Original Article

Constructing a predictive model for early-onset sepsis in neonatal intensive care unit newborns based on SHapley Additive exPlanations explainable machine learning

Xuefeng Tan[1], Xiufang Zhang[1], Jie Chai[1], Wenjuan Ji[1], Jinling Ru[1], Cuilin Yang[1], Wenjing Zhou[1], Jing Bai[1], Yueling Xiong[2]

[1]Department of Laboratory Medicine, The People's Hospital, Bozhou, China; [2]Translational Medicine Center, The Second Affiliated Hospital, Wannan Medical College, Wuhu, China

Contributions: (I) Conception and design: X Tan, X Zhang; (II) Administrative support: Y Xiong; (III) Provision of study materials or patients: X Tan, J Bai, C Yang; (IV) Collection and assembly of data: X Tan, J Chai, W Zhou; (V) Data analysis and interpretation: X Tan, W Ji, J Ru; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Yueling Xiong, MSc. Translational Medicine Center, The Second Affiliated Hospital, Wannan Medical College, 10 Kangfu Road, Wuhu 241001, China. Email: xyl2022@wnmc.edu.cn.

Background: The clinical characteristics of neonatal sepsis (NS) are subtle and non-specific, posing a serious threat to the lives of newborn infants. Early-onset sepsis (EOS) is sepsis that occurs within 72 hours after birth, with a high mortality rate. Identifying key factors of NS and conducting early diagnosis are of great practical significance. Thus, we developed a robust machine learning (ML) model for the early prediction of EOS in neonates admitted to the neonatal intensive care unit (NICU), investigated the pivotal risk factors associated with EOS development, and provided interpretable insights into the model's predictions.

Methods: A retrospective cohort study was conducted. This includes 668 newborns (EOS and non-EOS) admitted to the NICU of Bozhou People's Hospital from January to December 2023, excluding 72 newborns born more than three days ago and 166 newborns with medical record data missing more than 30%. Finally, 430 newborns (EOS and non-EOS) were included in the study. Clinical case data were meticulously analyzed, and the dataset was randomly partitioned, allocating 75% for model training and the remaining 25% for test. Data preprocessing was meticulously performed using R language, and the least absolute shrinkage and selection operator (LASSO) regression was implemented to select salient features, mitigating the risk of overfitting. Six ML models were leveraged to forecast the incidence of EOS in neonates. The predictive performance of these models was rigorously evaluated using the receiver operating characteristic (ROC) curve and precision-recall (PR) curve. Furthermore, the SHapley Additive exPlanations (SHAP) framework was employed to provide intuitive explanations for the predictions made by the Categorical Boosting (CatBoost) model, which emerged as the top performer.

Results: The ROC area under the curve (ROCAUC) of six ML models, CatBoost, random forest (RF), eXtreme Gradient Boosting (XGBoost), multilayer perceptron (MLP), support vector machine (SVM), logistic regression (LR) all exceeded 0.900 on the test set. Especially the CatBoost model exhibited superior performance, with favorable outcomes in calibration, decision curve analysis (DCA), and learning curves. Notably, the ROCAUC attained 0.975, and the area under the PR curve (PRAUC) reached 0.947, signifying a high degree of predictive accuracy. Utilizing the SHAP method, seven key features were identified and ranked by their importance: respiratory rate (RR), procalcitonin (PCT), nasal congestion (NC), yellow staining (YS), white blood cell count (WBC), fever, and amniotic fluid turbidity (AFT).

Conclusions: By constructing a precision-oriented ML model and harnessing the SHAP method for interpretability, this study effectively identified crucial risk factors for EOS development in neonates. This approach enables early prediction of EOS risk, thereby facilitating timely and targeted clinical interventions for precise diagnosis and treatment.

Keywords: Newborns; machine learning (ML); early-onset sepsis (EOS); SHapley Additive exPlanations (SHAP)

Submitted Jul 18, 2024. Accepted for publication Nov 05, 2024. Published online Nov 26, 2024.
doi: 10.21037/tp-24-278
View this article at: https://dx.doi.org/10.21037/tp-24-278

Introduction

Neonatal sepsis (NS) refers to nonspecific bacteremia and a systemic inflammatory response syndrome occurring in newborns post-birth and remains a significant cause of neonatal mortality, with an 8% mortality rate (1,2). Early-onset sepsis (EOS), occurring within the first 72 hours, often presents with subtle, non-specific symptoms, leading to rapid progression and higher risks of delayed treatment and adverse outcomes (3,4), making timely diagnosis challenging. The condition can rapidly deteriorate into septic shock and multi-organ failure, leaving survivors with potential severe sequelae (5,6). An international study found a 3.2% mortality rate for full-term newborns with EOS (7). Despite advances reducing EOS through improved obstetrical care and antibiotics, newborns remain vulnerable due to immature immune systems. The lack of clear clinical markers complicates diagnosis and treatment, impacting outcomes (8,9). Understanding the clinical nuances of EOS and achieving early diagnosis is crucial for optimal disease management. Currently, blood culture serves as the primary diagnostic tool for NS, despite its limitations (10,11). The positivity rate of blood cultures is influenced by various factors, such as specimen volume, prior antibiotic exposure, and time to sample processing, etc. These factors lead to a slow turnaround time for blood cultures, typically two days, and low sensitivity, especially in EOS cases. They are even less effective at detecting slow-growing or fastidious bacteria (12). Due to blood volume limitations in newborns, especially those with low birth weights, blood culture sensitivity is reduced; thus, each sample must be at least 1 mL (13), and a negative result does not exclude sepsis. Early identification and prompt treatment of EOS are crucial for improving neonatal outcomes.

Highlight box

Key findings

- The respiratory rate of neonatal intensive care unit newborns is the most important feature for model prediction, lower respiratory rates are more likely to be predicted as early-onset sepsis (EOS) by the model. However, the findings of this study still need further validation as they have not been confirmed elsewhere.

What is known and what is new?

- Currently, most studies on neonatal EOS are based on the criterion of positive blood culture. There are many factors that affect the positive rate of blood culture, such as specimen volume, previous antibiotic exposure, and sample processing time, which lead to longer blood culture time; the sensitivity is low, especially in children with EOS, and the detection rate of bacteria with slow growth rate and harsh culture conditions is even lower. Due to limitations in the amount of blood taken from newborns, especially those with low, extremely low, or ultra-low birth weight, the sensitivity of blood culture is worse, and negative results cannot completely rule out sepsis, leading to delayed diagnosis and treatment in some patients.
- The innovation of this study is to establish the features required for machine learning models, such as clinical manifestations and laboratory indicators of newborns, which can be obtained at the time of patient admission. Then, by inputting these features into the model, prediction results can be obtained, providing reference for clinical doctors to quickly and accurately diagnose neonatal EOS. Secondly, applying the SHapley Additive exPlanations to explain the model makes its predictions more transparent and reliable, allowing us to better understand the prediction process of the model.

What is the implication, and what should change now?

- It can assist clinicians in the early diagnosis and treatment of neonatal EOS, reducing the adverse consequences caused by delayed diagnosis and treatment, and facilitating a more rational allocation of medical resources in clinical practice.

Machine learning (ML), a pivotal subset and the most prevalent branch within the artificial intelligence (AI) realm, has demonstrated substantial maturity in enhancing diagnostic precision and expediency, augmenting physician decision-making, forecasting disease trajectories, and facilitating drug discovery (14-16). ML excels at navigating through voluminous and intricate medical datasets, distilling pivotal insights through iterative training processes. These ML models meticulously discern and categorize disease characteristics and patterns, enabling precise disease identification and classification (17-19). The efficacy of these predictive models is rigorously evaluated using meticulous metrics, including accuracy, sensitivity, specificity, and area under the curve (AUC) (20,21). A higher score in these metrics typically translates to superior predictive performance, underscoring the reliability and robustness of the ML models. Within the ML taxonomy, various learning paradigms exist, notably supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning (22). Notably, supervised and unsupervised learning methodologies hold sway in the realm of medical data processing (23-26). Supervised learning relies heavily on labeled training data to construct predictive models for unlabeled test data, while unsupervised learning processes unlabeled data, providing unique insights into disease complexity (27).

In the realm of NS diagnosis, many researchers have utilized ML techniques to conduct detailed analysis of objective NS data. Meeus *et al.* (28) for instance, leveraged AI technology to devise a predictive model for late-onset sepsis (LOS) and necrotizing enterocolitis (NEC) among infants under neonatal intensive care. Their ML model, rigorously evaluated on an independent test set of 148 patients, exhibited a commendable overall sensitivity of 69% for the onset of LOS/NEC, underscoring its potential clinical utility. Similarly, Lyra *et al.* (29) trained two comprehensive public datasets encompassing adult and neonatal patients using ML techniques. Subsequently, they validated their model using synthetic clinical data, achieving notable performance metrics, including a receiver operating characteristic AUC (ROCAUC) of 0.91 and a precision-recall AUC (PRAUC) of 0.38. Nonetheless, it is noteworthy that the ML methodologies utilized in these studies exhibit a degree of homogeneity, suggesting that there may be avenues for optimization and improvement in prediction outcomes. As such, future research endeavors should explore diverse ML techniques and strategies, aiming to refine prediction models and further elevate the diagnostic accuracy of NS. In this study, we endeavor to construct a diverse array of ML models, utilizing neonates with clinically confirmed NS within the first 72 hours of life. Using clinical diagnosis as the gold standard, we included EOS and non-EOS patients as our study cohort. Furthermore, we employ the SHAP (SHapley Additive exPlanations) (30) technique to gain insightful explanations of these models. The aim of this study is to achieve the following goals: (I) early prediction of EOS; (II) identification risk factor; and (III) apply SHAP for model insight. Thus, providing reference for clinical diagnosis and treatment (30,31). We present this article in accordance with the TRIPOD reporting checklist (available at https://tp.amegroups.com/article/view/10.21037/tp-24-278/rc).

Methods

Research object

A retrospective analysis was conducted on the clinical case data of 668 neonatal intensive care unit (NICU) neonatal patients admitted to Bozhou People's Hospital from January to December 2023, excluding 72 newborns born more than three days ago and 166 newborns with missing medical records exceeding 30%, a total of 430 newborns (EOS and non-EOS) were included in the study. These patients were randomly split into a training set of 322 cases and a test set of 108 cases at a ratio of 75% and 25%, respectively. The training set was further divided into an EOS group of 79 cases and a non-EOS group of 243 cases based on the occurrence of EOS in neonates. Given the issue of data imbalance, after filling in missing data, a random oversampling technique was applied to balance the data, resulting in 243 cases in both the EOS and non-EOS groups within the training set. Inclusion criteria: (I) patients meeting the EOS-related clinical diagnostic criteria outlined in the "Expert Consensus on Diagnosis and Treatment of Neonatal Sepsis" in 2019 (12). (II) Patients with complete data or missing data less than 30% of the total. Clinical diagnostic criteria for neonatal EOS: diagnosis is made when clinical abnormalities are present along with any of the following: (i) 2 positive non-specific blood tests; (ii) cerebrospinal fluid (CSF) examination showing changes of purulent meningitis; (iii) detection of pathogenic bacterial DNA in blood. Exclusion criteria: (I) patients with missing data exceeding 30% of the total; (II) patients beyond 3 days of age. The specific data preprocessing and experimental process are shown in *Figure 1*. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of the Bozhou People's Hospital (approval No. 2024-149). This study has obtained informed consent from the patients. The data we used in our study were all anonymous information of patients, and the unique identification information of patients was removed to ensure their privacy and security.

Data collection

Electronic medical records of NICU neonatal patients in the hospital were retrospectively collected. The clinical data of neonates included gestational age (GA), birth weight (BW), respiratory rate (RR), pulse, sex, yellow staining (YS), nasal congestion (NC), skin cyanosis (SC), and fever. The

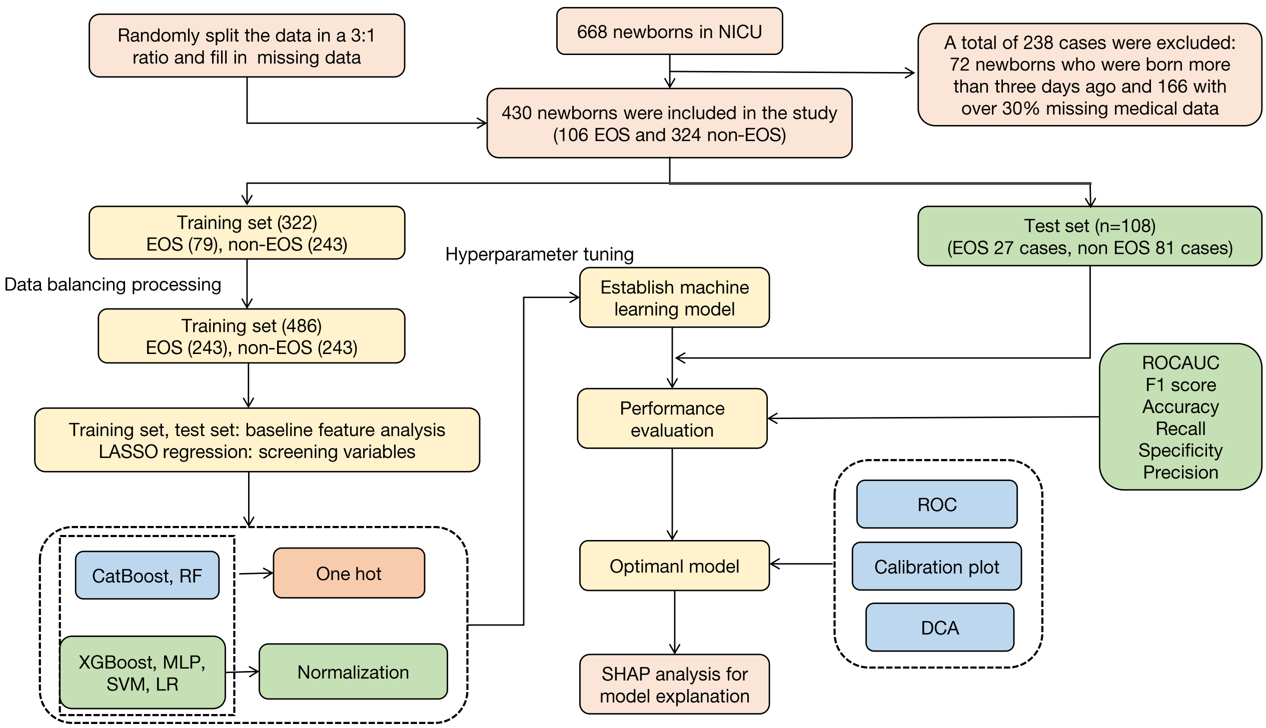

Figure 1 Data preprocessing and experimental flowchart. NICU, neonatal intensive care unit; EOS, early-onset sepsis; LASSO, least absolute shrinkage and selection operator; CatBoost, Categorical Boosting; RF, random forest; XGBoost, eXtreme Gradient Boosting; MLP, multilayer perceptron; SVM, support vector machine; LR, logistic regression; SHAP, SHapley Additive exPlanations; ROC, receiver operating characteristic; DCA, decision curve analysis; ROCAUC, receiver operating characteristic area under the curve.

clinical data of pregnant mothers encompassed amniotic fluid turbidity (AFT), abnormal thyroid function (ATF), history of miscarriage (HOM), anemia, gestational diabetes (GD), and fever during pregnancy (FDP). Laboratory indicators included procalcitonin (PCT), white blood cell count (WBC), and C-reactive protein (CRP).

Statistical analysis

Using R version 4.3.2, we undertook a comprehensive data preprocessing and analysis pipeline. For missing values, we imputed numerical variables with the median and categorical variables with the mode. We balanced the dataset using random oversampling. To compare baseline characteristics between the EOS and non-EOS groups, we used the Shapiro-Wilk test for normality, presenting non-normally distributed data as M (Q1, Q3) and comparing groups with the Wilcoxon rank-sum test. Categorical data were expressed as [n (%)] and analyzed with the Chi-squared test, with statistical significance set at $P<0.05$. Subsequently, we performed one-hot encoding on discrete categorical data and standardized continuous data. We used the least absolute shrinkage and selection operator (LASSO) regression with 10-fold cross-validation to select features. In Python 3.8.0, we analyzed the correlations among the retained features. We built six ML models: support vector machine (SVM), logistic regression (LR), random forest (RF), eXtreme Gradient Boosting (XGBoost), multilayer perceptron (MLP), and Categorical Boosting (CatBoost). For RF and CatBoost, continuous data remained unstandardized to aid interpretation. Each model underwent hyperparameter tuning with 5-fold cross-validation and was validated using the test set. Model evaluation utilized ROC curves, PR curves, and metrics from the confusion matrix. The best-performing algorithm was selected to build the final model, and decision curve analysis (DCA), calibration curves, and learning curves were plotted. SHAP was used to interpret the model results.

Results

Comparison of baseline data between neonatal EOS group and non-EOS group patients in the training set

A comparison of baseline data between the neonatal EOS group and the non-EOS group in the training set revealed significant differences in RR, pulse, YS, NC, fever, AFT, ATF, FDP, PCT, WBC, CRP. In both groups, there were no significant differences in GA, BW, sex, SC, HOM, anemia, and GD, with statistical significance defined as P<0.05. These findings are summarized in *Table 1*.

Feature selection using LASSO regression on the training set

Employing the EOS status from the training set as the outcome variable, we employed Lasso regression (with the selection criterion set as one standard error to the right of the minimum lambda) to screen the features. Ultimately, seven features with non-zero regression coefficients were included in the model: RR, PCT, NC, YS, WBC, fever, and AFT. The 10-fold cross-validation error curve for the Lasso regression is shown in *Figure 2A*, and the LASSO regression coefficient path is depicted in *Figure 2B*.

Model construction and ROC curve analysis

Utilizing the seven features selected by LASSO regression in Python 3.8.0, we constructed SVM, LR, RF, XGBoost, MLP, and CatBoost ML models. Hyperparameter tuning was performed for each model using grid search with 5-fold cross-validation to build the optimal models. Subsequently, ROC curves for predicting EOS on both the training and

Table 1 Comparison of baseline data between neonatal EOS group and non-EOS group patients in the training set

Clinical data	Total (n=486)	EOS (n=243)	Non-EOS (n=243)	P
Neonate				
GA	39.4 [38.3, 40.1]	39.4 [38.3, 40.1]	39.3 [38.1, 40.2]	0.52
BW	3.3 [3.0, 3.6]	3.3 [3.1, 3.5]	3.3 [2.9, 3.6]	0.35
RR	54 [50, 62]	50 [48, 53]	62 [55, 62]	<0.001
Pulse	134 [132, 138]	132 [132, 138]	136 [132, 140]	<0.001
Sex				0.79
Male	264 (54.3)	130 (53.5)	134 (55.1)	
Female	222 (45.7)	113 (46.5)	109 (44.9)	
YS				<0.001
No	436 (89.7)	236 (97.1)	200 (82.3)	
Yes	50 (10.3)	7 (2.9)	43 (17.7)	
NC				<0.001
No	355 (73.0)	206 (84.8)	149 (61.3)	
Yes	131 (27.0)	37 (15.2)	94 (38.7)	
SC				0.09
No	468 (96.3)	230 (94.7)	238 (97.9)	
Yes	18 (3.7)	13 (5.3)	5 (2.1)	
Fever				<0.001
No	446 (91.8)	209 (86.0)	237 (97.5)	
Yes	40 (8.2)	34 (14.0)	6 (2.5)	
Pregnant mother				
AFT				<0.001
No	417 (85.8)	229 (94.2)	188 (77.4)	
Yes	69 (14.2)	14 (5.8)	55 (22.6)	

Table 1 (*continued*)

Table 1 (*continued*)

Clinical data	Total (n=486)	EOS (n=243)	Non-EOS (n=243)	P
ATF				0.047
No	420 (86.4)	218 (89.7)	202 (83.1)	
Yes	66 (13.6)	25 (10.3)	41 (16.9)	
HOM				0.20
No	281 (57.8)	133 (54.7)	148 (60.9)	
Yes	205 (42.2)	110 (45.3)	95 (39.1)	
Anemia				0.26
No	187 (38.5)	87 (35.8)	100 (41.2)	
Yes	299 (61.5)	156 (64.2)	143 (58.8)	
GD				0.48
No	352 (72.4)	180 (74.1)	172 (70.8)	
Yes	134 (27.6)	63 (25.9)	71 (29.2)	
FDP				0.04
No	49 (10.1)	17 (7.0)	32 (13.2)	
Yes	437 (89.9)	226 (93.0)	211 (86.8)	
Laboratory indicators				
PCT (ng/mL)	0.56 [0.20, 1.88]	1.33 [0.54, 2.86]	0.28 [0.12, 0.60]	<0.001
WBC ($\times 10^9$/L)	14.48 [12.89, 18.74]	15.55 [14.43, 21.00]	14.48 [11.76, 16.74]	<0.001
CRP (mg/L)	2.14 [0.00, 6.00]	2.14 [0.80, 6.20]	2.14 [0.00, 5.55]	0.01

Data are presented as M [Q1, Q3] and n (%). P<0.05 indicates a statistically significant difference. EOS, early-onset sepsis; GA, gestational age; BW, birth weight; RR, respiratory rate; YS, yellow staining; NC, nasal congestion; SC, skin cyanosis; AFT, amniotic fluid turbidity; ATF, abnormal thyroid function; HOM, history of miscarriage; GD, gestational diabetes; FDP, fever during pregnancy; PCT, procalcitonin; WBC, white blood cell count; CRP, C-reactive protein.

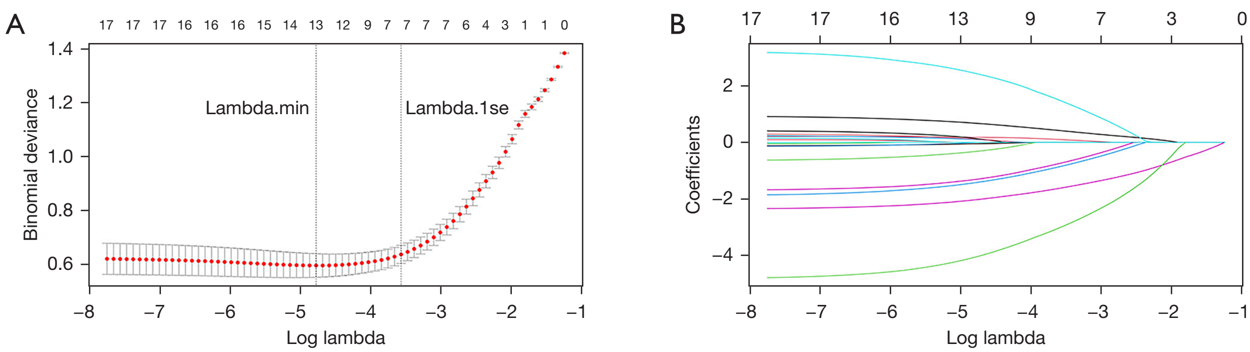

Figure 2 LASSO regression cross validation error curve and coefficient path diagram. (A) LASSO regression 10-fold cross-validation error curve. (B) LASSO regression coefficient path plot. lambda.min, lambda value at minimum deviation; lambda.1se, one standard error to the right of the minimum lambda; LASSO, least absolute shrinkage and selection operator.

test sets were plotted for each model. Notably, the CatBoost model demonstrated the highest ROCAUC values of 0.999 and 0.975 for the training and test sets, respectively, as shown in *Figure 3A,3B*. The performance metrics including ROCAUC, F1 score, accuracy, recall, specificity, and precision were calculated from the confusion matrices of multiple models, with CatBoost consistently outperforming the others, as summarized in *Table 2*.

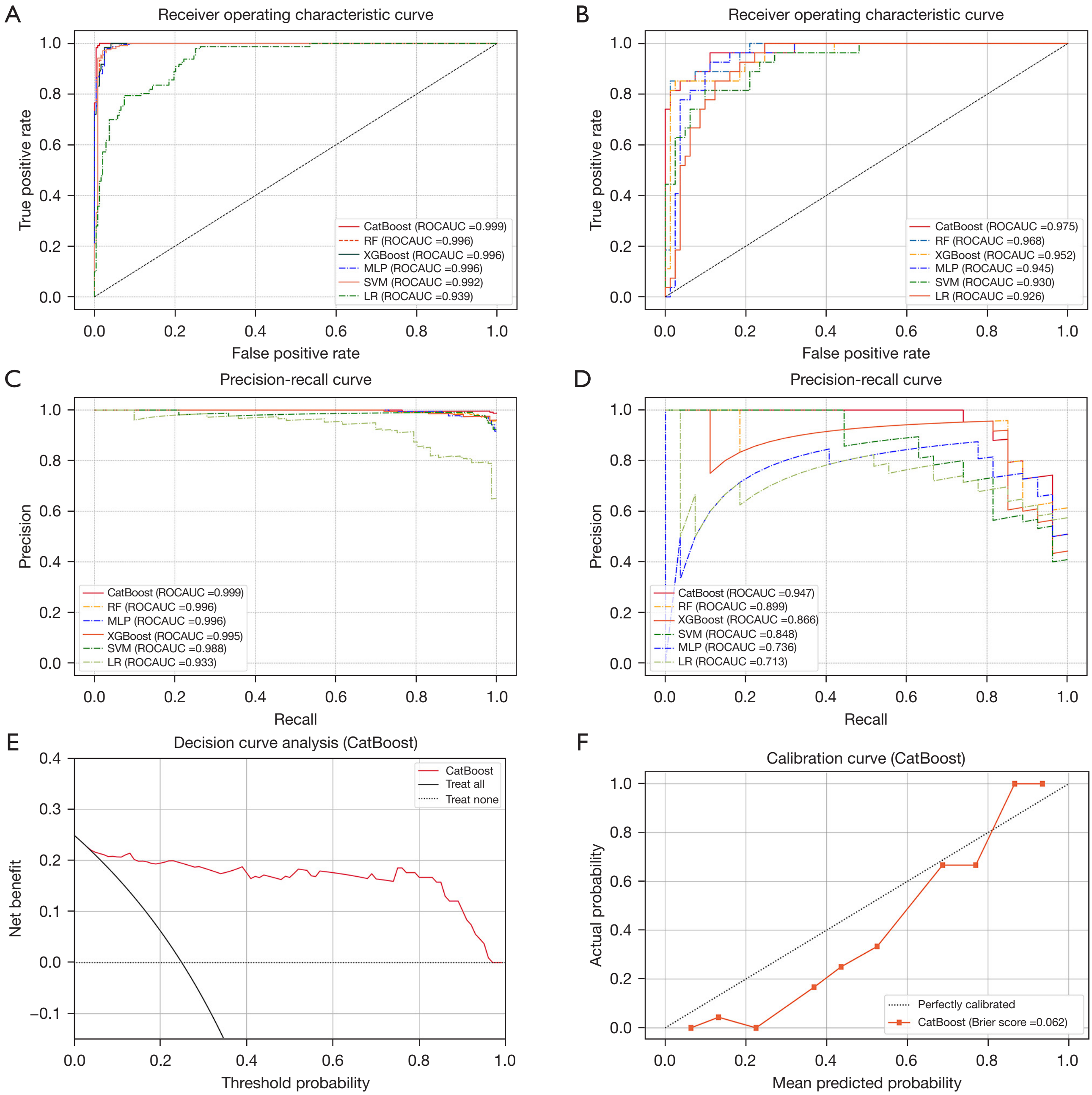

Figure 3 ROC and PR curves of multiple models and DCA and calibration curves of CatBoost mode. (A) ROC curves of multiple models in the training set. (B) ROC curves of multiple models in the test set. (C) PR curves of multiple models in the training set. (D) PR curves of multiple models in the test set. (E) Test set DCA curve. (F) Test set calibration curve. CatBoost, Categorical Boosting; ROCAUC, receiver operating characteristic area under the curve; RF, random forest; XGBoost, eXtreme Gradient Boosting; MLP, multilayer perceptron; SVM, support vector machine; LR, logistic regression; PR, precision-recall.

PR curve analysis

The PR curves for EOS occurrences in the training and test sets of various models are plotted, with the CatBoost model achieving the highest PRAUC scores of 0.999 and 0.947 in the training and test sets respectively, as shown in *Figure 3C,3D*.

Table 2 Performance indicators of multiple models with EOS occurring in the test sets

Performance index	ROCAUC	F1 score	Accuracy	Recall	Specificity	Precision
CatBoost	0.975	0.842	0.917	0.889	0.926	0.800
RF	0.968	0.821	0.907	0.852	0.926	0.793
XGBoost	0.952	0.821	0.907	0.852	0.926	0.793
MLP	0.945	0.772	0.880	0.815	0.901	0.733
SVM	0.930	0.755	0.880	0.741	0.926	0.769
LR	0.926	0.737	0.861	0.778	0.889	0.700

EOS, early-onset sepsis; ROCAUC, receiver operating characteristic area under the curve; CatBoost, Categorical Boosting; RF, random forest; XGBoost, eXtreme Gradient Boosting; MLP, multilayer perceptron; SVM, support vector machine; LR, logistic regression.

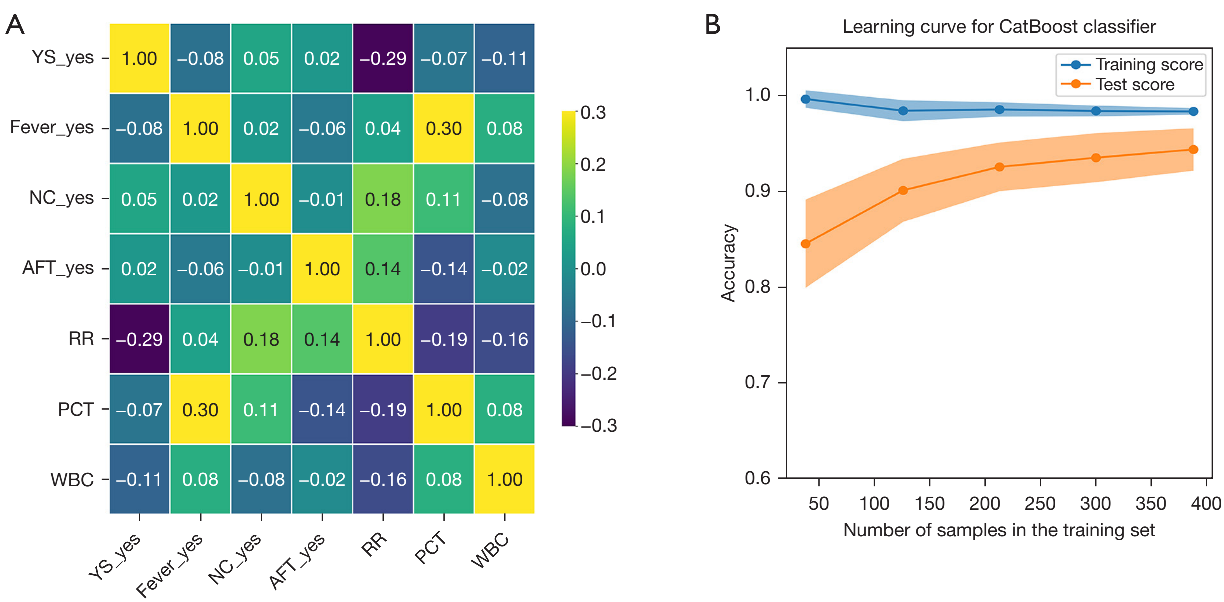

Figure 4 Correlation analysis and learning curve of CatBoost model features. (A) Feature correlation analysis chart. (B) Learning curve. YS, yellow staining; NC, nasal congestion; AFT, amniotic fluid turbidity; RR, respiratory rate; PCT, procalcitonin; WBC, white blood cell count; CatBoost, Categorical Boosting.

Analysis of DCA curve and calibration curve

Draw the DCA curve and calibration curve for the CatBoost model on the test set for EOS occurrences. The DCA curve shows that the model can provide a higher net benefit for clinical applications over a wide range of threshold probabilities (with threshold probability on the x-axis and net benefit on the y-axis), as shown in *Figure 3E*. The calibration curve demonstrates good agreement between the predicted curve and the ideal curve (with predicted probability on the x-axis and actual probability on the y-axis), as illustrated in *Figure 3F*.

Learning curve analysis

The correlation coefficients among the seven features were all below 0.3, and the correlation among features is illustrated in *Figure 4A*. The learning curve of the CatBoost model on the training set indicates good model fitting (with the number of training set samples on the x-axis and the model accuracy on the y-axis). The two learning curves gradually converge as the number of training samples increases, suggesting that the model is well-fitted and possesses good generalization ability, as shown in *Figure 4B*.

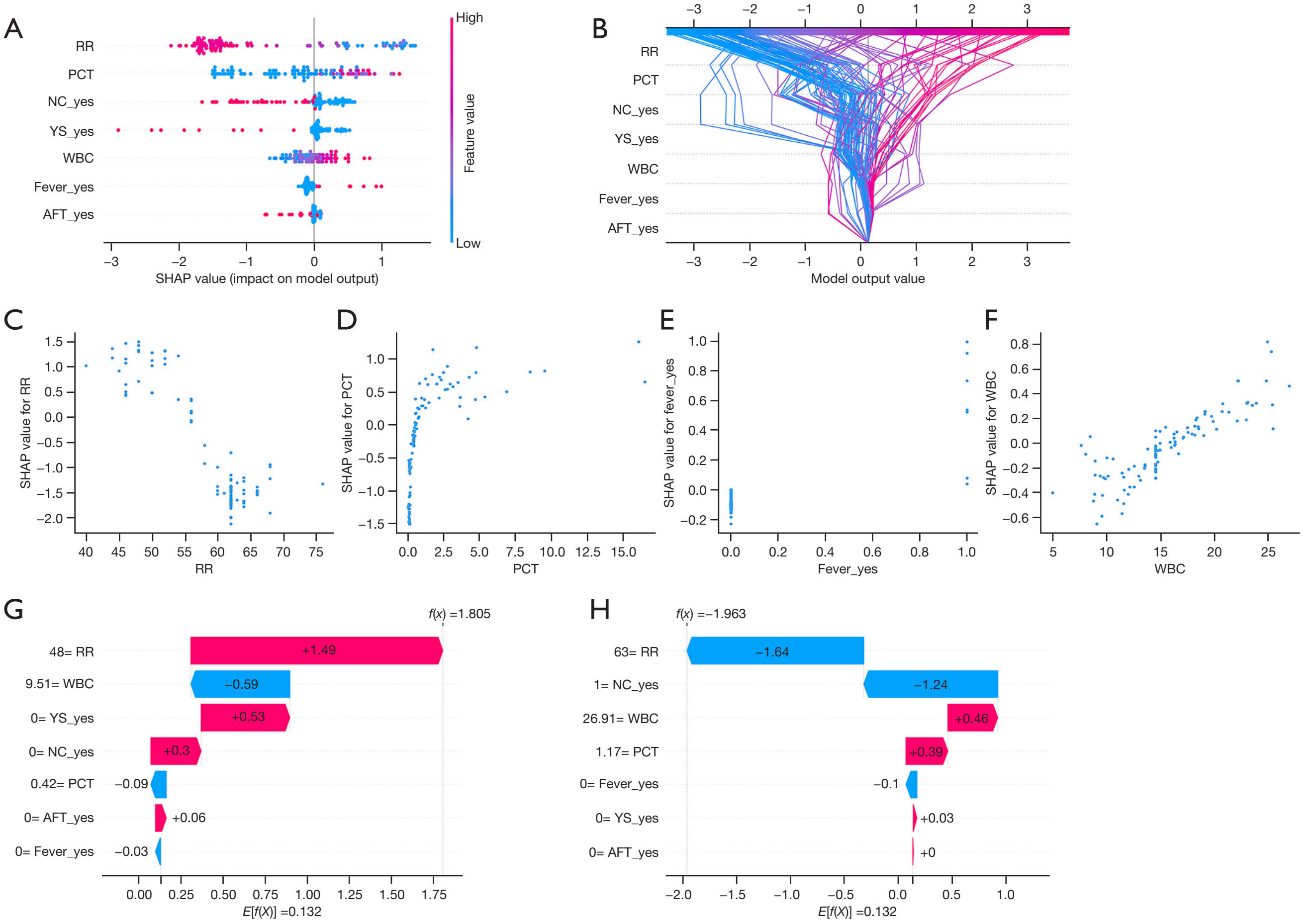

Figure 5 Explanation of CatBoost model based on SHAP values. (A) Summary plot based on SHAP values. (B) Decision diagram based on SHAP values. (C-F) The dependency graphs between the model features (RR, PCT, fever, WBC) based on SHAP values and the predicted results, respectively. (G) The SHAP waterfall plot predicted by the model for positive patients. (H) The SHAP waterfall plot predicted by the model for negative patients. RR, respiratory rate; PCT, procalcitonin; NC, nasal congestion; YS, yellow staining; WBC, white blood cell count; AFT, amniotic fluid turbidity; SHAP, SHapley Additive exPlanations.

Model interpretation based on SHAP values

Analysis of feature importance and its impact on prediction results of CatBoost model based on SHAP values

The SHAP method offers a visual explanation scheme by plotting a summary plot based on SHAP values to visualize the individual impact of all features on the prediction results, as shown in *Figure 5A*. Here, the SHAP values are on the horizontal axis, and the features are on the vertical axis. Each point represents the SHAP value of a feature for a sample, with these points categorized and arranged by features. The redder the color, the higher the feature value, and the bluer the color, the lower the feature value. The results show that the most important feature ranked first for model prediction is RR, indicating its greatest contribution to the model prediction. When a patient's RR is high, it mostly makes a negative contribution to the model prediction, and when it is low, it mostly makes a positive contribution.

Decision analysis of CatBoost model based on SHAP values

Plotting a decision graph based on SHAP values is used to visualize the process of the model making decisions for all patient data instances in the entire test set. As can be seen from *Figure 5B*, the model makes all predictions based on the contributions of individual features. From bottom to top in the decision-making process, the contributions of features are added sequentially until the final prediction is made at the top.

Feature dependence analysis of CatBoost model based on SHAP values

Drawing a dependency graph between the features (RR, PCT, fever, WBC) and prediction outcomes based on SHAP values. *Figure 5C* shows that when the RR is below 55 breaths per minute, it contributes positively to the model's prediction, and negatively above 55 breaths per minute. *Figure 5D* indicates that PCT contributes negatively to the model's prediction when it is below approximately 0.8 ng/mL, and positively above 0.8 ng/mL. *Figure 5E* displays that fever contributes positively to the model's prediction, while the absence of fever contributes negatively. *Figure 5F* shows that WBC mostly contributes negatively to the model's prediction when it is below 15×10^9/L, and mostly positively above 15×10^9/L.

Single sample explanation

Drawing a waterfall plot for a single patient's risk of EOS based on SHAP values (where red indicates positive contributions from features, blue indicates negative contributions, and the width of the bars represents the magnitude of each feature's contribution). *Figure 5G* shows the SHAP waterfall plot for a patient predicted to be positive by the model. It can be observed that despite WBC featuring a significant negative contribution in the model's prediction, RR, YS, and NC all contributed positively, ultimately leading to the patient being predicted as positive. *Figure 5H* presents the SHAP waterfall plot for a patient predicted to be negative by the model. In the process of predicting this patient as negative, RR and nasal congestion made substantial negative contributions.

Discussion

Although the diagnostic criteria for NS in China have been continuously optimized and are actively guiding clinical practice, new challenges persist. The clinical utility of biomarkers, notably CRP, WBC, and PCT, remains debated (32). Poor responsiveness is a key clinical feature of EOS, but accurately identifying it is complex due to potential confounding factors such as sleep states, hypoxemia, and hypoglycemia (33). These factors complicate clinical assessment and increase the risk of misdiagnosis. Consequently, the pursuit of an early and precise diagnostic approach that integrates clinical manifestations and laboratory examination for neonatal EOS has become a shared goal among healthcare professionals. Globally, AI technology experienced an unprecedented surge in development, with ML-based clinical prediction models becoming ubiquitous in diagnosing and assessing the prognosis of various diseases, including cancer, diabetes, and severe infections (34,35). However, the "black box" nature of these models posed significant challenges in explaining their internal workings and prediction logic. This lack of transparency is particularly problematic in medicine, where practitioners need clear insights into how models generate predictions based on patient-specific attributes and which features most influence outcomes. Consequently, integrating ML with medical diagnostics while enhancing interpretability through techniques like SHAP became a critical research focus. Based on Shapley values introduced by Shapley in 1952 (36), SHAP values precisely measure each feature's contribution to the prediction, revealing feature interactions and providing a comprehensive explanation of the model's decision process. This approach enhances model transparency and improves diagnostic accuracy and clinical decision-making.

ML models each possess unique strengths (37): SVM excels in high-dimensional data classification and is suitable for small sample sets; LR is concise and easy to understand, making it ideal for linearly separable problems; RF enhances prediction stability and accuracy through integrating multiple decision trees, displaying robustness against noise; XGBoost optimizes the gradient boosting method for computational efficiency and automatic data issue handling; MLP is versatile in fitting complex patterns but prone to overfitting; CatBoost, also based on gradient boosting, specifically optimizes categorical feature processing, reducing overfitting risks, and is particularly adept at classification tasks. Given the distinct advantages and limitations of these models, the choice of an appropriate model necessitates consideration of the specific problem characteristics, data scale, and computational resources. In this study, we utilized LASSO regression to select features with the greatest predictive power for the outcome variable, ultimately ranking the features in order of importance as RR, PCT, NC, YS, WBC, fever, and AFT. These features were then employed to construct SVM, LR, RF, XGBoost, MLP, and CatBoost ML models. While the traditional LR model demonstrated commendable performance, the other ML models, particularly CatBoost, outperformed it significantly. The CatBoost model achieved a test set ROCAUC of 0.975, accuracy of 0.917, sensitivity (recall) of 0.889, specificity of 0.926, precision of 0.800, and F1 score of 0.842, slightly surpassing the accuracy of 87.07%, sensitivity of 83.33%, and specificity of 92.00% reported in domestic studies on neonatal EOS prediction models. The PR curves of each model in this study also indicated that the CatBoost model performed best. The calibration curve demonstrated good agreement, and the DCA curve suggested that the model could provide substantial net benefit for clinical applications. Furthermore, the learning curve showed good model fitting, indicating robust generalization ability.

The CatBoost model's beeswarm plot based on SHAP values in this study revealed that increased PCT and WBC levels, along with the presence of fever, positively contributed to predicting EOS, while elevated RR, NC, YS, and turbid amniotic fluid (pregnant woman) were more likely to predict non-EOS. The CatBoost model's decision plot based on SHAP values provided a more intuitive visualization of the entire decision-making process, highlighting RR as a crucial factor ranking first in SHAP value order. Additionally, the CatBoost model's feature dependence analysis chart showed that neonatal RR had a positive contribution to model predictions when below 55 breaths per minute, but a negative contribution above this threshold. One of the common clinical manifestations in neonatal EOS is increased RR. This is due to sepsis causing a systemic inflammatory response, which leads to symptoms such as tachycardia and tachypnea. Infection and inflammation increase metabolic demands, and the body may compensate by increasing RR to meet higher oxygen demands. However, it is noteworthy that in newborns, especially premature infants or those with very low birth weights, the clinical manifestations may be atypical when facing severe infections or illnesses, and they may exhibit decreased RR or other non-specific symptoms. Fever was more common in EOS neonates, consistent with previous studies (6,38). The presence of nasal congestion, yellow staining, and turbid amniotic fluid (Pregnant woman) was more likely to predict non-EOS, potentially due to clinical staff's proactive interventions upon observing these symptoms, thereby reducing the occurrence of EOS. PCT and WBC are non-specific laboratory indicators for EOS. While PCT, as an infection marker, rises rapidly during bacterial infections (39), its early physiological elevation in neonates limits its diagnostic accuracy for EOS (40). In this study, PCT elevations to around 0.8 ng/mL positively contributed to EOS prediction, indicating that PCT still holds some value in predicting neonatal EOS, in line with previous research (41). WBC, a commonly used indicator, has low sensitivity and cannot be solely relied upon for diagnosing NS (42). In this study, WBC values below 15×10^9/L mostly had a negative contribution to model predictions, while values above 15×10^9/L shifted to a positive contribution. Neonatal EOS is not a single, static disease but a dynamic, continuous inflammatory response. This dynamic nature underscores the inadequacy of relying solely on a single biomarker for diagnosis, as different pathogens, immune statuses, and sepsis durations alter the systemic immune response (6), emphasizing the importance of multi-indicator comprehensive assessments.

This study has certain limitations: (I) the small amount of data included in the study failed to provide more detailed explanations for the prediction results, and being a single-center study, there is potential bias in the research results. It is necessary to conduct larger-scale, multi-center studies for further verification; (II) a web calculator for neonatal EOS based on clinical diagnostics was not developed to provide real-time predictions for clinical NICU patients, which would be more conducive to timely clinical diagnosis and treatment as well as evaluation of treatment effects. These are key directions for our future research work.

Conclusions

In summary, the present study successfully devised a ML model tailored for predicting EOS among neonates admitted to the NICU. The employment of SHAP methodology significantly enhances the transparency of this intricate model by elucidating its prediction outcomes (43). Notably, crucial clinical indicators including RR, nasal congestion, yellow staining, fever, and turbid amniotic fluid status can be readily assessed upon neonatal admission. Furthermore, prompt acquisition of PCT and WBC counts through the expedited emergency laboratory pathway subsequent to routine blood sampling, as per clinical directives, facilitates the integration of these biomarkers into the model. Consequently, the application of these comprehensive data inputs to the model yields predictive SHAP values, empowering clinicians to accurately identify EOS in neonates at an early stage during their hospital stay.

Acknowledgments

Funding: This article was supported by the Department of Education University Research Project in Anhui Province, China (No. 2023AH051748), the Scientific Research Project of Wuhu Health Commission (No. WHWJ2023y012), and the Young and Middle-aged Scientific Research Fund Project of Wannan Medical College (No. WK2022F43).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tp.amegroups.com/article/view/10.21037/tp-24-278/rc

Data Sharing Statement: Available at https://tp.amegroups.com/article/view/10.21037/tp-24-278/dss

Peer Review File: Available at https://tp.amegroups.com/article/view/10.21037/tp-24-278/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tp.amegroups.com/article/view/10.21037/tp-24-278/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of the Bozhou People's Hospital (approval No. 2024-149). This study has obtained informed consent from the patients.

References

1. Attia Hussein Mahmoud H, Parekh R, Dhandibhotla S, et al. Insight Into Neonatal Sepsis: An Overview. Cureus 2023;15:e45530.
2. Roca A, Camara B, Bognini JD, et al. Effect of Intrapartum Azithromycin vs Placebo on Neonatal Sepsis and Death: A Randomized Clinical Trial. JAMA 2023;329:716-24.
3. Gad A, Alkhdr M, Terkawi R, et al. Associations between maternal bacteremia during the peripartum period and early-onset neonatal sepsis: a retrospective cohort study. BMC Pediatr 2024;24:526.
4. Agudelo-Pérez S, Moreno AM, Martínez-Garro J, et al. 16S rDNA Sequencing for Bacterial Identification in Preterm Infants with Suspected Early-Onset Neonatal Sepsis. Trop Med Infect Dis 2024;9:152.
5. Chen J, Fang X, Liu W, et al. Exploring factors influencing delayed first antibiotic treatment for suspected early-onset sepsis in preterm newborns: a study before quality improvement initiative. BMC Pediatr 2024;24:407.
6. Barboza AZ, Flannery DD, Shu D, et al. Trends in C-Reactive Protein Use in Early-onset Sepsis Evaluations and Associated Antibiotic Use. J Pediatr 2024;273:114153.
7. Giannoni E, Dimopoulou V, Klingenberg C, et al. Analysis of Antibiotic Exposure and Early-Onset Neonatal Sepsis in Europe, North America, and Australia. JAMA Netw Open 2022;5:e2243691.
8. Shi W, Chen Z, Shi L, et al. Early Antibiotic Exposure and Bronchopulmonary Dysplasia in Very Preterm Infants at Low Risk of Early-Onset Sepsis. JAMA Netw Open 2024;7:e2418831.
9. Witt LT, Greenfield KG, Knoop KA. Streptococcus agalactiae and Escherichia coli induce distinct effector γδ T cell responses during neonatal sepsis. iScience 2024;27:109669.
10. Li J, Xiang L, Chen X, et al. Global, regional, and national burden of neonatal sepsis and other neonatal infections, 1990-2019: findings from the Global Burden of Disease Study 2019. Eur J Pediatr 2023;182:2335-43.
11. Dalut L, Brunhes A, Cambier S, et al. Early-onset neonatal sepsis: Effectiveness of classification based on ante- and intrapartum risk factors and clinical monitoring. J Gynecol Obstet Hum Reprod 2024;53:102775.
12. Department of Neonatology, Science Branch, Chinese Medical Association; Infection Committee of neonatal pediatricians branch, Chinese Medical Association. Expert consensus on diagnosis and treatment of neonatal sepsis (2019 Edition). Chinese J Pediatr 2019;57:252-7.
13. Polin RA. Management of neonates with suspected or proven early-onset bacterial sepsis. Pediatrics 2012;129:1006-15.
14. Grill A, Goeral K, Leitich H, et al. Maternal biomarkers in predicting neonatal sepsis after preterm premature rupture of membranes in preterm infants. Acta Paediatr 2024;113:962-72.

15. Kallonen A, Juutinen M, Värri A, et al. Early detection of late-onset neonatal sepsis from noninvasive biosignals using deep learning: A multicenter prospective development and validation study. Int J Med Inform 2024;184:105366.
16. Honoré A, Forsberg D, Adolphson K, et al. Vital sign-based detection of sepsis in neonates using machine learning. Acta Paediatr 2023;112:686-96.
17. Parkinson E, Liberatore F, Watkins WJ, et al. Gene filtering strategies for machine learning guided biomarker discovery using neonatal sepsis RNA-seq data. Front Genet 2023;14:1158352.
18. Kumar R, Kausch SL, Gummadi AKS, et al. Inflammatory biomarkers and physiomarkers of late-onset sepsis and necrotizing enterocolitis in premature infants. Front Pediatr 2024;12:1337849.
19. Persad E, Jost K, Honoré A, et al. Neonatal sepsis prediction through clinical decision support algorithms: A systematic review. Acta Paediatr 2021;110:3201-26.
20. Huang B, Wang R, Masino AJ, et al. Aiding clinical assessment of neonatal sepsis using hematological analyzer data with machine learning techniques. Int J Lab Hematol 2021;43:1341-56.
21. Qiao T, Tu X. A practical predictive model to predict 30-day mortality in neonatal sepsis. Rev Assoc Med Bras (1992) 2024;70:e20231561.
22. Zhang P, Wang Z, Qiu H, et al. Machine learning applied to serum and cerebrospinal fluid metabolomes revealed altered arginine metabolism in neonatal sepsis with meningoencephalitis. Comput Struct Biotechnol J 2021;19:3284-92.
23. Ye R Z, Bahig H, Wong P. PD-0406 Unsupervised leaning of biometric features predicts metastatic head and neck cancer progression. Radiotherapy and Oncology 2023;182:S308-S309.
24. Padegal G, Rao MK, Boggaram Ravishankar OA, et al. Analysis of RNA-Seq data using self-supervised learning for vital status prediction of colorectal cancer patients. BMC Bioinformatics 2023;24:241.
25. Khowaja A, Zou B, Kui X. Enhancing cervical cancer diagnosis: Integrated attention-transformer system with weakly supervised learning. Image and Vision Computing 2024,149:105193.
26. Gutiérrez Y, Arevalo J, Martínez F. A contrastive weakly supervised learning to characterize malignant prostate lesions in BP-MRI. Biomedical Signal Processing and Control 2024;96:106584.
27. Hang Y, Qu H, Yang J, et al. Exploration of programmed cell death-associated characteristics and immune infiltration in neonatal sepsis: new insights from bioinformatics analysis and machine learning. BMC Pediatr 2024;24:67.
28. Meeus M, Beirnaert C, Mahieu L, et al. Clinical Decision Support for Improved Neonatal Care: The Development of a Machine Learning Model for the Prediction of Late-onset Sepsis and Necrotizing Enterocolitis. J Pediatr 2024;266:113869.
29. Lyra S, Jin J, Leonhardt S, et al. Early Prediction of Neonatal Sepsis From Synthetic Clinical Data Using Machine Learning. Annu Int Conf IEEE Eng Med Biol Soc 2023;2023:1-4.
30. Zou Y, Shi Y, Sun F, et al. Extreme gradient boosting model to assess risk of central cervical lymph node metastasis in patients with papillary thyroid carcinoma: Individual prediction using SHapley Additive exPlanations. Comput Methods Programs Biomed 2022;225:107038.
31. Nohara Y, Matsumoto K, Soejima H, et al. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput Methods Programs Biomed 2022;214:106584.
32. Stocker M, van Herk W, El Helou S, et al. C-Reactive Protein, Procalcitonin, and White Blood Count to Rule Out Neonatal Early-onset Sepsis Within 36 Hours: A Secondary Analysis of the Neonatal Procalcitonin Intervention Study. Clin Infect Dis 2021;73:e383-90.
33. Xie L, Ding L, Tang L, et al. A real-world cost-effectiveness study of vancomycin versus linezolid for the treatment of late-onset neonatal sepsis in the NICU in China. BMC Health Serv Res 2023;23:771.
34. Ma M, Liu R, Wen C, et al. Predicting the molecular subtype of breast cancer and identifying interpretable imaging features using machine learning algorithms. Eur Radiol 2022;32:1652-62.
35. Deberneh HM, Kim I. Prediction of Type 2 Diabetes Based on Machine Learning Algorithm. Int J Environ Res Public Health 2021;18:3317.
36. Shapley LS. A Value for N-Person Games. Santa Moni-ca, CA: RAND Corporation; 1952.
37. Parvin SA, Saleena B. Analysis of machine learning and deep learning prediction models for sepsis and neonatal sepsis: A systematic review. ICT Express 2023;9:1215-25.
38. Chen X, He H, Wei H, et al. Risk factors for death caused by early onset sepsis in neonates: a retrospective cohort study. BMC Infect Dis 2023;23:844.
39. Jaiswal SR, Kota V, Rai R. Severe Hypertriglyceridemia in a Neonate Secondary to Septicemia and Acute Kidney

Injury. Indian J Pediatr 2023;90:1264.
40. Dammann O, Stansfield BK. Neonatal sepsis as a cause of retinopathy of prematurity: An etiological explanation. Prog Retin Eye Res 2024;98:101230.
41. Gopal N, Chauhan N, Jain U, et al. Advancement in biomarker based effective diagnosis of neonatal sepsis. Artif Cells Nanomed Biotechnol 2023;51:476-90.
42. Li T, Li X, Zhu Z, et al. Clinical value of procalcitonin-to-albumin ratio for identifying sepsis in neonates with pneumonia. Ann Med 2023;55:920-5.
43. Zhang J, Ma X, Zhang J, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. J Environ Manage 2023;332:117357.

Cite this article as: Tan X, Zhang X, Chai J, Ji W, Ru J, Yang C, Zhou W, Bai J, Xiong Y. Constructing a predictive model for early-onset sepsis in neonatal intensive care unit newborns based on SHapley Additive exPlanations explainable machine learning. Transl Pediatr 2024;13(11):1933-1946. doi: 10.21037/tp-24-278