



## OPEN ACCESS

## EDITED BY

Jia-Yuh Chen,  
Chung Shan Medical University, Taiwan

## REVIEWED BY

Gabriela Corina Zaharie,  
University of Medicine and Pharmacy Iuliu  
Hatieganu, Romania  
Yuh-Jyh Lin,  
National Cheng Kung University, Taiwan

## \*CORRESPONDENCE

Maryam Vafapour  
✉ m.vafapour.ped@gmail.com  
Reza Bahrami  
✉ r.bahrami.neo@gmail.com

RECEIVED 02 November 2024

ACCEPTED 04 February 2025

PUBLISHED 25 April 2025

## CITATION

Talebi H, Dastgheib SA, Vafapour M, Bahrami R, Golshan-Tafti M, Danaei M, Azizi S, Shahbazi A, Pourkazemi M, Yeganegi M, Shiri A, Masoudi A, Rashnavadi H and Neamatzadeh H (2025) Advancements in biomarkers and machine learning for predicting of bronchopulmonary dysplasia and neonatal respiratory distress syndrome in preterm infants. *Front. Pediatr.* 13:1521668. doi: 10.3389/fped.2025.1521668

## COPYRIGHT

© 2025 Talebi, Dastgheib, Vafapour, Bahrami, Golshan-Tafti, Danaei, Azizi, Shahbazi, Pourkazemi, Yeganegi, Shiri, Masoudi, Rashnavadi and Neamatzadeh. 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.

# Advancements in biomarkers and machine learning for predicting of bronchopulmonary dysplasia and neonatal respiratory distress syndrome in preterm infants

Hanieh Talebi<sup>1</sup>, Seyed Alireza Dastgheib<sup>2</sup>, Maryam Vafapour<sup>3\*</sup>, Reza Bahrami<sup>4\*</sup>, Mohammad Golshan-Tafti<sup>5</sup>, Mahsa Danaei<sup>6</sup>, Sepideh Azizi<sup>7</sup>, Amirhossein Shahbazi<sup>8</sup>, Melina Pourkazemi<sup>9</sup>, Maryam Yeganegi<sup>10</sup>, Amirmasoud Shiri<sup>11</sup>, Ali Masoudi<sup>12</sup>, Heewa Rashnavadi<sup>13</sup> and Hossein Neamatzadeh<sup>14</sup>

<sup>1</sup>Clinical Research Development Unit, Fatemeh Hospital, Hamadan University of Medical Sciences, Hamadan, Iran, <sup>2</sup>Department of Medical Genetics, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran, <sup>3</sup>Firoozabadi Clinical Research Development Unit, Department of Pediatrics, Iran University of Medical Sciences, Tehran, Iran, <sup>4</sup>Neonatal Research Center, Shiraz University of Medical Sciences, Shiraz, Iran, <sup>5</sup>Department of Pediatrics, Islamic Azad University of Yazd, Yazd, Iran, <sup>6</sup>Department of Obstetrics and Gynecology, School of Medicine, Iran University of Medical Sciences, Tehran, Iran, <sup>7</sup>Shahid Akbarabadi Clinical Research Development Unit, Iran University of Medical Sciences, Tehran, Iran, <sup>8</sup>School of Medicine, Ilam University of Medical Sciences, Ilam, Iran, <sup>9</sup>School of Medicine, Iran University of Medical Sciences, Tehran, Iran, <sup>10</sup>Department of Obstetrics and Gynecology, School of Medicine, Iranshahr University of Medical Sciences, Iranshahr, Iran, <sup>11</sup>School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran, <sup>12</sup>School of Medicine, Shahid Sadoughi University of Medical Sciences, Yazd, Iran, <sup>13</sup>School of Medicine, Tehran University of Medical Sciences, Tehran, Iran, <sup>14</sup>Mother and Newborn Health Research Center, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

Recent advancements in biomarker identification and machine learning have significantly enhanced the prediction and diagnosis of Bronchopulmonary Dysplasia (BPD) and neonatal respiratory distress syndrome (nRDS) in preterm infants. Key predictors of BPD severity include elevated cytokines like Interleukin-6 (IL-6) and Tumor Necrosis Factor-alpha (TNF- $\alpha$ ), as well as inflammatory markers such as the Neutrophil-to-Lymphocyte Ratio (NLR) and soluble gp130. Research into endoplasmic reticulum stress-related genes, differentially expressed genes, and ferroptosis-related genes provides valuable insights into BPD's pathophysiology. Machine learning models like XGBoost and Random Forest have identified important biomarkers, including CYR1, GALNT14, and OLAH, improving diagnostic accuracy. Additionally, a five-gene transcriptomic signature shows promise for early identification of at-risk neonates, underscoring the significance of immune response factors in BPD. For nRDS, biomarkers such as the lecithin/sphingomyelin (L/S) ratio and oxidative stress indicators have been effectively used in innovative diagnostic methods, including attenuated total reflectance Fourier transform infrared spectroscopy (ATR-FTIR) and high-content screening for ABCA3 modulation. Machine learning algorithms like Partial Least Squares Regression (PLSR) and C5.0 have shown potential in accurately identifying critical health indicators. Furthermore, advanced feature extraction methods for analyzing neonatal cry signals offer a non-invasive means to differentiate between conditions like

sepsis and nRDS. Overall, these findings emphasize the importance of combining biomarker analysis with advanced computational techniques to improve clinical decision-making and intervention strategies for managing BPD and nRDS in vulnerable preterm infants.

#### KEYWORDS

biomarkers, bronchopulmonary dysplasia, neonatal respiratory distress syndrome, machine learning, predictive models, preterm infants

## 1 Introduction

Bronchopulmonary dysplasia (BPD) and neonatal respiratory distress syndrome (nRDS) are major complications in preterm infants, often resulting in long-term health issues (1, 2). Identifying reliable biomarkers for these conditions is crucial for improving clinical outcomes. Studies have highlighted potential biomarkers and mechanisms involved in their pathogenesis. Endothelin-1 (ET-1) is a promising biomarker for predicting BPD, as it is associated with bronchoconstriction and pulmonary hypertension, with elevated levels indicating early risk in preterm infants with nRDS (3–5). Interleukin-6 (IL-6) is another significant biomarker, with higher serum levels correlating with BPD development, underscoring the role of inflammation in lung injury (6). Oxidative stress is also critical, as preterm infants have immature antioxidant defenses, leading to increased lung tissue damage (7, 8). Clara cell protein expression has been investigated as a potential biomarker for lung injury and BPD risk (9). Furthermore, microbial colonization, particularly by *Pneumocystis jirovecii*, may exacerbate respiratory distress in preterm infants through mechanisms like surfactant disruption and mucus overproduction (10, 11). This underscores the need for biomarkers that reflect both inflammatory responses and microbial presence to effectively assess BPD risk.

Recent advancements in predicting BPD and nRDS in preterm infants utilize biomarkers and machine learning techniques to enhance early diagnosis and clinical outcomes (12, 13). Machine learning algorithms have significantly improved predictive capabilities for these conditions. Studies employing weighted gene co-expression network analysis (WGCNA) and machine learning have identified novel serum biomarkers linked to BPD, such as TMCC2, GYPA, and BPGM (14). This approach enhances prediction accuracy and integrates complex datasets, deepening understanding of factors influencing BPD and nRDS (12, 15, 16). Combining clinical data with machine learning models has shown promise in predicting outcomes for neonates with respiratory distress. Analyzing clinical signs and laboratory results through machine learning can inform predictive models, assisting clinicians in managing preterm infants and leading to personalized treatment plans that improve survival rates and reduce long-term complications associated with BPD and nRDS (17–19).

This review aims to systematically explore and synthesize existing literature on advancements in biomarkers and machine learning algorithms that improve the prediction and early identification of BPD and nRDS in preterm infants. By

examining the interplay between biological markers and machine learning techniques, this review seeks to highlight innovative approaches that may enhance clinical outcomes, inform diagnostic strategies, and facilitate timely management of these prevalent neonatal conditions. Additionally, the review will identify research gaps, outline future investigation directions, and propose recommendations for integrating these technological advancements into clinical practice.

## 2 Bronchopulmonary dysplasia

### 2.1 Cytokines and inflammatory biomarkers

BPD is a significant complication in preterm infants, marked by chronic lung inflammation and injury, with cytokines and inflammatory biomarkers playing crucial roles in its development (20, 21). IL-6 is notably elevated in infants who develop BPD, with levels above  $46.125 \text{ pg}\cdot\text{ml}^{-1}$  correlating with increased risk, supported by a high area under the curve (AUC) of 0.849 (22). Tumor Necrosis Factor-alpha (TNF- $\alpha$ ) is also commonly elevated, disrupting lung development through inflammatory pathways. Acute inflammation is linked to Interleukin-1 (IL-1) and Interleukin-8 (IL-8), while reduced levels of the anti-inflammatory cytokine Interleukin-10 (IL-10) may predispose newborns to BPD by failing to regulate inflammation effectively (23–25). Inflammatory biomarkers such as the Neutrophil-to-Lymphocyte Ratio (NLR), soluble gp130, and 8-Hydroxy-2'-deoxyguanosine (8-OHdG) are emerging as predictors of BPD severity, with elevated NLR indicating systemic inflammation and high soluble gp130 levels reflecting inflammatory response involvement (26, 27). The inflammatory response in BPD includes immune cell infiltration, especially monocytes and neutrophils, which can further damage tissue. Additionally, the combination of hyperoxia from supplemental oxygen and inflammation exacerbates oxidative stress, contributing significantly to lung injury and BPD pathogenesis. A study by Gao et al. explored the predictive value of umbilical cord blood Interleukin-6 (UCB IL-6) in 414 preterm infants born before 32 weeks. The study found a significant correlation between UCB IL-6 levels and BPD severity, achieving an area under the receiver operating characteristic curve (AUROC) of 0.815, particularly differentiating Grade 2–3 BPD patients. Four machine learning models—XGBoost, CatBoost, LightGBM, and Random Forest—produced micro-average AUROC values of 0.841, 0.870, 0.851, and 0.878, respectively, with UCB IL-6

consistently identified as the most important feature across all models. This indicates that UCB IL-6 is a valuable biomarker for assessing the risk of severe BPD in preterm infants, aiding clinical decision-making (7).

## 2.2 Gene expression and stress-related biomarkers

### 2.2.1 Endoplasmic reticulum stress-related genes (ERSGs)

Recent studies highlight the involvement of endoplasmic reticulum (ER) stress-related genes (ERSGs) in BPD pathogenesis. ER stress, caused by the accumulation of misfolded proteins in the ER, disrupts cellular function, which is particularly relevant for lung development in premature infants (28–30). Research shows a strong correlation between ERSG expression and various immune cells, like neutrophils and macrophages, suggesting that ER stress significantly affects immune responses in BPD patients. ERSGs are crucial for essential processes supporting alveolar development, including mitochondrial function maintenance, oxidative stress regulation, and inflammation control. Infants with BPD exhibit significant dysregulation of ERSGs compared to those without, indicating a broader role for ER stress in the disease's progression and severity (31, 32). Understanding ERSG implications may lead to novel therapeutic strategies to mitigate lung damage in infants at risk for or affected by BPD. Ziyu Tao et al. examined the relationship between ERSGs and BPD using interpretable machine learning on the GSE32472 dataset. They identified two molecular clusters linked to ER stress in BPD patients, with notable immune cell infiltration and altered ERSG expression compared to controls. Among the models tested, the support vector machine (SVM) showed superior discriminative ability, particularly when combined with five specific genes, indicating its potential for clinical risk assessment and therapeutic interventions in BPD (32). These findings enhance our understanding of BPD mechanisms and highlight biomarkers that may assist in its management, underscoring ERSGs' role in the disease's pathophysiology and their potential as predictive markers.

### 2.2.2 Differentially expressed genes (DEGs)

Recent studies have identified differentially expressed genes (DEGs) associated with BPD, improving our understanding of its pathophysiology and potential therapeutic targets. One study found 49 cluster-specific DEGs by integrating genes related to ER stress with those linked to BPD, highlighting the complex relationship between ER stress and immune responses in the disease. Additionally, RNA sequencing of cord blood revealed 1,685 genes involved in pathways that inhibit lymphopoiesis and adaptive immune responses, with 471 DEGs significant at  $p < 0.01$  and a total of 1,685 at  $p < 0.05$ . These DEGs were enriched in crucial biological pathways, particularly T cell receptor signaling and inflammatory responses such as IL-6/JAK/STAT3 signaling. Key genes associated with lung alveolarization

and BPD development included SPARC and AGER (33). Moreover, elevated neutrophil activation and altered T cell homeostasis were observed in BPD infants, indicating a heightened inflammatory response that may contribute to disease progression. The correlation between DEGs and immune cell types, such as neutrophils and macrophages, suggests these genes could influence immune infiltration patterns in the lungs of affected infants (32, 34). The study by Luo et al. explores potential biomarkers in the peripheral blood of neonates with BPD. Utilizing the Gene Expression Omnibus dataset GSE32472, the researchers identified 470 DEGs and conducted WGCNA, revealing 1,351 significant genes. The intersection of DEGs and WGCNA modules yielded 273 common genes, which were subjected to functional enrichment analysis. To identify hub genes, the researchers applied three machine learning algorithms—SVM-RFE, LASSO, and Random Forest—resulting in the identification of CYYR1, GALNT14, and OLAH as potential BPD biomarkers. Additionally, flunisolide, budesonide, and beclomethasone were predicted as therapeutic drugs linked to these biomarkers (14). The findings indicate that CYYR1, GALNT14, and OLAH may serve as diagnostic indicators for BPD, thereby enhancing clinical diagnosis and preventive strategies while highlighting the relevance of gene expression profiles in understanding the condition.

### 2.2.3 Ferroptosis-related genes

Ferroptosis, an iron-dependent form of cell death characterized by lipid peroxidation, has emerged as a significant factor in various respiratory diseases, such as acute respiratory distress syndrome (ARDS), asthma, and acute lung injury (35). Recent findings indicate its involvement in the pathogenesis of BPD, particularly in premature infants. Studies have shown elevated free iron levels in the umbilical cord blood of these infants, with cumulative enteral iron increasing BPD risk. Additionally, abnormal iron accumulation has been detected in the lungs of BPD models, and therapies targeting ferroptosis demonstrate potential in mitigating hyperoxia-induced lung damage, suggesting a critical role for this regulated cell death pathway in respiratory health and disease (36, 37). Recent research by Fang et al. has highlighted the potential of ferroptosis-related genes as diagnostic biomarkers for BPD. Utilizing datasets from the GEO and FerrDb databases, the study identified 23 ferroptosis-related differentially expressed mRNAs (FRDE-mRNAs) primarily involved in biological processes such as autophagy, fatty acid metabolism, and ferroptosis itself. Four hub genes—LPIN1, ACADSB, WIPI1, and SLC7A11—were selected through machine learning algorithms including LASSO, SVM-RFE, and Random Forest to construct a diagnostic nomogram, which demonstrated strong predictive performance as evidenced by satisfactory receiver operating characteristic (ROC) and calibration curves. Furthermore, the study assessed immune cell infiltration in BPD, revealing significant differences in eight immune cell markers between BPD patients and controls. This research not only elucidates the role of ferroptosis in BPD but also identifies key biomarkers that could facilitate timely diagnosis and intervention, providing a foundation for future investigations into the underlying

mechanisms of the disease. Their research led to the development of a diagnostic nomogram, pinpointing the relevance of ferroptosis in BPD (38).

## 2.2.4 Cuproptosis-related genes (CRGs)

Cuproptosis is a recently discovered mechanism of cell death distinct from traditional apoptosis, triggered by the accumulation of copper ions within cells, resulting in toxic effects. This mechanism interacts uniquely with lipid-acylated TCA cycle proteins, and the disruption of these proteins due to elevated copper levels sets off a chain of molecular events leading to cell death (39, 40). Research into Cuproptosis has opened new avenues for understanding cell death mechanisms and developing therapeutic strategies to regulate copper levels in various diseases. In the context of bronchopulmonary BPD, especially among preterm infants, Cuproptosis-Related Genes (CRGs) such as NFE2L2 and GLS have been identified as differentially expressed between affected individuals and healthy controls, with these genes located on chromosome 2 showing a positive correlation and ties to immune responses (41, 42). Through bioinformatics analyses, including the GSE108754 dataset, researchers, notably Mingxuan Jia et al., uncovered differential expression patterns of CRGs. In particular, NFE2L2 was more expressed in control samples, while GLS was prominent in treatment groups, and immune infiltration studies revealed significant differences in monocyte cell populations. Applying weighted correlation network analysis (WGCNA) identified the top 25% of gene expressions, leading to the identification of five key marker genes—NFATC3, ERMN, PLA2G4A, MTMR9LP, and LOC440700—using various machine learning algorithms like Generalized Linear Models, Random Forest, SVM, and Extreme Gradient Boosting. The GLM model showed superior accuracy across validation datasets, suggesting these genes could be promising targets for early diagnosis and targeted therapies for BPD, highlighting the significant role of novel cell death mechanisms in the disease (41).

## 2.3 Genetic biomarkers and signatures

### 2.3.1 Genetic risk factors

Studies on heritability, including twin analyses, indicate that genetic factors contribute to 53%–79% of BPD susceptibility (43, 44), underscoring the role of familial influences. Investigations into candidate genes have revealed critical genetic variants, such as mutations in SPOCK2, which is vital for alveolarization, and surfactant-associated genes like ABCA3, linked to both nRDS and BPD development due to their impact on surfactant metabolism and lung development (45, 46). Although genome-wide association studies have struggled to consistently identify common genetic variations related to BPD, the complex genetic landscape appears to be influenced primarily by multiple rare variants (45). Understanding these genetic predispositions is crucial for identifying at-risk infants and developing prevention strategies. A study by Dai et al. sought to enhance the early prediction of BPD in premature infants by integrating genetic factors with

clinical data. Analyzing a cohort of 245 infants born before 32 weeks gestation through exome sequencing, they identified 30 genes associated with BPD and 21 linked to severe BPD. Their predictive models, incorporating these genetic risk gene sets with clinical factors, achieved an AUROC of 0.915 for BPD, outperforming the clinical-only model (AUROC 0.814,  $P=0.013$ ), and a similar improvement was seen in the severe BPD model (AUROC 0.907 vs. 0.826,  $P=0.016$ ). These findings underscore the significant role of genetic factors in BPD susceptibility and indicate that the proposed predictive model may facilitate better risk stratification in premature infants (47).

### 2.3.2 Transcriptomic gene signature

Gene expression profiling provides deep insights into cellular functionality across various tissues, particularly through whole genome microarray analysis, which is the leading method for evaluating transcriptomic variations in specific populations. Although transcriptome-based risk and prediction scores show promise for various adult diseases, their application in BPD has not been extensively explored (48). A recent study by Alvaro Moreira et al. aimed to develop a peripheral blood transcriptomic gene signature using artificial intelligence to identify preterm neonates at risk for BPD. By analyzing whole blood microarray data from 97 very low birth weight neonates on day five of life, the research defined BPD based on the necessity of positive pressure ventilation or supplemental oxygen by 28 days of age. The study found significant differences in gestational age and birth weight between neonates diagnosed with and without BPD. Out of 33,252 genes assessed, 4,523 showed a false discovery rate of less than 1%. Machine learning models that combined five genes achieved AUROC scores between 85.8% and 96.1%, demonstrating strong predictive accuracy. The identified pathways related to T cell development underscore the role of immune factors in BPD's pathogenesis. This study successfully creates a five-gene blood signature for early BPD prediction, emphasizing improved early identification and intervention strategies (49). The research utilized a rigorous methodology with careful participant assignment for model training and validation, ensuring reliability when compared to established predictors like gestational age and birth weight. Additionally, *post hoc* analyses suggest that these transcriptomic signatures retain predictive accuracy among high-risk neonates, and adherence to the TRIPOD reporting guidelines underscores the study's commitment to methodological transparency, enhancing its significance in the field.

## 2.4 Biomarkers and severities of BPD

To understand how biomarkers predict the severities of BPD it is essential to categorize the disease manifestations and examine the biomarkers linked to each. BPD can be classified into three main phenotypes based on clinical presentation: (1) Lung parenchymal disease, (2) Pulmonary vascular disease, and (3) Airway disease. While most patients exhibit significant overlap among these phenotypes, biomarkers can effectively identify the

primary severities of the condition, each reflecting unique underlying mechanisms and inflammatory profiles.

1. **Lung Parenchymal Disease:** This phenotype is characterized by impaired alveolar development and dysregulated inflammation. Biomarkers like IL-6 are elevated in infants with lung parenchymal disease and correlate with lung injury severity. Levels of IL-6 above  $46.125 \text{ pg}\cdot\text{ml}^{-1}$  indicate an increased risk of lung parenchymal disease, supported by an AUC of 0.849. The NLR predicts systemic inflammation, while soluble gp130 indicates inflammatory response involvement. Dysregulation of ERSGs also signifies broader roles in disease progression. Understanding these biomarkers can help identify at-risk infants and guide targeted interventions.
2. **Pulmonary Vascular Disease:** This phenotype features abnormal pulmonary vascular development and increased resistance. Inflammatory cytokines like TNF- $\alpha$  disrupt normal lung development, affecting vascular remodeling. Biomarkers linked to the vascular endothelium, such as angiopoietins and VEGF, provide insights into this phenotype, with elevated levels correlating with altered vascular development and increased risk of pulmonary hypertension in preterm infants. Additionally, ferroptosis-related genes indicate that rising iron levels contribute to oxidative stress and vascular injury. Recognizing these biomarkers can enhance our understanding and therapeutic strategies for pulmonary vascular disease in BPD.
3. **Airway Disease:** This phenotype involves airway inflammation, hyperreactivity, and obstructive pathology and is often associated with altered immune responses and infiltration of specific immune cells, such as neutrophils and macrophages. DEGs related to inflammatory pathways, particularly those in T cell receptor signaling and IL-6/JAK/STAT3 signaling, are vital for predicting airway disease. Elevated IL-8 and reduced IL-10 levels may predispose infants to inflammation and hyperreactivity. Recent studies have identified biomarkers like CYYR1, GALNT14, and OLAH as potential diagnostic indicators for airway disease, linked to inflammatory processes and immune cell infiltration. Understanding these biomarkers can facilitate early identification and intervention, ultimately improving clinical outcomes for affected infants.

### 3 Neonatal respiratory distress syndrome

#### 3.1 Lecithin/sphingomyelin ratio (L/S ratio) biomarkers

nRDS is closely associated with fetal lung maturity, which can be evaluated through the lecithin/sphingomyelin (L/S) ratio in amniotic fluid (50, 51). This ratio is a key indicator of surfactant production, vital for lung function after birth, as lecithin is a primary component of pulmonary surfactant produced by type II pneumocytes (52). A higher L/S ratio indicates more surfactant and a reduced risk of RDS, with a ratio of 2.4 or higher suggesting mature lungs and low RDS risk. Conversely, a ratio

below 1.5 indicates immature lungs and significantly increases RDS risk, while transitional lung maturity is reflected by ratios between 1.5 and 1.9, where some infants may still develop RDS but often recover. The L/S ratio is determined via amniocentesis and thin-layer chromatography, and for at-risk pregnancies, corticosteroids like betamethasone may be given to enhance surfactant production (53, 54). A study by Ahmed et al. introduced a novel diagnostic method that combines attenuated total reflectance Fourier transform infrared spectroscopy (ATR-FTIR) with machine learning to address the lack of an effective point-of-care (POC) diagnostic tool for nRDS. The diagnosis relies on the L/S ratio, with a critical threshold below 2.2 indicating nRDS. Researchers recorded ATR-FTIR spectra for L/S ratios from 1.0–3.4 using purified reagents, and employed principal component regression (PCR) and partial least squares regression (PLSR) to calibrate predictive models based on 155 raw baselined and second derivative spectra. They validated the models using an additional 104 spectra. The best-performing model was a three-factor PLSR model using second derivative spectra, achieving an  $R^2$  of 0.967 and a mean square error (MSE) of 0.014. Predicted L/S ratios ranged from 1.0–3.4, with a prediction interval of +0.29 to -0.37 and a mean interval around the critical ratio of 2.2. The study also examined the effectiveness of mid-infrared (IR) spectroscopy for diagnosing nRDS in preterm infants by measuring key biomarkers, lecithin (L) and sphingomyelin (S). A lung lipid model containing five surfactant lipids was used to train machine learning models for predicting lipid concentrations and the L/S ratio. Advanced statistical techniques, including jackknife and bootstrap methods, quantified prediction uncertainty, indicating the L/S ratio can be determined with approximately  $\pm 0.3 \text{ mol/mol}$  uncertainty. The study identified five significant wavenumbers related to the machine learning models, improving nRDS diagnostics. These findings highlight the potential of integrating ATR-FTIR with machine learning to develop a reliable POC device for detecting and quantifying various biomarkers using distinct mid-infrared spectral signatures, advancing real-time monitoring and outcomes in neonatal care (55).

#### 3.2 Oxidative stress biomarkers genetic polymorphisms

Research into the relationship between single-nucleotide polymorphisms (SNPs) in antioxidant enzymes and oxidative stress markers in neonates is crucial, particularly related to nRDS. Oxidative stress, resulting from an imbalance between reactive oxygen species (ROS) and antioxidant defenses, is notably heightened in preterm infants, making them more vulnerable to RDS. Key antioxidant enzymes like superoxide dismutase (SOD) and catalase (CAT) play essential roles in reducing oxidative damage. Recent studies have shown that preterm neonates with RDS often exhibit increased levels of oxidative stress markers, such as malondialdehyde (MDA) and hydrogen peroxide ( $\text{H}_2\text{O}_2$ ), alongside decreased SOD and CAT activity. Elevated advanced oxidation protein products (AOPPs)

and signals of oxidative DNA damage have been correlated with more severe RDS cases, indicating a strong link between oxidative stress and disease severity (56, 57). SNPs within genes encoding these antioxidant enzymes may affect their activity and expression, influencing a neonate's ability to manage oxidative stress, with specific genetic variations potentially leading to higher stress levels and worse outcomes. Understanding these genetic factors could aid in identifying at-risk neonates and guide targeted interventions, such as antioxidant therapy, while also serving as a prognostic tool for predicting complications related to oxidative stress in this population (58). A study by Sridharan et al. explored the use of machine learning algorithms to predict oxidative stress biomarkers and SNPs associated with RDS and significant alterations in liver functions. The study found that the C5.0 algorithm had the best predictive capability for liver function alterations, achieving an AUC of 0.63, with catalase being a critical predictor. Conversely, the Bayesian network was most effective for predicting RDS, with an AUC of 0.6 and ENOS1 identified as the key predictor (59).

### 3.3 ABCA3 protein modulation

The modulation of the ABCA3 protein is crucial in the development of nRDS, particularly in preterm infants (60). As a member of the ATP-binding cassette transporter family, ABCA3 plays an essential role in pulmonary surfactant metabolism, which helps to reduce alveolar surface tension and prevent lung collapse. Mutations in the ABCA3 gene can lead to surfactant deficiencies, significantly influencing the risk of RDS; studies indicate that 14.3% of RDS infants have these mutations compared to only 3.7% of non-RDS infants. While severe forms of neonatal RDS are more commonly associated with homozygous or compound heterozygous mutations, single mutations are more frequent and account for approximately 10.9% of the attributable risk for RDS in term and late preterm infants. Identifying ABCA3 mutations may serve as a critical marker for recognizing neonates at risk for serious respiratory complications, emphasizing the need for genetic testing to inform clinical management, including considerations for extracorporeal membrane oxygenation (ECMO) interventions (61). Moreover, ABCA3 mutations have been linked to chronic interstitial lung disease later in life, suggesting that early genetic screening could have significant long-term benefits for pediatric health by differentiating genetic causes of neonatal respiratory failure from other etiologies, thereby guiding tailored therapeutic approaches (62). In a recent study by Maria Forstner et al., researchers explored high-content screening to discover pharmacologic modulators for ABCA3 deficiency, a condition linked to respiratory distress syndrome in newborns and interstitial lung disease in children. Given the lack of curative options apart from lung transplantation, the study aimed to develop a cell-based assay powered by machine learning to identify morphological differences between ABCA3 wild-type and mutant cells. Screening 1,280 FDA-approved small molecules led to the discovery of cyclosporin A as a novel and effective

corrector for several ABCA3 variants, although not all. Functional assays further supported these findings, indicating that cyclosporin A could be a promising candidate for orphan drug evaluation and may facilitate controlled repurposing trials for patients with ABCA3-related diseases, addressing a significant gap in therapeutic strategies (63).

### 3.4 Crying as a biomarker

Crying in neonates, particularly preterm infants, has emerged as a valuable biomarker for assessing conditions like nRDS (64). Research has shown a strong correlation between the acoustic characteristics of an infant's cry, particularly the fundamental frequency (F0), and their physiological state, with higher F0 levels indicating greater distress affected by the autonomic nervous system's reaction to stress. Recent studies using multimodal data collection techniques—including cry analysis, electroencephalography (EEG), and near-infrared spectroscopy (NIRS)—have demonstrated that crying can accurately reflect an infant's emotional and physical condition, achieving up to 93% accuracy in distress classification. Machine learning methods have also been leveraged to differentiate between cries related to RDS and those linked to other conditions, such as sepsis, with impressive accuracy rates of 95.3%. This emphasizes the potential for developing automated diagnostic tools that rely on cry analysis (65). Understanding the nuances of cry acoustics could transform clinical practice by providing a non-invasive method to monitor preterm infant health and enable earlier interventions for RDS symptoms. Moreover, the Newborn Cry Diagnostic System (NCDS) utilizes machine learning to assess newborn cries for various health conditions, and recent advancements by Zahra Khalilzad et al. have focused on distinguishing between sepsis and RDS. Their research employed advanced machine learning techniques, particularly Multilayer Perceptron (MLP) and SVM, to analyze cry signals through musical features like the Harmonic Ratio (HR) and speech processing via Gammatone Frequency Cepstral Coefficients (GFCCs). By integrating features from both approaches and refining hyperparameters, they improved diagnostic performance, with SVM achieving a 95.3% accuracy, surpassing the MLP's accuracy of 92.49%, and underscoring the system's capability for effective class separation. This work illustrates the promise of combining diverse feature extraction methods and neural network classifiers to advance cry signal analysis and lays the groundwork for future research involving larger datasets and a wider range of pathologies (66).

### 3.5 Biomarkers and phenotypes of nRDS

To better understand nRDS and its diverse phenotypes, it is essential to explore how specific biomarkers can predict the condition's various manifestations. nRDS is not a uniform diagnosis; it represents a spectrum of clinical presentations influenced by factors such as gestational age, genetic predispositions, and environmental conditions.

1. **L/S Ratio and Phenotype Correlation:** The L/S ratio is a key biomarker for assessing fetal lung maturity, with predictive capabilities beyond risk classification for nRDS. Different nRDS phenotypes—mild, moderate, or severe—may correlate with varying L/S ratios. Infants with ratios near the critical threshold of 2.2 often exhibit milder symptoms and a higher chance of spontaneous recovery, while those with ratios below 1.5 are at greater risk for severe manifestations requiring intensive support. Future studies should stratify nRDS phenotypes based on L/S ratios and correlate them with clinical outcomes to enable tailored management strategies.
2. **Oxidative Stress Biomarkers and Clinical Severity:** Further exploration of oxidative stress biomarkers, such as MDA and H<sub>2</sub>O<sub>2</sub>, is needed to predict nRDS severity. Elevated levels of these markers may indicate increased oxidative stress and serve as predictors of severe respiratory distress. Establishing thresholds for these biomarkers alongside genetic polymorphisms related to antioxidant enzyme activity could help clinicians identify infants at risk for severe nRDS, allowing for proactive interventions. Additionally, examining the interplay between oxidative stress and inflammatory responses could shed light on their roles in different nRDS presentations.
3. **ABCA3 Mutations and Long-term Outcomes:** Identifying ABCA3 mutations as a risk marker for nRDS underscores the importance of genetic screening in predicting both immediate respiratory issues and long-term pulmonary health. Different mutations can result in varying degrees of surfactant dysfunction, leading to distinct nRDS phenotypes. For instance, severe mutations may cause early respiratory failure, while milder mutations could result in a gradual onset of symptoms. Understanding the specific ABCA3 mutations linked to various nRDS phenotypes can inform therapeutic approaches and help families anticipate potential long-term outcomes, including chronic lung disease risks.
4. **Crying as a Multifaceted Biomarker:** Analyzing the acoustic characteristics of crying may provide insights into the physiological and emotional states of neonates, potentially offering predictive value for different nRDS phenotypes. Research could differentiate cry characteristics associated with mild vs. severe nRDS and how these differ from cries linked to other neonatal conditions, like sepsis. Utilizing machine learning algorithms to analyze cry features in larger, diverse populations could lead to a robust framework for using crying as a diagnostic tool, accounting for the nuances of different nRDS manifestations.
5. **Integrative Approaches to Biomarker Prediction:** A comprehensive approach that integrates multiple biomarkers—such as the L/S ratio, oxidative stress markers, ABCA3 genetic variations, and acoustic cry analysis—could significantly enhance predictive accuracy for various nRDS phenotypes. Machine learning models incorporating data from these sources may uncover complex interactions and improve risk stratification. Future research should validate these integrative models in clinical settings, potentially leading to a

multi-faceted diagnostic tool for real-time risk assessment and individualized treatment plans for neonates at risk of nRDS.

## 4 Discussion

Our review of biomarkers and machine learning approaches for predicting BPD and nRDS revealed a range of promising methodologies and results. **Table 1** summarizes key findings in BPD prediction, highlighting the strong predictive capabilities of cytokines and inflammatory biomarkers. Machine learning models such as XGBoost and Random Forest achieved micro-average AUROC values between 0.841 and 0.878, with UCB IL-6 recognized as a critical feature. ERSGs genes demonstrated superior discriminative abilities when analyzed with SVM and specific genes. DEGs analysis identified hub genes via SVM-RFE and LASSO techniques, showcasing how advanced computational methods can reveal significant biological markers. A diagnostic nomogram based on ferroptosis-related genes showed strong ROC and calibration curves, while a genetic biomarker signature achieved AUROC values of 0.915 for general BPD and 0.907 for severe cases, surpassing traditional clinical models. In **Table 2**, focused on nRDS, the (L/S ratio, modeled with PLSR, produced excellent results ( $R^2$  of 0.967). The analysis of oxidative stress biomarkers demonstrated the effectiveness of the C5.0 algorithm for liver function changes and highlighted the Bayesian network's role in RDS prediction. Innovations like high-content screening for cyclosporin A targeting ABCA3 variants show the potential of machine learning in drug discovery. The differentiation of cries associated with RDS through SVM and MLP methods underscores the diverse applications of these technologies in clinical settings.

Recent research has improved our understanding of biomarkers and risk factors for BPD in preterm infants, especially those born before 32 weeks gestation. UCB IL-6 stands out as a promising biomarker with an AUROC of 0.815, strongly correlating with BPD severity. Machine learning models, including XGBoost, CatBoost, LightGBM, and Random Forest, further refined assessments, yielding AUROC values of 0.841–0.878 while consistently identifying UCB IL-6 as the most significant feature. Investigations into ERSGs revealed distinct molecular clusters with notable immune cell infiltration and dysregulated ER stress gene expression. Machine learning algorithms identified potential biomarkers like hub genes CYYR1, GALNT14, and OLAH, as well as therapeutic options such as flunisolide and budesonide (14). An evaluation of ferroptosis identified 23 differentially expressed mRNAs related to fatty acid metabolism, with hub genes LPIN1, ACADSB, WIPI1, and SLC7A11 contributing to a diagnostic nomogram (38). Whole exome sequencing provided insights into genetic risk gene sets tied to BPD and severe BPD (sBPD), achieving AUROC values of 0.915 and 0.907, respectively (47). Additionally, analysis of cuproptosis-related genes revealed differential expression patterns, particularly increased GLS levels in affected groups (41). Through WGCNA and various machine learning algorithms, promising predictive markers like NFATC3, ERMN, PLA2G4A,

TABLE 1 Comprehensive overview of biomarkers and machine learning approaches in predicting BPD.

Biomarker	Details	Machine learning results
Cytokines and Inflammatory Biomarkers	BPD is characterized by chronic lung inflammation with elevated IL-6 levels correlating with increased risk (AUC: 0.849). TNF- $\alpha$ also elevated; acute inflammation linked to IL-1 and IL-8; reduced IL-10 may predispose infants to BPD.	4 models (XGBoost, CatBoost, LightGBM, Random Forest) with micro-average AUROC values: 0.841, 0.870, 0.851, 0.878 respectively. UCB IL-6 identified as most important feature across models.
Endoplasmic Reticulum Stress-Related Genes	ER stress contributes to BPD pathogenesis by disrupting cellular function. Dysregulation of ERSGs noted in BPD infants compared to controls. Understanding these genes might lead to novel therapeutic strategies.	SVM showed superior discriminative ability when combined with five specific genes in assessing clinical risks.
Differentially Expressed Genes (DEGs)	49 cluster-specific DEGs identified; RNA sequencing revealed 471 DEGs ( $p < 0.01$ ), enriched in pathways like IL-6/JAK/STAT3 signaling. Genes SPARC and AGER are linked to lung alveolarization and BPD development.	Identified hub genes (CYR1, GALNT14, OLAH) using SVM-RFE, LASSO, and Random Forest from WGCNA analysis.
Ferroptosis-Related Genes	Ferroptosis implicated in BPD; elevated free iron and abnormal iron accumulation observed. Key biomarkers identified could facilitate diagnosis and intervention.	Constructed a diagnostic nomogram with four hub genes (LPIN1, ACADSB, WIP1, SLC7A11) showing strong predictive performance (ROC and calibration curves).
Cuproptosis-Related Genes (CRGs)	Differential expression patterns of CRGs observed; genes associated with immune responses. Identified five key marker genes (NFATC3, ERMN, PLA2G4A, MTMR9LP, LOC440700) through WGCNA and various machine learning algorithms.	GLM model showed superior accuracy across validation datasets for early diagnosis and targeted therapies in BPD.
Genetic Biomarkers and Signatures	Genetic factors contribute to 53%–79% of BPD susceptibility; mutations in SPOCK2 and genes like ABCA3 identified. A study integrated genetic factors with clinical data to enhance prediction of BPD in preterm infants.	AUROC of 0.915 for BPD; AUROC of 0.907 for severe BPD, outperforming clinical-only models ( $P = 0.013$ for BPD; $P = 0.016$ for severe BPD).
Transcriptomic Gene Signature	Developed a five-gene blood signature for early prediction of BPD; significant differences found in gestational age and birth weight.	AUC scores between 85.8% and 96.1% for predictive accuracy.

TABLE 2 Comprehensive overview of biomarkers and machine learning approaches in predicting nRDS.

Biomarker	Details	Machine learning results
Lecithin/Sphingomyelin Ratio (L/S Ratio)	<ul style="list-style-type: none"> <li>Key indicator of fetal lung maturity and surfactant production.</li> <li>Higher ratios (<math>\geq 2.4</math>) indicate lower RDS risk; lower ratios (<math>&lt; 1.5</math>) indicate a higher risk. Transitional maturity is reflected in ratios of 1.5–1.9.</li> <li>Determined by amniocentesis and thin-layer chromatography.</li> <li>Corticosteroids (e.g., betamethasone) may be administered in at-risk pregnancies.</li> </ul>	ATR-FTIR spectra predicted L/S ratios with a three-factor PLSR model, $R^2$ of 0.967, MSE of 0.014.
Oxidative Stress Biomarkers (OSBs)	<ul style="list-style-type: none"> <li>SNPs in antioxidant enzymes (e.g., SOD, CAT) are linked to oxidative stress, heightened in preterm infants vulnerable to RDS.</li> <li>Preterm neonates with RDS exhibit increased oxidative stress markers (e.g., MDA, H<sub>2</sub>O<sub>2</sub>) and decreased antioxidant activity.</li> <li>Elevated AOPPs and oxidative DNA damage correlate with severe RDS.</li> </ul>	C5.0 algorithm had the best predictive capability for liver function alterations (AUC 0.63). Bayesian network most effective for RDS (AUC 0.6).
ABCA3 Protein Modulation	<ul style="list-style-type: none"> <li>ABCA3 protein is essential for pulmonary surfactant metabolism.</li> <li>Mutations in the ABCA3 gene lead to an increased risk of RDS; ~14.3% of RDS infants carry mutations.</li> <li>Genetic testing for ABCA3 mutations can identify at-risk neonates and inform clinical decisions.</li> </ul>	High-content screening identified cyclosporin A as a potential corrector for ABCA3 variants.
Crying	<ul style="list-style-type: none"> <li>Neonatal crying correlates with physiological states and distress; higher fundamental frequency (F0) levels indicate greater distress.</li> <li>Multimodal data, including cry analysis and EEG/NIRS, improved distress classification accuracy (up to 93%).</li> </ul>	SVM achieved 95.3% accuracy in differentiating cries related to RDS and other conditions; MLP achieved 92.49% accuracy.

MTMR9LP, and LOC440700 were identified. These studies highlight the complex pathophysiology of BPD and the essential role of immunological, genetic, and cellular mechanisms in improving early diagnosis and therapy for this vulnerable population (41).

Recent advancements in diagnosing nRDS have emphasized integrating technologies like attenuated total reflectance ATR-FTIR and machine learning, along with innovative cry-based diagnostics and genetic screening. The combination of

ATR-FTIR and machine learning has shown exceptional effectiveness in predicting the L/S ratio—a key biomarker for nRDS—achieving high correlation metrics with an  $R^2$  of 0.967, positioning it as a promising point-of-care tool for timely interventions (55). Furthermore, research into oxidative stress biomarkers and single-nucleotide polymorphisms through various machine learning approaches highlights the complexities of neonatal conditions, revealing predictive capabilities linked to genetic factors affecting respiratory and liver health (59). The

innovative analysis of cry signals with advanced classifiers suggests vocalizations may serve as potential biomarkers, reflecting the multifaceted nature of neonatal diagnostics and the need for comprehensive screening strategies (66).

## 5 Clinical implications

The advancements in machine learning applications and biomarker identification for BPD and nRDS have profound clinical implications for improving outcomes in at-risk preterm infants. The detection of umbilical cord blood IL-6 as a significant biomarker, with an AUROC of 0.815, equips clinicians with a reliable tool for early diagnosis and risk stratification, especially for infants born before 32 weeks gestation. Utilizing machine learning models such as XGBoost, CatBoost, LightGBM, and Random Forest enhances predictive accuracy, paving the way for personalized interventions and better monitoring of high-risk patients. Furthermore, insights into ERSGs and potential therapeutic targets pave the way for tailored treatment strategies. In nRDS, the integration of attenuated total reflectance ATR-FTIR spectroscopy with machine learning to predict the L/S ratio represents a significant leap in point-of-care diagnostics, enabling timely interventions that can enhance respiratory outcomes. The exploration of oxidative stress biomarkers alongside genetic variations emphasizes the importance of personalized medicine in managing neonatal respiratory conditions. Additionally, investigating cry signals as potential biomarkers introduces an innovative perspective to neonatal assessments, showing that auditory cues may enhance diagnostic practices. These developments underscore the need for ongoing research to seamlessly integrate advanced methodologies into routine clinical practice, ultimately improving screening, early diagnosis, and intervention strategies for vulnerable neonatal populations and enhancing long-term health outcomes.

## 6 Limitations of current research

Despite the promising advancements in identifying biomarkers and employing machine learning techniques for diagnosing BPD and nRDS, several limitations persist. Bronchopulmonary dysplasia (BPD) is a multifaceted condition influenced by various antenatal and postnatal factors in preterm infants, which complicates the establishment of clear causal relationships in studies. The complexity of biological pathways involved in BPD and nRDS raises concerns about the robustness of identified biomarkers, as variations in gene expression and external factors like maternal health, environmental conditions, and neonatal care can confound results. While the integration of machine learning with biomarkers has enhanced the precise prediction of risk for BPD, these models likely did not establish causal inference, which must be acknowledged as a limitation of the study. Machine learning models, while demonstrating strong predictive capabilities, face challenges related to interpretability and the risk of overfitting, particularly when trained on small or

homogeneous datasets. The use of umbilical cord blood levels of biomarkers such as IL-6 may also be influenced by intra- and inter-individual variability, further questioning the generalizability of findings. Additionally, the incorporation of molecular data from ERSGs and cuproptosis-related analyses presents challenges in standardizing protocols and interpreting the clinical relevance of hub genes. Ethical and logistical concerns regarding data privacy and access arise with the integration of genetic information from whole exome sequencing, potentially hindering widespread application in clinical settings. Finally, while innovative diagnostic methods like ATR-FTIR and cry-based diagnostics hold promise for early detection, their implementation may be constrained by resource availability, training requirements for healthcare personnel, and the need for validation in broader, heterogeneous clinical cohorts. These factors highlight the necessity for ongoing research and validation to address current limitations and ensure the effective translation of these advanced techniques into clinical practice for improved outcomes in vulnerable neonatal populations.

## 7 Conclusion

The review underscores the transformative potential of machine learning and advanced diagnostic techniques in predicting and understanding BPD and nRDS in preterm infants. The identification of robust biomarkers, including umbilical cord blood levels of IL-6 and various genetic markers, supports a multifaceted approach that integrates immunological, genetic, and cellular mechanisms. Machine learning models, such as XGBoost, CatBoost, LightGBM, and Random Forest, have shown substantial improvements in predictive accuracy, as indicated by their AUROC values, highlighting their clinical applicability. The investigation of ER stress-related genes and promising results from analyses of ferroptosis and cuproptosis-related genes further reveal the complexity of BPD pathophysiology, suggesting new therapeutic interventions. The use of advanced technologies like ATR-FTIR for L/S ratio prediction, along with cry-based diagnostics and high-content screening, represents a shift toward precision medicine in neonatal care. These advancements enhance understanding, early diagnosis, and treatment strategies for neonatal conditions, leading to improved clinical outcomes. The collective research indicates a significant move towards innovative methodologies, emphasizing the importance of interdisciplinary approaches that leverage machine learning and biomarker discovery.

## Author contributions

HT: Data curation, Formal analysis, Writing – review & editing. SD: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. MV: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. RB: Conceptualization, Writing – original draft, Writing – review & editing. MG-T: Data curation, Validation, Writing – original draft, Writing – review & editing. MD: Conceptualization, Methodology,

Writing – original draft, Writing – review & editing. SA: Methodology, Writing – original draft, Writing – review & editing. ASha: Investigation, Methodology, Writing – original draft, Writing – review & editing. MP: Validation, Visualization, Writing – review & editing. MY: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. AShi: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. AM: Conceptualization, Writing – original draft, Writing – review & editing. HR: Validation, Writing – original draft, Writing – review & editing. HN: Investigation, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

## Acknowledgments

We would like to thank Dr. Hadi Zohouri for his insightful discussions and valuable feedback.

## References

- Bahrami R, Golshan-Tafti M, Dastgheib SA, Alijanpour K, Yeganegi M, Lookzadeh MH, et al. A comprehensive consolidation of data on the relationship between surfactant protein-B (SFTPB) polymorphisms and susceptibility to bronchopulmonary dysplasia. *Fetal Pediatr Pathol.* (2024) 43:436–54. doi: 10.1080/15513815.2024.2400145
- Golshan-Tafti M, Bahrami R, Dastgheib SA, Lookzadeh MH, Mirjalili SR, Yeganegi M, et al. A comprehensive compilation of data on the relationship between surfactant protein-B (SFTPB) polymorphisms and susceptibility to neonatal respiratory distress syndrome. *Fetal Pediatr Pathol.* (2024) 43:399–418. doi: 10.1080/15513815.2024.2390932
- El Shemi MS, Tawfik S, Khafagy SM, Hamza MT, Youssef AMA. Endothelin 1 as a predictor marker for bronchopulmonary dysplasia in preterm neonates with respiratory distress syndrome. *J Neonatal Perinatal Med.* (2017) 10:79–83. doi: 10.3233/NPM-1653
- Gien J, Tseng N, Seedorf G, Kuhn K, Abman SH. Endothelin-1–Rho kinase interactions impair lung structure and cause pulmonary hypertension after bleomycin exposure in neonatal rat pups. *Am J Physiol—Lung Cell Mol Physiol.* (2016) 311:L1090. doi: 10.1152/AJPLUNG.00066.2016
- Mokhtary-Hassanabad A, Mirjalili SR, Lookzadeh MH, Noorishadkam M. Comparison of treatment outcomes of high-flow nasal cannula and nasal continuous positive airway pressure in preterm neonates with respiratory distress syndrome in the NICU of shahid sadoughi hospital in Yazd. *World J Peri Neonatol.* (2024) 6:75–82. doi: 10.18502/WJPN.V6I2.15488
- Wang M, Luo C, Shi Z, Cheng X, Lei M, Cao W, et al. The relationship between cord blood cytokine levels and perinatal characteristics and bronchopulmonary dysplasia: a case–control study. *Front Pediatr.* (2022) 10:807932. doi: 10.3389/fped.2022.807932
- Gao L, Yang P, Luo C, Lei M, Shi Z, Cheng X, et al. Machine learning predictive models for grading bronchopulmonary dysplasia: umbilical cord blood IL-6 as a biomarker. *Front Pediatr.* (2023) 11:1301376. doi: 10.3389/fped.2023.1301376
- Noorishadkam M, Ekraminasab S, Shams SE, Neamatzadeh H. Budesonide-surfactant therapy for neonatal respiratory distress syndrome in preterm infants: a systematic review and meta-analysis of respiratory outcomes. *World J Peri Neonatol.* (2024) 7:16–27. doi: 10.18502/WJPN.V7I1.17324
- Guzmán-Bárceñas J, Calderón-Moore A, Baptista-González H, Irlés C. Clara cell protein expression in mechanically ventilated term and preterm infants with respiratory distress syndrome and at risk of bronchopulmonary dysplasia: a pilot study. *Can Respir J.* (2017) 2017:8074678. doi: 10.1155/2017/8074678

## Conflict of interest

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

## Generative AI statement

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

## Publisher's note

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

- Rojas P, Friaza V, García E, De La Horra C, Vargas SL, Calderón EJ, et al. Early acquisition of pneumocystis jirovecii colonization and potential association with respiratory distress syndrome in preterm newborn infants. *Clin Infect Dis.* (2017) 65:976–81. doi: 10.1093/CID/CIX454
- Xue T, Ma Z, Liu F, Du W, He L, Wang J, et al. Pneumocystis jirovecii colonization and its association with pulmonary diseases: a multicenter study based on a modified loop-mediated isothermal amplification assay. *BMC Pulm Med.* (2020) 20:1–13. doi: 10.1186/S12890-020-1111-4/TABLES/6
- Leigh RM, Pham A, Rao SS, Vora FM, Hou G, Kent C, et al. Machine learning for prediction of bronchopulmonary dysplasia-free survival among very preterm infants. *BMC Pediatr.* (2022) 22:542. doi: 10.1186/S12887-022-03602-W
- He W, Zhang L, Feng R, Fang WH, Cao Y, Sun SQ, et al. Risk factors and machine learning prediction models for bronchopulmonary dysplasia severity in the Chinese population. *World J Pediatr.* (2023) 19:568. doi: 10.1007/S12519-022-00635-0
- Luo L, Luo F, Wu C, Zhang H, Jiang Q, He S, et al. Identification of potential biomarkers in the peripheral blood of neonates with bronchopulmonary dysplasia using WGCNA and machine learning algorithms. *Medicine.* (2024) 103:e37083. doi: 10.1097/MD.00000000000037083
- Chioma R, Sbordone A, Patti ML, Perri A, Vento G, Nobile S. Applications of artificial intelligence in neonatology. *Appl Sci.* (2023) 13:3211. doi: 10.3390/AP13053211
- Keles E, Bagci U. The past, current, and future of neonatal intensive care units with artificial intelligence: a systematic review. *npj Digit Med.* (2023) 6:1–36. doi: 10.1038/s41746-023-00941-5
- Moreira AG, Husain A, Knake LA, Aziz K, Simek K, Valadie CT, et al. A clinical informatics approach to bronchopulmonary dysplasia: current barriers and future possibilities. *Front Pediatr.* (2024) 12:1221863. doi: 10.3389/fped.2024.1221863
- Montagna S, Magno D, Ferretti S, Stelluti M, Gona A, Dionisi C, et al. Combining artificial intelligence and conventional statistics to predict bronchopulmonary dysplasia in very preterm infants using routinely collected clinical variables. *Pediatr Pulmonol.* (2024) 59:3400–3409. doi: 10.1002/PPUL.27216
- Kim J, Villarreal M, Arya S, Hernandez A, Moreira A. Bridging the gap: exploring bronchopulmonary dysplasia through the lens of biomedical informatics. *J Clin Med.* (2024) 13:1077. doi: 10.3390/JCM13041077
- Rivera L, Siddaiah R, Oji-Mmuo C, Silveyra GR, Silveyra P. Biomarkers for bronchopulmonary dysplasia in the preterm infant. *Front Pediatr.* (2016) 4:190469. doi: 10.3389/fped.2016.00033

21. Shams SE, Dastgheib SA, Mousavi-Beni SA, Hosein Lookzadeh M, Mirjalili SR, Golshan-Tafti M, et al. Association of TNF- $\alpha$  genetic variants with neonatal bronchopulmonary dysplasia: consolidated results. *Front Pediatr.* (2024) 12:1511355. doi: 10.3389/fped.2024.1511355
22. Alia A, Kadi F, Yuniati T, Primadi A, Hidajat S, Sukadi A. Relationship between serum interleukin-6 levels and bronchopulmonary dysplasia in preterm infants at 28–34 weeks' gestation with respiratory distress syndrome. *Am J Clin Med Res.* (2019) 7:26–30. doi: 10.12691/AJCMR-7-1-5
23. Bose CL, Laughon MM, Allred EN, O'Shea TM, Van Marter LJ, Ehrenkranz RA, et al. Systemic inflammation associated with mechanical ventilation among extremely preterm infants. *Cytokine.* (2013) 61:315–22. doi: 10.1016/J.CYTO.2012.10.014
24. Balany J, Bhandari V. Understanding the impact of infection, inflammation, and their persistence in the pathogenesis of bronchopulmonary dysplasia. *Front Med.* (2015) 2:171424. doi: 10.3389/fmed.2015.00090
25. Owen JC, Garrick SP, Peterson BM, Berger PJ, Nold MF, Sehgal A, et al. The role of interleukin-1 in perinatal inflammation and its impact on transitional circulation. *Front Pediatr.* (2023) 11:1130013. doi: 10.3389/fped.2023.1130013
26. Sun Y, Chen C, Zhang X, Weng X, Sheng A, Zhu Y, et al. High neutrophil-to-lymphocyte ratio is an early predictor of bronchopulmonary dysplasia. *Front Pediatr.* (2019) 7:464. doi: 10.3389/fped.2019.00464
27. Cakir U, Tayman C, Tugcu AU, Yildiz D. Role of systemic inflammatory indices in the prediction of moderate to severe bronchopulmonary dysplasia in preterm infants. *Arch Bronconeumol.* (2023) 59:216–22. doi: 10.1016/J.ARBRRES.2023.01.003
28. Wu T-J, Teng M, Jing X, Pritchard KA, Day BW, Naylor S, et al. Endoplasmic reticulum stress in bronchopulmonary dysplasia: contributor or consequence? *Cells.* (2024) 13:1774. doi: 10.3390/CELLS13211774
29. Pritchard KA, Jing X, Teng M, Wells C, Jia S, Afolayan AJ, et al. Role of endoplasmic reticulum stress in impaired neonatal lung growth and bronchopulmonary dysplasia. *PLoS One.* (2022) 17:e0269564. doi: 10.1371/JOURNAL.PONE.0269564
30. Golshan-Tafti M, Bahrami R, Dastgheib SA, Hosein Lookzadeh M, Mirjalili SR, Yeganegi M, et al. The association between VEGF genetic variations and the risk of bronchopulmonary dysplasia in premature infants: a meta-analysis and systematic review. *Front Pediatr.* (2024) 12:1476180. doi: 10.3389/FPED.2024.1476180
31. Yang B, Yang L, Wang Y, Maddison LA, Tang Z, Haigh S, et al. Macrophages and neutrophils are necessary for ER stress-induced  $\beta$  cell loss. *Cell Rep.* (2022) 40:111255. doi: 10.1016/J.CELREP.2022.111255
32. Tao Z, Mao Y, Hu Y, Tang X, Wang J, Zeng N, et al. Identification and immunological characterization of endoplasmic reticulum stress-related molecular subtypes in bronchopulmonary dysplasia based on machine learning. *Front Physiol.* (2023) 13:1084650. doi: 10.3389/fphys.2022.1084650
33. Cho HY, Wang X, Campbell MR, Panduri V, Coviello S, Caballero MT, et al. Prospective epigenome and transcriptome analyses of cord and peripheral blood from preterm infants at risk of bronchopulmonary dysplasia. *Sci Rep.* (2023) 13:12262. doi: 10.1038/S41598-023-39313-0
34. Toldi G, Hummler H, Pillay T. T lymphocytes, multi-omic interactions and bronchopulmonary dysplasia. *Front Pediatr.* (2021) 9:694034. doi: 10.3389/FPED.2021.694034
35. Zhu L, Zhou J, Yu C, Gu L, Wang Q, Xu H, et al. Unraveling the molecular regulation of ferroptosis in respiratory diseases. *J Inflamm Res.* (2024) 17:2531. doi: 10.2147/JIR.S457092
36. Garcia MR, Comstock BA, Patel RM, Tolia VN, Josephson CD, Georgieff MK, et al. Iron supplementation and the risk of bronchopulmonary dysplasia in extremely low gestational age newborns. *Pediatr Res.* (2023) 93:701–7. doi: 10.1038/S41390-022-02160-2
37. Zhang Z, Chen K, Pan D, Liu T, Hang C, Ying Y, et al. A predictive model for preterm infants with bronchopulmonary dysplasia based on ferroptosis-related lncRNAs. *BMC Pulm Med.* (2023) 23:367. doi: 10.1186/S12890-023-02670-7
38. Fang C, Tu H, Li R, Bi D, Shu G. Bronchopulmonary dysplasia: analysis and validation of ferroptosis-related diagnostic biomarkers and immune cell infiltration features. *Pediatr Res.* (2024) 96:1673–80. doi: 10.1038/S41390-024-03249-6
39. Pan C, Ji Z, Wang Q, Zhang Z, Wang Z, Li C, et al. Cuproptosis: mechanisms, biological significance, and advances in disease treatment—a systematic review. *CNS Neurosci Ther.* (2024) 30:e70039. doi: 10.1111/CNS.70039
40. Hao D, Luo W, Yan Y, Zhou J. Focus on cuproptosis: exploring new mechanisms and therapeutic application prospects of cuproptosis regulation. *Biomed Pharmacother.* (2024) 178:117182. doi: 10.1016/J.BIOPHA.2024.117182
41. Jia M, Li J, Zhang J, Wei N, Yin Y, Chen H, et al. Identification and validation of cuproptosis related genes and signature markers in bronchopulmonary dysplasia disease using bioinformatics analysis and machine learning. *BMC Med Inform Decis Mak.* (2023) 23:69. doi: 10.1186/S12911-023-02163-X
42. Wang X, Cho HY, Campbell MR, Panduri V, Coviello S, Caballero MT, et al. Epigenome-wide association study of bronchopulmonary dysplasia in preterm infants: results from the discovery-BPD program. *Clin Epigenetics.* (2022) 14:1–20. doi: 10.1186/s13148-021-01220-4
43. Yu KH, Li J, Snyder M, Shaw GM, O'Brodovich HM. The genetic predisposition to bronchopulmonary dysplasia. *Curr Opin Pediatr.* (2016) 28:318. doi: 10.1097/MOP.0000000000000344
44. Hoffmann TJ, Shaw GM, Stevenson DK, Wang H, Quintance CC, Oehlert J, et al. Copy number variation in bronchopulmonary dysplasia. *Am J Med Genet A.* (2014) 164:2672. doi: 10.1002/AJMG.A.36659
45. Lavoie PM, Rayment JH. Genetics of bronchopulmonary dysplasia: an update. *Semin Perinatol.* (2023) 47:151811. doi: 10.1016/j.semper.2023.151811
46. Hadchouel A, Durrmeyer X, Bouzigon E, Incitti R, Huusko J, Jarreau PH, et al. Identification of SPOCK2 as a susceptibility gene for bronchopulmonary dysplasia. *Am J Respir Crit Care Med.* (2011) 184:1164. doi: 10.1164/rccm.201103-0548OC
47. Dai D, Chen H, Dong X, Chen J, Mei M, Lu Y, et al. Bronchopulmonary dysplasia predicted by developing a machine learning model of genetic and clinical information. *Front Genet.* (2021) 12:689071. doi: 10.3389/fgene.2021.689071
48. Moreira A, Tovar M, Smith AM, Lee GC, Meunier JA, Cheema Z, et al. Development of a peripheral blood transcriptomic gene signature to predict bronchopulmonary dysplasia. *Am J Physiol Lung Cell Mol Physiol.* (2023) 324:L76–87. doi: 10.1152/ajplung.00250.2022
49. Moreira AG, Arora T, Arya S, Winter C, Valadie CT, Kwinta P. Leveraging transcriptomics to develop bronchopulmonary dysplasia endotypes: a concept paper. *Respir Res.* (2023) 24:1–9. doi: 10.1186/s12931-023-02596-y
50. Daya A, Lookzadeh MH, Noorishadkam M, Mirjalili SR, Ekraminasab S. The effect of the Iranian surfactant (berakursf) in the treatment of respiratory distress syndrome in premature neonates. *World J Peri Neonatol.* (2023) 6:10–6. doi: 10.18502/WJPN.V6i1.14248
51. Safa FB, Noorishadkam M, Lookzadeh MH, Mirjalili SR, Ekraminasab S. Budesonide and surfactant combination for treatment of respiratory distress syndrome in preterm neonates and evaluation outcomes. *J Clin Neonatol.* (2023) 12:135–41. doi: 10.4103/jcn.jcn\_52\_23
52. Ahmed W, Veluthandath AV, Madsen J, Clark HW, Dushianthan A, Postle AD, et al. Towards quantifying biomarkers for respiratory distress in preterm infants: machine learning on mid infrared spectroscopy of lipid mixtures. *Talanta.* (2024) 275:126062. doi: 10.1016/j.talanta.2024.126062
53. Besnard AE, Wirjosekarto SAM, Broeze KA, Opmeer BC, Mol BWJ. Lecithin/sphingomyelin ratio and lamellar body count for fetal lung maturity: a meta-analysis. *Eur J Obstet Gynecol Reprod Biol.* (2013) 169:177–83. doi: 10.1016/j.ejogrb.2013.02.013
54. Wijnberger LDE, De Kleine M, Voorbij HAM, Arabin B, Bruinse HW, Visser GHA, et al. Prediction of fetal lung immaturity using gestational age, patient characteristics and fetal lung maturity tests: a probabilistic approach. *Arch Gynecol Obstet.* (2010) 281:15. doi: 10.1007/s00404-009-1033-0
55. Ahmed W, Veluthandath AV, Rowe DJ, Madsen J, Clark HW, Postle AD, et al. Prediction of neonatal respiratory distress biomarker concentration by application of machine learning to mid-infrared spectra. *Sensors.* (2022) 22. doi: 10.3390/S22051744
56. Lembo C, Buonocore G, Perrone S. Oxidative stress in preterm newborns. *Antioxidants.* (2021) 10:1672. doi: 10.3390/antiox10111672
57. Elkabany ZA, El-Farrash RA, Shinkar DM, Ismail EA, Nada AS, Farag AS, et al. Oxidative stress markers in neonatal respiratory distress syndrome: advanced oxidation protein products and 8-hydroxy-2-deoxyguanosine in relation to disease severity. *Pediatr Res.* (2020) 87:74. doi: 10.1038/s41390-019-0464-y
58. Dani C, Poggi C. The role of genetic polymorphisms in antioxidant enzymes and potential antioxidant therapies in neonatal lung disease. *Antioxid Redox Signal.* (2014) 21:1863. doi: 10.1089/ars.2013.5811
59. Sridharan K, Sekaran K, Doss C GP, Jufairi MA. Machine learning algorithms and computational validation of single-nucleotide polymorphisms of antioxidant enzymes and oxidative stress markers in neonates. *Biomark Med.* (2023) 17:369–78. doi: 10.2217/bmm-2023-0051
60. Magnani JE, Donn SM. Persistent respiratory distress in the term neonate: genetic surfactant deficiency diseases. *Curr Pediatr Rev.* (2020) 16:17–25. doi: 10.2174/1573396315666190723112916
61. Garmany TH, Moxley MA, White FV, Dean M, Hull WM, Whitsett JA, et al. Surfactant composition and function in patients with ABCA3 mutations. *Pediatr Res.* (2006) 59:801–5. doi: 10.1203/01.pdr.0000219311.14291.df
62. Bullard JE, Wert SE, Whitsett JA, Dean M, Noguee LM. ABCA3 mutations associated with pediatric interstitial lung disease. *Am J Respir Crit Care Med.* (2005) 172:1026. doi: 10.1164/rccm.200503-504OC
63. Forstner M, Lin S, Yang X, Kinting S, Rothenaigner I, Schorpp K, et al. High-content screening identifies cyclosporin A as a novel ABCA3-specific molecular corrector. *Am J Respir Cell Mol Biol.* (2022) 66:382–90. doi: 10.1165/rccb.2021-0223OC
64. Khalilzad Z, Kheddache Y, Tadj C. An entropy-based architecture for detection of sepsis in newborn cry diagnostic systems. *Entropy.* (2022) 24:1194. doi: 10.3390/e24091194
65. Khalilzad Z, Tadj C. Using CCA-fused cepstral features in a deep learning-based cry diagnostic system for detecting an ensemble of pathologies in newborns. *Diagnostics.* (2023) 13:879. doi: 10.3390/DIAGNOSTICS13050879
66. Khalilzad Z, Hasasneh A, Tadj C. Newborn cry-based diagnostic system to distinguish between sepsis and respiratory distress syndrome using combined acoustic features. *Diagnostics.* (2022) 12:2802. doi: 10.3390/DIAGNOSTICS12112802