under the curve (AUC) was used to evaluate the model's predictive performance, and the optimal threshold was determined by maximizing the Youden index (sensitivity + specificity - 1). A two-tailed $P < 0.05$ was considered statistically significant.

Results

Patient characteristics

This retrospective study included 292 patients from The Second Hospital of Nanjing as the derivation cohort for constructing and validating the prognostic model. Among these 292 patients, 220 survived and 72 died. The mean age of the patients was 65.51 years (standard deviation 10.62, range 25–88 years), with 162 female patients, accounting for 55% of the cohort. The demographic and clinical characteristics of the deceased and surviving groups in the derivation cohort are presented in [Table 1](#). A comparison of demographic and clinical features between the training set, internal validation set, and external validation set can be found in [Supplementary Table S1](#). Details of the study design are displayed in [Figure 1](#).

Model development and performance comparison

First, the Boruta algorithm was used to select 12 of the most predictive features from the initial 24, with their importance ranked based on *Z*-scores, as shown in [Figure 2](#). Using clinical data collected at patient admission, 10 machine learning models (ML models) were constructed to predict patient prognosis. The primary evaluation metric for this study was AUC, with accuracy, precision, recall, and F1 score used as secondary metrics. The performance of these 10 models is detailed in [Supplementary Table S2](#). Among

all models, the top five performers were: KNN (AUC = 0.922, 12 features), SVM (AUC = 0.912, 11 features), RF (AUC = 0.911, 10 features), XGBoost (AUC = 0.911, 7 features), and Voting Classifier (AUC = 0.901, 7 features). The ROC curves for the top five ML models are shown in [Figure 3](#). As the number of included features increased, the AUC showed a general upward trend. For these five models, the performance (AUC) of the Random Forest model with varying numbers of features is shown in [Figure 4A](#), and the other four evaluation metrics are provided in [Supplementary Figure S1](#).

To further optimize the model, the Delong non-parametric test was used to compare models with different numbers of features. For example, the AUC difference between the KNN model (AUC = 0.922, 12 features) and KNN (AUC = 0.905, 7 features) was not significant ($p = 0.433$). Similarly, the difference between the RF model ($\Delta\text{AUC} = 0.007$, $p = 0.653$, 7 and 10 features) and the SVM model ($\Delta\text{AUC} = 0.051$, $p = 0.127$, 7 and 11 features) was also not statistically significant. Finally, we fixed the number of features at seven, which were: Consciousness, AST, U, CRRT, LDH, L, and MV. Considering the simplicity of the features, we ultimately selected the XGBoost model (AUC = 0.911, seven features) as the final predictive model.

External validation of the final model

The model performed equally well in the external validation set. The final XGBoost model achieved an AUC of 0.891 (95% CI: 0.786–0.977), an accuracy of 0.894 (95% CI: 0.837–0.952), a precision of 0.907 (95% CI: 0.866–0.955), a recall of 0.894 (95% CI: 0.837–0.952), and an F1 score of 0.884 (95% CI: 0.807–0.949), indicating that the model performed well in both internal and external validations. The DCA curve is shown in [Figure 4B](#).

Comparison of SOFA and APACHE II scores

The prognostic performance of the SOFA and APACHE II scores was further evaluated and compared with the final selected Seven-feature model. The model incorporating the SOFA score achieved an AUC of 0.872, accuracy of 0.870, precision of 0.815, recall of 0.611, and F1 score of 0.698. In contrast, the model incorporating the APACHE II score showed an AUC of 0.845, accuracy of 0.819, precision of 0.807, recall of 0.347, and F1 score of 0.485. In the low-risk group (SOFA score ≤ 6), the model achieved an AUC of 0.8376 (95% CI: 0.7571–0.9048). In the high-risk group (SOFA score ≥ 7), the AUC was 0.6000 (95% CI: 0.3617–0.8257). These results indicate that the model performed significantly better in the low-risk group compared to the high-risk group.

Model interpretation

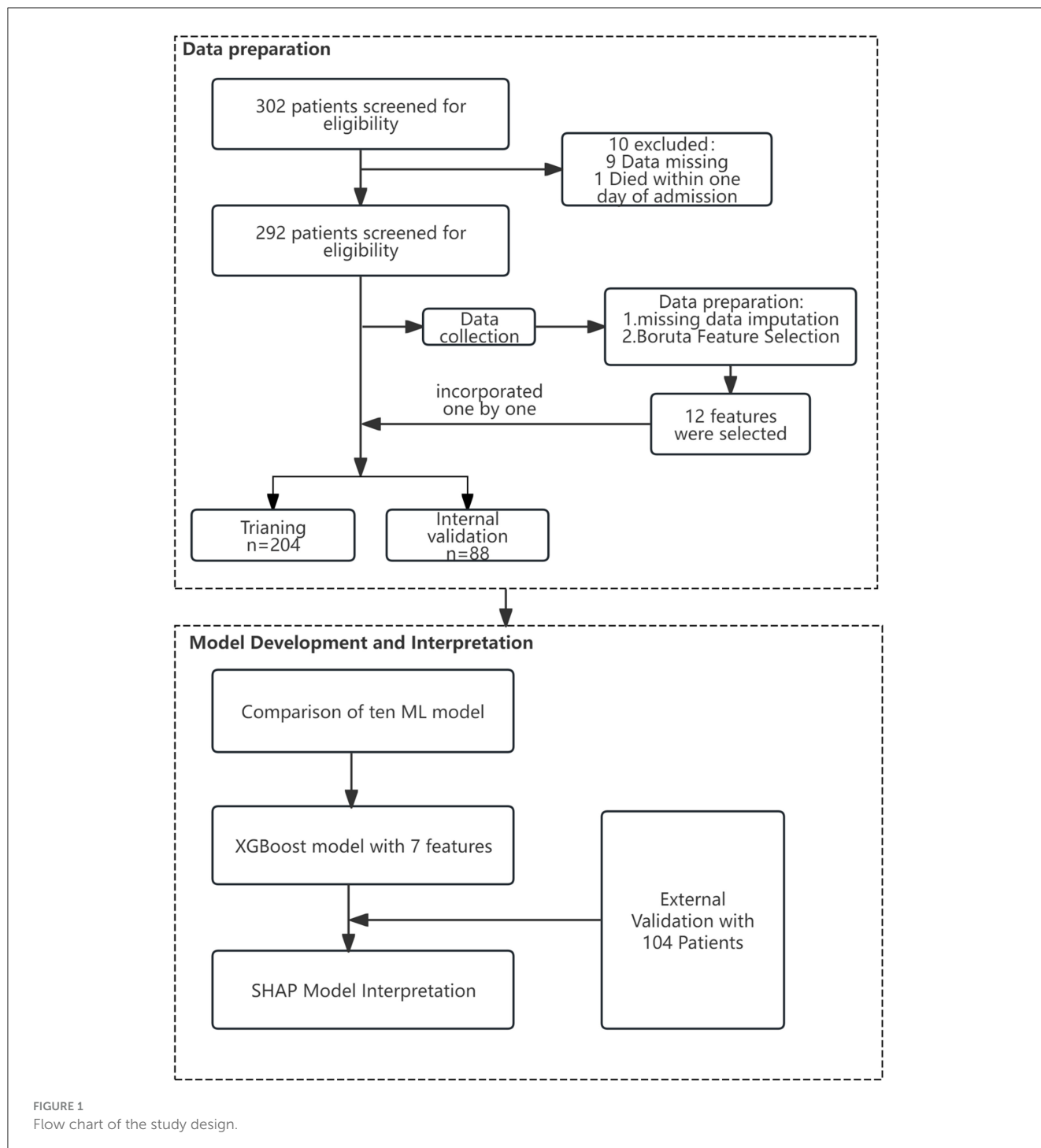
To make the model's prognoses more understandable for clinicians, this study used the SHAP algorithm to explain the importance of each predictor in the XGBoost model's prognoses. At the global interpretation level, the variable importance plot, ranked in descending order of contribution, demonstrated the

TABLE 1 Comparison of demographic and clinical characteristics between the survived and died groups.

Characteristic	Survived <i>N</i> = 220	Died <i>N</i> = 72	<i>p</i> -value ^a
Sex, n (%)			0.28
Female	126 (57%)	36 (50%)	
Male	94 (43%)	36 (50%)	
Age, median [Q1, Q3]	67.00 [56.00, 71.00]	70.00 [65.50, 75.00]	<0.001
APACHEII, median [Q1, Q3]	11.00 [8.00, 16.00]	18.00 [14.00, 23.00]	<0.001
SOFA, median [Q1, Q3]	3.00 [2.00, 4.00]	7.00 [5.00, 10.00]	<0.001
MV, n (%)			<0.001
Not applied	209 (95%)	35 (49%)	
Applied	11 (5.0%)	37 (51%)	
HF, n (%)			0.079
Not applied	214 (97%)	66 (92%)	
Applied	6 (2.7%)	6 (8.3%)	
CRRT, n (%)			<0.001
Not applied	207 (94%)	36 (50%)	
Applied	13 (5.9%)	36 (50%)	
Underlying disease, n (%)			0.22
No comorbidity	119 (54%)	33 (46%)	
Comorbidity	101 (46%)	39 (54%)	
Consciousness, n (%)			<0.001
No change	186 (85%)	15 (21%)	
Changed	34 (15%)	57 (79%)	
T (°C), median [Q1, Q3]	37.90 [36.80, 38.70]	38.50 [37.40, 38.85]	0.005
HR (bpm), median [Q1, Q3]	82.00 [69.50, 90.00]	89.00 [80.00, 104.50]	<0.001
MAP (mmHg), median [Q1, Q3]	82.00 [75.00, 91.00]	84.00 [72.50, 91.50]	0.50
WBC (10 ⁹ /L), median [Q1, Q3]	3.30 [2.07, 5.86]	3.12 [1.87, 4.45]	0.24
N (10 ⁹ /L), median [Q1, Q3]	2.23 [1.07, 4.41]	2.19 [1.15, 3.39]	0.53
L (10 ⁹ /L), median [Q1, Q3]	0.64 [0.43, 1.08]	0.47 [0.31, 1.04]	0.004
PLT (10 ⁹ /L), median [Q1, Q3]	53.00 [36.00, 69.50]	36.50 [27.50, 52.00]	<0.001
U (mmol/L), median [Q1, Q3]	4.95 [3.49, 7.33]	7.62 [5.51, 10.56]	<0.001
TBil (μmol/L), median [Q1, Q3]	7.90 [5.80, 10.70]	9.10 [6.75, 11.75]	0.014
ALT (U/L), median [Q1, Q3]	48.20 [32.35, 80.75]	85.70 [51.05, 146.65]	<0.001
AST (U/L), median [Q1, Q3]	105.20 [60.95, 186.50]	321.20 [194.20, 624.45]	<0.001
ALB (g/L), median [Q1, Q3]	35.30 [32.00, 38.10]	32.80 [30.00, 37.40]	0.007
LDH (U/L), median [Q1, Q3]	493.00 [357.00, 866.50]	1,099.00 [712.00, 1,919.00]	<0.001
ALP (U/L), median [Q1, Q3]	64.00 [48.85, 80.50]	69.00 [55.00, 100.00]	0.020
CR (μmol/L), median [Q1, Q3]	72.10 [59.05, 86.70]	88.50 [71.50, 119.50]	<0.001
D-Dimer (μg/L FEU), median [Q1, Q3]	1.77 [0.83, 4.14]	4.00 [1.78, 13.59]	<0.001
PNI, median [Q1, Q3]	39.33 [36.20, 42.53]	36.95 [33.10, 39.88]	<0.001

^aPearson's Chi-squared test; Wilcoxon rank sum test; Fisher's exact test.

MV, Mechanical Ventilation; HF, High Flow Oxygen Therapy; CRRT, Continuous Renal Replacement Therapy; T, Temperature; HR, Heart Rate; MAP, Mean Arterial Pressure; WBC, White Blood Cell Count; N, Neutrophil Count; L, Lymphocyte Count; PLT, Platelet Count; U, Urea; TBil, Total Bilirubin; ALT, Alanine Aminotransferase; AST, Aspartate Aminotransferase; ALB, Albumin; LDH, Lactate Dehydrogenase; ALP, Alkaline Phosphatase; CR, Creatinine; D-Dimer, D-Dimer; PNI, Prognostic Nutritional Index.



importance of each variable in the model, as shown in Figures 5A, B. In the model, death was defined as the positive class, and survival as the negative class. The study found that changes in the patient’s consciousness at admission were the most predictive of mortality risk, followed by AST, U, CRRT, LDH, L, and MV.

To further explore the positive and negative correlations of each feature with the prognosis outcomes, SHAP dependency plots were used to show how each feature influences the model’s output, with comparisons of the actual values and SHAP values

for the seven features, as shown in Figure 6A. In the context of mortality risk prognosis, features with SHAP values greater than zero were associated with an increased probability of the positive class, i.e., a higher risk of death. For instance, if a patient exhibited altered consciousness, the corresponding SHAP value would exceed zero, indicating that this feature contributed to a higher predicted likelihood of mortality. Similarly, urea levels above 4.966 mmol/L were associated with a shift in the model’s prognosis toward the “death” category, reflecting an elevated risk of death.

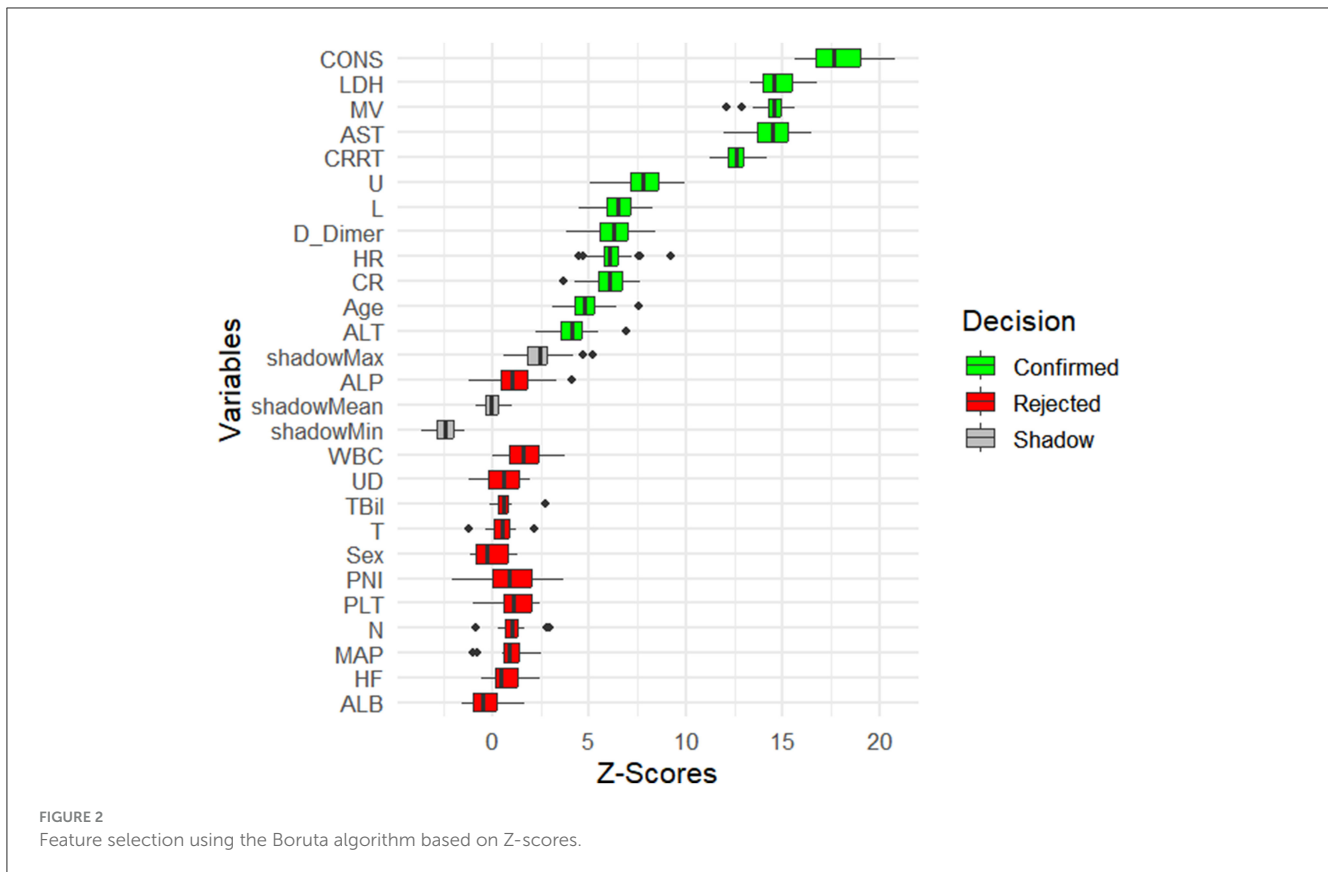


FIGURE 2
Feature selection using the Boruta algorithm based on Z-scores.

Local explanations provided insights into individual patient prognoses by visualizing the contribution of each feature. As illustrated in Figure 6B, one patient was predicted to have a 12% probability of mortality, placing them in the “death” category. In contrast, Figure 6C demonstrates another patient with an 88% probability of survival. The figures reveal that urea (U) had minimal influence on the prognosis for this patient. Instead, features such as consciousness, AST, LDH, CRRT, and MV contributed to pushing the prognosis toward the “survival” category, while the feature L shifted the prognosis toward the “death” category. These interpretations highlight how changes in specific features can meaningfully alter an individual’s predicted probability of survival or death.

Clinical application of the app

To support clinical use, this study developed a web application based on the Streamlit framework, offering clinicians a tool for real-time, individualized prognostic prognoses (Figure 7). By entering seven key feature values, the app automatically calculates the patient’s mortality risk and compares it against a predefined decision threshold. If the predicted probability exceeds the threshold, the patient is classified as being at high risk of death. Additionally, a waterfall plot is provided, with blue bars indicating features that push the outcome toward “survival,” and red bars representing those that push it toward “death.” The

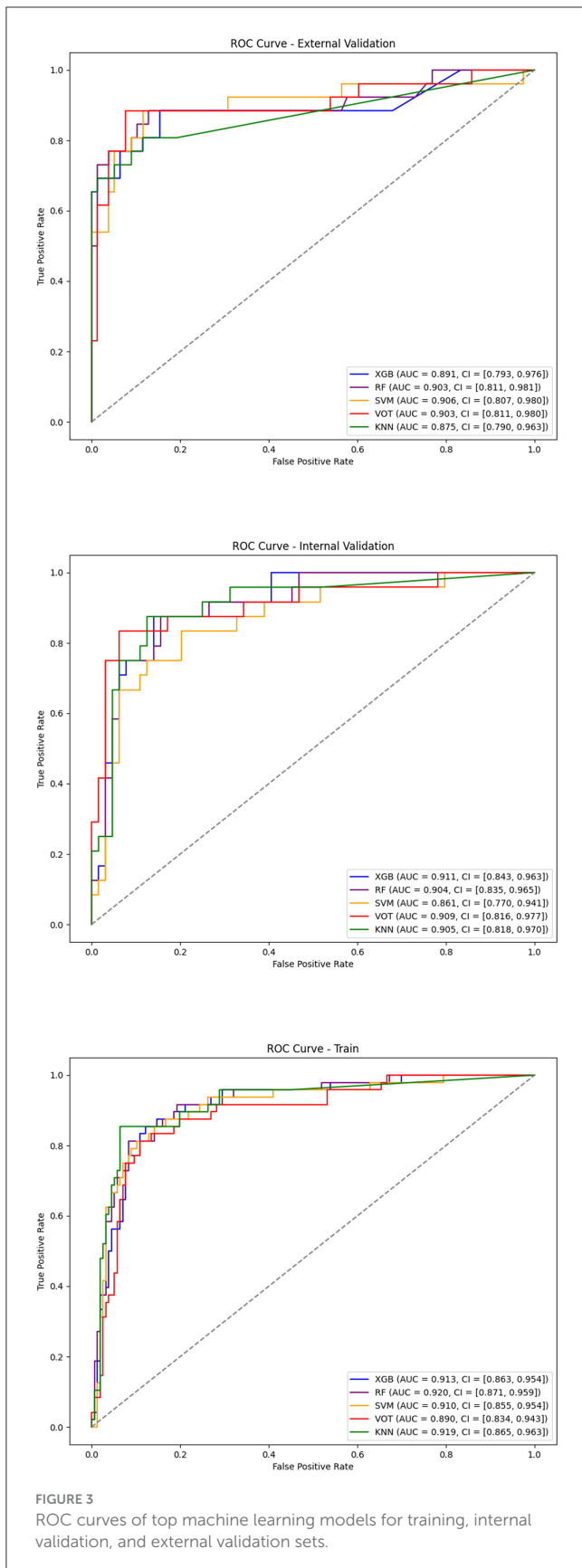
web application can be accessed at the following link: <https://jft6k52hfhpem8fqrlympd.streamlit.app/>.

Discussion

Main findings

This study successfully developed an XGBoost-based machine learning model, which demonstrated excellent predictive performance in both internal and external validation cohorts (Internal Validation: AUC = 0.911; External Validation: AUC = 0.891). These findings have significant clinical and theoretical implications. First, the model simplifies complex information by selecting a small number of high-value features, maintaining high accuracy while reducing the difficulty and cost of data acquisition, making it more feasible for clinical application. This is particularly important because existing prognostic models often rely on a large volume of laboratory and imaging data that are not easily accessible, limiting their use in actual clinical practice (Chen et al., 2023; Wang et al., 2024; Zheng et al., 2023; Zhou et al., 2024).

In this study, SOFA and APACHE II scores, which are traditional tools for assessing disease severity (Cheon et al., 2023; Kumar et al., 2020), were widely used in clinical settings and have been proven valuable for prognostic evaluation. In the external validation, the model based on the SOFA score performed well, with an AUC of 0.872, accuracy of 0.870, and



precision of 0.815. This indicates that the SOFA score effectively captures disease severity and accurately predicts patient prognosis. However, although the APACHE II score is widely recognized in clinical practice, its performance in this study was relatively inferior, with an AUC of only 0.845, a recall of 0.347, and an F1 score of 0.485. This suggests that the APACHE II score is inadequate for identifying high-risk patients (i.e., positive class samples). In contrast, the final model, which incorporated multiple features, significantly outperformed the single-score models in external validation, achieving an AUC of 0.891 and an F1 score of 0.884, demonstrating superior predictive power and ability to identify positive samples. Therefore, while SOFA and APACHE II scores have clinical value, our multi-feature model provides a more accurate and reliable tool for prognostic assessment of SFTS patients in a multicenter setting.

To further understand the contribution of individual features, SHAP analysis was performed. This analysis revealed the importance of each feature in the model's prognostic predictions, enhancing our understanding of the factors influencing patient outcomes. Among all the features included in the model, changes in consciousness and aspartate aminotransferase (AST) were identified as the most predictive features. AST, traditionally recognized as an indicator of liver dysfunction and systemic inflammatory response, typically reflects multi-organ damage when elevated, particularly in the context of viral infections. However, in our study, elevated AST levels were associated with improved survival, a finding that contrasts with conventional clinical expectations. This paradoxical result may be attributed to the unique characteristics of our study population or the specific disease context. For instance, mild to moderate AST elevation in our cohort might indicate an adaptive metabolic response or a marker of effective treatment rather than severe organ damage. Additionally, the interaction between AST and other clinical features (e.g., early interventions such as antiviral therapy) could have contributed to this observed association. While previous studies have shown that AST levels are closely related to multi-organ failure in SFTS patients, our findings suggest a more nuanced relationship between AST elevation and patient outcomes, highlighting the need for further investigation into the underlying mechanisms (Du et al., 2024; He et al., 2021; Wang et al., 2023). Altered consciousness may be an early signal of central nervous system involvement, indicating that the disease has progressed to a critical stage, aligning with reports in other studies that highlight neurological symptoms as indicators of poor prognosis (He et al., 2021; Xu et al., 2021).

Notably, the Boruta-selected features (e.g., consciousness, AST, CRRT) align with key predictors identified by clinical experts, such as neurological status and organ dysfunction markers. However, the algorithm also highlighted less intuitive features (e.g., urea), which may reflect novel interactions in SFTS pathophysiology. This synergy between data-driven selection and clinical expertise strengthens the model's biological plausibility.

Similarly, the importance of lactate dehydrogenase (LDH) and high-flow oxygen therapy (HF) was validated in the model. LDH is a marker of cell damage and tissue necrosis, and its

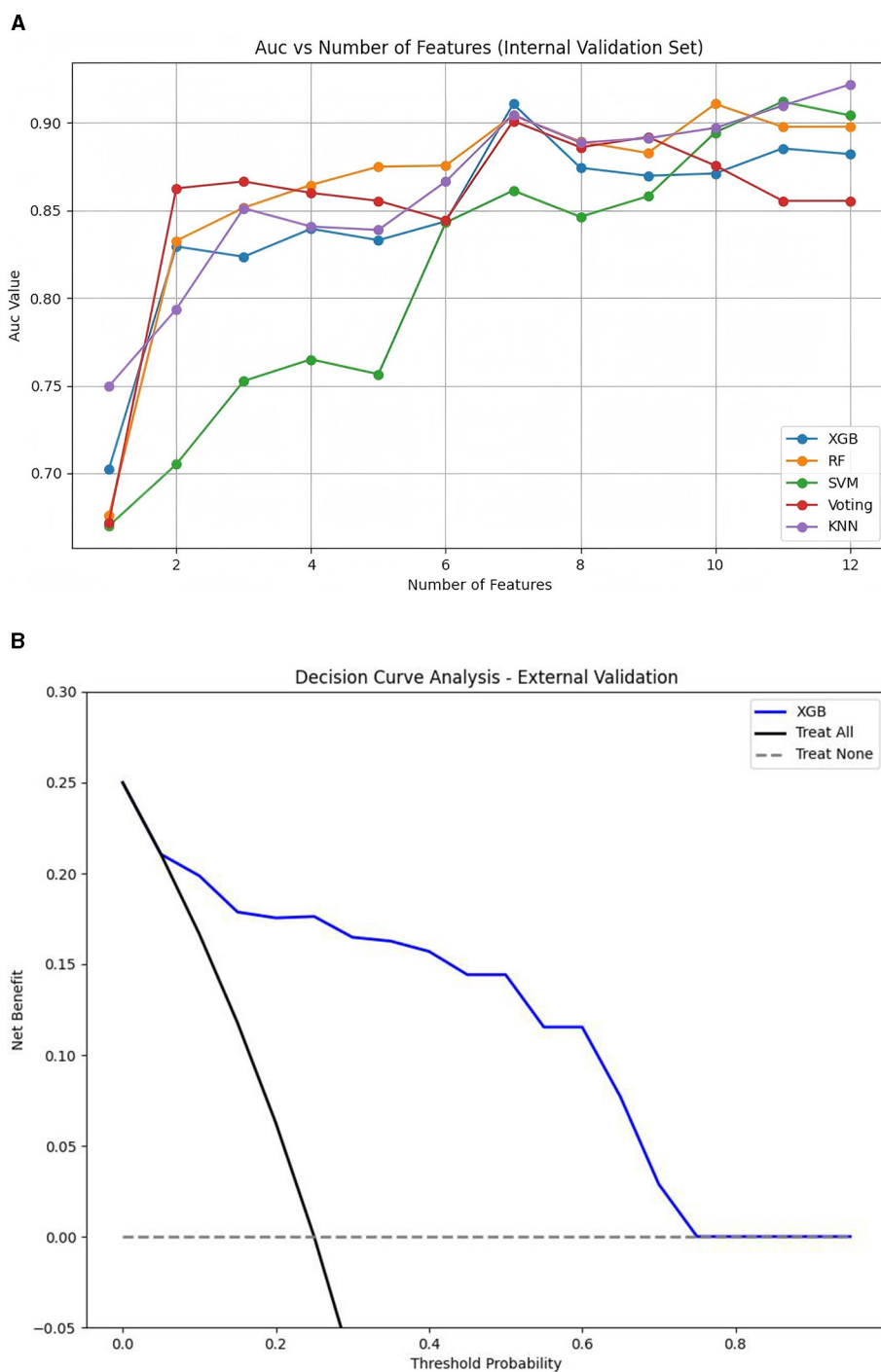
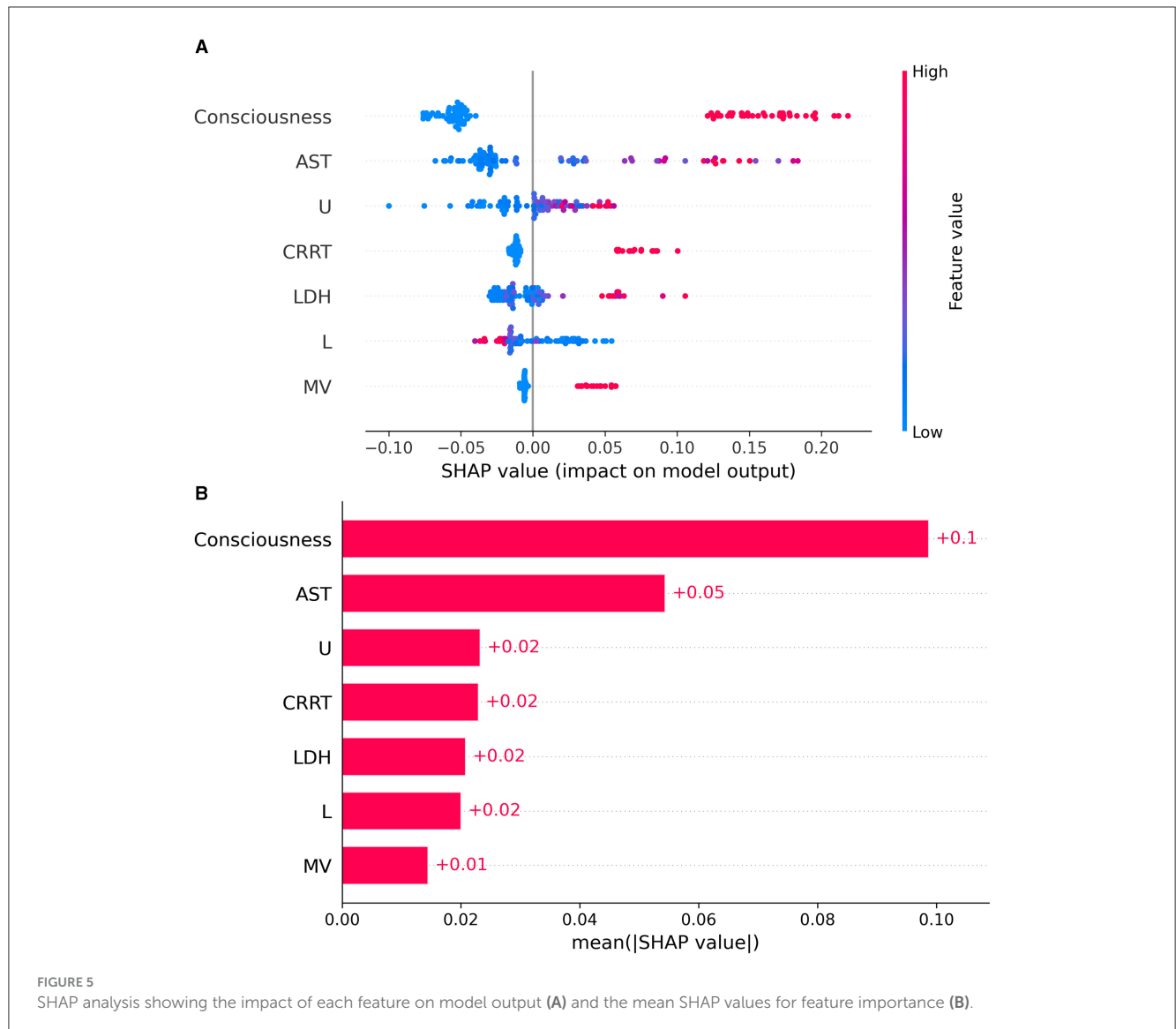


FIGURE 4
AUC vs. number of features for different models in the internal validation set (A) and decision curve analysis (DCA) for the final XGBoost model in the external validation set (B).

elevation reflects the extent of viral damage to multiple organs (He et al., 2021; Jia et al., 2017). The application of high-flow oxygen therapy, as a common respiratory support measure, indicates the severity of respiratory system impairment, and this variable showed a significant impact on prognostic prognosis in the model (Jia et al., 2017). Furthermore, the use of continuous renal replacement

therapy (CRRT), often seen in patients with renal failure, not only reflected the severity of the disease but also indirectly indicated the level of medical intervention the patient received. This suggests that future model development may benefit from incorporating more features related to therapeutic interventions, which could further enhance the model's predictive ability (Li et al., 2024a).



Limitations

Despite the multi-level analysis demonstrating the potential of the XGBoost model for SFTS prognosis, several limitations need to be addressed.

First, data source limitations: The data in this study were collected from two centers in eastern China. The epidemiological characteristics and viral strains of SFTS may vary by region, and the model's performance in other areas (such as Japan and South Korea) or different healthcare settings remains unclear. Previous studies have shown that machine learning models may perform differently across regions and ethnic backgrounds, suggesting that future research should include more geographically and ethnically diverse data for validation (Cui et al., 2024; Li et al., 2024b). Additionally, while the model incorporates post-discharge follow-up data to predict long-term outcomes, its current validation focuses on in-hospital mortality and short-term survival. Future studies should explicitly evaluate performance in long-term recovery scenarios.

Second, the “black box” issue of machine learning models: Although SHAP provided some level of explanation for the feature contributions in the XGBoost model, this interpretation still relies on feature importance rankings and does not fully reveal the complex interactions between features. For example, the explanations for CRRT and LDH are primarily focused on their individual effects, without exploring how these features interact with other factors in a multi-feature context (Hong et al., 2022; Manikandan et al., 2024). Therefore, future research could incorporate causal inference models or complex network analysis to further elucidate the interactions and mechanisms between features.

Third, the inherent bias of retrospective studies: As a retrospective cohort study, the electronic medical record data used may contain incomplete or inconsistent records, potentially introducing bias. Recent evidence-based studies suggest that constructing multi-factor confounding models can reduce bias in retrospective analyses, but this approach was not fully adopted in our study (Li et al., 2024b; Manikandan et al.,

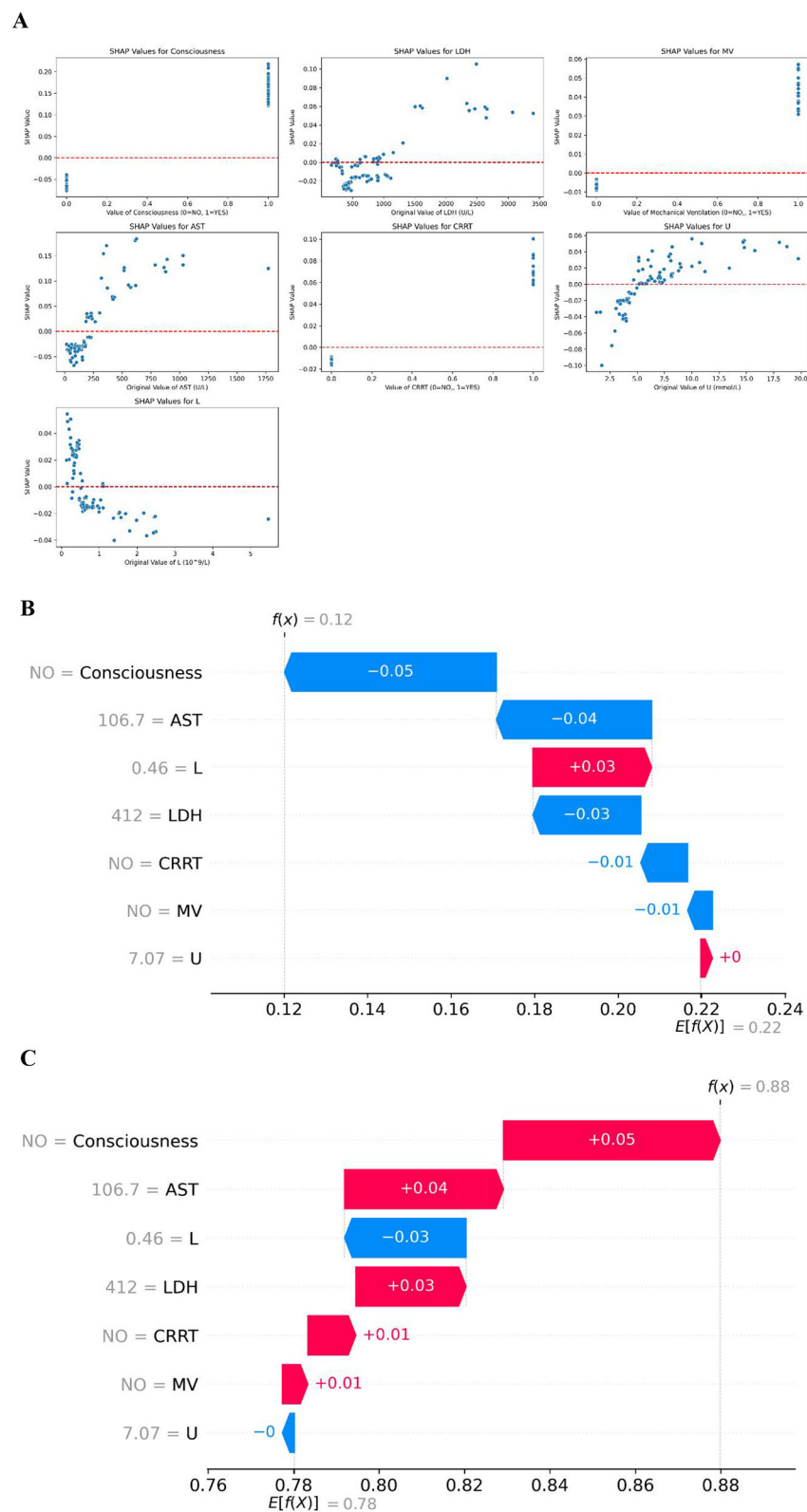


FIGURE 6 SHAP dependency plots and local explanation for individual prognoses. **(A)** SHAP dependency plots for the seven selected features, illustrating the relationship between feature values and their impact on model output. **(B)** SHAP local explanation for a patient with a predicted 12% probability of death, showing the contribution of each feature to the prognosis. **(C)** SHAP local explanation for a patient with a predicted 88% probability of survival, showing the contribution of each feature to the prognosis.

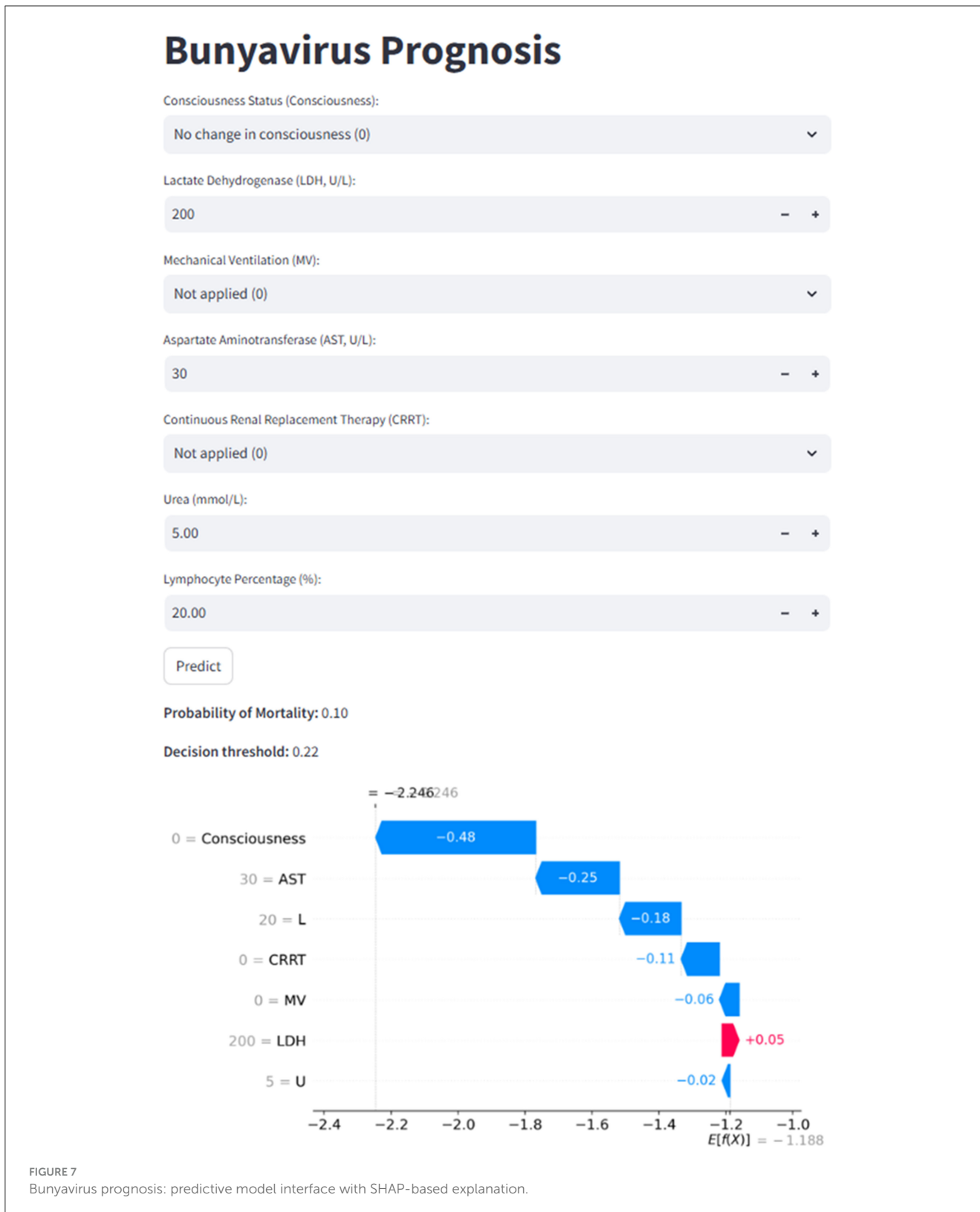


FIGURE 7
 Bunyavirus prognosis: predictive model interface with SHAP-based explanation.

2024). Additionally, differences in treatment management between hospitals may affect patient outcomes, and this heterogeneity was not fully quantified in the current model (Manikandan et al., 2024). Future research should incorporate more treatment-related information.

Finally, while the model demonstrated real-time applicability with rapid computation (0.5 s per prognosis) and minimal feature requirements, its deployment in clinical settings must adhere to data privacy regulations (e.g., China's PIPL and GDPR). All patient data were anonymized and encrypted during this study, and future

implementations will require strict access controls and ongoing compliance audits.

Conclusion

This study developed an XGBoost-based prognostic model for SFTS, providing both global and local explanations through SHAP, demonstrating efficient and transparent predictive capabilities. The development and validation of the model not only confirmed XGBoost's efficiency and accuracy in handling complex medical data but also highlighted the potential of incorporating interpretability into machine learning models. Although the model performed well on data from two hospitals, further validation of its external generalizability is necessary, especially in different regions and larger sample sizes. Future studies should consider incorporating multi-center, large-scale prospective cohort studies and optimizing the model with more treatment-related features. Additionally, combining complex network analysis and causal inference methods could further reveal the interactions between key features and the pathological mechanisms of the disease, providing a more scientific basis for personalized medicine and targeted interventions.

Data availability statement

The raw data and analysis code can be obtained by contacting the corresponding author, Yi-Shan Zheng, at 284159264@qq.com.

Ethics statement

The studies involving humans were approved by the Ethics Committee of The Second Hospital of Nanjing (2023-L-S-023) and the Ethics Committee of The First Affiliated Hospital of Wannan Medical College (202124). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

J-SX: Conceptualization, Visualization, Writing – original draft. KY: Conceptualization, Data curation, Writing – original draft. BQ:

Writing – original draft. JX: Data curation, Writing – original draft. Y-SZ: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by The General Program of Jiangsu Commission of Health (No. M2021088); Nanjing Infectious Disease Clinical Medical Center; and Innovation Center for Infectious Disease of Jiangsu Province (No. CXZX202232).

Acknowledgments

We appreciated the contributions of all participants in this study and all study staff.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2025.1557922/full#supplementary-material>

References

- Azodi, C. B., Tang, J., and Shiu, S. H. (2020). Opening the black box: interpretable machine learning for geneticists. *Trends Genet.* 36, 442–455. doi: 10.1016/j.tig.2020.03.005
- Berkelmans, G. F. N., Read, S. H., Gudbjörnsdóttir, S., Wild, S. H., Franzen, S., van der Graaf, Y., et al. (2022). Population median imputation was noninferior to complex approaches for imputing missing values in cardiovascular prediction models in clinical practice. *J. Clin. Epidemiol.* 145, 70–80. doi: 10.1016/j.jclinepi.2022.01.011
- Chen, X., Sun, J., Wang, Z., Fan, Y., and Qiao, J. (2023). "Deep learning and XGBoost based prediction algorithm for esophageal varices," in *Signal and Information Processing, Networking and Computers. Lecture Notes in Electrical Engineering, Vol 917*, eds J. Sun, Y. Wang, M. Huo, and L. Xu (Singapore: Springer). doi: 10.1007/978-981-19-3387-5_134
- Cheon, S., Kwon, H., and Ryu, J.-A. (2023). Evaluating the prognostic efficacy of scoring systems in neurocritical and neurosurgical care: an insight into APACHE II, SOFA, and GCS. *J. Neurointens. Care* 6, 123–129. doi: 10.32587/jnic.2023.00696
- Cui, Q., Pu, J., Li, W., Zheng, Y., Lin, J., Liu, L., et al. (2024). Study on risk factors of impaired fasting glucose and development of a prediction model based on extreme gradient boosting algorithm. *Front. Endocrinol.* 15:1368225. doi: 10.3389/fendo.2024.1368225

- Du, Q., Yu, J., Chen, Q., Chen, X., Jiang, Q., Deng, L., et al. (2024). Clinical characteristics and influencing factors of severe fever with thrombocytopenia syndrome complicated by viral myocarditis: a retrospective study. *BMC Infect. Dis.* 24:240. doi: 10.1186/s12879-024-09096-4
- Fan, Y., Liu, M., and Sun, G. (2023). An interpretable machine learning framework for diagnosis and prognosis of COVID-19. *PLoS ONE* 18:e0291961. doi: 10.1371/journal.pone.0291961
- Feretakis, G., Sakagianni, A., Anastasiou, A., Kapogianni, I., Bazakidou, E., Koufopoulos, P., et al. (2024). Integrating shapley values into machine learning techniques for enhanced predictions of hospital admissions. *Appl. Sci.* 14:5925. doi: 10.3390/app14135925
- Gai, Z. T., Zhang, Y., Liang, M. F., Jin, C., Zhang, S., Zhu, C. B., et al. (2012). Clinical progress and risk factors for death in severe fever with thrombocytopenia syndrome patients. *J. Infect. Dis.* 206, 1095–1102. doi: 10.1093/infdis/jis472
- Haider, R. Z., Ujjan, I. U., Khan, N. A., Urrechaga, E., and Shamsi, T. S. (2022). Beyond the in-practice CBC: the research CBC parameters-driven machine learning predictive modeling for early differentiation among leukemias. *Diagnostics* 12:138. doi: 10.3390/diagnostics12010138
- He, F., Zheng, X., and Zhang, Z. (2021). Clinical features of severe fever with thrombocytopenia syndrome and analysis of risk factors for mortality. *BMC Infect. Dis.* 21:1253. doi: 10.1186/s12879-021-06946-3
- Hong, W., Zhou, X., Jin, S., Lu, Y., Pan, J., Lin, Q., et al. (2022). A comparison of XGBoost, random forest, and nomograph for the prediction of disease severity in patients with COVID-19 pneumonia: implications of cytokine and immune cell profile. *Front. Cell. Infect. Microbiol.* 12:819267. doi: 10.3389/fcimb.2022.819267
- Huang, X., Ding, S., Jiang, X., Pang, B., Zhang, Q., Li, C., et al. (2019). Detection of SFTS virus RNA and antibodies in severe fever with thrombocytopenia syndrome surveillance cases in endemic areas of China. *BMC Infect. Dis.* 19:476. doi: 10.1186/s12879-019-4068-2
- Huyut, M. T. (2023). Automatic detection of severely and mildly infected COVID-19 patients with supervised machine learning models. *Ing. Rech. Biomed.* 44:100725. doi: 10.1016/j.irbm.2022.05.006
- Huyut, M. T., and Huyut, Z. (2021). Forecasting of oxidant/antioxidant levels of COVID-19 patients by using expert models with biomarkers used in the diagnosis/prognosis of COVID-19. *Int. Immunopharmacol* 100:108127. doi: 10.1016/j.intimp.2021.108127
- Huyut, M. T., and Huyut, Z. (2023). Effect of ferritin, INR, and D-dimer immunological parameters levels as predictors of COVID-19 mortality: a strong prediction with the decision trees. *Heliyon* 9:e14015. doi: 10.1016/j.heliyon.2023.e14015
- Huyut, M. T., and Ilkbahar, F. (2021). The effectiveness of blood routine parameters and some biomarkers as a potential diagnostic tool in the diagnosis and prognosis of Covid-19 disease. *Int. Immunopharmacol* 98:107838. doi: 10.1016/j.intimp.2021.107838
- Huyut, M. T., and Kocaturk, I. (2022). The effect of some symptoms and features during the infection period on the level of anxiety and depression of adults after recovery from COVID-19. *Curr. Psychiat. Res. Re.* 18, 151–163. doi: 10.2174/2666082218666220325105504
- Huyut, M. T., and Ustundag, H. (2022). Prediction of diagnosis and prognosis of COVID-19 disease by blood gas parameters using decision trees machine learning model: a retrospective observational study. *Med. Gas Res.* 12, 60–66. doi: 10.4103/2045-9912.326002
- Huyut, M. T., and Velichko, A. (2022). Diagnosis and prognosis of COVID-19 disease using routine blood values and LogNNet neural network. *Sensors* 22:4820. doi: 10.3390/s22134820
- Huyut, M. T., and Velichko, A. (2023). LogNNet model as a fast, simple and economical AI instrument in the diagnosis and prognosis of COVID-19. *MethodsX* 10:102194. doi: 10.1016/j.mex.2023.102194
- Huyut, M. T., Velichko, A., and Belyaev, M. (2022). Detection of risk predictors of COVID-19 mortality with classifier machine learning models operated with routine laboratory biomarkers. *Appl. Sci.* 12:12180. doi: 10.3390/app122312180
- Jia, B., Yan, X., Chen, Y., Wang, G., Liu, Y., Xu, B., et al. (2017). A scoring model for predicting prognosis of patients with severe fever with thrombocytopenia syndrome. *PLoS Negl. Trop. Dis.* 11:e0005909. doi: 10.1371/journal.pntd.0005909
- Kawakami, E., Tabata, J., Yanaiharu, N., Ishikawa, T., Koseki, K., Iida, Y., et al. (2019). Application of artificial intelligence for preoperative diagnosis and prognostic prediction in epithelial ovarian cancer based on blood biomarkers. *Clin. Cancer Res.* 25, 3006–3015. doi: 10.1158/1078-0432.CCR-18-3378
- Kim, K. H., Yi, J., Kim, G., Choi, S. J., Jun, K. I., Kim, N. H., et al. (2013). Severe fever with thrombocytopenia syndrome, South Korea, 2012. *Emerg. Infect. Dis.* 19, 1892–1894. doi: 10.3201/eid1911.130792
- Kumar, S., Gattani, S. C., Baheti, A. H., and Dubey, A. (2020). Comparison of the performance of APACHE II, SOFA, and mNUTRIC scoring systems in critically ill patients: a 2-year cross-sectional study. *Indian J. Crit. Care Med.* 24, 1057–1061. doi: 10.5005/jp-journals-10071-23549
- Li, X. Q., Wang, R. Q., Wu, L. Q., and Chen, D. M. (2024a). Transfer learning-enabled outcome prediction for guiding CRRT treatment of the pediatric patients with sepsis. *BMC Med. Inform. Decis. Mak.* 24:266. doi: 10.1186/s12911-024-02623-y
- Li, Y., Du, C., Ge, S., Zhang, R., Shao, Y., Chen, K., et al. (2024b). Hematoma expansion prediction based on SMOTE and XGBoost algorithm. *BMC Med. Inform. Decis. Mak.* 24:172. doi: 10.1186/s12911-024-02561-9
- Liu, Q., He, B., Huang, S. Y., Wei, F., and Zhu, X. Q. (2014). Severe fever with thrombocytopenia syndrome, an emerging tick-borne zoonosis. *Lancet Infect. Dis.* 14, 763–772. doi: 10.1016/S1473-3099(14)70718-2
- Liu, Z., Jiang, Z., Zhang, L., Xue, X., Zhao, C., Xu, Y., et al. (2023). A model based on meta-analysis to evaluate poor prognosis of patients with severe fever with thrombocytopenia syndrome. *Front. Microbiol.* 14:1307960. doi: 10.3389/fmicb.2023.1307960
- Manikandan, G., Pragadeesh, B., Manojkumar, V., Karthikeyan, A. L., Manikandan, R., and Gandomi, A. H. (2024). Classification models combined with boruta feature selection for heart disease prediction. *Informatics in Medicine Unlocked* 44:101442. doi: 10.1016/j.imu.2023.101442
- Mertoglu, C., Huyut, M. T., Olmez, H., Tosun, M., Kantarci, M., and Coban, T. A. (2022). COVID-19 is more dangerous for older people and its severity is increasing: a case-control study. *Med. Gas Res.* 12, 51–54. doi: 10.4103/2045-9912.325992
- Mosca, E., Szigeti, F., Tragianni, S., Gallagher, D., and Groh, G. (2022). *SHAP-Based Explanation Methods: A Review for NLP Interpretability*. Gyeongju, Republic of Korea. Available online at: <https://aclanthology.org/2022.coling-1.406/>
- Santos-Silva, M. A., Sousa, N., and Sousa, J. C. (2024). Artificial intelligence in routine blood tests. *Front. Med. Eng.* 2:1369265. doi: 10.3389/fmede.2024.1369265
- Soerensen, P. D., Christensen, H., Gray Worsoe Laursen, S., Hardahl, C., Brandslund, I., and Madsen, J. S. (2022). Using artificial intelligence in a primary care setting to identify patients at risk for cancer: a risk prediction model based on routine laboratory tests. *Clin. Chem. Lab. Med.* 60, 2005–2016. doi: 10.1515/cclm-2021-1015
- Tahir Huyut, M., Huyut, Z., Ilkbahar, F., and Mertoglu, C. (2022). What is the impact and efficacy of routine immunological, biochemical and hematological biomarkers as predictors of COVID-19 mortality? *Int. Immunopharmacol* 105:108542. doi: 10.1016/j.intimp.2022.108542
- Tian, D., Yan, H. J., Huang, H., Zuo, Y. J., Liu, M. Z., Zhao, J., et al. (2023). Machine learning-based prognostic model for patients after lung transplantation. *JAMA Netw. Open* 6:e2312022. doi: 10.1001/jamanetworkopen.2023.12022
- Üstündag, H., Mertoglu, C., and Huyut, M. T. (2023). Oxyhemoglobin dissociation curve in COVID-19 patients. *Meandros Med. Dental J.* 24, 58–64. doi: 10.4274/meandros.galenos.2023.87049
- Velichko, A., Huyut, M. T., Belyaev, M., Izotov, Y., and Korzun, D. (2022). Machine learning sensors for diagnosis of COVID-19 disease using routine blood values for internet of things application. *Sensors* 22:7886. doi: 10.3390/s22207886
- Wang, B., He, Z., Yi, Z., Yuan, C., Suo, W., Pei, S., et al. (2021). Application of a decision tree model in the early identification of severe patients with severe fever with thrombocytopenia syndrome. *PLoS ONE* 16:e0255033. doi: 10.1371/journal.pone.0255033
- Wang, L., Liu, Y., Qu, R., and Zou, Z. (2023). Serum mAST/ALT ratio had high predictive value for adverse outcome of severe fever with thrombocytopenia syndrome with severe condition. *BMC Infect. Dis.* 23:168. doi: 10.1186/s12879-023-08121-2
- Wang, Y., Gao, Z., Zhang, Y., Lu, Z., and Sun, F. (2024). Early sepsis mortality prediction model based on interpretable machine learning approach: development and validation study. *Intern. Emerg. Med.* doi: 10.1007/s11739-024-03732-2
- Wu, C. C., Yeh, W. C., Hsu, W. D., Islam, M. M., Nguyen, P. A. A., Poly, T. N., et al. (2019). Prediction of fatty liver disease using machine learning algorithms. *Comput. Methods Programs Biomed.* 170, 23–29. doi: 10.1016/j.cmpb.2018.12.032
- Wu, J., Zan, X., Gao, L., Zhao, J., Fan, J., Shi, H., et al. (2019). A machine learning method for identifying lung cancer based on routine blood indices: qualitative feasibility study. *JMIR Med. Inform.* 7:e13476. doi: 10.2196/13476
- Xu, Y., Shao, M., Liu, N., Dong, D., Tang, J., and Gu, Q. (2021). Clinical feature of severe fever with thrombocytopenia syndrome (SFTS)-associated encephalitis/encephalopathy: a retrospective study. *BMC Infect. Dis.* 21:904. doi: 10.1186/s12879-021-06627-1
- Yan, W., Shi, H., He, T., Chen, J., Wang, C., Liao, A., et al. (2021). Employment of artificial intelligence based on routine laboratory results for the early diagnosis of multiple myeloma. *Front. Oncol.* 11:608191. doi: 10.3389/fonc.2021.608191
- Yu, F., and Morita, K. (2019). “Laboratory diagnosis for SFTS,” in *Severe Fever with Thrombocytopenia Syndrome*, ed M. Saijo (Singapore: Springer). doi: 10.1007/978-981-13-9562-8_14
- Yu, X. J., Liang, M. F., Zhang, S. Y., Liu, Y., Li, J. D., Sun, Y. L., et al. (2011). Fever with thrombocytopenia associated with a novel bunyavirus in China. *New Engl. J. Med.* 364, 1523–1532. doi: 10.1056/NEJMoa1010095
- Zheng, J., Li, J., Zhang, Z., Yu, Y., Tan, J., Liu, Y., et al. (2023). Clinical data based XGBoost algorithm for infection risk prediction of patients with decompensated cirrhosis: a 10-year (2012–2021) multicenter retrospective case-control study. *BMC Gastroenterol.* 23:310. doi: 10.1186/s12876-023-02949-3
- Zhou, S., Lu, Z., Liu, Y., Wang, M., Zhou, W., Cui, X., et al. (2024). Interpretable machine learning model for early prediction of 28-day mortality in ICU patients with sepsis-induced coagulopathy: development and validation. *Eur. J. Med. Res.* 29:14. doi: 10.1186/s40001-023-01593-7