Available Online: 19 November 2026 DOI: 10.54254/3049-5458/2026.39356

Development and validation of multi-algorithm machine learning models for predicting phototherapy efficacy in premature infants with hyperbilirubinemia and nursing decision recommendations

Qiufen Shu[1], *Xiuping Huang*[2,*]

[1]Youjiang Medical University for Nationalities, Baise, China
[2]Department of Pediatrics, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China

*Corresponding Author. Email: hxiup1986@yeah.net

Abstract. Objective: To develop multi-algorithm machine learning models for predicting the efficacy of phototherapy in preterm infants with hyperbilirubinemia. Methods: A total of 250 preterm infants who received phototherapy were retrospectively enrolled and randomly divided into training and validation sets at a ratio of 7:3. A reduction of ≥ 20% in serum total bilirubin after 24 h of phototherapy was defined as an effective response. Following variable selection, three predictors—peak bilirubin level, albumin level, and multiple pregnancy—were included in the models. Logistic regression, Lasso, support vector machine, and naive Bayes models were developed and evaluated using the Area under the Receiver Operating Characteristic Curve (AUC), calibration curves, and decision curve analysis. SHAP analysis was also performed. Results: In the validation set, the logistic regression model achieved the highest AUC of 0.905 (95% CI: 0.834–0.975), while the AUCs of the other models ranged from 0.884 to 0.903. All four models demonstrated good calibration and positive net benefits across a relatively wide range of threshold probabilities. SHAP analysis identified the three aforementioned variables as the leading contributors to model predictions. Conclusion: The multi-algorithm models demonstrated good predictive performance, with logistic regression achieving the best overall performance. SHAP analysis provides a quantitative basis for identifying key risk variables and informing stratified nursing decisions. However, the findings require further external validation.

Keywords: preterm infants, hyperbilirubinemia, phototherapy efficacy, machine learning, nursing interventions

1. Introduction

Neonatal hyperbilirubinemia is one of the most common clinical conditions during the neonatal period and is particularly prevalent among preterm infants. Its incidence, age at onset, and severity are closely associated with gestational age, with lower gestational age generally corresponding to earlier onset and greater severity. Severe hyperbilirubinemia can cause irreversible neurological damage, including acute bilirubin

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

encephalopathy and kernicterus, and represents a substantial global burden of neonatal disease [1]. The activity of hepatic bilirubin UDP-Glucuronosyltransferase (UGT1A1) undergoes developmental maturation and can be divided into four stages: mid-gestation, late gestation, the neonatal to early infancy period, and maturity. During the first three stages, UGT1A1 activity is approximately 0.1%, 0.1%–1%, and 1%–100% of its mature level, respectively. Preterm birth can also affect its postnatal maturation [2]. Meanwhile, preterm infants have limited albumin synthesis capacity and significantly reduced bilirubin metabolism and conjugation, making them more susceptible to severe hyperbilirubinemia. Furthermore, their blood–brain barrier is incompletely developed, resulting in substantially increased susceptibility to bilirubin-induced neurotoxicity [3]. A nationwide survey in Japan reported that the incidence of clinical kernicterus among preterm infants born at < 30 weeks of gestation was approximately 0.18% (1.8 per 1,000 live births) [4]. Once it occurs, kernicterus can cause irreversible neurological damage, including cerebral palsy, hearing loss, and intellectual impairment.

Phototherapy is currently the most commonly used and preferred first-line treatment for neonatal hyperbilirubinemia because of its effectiveness [5]. However, despite more than 60 years of clinical application [6], the management of phototherapy in preterm infants continues to face two major challenges in clinical practice. First, the initiation and management of phototherapy are largely guided by standardized guidelines or decision-making tools, leaving limited room for individualized treatment [7]. Although existing guidelines and expert recommendations provide phototherapy indications based on Serum Total Bilirubin (TSB) levels, the populations on which these recommendations are based do not fully correspond to the population included in the present study. For infants with a gestational age of ≥ 35 weeks, phototherapy thresholds are primarily based on the guidelines of the American Academy of Pediatrics (AAP) [8]. For preterm infants with a gestational age of < 35 weeks, relevant expert recommendations are generally followed [9], with phototherapy also serving as a first-line intervention for this population. In addition, preterm infants exhibit considerable individual variation in their responses to phototherapy. Even among infants with the same TSB level, the rate of bilirubin reduction following phototherapy may differ substantially [10]. Second, TSB, as the principal indicator guiding clinical decisions, has inherent limitations [11]. It reflects only the total amount of bilirubin in the blood and cannot fully represent the actual burden of unbound bilirubin, which is the key substance responsible for neurotoxicity. Furthermore, TSB alone cannot accurately predict an individual infant's response to phototherapy [12].

In recent years, machine learning techniques have demonstrated considerable advantages in predicting neonatal diseases [13]. Compared with traditional univariate analytical methods, machine learning can simultaneously capture complex nonlinear relationships among multiple variables, thereby identifying combined effects that may not be revealed by individual indicators [14]. Previous studies have developed neonatal disease prediction models using machine learning algorithms such as random forests and have achieved good predictive performance. However, to the best of our knowledge, relatively few studies have developed predictive models specifically for phototherapy efficacy in preterm infants. Moreover, the black-box nature of machine learning models has long hindered their clinical translation. The introduction of SHapley Additive exPlanations (SHAP) offers an effective approach to addressing this limitation. It quantifies the contribution of each feature to model predictions at both the global and individual levels, thereby improving the interpretability of predictive results [15].

Meanwhile, nursing interventions play an indispensable role in the management of phototherapy in preterm infants. The management of adverse reactions during phototherapy, such as fever, rash, diarrhea, and increased insensible water loss, as well as parental anxiety, depends on nurses' proactive assessment and meticulous interventions. Current nursing practices in Neonatal Intensive Care Units (NICUs) largely follow fixed

protocols and lack differentiated nursing strategies based on individual risk profiles, making it difficult to meet the demands of precision nursing. Integrating machine learning prediction models into nursing practice and implementing stratified nursing interventions based on the key risk variables identified by these models may help overcome the limitations of traditional standardized nursing approaches.

Against this background, the present study aimed to develop predictive models for phototherapy efficacy in preterm infants using multiple machine learning algorithms, systematically evaluate the discrimination, calibration, and clinical net benefit of each model, and quantify the risk contributions of predictive variables through SHAP analysis. Based on these findings, this paper further soughts to propose targeted nursing decision-making recommendations based on risk stratification, thereby providing a theoretical foundation and practical tools for the precision management of hyperbilirubinemia in preterm infants.

2. Materials and methods

2.1. Study design and participants

This retrospective cohort study included 250 preterm infants who were admitted to the Neonatal Intensive Care Unit (NICU) of a tertiary Grade A hospital (the hospital name has been withheld to protect privacy) and received phototherapy between January 2019 and December 2024. The inclusion criteria were as follows: (1) gestational age < 37 weeks; (2) meeting the indications for phototherapy according to the 2014 Expert Consensus on the Diagnosis and Treatment of Neonatal Hyperbilirubinemia [16]; and (3) availability of complete clinical records. The exclusion criteria were as follows: (1) severe congenital malformations; (2) confirmed obstructive jaundice, such as biliary atresia; (3) requirement for exchange transfusion during phototherapy; and (4) more than 20% of clinical data missing. This study was approved by the hospital's ethics committee (approval No. 2026091701). Informed consent was waived because of the retrospective study design.

2.2. Phototherapy protocol and assessment of treatment efficacy

All infants received LED blue-light phototherapy at a wavelength of 425–475 nm and an irradiance of 10–30 μW/(cm²·nm). The distance between the light source and the infant was maintained at 30–40 cm, and phototherapy was administered continuously for 24 h. During treatment, protective eye shields were used to cover the eyes, and the perineal area was covered with a diaper. Treatment efficacy was assessed based on the percentage reduction in Serum Total Bilirubin (TSB) after 24 h of phototherapy, calculated as follows: TSB reduction (%) = [(TSB before phototherapy − TSB after 24 h of phototherapy) / TSB before phototherapy] × 100%. A reduction of ≥ 20% was defined as an effective response (group = 0), whereas a reduction of < 20% was defined as an ineffective response (group = 1). It should be noted that the ≥ 20% reduction in TSB after 24 h of phototherapy was adopted as an operational cutoff for assessing treatment efficacy in this study. This threshold has not yet been established through a universally accepted evidence-based consensus, and model performance under different cutoff values requires further validation.

2.3. Data collection

A total of 21 clinical variables were collected from the electronic medical record system, covering the following four categories: (1) Demographic characteristics: sex, gestational age, and birth weight; (2) Perinatal factors: degree of amniotic fluid contamination, Apgar scores at 1 and 5 min, passage of urine and meconium, mode of delivery, and advanced maternal age; (3) Comorbidities: ventilator-associated pneumonia, gestational

hypertension, multiple pregnancy, gestational diabetes mellitus, Respiratory Distress Syndrome (RDS), and sepsis; (4) Laboratory indicators: albumin, Lactate Dehydrogenase (LDH), peak bilirubin, C-Reactive Protein (CRP), hemoglobin, and the Bilirubin-to-Albumin (B/A) ratio. For each laboratory indicator, the most recent available test result before the initiation of phototherapy was used as the value for analysis. Due to the way information was recorded in the electronic medical records, the actual sampling time of certain individual test results could not be retrospectively verified on a case-by-case basis (see Section 4.5, Limitations). It should also be noted that the variable name "peak bilirubin" was retained from the original field name in the modeling dataset. The recorded values of this field were directly used in the analysis, with the most recent available test result before the initiation of phototherapy serving as its operational definition. The highest bilirubin level during hospitalization was not recalculated. Owing to the limitations of the electronic medical records, it was not possible to retrospectively verify on a case-by-case basis whether the recorded values of this field corresponded to the actual peak bilirubin levels during hospitalization.

2.4. Dataset partitioning

The 250 infants were divided into a training set (n = 175) and a validation set (n = 75) at a ratio of 7:3. There was no statistically significant difference in the distribution of treatment outcomes between the two sets (χ^2 = 3.329, p = 0.068). The training set was used for univariate and multivariate regression analyses and model development, whereas the validation set was used to evaluate model performance. Among the 175 infants in the training set, 76 experienced ineffective phototherapy. With three predictive variables, the Events-Per-Variable (EPV) ratio was approximately 25, meeting the minimum event requirements for fitting a logistic regression model. An EPV of ≥ 10 is generally considered adequate. Therefore, the sample size was considered sufficient to support model development and validation.

2.5. Variable selection

Variable selection was performed exclusively within the training set (n = 175). First, univariate analyses were conducted on all 21 variables in the training set. The normality of continuous variables was assessed using the Shapiro–Wilk test. Independent-samples t-tests were used for approximately normally distributed variables, whereas the Mann–Whitney U test was applied to variables that did not follow an approximately normal distribution. Categorical variables were analyzed using the chi-square test or Fisher's exact test. Variables with $p < 0.05$ in the univariate analyses were subsequently entered into a multivariate logistic regression model. Independent risk factors were identified using $p < 0.05$ as the criterion for variable retention. Since there were only four candidate variables, no stepwise regression procedure was performed. The univariate analyses revealed statistically significant differences between the effective and ineffective groups in four variables: albumin, peak bilirubin, B/A ratio, and multiple pregnancy (all $p < 0.05$). Four additional variables—gestational age, lactate dehydrogenase, 5-min Apgar score, and gestational diabetes mellitus—had p values ranging from 0.05 to 0.10 (0.086, 0.053, 0.081, and 0.064, respectively) and were therefore not included in the subsequent analyses. Considering that the derived ratio variable (B/A ratio) was mathematically coupled with peak bilirubin and albumin (B/A ratio = serum total bilirubin / albumin), a multicollinearity assessment was performed. The Pearson correlation coefficient between the B/A ratio and peak bilirubin was r = 0.906. After the B/A ratio was included in the model, the Variance Inflation Factors (VIFs) were 21.20 for the B/A ratio, 18.57 for peak bilirubin, and 3.81 for albumin. The VIFs of the first two variables substantially exceeded the threshold of 10, indicating severe multicollinearity. To avoid the potential effects of multicollinearity on coefficient estimation and model stability, the B/A ratio was excluded. Ultimately, peak bilirubin, albumin, and multiple pregnancy were retained as the three predictor variables for model development.

2.6. Model development

Based on the selected independent risk factors, four machine learning models were developed separately in the R environment: (1) Logistic regression model: Used as the baseline model, with parameters estimated using the maximum likelihood method without applying any regularization penalty. (2) Lasso regression model (essentially L1-regularized logistic regression): An L1 regularization term was introduced to shrink the coefficients. The penalty parameter λ was determined through five-fold cross-validation within the training set by minimizing the deviance and applying the one-standard-error (1-SE) rule. The final λ value in this study was 0.0539 [17]. (3) Support Vector Machine (SVM) model [18]: A Radial Basis Function (RBF) kernel was used [19]. The kernel width parameter σ and penalty parameter C were optimized through grid-based cross-validation within the training set. The final values were σ = 0.05 and C = 1. Probability outputs were calibrated using Platt scaling [20]. (4) Naive Bayes model: Based on the assumption of conditional independence among features, the conditional probability distributions of individual features were modeled using Gaussian distributions. Logistic regression and naive Bayes were fitted using the original measurement units, and their coefficients were therefore expressed in the original units. For Lasso regression and SVM, continuous variables were internally standardized during model fitting; consequently, their coefficients and regularization parameters did not have interpretations in the original measurement units. All models were trained on the training set and independently evaluated on the validation set. Internal hyperparameter optimization was performed using cross-validation within the training set. The alternative model used in the sensitivity analysis, in which the B/A ratio replaced peak bilirubin, was fitted using the same procedure as the primary model (maximum likelihood estimation without regularization penalties) to ensure direct comparability between the two models.

2.7. Model evaluation

Model performance was evaluated across four dimensions: (1) Discrimination: Receiver operating characteristic (ROC) curves were plotted for the validation set, and the Area Under the Curve (AUC) and its 95% Confidence Interval (CI) were calculated using the DeLong method [21]. An AUC > 0.8 was considered indicative of good discrimination. (2) Calibration: Calibration curves were plotted, and the Hosmer–Lemeshow (H-L) goodness-of-fit test was performed. A p value > 0.05 indicated good model calibration [22]. The Brier score was also calculated to measure the mean squared error between predicted probabilities and observed outcomes. The Brier score ranges from 0 to 1, with lower values indicating better calibration. (3) Clinical utility: Decision Curve Analysis (DCA) was performed to evaluate the net benefit of each model across different threshold probabilities [23]. (4) Training-set model performance: The AUC in the training set was briefly reported to assess the potential presence of overfitting. In addition, the following supplementary analyses were conducted to provide a more robust assessment of model performance: (5) Bootstrap resampling with 1,000 iterations was used to estimate the 95% CI of the validation-set AUC, while the Out-of-Fold (OOF) cross-validation AUC was calculated to assess optimism in the training-set performance. (6) Calibration performance was further assessed by reporting the calibration intercept (ideal value: 0) and calibration slope (ideal value: 1). (7) Pairwise comparisons of the AUCs of the four models were performed using the DeLong test [21]. Subsequent SHAP analyses were conducted using the recommended model, namely logistic regression. The rationale for this recommendation is provided in Section 3.2.

2.8. Shap-based interpretability analysis

Based on the Shapley value concept from cooperative game theory, SHAP assigns a contribution value to each feature for every individual sample, indicating the direction and magnitude of that feature's influence on the

prediction [15, 24]. In this study, SHAP analysis was performed at both the global and local levels. At the global level, feature importance bar plots and beeswarm plots were used to illustrate the mean contribution of each feature across all samples and the distribution of contribution directions. At the local level, force plots and waterfall plots were used to decompose the contributions of individual features for specific patients, thereby providing individualized explanations of model predictions.

2.9. Statistical analysis

All data analyses and model development were performed in the R environment using the glm, glmnet, kernlab, klaR, caret, pROC, rmda, shapviz, and kernelshap packages. SHAP values were calculated using Kernel SHAP (kernelshap) and visualized with the shapviz package. DCA was performed using the rmda package, while ROC curves and their 95% CIs were calculated using the pROC package based on the DeLong method. Continuous variables were expressed as mean ± standard deviation or median (interquartile range), as appropriate. Between-group comparisons were performed using the t-test or Mann–Whitney U test. Categorical variables were expressed as frequencies (percentages), with between-group comparisons conducted using the chi-square test or Fisher's exact test. Variables were retained in the multivariate logistic regression model based on a significance threshold of $p < 0.05$. A p value < 0.05 was considered statistically significant. The reporting of this prediction model followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement [25]. None of the 21 clinical variables collected from the 250 enrolled infants had missing values; therefore, no data imputation was performed. Random seeds were fixed for all procedures involving randomization, including cross-validation fold allocation and SHAP value computation, to ensure reproducibility. Reproducibility was addressed at two levels. First, at the data level, de-identified raw data and the files documenting the training and validation set allocations can be obtained from the corresponding author upon reasonable request, subject to ethical approval requirements and the protection of participant privacy. Second, at the algorithmic level, the complete scripts covering primary model fitting on the training set and subsequent evaluation on the validation set are provided. These scripts specify the software packages, fitting procedures, and final hyperparameter values used for each model. Running the scripts in an R 4.4.0 environment enables reproduction of the validation-set AUCs and their corresponding 95% CIs.

3. Results

3.1. Baseline characteristics

A total of 250 preterm infants were included in this study, of whom 132 (52.8%) were in the effective phototherapy group (group = 0) and 118 (47.2%) were in the ineffective phototherapy group (group = 1). The cohort was randomly divided at a ratio of 7:3 into a training set comprising 175 infants (99 effective cases, 56.6%; 76 ineffective cases, 43.4%) and a validation set comprising 75 infants (33 effective cases, 44.0%; 42 ineffective cases, 56.0%). The difference in the distribution of treatment outcomes between the two sets was not statistically significant ($\chi^2 = 3.329$, $p = 0.068$), although the proportion of ineffective outcomes was relatively higher in the validation set. No statistically significant differences were observed in the other baseline variables between the two sets (all $p > 0.05$), suggesting that the validation set was at least as challenging to predict as the training set.

A comparison of the baseline characteristics of the entire cohort ($n = 250$) is presented in Table 1. Seven variables showed statistically significant differences between the effective and ineffective groups: gestational age (243.1 ± 12.6 days vs. 235.9 ± 14.8 days, $p < 0.001$), 5-min Apgar score ($p = 0.039$), multiple pregnancy

(17.4% vs. 42.4%, $p < 0.001$), albumin (36.0 ± 6.0 g/L vs. 29.9 ± 6.5 g/L, $p < 0.001$), lactate dehydrogenase (628.0 ± 399.1 U/L vs. 531.4 ± 355.9 U/L, $p = 0.045$), peak bilirubin (78.5 ± 45.9 μmol/L vs. 141.5 ± 59.2 μmol/L, $p < 0.001$), and the B/A ratio (2.2 ± 1.3 vs. 4.8 ± 2.1, $p < 0.001$). These comparisons represent descriptive analyses of the entire cohort. Variable selection for model development was performed exclusively within the training set (see Section 2.5). In the training set, albumin, peak bilirubin, B/A ratio, and multiple pregnancy were the four variables with $p < 0.05$ in the univariate analyses. Multivariate regression was therefore based on these four variables, with the collinear B/A ratio subsequently excluded, leaving three variables in the final model. It should be noted that the prevalence of ventilator-associated pneumonia in this cohort was relatively high (59.6%). However, no statistically significant difference was observed between the effective and ineffective groups (60.6% vs. 58.5%, $p = 0.732$). This variable was not included in the final prediction model, and no clinical causal inferences regarding it are made in the present study; it is reported solely as part of the baseline characteristics.

The four variables with $p < 0.05$ in the univariate analyses of the training set (albumin, peak bilirubin, B/A ratio, and multiple pregnancy) were entered into a multivariate logistic regression model. Following collinearity assessment, the highly collinear derived variable (B/A ratio) was excluded because of its strong correlation with peak bilirubin (Pearson $r = 0.906$) and a Variance Inflation Factor (VIF) > 10 (see Section 2.5). Ultimately, three variables—peak bilirubin, albumin, and multiple pregnancy—were included in the prediction model. The multivariate logistic regression results showed that each 1 μmol/L increase in peak bilirubin was associated with a 2.9% increase in the risk of ineffective phototherapy (OR = 1.029, 95% CI: 1.019–1.039, $p < 0.001$). Each 1 g/L increase in albumin was associated with an approximately 20% reduction in the risk of an ineffective response (OR = 0.803, 95% CI: 0.743–0.868, $p < 0.001$). Preterm infants from multiple pregnancies had 2.9 times the risk of an ineffective phototherapy response compared with those from singleton pregnancies (OR = 2.855, 95% CI: 1.091–7.473, $p = 0.033$).

The prediction equation was as follows: $\text{logit}(p = \text{invalid}) = 3.489 + 0.0290 \times \text{peak bilirubin} - 0.2194 \times \text{albumin} + 1.049 \times \text{multiple pregnancy}$.

Table 1. Descriptive comparison of baseline characteristics between the effective and ineffective phototherapy groups in the entire cohort ($n = 250$)

Variable	Overall (n = 250)	Effective group (n = 132)	Ineffective group (n = 118)	Test statistic	p value
Sex, n (%)				$\chi^2 = 0.004$	0.950
Female	134 (53.6)	71 (53.8)	63 (53.4)		
Male	116 (46.4)	61 (46.2)	55 (46.6)		
Gestational age (days), mean ± SD	239.7 ± 14.1	243.1 ± 12.6	235.9 ± 14.8	t = 4.13	< 0.001
Birth weight (kg), mean ± SD	2.1 ± 0.5	2.2 ± 0.5	2.1 ± 0.6	t = 0.77	0.441
Amniotic fluid contamination, n (%)					0.669
Absent	245 (98.0)	130 (98.5)	115 (97.5)		
Present	5 (2.0)	2 (1.5)	3 (2.5)		
1-min Apgar score, n (%)					0.860
5–7	37 (14.8)	19 (14.4)	18 (15.3)		
8–10	213 (85.2)	113 (85.6)	100 (84.7)		
5-min Apgar score, n (%)				$\chi^2 = 6.485$	0.039

Table 1. Continued

8	21 (8.4)	10 (7.6)	11 (9.3)		
9	97 (38.8)	61 (46.2)	36 (30.5)		
10	132 (52.8)	61 (46.2)	71 (60.2)		
Passage of urine and stool, *n* (%)				$\chi^2 = 0.319$	0.572
Absent	133 (53.2)	68 (51.5)	65 (55.1)		
Present	117 (46.8)	64 (48.5)	53 (44.9)		
Ventilator-associated pneumonia, *n* (%)				$\chi^2 = 0.118$	0.732
Absent	101 (40.4)	52 (39.4)	49 (41.5)		
Present	149 (59.6)	80 (60.6)	69 (58.5)		
Gestational hypertension, *n* (%)				$\chi^2 = 0.100$	0.752
Absent	167 (66.8)	87 (65.9)	80 (67.8)		
Present	83 (33.2)	45 (34.1)	38 (32.2)		
Multiple pregnancy, *n* (%)				$\chi^2 = 18.76$	< 0.001
No	177 (70.8)	109 (82.6)	68 (57.6)		
Yes	73 (29.2)	23 (17.4)	50 (42.4)		
Gestational diabetes mellitus, *n* (%)				χ^2=3.52	0.061
Absent	153 (61.2)	88 (66.7)	65 (55.1)		
Present	97 (38.8)	44 (33.3)	53 (44.9)		
Mode of delivery, *n* (%)				$\chi^2 = 1.262$	0.261
Vaginal delivery	83 (33.2)	48 (36.4)	35 (29.7)		
Cesarean section	167 (66.8)	84 (63.6)	83 (70.3)		
Respiratory distress syndrome, *n* (%)				$\chi^2 = 1.325$	0.250
Absent	143 (57.2)	80 (60.6)	63 (53.4)		
Present	107 (42.8)	52 (39.4)	55 (46.6)		
Sepsis, *n* (%)				$\chi^2 = 1.282$	0.258
Absent	224 (89.6)	121 (91.7)	103 (87.3)		
Present	26 (10.4)	11 (8.3)	15 (12.7)		
Advanced maternal age, *n* (%)				$\chi^2 = 1.188$	0.276
No	176 (70.4)	89 (67.4)	87 (73.7)		
Yes	74 (29.6)	43 (32.6)	31 (26.3)		
Albumin (g/L), mean ± SD	33.1 ± 6.9	36.0 ± 6.0	29.9 ± 6.5	t = 7.61	<0.001
Lactate dehydrogenase (U/L), mean ± SD	582.4 ± 381.6	628.0 ± 399.1	531.4 ± 355.9	t = 2.02	0.045
Peak bilirubin (μmol/L), mean ± SD	108.3 ± 61.2	78.5 ± 45.9	141.5 ± 59.2	t = −9.33	< 0.001
C-reactive protein (mg/L), median (IQR)	4.4 (0.1, 9.9)	4.5 (0.1, 10.1)	4.1 (0.1, 9.3)	Z = −0.30	0.764

Hemoglobin (g/L), mean ± SD	162.9 ± 28.0	160.9 ± 28.4	165.2 ± 27.4	t = −1.24	0.217
B/A ratio, mean ± SD	3.4 ± 2.2	2.2 ± 1.3	4.8 ± 2.1	t = −11.82	< 0.001

3.2. Model performance

Based on the independent risk factors identified through multivariate regression analysis, four predictive models were developed: logistic regression, Lasso regression, Support Vector Machine (SVM), and naive Bayes. The cross-validated area under the curve (OOF AUC) of each model within the training set was comparable to its AUC in the validation set, with no apparent evidence of overfitting.

The validation results (Table 2 and Figure 1) showed that the logistic regression model achieved the best discriminative performance, with an AUC of 0.905 (95% CI: 0.834–0.975). The Lasso regression model ranked second, with an AUC of 0.903 (95% CI: 0.832–0.974), followed by the SVM model [AUC = 0.892 (95% CI: 0.816–0.967)] and the naive Bayes model [AUC = 0.884 (95% CI: 0.806–0.961)]. Pairwise DeLong tests revealed no statistically significant differences in AUC between any two models (logistic regression vs. Lasso, $p = 0.64$; logistic regression vs. SVM, $p = 0.30$; logistic regression vs. naive Bayes, $p = 0.17$). Their 95% CIs also overlapped. Considering discrimination, model simplicity, and interpretability together, logistic regression was recommended as the preferred prediction model in this study.

Regarding calibration, all four models had H-L test p values greater than 0.05 (logistic regression, $p = 0.081$; Lasso, $p = 0.175$; SVM, $p = 0.079$; naive Bayes, $p = 0.395$), indicating no statistically significant discrepancies between predicted probabilities and observed outcomes. It should be noted that the H-L test has limited power when applied to a small validation set; therefore, $p > 0.05$ indicates that no significant lack of fit was detected. All four models achieved relatively low Brier scores (0.133 for logistic regression, 0.131 for Lasso regression, 0.144 for SVM, and 0.146 for naive Bayes), suggesting relatively small discrepancies between predicted probabilities and actual outcomes. The calibration curves (Figure 2) showed that the predicted probability curves of the logistic regression and Lasso regression models were closest to the ideal calibration line (45-degree diagonal), particularly in the moderate-to-high probability ranges, where both models demonstrated excellent calibration performance. The separate calibration curve for the best-performing model (logistic regression) further confirmed its calibration stability across different probability ranges. The logistic regression model had a calibration intercept of 0.663 and a calibration slope of 0.949. The slope was close to the ideal value of 1, suggesting general agreement between predicted probabilities and observed risks, although slight overestimation was observed at the extreme ends of the predicted probability range.

The Decision Curve Analysis (DCA) curves (Figure 1B) showed that all four models yielded positive net benefits across threshold probabilities ranging from 5% to 85%. Logistic regression and Lasso regression achieved the highest net benefits across most threshold ranges, indicating that applying these models in clinical decision-making could provide greater net benefit than either the "treat-all" or "treat-none" strategy. To assess the robustness of the collinearity-handling approach, two sensitivity analyses were conducted. First, peak bilirubin was replaced with the B/A ratio, while the other variables remained unchanged, to construct an alternative three-variable model. Its validation-set AUC was 0.900, almost identical to that of the primary model (0.905), suggesting substantial overlap in the information provided by peak bilirubin and the B/A ratio and indicating that the variable selection in the primary model did not result in any substantial loss of predictive performance. Second, gestational age was additionally included in the alternative model to construct a four-variable model, which achieved a validation-set AUC of 0.942. Both sensitivity analyses used the same fitting procedure as the primary model.

Table 2. Performance metrics of the models in the validation set

Model	AUC (95% CI)	H-L statistic	H-L *p* value	Brier score
Logistic regression	0.905 (0.834-0.975)	14.02	0.081	0.133
Lasso regression	0.903 (0.832-0.974)	11.50	0.175	0.131
SVM	0.892 (0.816-0.967)	14.12	0.079	0.144
Naive Bayes	0.884 (0.806-0.961)	8.41	0.395	0.146

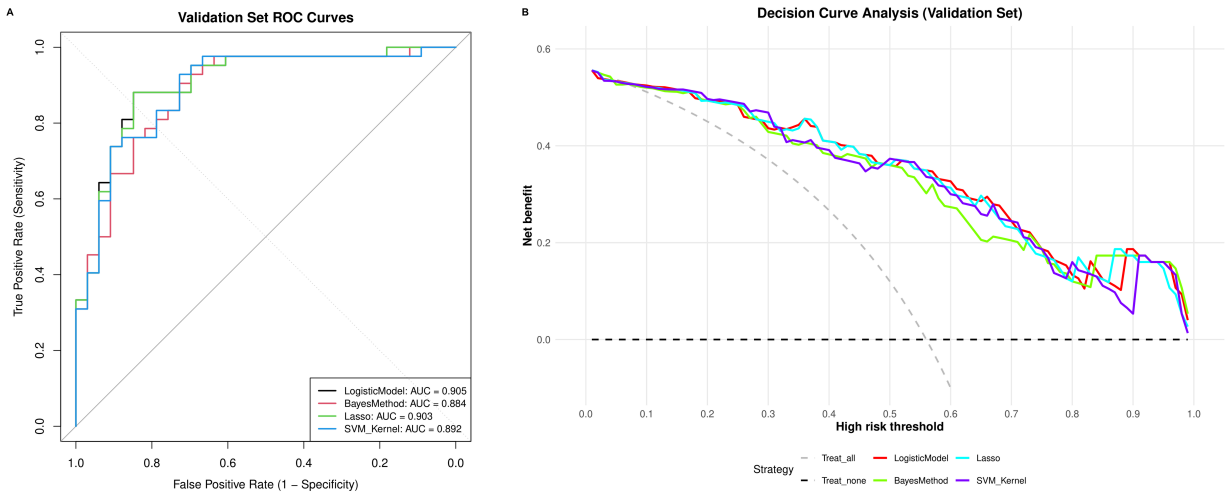


Figure 1. ROC curves and decision curves of the models in the validation Set. (A) ROC curves of the four machine learning models; (B) Decision Curve Analysis (DCA)

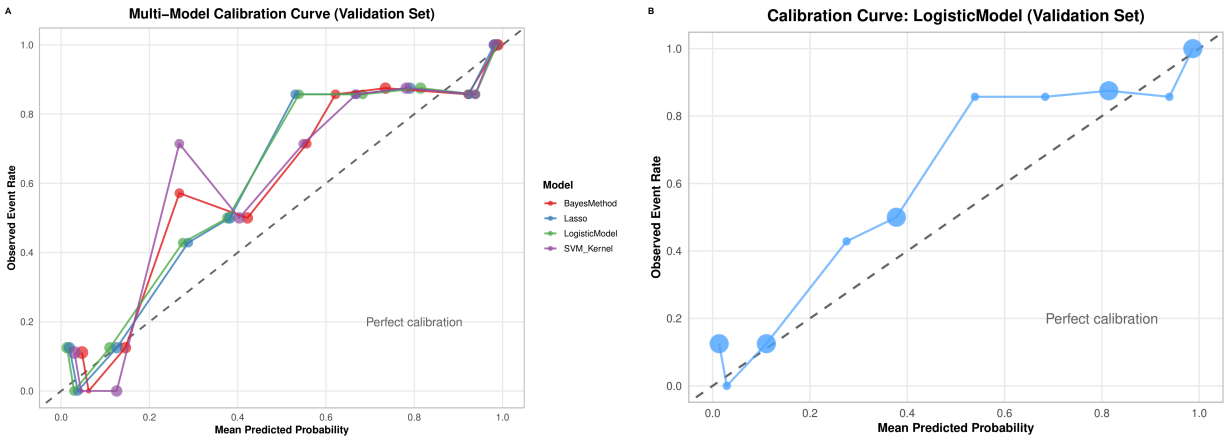


Figure 2. Calibration curves of the models in the validation set. (A) Comparison of the calibration curves of the four models; (B) individual calibration curve of the best-performing model (logistic regression)

3.3. Shap-based interpretability analysis

SHAP analysis was performed on the recommended model (logistic regression). The global feature importance analysis (Figure 3A) showed that peak bilirubin had the highest mean absolute SHAP value (0.207), ranking first in terms of contribution, followed by albumin (0.173) and multiple pregnancy (0.057). The beeswarm plot

(Figure 3B) further illustrated the direction and distribution of feature contributions. Peak bilirubin and multiple pregnancy showed predominantly positive SHAP values, with red dots concentrated in the positive SHAP region. This indicates that higher peak bilirubin levels and multiple pregnancy shifted the model predictions toward an ineffective phototherapy outcome. Albumin was negatively associated with the predicted outcome: individuals with low albumin levels (blue dots) were concentrated in the positive SHAP region, indicating that low albumin levels shifted predictions toward an ineffective phototherapy outcome and were associated with an increased risk of phototherapy failure. It should be noted that the present model is a linear logistic regression model; therefore, the SHAP feature importance ranking was broadly consistent with the relative contributions of the standardized regression coefficients. The added value of SHAP analysis lies primarily in its ability to provide individual-level contribution decomposition (Figure 4), thereby facilitating nursing communication and risk-stratified decision-making.

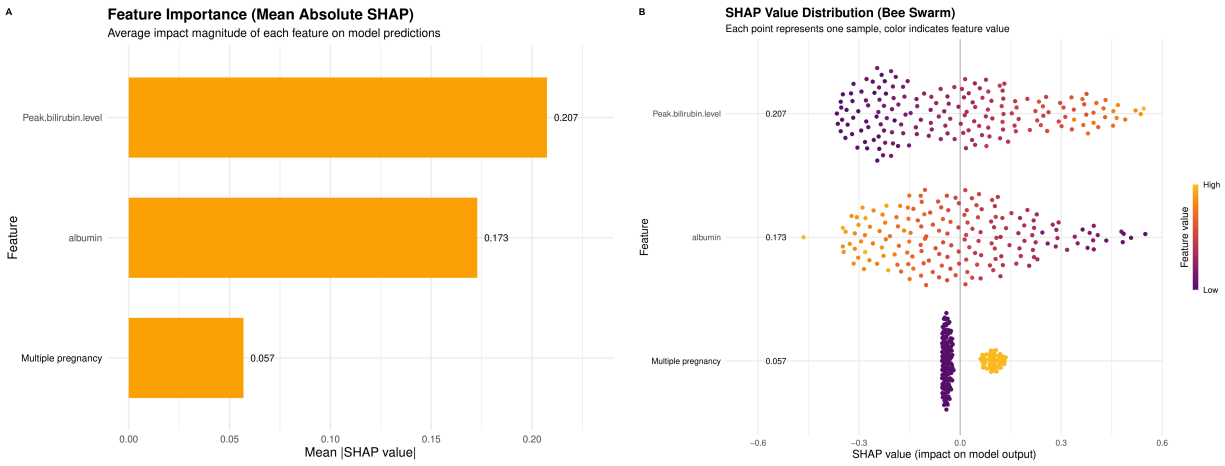


Figure 3. Global SHAP interpretability analysis. (A) feature importance bar plot; (B) SHAP beeswarm plot

The local feature interpretation analysis (Figure 4) presents SHAP force plots and waterfall plots for two representative individuals. Figures 4A–B show the SHAP analysis results for a high-risk infant. This infant had a peak bilirubin level of 187.9 μmol/L and an albumin level of only 16.9 g/L. These features jointly increased the predicted probability of an ineffective phototherapy response to above 0.99, and the model correctly predicted an ineffective outcome. Among the contributing features, albumin had the largest positive contribution (+0.348), followed by peak bilirubin (+0.233), with both jointly increasing the predicted probability. Figures 4C–D present the SHAP analysis results for a low-risk infant. This infant had a peak bilirubin level of only 26.4 μmol/L and an albumin level of 37.6 g/L. These low-risk characteristics reduced the predicted probability to below 0.02, and the model correctly predicted an effective phototherapy response. Among the contributing features, peak bilirubin had the most pronounced negative contribution (−0.271), followed by albumin (−0.116).

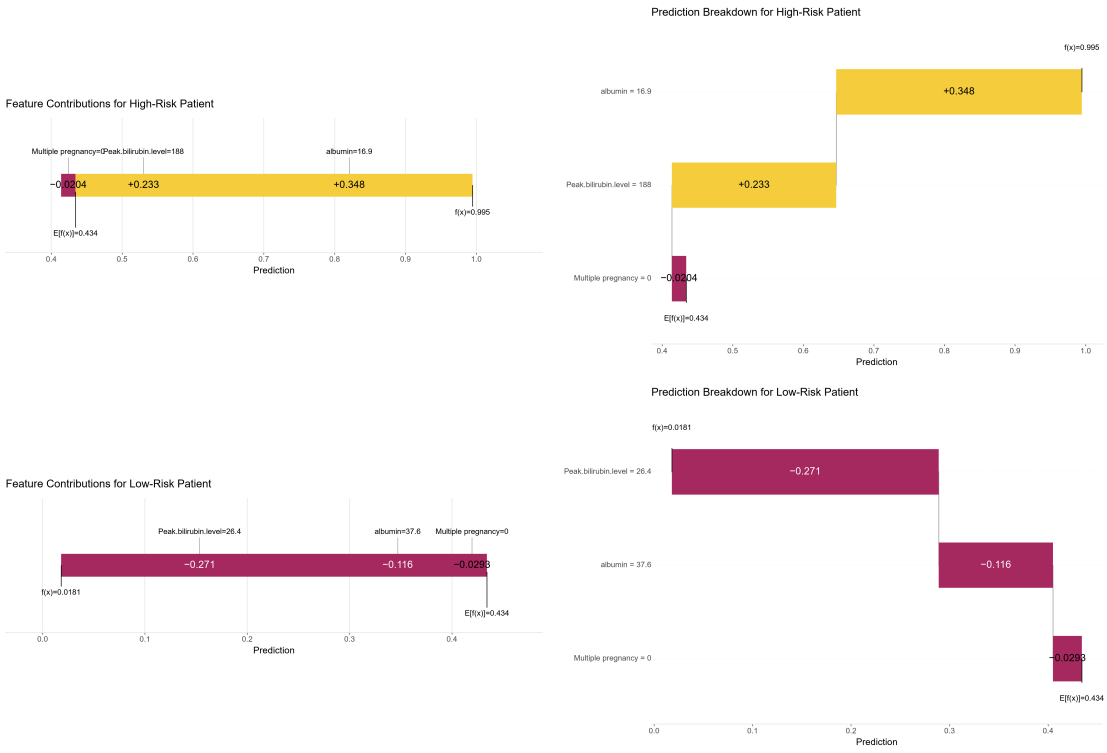


Figure 4. Local SHAP interpretability analysis. (Above) SHAP force plot and waterfall plot for the high-risk infant; (Below) SHAP force plot and waterfall plot for the low-risk infant

4. Discussion

4.1. Main findings

This study used clinical data from 250 preterm infants. Variable selection was performed exclusively in the training set (n = 175). Univariate analysis identified four variables—albumin, peak bilirubin, the bilirubin-to-albumin ratio (B/A ratio), and multiple pregnancy—all with $p < 0.05$. After the B/A ratio, which was highly collinear with peak bilirubin, was excluded following collinearity assessment, four machine learning prediction models were developed using three variables: peak bilirubin, albumin, and multiple pregnancy. The validation results showed that the logistic regression model had the best overall performance (AUC = 0.905), while the AUCs of the other three models were all above 0.88. All four models demonstrated good calibration (H-L $p > 0.05$) and positive clinical net benefit. SHAP analysis further identified peak bilirubin, albumin, and multiple pregnancy as the three most influential predictors, providing a quantitative basis for clinical risk identification and nursing interventions.

Compared with previous studies, the model performance in this study was comparable to that reported in the existing literature. Liu et al. [26] developed a prediction model for phototherapy failure in neonates based on admission indicators, with AUCs of 0.958 (95% CI: 0.924–0.993) and 0.961 (95% CI: 0.914–1.000) in the training and validation sets, respectively. Daunhawer et al. [27] applied the Early Phototherapy Prediction tool (EPPT) to another independent cohort (n = 1,109), achieving an AUROC of 84.6%, which increased to 88.8% after model retraining. Keren et al. [28] developed a prediction model based on the hour-specific bilirubin percentile nomogram proposed by Bhutani et al. [29], which was primarily intended for term and late preterm infants. Another study predicting the need for phototherapy in preterm infants included fewer variables and

did not report individual-level interpretability analyses [30]. The validation-set AUC in the present study was 0.905, placing it in the same general range as those reported in the aforementioned studies. Its main additional contribution is the provision of an individual-level SHAP interpretation framework. It should be noted that the prediction outcomes, gestational age distributions of the study populations, and model development procedures in these studies were not fully consistent with the 24-hour phototherapy efficacy assessment used in the present study. Therefore, direct comparisons of AUC values should be interpreted as reference comparisons only. In addition, the cross-validation AUCs in the training set were comparable to the validation-set AUCs for all four models, with no apparent overfitting. However, the validation set was relatively small (n = 75), and the 95% CIs for the AUC estimates were wide. The stability of these estimates should therefore be interpreted with caution.

4.2. Clinical significance of the key predictors

SHAP analysis showed that peak bilirubin was the most important predictor of phototherapy efficacy in preterm infants. Peak bilirubin directly reflects the infant's bilirubin burden. Previous studies have reported a predictable nonlinear relationship between the extent of serum bilirubin reduction after phototherapy and the pre-treatment bilirubin concentration. The pre-treatment concentration is a major determinant of phototherapy efficiency and can be used to individualize phototherapy duration and light dosage [31]. This finding is consistent with the established understanding that bilirubin burden is a central indicator in phototherapy decision-making. Albumin ranked second in terms of contribution and may affect phototherapy efficacy in two ways. First, as a carrier protein, it influences bilirubin transport and excretion. Second, hypoalbuminemia may indicate impaired hepatic synthetic function and is often accompanied by reduced overall metabolic capacity [32]. Multiple pregnancy ranked third. Multiple pregnancies are often associated with placental insufficiency and intrauterine growth restriction, and infants from multiple pregnancies may have relatively limited bilirubin metabolic capacity. Importantly, in the present dataset, the B/A ratio was highly collinear with peak bilirubin (r = 0.906, VIF = 21.20), and the two variables were therefore unsuitable for simultaneous inclusion in the model. Peak bilirubin was selected as the representative variable to reflect bilirubin burden directly. It should be emphasized that the B/A ratio is an important surrogate indicator of unbound bilirubin burden in the assessment of neurotoxicity risk in preterm infants [33, 34]. Its exclusion as an independent predictor in this study does not imply that its clinical value is unimportant; rather, the decision was made to avoid unstable coefficient estimates resulting from its high collinearity with peak bilirubin.

During variable selection, gestational age was significantly associated with phototherapy efficacy in the baseline comparison of the full cohort ($p < 0.001$). However, it did not meet the $p < 0.05$ threshold in the univariate screening conducted in the training set ($p = 0.086$) and was therefore not included in the multivariable regression model. The primary model ultimately included three variables—peak bilirubin, albumin, and multiple pregnancy—to maintain consistency and simplicity in relation to SHAP interpretability analysis and potential clinical application. It should be noted that, in a sensitivity analysis, adding gestational age to the three-variable primary model (peak bilirubin, albumin, and multiple pregnancy) yielded a four-variable model with a validation-set AUC of 0.951, higher than that of the three-variable primary model (0.905). Both bootstrap analysis and the DeLong test indicated that this difference was statistically significant (p = 0.011 and 0.007, respectively). However, this result should be interpreted with caution for two reasons. First, the variable set in the primary model was determined according to the prespecified training-set selection criterion ($p < 0.05$). Selecting a model based on its performance in the validation set would violate the TRIPOD principle that the validation set should remain independent and would introduce information leakage. It should be clarified that the differences in discrimination among the four algorithms were not

statistically significant, with all pairwise DeLong test *p* values greater than 0.05. Thus, the model recommendation in this study did not itself involve selecting the best-performing model on the basis of validation-set AUC. Second, the observed difference was based on a validation set of only 75 infants, limiting the stability of the estimate. Accordingly, the three-variable primary model was retained in accordance with the prespecified criteria. The potential incremental value of gestational age requires further evaluation in prospective studies with larger sample sizes (see the limitations in Section 4.5). It is also worth noting that previous studies have suggested that the apparent effects of variables such as birth weight, gestational age, and postnatal age on phototherapy efficiency may be explained by differences in pre-treatment bilirubin concentrations [31]. In contrast, in the present study, gestational age still provided incremental discriminatory value after adjustment for peak bilirubin (see above). This discrepancy may be attributable to differences in outcome definitions (absolute reduction over 24 hours versus binary classification of treatment efficacy) and study population characteristics.

4.3. Recommendations for risk-stratified nursing decisions based on SHAP analysis

The SHAP analysis provides a scientific basis for developing refined nursing decision-making recommendations based on risk stratification. According to the key risk variables identified by the model and the directions of their contributions, this study proposes a "risk stratification–targeted intervention" nursing framework that classifies infants into high-risk and low-risk groups and implements differentiated nursing plans. The proposed clinical implementation pathway is as follows: at the initiation of phototherapy, the primary nurse completes a risk assessment and assigns the infant to a group using the model scoring sheet ; infants in the high-risk group receive focused follow-up from NICU specialist nurses and are included in joint medical and nursing ward rounds for reassessment; all nursing interventions are documented in the nursing record, and handovers focus on the risk group and key variables (peak bilirubin, albumin, and multiple pregnancy); the departmental quality control team reviews risk-stratification implementation rates and phototherapy-related adverse events monthly to establish a continuous quality improvement cycle. A second assessment is performed 24 hours after phototherapy based on the actual reduction in TSB, and the nursing plan is adjusted across risk groups when necessary.

4.3.1. Nursing recommendations for high-risk infants

Criteria for the high-risk group: a model-predicted probability of ineffective phototherapy > 0.5. Based on this criterion, among the 75 infants in the validation set, 35 were classified as high-risk, of whom 31 experienced ineffective phototherapy (88.6%). The low-risk group comprised 40 infants, of whom 11 experienced ineffective phototherapy (27.5%). These findings suggest that probability-based stratification by the model can effectively distinguish differences in phototherapy response. The nursing priorities are as follows:

(1) Optimization of the phototherapy regimen: for infants predicted by the model to be at high risk, particularly those with high peak bilirubin levels and low albumin levels, the nursing team should proactively communicate with the medical team and, in accordance with medical orders, assess whether the phototherapy regimen should be adjusted (e.g., increasing irradiance to 30 μW/(cm²·nm) and reducing the distance from the light source to 30 cm). The frequency of TSB monitoring should also be increased to once every 4–6 hours [5]. Meanwhile, the reduction in TSB after 24 hours of phototherapy should be closely monitored. If the reduction does not reach the 20% efficacy threshold, the physician should be notified promptly so that the treatment regimen can be adjusted.

(2) Management of albumin levels: SHAP analysis showed that low albumin was an important factor contributing to ineffective phototherapy. For infants with albumin levels < 30 g/L, nursing staff should check whether human serum albumin infusion has been prescribed. During infusion, transfusion procedures must be

strictly followed, the infusion rate controlled, and potential allergic reactions monitored. For infants who can tolerate enteral feeding, early initiation of feeding and gradual increases in milk volume should be encouraged to promote albumin synthesis through enteral nutrition.

(3) Special nursing care for infants from multiple pregnancies: preterm infants from multiple pregnancies often have insufficient intrauterine nutritional reserves and placental insufficiency. Nursing care should therefore place greater emphasis on temperature management (maintaining an axillary temperature of 36.5–37.5°C), reducing insensible water loss (maintaining humidity at 60%–80%), and accurately recording 24-hour fluid intake and output. As insensible water loss may increase by approximately 20%–30% during phototherapy, fluid intake should be adjusted appropriately, with concurrent monitoring of weight changes and electrolyte levels.

(4) Preventive nursing care for adverse reactions: high-risk infants may require longer phototherapy, which increases their risk of adverse reactions. The nursing team should establish preventive measures, including monitoring body temperature every 2 hours and promptly managing fever by loosening swaddling and adjusting the incubator temperature; inspecting the skin every 4 hours to facilitate early identification of rashes and Bronze Baby syndrome; checking the position of eye shields at every nursing shift to prevent phototoxic retinal injury; and documenting bowel movements, providing enhanced perineal care for infants with diarrhea, and monitoring indicators of dehydration.

4.3.2. Nursing recommendations for low-risk infants

Criteria for the low-risk group: a model-predicted probability of effective phototherapy > 0.5. The nursing priorities are as follows:

(1) Implementation of the standard phototherapy regimen: low-risk infants can receive the standard phototherapy regimen (irradiance of 10–15 μW/(cm²·nm)) to avoid unnecessary adverse reactions associated with excessive phototherapy. Nursing staff should administer phototherapy in accordance with medical orders, ensure an adequate duration of continuous exposure, and monitor changes in TSB over time. Once TSB has decreased to a safe range, the physician should be notified promptly to assess whether phototherapy can be discontinued.

(2) Family health education and psychological support: the relatively stable condition of low-risk infants provides a favorable opportunity for family health education. Nursing staff should explain the basic knowledge of hyperbilirubinemia, the necessity and safety of phototherapy, and key aspects of home observation, such as how to monitor skin jaundice, feeding techniques, and follow-up arrangements. These measures can help alleviate family anxiety and improve adherence to care after discharge.

(3) Promotion of early feeding and excretion: breastfeeding or formula feeding should be encouraged on demand to ensure adequate caloric intake [35]. Tube feeding should be used for preterm infants with uncoordinated sucking and swallowing functions. Nursing staff should document urine and stool output and bowel movement frequency at every shift, assess intestinal motility, and promote bilirubin excretion through the intestinal tract [9].

(4) Developmental supportive care: Kangaroo Mother Care during interruptions in phototherapy may help promote parent–infant bonding and stabilize vital signs [36]. The NICU environment should be kept quiet (noise level < 45 dB), and nursing procedures should be clustered to reduce stimulation from light and touch and support neurobehavioral development.

To further objectively evaluate the actual effectiveness of the risk-stratified nursing strategy, the following measurable outcome indicators are recommended for inclusion in clinical implementation: the incidence of phototherapy-related adverse events (e.g., fever, rash, and diarrhea), the rate of TSB reduction after phototherapy, mean length of hospital stay, and family anxiety scores measured using the State-Trait Anxiety

Inventory (STAI). These indicators can be used to assess the impact of risk stratification on nursing quality and infant outcomes.

4.4. The value of SHAP analysis in supporting nursing decisions

Traditional nursing decisions regarding phototherapy in preterm infants rely primarily on nurses' subjective experience and TSB as a single indicator, lacking quantitative integration of multidimensional risk factors. SHAP analysis provides support for nursing decisions at two levels. At the population level, the ranking of feature importance helps nursing teams identify the risk factors that warrant priority attention and understand the directions of their effects, allowing limited nursing resources to be focused on monitoring and interventions targeting variables with high contributions. At the individual level, force plots and waterfall plots can visually illustrate the risk profile of a specific infant, enabling nursing staff to understand "why this infant is predicted to be at high risk" and thereby deliver more precise, individualized nursing care [15].

Furthermore, the local interpretability of SHAP analysis provides a standardized tool for nursing handovers and risk assessments. By displaying the specific contribution of each feature to the predicted probability, nursing staff can quickly communicate an infant's key risk information during handovers, reducing the likelihood of information being omitted. This data-driven approach to risk communication can help improve team collaboration and efficiency.

4.5. Study limitations

This study has several limitations. First, it was a single-center retrospective study with a limited sample size (n = 250), and the validation set included only 75 infants. The external validity and generalizability of the model require further evaluation in multicenter prospective studies. Second, certain factors that may affect phototherapy efficacy were not included, such as the actual irradiance of the phototherapy equipment, feeding methods, and fluid intake. Future studies should collect these process-related variables. Third, the SHAP analysis was based on a logistic regression model, which assumes linear relationships between variables and may not adequately capture complex nonlinear interactions. Future studies could consider SHAP analysis based on tree-based models, such as XGBoost, to verify the consistency of the findings. Fourth, the proposed risk-stratified nursing intervention strategy has not yet been validated through clinical implementation. Its effectiveness should be evaluated in future randomized controlled trials or before-and-after controlled studies. Fifth, this study specified that laboratory indicators should be based on the most recent available test results before the initiation of phototherapy. However, due to the way electronic medical records were maintained, the actual sampling time for each indicator could not be retrospectively verified for every infant. If any bilirubin measurements were in fact obtained after the outcome assessment, this could have introduced information bias (look-ahead bias). Future prospective studies should standardize the timing of sample collection to ensure that all measurements are obtained before phototherapy begins. Sixth, infants who required exchange transfusion during phototherapy were excluded, but the specific number and baseline characteristics of these excluded infants could not be reported, potentially introducing selection bias. Seventh, a reduction in TSB of ≥ 20% after 24 hours of phototherapy was used as the operational cutoff for determining treatment efficacy. This threshold has not yet been established through a consistent evidence-based consensus. Due to the limitations of the retrospective data, sensitivity analyses using alternative thresholds could not be conducted. Model performance under different efficacy thresholds therefore remains to be validated. Eighth, the phototherapy indications in this study were based on the 2014 *Expert Consensus on the Diagnosis and Treatment of Neonatal Hyperbilirubinemia* [16]. However, much of the available evidence on intervention thresholds comes from populations with gestational ages of ≥ 35 weeks, such as those covered by the

American Academy of Pediatrics guidelines [8], and its applicability to preterm infants with gestational ages of < 35 weeks is limited. China issued an updated *Guidelines for the Diagnosis and Treatment of Neonatal Hyperbilirubinemia* in 2025 [37]. Future studies should redefine phototherapy indications and inclusion criteria for preterm infants in accordance with current guidelines to evaluate the applicability of the present model under the updated criteria.

5. Conclusion

This study developed a multi-algorithm machine learning model to predict phototherapy efficacy in preterm infants with hyperbilirubinemia. All four models demonstrated good discrimination and calibration, with the logistic regression model showing the best overall performance (AUC = 0.905). SHAP analysis identified peak bilirubin, albumin, and multiple pregnancy as the key predictive variables, providing a quantitative basis for clinical risk identification. The resulting recommendations for risk-stratified nursing decisions provide a theoretical basis and preliminary practical framework for the transition from standardized nursing care to precision nursing, offering a new approach to the precision management of hyperbilirubinemia in preterm infants. Further multicenter external validation is needed to assess the generalizability of the model, together with clinical evaluations of the implementation and effectiveness of the proposed nursing interventions.

Author contribution

The first author was responsible for the study design, data analysis, and manuscript preparation. The corresponding author provided research supervision and revised the manuscript. All authors have read and approved the final manuscript.

Funding

Scientific Research Project of Youjiang Medical University for Nationalities (YZCXJH2026016)

Data availability statement

The data used in this study were obtained from the hospital's electronic medical record system. Due to ethical approval requirements and patient privacy protection considerations, the data are not publicly available. Reasonable requests for access to the data may be submitted to the corresponding author.

Ethics approval

This study was approved by the hospital's Ethics Committee (Approval No. 2026091701). Informed consent was waived because of the retrospective study design.

References

[1] Bhutani, V. K., Zipursky, A., Blencowe, H., Khanna, R., Sgro, M., Ebbesen, F., Bell, J., Mori, R., Slusher, T. M., Fahmy, N., Paul, V. K., Du, L., Okolo, A. A., de Almeida, M.-F., Olusanya, B. O., Kumar, P., Cousens, S., & Lawn, J. E. (2013). Neonatal hyperbilirubinemia and Rhesus disease of the newborn: Incidence and impairment estimates for 2010 at regional and global levels. *Pediatric Research, 74*(Suppl 1), 86–100.

[2] Kawade, N., & Onishi, S. (1981). The prenatal and postnatal development of UDP-glucuronyltransferase activity towards bilirubin and the effect of premature birth on this activity in the human liver. *Biochemical Journal, 196*(1), 257–260.

[3] Watchko, J. F. (2016). Bilirubin-induced neurotoxicity in the preterm neonate. *Clinics in Perinatology, 43*(2), 297–311.

[4] Morioka, I., Nakamura, H., Koda, T., Ishida, A., Ohama, Y., Iwamoto, S., & Ibara, S. (2015). Current incidence of clinical kernicterus in preterm infants in Japan. *Pediatrics International, 57*(3), 494–497.

[5] Bhutani, V. K., & Committee on Fetus and Newborn. (2011). Technical report: Phototherapy to prevent severe neonatal hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. *Pediatrics, 128*(4), e1046–e1052.

[6] Hansen, T. W. R., Maisels, M. J., Ebbesen, F., & Bratlid, D. (2020). Sixty years of phototherapy for neonatal jaundice – from serendipitous observation to standardized treatment and rescue for millions. *Journal of Perinatology, 40*(2), 180–193.

[7] Arain, Y., Banda, J. M., Faulkenberry, J., & Cousins, B. (2020). Clinical decision support tool for phototherapy initiation in preterm infants. *Journal of Perinatology, 40*(10), 1518–1523.

[8] Kemper, A. R., Newman, T. B., Slaughter, J. L., Wade, K. C., Lauer, B., Bena, J., & AAP Clinical Practice Guideline Revision Panel. (2022). Clinical practice guideline revision: Management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. *Pediatrics, 150*(3), e2022058859.

[9] Maisels, M. J., Watchko, J. F., Bhutani, V. K., & Stevenson, D. K. (2012). An approach to the management of hyperbilirubinemia in the preterm infant less than 35 weeks of gestation. *Journal of Perinatology, 32*(9), 660–664.

[10] Wei, H., Zhang, T., Ji, R., Li, X., Wang, Y., & Chen, L. (2025). Association between red blood cell distribution width coefficient of variation and post-treatment bilirubin decline velocity in neonatal hyperbilirubinemia. *Frontiers in Pediatrics, 13*, 1690164.

[11] Maisels, M. J., Bhutani, V. K., Bogen, D., Newman, T. B., Stark, A. R., & Watchko, J. F. (2009). Hyperbilirubinemia in the newborn infant ≥35 weeks' gestation: An update with clarifications. *Pediatrics, 124*(4), 1193–1198.

[12] Ahlfors, C. E., & Wennberg, R. P. (2004). Bilirubin-albumin binding and neonatal jaundice. *Seminars in Perinatology, 28*(5), 334–339.

[13] Daunhawer, I., Kasser, S., Koch, G., Wellmann, S., & Rellstab, S. (2019). Enhanced early prediction of clinically relevant neonatal hyperbilirubinemia with machine learning. *Pediatric Research, 86*(1), 122–127.

[14] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine, 380*(14), 1347–1358.

[15] Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, R. P., Liston, D. E., Low, D. K. W., Newman, S.-F., Kim, J., Butler, J. N., & Lee, S.-I. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering, 2*(10), 749–760.

[16] Subspecialty Group of Neonatology, Society of Pediatrics, Chinese Medical Association. (2014). Expert consensus on the diagnosis and treatment of neonatal hyperbilirubinemia. *Chinese Journal of Pediatrics, 52*(10), 745–748.

[17] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Methodological, 58*(1), 267–288.

[18] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

[19] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* (pp. 144–152). ACM.

[20] Platt, J. C. (1999). Probabilities for SV machines. In Smola, A. J., Bartlett, P., Schölkopf, B., & Schuurmans, D. (Eds.), *Advances in large margin classifiers* (pp. 61–74). MIT Press.

[21] DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44*(3), 837–843.
[22] Hosmer, D. W., & Lemeshow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics – Theory and Methods, 9*(10), 1043–1069.
[23] Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making, 26*(6), 565–573.
[24] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates.
[25] Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Annals of Internal Medicine, 162*(1), 55–63.
[26] Liu, Q., Tang, Z., Li, H., Zhang, Y., Wang, X., & Chen, J. (2022). The development and validation of a predictive model for neonatal phototherapy outcome using admission indicators. *Frontiers in Pediatrics, 10*, 745423.
[27] Daunhawer, I., Schumacher, K., Badura, A., Koch, G., Wellmann, S., & Rellstab, S. (2023). Validating the early phototherapy prediction tool across cohorts. *Frontiers in Pediatrics, 11*, 1229462.
[28] Keren, R., Luan, X., Friedman, S., Saddlemire, S., Cnaan, A., & Bhutani, V. K. (2008). A comparison of alternative risk-assessment strategies for predicting significant neonatal hyperbilirubinemia in term and near-term infants. *Pediatrics, 121*(1), e170–e179.
[29] Bhutani, V. K., Johnson, L., & Sivieri, E. M. (1999). Predictive ability of a predischarge hour-specific serum bilirubin for subsequent significant hyperbilirubinemia in healthy term and near-term newborns. *Pediatrics, 103*(1), 6–14.
[30] Raba, A. A., O'Sullivan, A., & Miletin, J. (2020). Prediction of the need for phototherapy during hospital stay in preterm infants by transcutaneous bilirubinometry. *Early Human Development, 146*, 105029.
[31] Jährig, K., Jährig, D., & Meisel, P. (1982). Dependence of the efficiency of phototherapy on plasma bilirubin concentration. *Acta Paediatrica Scandinavica, 71*(2), 293–299.
[32] Soeters, P. B., Wolfe, R. R., & Shenkin, A. (2019). Hypoalbuminemia: Pathogenesis and clinical significance. *JPEN Journal of Parenteral and Enteral Nutrition, 43*(2), 181–193.
[33] Ahlfors, C. E. (2010). Predicting bilirubin neurotoxicity in jaundiced newborns. *Current Opinion in Pediatrics, 22*(2), 129–133.
[34] Watchko, J. F., & Maisels, M. J. (2014). The enigma of low bilirubin kernicterus in premature infants: Why does it still occur, and is it preventable? *Seminars in Perinatology, 38*(7), 397–406.
[35] Gartner, L. M., Morton, J., Lawrence, R. A., Naylor, A. J., O'Hare, D., Schanler, R. J., & Eidelman, A. I. (2005). Breastfeeding and the use of human milk. *Pediatrics, 115*(2), 496–506.
[36] Conde-Agudelo, A., & Díaz-Rossello, J. L. (2016). Kangaroo mother care to reduce morbidity and mortality in low birthweight infants. *Cochrane Database of Systematic Reviews, 2016*(8), CD002771.
[37] Subspecialty Group of Neonatology, Society of Pediatrics, Chinese Medical Association, & Editorial Board of the Chinese Journal of Pediatrics. (2025). Guidelines for the diagnosis and treatment of neonatal hyperbilirubinemia (2025). *Chinese Journal of Pediatrics, 63*(4), 338–350.