



## OPEN ACCESS

## EDITED BY

Luigi Vetrugno,  
University of Studies G. d'Annunzio Chieti and  
Pescara, Italy

## REVIEWED BY

Yunhuan Liu,  
Tongji University, China  
Fabrizio Tritapepe,  
University of Studies G. d'Annunzio Chieti and  
Pescara, Italy

## \*CORRESPONDENCE

Hongxia Zhang  
✉ aerkaterru@163.com

RECEIVED 16 August 2024

ACCEPTED 18 October 2024

PUBLISHED 01 November 2024

## CITATION

Huang C, Ha X, Cui Y and Zhang H (2024) A study of machine learning to predict NRDS severity based on lung ultrasound score and clinical indicators.  
*Front. Med.* 11:1481830.  
doi: 10.3389/fmed.2024.1481830

## COPYRIGHT

© 2024 Huang, Ha, Cui and Zhang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# A study of machine learning to predict NRDS severity based on lung ultrasound score and clinical indicators

Chunyan Huang<sup>1,2</sup>, Xiaoming Ha<sup>1</sup>, Yanfang Cui<sup>1</sup> and Hongxia Zhang<sup>1\*</sup>

<sup>1</sup>Department of Ultrasound, Yantaishan Hospital, Yantai, China, <sup>2</sup>Medical Impact and Nuclear Medicine Program, Binzhou Medical University, Yantai, China

**Objective:** To develop predictive models for neonatal respiratory distress syndrome (NRDS) using machine learning algorithms to improve the accuracy of severity predictions.

**Methods:** This double-blind cohort study included 230 neonates admitted to the neonatal intensive care unit (NICU) of Yantaishan Hospital between December 2020 and June 2023. Of these, 119 neonates were diagnosed with NRDS and placed in the NRDS group, while 111 neonates with other conditions formed the non-NRDS (N-NRDS) group. All neonates underwent lung ultrasound and various clinical assessments, with data collected on the oxygenation index (OI), sequential organ failure assessment (SOFA), respiratory index (RI), and lung ultrasound score (LUS). An independent sample test was used to compare the groups' LUS, OI, RI, SOFA scores, and clinical data. Use Least Absolute Shrinkage and Selection Operator (LASSO) regression to identify predictor variables, and construct a model for predicting NRDS severity using logistic regression (LR), random forest (RF), artificial neural network (NN), and support vector machine (SVM) algorithms. The importance of predictive variables and performance metrics was evaluated for each model.

**Results:** The NRDS group showed significantly higher LUS, SOFA, and RI scores and lower OI values than the N-NRDS group ( $p < 0.01$ ). LUS, SOFA, and RI scores were significantly higher in the severe NRDS group compared to the mild and moderate groups, while OI was markedly lower ( $p < 0.01$ ). LUS, OI, RI, and SOFA scores were the most impactful variables for the predictive efficacy of the models. The RF model performed best of the four models, with an AUC of 0.894, accuracy of 0.808, and sensitivity of 0.706. In contrast, the LR, NN, and SVM models have lower AUC values than the RF model with 0.841, 0.828, and 0.726, respectively.

**Conclusion:** Four predictive models based on machine learning can accurately assess the severity of NRDS. Among them, the RF model exhibits the best predictive performance, offering more effective support for the treatment and care of neonates.

## KEYWORDS

NRDS, machine learning, risk factor, clinical indicator, prediction model

## Introduction

Neonatal respiratory distress syndrome (NRDS) significantly contributes to neonatal mortality, primarily due to a deficiency of pulmonary surfactant, a critical factor in its complex pathogenesis (1). Timely diagnosis and intervention are crucial for improving the prognosis of affected neonates. Currently, the primary methods for diagnosing NRDS include clinical symptom assessment, imaging studies, and blood gas analysis to monitor respiratory and multi-organ function (2, 3). OI and RI are critical indicators of pulmonary function, widely used to assess the severity of respiratory distress in neonates, and are strongly associated with pulmonary developmental abnormalities (4, 5). These metrics objectively reflect the oxygenation status of neonates and are vital tools for assessing pulmonary ventilation and gas exchange.

The SOFA score is a clinical tool used to evaluate the degree of multi-organ dysfunction in pediatric patients, widely applied in intensive care units and for assessing critically ill neonates. In clinical practice, the severity of NRDS is typically assessed using the OI, RI, and SOFA scoring systems (6). Pulmonary ultrasound effectively reflects pulmonary ventilation, and LUS is well-documented for quantifying ventilation status, particularly in adults (7). Several scoring systems are available for quantitatively assessing pulmonary ultrasound in adults. Studies indicate that LUS scoring in pulmonary ultrasound examinations effectively reflects NRDS severity (8). In some regional hospital ICUs, pulmonary ultrasound has almost replaced traditional chest radiography (9). However, some studies suggest that due to various factors influencing pulmonary ultrasound results, traditional chest radiographs may be more effective for diagnosing NRDS. The clinical applications of quantitative ultrasound methods for diagnosing pulmonary diseases and assessing their severity are now widespread.

Recent years have shown significant potential for applying artificial intelligence in the medical field (10, 11). Machine learning, a vital tool for data mining, offers greater flexibility and scalability than traditional statistical methods, effectively handling multivariable interactions and collinearity (12). Machine learning uses existing medical testing or patient survey data to establish risk models, enabling disease prediction, diagnosis, and severity assessment (13). Research shows that machine learning algorithms in AI can develop efficient diagnostic and predictive tools, improving tumor diagnosis accuracy by 15 to 20% (14–16). In many prospective studies, machine learning models perform better than medical experts (17, 18). This could be due to the reduced human intervention in AI, minimizing biases and subjective errors in predictions.

This study aims to develop four machine learning models to predict NRDS severity based on clinical indicators and LUS, identifying the most effective models to support NRDS treatment. It will enhance understanding of the relationship between pulmonary pathology, respiratory function, and systemic organ function in neonates with NRDS. This will provide clinicians with a more precise basis for assessing affected neonates and offer a scientific foundation for diagnosing and treating respiratory diseases.

## Methods

### Experimental design

This double-masked cohort study was conducted in the NICU to assess NRDS severity in neonates using LUS and other clinical indicators.

The entire research process was conducted within the NICU to ensure legality and ethical integrity. NICU clinical management adhered to local guidelines, with the study not directly influencing clinical practices. Researchers ensured neonates received appropriate treatment and care while collecting clinical data and LUS scores for subsequent analysis. Researchers used the STROBE checklist during manuscript preparation to ensure the study's reliability and scientific rigor.

### Patients

This study involved 336 pediatric patients admitted to the NICU at Yantaishan Hospital from December 2020 to July 2023. Neonates were selected based on strict inclusion criteria to ensure they met the diagnostic standards for NRDS. From the original cohort, 106 patients were excluded for reasons including an unclear diagnosis, withdrawal from treatment, congenital anomalies, tuberculosis, or congestive heart failure. As a result, 230 neonates remained in the study, with 119 identified as having NRDS and 111 categorized as N-NRDS. The diagnostic procedures were carried out using a Philips CX50 portable ultrasound machine equipped with a linear array probe, operating at a frequency range of 8 to 12 MHz. The study protocol adhered to medical ethical standards and received approval from the Medical Ethics Committee of Yantai Mountain Hospital (approval number: 20220001), with informed consent obtained from the patients' families. All data collection and processing followed relevant guidelines and standards.

### Inclusion criteria and exclusion criteria

The inclusion criteria for this study were grounded in the Berlin criteria, encompassing clinical symptoms, arterial blood gas analysis, and chest X-ray results. A diagnosis of NRDS was confirmed by evaluating clinical signs, arterial blood gas measurements, and radiographic evidence. Eligible neonates were those born within 6 to 12 h before disease onset, showing severe respiratory distress, cyanosis, expiratory grunting, inspiratory retractions, and related symptoms. Arterial blood gas analysis required evidence of hypoxemia, defined as an arterial oxygen partial pressure below 60 mmHg, or hypercapnia, indicated by a carbon dioxide partial pressure exceeding 50 mmHg. Chest X-rays showed decreased lung transparency, increased lung markings, evenly distributed granulation and reticular opacities, blurred cardiac and diaphragmatic outlines, bronchial aeration, and signs of severe pneumonia in both lung fields. Inclusion required that the neonate have a complete clinical medical record.

The exclusion criteria for this study include neonates with congenital disorders such as complex respiratory malformations, congenital heart disease, congenital mental disorders, or chromosomal abnormalities. Neonates showing signs of cardiogenic pulmonary edema or persistent pulmonary hypertension. Neonates with severe intracranial hemorrhage, sepsis, severe hypoxic-ischemic encephalopathy, severe sepsis, septic shock, or other serious illnesses and multi-system complications. Neonates diagnosed with pulmonary hemorrhage, neonatal wet lung, pneumothorax, meconium aspiration syndrome, neonatal infectious pneumonia, or other respiratory disorders. Additionally, neonates were excluded if factors led to increased measurement errors or significantly compromised ultrasound image quality.

## Observation and evaluation indicator

Lung ultrasound score has diverse applications in clinical assessment (19). This study employed a 12-region scoring system, dividing the pediatric chest into 12 zones, with each lung region systematically scanned and scored based on specific ultrasound images, leading to the calculation of the LUS score. The scoring criteria outlined in Table 1 provide a standardized approach for the quantitative evaluation of pulmonary diseases and their severity. Continuous, smooth A-lines or the presence of fewer than three isolated B-lines are assigned a score of 0, indicating normal lung status. A score of 1 is given when three or more scattered, distinct B-lines reflect a mild abnormality. A score of 2 is assigned for diffusely confluent B-lines appearing in a waterfall pattern, indicating moderate abnormality. Lung consolidation, with a score of 3, indicates severe abnormality. The total LUS score ranges from 0 to 36, as depicted in Figure 1, which presents representative images from our institution's assessment of NRDS scores between 0 and 3 (20).

The OI is a critical measure for assessing the severity of NRDS, calculated as the ratio of arterial oxygen partial pressure ( $\text{PaO}_2$ ) to the fraction of inspired oxygen ( $\text{FiO}_2$ ) (21). The formula for its calculation is  $\text{OI} = \text{PaO}_2 / \text{FiO}_2$ . Typically, the normal range for the oxygenation index is between 400 and 500 mm Hg; lower scores suggest worsening pulmonary function and oxygenation capacity in the neonate, thereby indicating the severity of the condition. RI is crucial for assessing children's respiratory function. It is calculated as the ratio of the alveolar-arterial oxygen pressure difference ( $\text{A-aDO}_2$ ) to  $\text{PaO}_2$ . The formula for its calculation is  $\text{RI} = \text{A-aDO}_2 / \text{PaO}_2$ . RI between 0.8 and 1.2 is considered normal, indicating good respiratory function, regular lung activity, and sufficient gas exchange. An RI between 1.2 and 1.6 suggests mild abnormalities, potentially indicating mild dyspnea due to respiratory muscle fatigue or upper respiratory tract infection. An RI of 1.6 to 2.0 indicates moderate abnormalities, usually signifying moderate dyspnea and pulmonary function decline, possibly due to conditions like chronic obstructive pulmonary disease. An RI exceeding 2.0, particularly above 3.0, indicates severe abnormalities, often reflecting significant respiratory dysfunction, potentially due to severe pulmonary diseases or respiratory acidosis, leading to substantial dyspnea and critically impaired lung function.

The SOFA score includes six systems: respiratory, coagulation, liver, circulatory, central nervous, and renal (22). Each system is scored from 0 to 4, with the total SOFA score being the sum of these scores. Table 2 shows that when evaluating the SOFA score in children with NRDS, special attention should be given to each system's score, as a higher total score indicates more severe organ dysfunction. According to the criteria, the SOFA score typically ranges from 0 to 24.

## Machine learning model

Figure 2 shows the flowchart of this study. This study addresses a binary classification problem ideally suited for supervised learning in

TABLE 1 LUS evaluation criteria (20).

|   | Score |
|---|-------|
| Lung consolidation                              | 3     |
| Numerous integrated B-lines (waterfall sign)    | 2     |
| Scattered on clear B-lines ( $\geq 3$ )         | 1     |
| Smooth A-lines or fewer than 3 isolated B-lines | 0     |

machine learning. Supervised learning techniques encompass a variety of algorithms such as linear and LR, SVM, Naive Bayes, extreme gradient boosting, decision trees, and RF. In this research, we selected LR, RF, NN, and SVM as the training models commonly used in medical treatment.

## Machine learning model building process

Data sets with statistically significant differences were selected to construct machine learning metrics. Subsequently, 161 cases (70%) were randomly chosen from the entire sample as the training set. The remaining 69 cases (30%) were used as the test set. The training set data underwent five-fold cross-validation to achieve the optimal training model. The advantage of cross-validation is that it uses all data for both training and validation sets, with each subset independently validated, providing a more robust reflection of the training model. Finally, an independent external validation set was selected, consisting of 44 NRDS and 43 N-NRDS patients, to test and evaluate the model's accuracy.

## Statistical analysis

Data entry was performed using SPSS software. Continuous variables that followed a normal or approximately normal distribution were expressed as mean  $\pm$  standard deviation ( $\bar{x} \pm s$ ), with group comparisons made using the independent samples t-test. For continuous variables not following a normal distribution, data were reported as median (interquartile range), and group differences were assessed using non-parametric rank-sum tests. All data analyses and the construction of machine learning models were performed using R software (version 4.0.3). All models were built using the Caret package, with the importance of predictive variables assessed through the vamp function within Caret. In this study, a  $p$ -value of  $<0.05$  was considered statistically significant.

## Results

### Comparison of primary clinical data

Table 3 presents the baseline characteristics of the two groups. In the NRDS cohort, 47.90% of the participants were male, and the median gestational age was 32 (27,36) weeks. The mean birth weight was  $2013 \pm 622$  g, and 60.50% of deliveries were by cesarean section. In the N-NRDS group, the proportion of males was 60.30%, the median gestational age was 35 (30,39) weeks, the mean birth weight was  $2,291 \pm 741$  g, and cesarean deliveries accounted for 54.05% of the total deliveries.

### Comparison of clinical indicators among different cohorts

The LUS and clinical score variations across different patient groups were compared, as shown in Table 4. The NRDS group exhibited higher LUS, SOFA, and RI values than the N-NRDS group, while OI values were lower ( $p < 0.01$ ).

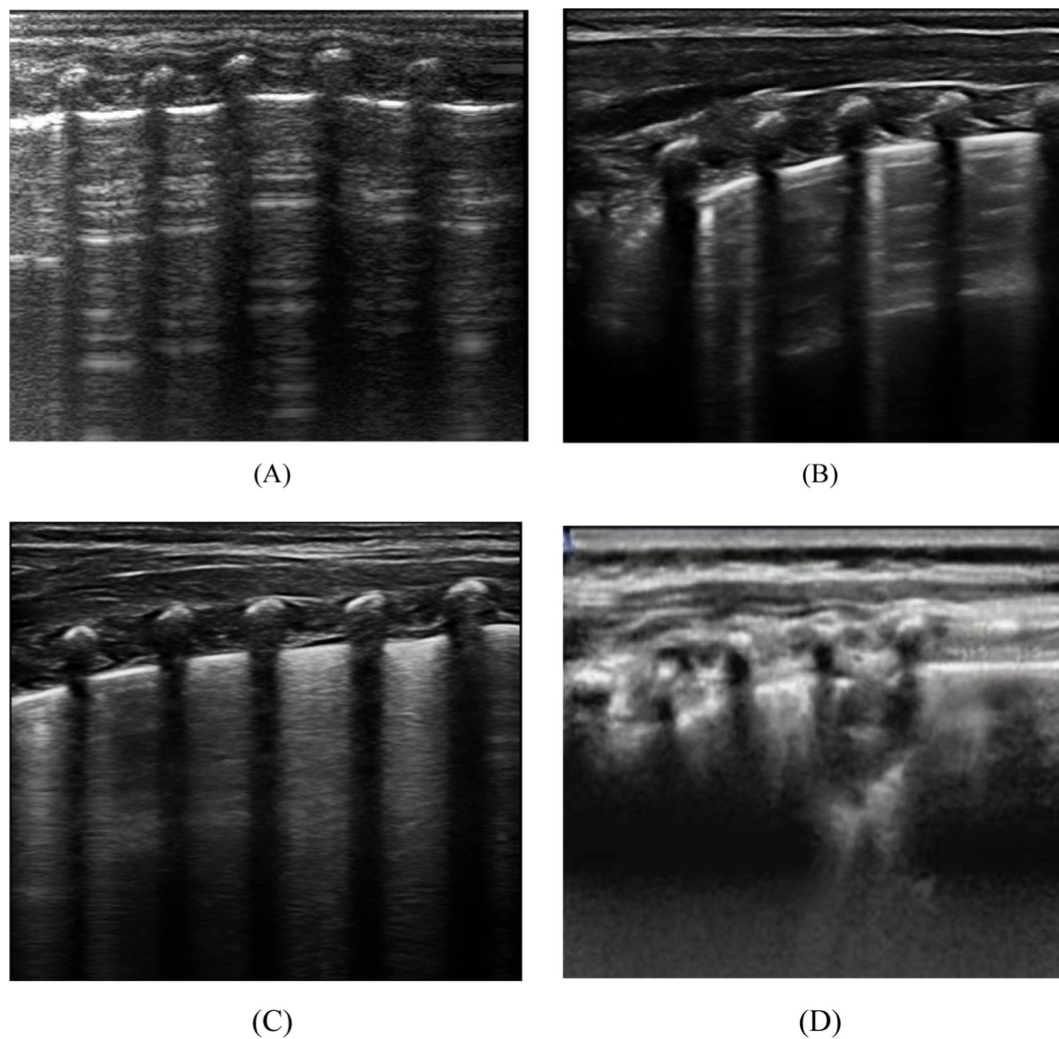


FIGURE 1

Depicts the four distinct lung ultrasound scoring patterns. (A) A score of 0 is assigned when only A lines are visible or fewer than 3 B lines. A vertical arrow indicates A-lines. (B) A score of 1 is given when there are 3 or more well-spaced B-lines, shown with horizontal arrows. (C) Score 2 was line B with many fusions, with or without subpleural consolidation. (D) 3 points are defined as "hepatoid" lung tissue or accompanied by an air bronchogram, represented by an oval.

## Comparison of different severity of NRDS patients

Patients with NRDS were categorized based on the severity of their pulmonary disease into mild, moderate, and severe groups. Table 5 displays LUS levels and clinical scores corresponding to various degrees of NRDS, revealing significant statistical differences among the groups ( $p < 0.01$ ). The severe NRDS group showed significantly elevated LUS, SOFA, and RI scores relative to the mild and moderate groups, whereas the OI was notably lower.

## Model training process

To develop the NRDS predictive model, four commonly used machine learning algorithms were employed: LR, RF, NN, and SVM. Based on this foundation, ensemble and cascading methods

were used to combine the individual classification models, and the performance of the integrated classifier was subsequently evaluated.

## Importance of predictors of different models

We visualized the top eight variables according to their weights. The length of the bars is directly proportional to the importance of each variable. The results indicate that the model's key variables fall into two groups, with the first group—LUS, OI, SOFA, and RI—significantly impacting predictive performance. LUS is the most influential variable in the LR, RF, and NN models. LUS is the most influential variable for the LR, RF, and NN models. The second group consists of cesarean birth, gestational age, weight, and gender, which have a relatively minor impact on prediction; detailed information is provided in Figure 3.

TABLE 2 SOFA score table.

| System                        | Evaluation index                          | Score |
|-------------------------------|---|-------|
| <b>Coagulation system</b>     |   |       |
|                               | >150                                      | 0     |
|                               | 100 ~ 150                                 | 1     |
|                               | 50 ~ 100                                  | 2     |
|                               | 20 ~ 50                                   | 3     |
|                               | <20                                       | 4     |
| <b>Central nervous system</b> |   |       |
|                               | <110                                      | 0     |
|                               | 110 ~ 170                                 | 1     |
|                               | 171 ~ 299                                 | 2     |
|                               | 300 ~ 440                                 | 3     |
|                               | >440                                      | 4     |
| <b>Liver</b>                  |   |       |
|                               | <20                                       | 0     |
|                               | 20 ~ 32                                   | 1     |
|                               | 33 ~ 101                                  | 2     |
|                               | 102 ~ 204                                 | 3     |
|                               | >204                                      | 4     |
| <b>Respiratory system</b>     |   |       |
|                               | >400 mmHg                                 | 0     |
|                               | 300 ~ 400 mmHg                            | 1     |
|                               | 200 ~ 300 mmHg                            | 2     |
|                               | 100 ~ 200 mmHg                            | 3     |
| <b>Kidney function</b>        |   |       |
|                               | <110                                      | 0     |
|                               | 110 ~ 170                                 | 1     |
|                               | 171 ~ 299                                 | 2     |
|                               | 300 ~ 440                                 | 3     |
|                               | >440                                      | 4     |
| <b>Cardiovascular system</b>  |   |       |
|                               | Map $\geq$ 70 mmHg                        | 0     |
|                               | Map < 70 mmHg                             | 1     |
|                               | Dopamine $\leq$ 5 or dobutamine dose      | 2     |
|                               | Dopamine > 5 or norepinephrine $\leq$ 0.1 | 3     |
|                               | Dopamine > 15 or norepinephrine > 0.1     | 4     |

### Evaluation results of different machine learning algorithms

This research utilizes five key performance metrics, the AUC, accuracy, sensitivity, specificity, and the F1 score, to assess the effectiveness of four machine learning algorithms, as detailed in Table 6. ROC curves were plotted using models constructed from the training dataset (Figure 4), with test set AUC values as follows: LR (AUC=0.841, 95% CI: 0.712–0.881), RF (AUC=0.894, 95% CI: 0.811–0.976), NN (AUC=0.828, 95% CI: 0.701–0.935), and SVM (AUC=0.726, 95% CI: 0.623–0.887). The RF model demonstrated the highest performance of

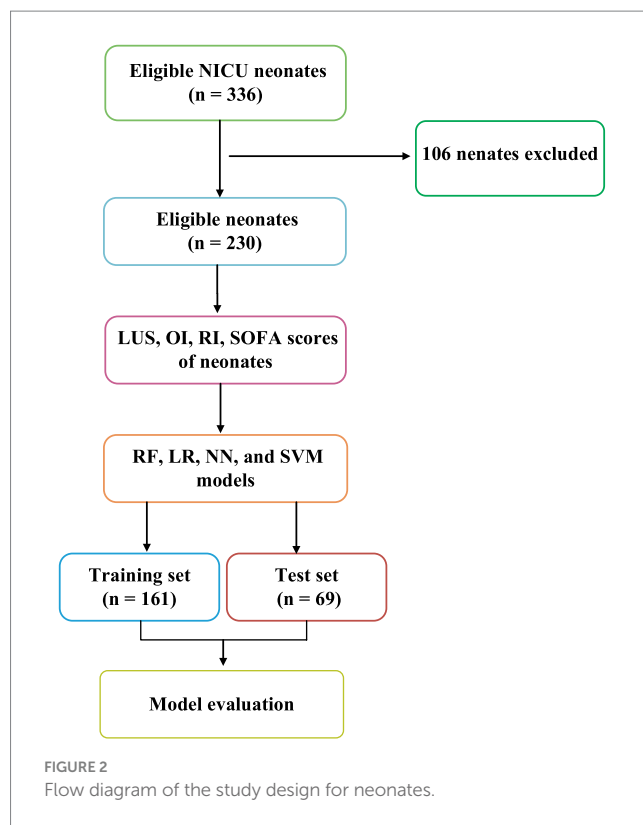


TABLE 3 Characteristic parameters of patients.

|  | NRDS group<br><i>n</i> = 119 | N-NRDS group<br><i>n</i> = 111 | <i>P</i> -value |
|--|------------------------------|--------------------------------|-----------------|
| <b>Sex (%)</b>                           |                              |                                |                 |
| Male                                     | 57 (47.90%)                  | 67 (60.30%)                    | 0.131           |
| Cesarean section (%)                     | 72 (60.50%)                  | 60 (54.05%)                    | <0.01           |
| Weight (g)                               | 2013 ± 622                   | 2,291 ± 741                    | <0.01           |
| Gestational age (range)                  | 32 (27,36)                   | 35 (30,39)                     | <0.01           |
| <Number of children at 28 weeks          | 76 (63.87%)                  | 73 (65.77%)                    | 0.107           |
| <The Number of children with 1,500 grams | 51 (42.86%)                  | 40 (36.04%)                    | 0.013           |
| <b>Apgar score</b>                       |                              |                                |                 |
| 1 min ( $\leq$ 7)                        | 74 (62.18%)                  | 55 (49.55)                     | 0.610           |
| 5 min ( $\leq$ 7)                        | 48 (40.33%)                  | 26 (18.92%)                    | 0.861           |

Values are expressed as mean ± standard deviation, median (25th percentile; 75th percentile), or Number (%); the difference between groups was statistically significant (*p* < 0.05).

the four models, with an AUC of 0.894, accuracy of 0.808, and sensitivity of 0.706. In contrast, the AUC values for the LR, NN, and SVM models were notably lower. Given the superior performance of the RF model, it was selected as the final model for this study.

### Verification of the model

This study used calibration curves to evaluate the accuracy of the model. Figure 5 depicts the calibration curves for the training, internal validation, and external validation sets. The results show that the

TABLE 4 LUS and clinical scores of children in different cohorts.

|      | Total<br>n = 330 | NRDS<br>group<br>n = 119 | N-NRDS<br>group<br>n = 111 | t/Z   | P     |
|------|------------------|--------------------------|----------------------------|-------|-------|
| LUS  | 21.33±7.12       | 21.88±6.87               | 20.51±7.71                 | -0.33 | <0.01 |
| OI   | 181.7±75.7       | 179.2±75.3               | 183.5±76.3                 | 1.91  | <0.01 |
| SOFA | 16.68±4.21       | 17.3±4.83                | 16.2±3.81                  | 0.97  | <0.01 |
| RI   | 1.51±0.52        | 1.65±0.56                | 1.29±0.49                  | -0.59 | <0.01 |

NRDS is neonatal respiratory distress syndrome, LUS is lung ultrasound score, OI is oxygenation index, RI is respiratory index, and SOFA is sequential organ failure assessment.

TABLE 5 Comparison of LUS and clinical index scores of NRDS patients with different severity.

|          |    | LUS                     | OI                      | SOFA                    | RI                     |
|----------|----|-------------------------|-------------------------|-------------------------|------------------------|
| Mild     | 38 | 14.68±6.73              | 233.1±83.6              | 10.79±4.01              | 1.19±0.79              |
| Moderate | 48 | 20.01±5.47 <sup>a</sup> | 161.1±69.1 <sup>a</sup> | 15.01±4.83 <sup>a</sup> | 1.55±0.56 <sup>a</sup> |
| Serious  | 33 | 27.12±7.1 <sup>b</sup>  | 111.7±57.8 <sup>b</sup> | 19.36±5.11 <sup>b</sup> | 2.21±1.12 <sup>b</sup> |
| F        |    | 11.22                   | 17.32                   | 9.76                    | 4.53                   |
| P        |    | <0.01                   | <0.01                   | <0.01                   | <0.01                  |

Compared with the mild group, <sup>a</sup>P<0.01. Compared with the moderate group, <sup>b</sup>P<0.01.

nomogram’s predictive performance across all sets is closely aligned with ideal performance.

## Discussion

In recent years, NRDS has garnered widespread attention. NRDS manifests as progressive respiratory distress and respiratory failure shortly after birth (23). Previous studies, both domestic and international, have shown that preterm neonates are at significantly higher risk of developing NRDS, posing a severe threat to neonatal health (24). For example, preterm neonates born at 28–29 weeks have an 81% probability of developing NRDS (25). Similar studies in Europe have yielded consistent results, showing an 80% incidence of NRDS among preterm neonates born before 28 weeks. Without timely intervention, NRDS can lead to bronchopulmonary dysplasia and even pose a life-threatening risk to the neonatal (26, 27). Therefore, early identification and timely intervention in neonates with NRDS are crucial for improving survival rates and outcomes.

Among clinical indicators, OI, SOFA, and RI are crucial for assessing NRDS in patients (28). OI assesses neonates’ respiratory function and

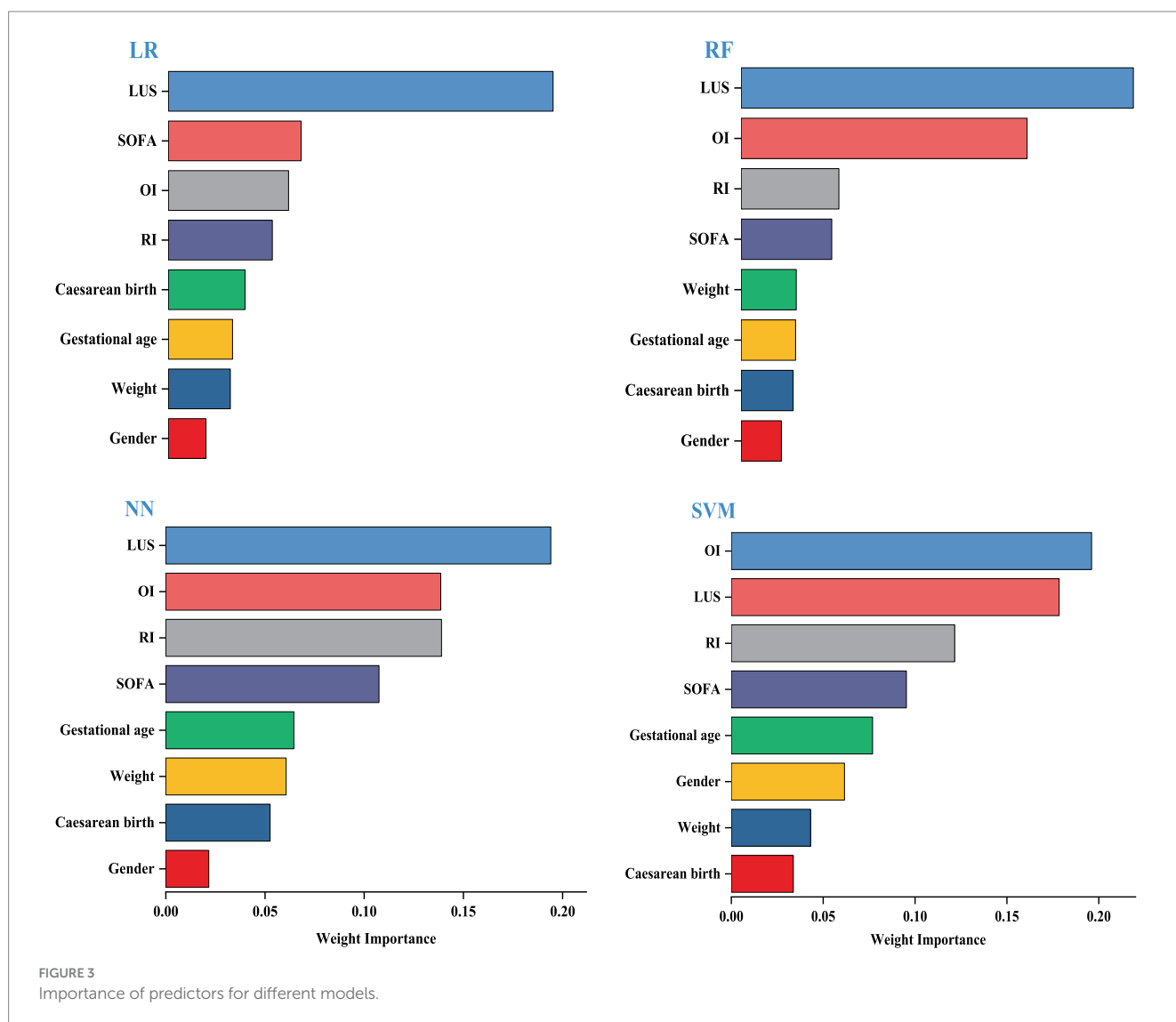
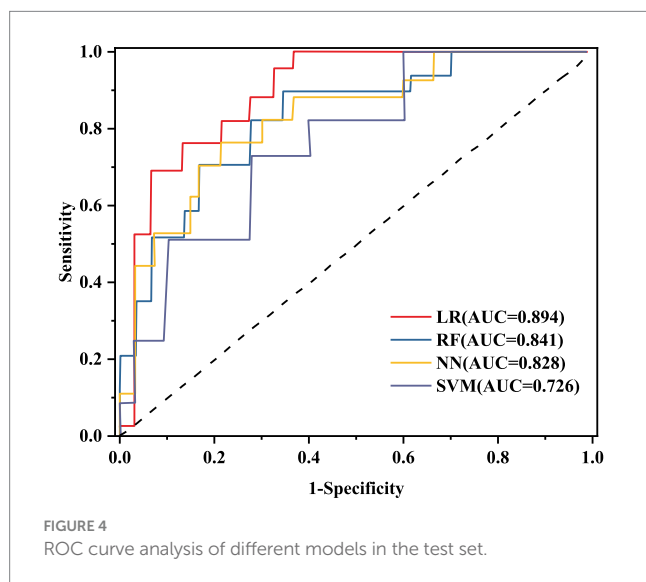


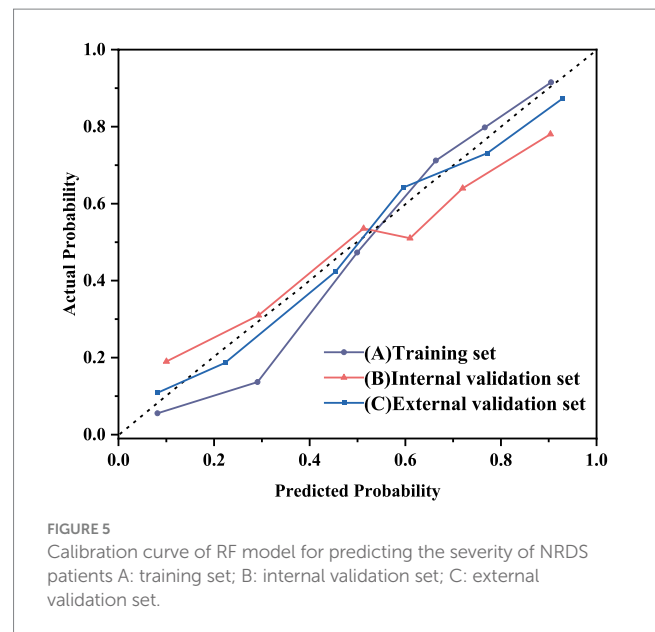
TABLE 6 The performance of different machine learning algorithm models.

|            | AUC   | Accuracy | Sensitivity | Specificity | F1    |
|------------|-------|----------|-------------|-------------|-------|
| <b>LR</b>  |       |          |             |             |       |
| Train      | 0.870 | 0.818    | 0.691       | 0.861       | 0.727 |
| Test       | 0.841 | 0.771    | 0.613       | 0.877       | 0.661 |
| <b>RF</b>  |       |          |             |             |       |
| Train      | 0.872 | 0.801    | 0.679       | 0.871       | 0.704 |
| Test       | 0.894 | 0.808    | 0.706       | 0.867       | 0.727 |
| <b>NN</b>  |       |          |             |             |       |
| Train      | 0.847 | 0.778    | 0.632       | 0.733       | 0.681 |
| Test       | 0.828 | 0.786    | 0.593       | 0.851       | 0.653 |
| <b>SVM</b> |       |          |             |             |       |
| Train      | 0.670 | 0.731    | 0.754       | 0.776       | 0.693 |
| Test       | 0.726 | 0.766    | 0.704       | 0.801       | 0.686 |



oxygenation levels, making it a key parameter for determining NRDS severity and predicting outcomes (29). Monitoring OI during NRDS treatment helps assess the neonate's oxygenation status promptly, enabling adjustments in treatment to enhance respiratory support and improve oxygenation. SOFA is a valuable clinical tool for evaluating the severity of multiple organ dysfunction in neonates with NRDS (30). SOFA scores various organs to assess disease severity and prognosis, offering crucial guidance for clinical decision-making. RI is another essential indicator for NRDS, calculated using gas exchange principles to reflect the lung's ability to diffuse oxygen and exchange gasses. In healthy individuals, RI values generally range from 0.8 to 1.2. An RI value exceeding this range may signal problems like inadequate oxygenation or carbon dioxide retention, which could lead to respiratory failure.

Data analysis demonstrated that the NRDS group had higher LUS, RI, and SOFA scores than the N-NRDS group, while OI values were lower in the NRDS group. This implies that more severe cases of NRDS at admission are associated with reduced OI values and worse prognoses. The results suggest that integrating LUS, OI, RI, and SOFA scores could effectively predict the severity of NRDS, potentially



leading to improved treatment strategies and better clinical outcomes. The NRDS neonates were categorized into mild, moderate, and severe groups based on pulmonary disease severity. Results demonstrated that neonates with severe NRDS had significantly higher LUS, SOFA, and RI scores and the lowest OI values compared to those with mild and moderate NRDS, with scores of  $27.12 \pm 7.1$ ,  $111.7 \pm 57.8$ ,  $19.36 \pm 5.11$ , and  $2.21 \pm 1.12$ , respectively ( $p < 0.01$ ).

The increasing incidence of neonatal pulmonary infections has made managing high-dimensional data with traditional statistical models increasingly tricky (29). In contrast, machine learning leverages large datasets to build highly efficient and precise mathematical models, reshaping traditional clinical practices and offering potentially optimal treatment solutions for patients. The integration of machine learning into clinical practice offers significant potential. For example, Khamzin et al. (31) investigated the use of machine learning for the intelligent analysis of CT and MRI images in medical imaging. Likewise, McCoubrey et al. (32) utilized machine learning to examine gut microbiome characteristics for evaluating drug metabolism. Currently, machine learning is also being used for sensitivity analysis in pneumonia. For example, Lu et al. (33) and Liu et al. (34) integrated machine learning with Raman spectroscopy and whole-genome sequencing to evaluate resistance levels in *Klebsiella pneumoniae*. Nevertheless, Raman spectroscopy and whole-genome sequencing still need to be standard practices in microbial identification.

The use of artificial intelligence methods, such as machine learning, in clinical practice remains in its early exploratory phase. Combining clinical expertise with ultrasound data shows promising potential as machine learning algorithms become more sophisticated and reliable. Various machine learning algorithms differ in characteristics such as fitting performance, algorithmic complexity, and the capacity to manage multi-feature data. This study aims to develop a low-cost, high-precision diagnostic system for NRDS to aid clinicians in making timely and accurate decisions. To achieve this, the study constructed four machine-learning models to predict the severity of NRDS, including LR, RF, NN, and SVM. Among these models, the RF model performed the best in the training set, achieving an AUC value of 0.894. The RF model outperformed logistic regression, neural networks, and

support vector machines in terms of speed and accuracy with large datasets. Conversely, the minimum prediction accuracy of SVM is 0.766, compared to 0.771 for LR, 0.808 for RF, and 0.786 for NN. This performance discrepancy may be attributed to the small sample size, the limited number of variables, and the relatively homogeneous study population. Furthermore, the RF model offers several advantages, including fewer statistical assumptions, robust noise tolerance, reduced risk of overfitting, and minimal need for parameter tuning.

## Conclusion

This study created four machine-learning models by combining clinical indicators with imaging features. The results show that the RF model exhibited the highest performance and is particularly effective in predicting NRDS severity. Furthermore, analyzing variable weights across the four algorithms revealed key predictors, including LUS, OI, SOFA, and RI. Using machine learning models can make more accurate clinical decisions for NRDS patients.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

This retrospective observational study received approval from the Clinical Ethics Committee of Yantai Shan Hospital in Shandong Province, China (Ethical Review No. 20220001). Informed consent was obtained from the guardians of all participating children. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

## References

- Ma H, Yan W, Liu J. Diagnostic value of lung ultrasound for neonatal respiratory distress syndrome: a meta-analysis and systematic review. *Med Ultrason.* (2020) 22:325–33. doi: 10.11152/mu-2485
- Vetrugno L, Sala A, Deana C, Meroi F, Grandesso M, Maggiore SM, et al. Quality of life 1 year after hospital discharge in unvaccinated pregnant women with COVID-19 respiratory symptoms: a prospective observational study (ODISSEA-PINK study). *Front Med.* (2023) 10:1225648. doi: 10.3389/fmed.2023.1225648
- Vetrugno L, Castaldo N, Fantin A, Deana C, Cortegiani A, Longhini F, et al. Ventilatory associated barotrauma in COVID-19 patients: a multicenter observational case control study (COVI-MIX-study). *Pulmonology* (2023) 29:457–468. doi: 10.1016/j.pulmoe.2022.11.002
- De Luca D, Autilio C, Pezza L, Shankar-Aguilera S, Tingay DG, Carnielli VP. Personalized medicine for the management of RDS in preterm neonates. *Neonatology.* (2021) 118:127–38. doi: 10.1159/000513783
- Aşar S, Rahim F, Rahimi P, Acicbe Ö, Tontu F, Çukurova Z. Novel oxygenation and saturation indices for mortality prediction in COVID-19 ARDS patients: the impact of driving pressure and mechanical power. *J Intensive Care Med.* (2024) 39:595–608. doi: 10.1177/08850666231223498
- Shi S, Guo J, Fu M, Liao L, Tu J, Xiong J, et al. Evaluation of the neonatal sequential organ failure assessment and mortality risk in neonates with respiratory distress

## Author contributions

CH: Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft. XH: Conceptualization, Data curation, Project administration, Resources, Writing – review & editing. YC: Funding acquisition, Software, Supervision, Visualization, Writing – original draft. HZ: Data curation, Funding acquisition, Investigation, Resources, Validation, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by Yantai Science and Technology Bureau (grant number: 2019YD031).

## Acknowledgments

We express our gratitude to all the neonatal and their families who took part in this study. We also extend our appreciation to the reviewers for their insightful comments and suggestions.

## Conflict of interest

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

## Publisher's note

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

syndrome: a retrospective cohort study. *Front Pediatr.* (2022) 10:911444. doi: 10.3389/fped.2022.911444

7. Biasucci DG, Loi B, Centorrino R, Raschetti R, Piastra M, Pisapia L, et al. Ultrasound-assessed lung aeration correlates with respiratory system compliance in adults and neonates with acute hypoxemic restrictive respiratory failure: an observational prospective study. *Respir Res.* (2022) 23:360. doi: 10.1186/s12931-022-02294-1

8. Zhang J, Liu H, Zhang Y, Zhu W, Liu Y, Han T. Prospective, non-blinded, randomized controlled trial on early administration of pulmonary surfactant guided by lung ultrasound scores in very preterm infants: study protocol. *Front Pediatr.* (2024) 12:1411068. doi: 10.3389/fped.2024.1411068

9. De Luca D, Bonadies L, Alonso-Ojembarrena A, Martino D, Gutierrez-Rosa I, Loi B, et al. Quantitative lung ultrasonography to guide surfactant therapy in neonates born late preterm and later. *JAMA Netw Open.* (2024) 7:e2413446. doi: 10.1001/jamanetworkopen.2024.13446

10. Lee D, Yoon SN. Application of artificial intelligence-based Technologies in the Healthcare Industry: opportunities and challenges. *Int J Environ Res Public Health.* (2021) 18:271. doi: 10.3390/ijerph18010271

11. Yoon M, Park JJ, Hur T, Hua C-H, Hussain M, Lee S, et al. Application and potential of artificial intelligence in heart failure: past, present, and future. *Int J Heart Fail.* (2023) 6:11–9. doi: 10.36628/ijhf.2023.0050



12. Sarker M. Revolutionizing healthcare: the role of machine learning in the health sector. *J Artif Intell General Sci.* (2024) 2:36–61. doi: 10.60087/jaigs.v2i1.96
13. Zeng Q, Sun W, Xu J, Wan W, Pan L. Machine learning-based medical imaging detection and diagnostic assistance. *Int J Comput Sci Informat Technol.* (2024) 2:36–44. doi: 10.62051/ijcsit.v2n1.05
14. Agarwal S, Yadav AS, Dinesh V, Vatsav KSS, Prakash KSS, Jaiswal S. By artificial intelligence algorithms and machine learning models to diagnosis cancer. *Mater Today Proc.* (2023) 80:2969–75. doi: 10.1016/j.matpr.2021.07.088
15. Iqbal MJ, Javed Z, Sadia H, Qureshi IA, Irshad A, Ahmed R, et al. Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. *Cancer Cell Int.* (2021) 21:270. doi: 10.1186/s12935-021-01981-1
16. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol.* (2019) 16:391–403. doi: 10.1038/s41585-019-0193-3
17. Zachariah FJ, Rossi LA, Roberts LM, Bosserman LD. Prospective comparison of medical oncologists and a machine learning model to predict 3-month mortality in patients with metastatic solid tumors. *JAMA Netw Open.* (2022) 5:e2214514. doi: 10.1001/jamanetworkopen.2022.14514
18. Vetrugno L, Biasucci DG, Deana C, Spadaro S, Lombardi FA, Longhini F, et al. Lung ultrasound and supine chest X-ray use in modern adult intensive care: mapping 30 years of advancement (1993–2023). *Ultrasound J.* (2024) 16:7. doi: 10.1186/s13089-023-00351-4
19. Demi L, Wolfram F, Klersy C, De Silvestri A, Ferretti VV, Muller M, et al. New international guidelines and consensus on the use of lung ultrasound. *J Ultrasound Med.* (2023) 42:309–44. doi: 10.1002/jum.16088
20. Huang C, Zhang S, Ha X, Cui Y, Zhang H. The value of lung ultrasound score in neonatal respiratory distress syndrome: a prospective diagnostic cohort study. *Front Med.* (2024) 11:1357944. doi: 10.3389/fmed.2024.1357944
21. Shen H, Qu D, Na W, Liu S, Huang S, Hui Y. Comparison of the OI and PaO<sub>2</sub>/FiO<sub>2</sub> score in evaluating PARDS requiring mechanical ventilation. *Pediatr Pulmonol.* (2021) 56:1182–8. doi: 10.1002/ppul.25194
22. Moreno R, Vincent J-L, Matos R, Mendonça A, Cantraine F, Thijs L, et al. The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicentre study. *Intensive Care Med.* (1999) 25:686–96. doi: 10.1007/s001340050931
23. Weng M, Wang J, Yin J, He L, Yang H, He H. Maternal prenatal systemic inflammation indexes predicts premature neonatal respiratory distress syndrome. *Sci Rep.* (2024) 14:18129. doi: 10.1038/s41598-024-68956-w
24. Bi Y, Yu W, Bian W, Jin M, He Y, Wang J, et al. Metabolic and microbial dysregulation in preterm infants with neonatal respiratory distress syndrome: an early developmental perspective. *J Proteome Res.* (2024) 23:3460–8. doi: 10.1021/acs.jproteome.4c00114
25. Li H-X, Gao C-J, Cheng S, Mao Z-L, Wang H-Y. Risk factors for respiratory assistance in premature infants. *Exp Ther Med.* (2021) 21:237–1. doi: 10.3892/etm.2021.9668
26. Shen L, Cai N, Wan S, Chen S. Development and validation of a predictive model for early diagnosis of neonatal acute respiratory distress syndrome based on the Montreux definition. *Front Pediatr.* (2023) 11:1276915. doi: 10.3389/fped.2023.1276915
27. Boussuges A, Fourdrain A, Leone M, Brioude G, Menard A, Zieleskiewicz L, et al. Diagnosis of hemidiaphragm paralysis: refine ultrasound criteria. *Front Med.* (2024) 11:1416520. doi: 10.3389/fmed.2024.1416520
28. Villar J, González-Martín JM, Ambrós A, Mosteiro F, Martínez D, Fernández L, et al. Stratification for identification of prognostic categories in the acute RESpiratory distress syndrome (SPIRES) score. *Crit Care Med.* (2021) 49:e920–e930. doi: 10.1097/CCM.0000000000005142
29. Huber W, Findeisen M, Lahmer T, Herner A, Rasch S, Mayr U, et al. Prediction of outcome in patients with ARDS: a prospective cohort study comparing ARDS-definitions and other ARDS-associated parameters, ratios and scores at intubation and over time. *PLoS One.* (2020) 15:e0232720. doi: 10.1371/journal.pone.0232720
30. Saleh A, Ahmed M, Sultan I, Abdel-lateif A. Comparison of the mortality prediction of different ICU scoring systems (APACHE II and III, SAPS II, and SOFA) in a single-center ICU subpopulation with acute respiratory distress syndrome. *Egypt J Chest Dis Tuberc.* (2015) 64:843–8. doi: 10.1016/j.ejcdt.2015.05.012
31. Khamzin S, Dokuchaev A, Bazhutina A, Chumarnaya T, Zubarev S, Lyubimtseva T, et al. Machine learning prediction of cardiac resynchronisation therapy response from combination of clinical and model-driven data. *Front Physiol.* (2021) 12:753282. doi: 10.3389/fphys.2021.753282
32. McCoubrey LE, Thomaidou S, Elbadawi M, Gaisford S, Orlu M, Basit AW. Machine learning predicts drug metabolism and bioaccumulation by intestinal microbiota. *Pharmaceutics.* (2021) 13:2001. doi: 10.3390/pharmaceutics13122001
33. Lu J, Chen J, Liu C, Zeng Y, Sun Q, Li J, et al. Identification of antibiotic resistance and virulence-encoding factors in *Klebsiella pneumoniae* by Raman spectroscopy and deep learning. *Microb Biotechnol.* (2022) 15:1270–80. doi: 10.1111/1751-7915.13960
34. Liu W, Ying N, Mo Q, Li S, Shao M, Sun L, et al. Machine learning for identifying resistance features of *Klebsiella pneumoniae* using whole-genome sequence single nucleotide polymorphisms. *J Med Microbiol.* (2021) 70:001474. doi: 10.1099/jmm.0.001474