Cardiology in the Young

cambridge.org/cty

Prediction of extubation failure in the paediatric cardiac ICU using machine learning and high-frequency physiologic data

Sydney R. Rooney[1,2], Evan L. Reynolds[3], Mousumi Banerjee[3,4], Sara K. Pasquali[2,4], John R. Charpie[2], Michael G. Gaies[2,4] and Gabe E. Owens[2]

[1]Department of Pediatrics, Children's Hospital of Pittsburgh, Pittsburgh, PA, USA; [2]Department of Pediatrics, Division of Cardiology, C.S. Mott Children's Hospital, University of Michigan, Ann Arbor, MI, USA; [3]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, USA and [4]Center for Healthcare Outcomes and Policy, University of Michigan, Ann Arbor, MI, USA

Original Article

Cite this article: Rooney SR, Reynolds EL, Banerjee M, Pasquali SK, Charpie JR, Gaies MG, and Owens GE (2022) Prediction of extubation failure in the paediatric cardiac ICU using machine learning and high-frequency physiologic data. *Cardiology in the Young* **32**: 1649–1656. doi: 10.1017/S1047951121004959

Received: 3 April 2021
Revised: 3 November 2021
Accepted: 21 November 2021
First published online: 20 December 2021

Keywords:
Airway extubation; artificial respiration; quality of healthcare; clinical decision-making; machine learning

Author for correspondence:
G. Owens, MD, PhD, C.S. Mott Children's Hospital, 1540 E Hospital Drive, Ann Arbor, MI 48109, USA. Tel: (734) 936-8997;
Fax: 734-936-9470.
E-mail: gabeo@med.umich.edu

Abstract

Background: Cardiac intensivists frequently assess patient readiness to wean off mechanical ventilation with an extubation readiness trial despite it being no more effective than clinician judgement alone. We evaluated the utility of high-frequency physiologic data and machine learning for improving the prediction of extubation failure in children with cardiovascular disease. *Methods:* This was a retrospective analysis of clinical registry data and streamed physiologic extubation readiness trial data from one paediatric cardiac ICU (12/2016-3/2018). We analysed patients' final extubation readiness trial. Machine learning methods (classification and regression tree, Boosting, Random Forest) were performed using clinical/demographic data, physiologic data, and both datasets. Extubation failure was defined as reintubation within 48 hrs. Classifier performance was assessed on prediction accuracy and area under the receiver operating characteristic curve. *Results:* Of 178 episodes, 11.2% (N = 20) failed extubation. Using clinical/demographic data, our machine learning methods identified variables such as age, weight, height, and ventilation duration as being important in predicting extubation failure. Best classifier performance with this data was Boosting (prediction accuracy: 0.88; area under the receiver operating characteristic curve: 0.74). Using physiologic data, our machine learning methods found oxygen saturation extremes and descriptors of dynamic compliance, central venous pressure, and heart/respiratory rate to be of importance. The best classifier in this setting was Random Forest (prediction accuracy: 0.89; area under the receiver operating characteristic curve: 0.75). Combining both datasets produced classifiers highlighting the importance of physiologic variables in determining extubation failure, though predictive performance was not improved. *Conclusion:* Physiologic variables not routinely scrutinised during extubation readiness trials were identified as potential extubation failure predictors. Larger analyses are necessary to investigate whether these markers can improve clinical decision-making.

Mechanical ventilation is a routine therapy in the paediatric cardiac intensive care unit. Separation from mechanical ventilation represents a vulnerable transition for patients with critical cardiovascular disease. Extubation can exacerbate underlying cardiac dysfunction and worsen other circulatory pathophysiology, potentially necessitating reintubation (extubation failure) and leading to poor outcomes. Children experiencing extubation failure are more likely to have prolonged duration of mechanical ventilation and length of stay, increased likelihood for tracheostomy, higher mortality rates, and higher hospital costs.[1–4] Thus, determining new approaches to predict extubation readiness and avoid extubation failure may help improve clinical outcomes and resource utilisation.

Efforts to improve medical decision-making around the time of extubation have largely been unsuccessful to date. Extubation readiness trials are commonly used in paediatric intensive care, but previous work demonstrates that standard extubation readiness trials fail to enhance clinician judgement.[5] Evidence to support any given method in paediatric intensive care patients remains limited.[5–8] The typical variables assessed during an extubation readiness trial – oxygen saturation, respiratory rate, heart rate, and tidal volume[5,7,9] – can vary greatly both at baseline and in response to stress amongst patients with structural heart disease. An increasing number of paediatric cardiac intensive care units have access to high-frequency clinical data systems, which can lead to more intricate characterisation of these variables. However, subtle changes in physiologic variables may not be easily identifiable to clinicians in real time. Tree-based machine learning methods offer an intuitive framework for classification and are adept at making predictions using complex data.[10] Capitalising on these methods and the availability of high-frequency clinical data may help to determine physiologic patterns that predict extubation failure across patients with varying cardiac lesions.

© The Author(s), 2021. Published by Cambridge University Press.

In this context, we applied tree-based machine learning methods to develop and evaluate a prediction tool for extubation failure in the cardiac intensive care unit. To accomplish this analysis, we linked our local Pediatric Cardiac Critical Care Consortium (PC^4) clinical registry data with continuously captured physiologic data from the T3 software platform (Etiometry, Inc., Boston, MA). We hypothesised that we could identify physiologic variables during extubation readiness trials associated with extubation failure in cardiac intensive care unit patients and develop a tool that could be implemented in the cardiac intensive care unit to improve assessment of extubation readiness.

Materials and methods

Data sources

The Pediatric Cardiac Critical Care Consortium clinical registry is a voluntary quality improvement collaborative that collects data on all patients with primary cardiac disease admitted to the cardiac intensive care unit attending service of participating hospitals.[11] Participating hospitals are audited on a regular schedule and audit results suggest complete, accurate and timely submission of data across hospitals, with the most recent published results demonstrating a major discrepancy rate of 0.6% across 29,476 fields.[12] Patient demographics, comorbidities, surgical procedures/interventions, critical care therapies, and complications are collected in the database. We abstracted clinical information about patients from our institution only for our analyses.

T3 (*Etiometry, Inc.; Boston, MA)* is a commercial software application that allows for the continuous collection and visualisation of physiologic data from patients in an intensive care setting.[13] Data at our institution are streamed at 1-minute intervals from physiologic monitors, ventilator data, and other inputs, providing a higher frequency dataset compared to the electronic medical record. The T3 data at the University of Michigan were cleaned through joint efforts of our analysis team and *Etiometry, Inc.* All physiologic data were filtered for the removal of implausible physiologic data. T3 ventilator data were compared to the clinical registry data to make sure the data were clinically aligned. Any runs that were tagged as incongruent between T3 and Pediatric Cardiac Critical Care Consortium by the programme we developed to compare them were examined further and brought to the attention of *Etiometry, Inc.* Multiple updates of the data warehouse took place in order to obtain the most accurate data possible.

The University of Michigan Medical School Institutional Review Board (IRBMED) provided oversight for this study, which was approved with waiver of informed consent (Approval Number: HUM00153234).

Inclusion and exclusion criteria

The study population included patients admitted to the paediatric cardiac intensive care unit at the University of Michigan between 12/2016 and 3/2018 for both medical and surgical indications. The episode of analysis was a course of mechanical ventilation. We retrospectively analysed patients who had at least one ventilation episode ≥12 hours and an extubation readiness trial (minimum: 30 minutes; maximum: 3 hours) within 14 hours of extubation. Timing of the extubation is at the discretion of the clinical team after utilising all clinical inputs (including vent setting, sedative medication load, amount of endotracheal tube leak). We used 12 hours as the minimum episode length as literature has shown that risk of extubation failure increases with length of ventilation,[14] so those extubated shortly after intubation likely have lower extubation failure rates and represent a simpler clinical decision for physicians. Only the first qualifying mechanical ventilation episode that ended in a planned extubation during each hospitalisation of our qualifying patients was analysed. Patients were excluded from the analysis if they were ventilated via tracheostomy or extubated due to withdrawal of care.

Outcome and predictor variables

The primary outcome of this analysis was extubation failure. Extubation failure was defined as reintubation within 48 hours of a planned extubation.[11] We utilised local Pediatric Cardiac Critical Care Consortium registry data for patient and clinical outcome data. We included admission type (medical vs. surgical) and timing of ventilation relative to surgery (pre- vs. post-operative) in the analysis. History of a paralysed diaphragm, vocal cord dysfunction, extracorporeal membrane oxygenation, and number of cardiac ventricles was recorded. Mechanical ventilation-related variables included the length of ventilation prior to extubation, intercurrent surgical procedure during the episode, and the number of previous extubations the patient had during that cardiac intensive care unit encounter. Other clinical and demographic variables were included that were readily available in the database and thought to be clinically applicable.

Though T3 collects data over the entirety of each patient's cardiac intensive care unit hospitalisation, we focused our analysis on physiologic data during the last extubation readiness trial prior to extubation of each patient's qualifying run of mechanical ventilation. At our institution, extubation readiness trials are performed in appropriate patients by turning off the inspiratory pressure and mandatory vent rate for patients on minimal positive end-expiratory pressure (5–6 cmH_2O) and pressure support (8–10 cmH_2O) in order to simulate breathing on one's own for a minimum of 30 minutes. Patients may not undergo extubation readiness trial testing despite being on minimal settings at the discretion of the physician if there are factors outside of their ventilator support that may increase the risk of extubation failure (i.e., other signs of clinical worsening) or if they require to stay intubated for other aspects of their medical care. Inability for a patient to maintain these settings while maintaining normal vital signs/oxygen saturation for their physiology indicates a failed trial and leads to early termination. SAS Version 9.4 (SAS Institute, Cary, NC) was used to create a programme that automatically scans the T3 files to identify extubation readiness trials of appropriate length based on ventilator settings described above. The programme compared each patient's extubation readiness trial times to their extubation time as listed in Pediatric Cardiac Critical Care Consortium, and the extubation readiness trial closest to extubation was evaluated. This programme was validated through repeated comparison of extubation readiness trial timing with the electronic medical record.

Physiologic variables were selected from the T3 database a priori, based on clinical importance. Variables chosen were either ones typically assessed in clinical practice or predicted to be relevant to extubation readiness. The physiologic variables studied were heart rate, systolic/diastolic/mean blood pressure, respiratory rate, oxygen saturation, central venous pressure, dynamic compliance, volume delivered by the ventilator (i.e., the measured volume observed during the extubation readiness trial with pressure-supported spontaneous ventilation), fraction of inspired oxygen, oxygen saturation/fraction of inspired oxygen ratio, mean airway pressure, and positive end-expiratory pressure.

https://doi.org/10.1017/S1047951121004959 Published online by Cambridge University Press

To acknowledge that some of these physiologic variables vary across age groups, we analysed statistical summaries of these variables across the extubation readiness trial for each patient rather than solely the raw data itself: slope over the entire extubation readiness trial, average piece-wise slope, and the standard deviation of each variable. Specifically, in regard to slope, slope over the entire extubation readiness trial was calculated using ordinary least squares regression. Average piece-wise slope was calculated as the mean slope between each pair of consecutive measurements during the extubation readiness trial. For variables that did not depend on age, we also examined the mean, minimum, and maximum values in addition to the summary statistics used for their age-dependent counterparts.

Machine learning approach

We utilised three machine learning methods to predict extubation failure.[10] As an initial step, we performed classification and regression tree analyses on our data, which creates a single decision tree where each decision point/node is a split in the predictor variable that the algorithm determines will maximise within-node homogeneity of the resulting "daughter nodes."[15] We expanded on our analysis by also performing tree-based ensemble analyses, namely Boosting and Random Forest.[16–18] While classification and regression tree analyses tend to be the most clinically intuitive due to their output of a decision tree, these ensemble methods tend to create more robust classifiers as they use many trees (in this case, 1000 trees based on 1000 bootstrapped samples) to make predictions. Their output instead is a variable importance plot that gives weight to the different predictor variables based on many different trees. Our variable importance plots display the twenty highest-weighted variables.

We implemented tree-growing control by only considering splits that resulted in daughter nodes of at least five patients. In addition, for the classification and regression tree and Boosting algorithms, we only allowed nodes with 15 or more patients to be split and only considered splits that reduced node impurity by a factor of 0.01 (using the "cp" parameter in the rpart package). Classification and regression tree and Boosting used surrogate splits, and Random Rorest used a fast imputation approach to handle missing covariate information.

To enhance the training dataset for the classifiers both in terms of total sample size and number of events, we trained our algorithms using a bootstrapped sample (with replacement) from the original population. This final sample was 4× the size of the original study population.

We employed an iterative approach by first building classifiers based just on clinical data (Pediatric Cardiac Critical Care Consortium database). Next, we developed classifiers using solely the physiologic summary measures (T3 database). Lastly, we utilised machine learning to build classifiers based on the combined clinical and physiologic information (Pediatric Cardiac Critical Care Consortium + T3 database). In each of the above scenarios, classifiers were built using classification and regression tree, Random Forest, and Boosting.

Prediction performance was assessed through a 10-fold cross-validation stratified by extubation failure status. Metrics used to assess prediction performance included area under the receiver operating characteristic curve, prediction accuracy (prediction accuracy: percentage of the time the classifier correctly predicts a patient's extubation outcome), positive predictive value, and negative predictive value.

All analyses were performed using R version 4.1.1 using the rpart, randomForestSRC, adabag, mlr, and ggplot2 packages.[19–22]

Results

Population characteristics

There were 178 ventilator episodes from our institution meeting inclusion criteria for this study. The overall extubation failure rate for the cohort was 11.2% (N = 20 failures). Demographic and clinical characteristics of the patients are described in Table 1.

Machine learning outputs

Performance and prediction accuracy of all machine learning methods are summarised in Table 2. Classifier area under the receiver operating characteristic curve values ranged from 0.40 to 0.78, with Random Forest using both clinical and demographic data having the highest area under the receiver operating characteristic curve (area under the receiver operating characteristic curve: 0.78). Prediction accuracies were higher for ensemble methods (range: 0.88–0.89) than using classification and regression tree for all three data groups.

The classification tree based on the Pediatric Cardiac Critical Care Consortium-derived clinical and demographic information is displayed in Figure 1. This classifier split patients based on age, weight, length of admission, length of episode of mechanical ventilation, previous extubations in that encounter, and presence of pulmonary hypertension or a paralysed diaphragm. Applying the Random Forest algorithm to Pediatric Cardiac Critical Care Consortium data identified duration of mechanical ventilation, patient length and weight, age at time of cardiac intensive care unit admission, and number of ventricles as the five variables of greatest predictive value. Boosting identified similar Pediatric Cardiac Critical Care Consortium variables as being of highest importance, with the inclusion of race instead of number of ventricles.

The classification tree based on only the T3 physiologic data split patients on a variety of physiologic variables, with prominent inclusion of variables related to respiratory rate, central venous pressure, and oxygen saturation. When applying Random Forest methodology to T3 data, the variables of greatest importance in predicting extubation readiness were oxygen saturation mean, minimum, and maximum, in addition to standard deviation and maximum of dynamic compliance. When applying Boosting to the physiologic T3 data, the oxygen saturation mean, slopes of heart and respiratory rate, central venous pressure minimum, and age were identified as most important.

Applying classification and regression tree to both datasets (Pediatric Cardiac Critical Care Consortium and T3) generated a classifier with similar physiologic variables compared to the T3 classifier (Fig 2). The only clinical/demographic variables represented were patient weight and the duration of ventilation. Prediction accuracy metrics based on the 10-fold cross-validation are displayed in Table 2. When using Random Forest on the combined dataset, which had the highest area under the receiver operating characteristic curve, the variable importance plot (Fig 3) illustrates physiologic variables rather than demographic variables being the majority of the twenty highest-weighted variables in determining extubation readiness. Physiologic variables had overall higher importance compared to clinical and demographic data as well when using Boosting. However, overall predictive accuracy was similar when comparing the physiologic or combined dataset to clinical/demographic data alone.

https://doi.org/10.1017/S1047951121004959 Published online by Cambridge University Press

Table 1. Study population demographics and characteristics

Characteristic	Full cohort (n = 178)	No EF (n = 158)	EF (n = 20)
Age at cardiac intensive care unit admission, N (%)			
Preterm neonate (<30 days)	10 (5.6%)	10 (6.3%)	0 (0.0%)
Neonate (<30 days)	60 (33.7%)	49 (31.0%)	11 (55.0%)
Infant (30 days to 1 year)	73 (41.0%)	68 (43.0%)	5 (25.0%)
Child (1–18 years)	27 (15.2%)	23 (14.6%)	4 (20.0%)
Adult (>18 years)	8 (4.5%)	8 (5.1%)	0 (0.0%)
Sex, male, N (%)	102 (57.3%)	89 (56.3%)	13 (65.0%)
Race, N (%)			
Caucasian	111 (67.3%)	101 (67.8%)	10 (62.5%)
lack	24 (14.6%)	20 (13.4%)	4 (25.0%)
Asian	4 (2.4%)	3 (2.0%)	1 (6.3%)
Native American	2 (1.2%)	2 (1.3%)	0 (0.0%)
Native Pacific Islander	2 (1.2%)	2 (1.3%)	0 (0.0%)
Other/multiracial	22 (13.3%)	21 (14.1%)	1 (6.3%)
Missing	13	9	4
Ethnicity, N (%)			
Hispanic or Latino	11 (6.3%)	11 (7.1%)	0 (0.0%)
Missing	3	3	0
Patient category, N (%)			
1: Medical patient	23 (12.9%)	18 (11.4%)	5 (25.0%)
2: Pre-operative patient, STAT 1–3	3 (1.7%)	2 (1.3%)	1 (5.0%)
3: Pre-operative patient, STAT 4–5	7 (3.9%)	7 (4.4%)	0 (0.0%)
4: Post-operative patient, STAT 1–3	65 (36.5%)	60 (38.0%)	5 (25.0%)
5: Post-operative patient, STAT 4–5	78 (43.8%)	69 (43.7%)	9 (45.0%)
Unassigned	2 (1.1%)	2 (1.3%)	0 (0.0%)
Single versus double ventricle patient, N (%)			
1	61 (34.3%)	48 (30.4%)	13 (65.0%)
2	117 (65.7%)	110 (69.6%)	7 (35.0%)
Number of previous extubation in the cardiac intensive care unit encounter, N (%)			
0	165 (92.7%)	148 (93.7%)	17 (85.0%)
1	13 (7.3%)	10 (6.3%)	3 (15.0%)
Duration of mechanical ventilation (minutes), mean (sd)	6740.0 (6500.8)	6504.8 (6302.2)	8598.2(7836.3)
Airway anomaly, N (%)	14 (7.9%)	14 (8.9%)	0 (0.0%)
Non-airway anomaly, N (%)	30 (16.9%)	27 (17.1%)	3 (15.0%)
Chromosomal abnormality or syndrome, N (%)	41 (23.0%)	37 (23.4%)	4 (20.0%)
Paralysed diaphragm during the hospitalisation, N (%)	7 (3.9%)	4 (2.5%)	3 (15.0%)
Vocal cord dysfunction during the hospitalisation, N (%)	5 (2.8%)	5 (3.2%)	0 (0.0%)
ECMO*, N (%)	12 (6.7%)	10 (6.3%)	2 (10.0%)
Cardiac arrest*, N (%)	11 (6.2%)	11 (7.0%)	0 (0.0%)
Acute renal failure*, N (%)	1 (0.1%)	1 (0.1%)	0 (0.0%)
Acute decompensated heart failure*, N (%)	18 (10.1%)	15 (9.5%)	3 (15.0%)
Sepsis*, N (%)	4 (2.3%)	4 (2.5%)	0 (0.0%)

(*Continued*)

https://doi.org/10.1017/S1047951121004959 Published online by Cambridge University Press

Table 1. (*Continued*)

Characteristic	Full cohort (n = 178)	No EF (n = 158)	EF (n = 20)
Stroke*, N (%)	4 (2.3%)	4 (2.5%)	0 (0.0%)
Pulmonary hypertension*, N (%)	24 (13.5%)	22 (13.9%)	2 (10.0%)

*Before extubation in the same cardiac intensive care unit encounter as extubation

Table 2. Machine learning algorithm results using clinical/demographic data, physiologic data, and both data sources

Method	Accuracy	Negative predictive value	Positive predictive value	Area under the receiver operating characteristic curve
Clinical/demographic data				
Classification and regression tree:Pediatric Cardiac Critical Care Consortium	0.83	0.89	0.25	0.54
Random Forest: Pediatric Cardiac Critical Care Consortium	0.89	0.89	0.20	0.71
Boosting: Pediatric Cardiac Critical Care Consortium	0.88	0.90	0.45	0.74
Physiologic data				
Classification and regression tree: T3	0.83	0.90	0.18	0.40
Random Forest: T3	0.89	0.89	0.30	0.75
Boosting: T3	0.88	0.89	0.10	0.69
Combined data				
Classification and regression tree: Pediatric Cardiac Critical Care Consortium+T3	0.84	0.90	0.15	0.54
Random Forest: Pediatric Cardiac Critical Care Consortium+T3	0.89	0.89	0.10	0.78
Boosting: Pediatric Cardiac Critical Care Consortium+T3	0.89	0.89	0.20	0.71

Discussion

We applied machine learning methods to clinical registry and streamed physiologic data to build a classifier for prediction of extubation failure at the time of extubation readiness testing with the aim to identify extubations that are unlikely to be successful. Our results indicate that in patients with critical cardiovascular disease, unique, high-frequency physiologic variables might aid in the prediction of extubation failure. We demonstrated the potential importance of assessing physiologic variables that are not routinely analysed during standard extubation readiness trials as well as how computational methodologies of extracting information on routine vital signs may lead to increased discriminative capabilities. However, our most accurate classifier had an accuracy of 89%, and our population's extubation failure rate was 11%, indicating a similar degree of misidentification of ready subjects. While we believe that these methods show promise to improve clinical care, classifier performance must improve before being applied as a clinical tool.

Machine learning methods have been used to create prediction and decision-making tools in various clinical domains.[23–25] Examples include the prediction of periventricular leukomalacia after neonatal cardiac surgery[26,27] and the prediction of clinical deterioration in those with haematologic malignancies.[24] Additionally, analysis of granular physiologic data with machine learning methods can reveal physiologic changes that are not visible to clinicians. For instance, the HeRO score derived from high fidelity electrocardiogram data has helped clinicians identify neonatal sepsis and reduce mortality rates by 20%.[28] Uncovering these silent patterns illustrates the promise of these approaches.

Despite the evidence of machine learning's utility, there is currently a paucity of research to unlock the value of high-frequency physiologic data collected every day in paediatric cardiac intensive care units. The variable importance plots from our work demonstrate that physiologic variables might have more importance than clinical and demographic data, suggesting the value of analysing these variables with advanced statistical methodologies. This is particularly compelling as while we consider our T3 data as high frequency compared to electronic medical record documentation as it is streamed at 1-minute intervals, there are many platforms that collect even higher fidelity physiologic data. The physiologic complexity of paediatric cardiac patients and the wealth of available data from various monitoring techniques may make this patient population particularly suited to apply machine learning methods in an effort to develop predictive algorithms that improve patient care.[29]

Our work highlighted the importance of identifying unique features with biologic plausibility not robustly interrogated during traditional extubation readiness trials. While features such as central venous pressure, standard deviation of blood pressure, and extremes of oxygen saturation may be recorded or derivable from recorded measurements during an extubation readiness trial, most parameters to determine a "pass" or "fail" are related to respiratory mechanics and function. While these vital signs are often available during an extubation readiness trial, trend data and measures of variability may not be recognised. An advantage to the algorithms

https://doi.org/10.1017/S1047951121004959 Published online by Cambridge University Press

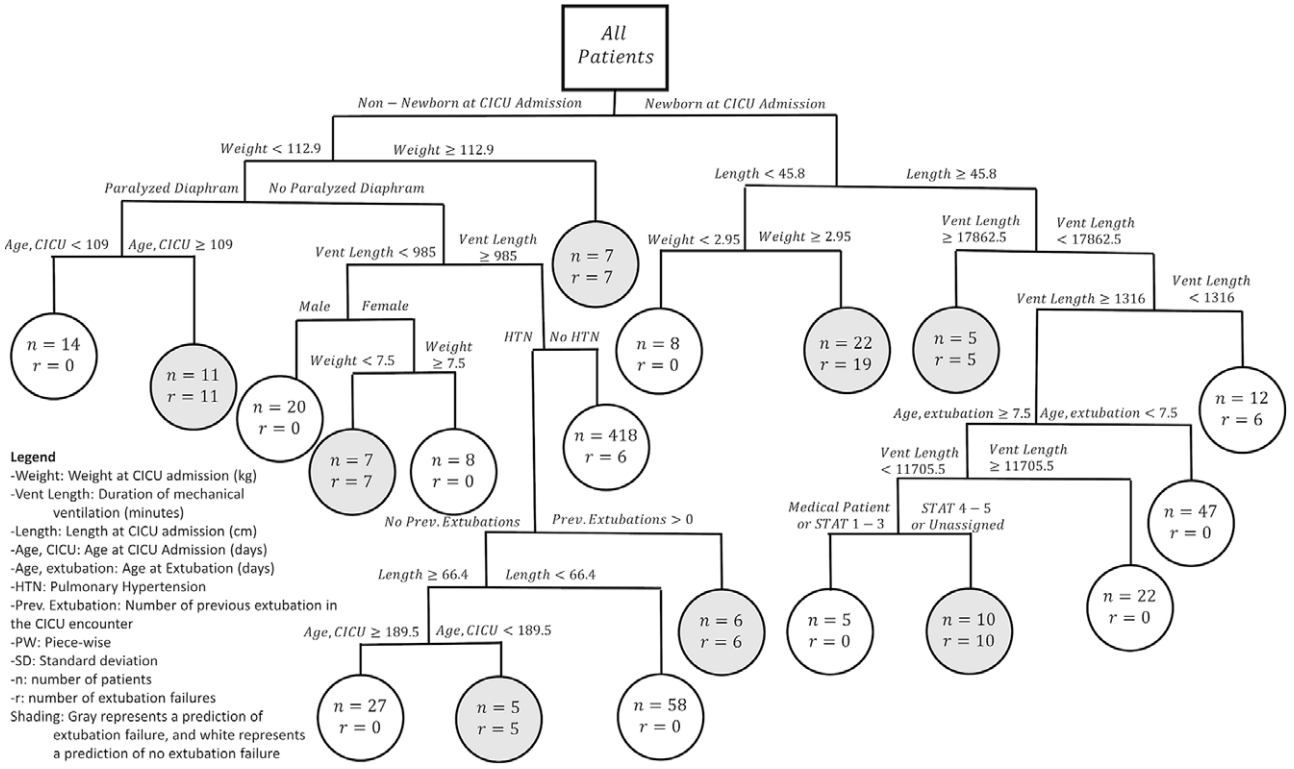


Figure 1. Classification and regression tree (CART) result using clinical/demographic data.

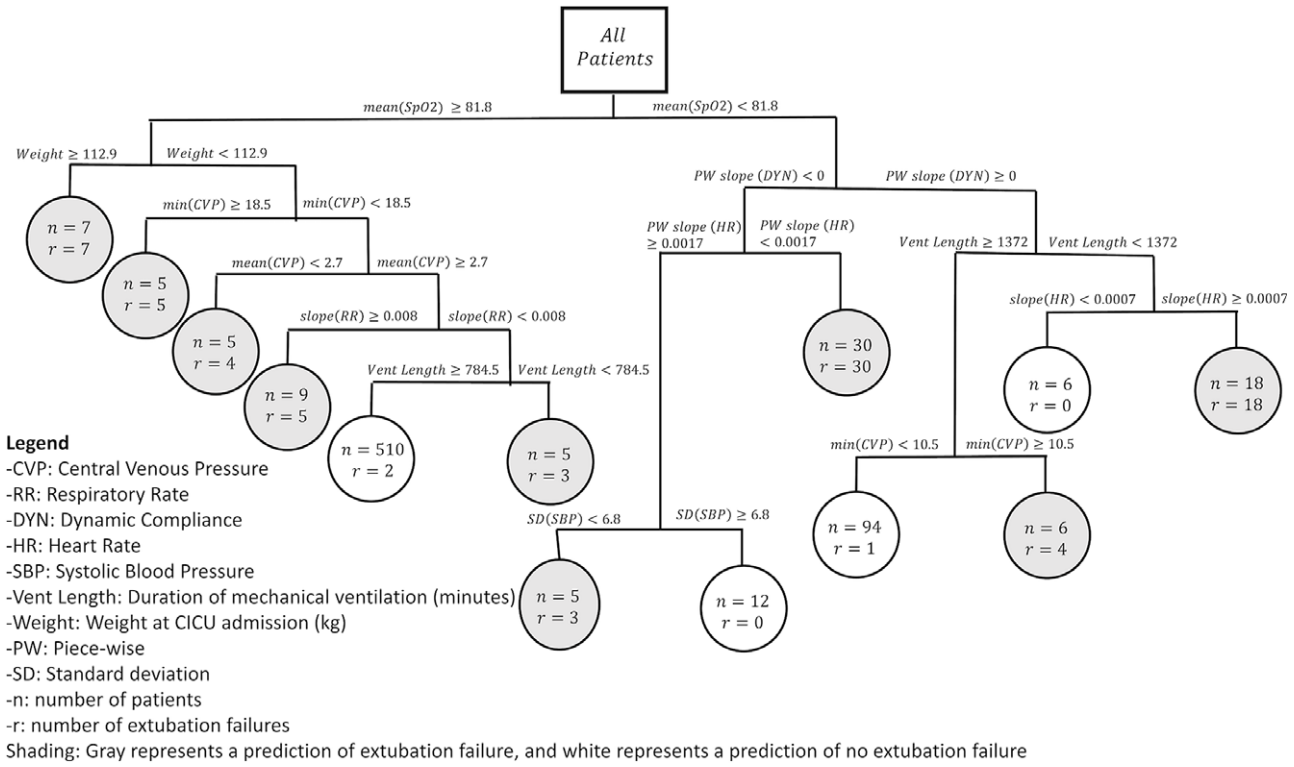


Figure 2. Classification and regression tree (CART) result using both physiologic and clinical/demographic data.

we created is that they not only include a larger number of variables and trends than clinicians can feasibly integrate in real-time for decision-making, but they also innately create cut points at different nodes that maximise clinical utility, which is important as we do not have literature on how many of these variable trends over an extubation readiness trial correlate to extubation readiness.

https://doi.org/10.1017/S1047951121004959 Published online by Cambridge University Press

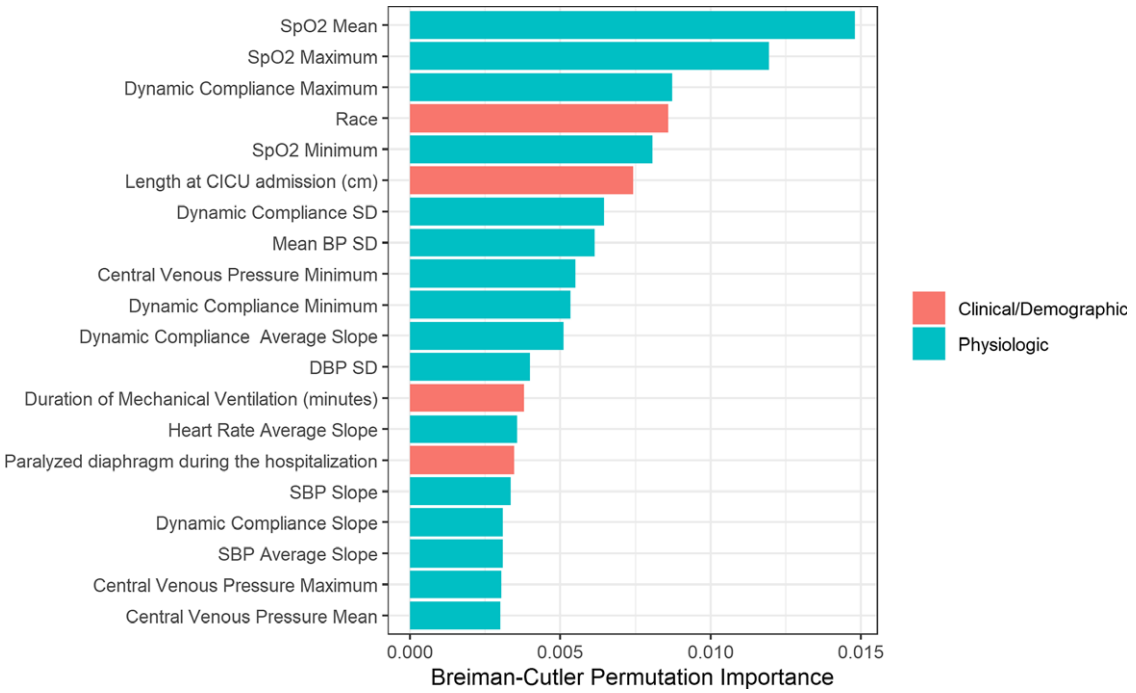


Figure 3. Random Forest variable importance plot using both physiologic and clinical/demographic data.

Furthermore, T3 has a graphical user interface for clinicians that can display real-time data and algorithm results. For example, the University of Michigan cardiac intensive care unit T3 graphical user interface currently displays an inadequate oxygen delivery index that is derived from streamed physiologic data. This interface increases the feasibility of a real-time extubation readiness algorithm. Optimising the number of physiologic clues clinicians can utilise in the decision-making process could potentially lead to more accurate predictions and improve clinical care.

There are important limitations to this analysis. Focusing the study on our institution limited the sample size and number of events. We addressed this by enhancing our learning set through a bootstrapped sample of the original population as well as employing training and a 10-fold cross-validation. Despite this, our classifiers' performance was below that necessary to support real-time clinical decision-making. Currently, it is logistically challenging to use the data from the T3 repository across institutions; however, we aim to use T3 data from multiple cardiac intensive care units in the future, allowing for larger, more heterogeneous populations that would improve the accuracy and applicability of these approaches.

Another limitation is that while physiologic variables were measured at a frequency of once per minute, they underwent conversion into summary statistics in order to be included in our analysis. While our techniques focused on optimising clinical interpretability, reducing the physiologic data to summary measures inherently creates limitations. It is possible that our methods resulted in loss of important clinical information and negatively affected our classifier performance. Also, we aimed for these summary statistics to allow us to compare physiologic variables across age groups that naturally have different baselines. However, we recognise this is an imperfect means of age adjustment as certain variables are known to also have different variability across age ranges.

When analysing potential predictors of extubation failure, we were limited to candidate variables that were either in the Pediatric Cardiac Critical Care Consortium or T3 clinical registry. Though comprehensive, there may be other significant predictors of extubation failure that were not included in our analyses. For instance, T3 as well as certain other high fidelity software systems in cardiac intensive care units currently are able to access NIRS data, which is a variable that has significant potential to predict outcomes but was not available in our dataset.

In summary, we applied machine learning methods in an attempt to generate information that could assist clinicians in the cardiac intensive care unit when assessing patients for extubation readiness. Several physiologic variables that are not routinely scrutinised during standard extubation readiness trials were identified as potential predictors of extubation failure. Our experience suggests that the application of machine learning to the plethora of data collected in the cardiac intensive care unit holds promise for predicting extubation failure, but it is subject to the same limitations of sample size as traditional statistical approaches. Larger patient cohorts might yield analyses to inform a clinically applicable tool that improves medical decision-making for mechanically ventilated patients.

Acknowledgements. We acknowledge the data collection team for Pediatric Cardiac Critical Care Consortium at the University of Michigan. Sydney Rooney at the time of the study was a research fellow supported by the Sarnoff Cardiovascular Research Foundation.

Financial support. This study was supported in part by funding from the University of Michigan Congenital Heart Center, CHAMPS for Mott, and the Michigan Institute for Clinical & Health Research (NIH/NCATS, grant number: UL1TR002240). Sara Pasquali receives funding from the National Heart, Lung, and Blood Institute (grant number: R01HL12226; PI Pasquali). Dr. Banerjee's research was partially funded by a grant BCBS 002983 from the Blue Cross Blue Shield Foundation of Michigan.

https://doi.org/10.1017/S1047951121004959 Published online by Cambridge University Press

Conflicts of interest. None.

References

1. Farias JA, Retta A, Alía I, et al. A comparison of two methods to perform a breathing trial before extubation in pediatric intensive care patients. Intensive Care Med 2001; 27: 1649–1654.
2. Kurachek SC, Newth CJ, Quasney MW, et al. Extubation failure in pediatric intensive care: a multiple-center study of risk factors and outcomes. Crit Care Med 2003; 31: 2657–2664.
3. Rothaar RC, Epstein SK. Extubation failure: magnitude of the problem, impact on outcomes, and prevention. Curr Opin Crit Care 2003; 9: 59–66.
4. Epstein SK, Ciubotaru RL, Wong JB. Effect of failed extubation on the outcome of mechanical ventilation. Chest 1997; 112: 186–192.
5. Randolph AG, Wypij D, Venkataraman ST, et al. Effect of mechanical ventilator weaning protocols on respiratory outcomes in infants and children: a randomized controlled trial. JAMA 2002; 288: 2561–2568.
6. Ferguson LP, Walsh BK, Munhall D, et al. A spontaneous breathing trial with pressure support overestimates readiness for extubation in children. Pediatr Crit Care Med 2011, 12, , e330–335.
7. Newth CJL, Venkataraman S, Willson DF, et al. Weaning and extubation readiness in pediatric patients. Pediatr Crit Care Med 2009; 10: 1–11.
8. Ferreira FV, Sugo EK, Aragon DC, et al. Spontaneous breathing trial for prediction of extubation success in pediatric patients following congenital heart surgery: a randomized, controlled trial. Pediatr Crit Care Med 2019; 20: 940–946.
9. Abu-Sultaneh S, Hole AJ, Tori AJ, et al. An interprofessional quality improvement initiative to standardize pediatric extubation readiness assessment. Pediatr Crit Care Med 2017, 18, , e463–e471.
10. Banerjee M, Reynolds E, Andersson HB, et al. Tree-based analysis. Circ Cardiovasc Qual Outcomes 2019; 12: e004879.
11. Gaies M, Cooper DS, Tabbutt S, et al. Collaborative quality improvement in the cardiac intensive care unit: development of the Paediatric Cardiac Critical Care Consortium (PC4). Cardiol Young 2015; 25: 951–957.
12. Gaies M, Donohue JE, Willis GM, et al. Data integrity of the Pediatric Cardiac Critical Care Consortium (PC4) clinical registry. Cardiol Young 2016; 26: 1090–1096.
13. T3 Data Aggregation & Visualization [Internet]. Etiometry. Retrieved October 29, 2018, from http://www.etiometry.com/our-platform/t3-visualization/
14. Gaies M, Tabbutt S, Schwartz SM, et al. Clinical epidemiology of extubation failure in the pediatric cardiac ICU: a report from the pediatric cardiac critical care consortium. Pediatr Crit Care Med 2015; 16: 837–845.
15. Breiman L, Friedman J, Olshen R, et al. Classification and Regression Trees. Wadsworth, Belmont, CA, 1984.
16. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, 2001.
17. Breiman L. Random forests. Mach Learn 2001; 45: 5–32.
18. Freund Y, Schapire R. Experiments with a new boosting algorithm. In Proceedings of the Thirteenth International Conference on Machine Learning. Morgan Kaufmann, San Francisco, CA, 1996: 148–156.
19. Ishwaran H, Kogalur U. Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC) [Internet], 2020. https://cran.r-project.org/package=randomForestSRC
20. Therneau T, Atkinson B, Ripley B. Rpart: Recursive Partitioning and Regression Trees. R Package Version 4.1-3, 2013. http://CRAN.R-project.org/package=rpart
21. Bischl B, Lang M, Schiffner J, et al. mlr: machine learning in R. J Mach Learn Res 2016; 17: 1–5.
22. Alfaro E, Gamez M, Garcia N. adabag: an R package for classification with boosting and bagging. J Stat Softw 2013, 54:1–35.
23. Norrie J. The challenge of implementing AI models in the ICU. Lancet Respir Med 2018; 6: 886–888.
24. Hu SB, Wong DJL, Correa A, et al. Prediction of clinical deterioration in hospitalized adult patients with hematologic malignancies using a neural network model. PloS One 2016; 11: e0161401.
25. Jalali A, Bender D, Rehman M, et al. Advanced analytics for outcome prediction in intensive care units. Annu Int Conf IEEE Eng Med Biol Soc 2016; 2016: 2520–2524.
26. Jalali A, Licht DJ, Nataraj C. Application of decision tree in the prediction of periventricular leukomalacia (PVL) occurrence in neonates after heart surgery. Annu Int Conf IEEE Eng Med Biol Soc 2012; 2012: 5931–5934.
27. Goodman KE, Lessler J, Cosgrove SE, et al. A clinical decision tree to predict whether a bacteremic patient is infected with an extended-spectrum β-lactamase-producing organism, 2016; 63: 896–903.
28. Fairchild KD. Predictive monitoring for early detection of sepsis in neonatal ICU patients. Curr Opin Pediatr 2013; 25: 172–179.
29. Olive MK, Owens GE. Current monitoring and innovative predictive modeling to improve care in the pediatric cardiac intensive care unit. Transl Pediatr 2018; 7: 120–128.

https://doi.org/10.1017/S1047951121004959 Published online by Cambridge University Press