Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Developing an extubation prediction model utilizing continuous physiologic data for neonates <28 weeks gestational age

Lindsey Knake

lindsey-knake@uiowa.edu

University of Iowa https://orcid.org/0000-0002-1033-0922

Avinash Mudireddy

https://orcid.org/0000-0002-5252-6990

Matthew Jones

Danielle Rios

University of Iowa https://orcid.org/0000-0002-2766-9244

James Blum

Craig Rusin

Milan Sonka

Article

Keywords:

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11236196/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** conflict of interest to disclose.

Abstract

Objective

To incorporate continuous vital sign and ventilator data to improve neonatal extubation failure prediction in high-risk neonates.

Study Design:

In a single-center retrospective cohort, neonates < 28 weeks' gestation with an extubation event and physiologic data were included. Extubation failure was defined as re-intubation within 7 days. The cohort was divided 75/25 into train-test sets with K-fold cross-validation, evaluating multiple machine learning algorithms. Model interpretability was assessed using SHapley Additive exPlanations (SHAP) values.

Results

Of 157 infants, 31 experienced extubation failure (mean time to failure, 58 hours). Using physiologic data and clinician-engineered features, logistic regression performed best (AUC 0.88). Top predictors included post-menstrual age at extubation, median FiO2, median SpO2, maximum heart rate, and minimum SpO2.

Conclusion

Incorporating physiologic data significantly improved extubation prediction over historical models. We propose optimizing negative predictive value to objectively guide earlier ventilator liberation, with future multicenter studies required to enhance performance and transferability.

Introduction

Extubation failure rates in neonates are reported between 20–40% in extremely low gestational age (ELGA) infants with failure rates increasing with decreasing gestational age (GA).(1) Extubation failure in preterm neonates is associated with increased risk of death and prolonged mechanical ventilation.(2, 3) Additionally, the prolonged course and multiple courses of mechanical ventilation are independently associated with chronic lung disease and/or bronchopulmonary dysplasia (BPD).(4)

Deciding to extubate an extremely preterm neonate is one of the most difficult decisions that neonatologists must make with limited scientific evidence available to aid in this decision-making.(3) A recent systematic review of extubation readiness tests and spontaneous breathing trials found that no tests have sufficient evidence to warrant routine use in neonatology.(5, 6, 7) Thus, developing a neonatal extubation prediction model as an objective clinical decision support tool may be useful to clinicians to

reduce mechanical ventilation days, accelerate hospital discharge, and reduce potential long-term complications that afflict neonates throughout their lives.(8) In adult intensive care units (ICUs), standardized weaning protocols have been shown to reproducibly reduce the duration of mechanical ventilation across multiple ICU settings, but few standardized protocols have been implemented in neonatology.(9, 10)

Multiple logistic regression (LR) and machine learning (ML) models have been developed to predict neonatal extubation success with area under the receiver operator curve (AUC) ranging from 0.72–0.94.(11, 12) However, none have been routinely adopted in clinical care and few have been thoroughly tested with multicenter external validation. One LR model does have a clinical decision support tool available online (AUC 0.77) and is undergoing multicenter evaluation. However, the AUC of the model dropped with external validation (AUC 0.72), which is common in single center-derived prediction models.(13, 14) There is concern that the low sensitivity of this model may lead to clinicians not trusting the model.

The ML model above is developed solely on static features and does not incorporate physiologic time-series data such as continuous vital signs. Models that have incorporated physiologic time-series data (heart rate variability or intermittent hypoxemia) have shown better performance metrics.(11, 12) In the adult literature, there have been multiple models using time-series data that have demonstrated considerable improvement with the integration of continuous ventilator and vital sign data.(15) We hypothesize that incorporating such physiologic time-series data may improve neonatal extubation prediction models.

Objective

To determine if physiologic time-series data may enhance neonatal prediction models by performing a single center pilot study of an extubation failure prediction model for high-risk neonates.

Methods

A single center retrospective cohort study was performed in an 88-bed level IV neonatal intensive care unit (NICU) at the University of Iowa. Patients were included if they were < 28 weeks gestational age (highest risk group of extubation failure) and had an extubation event with 24-hours of continuous vital sign and ventilator data available (collected in 2 second increments) from March 2020 – June 2023. Data was collected using Sickbay™(Medical Informatics Corp, Houston, TX, USA) an FDA-cleared Class II medical device for remote patient monitoring that collects data from bedside monitors and ventilators at native resolution, including waveforms, vital signs, alarms, and alarm settings.

Extubation failure was defined as the need for respiratory re-intubation within 7 days, which is consistent with published definitions.(1, 2, 3, 16) Only the first extubation attempt was included in this dataset and unplanned extubations were included only if the patient received a trial of non-invasive support. Per local unit protocol, extremely low birth weight (ELBW) infants are generally extubated to continuous positive

airway pressure (CPAP) delivered via a nasopharyngeal tube when weight is > 900g and non-invasive positive pressure ventilation (NIPPV) is provided via neurally adjusted ventilatory assist (NAVA) due to the concern for significant apnea, bradycardia, and desaturation events after extubation. The University of Iowa institutional review board approved this retrospective study as qualifying for a waiver of consent.

The 24 hours of continuous physiologic data prior to extubation were summarized with statistical metrics (min, max, median, and variance) and all missing values were dropped and thus not included in these metrics. Additional clinical features were engineered for this model (low respiratory rate, intermittent hypoxia, etc.). SMOTE was applied to these aggregate features and was not performed on the raw physiologic waveforms. Intermittent hypoxia (IH) was defined by using the number of SpO_2 values that dropped below a threshold of 80. In addition, the respiratory rate (RR) was split into three categories: low RR < 30 breaths per minute (bpm), normal RR classified as 30 - <60 bpm, and high RR $\geq$ 60 bpm. Additional clinical features chosen a priori by the clinical team based on previously published literature were included in the model. These included gestational age, highest oxygenation index and respiratory severity score in the first 12 hours of life, antenatal steroids, histologic chorioamnionitis (from available pathology reports), and pre-extubation pH.(13, 18, 19)

We divided the cohort into a 75/25 train-test split with a K-fold cross validation performed on the training data. Due to the highly imbalanced training data, Synthetic minority over-sampling techniques (SMOTE) were applied to the training set to generate synthetic samples for the extubation failure class.(17) A suite of models was trained on the resampled dataset. Hyperparameter optimization was conducted using randomized search with cross-validation with model performance evaluated using average AUC as the primary selection criterion.

Both physiologic and engineered features were passed through common ML models (random forest, extreme gradient boosting, artificial neural networks, gaussian naïve bayes) and compared to traditional LR modeling. Overall missingness of the static data was low (5.2%) and physiologic features had 0.38% of missingness. After aggregating the data into patient-level features, any remaining missing values were imputed using the median imputation across all features.

Descriptive statistics were used to compare demographic and clinical information utilizing chi square and Wilcoxon rank sum test for categorical and continuous variables, respectively. Model performance was assessed by quantifying discrimination using the AUC and model calibration was assessed numerically via Brier score and a logistic recalibration model. To determine an optimal classification threshold, we selected the probability cutoff that maximized the F1 score for clinical outcome, rather than using a default threshold. Model interpretability was evaluated using SHapley Additive exPlanations (SHAP), which estimate the contribution of each feature to individual predictions.(20) SHAP summary and dependence plots were generated to examine global feature importance and the direction and magnitude of feature effects.

Results

157 infants met inclusion criteria and 31 qualified as extubation failure in a naturally unbalanced data set with the average time to extubation failure of 58 hours. Infants with less exposure to antenatal steroids and a lower post-menstrual age (PMA) at extubation failed more often (Table 1). Univariate analysis found that SpO_2 median, minimum, and variance, and heart rate median and variance were significantly different between the groups.

Table 1
Descriptive statistics of standard demographics

Gestational age (weeks)	**Entire cohort (n = 157)**	**Extubation success (n = 126)**	**Extubation failure (n = 31)**	**P-value**
	25 [23–26]	**25 [23–26]**	**25 [23–26]**	**0.73**
21 to < 25 weeks	68 (43%)	55 (44%)	13 (42%)	0.02
Birth weight (grams)	688 [578–925]	689 [578–936]	680 [560–910]	0.71
Male	88 (56%)	70 (56%)	18 (58%)	0.80
Inborn	114 (73%)	90 (71%)	24 (77%)	0.50
Surfactant given[#]	150 (96%)	121 (97%)	29 (93%)	0.40
Histologic chorioamnionitis[#]	47 (30%)	36 (29%)	11 (35%)	0.45
Any antenatal steroids	141 (90%)	116 (92%)	25 (81%)	0.03
Highest OI in first 12 hours[*]	12 [7–22]	14 [7–22]	11 [7–16]	0.31
Highest RSS in first 12 hours[*]	5 [3–8]	5 [3–8]	4 [3–8]	1.0
Pre-extubation pH	7.36 [7.34–7.40]	7.37 [7.34–7.40]	7.35 [7.32–7.39]	0.05
Extubation data				
PMA at extubation (weeks)	30 [28–32]	30 [28–32]	28 [27–30]	0.001
Weight at extubation (g)	1170 [1025–1471]	1180 [1046–1545]	1150 [963–1297]	0.06
Periextubation steroids	105 (67%)	85 (67%)	20 (64%)	0.76
Continuous ventilator and physiologic data (24 hours prior to extubation)				
Values displayed as median [IQR] or n (%).				
PMA, post-menstrual age; OI, oxygenation index (MAP × FiO_2 × 100/PaO_2); RSS, respiratory severity score (MAP × highest FiO_2); MAP, mean airway pressure; HR, heart rate; RR, respiratory rate.				
[*]Missing values present, imputed to median.				
[#]Missing values present, imputed to mode.				

Gestational age (weeks)	Entire cohort (n = 157)	Extubation success (n = 126)	Extubation failure (n = 31)	P-value
	25 [23–26]	**25 [23–26]**	**25 [23–26]**	**0.73**
FiO_2 Median (%)	31 [26–40]	30 [25–40]	33 [30–39]	0.22
Vent MAP Median	10 [8–11]	10 [8–11]	10 [8–11]	0.63
SpO_2 Median	91 [89–95]	91 [90–95]	89 [89–91]	< 0.005
SpO_2 Min	42 [32–50]	43 [33–51]	35 [28–45]	0.03
SpO_2 Variance	39 [22–52]	37 [20–51]	50 [28–61]	0.01
HR Max	200 [194–207]	201 [195–208]	197 [192–203]	0.07
HR Median	159 [150–166]	158 [148–164]	162 [154–170]	0.01
HR Min	65 [53–85]	65 [50–87]	68 [61–83]	0.44
HR Var	126 [89–190]	137 [96–203]	96 [74–159]	0.01
RR Max	115 [106–125]	114 [106–124]	118 [110–126]	0.24
RR Median	44 [39–48]	44 [39–48]	42 [38–47]	0.67
RR Min	10 [0–14]	10 [2–13]	7 [0–16]	0.86
RR Var	170 [126–217]	161 [125–214]	192 [126–233]	0.22
Values displayed as median [IQR] or n (%).				
PMA, post-menstrual age; OI, oxygenation index (MAP × FiO_2 × 100/PaO_2); RSS, respiratory severity score (MAP × highest FiO_2); MAP, mean airway pressure; HR, heart rate; RR, respiratory rate.				
*Missing values present, imputed to median.				
#Missing values present, imputed to mode.				

Using physiologic data and clinician engineered features, we found that LR was the had the best discrimination (AUC 0.88) when compared to other ML models (Fig. 1), however, there was limited calibration (Brier 0.17, slope 0.66, intercept − 1.40). The top variables included in the model are displayed in Fig. 2 with PMA, fraction of inspired oxygen (FiO_2) median, SpO_2 median, heart rate maximum, SpO_2 minimum, GA, any antenatal steroid exposure, and SpO_2 variance being the top 8 variables of importance. A list of all variables included in the model is included in Supplemental Table 1.

The optimal F1 threshold for this model was 0.49 to achieve an F1 score of 0.54 with a precision or positive predictive value (PPV) of 0.43, negative predictive value (NPV) of 0.92, sensitivity (recall) of 0.75, and specificity of 0.75. To adjust our threshold for a more relevant clinical outcome of no false negative results (not predicting extubation success when a patient fails extubation), the threshold would be 0.43 for an F1 score of 0.63, Specificity of 0.72, PPV of 0.47, NPV of 1, and sensitivity (recall) of 1. Using the 40 patients hold out test set, Fig. 3 represents a two-by-two outcome table displaying the outcome changes when using a different threshold for the prediction. Figure 4 demonstrates the different outcomes with the different threshold response.

Discussion

Using a small single center dataset, we developed an extubation failure prediction model with a higher AUC than previously published models available for use (extubation.net). Of the top 10 performing variables, 5 were extracted from continuous vital sign data. In addition, we proposed creating a model prediction threshold that focuses on optimizing negative predictive values, which would result in the model predicting extubation success accurately with few false negatives. This would encourage liberalization from the ventilator sooner if the prediction model is predicting a low chance of extubation failure.

While no single performance metric is perfect when comparing prediction models(11), AUC is the most common metric reported in literature. However, it is important to understand the clinical context of the model and evaluation metrics (PPV, NPV, etc.). For example, the extubation model created by Brasher et al. achieved an AUC of 0.94 when they incorporated ventilator data and pulse oximetry data to calculate intermittent hypoxemia.(18) However, this study was limited to a sample size of 24 infants with no external validation, which suggests their model is overfit to their small sample size but supports our hypothesis that using physiologic time series data results in improved model performance. Our model also showed strong discrimination with a high AUC, but limited calibration (Brier score of 0.17), consistent with overfitting given the modest number of extubation failure events. Kanbar et al. incorporated the number of desaturations and respiratory rate profiles in their extubation prediction model (AUC 0.75), but stated that their model failed at trying to prioritize all clinical outcomes since they had a low PPV by classifying 66% of infants as extubation failure when they resulted in extubation success.(19) Our model had improved results with only 53% of infants predicted as an extubation failure resulting in extubation success (Fig. 4). However, many clinical prediction models (i.e., sepsis prediction model) have a low PPV because optimizing PPV must be balanced by minimizing false negatives or positives (i.e., false alarms).(21, 22)

The clinical utility of a model with a low PPV model should not be dismissed in this clinical condition. A model that can accurately predict extubation success (high NPV) is remarkably useful. If clinicians are provided with an objective prediction model suggesting that a patient is ready for extubation, they may be likely to extubate sooner and reduce the number of mechanical ventilation days which have been

associated with increased risk of BPD, worse neurodevelopmental outcomes, and prolonged LOS.(4, 23, 24, 25, 26)

Any extubation prediction model must also consider the risk of prolonged intubation duration. Infants that this model predicts to have a higher risk of extubation failure may require further lung optimization prior to extubation but still deserve a trial of extubation when clinicians deem appropriate. This model may provide an objective talking point with families about a predicted risk of extubation failure when performing a trial of extubation. Only one 2008 randomized controlled trial explored delaying extubation by 36 hours in < 28 week neonates and found similar outcomes in extubation failure and BPD, but the delayed extubation group had a decreased number of days requiring mechanical ventilation and decreased length of stay.(27)

A recent systematic review on extubation failure prediction models showed there is wide variability in the parameters included in extubation prediction models.(2) However, in a subgroup analysis of preterm infants with GA < 28 weeks, predictors associated with extubation failure included increased immaturity (low GA, birthweight, and PMA at extubation) and higher presurfactant FiO_2 and lower pre-extubation pH.(2) These clinical variables were consistent with our findings. However, we found physiologic variables such as heart rate and SpO_2 to also be top contributors to our model.

This study has several limitations. Given the single-center setting, relatively homogeneous population at the University of Iowa, and small sample size, subgroup analyses by race/ethnicity were not feasible due to risk of overfitting, and model fairness across sociodemographic groups could not be assessed. Due to this limited dataset, we focused on patients < 28 weeks gestational age who are at highest risk of extubation failure. We also used a SMOTE statistical technique to address the limited actual data, which has a risk of bias. However, this was only used on the training dataset and the models performed adequately on the hold out test dataset. However, given the small sample size our model is likely overfit to our single center data, thus, future larger multicenter studies are needed. Finally, while we examined a number of features, we did not include some other known clinical predictors of extubation success, such as lung ultrasound scores due to lack of availability in this population.(28, 29) In addition, further exploration of ML techniques can be used in larger databases such as utilizing masked autoencoders or using low-pass filtering pipelines to extract feature embeddings directly from physiological waveforms.(30, 31, 32)

As the immediate future step leading to broader validation of our approach, we propose developing this physiology-based clinical prediction model in a larger multicenter database as an additional objective tool for clinicians to use when making the decision to extubate. Prediction models should not be expected to replace clinical judgment, as they will not incorporate all the clinical context that a clinician may have. However, they may provide an objective score that can further help the difficult clinical decision making task by monitoring such objective scores over time and by contributing to understanding when a patient may be ready for a trial of extubation. Future work is needed to externally validate and refine this extubation prediction model. In this pilot study, a LR model had the best

performance, but more advanced ML modeling techniques will likely perform even better and offer additional transferability across institutions when trained on data with larger sample sizes from multiple centers.

Conclusion

Using a single-center pilot study, we developed an extubation prediction model that performs better than prior historical models due to the incorporation of time-series physiologic data and advanced statistical techniques. We propose developing an extubation prediction model that focuses on optimizing NPV with the goal to provide an objective measure to liberalize neonates from mechanical ventilation sooner. Future multicenter investigation will likely enhance model performance and transferability.

Declarations

Conflicts of Interest:

No competing conflicts of interest for any co-authors.

Use of Artificial Intelligence

The authors used a large language model solely for grammar and language editing during manuscript preparation. The tool was not used for study design, data analysis, interpretation of results, or drafting of scientific content. All AI-assisted edits were reviewed by the authors, who take full responsibility for the accuracy and integrity of the manuscript.

Funding:

Research reported in this publication was supported by the National Heart, Lung, And Blood Institute of the National Institutes of Health under Award Number K23HL179403. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions:

Dr. Lindsey Knake and Avinash Mudireddy had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Matthew Jones assisted with chart review and additional data collection. All co-authors contributed to the conception and design of the analysis. AM and LK performed the statistical data analysis, and all co-authors contributed to the interpretation of data. LAK drafted the manuscript and all co-authors gave approval of the final manuscript for publishing.

Availability of Data and Materials:

De-identified data, code, and logistic regression coefficients may be made available upon request to the authors. No prospective study protocol was prepared and the study was not registered, as this was an exploratory single-center pilot study intended to inform the design of future, larger investigations. Patients and the public were not involved in the design, conduct, reporting, or dissemination of this study.

References

1. Shalish W, Keszler M, Kovacs L, Chawla S, Latremouille S, Beltempo M et al. Age at First Extubation Attempt and Death or Respiratory Morbidities in Extremely Preterm Infants. J Pediatr 2023; 252: 124–130.e123.
2. Hoff Calegari L, Goyal M, Dutta S, Mukerji A. Predictors and outcomes of extubation failure in preterm neonates: A systematic review. Pediatrics 2025; 155(2): e2024068677.
3. Shalish W, Keszler M, Davis PG, Sant'Anna GM. Decision to extubate extremely preterm infants: art, science or gamble? Arch Dis Child Fetal Neonatal Ed 2022; 107(1): 105–112.
4. Jensen EA, DeMauro SB, Kornhauser M, Aghai ZH, Greenspan JS, Dysart KC. Effects of Multiple Ventilation Courses and Duration of Mechanical Ventilation on Respiratory Outcomes in Extremely Low-Birth-Weight Infants. JAMA Pediatr 2015; 169(11): 1011–1017.
5. Shalish W, Kanbar L, Kovacs L, Chawla S, Keszler M, Rao S et al. Assessment of Extubation Readiness Using Spontaneous Breathing Trials in Extremely Preterm Neonates. JAMA Pediatr 2020; 174(2): 178–185.
6. Shalish W, Latremouille S, Papenburg J, Sant'Anna GM. Predictors of extubation readiness in preterm infants: a systematic review and meta-analysis. Arch Dis Child Fetal Neonatal Ed 2019; 104(1): F89–f97.
7. Alarcon-Martinez T, Latremouille S, Kovacs L, Kearney RE, Sant'Anna GM, Shalish W. Clinical usefulness of reintubation criteria in extremely preterm infants: a cohort study. Arch Dis Child Fetal Neonatal Ed 2023; 108(6): 643–648.
8. Khodak I, Kahovec M, Romano V, Nielsen A, Day CL, Dylag AM. Reducing Failed Extubations in Preterm Infants Via Standardization and Real-Time Decision Support. Pediatrics 2024; 154(3).
9. McConville JF, Kress JP. Weaning patients from the ventilator. N Engl J Med 2012; 367(23): 2233–2239.
10. Gentle SJ, Travers CP, Winter L, Tucker J, Purvis D, Hoover A et al. Decreasing Exposure to Mechanical Ventilation in Extremely Preterm Infants. Pediatrics 2025; 156(6).
11. Husain A, Knake L, Sullivan B, Barry J, Beam K, Holmes E et al. AI models in clinical neonatology: a review of modeling approaches and a consensus proposal for standardized reporting of model performance. Pediatr Res 2025; 98(2): 412–422.

12. Jenkinson AC, Dassios T, Greenough A. Artificial intelligence in the NICU to predict extubation success in prematurely born infants. J Perinat Med 2024; 52(2): 119–125.
13. Gupta D, Greenberg RG, Sharma A, Natarajan G, Cotten M, Thomas R et al. A predictive model for extubation readiness in extremely preterm infants. J Perinatol 2019; 39(12): 1663–1669.
14. Dryer RA, Salem A, Saroha V, Greenberg RG, Rysavy MA, Chawla S et al. Evaluation and validation of a prediction model for extubation success in very preterm infants. J Perinatol 2022; 42(12): 1674–1679.
15. Torrini F, Gendreau S, Morel J, Carteaux G, Thille AW, Antonelli M et al. Prediction of extubation outcome in critically ill patients: a systematic review and meta-analysis. Crit Care 2021; 25(1): 391.
16. Shalish W, Kanbar L, Keszler M, Chawla S, Kovacs L, Rao S et al. Patterns of reintubation in extremely preterm infants: a longitudinal cohort study. Pediatr Res 2018; 83(5): 969–975.
17. Chawla N.V. KWB, L. O. Hall, W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res 2002; 16: 321–357.
18. Brasher M, Virodov A, Raffay TM, Bada HS, Cunningham MD, Bumgardner C et al. Predicting Extubation Readiness in Preterm Infants Utilizing Machine Learning: A Diagnostic Utility Study. J Pediatr 2024: 114043.
19. Kanbar LJ, Shalish W, Onu CC, Latremouille S, Kovacs L, Keszler M et al. Automated prediction of extubation success in extremely preterm infants: the APEX multicenter study. Pediatr Res 2023; 93(4): 1041–1049.
20. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell 2020; 2(1): 56–67.
21. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Intern Med 2021; 181(8): 1065–1070.
22. Lafon T, Weingart M, Vaidie J, Calfee CS, Jacob ST, Freund Y et al. Challenges in early detection and prognostication of sepsis: new approaches from the emergency department and intensive care unit. EClinicalMedicine 2026; 94: 103864.
23. Yang Y, Gu XY, Lin ZL, Pan SL, Sun JH, Cao Y et al. Effect of different courses and durations of invasive mechanical ventilation on respiratory outcomes in very low birth weight infants. Sci Rep 2023; 13(1): 18991.
24. Guillot M, Guo T, Ufkes S, Schneider J, Synnes A, Chau V et al. Mechanical Ventilation Duration, Brainstem Development, and Neurodevelopment in Children Born Preterm: A Prospective Cohort Study. J Pediatr 2020; 226: 87–95 e83.
25. Rocha G, Amaral R. Impact of prolonged invasive mechanical ventilation in extremely preterm infants on long-term neurodevelopmental and neurosensory outcomes: A narrative review. Early Hum Dev 2026; 217: 106519.
26. Park SJ, Bae MH, Jeong MH, Jeong SH, Lee N, Byun SY et al. Risk factors and clinical outcomes of extubation failure in very early preterm infants: a single-center cohort study. BMC Pediatr 2023;

23(1): 36.

27. Danan C, Durrmeyer X, Brochard L, Decobert F, Benani M, Dassieu G. A randomized trial of delayed extubation for the reduction of reintubation in extremely preterm infants. Pediatr Pulmonol 2008; 43(2): 117–124.
28. Gautam G, Priyadarshi M, Chaurasia S, Singh P, Basu S. Sonographic evaluation of lungs and diaphragm as a predictor for extubation failure in early preterm neonates: a prospective diagnostic accuracy study. Eur J Pediatr 2025; 184(11): 725.
29. Jiang L, Li F, Hong L, Yang X, Xiao L, Ke L et al. Application of a combined predictive model based on lung ultrasound score trajectory changes in deciding mechanical ventilator weaning for neonatal respiratory distress syndrome: a retrospective study. Front Med (Lausanne) 2026; 13: 1764757.
30. Woo G, Liu C, Kumar A, Xiong C, Savarese S, Sahoo D. Unified training of universal time series forecasting transformers. Forty-first International Conference on Machine Learning; 2024.
31. Chen M, Shen L, Li Z, Wang XJ, Sun J, Liu C. Visionts: Visual masked autoencoders are free-lunch zero-shot time series forecasters. arXiv preprint arXiv:240817253 2024.
32. He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked autoencoders are scalable vision learners. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022; p. 16000–16009.

Figures

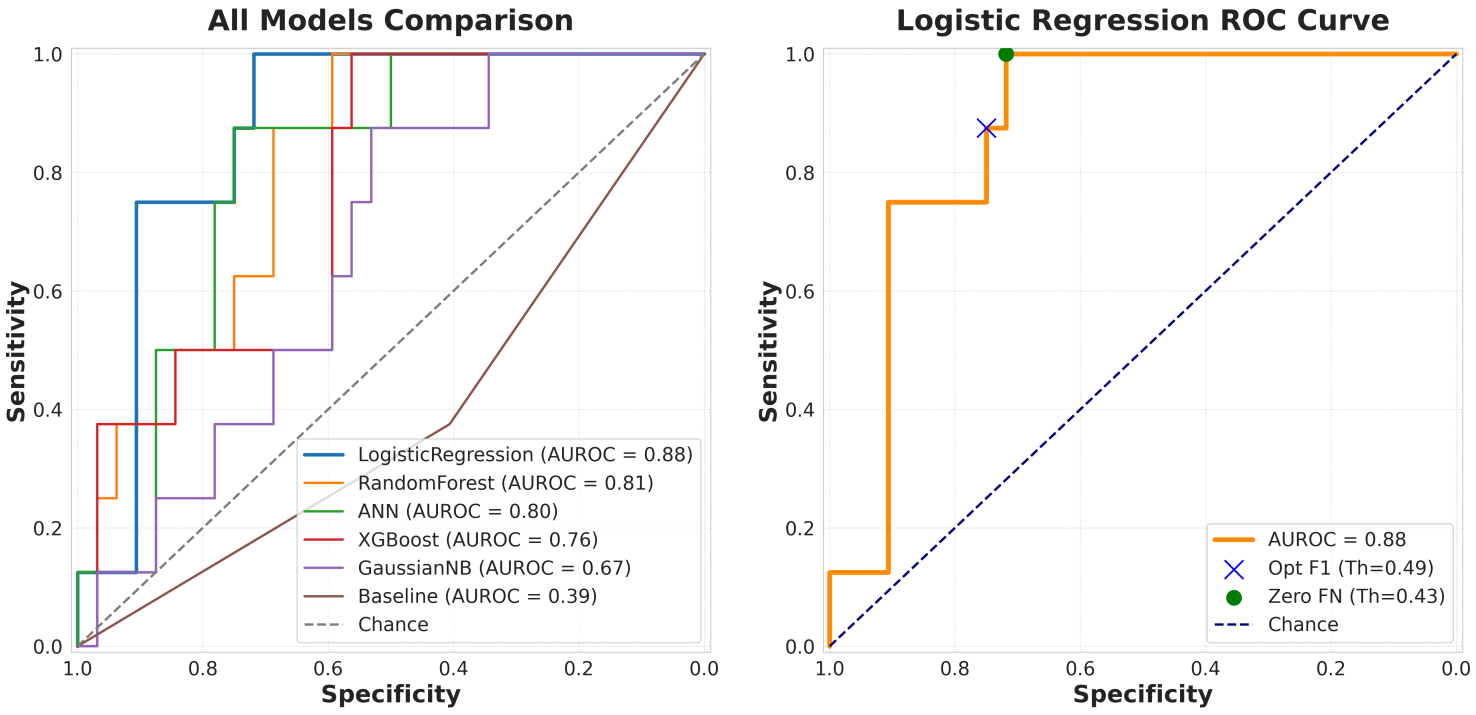


Figure 1

Displaying AUC of all ML models and LR Threshold Options

The AUC for all the models on the left is based on choosing the threshold for the optimal F1 score. AUC (Area under the curve), ANN (artificial Neural Network, XGBoost (extreme gradient boosting), GaussianNB (Gaussian Naïve Bayes).