


Combining artificial intelligence and conventional statistics to predict bronchopulmonary dysplasia in very preterm infants using routinely collected clinical variables

Sara Montagna PhD¹ | Dalila Magno MD^{2,3} | Stefano Ferretti PhD¹ |
 Michele Stelluti MS⁴ | Andrea Gona MD² | Camilla Dionisi MD^{2,5} |
 Giuliana Simonazzi PhD^{2,5} | Silvia Martini PhD^{2,3} | Luigi Corvaglia MD^{2,3} |
 Arianna Aceti PhD^{2,3} 

¹Department of Pure and Applied Sciences (DISPeA), University of Urbino Carlo Bo, Urbino, Italy

²Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy

³Neonatal Intensive Care Unit, IRCCS AOU BO, Bologna, Italy

⁴Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

⁵Obstetric Unit, IRCCS AOU BO, Bologna, Italy

Correspondence

Arianna Aceti, PhD, Neonatal Intensive Care Unit, IRCCS AOU Bologna, Via Massarenti, 13, 40138 Bologna, Italy.
 Email: arianna.aceti2@unibo.it

Funding information

None

Abstract

Background: Prematurity is the strongest predictor of bronchopulmonary dysplasia (BPD). Most previous studies investigated additional risk factors by conventional statistics, while the few studies applying artificial intelligence, and specifically machine learning (ML), for this purpose were mainly targeted to the predictive ability of specific interventions. This study aimed to apply ML to identify, among routinely collected data, variables predictive of BPD, and to compare these variables with those identified through conventional statistics.

Methods: Very preterm infants were recruited; antenatal, perinatal, and postnatal clinical data were collected. A BPD prediction model was built using conventional statistics, and nine supervised ML algorithms were applied for the same purpose: the results of the best-performing model were described and compared with those of conventional statistics.

Results: Both conventional statistics and ML identified the degree of immaturity (low gestational age and/or birth weight), need for mechanical ventilation, and absent or reversed end diastolic flow (AREDF) in the umbilical arteries as risk factors for BPD. Each of the two approaches also identified additional potentially predictive clinical variables.

Conclusion: ML algorithms might be useful to integrate conventional statistics in identifying novel risk factors, in addition to prematurity, for the development of BPD in very preterm infants. Specifically, the identification of AREDF status as an independent risk factor for BPD by both conventional statistics and ML highlights the opportunity to include detailed antenatal information in clinical predictive models for neonatal diseases.

KEYWORDS

artificial intelligence, bronchopulmonary dysplasia, machine learning, preterm infant

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Author(s). *Pediatric Pulmonology* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Despite continuous improvement in neonatal care, bronchopulmonary dysplasia (BPD) remains the most frequent complication of extremely preterm birth, with consequences that might impact on long-term growth, health, and neurodevelopment of affected infants.¹

The strongest risk factor for BPD is known to be prematurity, with the highest risk for infants at the lowest gestational ages (GA). The disease results by the combination of arrested fetal lung development at the alveolar phase due to preterm birth, together with an aberrant response in terms of lung repair to both antenatal and postnatal lung injury.² The constantly improving survival of extremely preterm infants leads to an increasing number of infants at risk for BPD, as these infants suffer significant lung immaturity and are likely to experience, during their stay in the Neonatal Intensive Care Unit (NICU), a huge number of interventions, such as mechanical ventilation, which guarantee survival but are detrimental to lung growth and repair.

The possibility to identify risk factors for BPD additional to the degree of prematurity would allow to tailor clinical care of high-risk infants through proven and effective preventive interventions.³ To this purpose, the National Institute of Child Health and Human Development (NICHD) developed a web-based risk calculator for BPD,⁴ which was recently updated to adapt to evolving BPD definition and respiratory care of preterm infants.^{5,6} To note, the prevalence of BPD varies greatly across centers, and risk factors for BPD may differ according to population characteristics and specific clinical practices, thus limiting a widespread applicability of algorithms or calculators like the one proposed by the NICHD.⁷

In recent years, several studies have tried to move beyond conventional statistical methods and their intrinsic limitations, attempting to define prediction models for clinical outcomes relevant to neonatal medicine through innovative methods based on artificial intelligence (AI), and machine learning (ML) in particular.^{8,9} In this respect, studies addressing BPD are few: most of them are targeted to the role of specific procedures or interventions in predicting BPD onset or severity, such as exome sequencing,¹⁰ gastric aspirates,¹¹ chest X-ray,¹² and modes of respiratory support¹³; so far, only two studies have applied ML techniques to build BPD predictive models based on routinely collected perinatal variables.^{14,15}

The potential for applying AI into NICU data analysis is huge and mostly related to the ability of AI to identify complex relationships among data which might be missed by conventional statistical methods, thus improving early diagnosis, guiding personalized treatment, and optimizing resource allocation. However, at present, several technical, ethical, and clinical limitations prevent routine application of AI methods into the NICU setting.¹⁶

The aim of the study was to apply ML to identify, among routinely collected antenatal, perinatal, and neonatal data, variables predictive of BPD development in very preterm infants, and to critically compare these variables with those identified through conventional statistics.

2 | METHODS

2.1 | Study population

The study was conducted at the level IV NICU of IRCCS AOU Bologna, Italy. Data collection had been approved by the Institutional Review Board (CE AVEC—study ID 76/2013/U/Sper). Parents and/or legal guardians of the recruited infants were asked to provide written informed consent for their children participation in the study.

Infants born between January 2007 and December 2017 with a GA < 32 weeks and/or birth weight (BW) < 1500 g were recruited; variables related to maternal health, pregnancy, and delivery, as well as variables describing neonatal health, were recorded as for standard clinical practice.

Specifically, the following prenatal and perinatal variables were collected: prenatal ultrasound data, including absent or reverse end diastolic flow in the umbilical artery (AREDF), and alterations in the blood flow in the ductus venosus (DV) and middle cerebral artery (MCA), clinical and/or histological chorioamnionitis, maternal hypertension (including pre-eclampsia, gestational hypertension, and chronic maternal hypertension), antenatal steroid prophylaxis, administration of magnesium sulfate (either given for maternal eclampsia or neonatal neuroprotection), fetal growth restriction (FGR), preterm prolonged rupture of membranes (pPROM) and type of delivery (vaginal delivery or cesarean section).

Neonatal data were collected at birth, during hospitalization, and at discharge. At birth, the following variables were recorded: GA, sex, twin status, 5'-Apgar score, cord blood pH and base excess (BE), presence of major congenital malformations and anthropometric measures (birth, length, and head circumference [HC]). Centiles and Z-scores for weight, length and HC at birth were calculated using the Italian Growth standards (INeS growth charts¹⁷).

BPD was defined, according to the criteria set by the 2018 NICHD workshop,¹⁸ as a persistent parenchymal lung disease, confirmed radiographically, in a preterm infant who required, at 36 weeks postmenstrual age, a definite degree of respiratory/oxygen support to maintain optimal arterial oxygen saturations.

During hospitalization, the occurrence of major comorbidities was also recorded: these included respiratory distress syndrome (RDS),¹⁹ need for and length of mechanical ventilation (MV), intraventricular hemorrhage (IVH),²⁰ periventricular leukomalacia (PVL), early-onset and late-onset sepsis (EOS and LOS), retinopathy of prematurity (ROP),²¹ patent ductus arteriosus (PDA) and necrotizing enterocolitis (NEC) and its stage.²²

At hospital discharge, data about nutrition and growth were recorded. Auxological assessment at discharge was carried out using the INTERGROWTH-21st charts, which are the most updated standards for measuring postnatal growth in preterm infants.²³ As for nutrition, infants were categorized as having received exclusive human milk (mother breast milk, donor milk, or both), exclusive formula milk, or mixed feeding.

2.2 | Data analysis

2.2.1 | Conventional statistics

All statistical analyses were carried out using IBM SPSS Statistics for Windows, Version 28.0 (IBM Corp.). Data distribution was examined through the Kolmogorov–Smirnov test. As most variables did follow a normal distribution, parametric tests were used.

Specifically, a univariate analysis was first performed to identify prenatal and neonatal variables potentially related to BPD. The *t* test for independent samples was used for continuous variables, and the chi-square test for dichotomous variables. Potential collinearity between independent variables to be included in the regression models was checked using the Pearson correlation coefficient or the point-biserial correlation coefficient as appropriate. Correlation was defined as “strong” when correlation coefficients were above 0.6. Variables that proved to be significantly different between groups (BPD vs. non-BPD) were used to build a logistic regression model. A *p*-value < .05 was considered statistically significant.

2.2.2 | Machine learning approach

ML allows for the analysis of large data sets to identify relationships between variables and detect crucial features in predictions. Its primary objective is to develop accurate and robust predictive models capable of identifying the effect of various factors or features on the likelihood of developing a specific neonatal disease. Since ML grounds on data, accuracy in data acquisition and sample size are crucial for ensuring models that manifest good performances. In Figure 1, the *t*-distributed stochastic neighbor embedding (*t*-SNE) visualization, a technique commonly employed to project high-dimensional data into a two-dimensional space, is presented. Each point depicted in the visualization corresponds to a data instance,

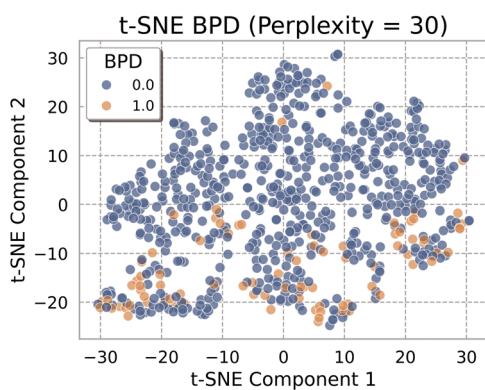


FIGURE 1 Data set *t*-distributed stochastic neighbor embedding (*t*-SNE) visualization: instances from different classes are closely clustered together in the two-dimensional space. However, this mixing of samples suggests overlapping or similar characteristics among the two classes in the high-dimensional feature space. BPD, bronchopulmonary dysplasia.

with its color indicating the respective class membership. This visualization method is instrumental in identifying latent patterns and structures within the data set, thereby facilitating its interpretation and analysis. Typically, instances sharing similarities tend to form clusters in close proximity. Application of *t*-SNE to our data set revealed inherent complexities, as evidenced by the observed overlap between the two distinct classes. Nevertheless, the conspicuous concentration of the BPD class within the bottom right region of the graph implies the presence of an inherent pattern unique to this class. This pattern holds potential significance for extracting distinguishing features that enable the differentiation of BPD patients from non-BPD individual.

In the following paragraphs, the ML pipeline designed to automate and streamline the process of building, training, evaluating, and deploying chosen ML models is described. All analyses were carried out using the Scikit-learn library.²⁴

Data preparation

The purpose of data preprocessing is to ensure that the data are carefully prepared and balanced, laying the groundwork for effective model training and validation. The data preparation phase was carried out as follows: a preliminary strategy of eliminating rows and columns with null values was applied, when these presented several null values greater than 50% of the samples. For the remaining instances, imputation was performed using the median or mode for the specific attribute, depending on the variable type. The detection of outliers was carried out upstream of the analysis. Depending on the characteristics of the column attributes, we opted for either normalization using the Min–Max Scaler or standardization. Since the data set shows a significant imbalance, during the analysis various data balancing techniques were used, either individually or in combination. Both Borderline-SMOTE (synthetic minority over-sampling technique) and synthetic minority over-sampling technique for nominal and continuous (SMOTENC) were used as oversampling methods,^{25,26} while undersampling was performed by randomly selecting records to be removed (Random Undersampling). The data set was divided into train (70%), validation (15%), and test (15%) sets.

The analysis process was conducted in two phases: one including all features and another removing clinical variables considered irrelevant to the analysis (initially there were over 70, reduced to 35).

Training

Different supervised ML algorithms were employed for classification tasks, aiming to exploit the unique properties inherent to each. Specifically, among the linear models, Support Vector Machines and Logistic Regression were utilized. For tree-based methodologies, Decision Tree was employed. *k*-Nearest Neighbors was used as an instance-based method and Gaussian Naive Bayes was adopted as the probabilistic modeling approach, while the adopted ensemble methods included Random Forest, Adaptive Boosting, Gradient Boosting, and Extreme Gradient Boosting (XGBoost).

Models were trained aiming to optimize hyperparameters using a randomized search cross-validation approach. During the training

process, the recall metric was selected as the metric to be optimized. Particularly in the medical field, recall is a priority as it is essential to correctly identify all patients with a specific disease. Although this may cause over-diagnosis, in many medical contexts, it is better to err on the side of caution rather than run the risk of not identifying a pathological condition.

Evaluation

Following the training phase, the top-performing models, characterized by their best hyperparameters, were evaluated on an independent test set. The main goal of this evaluation process was twofold: firstly, to maximize both accuracy and recall metrics, ensuring the model's ability to correctly classify instances of interest while maintaining overall predictive performance. Secondly, the goal was to mitigate the risk of overfitting, thereby promoting the development of a robust and generalizable model capable of effectively handling unseen data.

From the confusion matrices generated during the evaluation, a comprehensive array of performance metrics was computed. These metrics included accuracy, recall, precision, and F1 score, providing a holistic view of the model's classification performance across different categories. Additionally, the loss function and the area under the receiver operating characteristic (ROC-AUC) curves were calculated, offering further insights into model performance and discriminative power.

The results obtained for each ML model, with different pre-processing techniques, are shown in Table 1.

Model interpretability

During the testing phase, the most relevant features were identified, namely those that had a greater impact on predicting the target variables. The computation of feature importance is particularly useful for understanding which variables influence the model's predictions the most and to what extent. Feature importance estimation was conducted using various methods, specifically tailored for the ML algorithm, provided by a fitted attribute in the Scikit-learn library.

3 | RESULTS

3.1 | Study population

Over the 11-year study period, 709 infants fulfilling inclusion criteria were recruited (337 males, 47.5%). Forty-one infants (5.8%) did not survive the neonatal period.

Data collection was nearly complete for neonatal data, while suffered from some missing data related to prenatal variables, especially for outborn infants. The percentage description of prenatal and neonatal data that follows is thus referred to the actual available information for each item.

Six hundred thirty-two over 691 infants (91.5%) were inborn; most infants (565/626, 90.3%) had received at least one dose of antenatal steroid prophylaxis, while a minority (117/588 for whom

TABLE 1 Performance metrics of different supervised machine learning approaches in various test conditions.

Model	Accuracy	Recall	Precision	F1 score
<i>Original data set without preprocessing</i>				
Logistic Regression	0.887	0.385	0.556	0.455
Decision Tree	0.858	0.538	0.438	0.483
Random Forest	0.887	0.077	1.000	0.143
k-Nearest Neighbors	0.887	0.231	0.600	0.333
Support Vector Machines	0.906	0.385	0.714	0.500
Gaussian Naïve Bayes	0.679	1.000	0.277	0.433
Adaptive Boosting	0.877	0.154	0.500	0.235
Gradient Boosting	0.887	0.077	1.000	0.143
Extreme Gradient Boosting	0.887	0.308	0.571	0.400
<i>Test set Oversampling no scaling</i>				
Logistic Regression	0.830	0.692	0.391	0.500
Decision Tree	0.792	0.538	0.304	0.389
Random Forest	0.821	0.615	0.364	0.457
k-Nearest Neighbors	0.792	0.769	0.345	0.476
Support Vector Machines	0.849	0.692	0.429	0.529
Gaussian Naïve Bayes	0.660	0.923	0.255	0.400
Adaptive Boosting	0.783	0.615	0.308	0.410
Gradient Boosting	0.830	0.615	0.381	0.471
Extreme Gradient Boosting	0.906	0.615	0.615	0.615
<i>Test set Undersampling</i>				
Logistic Regression	0.896	0.462	0.600	0.522
Decision Tree	0.868	0.538	0.467	0.500
Random Forest	0.887	0.077	1.000	0.143
k-Nearest Neighbors	0.877	0.308	0.500	0.381
Support Vector Machines	0.896	0.308	0.667	0.421
Gaussian Naïve Bayes	0.679	1.000	0.277	0.433
Adaptive Boosting	0.868	0.077	0.333	0.125
Gradient Boosting	0.896	0.231	0.750	0.353
Extreme Gradient Boosting	0.877	0.462	0.500	0.480
<i>Test set Undersampling + Oversampling SMOTENC</i>				
Logistic Regression	0.830	0.692	0.391	0.500
Decision Tree	0.792	0.462	0.286	0.353
Random Forest	0.840	0.692	0.409	0.514
k-Nearest Neighbors	0.811	1.000	0.394	0.565
Support Vector Machines	0.783	0.923	0.353	0.511
Gaussian Naïve Bayes	0.670	0.923	0.353	0.511
Adaptive Boosting	0.783	0.615	0.308	0.410
Gradient Boosting	0.830	0.692	0.391	0.500
Extreme Gradient Boosting	0.858	0.692	0.450	0.545

Abbreviation: SMOTENC, synthetic minority over-sampling technique for nominal and continuous.

data on prophylaxis were available, 19.9%) had received antenatal magnesium sulfate.

A diagnosis of FGR was made for 147/589 (25%) infants, and pPROM was detected in 163/600 (27.2%) cases. A diagnosis of chorioamnionitis was made in only 15 cases (2.6%), but the number of missing data in this respect was quite high (121/709). Maternal hypertension was present in 163/627 cases (26%). Most infants (593/691, 85.8%) were born after a cesarean section.

Mean (standard deviation [SD]) GA was 29.3 (2.8) weeks and mean (SD) BW was 1157 (352) grams. Two hundred thirty-six (33.4%) infants had a BW < 1000 g (extremely low birth weight [ELBW]).

RDS was diagnosed in 619/677 (91.4%) infants, and BPD in 85/653 (13%) infants. None of the infants without RDS later developed BPD. The percentage of infants requiring MV was 34.5% (237/686).

As for comorbidities, a diagnosis of IVH was made in 177/676 (26.2%) infants (144 mild IVH, 33 severe IVH), while PVL was detected in 20/665 (3%) infants. EOS occurred in 38/667 (5.7%) infants, while LOS in 110/671 (16.4%) infants. ROP occurred in 87/656 (13.3%) infants. PDA was diagnosed in 330/676 infants (48.8%), required pharmacological treatment in 186 infants and surgery in 41 infants. A diagnosis of NEC any stage was made in 74/673 (11%) infants, while infants requiring surgery were 17/673 (2.5%).

3.2 | Data analysis using conventional statistics

At the univariate analysis, several prenatal and neonatal factors, as well as neonatal comorbidities, were found to be associated to BPD development (Table 2).

As for prenatal variables, AREDF status, alterations of the blood flow in the DV, and chorioamnionitis were more frequently detected in infants later developing BPD ($p = .015$, $<.001$, and $.014$, respectively). Vaginal delivery was more frequent in infants developing BPD compared to those who did not ($p = .002$). Quite unexpectedly, BPD occurred more frequently in singleton infants compared to twins ($p = .007$); to note, twin infants had slightly but significantly higher GA and BW compared to singletons (mean GA 29.7 [SD 2.7] vs. 29.1 [2.8] weeks, $p = .004$; mean BW 1237 [SD 341] vs. 1113 [SD 351] g, $p < .001$). No significant difference ($p = .346$) in terms of BPD incidence was documented between uncomplicated and complicated (i.e. triplets, twin-to-twin transfusion syndrome, monochorionic twins, IUGR or loss of one fetus) twin pregnancies.

Overall, infants developing BPD were significantly younger in terms of GA and smaller, both for weight, length, and HC at birth ($p < .001$ for all comparisons). In addition, BPD was significantly associated with lower 5'-Apgar score and higher BE ($p < .001$ for both comparisons). As for neonatal comorbidities, infants developing BPD were more likely to need MV ($p < .001$) and to require it for longer periods ($p < .001$). Infants developing BPD were also at higher risk of other clinical morbidities including IVH, PDA, ROP, EOS, LOS, and surgical NEC ($p < .001$ for all comparison, apart from surgical NEC, for which $p = .035$). As expected, none of the infants without RDS developed BPD ($p < .001$). Despite the

relatively long period of enrollment, which could have impacted on specific features of neonatal care, including for example respiratory and nutritional management, no effect of birth year on BPD risk was documented ($p = .234$).

To select variables to be included in the regression model, potential correlations among variables which had proven to be significant at the univariate analysis were tested: strong and significant correlations were documented between BW and GA ($r = .745$, $p < .001$) and among all anthropometric measures at birth (BW and length: $r = .904$, $p < .001$, BW and HC: $r = .845$, $p < .001$, length and HC: $r = .826$, $p < .001$); for this reason, only GA was included in the conventional statistics model. In addition, PDA and need for pharmacological treatment were strongly related ($r = .638$, $p < .001$), thus leading to the sole inclusion of the latter variable in the model. All the other potential correlations between variables proved to be mild ($r < .6$, $p < .05$) or nonsignificant ($p > .05$).

The final model was then built by including three prenatal variables (AREDF status, Doppler alterations in the DV, and chorioamnionitis), five birth variables (twin status, mode of delivery, 5' Apgar score, arterial cord blood BE, and GA), and six variables related to NICU stay (MV, IVH, LOS, ROP, PDA requiring pharmacological treatment, and surgical NEC). In the final model, four variables proved to be independently related to the occurrence of BPD: AREDF status, GA, MV, and LOS (Table 3).

3.3 | Data analysis using ML approaches

As shown in Table 1, nine supervised ML algorithms were employed; performances of different algorithms were evaluated, to assess which model would provide the best hyperparameters. Different ML algorithms showed variable performance metrics, with only four models demonstrating an F1 score above 0.5: the SVC and the XGB Classifier in the experiment involving oversampling of the data set using the SMOTENC algorithm, without applying any scaling technique, and the RF and XGB Classifier in the experiment where an integration of undersampling and oversampling with SMOTENC were applied. Among these, the only model also showing both recall and precision values above 0.5 was the XGB Classifier Oversampling no scaling.

Figure 2 depicts feature importance estimation and the ROC-AUC curve obtained with this latter model. As shown in the figure, the five most relevant variables which would predict BPD were ELBW, GA, AREDF status, magnesium sulfate prophylaxis, and MV. The area under the curve was 0.92.

Conventional statistics and the best-performing ML algorithm converged into identifying the degree of immaturity, depicted by low GA (and ELBW in the ML approach), MV, and AREDF status, as potential determinants of the risk of BPD. Furthermore, results of conventional statistics suggested an additional association between LOS and BPD, while the ML approach proposed magnesium sulfate antenatal prophylaxis as a potential risk factor for BPD.

Results of the three other models with recall and F1 score above 0.5, but precision below 0.5, are shown in Supplementary Figures E-1 to E-3.

TABLE 2 Univariate analysis of potential prenatal, perinatal and neonatal risk factors for bronchopulmonary dysplasia.

	BPD	No BPD	p Value
<i>Antenatal variables</i>			
Doppler alterations			
Umbilical artery AREFD	24/75 (32.0%)	100/518 (29.3%)	.015
DV alterations	12/75 (16.0%)	18/518 (3.5%)	<.001
MCA alterations	9/75 (12%)	35/517 (6.8%)	.152
Chorioamnionitis	5/68 (7.4%)	8/489 (1.6%)	.014
Maternal hypertension	20/76 (26.3%)	138/519 (26.6%)	1.000
Antenatal steroids	66/76 (86.8%)	471/518 (90.9%)	.294
Magnesium sulfate	15/69 (21.7%)	97/488 (19.9%)	.748
Fetal growth restriction	18/68 (26.5%)	123/480 (25.6%)	.882
pPROM	20/70 (28.6%)	138/499 (27.6%)	.887
<i>Birth variables</i>			
Inborn	77/85 (90.6%)	521/568 (91.7%)	.678
Twin status	19/85 (22.4%)	211/568 (37.1%)	.007
Vaginal delivery	21/85 (24.7%)	64/568 (11.3%)	.002
Gestational age	26.6 (2.1)	30.0 (2.4)	<.001
Weight	778 (251)	1245 (304)	<.001
Weight centile	40.37 (30.47)	45.06 (29.57)	.090
Weight Z score	-0.39 (1.14)	-0.22 (1.06)	.081
Birth weight<1000 g	70/85 (82.3%)	133/568 (23.4%)	<.001
Small for gestational age	19/84 (22.6%)	96/566 (17.0%)	.220
Length	32.5 (3.7)	38.0 (3.4)	<.001
Length centile	36.35 (30.95)	43.37 (30.55)	.029
Length Z score	-0.73 (1.42)	-0.30 (1.20)	.006
Head circumference (HC)	23.6 (2.6)	27.4 (2.3)	<.001
HC centile	39.82 (32.49)	49.07 (32.26)	.015
HC Z score	-0.42 (1.44)	-0.04 (1.27)	.013
Congenital malformations	9/85 (10.6%)	29/568 (5.1%)	.076
Female sex	35/85 (41.2%)	298/568 (52.5%)	.062
5'-Apgar score	7.40 (1.71)	8.53 (1.32)	<.001
Arterial cord blood pH	7.15 (0.79)	7.24 (0.52)	.086
Arterial cord blood BE	-6.23 (6.84)	-3.73 (3.85)	<.001
<i>Neonatal variables</i>			
Mortality	3/85 (3.5%)	6/568 (1.1%)	.100
RDS	84/84 (100%)	505/568 (88.9%)	<.001
Mechanical ventilation (MV)	70/85 (82.4%)	135/568 (23.8%)	<.001
Duration of MV (days)	31.6 (33.6)	9.7 (11.1)	<.001
Intraventricular hemorrhage	41/85 (48.2%)	122/568 (21.5%)	<.001
Periventricular leukomalacia	3/84 (3.6%)	16/562 (2.8%)	.726
Early onset sepsis	5/84 (6.0%)	31/564 (5.5%)	.800
Late onset sepsis	42/85 (49.4%)	66/567 (11.6%)	<.001

TABLE 2 (Continued)

	BPD	No BPD	p Value
Retinopathy of prematurity	41/84 (48.8%)	46/563 (8.2%)	<.001
PDA	66/84 (78.6%)	240/567 (42.3%)	<.001
PDA requiring drugs	49/83 (59.0%)	122/565 (21.6%)	<.001
PDA requiring surgery	19/83 (22.9%)	18/565 (3.2%)	<.001
NEC	14/85 (16.5%)	52/567 (9.1%)	.052
NEC requiring surgery	5/85 (5.9%)	10/566 (1.8%)	.035

Note: Variables are presented as number (percentage) or mean (standard deviation) as appropriate. A *p* value <.05 was considered as statistically significant.

Abbreviations: AREDF, absent or reversed end diastolic flow; BE, base excess; BPD, bronchopulmonary dysplasia; DV, ductus venosus; MCA, medium cerebral artery; NEC, necrotizing enterocolitis; PDA, patent ductus arteriosus; pPROM, preterm prolonged rupture of membranes; SIP, spontaneous intestinal perforation.

TABLE 3 Logistic regression model built to evaluate prenatal, perinatal and neonatal variable associated with the occurrence of bronchopulmonary dysplasia.

	B (S.E.)	EXP(B)	95% CI EXP(B)	p
AREDF	1.526 (0.496)	4.598	1.741-12.143	.002
DV alterations	0.615 (0.655)	1.850	0.512-6.681	.348
Chorioamnionitis	0.449 (0.822)	1.567	0.313-7.815	.585
Twin	0.162 (0.424)	1.175	0.512-2.700	.703
Mode of delivery	-0.384 (0.520)	0.681	0.246-1.886	.460
5'-Apgar score	-0.134 (0.123)	0.875	0.688-1.112	.275
Cord blood BE	-0.026 (0.036)	0.975	0.908-1.046	.476
Gestational age	-0.480 (0.111)	0.619	0.498-0.769	<.001
MV	0.888 (0.439)	2.430	1.028-5.745	.043
IVH	0.229 (0.411)	1.258	0.562-2.813	.577
LOS	1.031 (0.378)	2.803	1.335-5.885	.006
ROP	0.675 (0.420)	1.965	0.863-4.472	.108
Pharm_PDA	0.281 (0.371)	1.324	0.637-2.751	.452
Surgical NEC	-1.493 (0.836)	0.225	0.044-1.156	.074
Constant	10.9340 (3.192)	56356.502		<.001

Note: A *p* value < .05 was considered statistically significant.

Abbreviations: AREDF, absent or reversed end diastolic flow; BE, base excess; DV, ductus venosus; IVH, intraventricular hemorrhage; LOS, late onset sepsis; MV, mechanical ventilation; NEC, necrotizing enterocolitis; PDA, patent ductus arteriosus; ROP, retinopathy of prematurity.

According to these three additional algorithms, additional clinical variables, such as mode of delivery, PVL, ROP, LOS, and NEC, might be linked to BPD. However, the exact impact of these variables is difficult to interpret, as the precision of these models was suboptimal.

4 | DISCUSSION

The present study was aimed at evaluating and comparing the ability of conventional statistics and AI to identify, among routinely collected antenatal, perinatal, and neonatal variables, those predictive of BPD development in very preterm infants. The two approaches converged into identifying, beyond well-known risk factors such as immaturity and mechanical ventilation, also AREDF status, highlighting the opportunity of including prenatal variables into predictive models for neonatal diseases.

Recent years have witnessed the successful integration of AI into healthcare, particularly through ML, which excels in predictive tasks without explicit programming. ML's adaptability has transformed clinical medicine, even surpassing human performance in some instances, notably in computer-aided diagnosis systems.^{27,28} While most randomized controlled trials comparing AI to standard care were focused on intermediate clinical endpoints,²⁹ a recent randomized controlled trial even demonstrated the ability of an AI-enabled electrocardiogram, compared to conventional care, to reduce the risk of all-cause mortality in nearly 16,000 hospitalized patients.³⁰ Medical data are rapidly expanding with the development of new therapies and diagnostics. Health records also accumulate as patients age, develop comorbidities, and undergo more diagnostic testing. Traditional techniques are not equipped to manage this exponential information growth, while ML algorithms are ideally suited to handle abundant and heterogeneous data and may become the most feasible option available in many biomedical settings.³¹ These attributes are particularly useful for large and complex data sets, such as those related to neonatal diseases, including BPD. Neonates who develop BPD are usually hospitalized for several months, generating a great amount of data; furthermore, the complex interplay between prenatal factors and postnatal events that leads to the development of BPD might be more suitable for ML analysis rather than for conventional statistics, which might miss some unique insights.

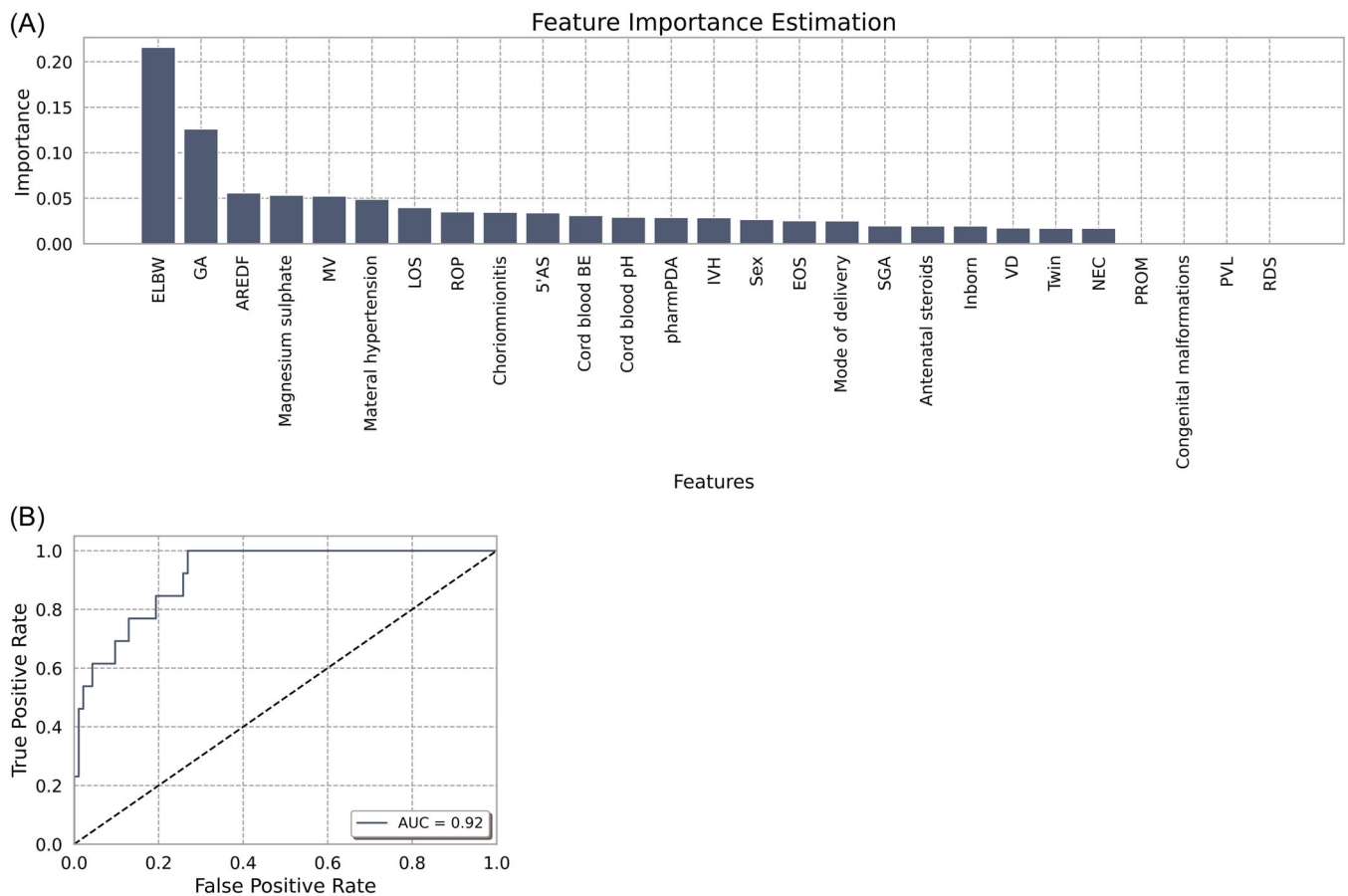


FIGURE 2 Feature importance estimation (A) and the area under the receiver operating characteristic (ROC–AUC) curve (B) obtained through the machine learning algorithm Extreme Gradient Boosting Classifier under the synthetic minority over-sampling technique for nominal and continuous (SMOTENC) oversampling preprocessing strategy. Features are ranked according to their contribution to predictive performance.

Previous studies, mainly based on conventional statistics, have suggested several perinatal and neonatal risk factors for BPD: most studies converge in identifying prematurity-related features (low GA and/or low BW) and the degree of respiratory support as the main risk factors for BPD occurrence and severity^{4,5,15}; furthermore, in analogy with the present study, some authors have proposed a role for other comorbidities, such as LOS and NEC, in increasing BPD risk.

As for papers dealing with ML algorithms to predict BPD, these are few, often based on small samples, and heterogenous in structural modelling and results.^{7,10–15,32} In addition, not all the available ML studies were aimed at investigating clinical predictors of BPD, but some of them were focused on evaluating specific variables, such as chest X-ray features, genomic profile, and BPD markers on biological specimens. To note, however, most ML studies agreed with our results in terms of neonatal variables predictive of BPD, including low GA and BW, need for respiratory support, and comorbidities such as LOS.¹⁵

Interestingly, none of the previous studies was specifically designed to include antenatal clinical and ultrasonographic variables among potential predictors of BPD. The appraisal of AREDF status among potential risk factors for BPD is in line with limited previous observations on the higher incidence of BPD in AREDF versus non-AREDF infants³³; at present, however, it is unclear whether the timing of onset of fetal growth

restriction and the degree of Doppler velocimetry alterations would have a specific impact on BPD risk. Indeed, the identification of AREDF status as an independent risk factor for BPD by both conventional statistics and ML algorithm highlights the opportunity to include detailed antenatal information into clinical predictive models for neonatal diseases. Further studies aimed at linking prenatal data with neonatal clinical outcomes should not describe only the occurrence and timing of fetal growth restriction,³⁴ but should also detail the features of blood flow alterations, as it is well known that the earliest fetal growth restriction with AREDF occurs, the worst will be the potential impact on fetal and neonatal wellbeing.³⁵

Magnesium sulfate prophylaxis was also identified by the ML algorithm as potentially related to BPD, while no difference between BPD and non-BPD infants was documented through conventional statistics (13.4 vs. 12.1%, $p = .748$). A potential explanation for this finding could be related to the so-called “overfitting”, which occurs when an algorithm gives undue weight to an important feature which is strongly associated with another relevant feature (i.e. magnesium sulfate prophylaxis, which is usually indicated in early preterm birth). The two main contributors to overfitting are selection bias and small data set size^{32,36}: indeed, the proportion of infants receiving antenatal magnesium sulfate prophylaxis was quite low, several prenatal data about outborn infants were missing

and we were unable to distinguish indications for treatment (pre-eclampsia vs. neuroprotection).

From a wider perspective, AI has been increasingly used across clinical specialties, showing positive outcomes primarily related to diagnostic yield or performance²⁹; as for neonatal medicine, AI has the potential to become a valuable diagnostic tool, offering several promising new insights into BPD and other neonatal diseases such as NEC, ROP, hypoxic-ischemic encephalopathy, and neurodevelopmental impairment.^{8,9,16,37} As for potential applications in BPD prediction, AI might be used to analyze and combine different antenatal risk factors involved in BPD onset with postnatal clinical, radiological, and laboratory parameters, to develop a prediction model for BPD, or even a score to stratify the severity of BPD. In addition, AI might help in discovering new risk factors involved in BPD pathogenesis, in describing different phenotypes of this heterogeneous disease (as suggested by the recent discovery of a restrictive BPD phenotype³⁸), or in applying novel treatments targeted to specific patients' characteristics. At present, BPD treatment is based on preventive strategies, broadly applied to the entire BPD population to limit injury, and promote repair, whose efficacy likely depends on individual factors which are not entirely understood.

Despite its enormous potential, AI solutions will likely never replace the demanding work of clinicians. Sullivan et al. summarized the major challenges of AI and ML through the acronymous "BARRIERS" (Babies, Analytics, Reactors, Reassurance, Integration, Equipment, Re-education, and Space)³⁹: the word refers to the difficulty of AI at standardizing non-specific neonatal clinical alterations and referring them to an exact disease, creating standardized models which would overcome the heterogeneity of clinical events, and displaying them to a broad range of clinicians, who should be compliant with the technology and confident with its results. Finally, limited space can be a logistic barrier in the NICU. Possible solutions to overcome these barriers are the use of ML-trained models to provide statistical guarantees on discovery findings and the establishment of a multidisciplinary team including healthcare professionals, patients' representatives, and data scientists,³² with complementary roles: patients generate data, healthcare professionals identify which data are useful and how to combine them, and clearly explain the investigated disease to data scientists, who build up a specific set of data.

Some limitations of the present study must be acknowledged: study recruitment encompasses a relatively long period, during which several changes have occurred in neonatal resuscitation guidelines and neonatal intensive care practice, including oxygen supplementation, modes of respiratory support, both at birth and during NICU stay, and nutritional practices. Even if in the present study birth year was not found to affect the risk of BPD, those factors might have had an additional impact on the risk of developing BPD³ and should be considered as potential confounders when planning future studies on this topic. In addition, recent years have witnessed a rising awareness of the importance of maintaining normothermia at birth,⁴⁰ as this is associated with improved clinical outcomes in preterm infants, including BPD.⁴¹ Data about temperature at NICU admission were not available for all the included infants, so we were unable to assess the specific impact of hypothermia on BPD in the study cohort.

5 | CONCLUSIONS

BPD remains one of the most challenging diseases to understand, prevent, and manage in the field of neonatology. An in-depth understanding of its antenatal, perinatal, and postnatal risk factors is increasingly needed to create a modern and more comprehensive definition of BPD, with a consequent impact on diagnosis and treatment.

Our study suggests that ML might discover new variables related to BPD development. Even if accounting for ML limitations and biases, we believe that many solutions are available to overcome these biases and make ML a reliable and essential component of clinical research.

AUTHOR CONTRIBUTIONS

Sara Montagna: Conceptualization; formal analysis; writing—original draft. **Dalila Magno:** Writing—original draft. **Stefano Ferretti:** Formal analysis; writing—review and editing. **Michele Stelluti:** Formal analysis; writing—review and editing. **Andrea Gona:** Data curation; writing—review and editing. **Camilla Dionisi:** Data curation; writing—review and editing. **Giuliana Simonazzi:** Supervision; writing—review and editing. **Silvia Martini:** Data curation; writing—review and editing. **Luigi Corvaglia:** Supervision; writing—review and editing. **Arianna Aceti:** Conceptualization; formal analysis; writing—original draft.

ACKNOWLEDGMENTS

The authors have no funding to report.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Arianna Aceti  <http://orcid.org/0000-0001-9274-985X>

REFERENCES

1. Duijts L, van Meel ER, Moschino L, et al. European respiratory society guideline on long-term management of children with bronchopulmonary dysplasia. *Eur Respir J*. 2020;55:1900788.
2. Thébaud B, Goss KN, Laughon M, et al. Bronchopulmonary dysplasia. *Nat Rev Dis Primers*. 2019;5(1):78.
3. Abiramalatha T, Ramaswamy VV, Bandyopadhyay T, et al. Interventions to prevent bronchopulmonary dysplasia in preterm neonates: an umbrella review of systematic reviews and meta-analyses. *JAMA Pediatrics*. 2022;176(5):502-516.
4. Laughon MM, Langer JC, Bose CL, et al. Prediction of bronchopulmonary dysplasia by postnatal age in extremely premature infants. *Am J Respir Crit Care Med*. 2011;183(12):1715-1722.
5. Greenberg RG, McDonald SA, Laughon MM, et al. Online clinical tool to estimate risk of bronchopulmonary dysplasia in extremely preterm infants. *Arch Disease Childhood Fetal Neonat Ed*. 2022;107(6):638-643.

6. Jensen EA, Dysart K, Gantz MG, et al. The diagnosis of bronchopulmonary dysplasia in very preterm infants an evidence-based approach. *Am J Respir Crit Care Med*. 2019;200(6):751-759.
7. Patel M, Sandhu J, Chou FS. Developing a machine learning-based tool to extend the usability of the NICHD BPD outcome estimator to the Asian population. *PLoS One*. 2022;17:e0272709.
8. McAdams RM, Kaur R, Sun Y, Bindra H, Cho SJ, Singh H. Predicting clinical outcomes using artificial intelligence and machine learning in neonatal intensive care units: a systematic review. *J Perinatol*. 2022;42(12):1561-1575.
9. Chioma R, Sbordone A, Patti ML, Perri A, Vento G, Nobile S. Applications of artificial intelligence in neonatology. *Appl Sci*. 2023;13:3211.
10. Dai D, Chen H, Dong X, et al. Bronchopulmonary dysplasia predicted by developing a machine learning model of genetic and clinical information. *Front Genet*. 2021;12:1-10.
11. Verder H, Heiring C, Ramanathan R, et al. Bronchopulmonary dysplasia predicted at birth by artificial intelligence. *Acta Paediatr*. 2021;110(2):503-509.
12. Xing W, He W, Li X, et al. Early severity prediction of BPD for premature infants from chest X-ray images using deep learning: A study at the 28th day of oxygen inhalation. *Comput Methods Programs Biomed*. 2022;221:106869.
13. Leigh RM, Pham A, Rao SS, et al. Machine learning for prediction of bronchopulmonary dysplasia-free survival among very preterm infants. *BMC Pediatr*. 2022;22(1):542.
14. Wu TY, Lin WT, Chen YJ, Chang YS, Lin CH, Lin YJ. Machine learning to predict late respiratory support in preterm infants: a retrospective cohort study. *Sci Rep*. 2023;13(1):2839.
15. Kostekci YE, Bakirarar B, Okulu E, Erdeve O, Atasay B, Arsan S. An early prediction model for estimating bronchopulmonary dysplasia in preterm infants. *Neonatology*. 2023;120:709-717.
16. Beam K, Sharma P, Levy P, Beam AL. Artificial intelligence in the neonatal intensive care unit: the time is now. *J Perinatol*. 2024;44(1):131-135.
17. Bertino E, Di Nicola P, Varalda A, Occhi L, Giuliani F, Coscia A. Neonatal growth charts. *J Matern Fetal Neonatal Med*. 2012;25(suppl 1):67-69.
18. Higgins RD, Jobe AH, Koso-Thomas M, et al. Bronchopulmonary dysplasia: executive summary of a workshop. *J Pediatr*. 2018;197:300-308.
19. Sweet DG, Carnielli VP, Greisen G, et al. European consensus guidelines on the management of respiratory distress syndrome: 2022 update. *Neonatology*. 2023;120(1):3-23.
20. Inder TE, Perlman JM, Volpe JJ. Preterm IVH/posthemorrhagic hydrocephalus. In *Volpe's Neurology of the Newborn*. 6th ed. Elsevier; 2018;637-698.
21. Chiang MF, Quinn GE, Fielder AR, et al. International classification of retinopathy of prematurity, third edition. *Ophthalmology*. 2021;128(10):e51-e68.
22. Walsh MC, Kliegman RM. Necrotizing enterocolitis: treatment based on staging criteria. *Pediatr Clin North Am*. 1986;33(1):179-201.
23. The Global Health Network. INTERGROWTH 21st. Standards and Tools. <https://intergrowth21.com/intergrowth-21st-applications-calculators>
24. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825-2830.
25. Raschka S, Mirjalili V. Python Machine Learning, 3rd Ed. Packt Publishing; 2019.
26. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. 2002.
27. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nature Med*. 2019;25(1):44-56.
28. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nature Med*. 2022;28(1):31-38.
29. Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. *Lancet Digital Health*. 2024;6(5):e367-e373.
30. Lin C-S, Liu W-T, Tsai D-J, et al. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. *Nature Med*. 2024;30:1461-1470.
31. Piccialli F, Somma VD, Giampaolo F, Cuomo S, Fortino G. A survey on deep learning in medicine: why, how and when? *Information Fusion*. 2021;66:111-137.
32. Shah M, Jain D, Prasath S, Dufendach K. Artificial intelligence in bronchopulmonary dysplasia-current research and unexplored frontiers. *Pediatr Res*. 2023;93(2):287-290.
33. Morsing E, Brodzki J, Thuring A, Maršal K. Infant outcome after active management of early-onset fetal growth restriction with absent or reversed umbilical artery blood flow. *Ultrasound Obstet Gynecol*. 2021;57(6):931-941.
34. Lees CC, Stampalija T, Baschat AA, et al. ISUOG practice guidelines: diagnosis and management of small-for-gestational-age fetus and fetal growth restriction. *Ultrasound Obstet Gynecol*. 2020;56(2):298-312.
35. Della Gatta AN, Aceti A, Spinedi SF, et al. Neurodevelopmental outcomes of very preterm infants born following early foetal growth restriction with absent end-diastolic umbilical flow. *Eur J Pediatr*. 2023;182(10):4467-4476.
36. Kernbach JM, Staartjes VE. Foundations of machine learning-based clinical prediction modeling: part II—Generalization and overfitting. *Acta Neurochir Suppl*. 2022;134:15-21.
37. Kwok TC, Henry C, Saffaran S, et al. Application and potential of artificial intelligence in neonatal Medicine. *Semin Fetal Neonat Med*. 2022;27(5):101346.
38. Shepherd EG, Clouse BJ, Hasenstab KA, et al. Infant pulmonary function testing and phenotypes in severe bronchopulmonary dysplasia. *Pediatrics*. 2018;141(5):e20173350.
39. Sullivan BA, Kausch SL, Fairchild KD. Artificial and human intelligence for early identification of neonatal sepsis. *Pediatr Res*. 2023;93(2):350-356.
40. Ramaswamy VV, de Almeida MF, Dawson JA, et al. Maintaining normal temperature immediately after birth in late preterm and term infants: a systematic review and meta-analysis. *Resuscitation*. 2022;180:81-98.
41. Mohamed SOO, Ahmed SMI, Khidir RJJ, et al. Outcomes of neonatal hypothermia among very low birth weight infants: a meta-analysis. *Matern Health Neonatol Perinatol*. 2021;7(1):14.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Montagna S, Magno D, Ferretti S, et al. Combining artificial intelligence and conventional statistics to predict bronchopulmonary dysplasia in very preterm infants using routinely collected clinical variables. *Pediatr Pulmonol*. 2024;1-10. doi:10.1002/ppul.27216