

RESEARCH

Open Access



A time series algorithm to predict surgery in neonatal necrotizing enterocolitis

Cheng Cui¹, Ling Qiu², Ling Li³, Fei-Long Chen⁴, Xiao Liu⁵, Huan Sun¹, Xiao-Chen Liu¹, Lei Bao^{1*} and Lu-Quan Li^{1*}

Abstract

Background Determining the optimal timing of surgical intervention for Neonatal necrotizing enterocolitis (NEC) poses significant challenges. This study develops a predictive model using the long short-term memory network (LSTM) with a focal loss (FL) to identify infants at risk of developing Bell IIB + NEC early and issue timely surgical warnings.

Methods Data from 791 neonates diagnosed with NEC are gathered from the Neonatal Intensive Care Unit (NICU), encompassing 35 selected features. Infants are categorized into those requiring surgical intervention ($n = 257$) and those managed medically ($n = 534$) based on the Mod-Bell criteria. A fivefold cross-validation approach is employed for training and testing. The LSTM algorithm is utilized to capture and utilize temporal relationships in the dataset, with FL employed as a loss function to address class imbalance. Model performance metrics include precision, recall, F1 score, and average precision (AP).

Results The model tested on a real dataset demonstrated high performance. Predicting surgical risk 1 day in advance achieved precision (0.913 ± 0.034), recall (0.841 ± 0.053), F1 score (0.874 ± 0.029), and AP (0.917 ± 0.025). The 2-days-in-advance predictions yielded (0.905 ± 0.036), recall (0.815 ± 0.057), F1 score (0.857 ± 0.035), and AP (0.905 ± 0.029).

Conclusion The LSTM model with FL exhibits high precision and recall in forecasting the need for surgical intervention 1 or 2 days ahead. This predictive capability holds promise for enhancing infants' outcomes by facilitating timely clinical decisions.

Keywords Deep learning, Long short-term memory network, Neonatal necrotizing enterocolitis, Predictive surgery, Auxiliary diagnosis

*Correspondence:

Lei Bao
400702@hospital.cqmu.edu.cn
Lu-Quan Li
liluquan123@163.com

¹Neonatal Diagnosis and Treatment Center of Children's Hospital of Chongqing Medical University, National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing Key Laboratory of Child Rare Diseases in Infection and Immunity, Chongqing 400014, China

²The First People's Hospital Of Longquanyi District, Chengdu 610100, China

³Guang'an District Maternal and Child Health Care and Family Planning Service Center, Chengdu 638000, China

⁴College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

⁵College of Safety Engineering, China University of Mining and Technology, Beijing 221116, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Necrotizing enterocolitis (NEC) is a severe intestinal disorder in newborns, occurring at a rate of 5–10% in extremely low birth weight infants [1–3]. Early symptoms include feeding intolerance, gastric retention and respiratory pauses. As the condition progresses, infants may develop abdominal distension, vomiting, bloody stools, and, in severe cases, intestinal perforation, necessitating emergency surgical intervention. The mortality rate for medical NEC is 23.5%, which increases to approximately 30–35% for infants requiring surgery [3, 4]. Even survivors may face complications such as short bowel syndrome, intestinal stenosis, and developmental delays in the nervous system [5, 6]. Treatment for NEC is typically approached through medical and surgical means. In cases where medical treatment fails, early surgery can salvage necrotic bowel segments [7] and reduce the risk of intestinal stenosis and full-thickness necrosis, thereby minimizing complications and mortality [8, 9]. Accurately and early predicting the need for surgical intervention in NEC holds significant importance.

Currently, the determination of whether surgical intervention is necessary for NEC involves both absolute and relative indications [10]. Absolute indication includes pneumoperitoneum [10], which is typically confirmed only after the disease has progressed to intestinal perforation. When conservative treatment fails and symptoms persist, relative indications for surgical intervention include portal venous gas, severe intestinal wall gas, non-leakage abdominal fluid accumulation, and elevated C-reactive protein (CRP) and procalcitonin (PCT), severe acidosis, and decreased platelet count [10, 11]. While these indicators strongly suggest surgery, their lower specificity and subjective nature contribute to controversy over the optimal timing.

Machine learning offers the potential to identify hidden information in large datasets, providing a new perspective for NEC intervention. Masi et al. [12] achieved 87.5% accuracy in classifying NEC using a fecal metagenomics sequencing-based prediction model with 48 samples. Similarly, Lin et al. [13] reported significant results with an NEC prediction system based on fecal microbiota analysis.

However, for NEC, the timing of surgery is determined based on the dynamic changes observed in clinical symptoms, blood tests, and imaging examinations. Therefore, compared to static data, time-series data holds greater value. The recurrent neural network (RNN) [14] is a machine learning algorithm capable of recognizing time series. However, issues such as gradient vanishing and exploding gradients arise when dealing with long sequences. Long short-term memory (LSTM) [15–17] avoids these problems and has been widely used in the medical field for predicting surgical complications

[18] and mortality [19]. Moreover, up to now, there are no reports on time series algorithms predicting NEC surgery.

In this study, we developed a new NEC prediction model using the LSTM algorithm and Focal Loss (FL) [20–22]. Our goal is to identify infants who will require surgical intervention for NEC and to provide an alert 1 or 2 days in advance. We also analyzed the significant features that contribute to the model's predictions.

Materials and methods

Study location and ethics

This retrospective study was conducted in compliance with Helsinki standards, with approval from the Ethics Committee of Children's Hospital of Chongqing Medical University (No.2023–594). The study was conducted at the Neonatal Treatment Center of the hospital, with consent obtained from parents or authorized guardians for the use of infants' data in research.

Inclusion and exclusion criteria

The study included neonates diagnosed with NEC admitted to the Neonatal Intensive Care Unit (NICU) at Children's Hospital of Chongqing Medical University between April 2017 and April 2022. Exclusion criteria encompassed cases of re-hospitalization, those with over 20% missing personal information, parents refusing surgery, as well as diagnoses of esophageal atresia, duodenal atresia, anal atresia, inguinal hernia incarceration, gastric wall developmental defects, Meckel's diverticulum, congenital megacolon, congenital hypertrophic pyloric stenosis, meconium peritonitis, intestinal torsion/congenital malrotation, and infants without NEC diagnosis.

Diagnostic criteria for NEC adhered to Mod-Bell criteria, involving one or more clinical signs (bilious gastric aspirate or emesis, abdominal distention, and occult and/or gross blood in stool (no fissures)), and the presence of at least one of the following three radiographic or sonographic findings: ① pneumatosis intestinalis, ② portal vein gas, and/or ③ pneumoperitoneum [23, 24].

Surgical interventions were determined by senior pediatric surgeons, guided by criteria including intestinal perforation or ineffectiveness of conservative medical treatment with worsening clinical status [7], supported by pathological biopsy. The inclusion and exclusion process of this study is shown in Fig. 1.

Feature selection

Feature selection focused on identifying variables influencing the decision for NEC surgery. Subjective indicators like abdominal distension and mental reactions were excluded due to the difficulty in quantification. Selected features included demographic data, routine stool test results, inflammatory markers, blood analyses, blood gas

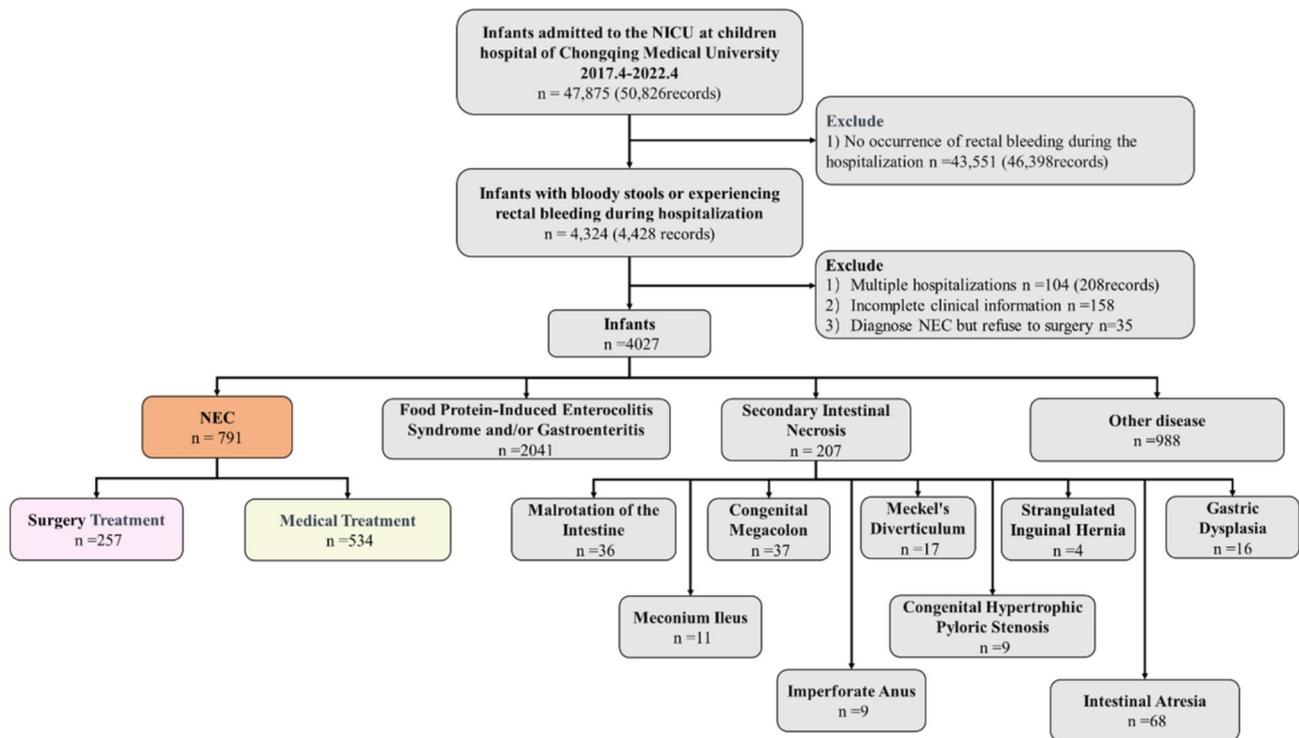


Fig. 1 Standardized exclusion criteria: the flowchart demonstrates the selection of research cases based on inclusion and exclusion criteria

analysis, abdominal ultrasound (AUS), and standardized abdominal X-ray (AR) data, totaling 35 features (Table 1). AUS provided insights into bowel characteristics such as echogenicity, peristalsis, bowel perfusion, bowel wall pneumatosis, free gas, and abdominal effusion. Therefore, these features were extracted from clinical reports [25]. AR assessments utilized the Duke Abdominal Assessment Scale (DASS) to standardize reporting and mitigate subjective bias in feature extraction [26].

Data preprocessing

The original dataset included intermittent laboratory and imaging examinations conducted by pediatricians based on the clinical condition of infants. To capture temporal changes in these features, we constructed a time-series feature set. Initially we applied one-hot encoding to the dataset labels. The time series was defined with start and end points: data collection ceased at the time of surgery for infants undergoing NEC surgery, and for those with medical NEC, it continued until the last positive fecal occult blood (OB) test. Due to the non-continuous nature of these data, we resampled the time axis at specific intervals. Different sampling intervals resulted in varying rates of missing data; shorter intervals extended the time series but increased missing data rates, while longer intervals reduced missing data rates but shortened the series, affecting temporal dependencies. For intervals of 1, 2, and 3 days, average missing data percentages were 72.01%, 66.38%, and 62.01%, respectively. Subsequently,

we compared the model performance across different sampling intervals. Considering the trade-off between time series length and the marginal benefit of reducing missing data rates, a 2-day sampling interval was selected. The missing data rates and model performance are presented in Figs. 2 and 3.

To address missing values on the time axis, we initially applied forward and backward filling techniques. Completely missing discrete features (e.g., gestational age and birth weight) were filled with -1 . Continuous features with missing values in both NEC surgical and non-NEC surgical groups were imputed using their respective means. This preprocessing step resulted in a complete time series dataset.

Model

LSTM is an RNN architecture widely used for sequence modeling and time series analysis [5, 6]. Unlike traditional RNNs, LSTM features gating mechanisms and memory cells. The gating mechanisms selectively retain or omit crucial information at each time step in longer sequences, while the memory cells are responsible for storing and updating the internal state as long-term memory. This selective mechanism enables LSTM networks to efficiently preserve essential information in extended sequences, overcoming the issue of vanishing gradients in RNN algorithms.

The gating mechanism primarily consists of three components: the forget gate (f), the input gate (i), and the

Table 1 Feature description

Type	Description	
Demographics	Gender	Delivery Method
	Pregnancy day	Gravida
	Birth Weight	Para
Blood	Blood oxygen partial pressure(PaO ₂)	White blood cell count (WBC)
	Actual base excess (ABE)	Percentage of eosinophils
	Concentration of hydrogen carbonate (HCO ₃ ⁻)	Neutrophil percentage (N%)
	pH	Platelet count(PLT)
	Carbon dioxide partial pressure(PaCO ₂)	Absolute value of neutrophil
	Lactic acid (Lac)	Immature Neutrophil/Total Neutrophil(I/T)
	Procalcitonin (PCT)	C-reactive protein(CRP)
Stool	Occult blood (OB)	Fecal leukocytes
	Fecal red blood cells	
Image	Bowel dilation (AUS)	Intestinal dilation (AR)
	Intestinal pneumatosis (AUS)	Widening of the interloop space (AR)
	Thickening of the bowel wall (AUS)	intestinal pneumatosis (AR)
	Portal venous gas (AUS)	Portal venous gas (AR)
	Intra-abdominal fluid accumulation (AUS)	Pneumoperitoneum (AR)
	Pneumoperitoneum (AUS)	
	Patent ductus arteriosus (PDA)	

output gate (o). These gates control information flow, allowing the network to determine what to remember or forget at each time step.

The forget gate f is a sigmoid layer that decides which information to discard from the cell state. At time step t , the forget gate f_t controls the extent to which the previous memory cell state C_{t-1} should be forgotten. It takes the input features x_t and the previous hidden state h_{t-1} as inputs and outputs a value between 0 and 1 for each element of the memory cell. 1 represents “completely retaining this value”, and 0 means “throwing this value completely”. The calculation method for the forget gate is as follows:

$$f_t = \sigma(W_f \cdot [h_{(t-1)}, x_t] + b_f)$$

where f_t is the output of the forget gate, W_f is the weight of the linear layer, h_{t-1} is the hidden state of the previous moment, x_t is the current input, and b_f represents the bias vectors.

The input gate i_t is a sigmoid layer that determines what information need to store in cell state. A tanh layer creates a vector \tilde{C}_t for the new candidate value, which can be added to the state. Then i_t and \tilde{C}_t are merged to update the state.

$$i_t = \sigma(W_i \cdot [h_{(t-1)}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{(t-1)}, x_t] + b_c)$$

Then, we can obtain the cell state C_t in step t by forgetting C_{t-1} and adding the input information \tilde{C}_t limitedly, as follows:

$$C_t = f_t \odot C_{(t-1)} + i_t \odot \tilde{C}_t$$

where \odot is hadamard product, denotes pointwise multiplication operation for two vectors. Ultimately, the output gate o_t determines the content of output, cell state pressing the value between -1 and 1 by Tanh, and multiply with o_t to obtain the hidden state h_t .

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \odot \tanh(C_t)$$

The loss function is a critical metric in neural network training, assessing the discrepancies between model predictions and intended outputs. A smaller loss function indicates that the model's predictions are closer to the actual values, reflecting better performance. Well-known loss functions include the mean squared error (MSE), commonly used in training regression models, and the widely applied cross-entropy (CE) loss, which is used in classification tasks.

Although CE is widely used, it has the property that easily classified instances result in a significant loss. This issue may have a negative impact on rarer classes (NEC surgery group). FL [27–31] addresses this problem by reshaping the CE function to assign less importance to easy examples and to focus more on harder examples. This reshaped loss function allows the model to better

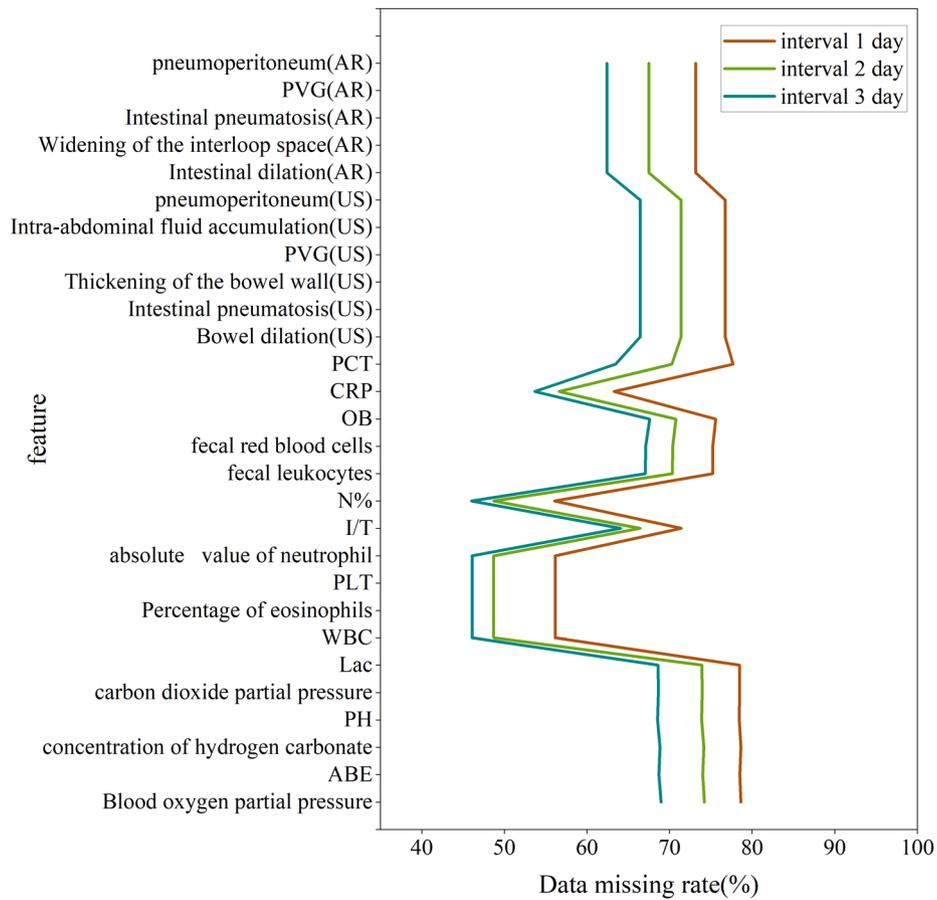


Fig. 2 The data missing rate at different sampling intervals: Shows the data missing situation of 35 features at sampling intervals of 1 day, 2 days, and 3 days, respectively

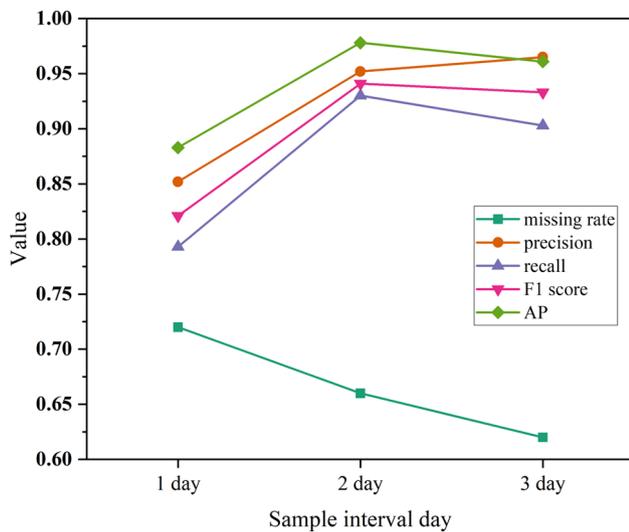


Fig. 3 Model performance at different sampling intervals and missing data rates

distinguish between different classes, especially rarer classes, resulting in improved overall performance.

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t), \gamma \geq 0$$

where p_t is the predicted probability for the class, α is a balancing variant, and γ is the tunable focusing parameter. The modulation factor aims to decrease the weights of easily categorized medical NEC infants during training, thereby directing the model focus toward more difficult-to-classify ones. In cases when an infant is misclassified and the predicted p_t is small, the value of the modulation factor is close to 1, resulting in minimal impact on the loss. When $\gamma=0$ and $\alpha=1$, the FL becomes equivalent to the CE function.

The entire training process of the model is shown in Fig. 4, which includes data processing, LSTM with FL model training, and result output. The parameters of the LSTM model are recorded in Table 2.

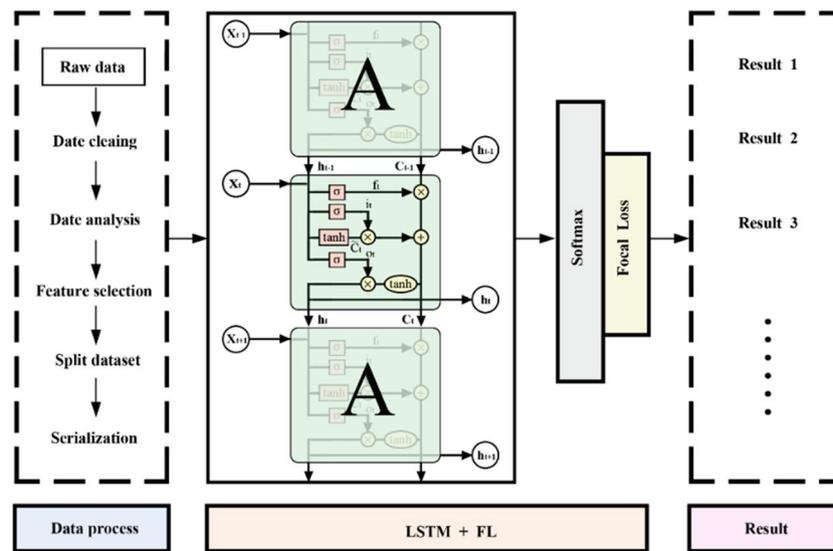


Fig. 4 Flowchart of the model prediction process. After preprocessing the raw data, selecting features, dividing into training and testing sets, and serializing, the data is input into the LSTM model with FL as the loss function, resulting in predictions

Table 2 The parameters of the LSTM model

Optimizer	Dropout	Epoch	Batch size	Feature dimension	Hidden dimension	Output dimension
Adam	0.5	100	128	20	128	2

Evaluation matrix and internal validation

The model’s evaluation utilized 5-fold cross-validation, where 80% of the data was used for training and the remaining 20% for testing. This approach ensured each fold’s training and testing sets encompassed different infants, enhancing the reliability of the results. To evaluate the model’s performance, we employ standard evaluation metrics, including precision, recall, and F1 score [32, 33]. In handling imbalanced datasets, the F1 score, a balanced measure derived from the harmonic mean of precision and recall, is commonly employed. The F1 score ranges from 0 to 1, with higher values indicating better performance. Let TP, TN, FP, and FN represent the true positive, true negative, false positive, and false negative in the confusion matrix. Three evaluation metrics can be obtained easily using the following formulas:

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

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

$$F1\ score = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

Additionally, the area under the curve (AUC) of the receiver operating characteristic (ROC) curve is

commonly used to evaluate classification models. However, the ROC curve effectiveness can be limited when there is an extreme imbalance between positive and negative samples.

To address this limitation, the precision-recall curve (PR curve) provides a more robust evaluation of performance on imbalanced datasets [34]. The PR curve is better suited for extreme imbalances, because it reflects the trade-off between precision and recall, highlighting the model’s effectiveness in balancing precision and recall. The AUC of the PR curve, known as average precision (AP), will be utilized to evaluate the model’s performance.

Statistical analysis

Quantitative data were presented as mean±standard deviation or median (interquartile range), while categorical data were expressed as frequency(percentage). Group comparisons were conducted using the chi-square test, Mann-Whitney U test, or t-test, with $P < 0.05$ considered statistically significant. Data analysis and model development were performed using Excel LTSC 2021 and Python 3.8.0 leveraging TensorFlow 2.4 as the deep learning framework. Feature importance was assessed through ANOVA analysis, Image processing and visualization were conducted using PowerPoint LTSC 2021, Photoshop 2023, and Origin 2019. The experimental procedures were conducted on a computer equipped with an AMD

Ryzen 5 5600 6-Core Processor, operating at 3.5 GHz CPU, 32 GB of RAM, and an AMD 5600XT GPU.

Results

Dataset description

We obtained medical records from a big data center comprising 47,875 hospitalized infants (50,826 records) admitted to the NICU between April 2017 and April 2022. Applying standardized exclusion criteria (Fig. 1), we excluded 104 cases (208 records) of readmitted infants, 158 infants with more than 20% missing personal information, and 35 cases of infants refusing surgery were excluded to ensure dataset integrity and quality. This study focused on predicting NEC surgery and early

warnings, 791 infants diagnosed with NEC, included 257 infants who underwent surgical intervention and 534 who received medical treatment.

Among those in the surgery group, 194 were premature, and 63 were full-term. While the medical group comprised 292 premature and 242 full-term infants. Table 3 presents a clinical feature comparison between the two groups.

Optimization and performance of focal loss function

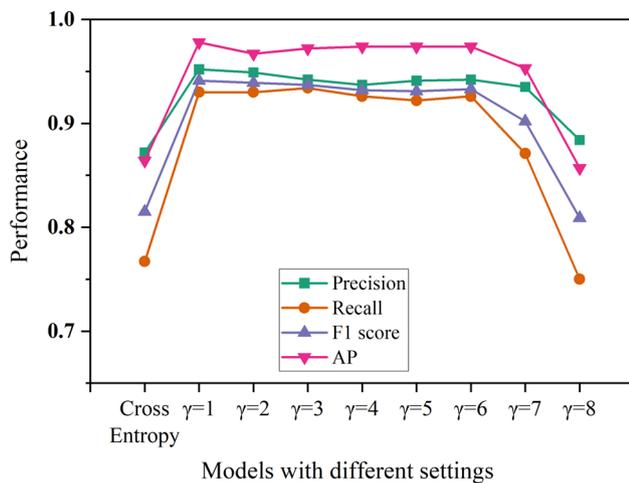
We investigated the impact of the hyperparameter γ on the performance of FL in our model. FL comprises two hyperparameters: α and γ . The hyperparameter γ dynamically adjusts the rate of weight reduction for

Table 3 Clinical features of 791 NEC infants

Characteristics	Medical NEC (n=534)	Surgical NEC (n=257)	P
Gender (male/female)	314/220	149/108	0.825
Birth weight, ($\bar{x} \pm s$, g)	2466.26 \pm 770.23	2089.97 \pm 804.88	<0.001
Gestational age, ($\bar{x} \pm s$, w)	36.12 \pm 3.24	34.07 \pm 3.72	<0.001
Arterial blood gas analysis			
PaO ₂ , [M(P ₂₅ , P ₇₅), mmHg]	79(65,93)	91(73,133)	<0.001
ABE, [M(P ₂₅ , P ₇₅), mmol/L]	-1.9(-4.4,0.4)	-2.1(-4.8,0)	0.002
HCO ₃ ⁻ , [M(P ₂₅ , P ₇₅), mmol/L]	22.4(20.6,25.2)	22.6(20.5,25.4)	0.419
pH, [M(P ₂₅ , P ₇₅)	7.42(7.36,7.47)	7.38(7.30,7.45)	<0.001
PaCO ₂ , [M(P ₂₅ , P ₇₅), mmHg]	36(30,41)	40(33,48)	<0.001
Lactic acid, [M(P ₂₅ , P ₇₅), mmol/L]	1.9(1.3,2.8)	1.4(0.9,2.1)	<0.001
Blood test			
WBC, [M(P ₂₅ , P ₇₅), 10 ⁹ /L]	11.6(9.1,14.5)	10.9(7.6,14.9)	0.013
Percentage of eosinophils, [M(P ₂₅ , P ₇₅), %]	3.0(1.0,5.0)	4.0(2.0,7.7)	<0.001
PLT, [M(P ₂₅ , P ₇₅), 10 ⁹ /L]	366(258,476)	258(169,380)	<0.001
Absolute value of neutrophil, [M(P ₂₅ , P ₇₅), 10 ⁹ /L]	4.9(3.4,7.2)	5.3(3.2,8.0)	0.151
VT, [M(P ₂₅ , P ₇₅), %]	6.1(4.4,8.0)	7.1(4.5,12.8)	<0.001
Percentage of neutrophil, [M(P ₂₅ , P ₇₅), %]	45(35,57)	52(41,63)	<0.001
CRP, [M(P ₂₅ , P ₇₅), mg/dl]	8(8,8)	8(8,21)	<0.001
PCT, [M(P ₂₅ , P ₇₅), ng/ml]	0.21(0.12,0.81)	0.46(0.18,3.53)	<0.001
Stool routine			
Fecal leukocytes, [M(P ₂₅ , P ₇₅), /HP]	1(0,2)	3(1,4)	<0.001
Fecal RBC, [M(P ₂₅ , P ₇₅), /HP]	1(0,2)	1(0,2)	<0.001
OB+, n(%)	0(0,1)	1(0,1)	<0.001
Abdominal ultrasound			
Bowel dilation, n(%)	168/534(31.5)	69/257(26.8)	0.185
Intestinal pneumatosis, n(%)	49/534(9.2)	35/257(13.3)	0.058
Thickening of the bowel wall, n(%)	62/534(11.6)	43/257(16.7)	0.047
Portal venous gas, n(%)	135/534(25.3)	53/257(20.6)	0.149
Intra-abdominal fluid accumulation, n(%)	143/534(26.8)	116/257(45.1)	<0.001
Pneumoperitoneum, n(%)	4/534(0.0)	11/257(4.3)	<0.001
Abdominal radiograph			
Intestinal dilation, n(%)	452/534(84.6)	202/257(78.6)	0.035
Widening of the interloop space, n(%)	170/534(31.8)	77/257(30.0)	0.594
Intestinal pneumatosis, n(%)	240/534(44.9)	100/257(38.9)	0.108
Portal venous gas, n(%)	36/534(6.8)	41/257(16.0)	<0.001
Pneumoperitoneum, n(%)	0/534(0.0)	22/257(8.6)	<0.001

Table 4 Performance comparison between models with different settings

Loss function		Precision	Recall	F1 score	AP
Cross Entropy		0.872 ± 0.042	0.767 ± 0.060	0.815 ± 0.044	0.864 ± 0.034
Focal Loss	$\gamma=1$	0.952 ± 0.015	0.930 ± 0.015	0.941 ± 0.012	0.978 ± 0.006
	$\gamma=2$	0.949 ± 0.009	0.930 ± 0.026	0.939 ± 0.015	0.967 ± 0.023
	$\gamma=3$	0.942 ± 0.016	0.934 ± 0.023	0.937 ± 0.013	0.972 ± 0.006
	$\gamma=4$	0.937 ± 0.013	0.926 ± 0.015	0.932 ± 0.006	0.974 ± 0.014
	$\gamma=5$	0.941 ± 0.011	0.922 ± 0.025	0.931 ± 0.013	0.974 ± 0.008
	$\gamma=6$	0.942 ± 0.020	0.926 ± 0.023	0.933 ± 0.007	0.974 ± 0.009
	$\gamma=7$	0.935 ± 0.037	0.871 ± 0.024	0.902 ± 0.020	0.953 ± 0.016
	$\gamma=8$	0.884 ± 0.069	0.750 ± 0.051	0.809 ± 0.034	0.857 ± 0.040

**Fig. 5** The performance of models with different settings: As γ gradually increases from 1 to 8, the overall performance of the model exhibits a fluctuating downward trend. Compared to the Cross-Entropy (CE) loss function, the focal loss (FL) function demonstrates superior performance

simple samples leveraging the rapid scaling property of the power function. In contrast, α balances the importance of samples from different categories but does not significantly improve overall model performance. Therefore, we focused on the influence of the γ on the experimental performance, determining the optimal value for subsequent research. We compared the model's performance using CE as the loss function against the model using FL. Table 4; Fig. 5 illustrate performance changes with varying γ .

As γ increases from 1 to 8, model performance trends to decrease. At $\gamma=1$, the model achieved its highest F1 score (0.941 ± 0.012) and AP (0.978 ± 0.006). In comparison, the model with CE as the loss function had a lower F1 score (0.815 ± 0.044) and AP (0.864 ± 0.034). These results indicate that using FL as a loss function helps the model focus more on learning difficult-to-classify samples, thereby enhancing classification performance.

Performance of surgical NEC early prediction

In this section, we evaluate the model's ability to predict the need for surgery 1–2 days in advance. We truncated

the end of the time series data by 24 h and 48 h, respectively, and processed the data with the model to generate surgical prediction probabilities using precision, recall, F1 score, and AP evaluations. Figure 6 compares the need for surgery predictions 1 and 2 days in advance, with values representing 5-fold averages. The PR curve in Fig. 7 provides an intuitive display of surgical prediction performance across the 5-fold results. For predicting surgery 1-day in advance, the model achieved high performance with a precision of 0.913 ± 0.034, recall 0.841 ± 0.053, F1 score 0.874 ± 0.029, and the AP was 0.917 ± 0.025. For 2 days in advance, the model also exhibited commendable performance, with a precision of 0.905 ± 0.036, recall of 0.815 ± 0.057, F1 score of 0.857 ± 0.035, and AP of 0.905 ± 0.029. A comparison of the forecasts indicates slightly better predictive performance for surgery 1 day in advance than for surgery 2-days in advance (Figs. 6 and 7). Figures 8 and 9 depict the confusion matrices for model classification, further illustrating the model's effectiveness.

Compare with traditional machine learning

Traditional machine learning models typically use cross-sectional data for predictions, whereas LSTM models leverage short-term sequential data. In order to compare the effects of these two types of data on the predictive outcomes, we selected four traditional machine learning models: Naïve Bayesian Model (NBM), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Light Gradient Boosting Machine (LightGBM). The comparison results are presented in Table 5. The results show that both SVM and LightGBM perform well, with LightGBM surpassing LSTM+FL in terms of precision. However, LSTM+FL demonstrates significant performance advantages in other metrics, highlighting its effectiveness in utilizing sequential data for NEC surgery prediction.

One-way analysis

Furthermore, we compare the performance of models trained using individual clinical features versus those trained with a combination of clinical and imaging data. Based on feature importance, we decide to use CRP, I/T,

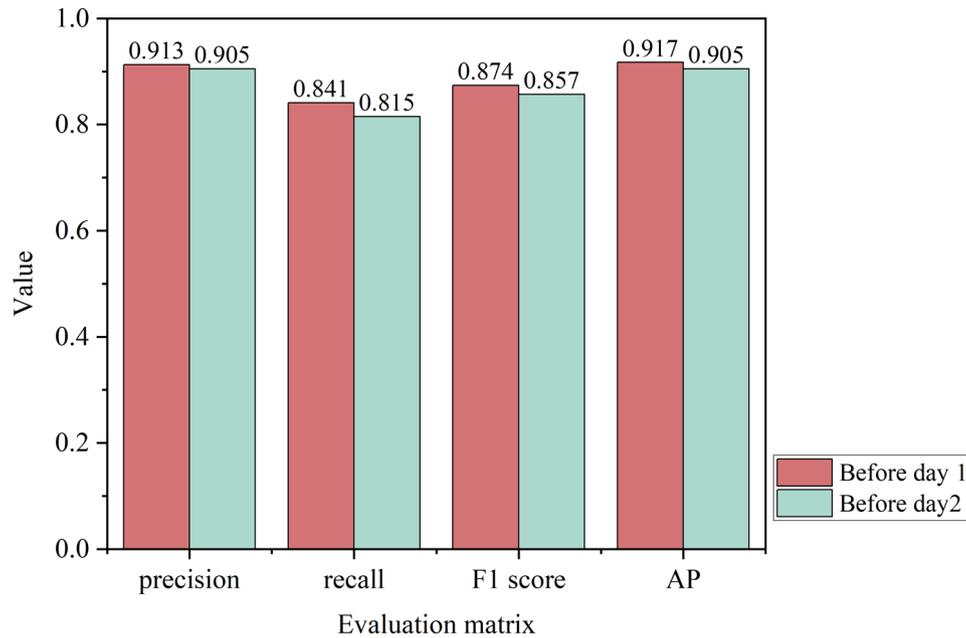


Fig. 6 Surgical Prediction PR Curve for 1 Day and 2 Days

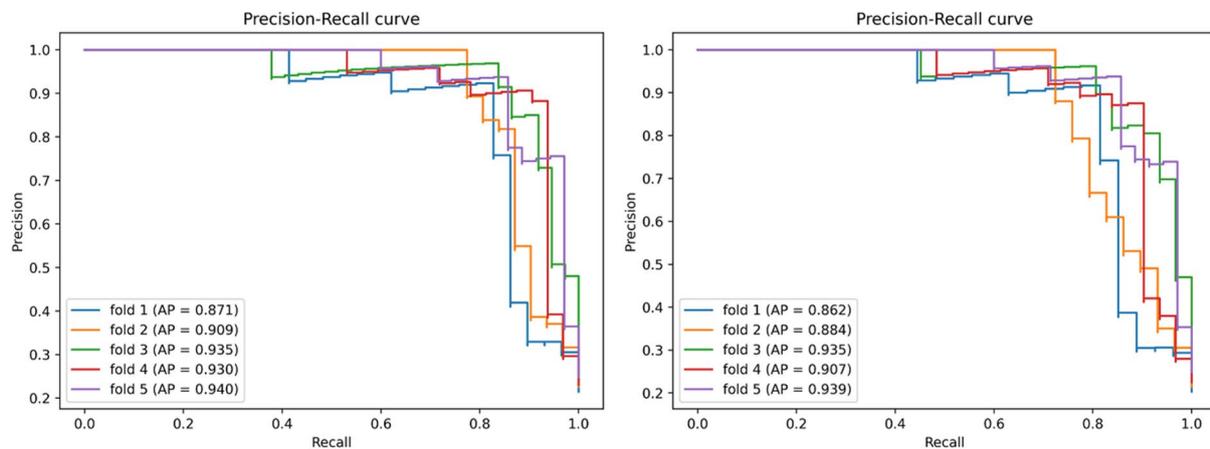


Fig. 7 The Average Performance in Predicting Surgery 1 or 2 Days in Advance. The left plot illustrates the performance of predicting NEC surgery 1 day in advance, while the right plot displays the performance of predicting NEC surgery 2 days in advance

PCT, and WBC to train the models separately. In terms of the AP evaluation metric, models trained using CRP alone achieve similar performance to those trained with combined clinical and imaging data. However, models trained with both clinical and imaging data still exhibit unparalleled performance advantages. The results are presented in Table 6.

Feature importance

Inflammatory markers (CRP, I/T, fecal leukocytes, PCT) made significant contributions to the model. Furthermore, eosinophil percentage, blood oxygen pressure, blood pH value, carbon dioxide pressure, and gestational age also played important roles in predicting NEC

surgery. A higher feature score indicates a greater contribution to the model, as shown in Fig. 10.

Discussion

Early prediction of the timing for NEC surgery and timely intervention are crucial for improving the prognosis of NEC. Early prediction of surgery can provide advance warning time, which helps reduce the time for observation and follow-up decisions, thereby potentially decreasing mortality and complications. In this study, we constructed a novel model for predicting NEC surgery based on LSTM to recognize time series and FL to address imbalanced data. This model holds the potential

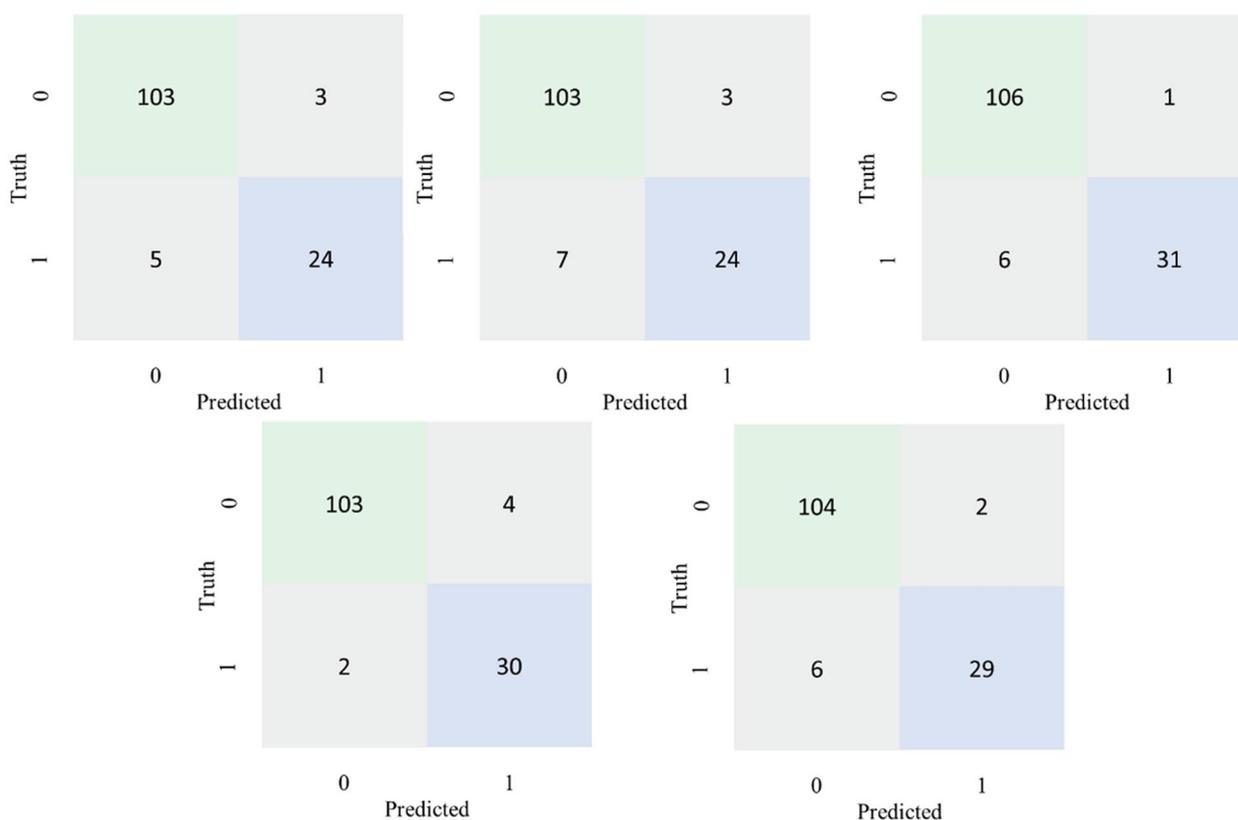


Fig. 8 Predict Confusion Matrix: 1 Day in Advance - Confusion matrix for predicting NEC surgery one day in advance

to enhance outcomes for NEC patients by facilitating timely interventions.

CRP emerges as a significant predictor in our model for determining the timing of NEC surgery. Previous research supports the predictive role of elevated CRP levels in NEC surgery and post-operative intestinal strictures [9, 35, 36], making it a common marker for monitoring and diagnosing NEC among surgeons [11]. The ratio of immature neutrophils to total neutrophils (I/T), which is very useful for distinguishing neonatal infections [37]. Elevated I/T levels, indicative of severe infection, also play a crucial role. Miner et al. [38] found that increased I/T levels help distinguish between Bell stage II and III of NEC. These easily accessible markers enhance the model’s applicability in clinical settings.

The duration of abnormal fecal leukocytes may also play a significant role in the model. In the medical NEC group, the duration of abnormal fecal leukocytes has a median of 4 days (interquartile range: 2–8 days), and a mean of 5.97 days. Conversely, the surgical intervention NEC group has a median duration of 10 days (interquartile range: 4–14 days), and a mean of 9.82 days. The Mann-Whitney U test indicates a significant difference ($p < 0.001$) between the two groups, suggesting the

duration of abnormal fecal leukocytes in NEC infants warrants further investigation.

The trend of PCT changes serves as a valuable indicator for predicting the timing of NEC surgery. PCT can be detected 2 h after bacterial infection and rapidly increases within 6 h post-infection [39]. Liebe et al. [40] suggested that infants with NEC may require surgery if PCT levels exceed 1.4 ng/ml. However, limited consideration is given to the trend of PCT changes. Turner et al. [41] monitored PCT in suspected NEC and septic children for three consecutive days, finding no statistically significant differences. Conversely, Cetinkaya et al. [42] discovered that in infants with NEC Bell stage III, PCT persists longer and decreases more slowly compared to sepsis. This highlights the potential impact of feature change trends on predictive outcomes.

Choosing to issue early warnings within 2 days is appropriate, as longer intervals might miss rapid changes in the patient’s condition. For instance, fulminant NEC can lead to death within 48 h of onset [43], so the model’s warning period should not exceed 48 h to ensure timely decisions. The advantages of early warnings include focused monitoring of high-risk infants, timely communication with families to gain their support, and individualized decision-making based on the warning

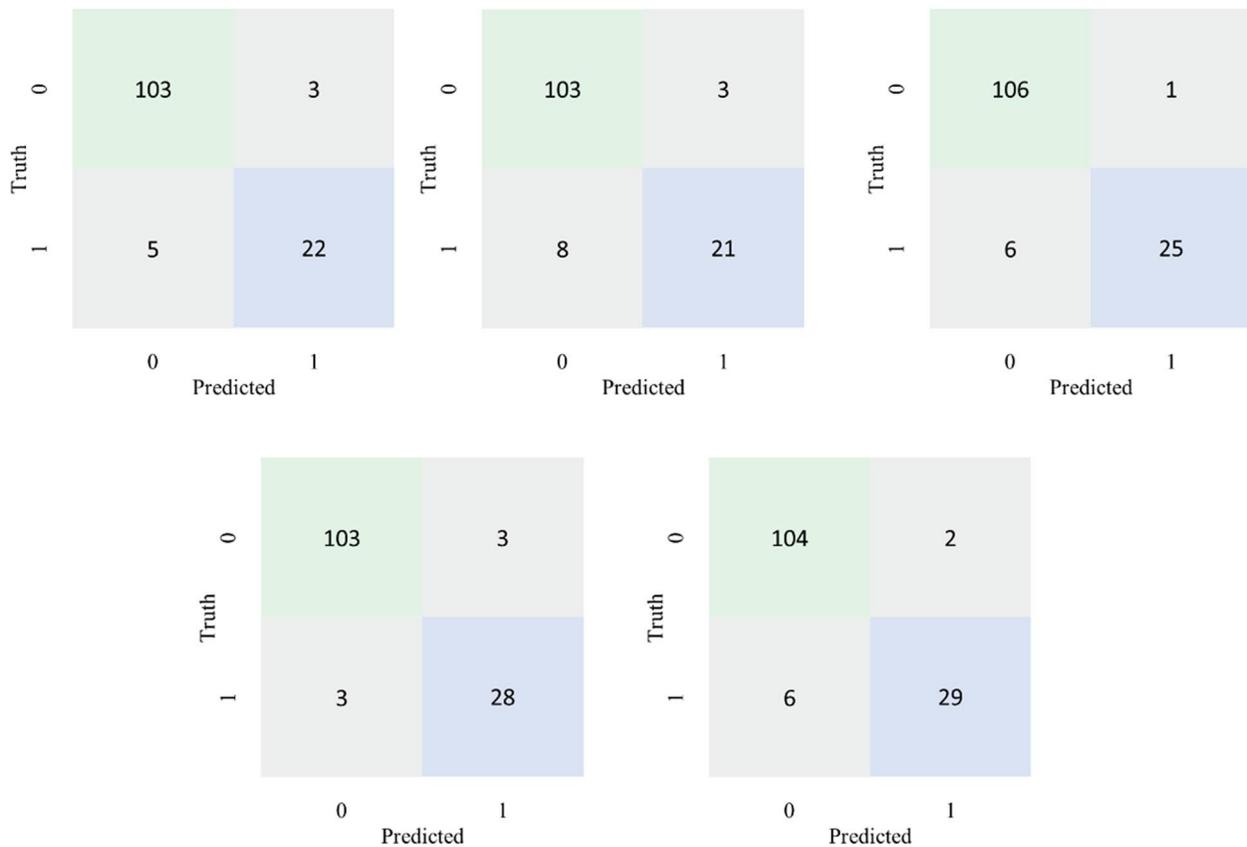


Fig. 9 Predict Confusion Matrix: 2 Day in Advance - Confusion matrix for predicting NEC surgery two days in advance

Table 5 Comparison results with machine learning models

Evaluation	NBM	SVM	KNN	LightGBM	LSTM + FL
Precision	0.835 ± 0.025	0.905 ± 0.033	0.886 ± 0.024	0.963 ± 0.010	0.952 ± 0.015
Recall	0.833 ± 0.029	0.902 ± 0.026	0.812 ± 0.032	0.915 ± 0.029	0.930 ± 0.015
F1 score	0.833 ± 0.027	0.903 ± 0.029	0.834 ± 0.033	0.934 ± 0.023	0.941 ± 0.012
AP	0.812 ± 0.041	0.922 ± 0.025	0.811 ± 0.026	0.885 ± 0.042	0.978 ± 0.006

Table 6 Comparison results with using single clinical feature

Evaluation metrics	CRP	I/T	PCT	WBC	LSTM + FL
Precision	0.893 ± 0.037	0.894 ± 0.033	0.891 ± 0.031	0.887 ± 0.034	0.952 ± 0.015
Recall	0.922 ± 0.012	0.802 ± 0.051	0.770 ± 0.095	0.739 ± 0.055	0.930 ± 0.015
F1 score	0.907 ± 0.016	0.844 ± 0.026	0.822 ± 0.049	0.804 ± 0.027	0.941 ± 0.012
AP	0.977 ± 0.010	0.955 ± 0.022	0.961 ± 0.023	0.932 ± 0.027	0.978 ± 0.006

information. However, there are also drawbacks. Establishing a time series with a 2-day interval requires monitoring the infant with at least one test every two days, such as blood analysis, blood gas analysis, stool tests, AUS or AR. This may lead to increasing the economic burden on patients.

The features collected for our model are less prone to interference, with quick test results and high usability. Some researchers have attempted to predict NEC surgery

using fecal microbiome metagenomics [13]. However, the neonatal gut microbiome is influenced by various factors such as delivery method, infections, feeding practices, antibiotic use, and probiotics [44]. Additionally, the high cost and time-consuming nature of metagenomic testing may limit its widespread application. Our model, based on blood analysis and abdominal imaging, provides timely results, strong usability, and is feasible even in primary care hospitals equipped with basic diagnostic

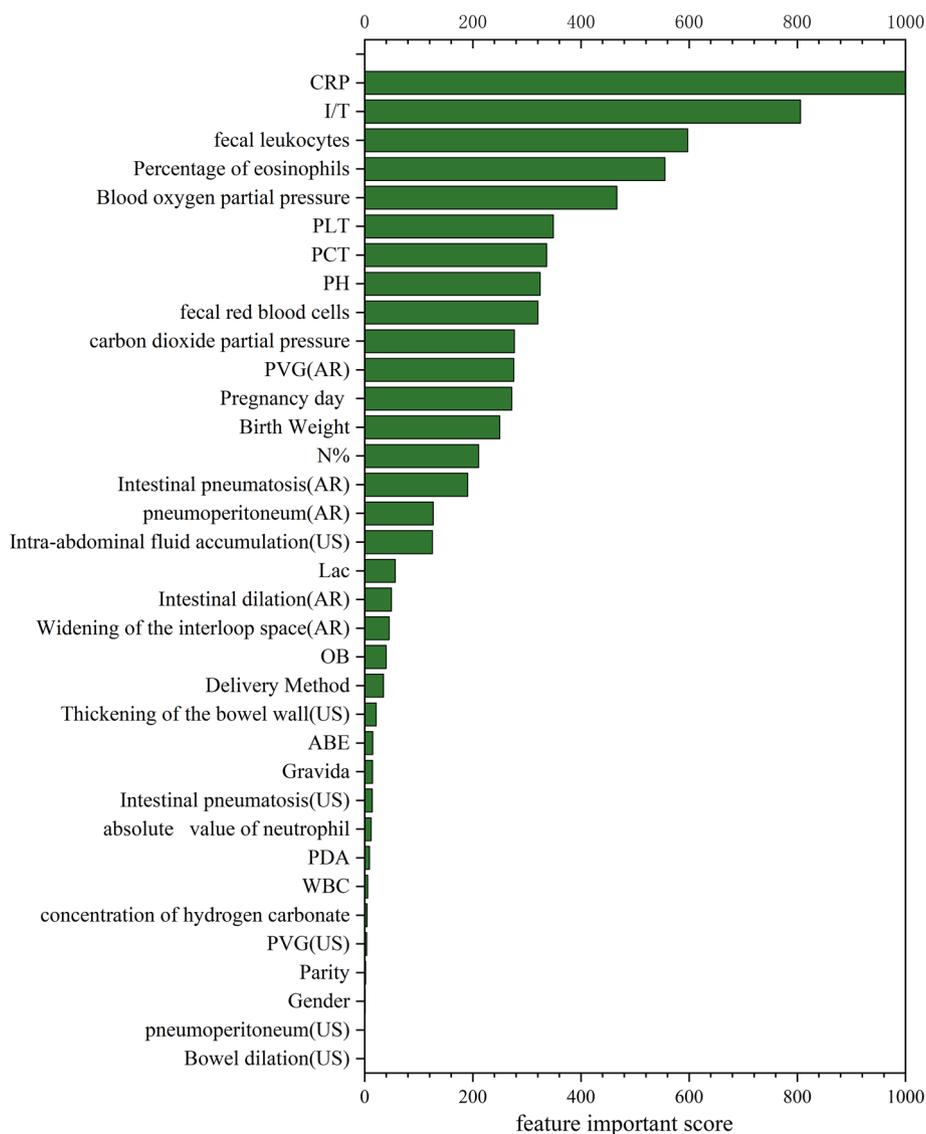


Fig. 10 Clinical Feature Contribution to the Model - Relative contributions of various features in predicting NEC surgery. A higher feature score indicates greater importance of the variable to the model

equipment. Compared to models based on computer vision [45], our approach considers the dynamic changes in features. For instance, abdominal imaging features critical for NEC diagnosis, such as portal venous gas, bowel wall thickening, and pneumatosis, can also appear in other conditions like ischemic bowel necrosis or food protein-induced enterocolitis [46, 47]. Portal venous gas caused by food protein-induced enterocolitis resolves faster with dietary management compared to NEC [47], and Sharma found that nearly half of infants with portal venous gas can survive without surgery [48]. Therefore, dynamically considering the changing trends and speed of imaging characteristics might be one of the directions to consider in the future.

Limitation

This study has several limitations. Firstly, the retrospective nature of the study results in a high rate of data missingness. While a 3-day sampling interval shows a decrease in data missingness compared to 2-day interval, it also results in reduced performance, likely due to the shorter time series and delayed capture of disease changes. Additionally, there was an uneven distribution in the frequency of examinations and tests conducted by doctors, especially for infants with severe conditions, requiring intensive follow-ups for surgical assessments [10]. Future research should implement specific follow-up protocols for high-risk infants to reduce missing data and consider adjusting algorithm weights for specific time periods to enhance generalization ability and

robustness. Besides, our imaging features were extracted from text reports generated using the DASS. Integrating imaging algorithm modules, such as computer vision techniques to extract latent information from AR [45], in future studies may help avoid subjectivity. Lastly, the LSTM algorithm can reduce the subjectivity in NEC surgical decision-making, but its explainability remains a topic of discussion in the academic community. Some researchers believe that high-quality machine learning models can be evidence-based even without explainability [49]. However, measures such as eliminating less relevant features and incorporating attention mechanisms are expected to reduce model complexity and enhance explainability in the future.

Conclusion

The LSTM algorithm is employed to construct a model for diagnosing and predicting the surgical risk of NEC, addressing imbalanced categories using FL. Results showed that this model can serve as an auxiliary tool for surgical decision-making in the NICU.

Acknowledgements

We extend our gratitude to Professor Xian-Ming Xu for assistant in data collection and to Ms. Ru-Yun Hou for providing language support.

Author contributions

CC: conceptualization, Methodology, Analysis, Writing. QL: Literature search, Editing, Visualization, statistical analysis. LL: Investigation, Data collection, Data preprocessing. FLC: Data preprocessing, Modeling, Writing. XL: Visualization, Formula. HS: Visualization, Validation. XCL: Visualization, LQL: conceptualization, funding, review and editing. LB: conceptualization, funding, supervision, validation, review and editing. All authors reviewed the manuscript.

Funding

This study was funded by the Natural Science Foundation of Chongqing City, China (cstc2021jcyj-msxmX0063), and the Municipal Science and Health Joint Project of Chongqing City, China (2022MSXM039).

Data availability

We fully acknowledge the significance of data sharing as stipulated by SCI journals. However, considering the privacy of patients, the raw data, which included hospital data containing sensitive patient information, will not be made public. Nevertheless, we have meticulously presented a comprehensive account of the experimental design, analysis, results, Should the esteemed editor and reviewers require further elucidation or specific inquiries pertaining to the data, we pledge our utmost commitment to providing detailed explanations and clarifications.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Children's Hospital of Chongqing Medical University (No.2023 – 594) on January 22, 2024. All parents of infants or their authorized guardians permission for the hospital to use patient data for research purposes. This study involves a retrospective analysis of medical records. The ethics committee waived the requirement for additional informed consent from parents or authorized guardians, as the study poses minimal risk to infants and maintains anonymity, in accordance with the principles outlined in the Declaration of Helsinki. Therefore, the ethics committee has approved the use of patient data for this study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 10 April 2024 / Accepted: 25 September 2024

Published online: 18 October 2024

References

1. Neu J, Walker WA. Necrotizing enterocolitis. *N Engl J Med*. 2011;364:255–64. <https://doi.org/10.1056/NEJMra1005408>.
2. Bell EF, Hintz SR, Hansen NI, Bann CM, Wyckoff MH, DeMauro SB, Walsh MC, Vohr BR, Stoll BJ, Carlo WA, et al. Mortality, In-Hospital morbidity, Care practices, and 2-Year outcomes for extremely Preterm infants in the US, 2013–2018. *JAMA*. 2022;327:248–63. <https://doi.org/10.1001/jama.2021.23580>.
3. Hein-Nielsen AL, Petersen SM, Greisen G. Unchanged incidence of necrotising enterocolitis in a tertiary neonatal department. *Dan Med J*. 2015;62:A5091.
4. Jones IH, Hall NJ. Contemporary Outcomes for Infants with necrotizing Enterocolitis-A systematic review. *J Pediatr*. 2020;220:86–e923. <https://doi.org/10.1016/j.jpeds.2019.11.011>.
5. McNelis K, Goddard G, Jenkins T, Poindexter A, Wessel J, Helmraht M, Poindexter B. Delay in achieving enteral autonomy and growth outcomes in very low birth weight infants with surgical necrotizing enterocolitis. *J Perinatol*. 2021;41:150–6. <https://doi.org/10.1038/s41372-020-00880-z>.
6. Duric B, Gray C, Alexander A, Naik S, Haffenden V, Yardley I. Effect of time of diagnosis to surgery on outcome, including long-term neurodevelopmental outcome, in necrotizing enterocolitis. *Pediatr Surg Int*. 2022;39:2. <https://doi.org/10.1007/s00383-022-05283-z>.
7. Munaco AJ, Veenstra MA, Brownie E, Danielson LA, Nagappala KB, Klein MD. Timing of optimal surgical intervention for neonates with necrotizing enterocolitis. *Am Surg*. 2015;81:438–43.
8. Ergenekon E, Tayman C, Özkan H. Turkish neonatal Society Necrotizing enterocolitis diagnosis, Treatment and Prevention Guidelines. *Turk Arch Pediatr*. 2021;56:513–24. <https://doi.org/10.5152/TurkArchPediatr.2021.21164>.
9. Zhang H, Chen J, Wang Y, Deng C, Li L, Guo C. Predictive factors and clinical practice profile for strictures post-necrotising enterocolitis. *Med (Baltim)*. 2017;96:e6273. <https://doi.org/10.1097/MD.00000000000006273>.
10. Robinson JR, Rellinger EJ, Hatch LD, Weitkamp J-H, Speck KE, Danko M, Blakely ML. Surgical necrotizing enterocolitis. *Semin Perinatol*. 2017;41:70–9. <https://doi.org/10.1053/j.semperi.2016.09.020>.
11. Gao J, Lai D, Tou J. Survey on surgical treatment of neonatal necrotizing enterocolitis in China 2022. *World J Pediatr Surg*. 2023;6:e000588. <https://doi.org/10.1136/wjps-2023-000588>.
12. Masi AC, Embleton ND, Lamb CA, Young G, Granger CL, Najera J, Smith DP, Hoffman KL, Petrosino JF, Bode L, et al. Human milk oligosaccharide DSLNT and gut microbiome in preterm infants predicts necrotising enterocolitis. *Gut*. 2021;70:2273–82. <https://doi.org/10.1136/gutjnl-2020-322771>.
13. Lin YC, Salleb-Aouissi A, Hooven TA. Interpretable prediction of necrotizing enterocolitis from machine learning analysis of premature infant stool microbiota. *BMC Bioinformatics*. 2022;23:104. <https://doi.org/10.1186/s12859-022-04618-w>.
14. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D*. 2020;404:132306. <https://doi.org/10.1016/j.physd.2019.132306>.
15. Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J. LSTM: a search space odyssey. *IEEE Trans Neural Netw Learn Syst*. 2017;28:2222–32. <https://doi.org/10.1109/TNNLS.2016.2582924>.
16. Yu Y, Si X, Hu C, Zhang J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput*. 2019;31:1235–70.
17. Siami-Namini S, Tavakoli N, Namin AS, IEEE. The Performance of LSTM and BiLSTM in Forecasting Time Series. 2019 *IEEE International Conference on Big Data (Big Data)*. Los Angeles, CA, USA: (2019). pp. 3285–3292 <https://doi.org/10.1109/BigData47090.2019.9005997>
18. Adhikari L, Ozrazgat-Baslanti T, Ruppert M, Madushani RWMA, Paliwal S, Hashemighouchani H, Zheng F, Tao M, Lopes JM, Li X, et al. Improved predictive models for acute kidney injury with IDEA: Intraoperative Data Embedded Analytics. *PLoS ONE*. 2019;14:e0214904. <https://doi.org/10.1371/journal.pone.0214904>.
19. Du H, Ghassemi MM, Feng M. The effects of deep network topology on mortality prediction. *Annu Int Conf IEEE Eng Med Biol Soc (2016)* 2016:2602–5. <https://doi.org/10.1109/EMBC.2016.7591263>

20. Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*. (2017). pp. 2980–2988 http://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html [Accessed November 24, 2023].
21. Peng H, Wu C, Xiao Y. CBF-IDS: addressing Class Imbalance using CNN-BiLSTM with focal loss in Network Intrusion Detection System. *Appl Sci*. 2023;13:11629. <https://doi.org/10.3390/app132111629>.
22. Prabhakar SK, Rajaguru H, Won D-O. Performance analysis of Hybrid Deep Learning models with attention mechanism positioning and focal loss for text classification. *Sci Program*. 2021;2021:1–12. <https://doi.org/10.1155/2021/2420254>.
23. Walsh MC, Kliegman RM. Necrotizing enterocolitis: treatment based on staging criteria. *Pediatr Clin North Am*. 1986;33:179–201. [https://doi.org/10.1016/s0031-3955\(16\)34975-6](https://doi.org/10.1016/s0031-3955(16)34975-6).
24. Li Q-Y, An Y, Liu L, Wang X-Q, Chen S, Wang Z-L, Li L-Q. Differences in the clinical characteristics of early- and late-onset necrotizing enterocolitis in full-term infants: a retrospective case-control study. *Sci Rep*. 2017;7:43042. <https://doi.org/10.1038/srep43042>.
25. Janssen Lok M, Miyake H, Hock A, Daneman A, Pierro A, Offringa M. Value of abdominal ultrasound in management of necrotizing enterocolitis: a systematic review and meta-analysis. *Pediatr Surg Int*. 2018;34:589–612. <https://doi.org/10.1007/s00383-018-4259-8>.
26. Morrison SC, Jacobson JM. The radiology of necrotizing enterocolitis. *Clin Perinatol*. 1994;21:347–63.
27. Petmezas G, Cheimariotis G-A, Stefanopoulos L, Rocha B, Paiva RP, Katsaggelos AK, Maglaveras N. Automated lung sound classification using a hybrid CNN-LSTM Network and focal loss function. *Sensors*. 2022;22:1232. <https://doi.org/10.3390/s22031232>.
28. Petmezas G, Haris K, Stefanopoulos L, Kilintzis V, Tzavelis A, Rogers JA, Katsaggelos AK, Maglaveras N. Automated Atrial Fibrillation detection using a hybrid CNN-LSTM Network on Imbalanced ECG datasets. *Biomed Signal Process Control*. 2021;63:102194. <https://doi.org/10.1016/j.bspc.2020.102194>.
29. Mukhoti J, Kulharia V, Sanyal A, Golodetz S, Torr P, Dokania P. Calibrating Deep Neural Networks using Focal Loss. *Advances in Neural Information Processing Systems*. Curran Associates, Inc. (2020). pp. 15288–15299 <https://proceedings.neurips.cc/paper/2020/hash/aeb7b30ef1d024a76f21a1d40e30c302-Abstract.html> [Accessed November 27, 2023].
30. Tran GS, Nghiem TP, Nguyen VT, Luong CM, Burie J-C. Improving accuracy of lung nodule classification using deep learning with focal loss. *J Healthc Eng*. 2019;2019:1–9. <https://doi.org/10.1155/2019/5156416>.
31. Yeung M, Sala E, Schönlieb C-B, Rundo L. Unified focal loss: generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput Med Imaging Graph*. 2022;95:102026. <https://doi.org/10.1016/j.compmedimag.2021.102026>.
32. Juba B, Le HS. Precision-Recall versus Accuracy and the Role of Large Data Sets. *Proceedings of the AAAI Conference on Artificial Intelligence* (2019) 33:4039–4048. <https://doi.org/10.1609/aaai.v33i01.33014039>
33. Powers D. Evaluation. From Precision, Recall and F-Measure to ROC, Informedness, Markedness & correlation. *J Mach Learn Technol*. 2011;2:37–63.
34. Qi Q, Luo Y, Xu Z, Ji S, Yang T. Stochastic Optimization of Areas Under Precision-Recall Curves with Provable Convergence. *Advances in Neural Information Processing Systems*. Curran Associates, Inc. (2021). pp. 1752–1765 https://proceedings.neurips.cc/paper_files/paper/2021/hash/0dd1bc593a91620daecf7723d2235624-Abstract.html [Accessed November 27, 2023].
35. Gaudin A, Farnoux C, Bonnard A, Alison M, Maury L, Biran V, Baud O. Necrotizing enterocolitis (NEC) and the risk of intestinal stricture: the value of C-reactive protein. *PLoS ONE*. 2013;8:e76858. <https://doi.org/10.1371/journal.pone.0076858>.
36. Duci M, Fascetti-Leon F, Erculiani M, Priante E, Cavicchiolo ME, Verlatto G, Gamba P. Neonatal independent predictors of severe NEC. *Pediatr Surg Int*. 2018;34:663–9. <https://doi.org/10.1007/s00383-018-4261-1>.
37. Krauel J, Salvado A, Mira A, Orellana N, Calvet I, Molina V, Lizarraga I. [Simultaneous determination of total and immature neutrophil C-reactive protein in normal, diseased, and infected newborn infants]. *Esp Pediatr*. 1987;27:257–60.
38. Miner CA, Fullmer S, Eggett DL, Christensen RD. Factors affecting the severity of necrotizing enterocolitis. *J Matern Fetal Neonatal Med*. 2013;26:1715–9. <https://doi.org/10.3109/14767058.2013.798283>.
39. Downes KJ, Fitzgerald JC, Weiss SL. Utility of Procalcitonin as a Biomarker for Sepsis in Children. *J Clin Microbiol*. 2020;58. <https://doi.org/10.1128/JCM.01851-19>.
40. Liebe H, Lewis S, Loerke C, Golubkova A, Leiva T, Stewart K, Sarwar Z, Gin A, Porter M, Chaaban H, et al. A retrospective Case Control Study Examining Procalcitonin as a Biomarker for Necrotizing enterocolitis. *Surg Infect (Larchmt)*. 2023;24:448–55. <https://doi.org/10.1089/sur.2022.366>.
41. Turner D, Hammerman C, Rudensky B, Schlesinger Y, Wine E, Muise A, Schimmel MS. Low levels of procalcitonin during episodes of necrotizing enterocolitis. *Dig Dis Sci*. 2007;52:2972–6. <https://doi.org/10.1007/s10620-007-9763-y>.
42. Cetinkaya M, Ozkan H, Köksal N, Akaci O, Ozgür T. Comparison of the efficacy of serum amyloid A, C-reactive protein, and procalcitonin in the diagnosis and follow-up of necrotizing enterocolitis in premature infants. *J Pediatr Surg*. 2011;46:1482–9. <https://doi.org/10.1016/j.jpedsurg.2011.03.069>.
43. Lin L, Xia X, Liu W, Wang Y, Hua Z. Clinical characteristics of neonatal fulminant necrotizing enterocolitis in a tertiary children's hospital in the last 10 years. *PLoS ONE*. 2019;14:e0224880. <https://doi.org/10.1371/journal.pone.0224880>.
44. Vandenplas Y, Carnielli VP, Ksiazek J, Luna MS, Migacheva N, Mosselmans JM, Picaud JC, Possner M, Singhal A, Wabitsch M. Factors affecting early-life intestinal microbiota development. *Nutrition*. 2020;78:110812. <https://doi.org/10.1016/j.nut.2020.110812>.
45. van Druten J, Sharif MS, Chan SS, Chong C, Abdalla H. A deep learning based suggested model to detect necrotising enterocolitis in abdominal radiography images. 2019 *International Conference on Computing, Electronics & Communications Engineering (ICCECE)*. IEEE (2019). pp. 118–123.
46. Mehr S, Kakakios A, Frith K, Kemp AS. Food protein-induced enterocolitis syndrome: 16-year experience. *Pediatrics*. 2009;123:e459–64.
47. Guo Y, Si S, Jia Z, Lv X, Wu H. Differentiation of food protein-induced enterocolitis syndrome and necrotizing enterocolitis in neonates by abdominal sonography. *J Pediatr (Rio J)*. 2021;97:219–24. <https://doi.org/10.1016/j.jpeds.2020.03.001>.
48. Sharma R, Tepas JJ 3rd, Hudak ML, Wludyka PS, Mollitt DL, Garrison RD, Bradshaw JA, Sharma M. Portal venous gas and surgical outcome of neonatal necrotizing enterocolitis. *J Pediatr Surg*. 2005;40:371–6. <https://doi.org/10.1016/j.jpedsurg.2004.10.022>.
49. McCoy LG, Brenna CTA, Chen SS, Vold K, Das S. Believing in black boxes: machine learning for healthcare does not need explainability to be evidence-based. *J Clin Epidemiol*. 2022;142:252–7. <https://doi.org/10.1016/j.jclinepi.2021.11.001>.

Publisher's note

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