34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Prediction of Extubation Readiness in Extreme Preterm Infants Based on Measures of Cardiorespiratory Variability*

Doina Precup, Carlos A. Robles-Rubio, *Student Member, IEEE*, Karen A. Brown, L. Kanbar, J. Kaczmarek, S. Chawla, G.M. Sant'Anna, and Robert E. Kearney, *Fellow, IEEE*

***Abstract*—The majority of extreme preterm infants require endotracheal intubation and mechanical ventilation (ETT-MV) during the first days of life to survive. Unfortunately this therapy is associated with adverse clinical outcomes and consequently, it is desirable to remove ETT-MV as quickly as possible. However, about 25% of extubated infants will fail and require re-intubation which is also associated with a 5-fold increase in mortality and a longer stay in the intensive care unit. Therefore, the ultimate goal is to determine the optimal time for extubation that will minimize the duration of MV and maximize the chances of success. This paper presents a new objective predictor to assist clinicians in making this decision. The predictor uses a modern machine learning method (Support Vector Machines) to determine the combination of measures of cardiorespiratory variability, computed automatically, that best predicts extubation readiness. Our results demonstrate that this predictor accurately classified infants who would fail extubation.**

I. INTRODUCTION

At birth, extreme preterm infants (≤28 weeks) have inconsistent respiratory drive, airway instability, surfactant deficiency, and immature lungs that frequently result in respiratory failure. Management of these infants is difficult and most of them will require endotracheal intubation and mechanical ventilation (ETT-MV) within the first days of life [1]. ETT-MV is an invasive therapy that is associated with adverse clinical outcomes including bronchopulmonary dysplasia, pneumonia, neurodevelopment problems, and increased mortality. In neonates prolonged invasive ventilation increases the incidence of neurodevelopment problems by a factor of 1.94 per 4 weeks of ventilation [1]. Consequently, clinicians try to remove ETT-MV as quickly as possible. However, as many as 25% of mechanically ventilated infants weighing <1250 g at birth need to be re-intubated following extubation[2]. Failure of extubation, defined as need for re-intubation within 72 hours after extubation, has been associated with higher death rate, increased length of hospital stay, and prolonged ventilation in both adult and pediatric populations [3]. Therefore physicians must determine the optimal timing for extubation which will minimize the duration of MV and maximize the chances of success. A variety of objective measures have been proposed to assist with this decision but none have proven to be clinically useful.

Recently, we explored the predictive power of indices of the variability of heart rate (HR) and respiratory (RV). In a retrospective study, we first showed that the combination of RV measurements with the results of a spontaneous breathing test (SBT), where ventilator support was temporarily removed by switching it to ETT continuous positive airway pressure (CPAP), had excellent positive predictive value (PPV=95%) and sensitivity (100%) [4]. In a subsequent prospective study of 56 preterm infants, heart rate variability (HRV) was computed from ECG and RV from the manual analysis of non-invasive respiratory measurements made using respiratory inductive plethysmography (RIP). Both HRV and RV were significantly lower in infants who failed extubation, providing a strong PPV and sensitivity but low specificity [5, 6]. A further analysis using more sophisticated, automated, algorithms to analyze the RIP data detected significant differences between indices of RV in infants who failed and those who were successfully extubated [7].

These results suggest that cardio-respiratory signals contain information that might be used to predict when extubation will be successful. However, the most effective way to extract this information remains to be determined. This study explores the utility of using machine learning methods to combine automatically determined features of HRV and RV to predict the success of extubation.

The paper is organized as follows: Section II describes the patient population and data acquisition. Section III describes the signal analysis methods used to define features. Section IV describes the machine learning methods used to develop

*Research supported in part by the Natural Sciences and Engineering Research Council of Canada. The work of C. A. Robles-Rubio was supported in part by the Mexican national Council for Science and Technology. GMS and JK were supported by the Research Institute of the Montreal Children's Hospital.

D. Precup is with the Department of Computer Science, McGill University, Montreal, Quebec, H3A 0E9, Canada (e-mail: dprecup@cs.mcgill.ca).

C. A. Robles-Rubio, L. Kanbar and R. E. Kearney are with the Department of Biomedical Engineering, McGill University, Montreal, Quebec, H3A 2B4, Canada (e-mail: carlos.roblesrubio@mail.mcgill.ca; robert.kearney@ mcgill.ca).

K. A. Brown is with the Department of Anesthesia, McGill University Health Center, Montreal, Quebec, H3A 2B4, Canada (e-mail: karen.brown@mcgill.ca).

J. Kaczmarek and G.M. Sant'anna are with the Department of Pediatrics, McGill University Health Center, Montreal, Quebec, H3A 2B4, Canada (e-mail: guilherme.santanna@mcgill.ca)

S. Chawla is with the Detroit Medical Center, Wayne State University, Detroit, MI, 48201, USA (e-mail: schawla@dmc.org).

the optimal predictor. Section V reports the performance results. Section VI provides concluding remarks.

II. Methods

This paper presents the results of a new analysis of data originally reported in [5, 6].

A. Study Population

All infants admitted to the Neonatal Intensive Care Unit (NICU) of the Royal Victoria Hospital (Montreal, Canada), Jewish General Hospital (Montreal, Canada), Montreal Children's Hospital (Montreal, Canada) and Detroit Medical Centre (Detroit, USA) with a birth weight ≤ 1250g and requiring MV were eligible for inclusion. The research ethics committee of each institution approved the study and written informed consent was obtained from parents.

The decision to extubate was made by the most responsible physician according to the following guidelines: 1) infants below 1000g were extubated with a mean airway pressure (MAP) ≤ 7 cmH_2O and $FiO_2 \leq 0.3$ and 2) infants ≥ 1000g were extubated with a MAP ≤ 8 cmH_2O and $FiO_2 \leq 0.3$. All infants were examined at the time of their first extubation from MV. Infants were excluded if they had any major congenital anomalies, congenital heart disease, cardiac arrhythmias, been administered vasopressor or sedative drugs at the time of extubation or were being extubated directly from high frequency oscillatory ventilation.

Post-extubation management involved the application of nasal CPAP or nasal intermittent positive pressure ventilation (NIPPV) using either bi-nasal prongs or a single nasopharyngeal tube. Infants were re-intubated if they met at least one of the following four criteria: 1) $FiO_2 > 0.5$ in order to maintain SpO_2 > 88% or partial pressure of oxygen (PaO_2) > 45 mmHg, 2) partial pressure of carbon dioxide ($PaCO_2$) > 55-60 mmHg with a pH < 7.25, 3) apnea requiring positive pressure ventilation with bag and mask or 4) significant evidence of increased respiratory distress including frequent retractions, grunting and chest wall distortion. Extubation failure was defined as the need for re-intubation within seventy-two hours of initial disconnection from ETT-MV.

The extubation and post-extubation management settings were not fixed; they were determined by each infant's attending physician. However, an analysis of the success and failure groups failed to detect any differences between the groups with respect to the mode of ventilation (SIMC or AC), ventilatory settings (PIP, PEEP, Rate, FiO_2), blood gases (SpO2, PCO_2, HCO_3 or BE) or the use of caffeine prior to extubation Nor, were there any differences in the mode of respiratory support provided after extubation

Data was acquired from 56 infants; 44 successfully extubated and 12 required re-intubation. Both groups had similar population characteristics, ventilator settings and blood gases prior to extubation.

B. Data Acquisition

Data were recorded during two phases immediately prior to extubation. Phase I comprised sixty minutes during which the infant was receiving ETT-MV. Phase II comprised a subsequent spontaneous breathing test (SBT) lasting for three minutes after the mode of ventilation was switched to ETT-CPAP. SBT trials in preterm infants must be kept short given the risks of atelectasis and lung collapse when the infant is breathing through a high resistance tube with no support. Previous studies have demonstrated a three minute interval to be safe. Patients remained supine throughout the recording.

ECG data was acquired from three leads placed on the infant's chest or limbs for heart rate detection and monitoring. Respiratory movements were measured with RIP (Respitrace, Vyasis, Yorba Linda,) using two respiband transducers. One respiband was placed around the infant's chest at the level of the nipple line to measure rib cage (RC) movements. The other respiband was placed around the infant's abdomen, half a centimeter above the umbilicus to measure abdominal (AB) movements.

C. Signal processing

The RIP and ECG signals were sampled at 1kHz but for analysis were decimated to 50Hz, the sampling rate for which our respiratory signal analysis algorithms were tuned and validated. The RIP and ECG were analyzed continuously starting at Phase I until the end of Phase II as follows.

D. Respiratory Signal Analysis

The respiratory pattern from the RIP signals was analyzed with AUREA [8], a novel system for Automated Unsupervised Respiratory Event Analysis that requires no human intervention. It characterizes respiratory activity in terms of a series of metrics that extract the amplitude, frequency and thoraco-abdominal asynchrony information from the RIP on a sample-by-sample basis as described in [8, 9]. These metrics provide reliable, quantitative measures of respiratory activity. They include:

Instantaneous respiratory frequency (f_{max}): defined as the frequency in the 0 – 2 Hz band with the most power. Values of f_{max} in the 0.4 – 2.0 Hz band correspond to the respiratory rate, while lower values are observed during movement artifacts. This is estimated by passing the RIP signals through a bank of band-pass filters with a bandwidth of 0.2 Hz; f_{max} is defined as the central frequency of the filter with the highest output power at each time. This yields an estimate accurate to within 0.1 Hz.

RMS metric (r^+): quantified the absolute amplitude information of the RIP signals; it is defined as sum of the root mean square values of RC and AB.

Movement Artifact metrics (m^{rc} and m^{ab}): compare the power in the movement artifact band (i.e., 0 – 0.4 Hz) to that in the regular breathing band. The metric for RC (m^{rc}) is calculated using the outputs of the filter bank described above as

$$m^{rc}[n] = \frac{\max_i \left\{\wp_i^{rc}[n]\right\}_{i\in I} - \max_i \left\{\wp_i^{rc}[n]\right\}_{i\in J}}{\max_i \left\{\wp_i^{rc}[n]\right\}_{i\in I} + \max_i \left\{\wp_i^{rc}[n]\right\}_{i\in J}},$$

where the sets I and J define the filter numbers that span the breathing and movement artifact bands respectively, $\wp_i^{rc}$ is the power of rc_i, the output of the i^{th} filter in either I or J, computed over a window of length N_M. The metric (m^{ab}) for AB is defined similarly. This metric is close to 1 during regular breathing and shifts towards -1 during movement artifacts.

Thoraco-Abdominal Asynchrony (ϕ) metric: estimates the phase between RC and AB using selectively filtered RIP signals to improve the signal-to-noise ratio. The filtered signals are converted to binary signals and the exclusive-OR (XOR) is computed between them at each sample. The asynchrony metric is the average of this binary signal over a window of length N_A, and is proportional to the phase shift

Pause metrics (p^{rc} and p^{ab}): provide measures of the RIP power in the regular breathing band (0–2.0 Hz [9]) relative to that expected for normal breathing. The pause metric for RC (p^{rc}) is defined by

$$p^{rc}[n] = \sqrt{\frac{\wp_B^{rc}[n]}{\underset{l \in [n-N_Q+1, n]}{median}\left\{\wp_B^{rc}[l]\right\}}}$$

where $\wp_B^{rc}$ is the power over a window of length $N_P << N_Q$ of the RC signal band-pass filtered on the regular breathing band, and N_Q is the length of the window used to estimate the median regular breathing power at each time. This metric is close to 1 during regular breathing and lower during pauses. The pause metric for AB (p^{ab}) is defined analogously

E. Cardiac Signal Analysis

The ECG signal was analyzed with the short-time Fourier transform (STFT) to obtain a time-varying representation of the frequency content. The instantaneous heart rate (h_{max}) was determined at each sample as the frequency with the greatest power in the 1.5 – 3.5 Hz band (i.e. 90 – 210 beats per minute). The STFT has a fixed relation between its window length and frequency resolution; long windows give higher frequency resolution but slower tracking; short windows give less frequency resolution faster tracking. We set the frequency resolution to 0.01 Hz, with a STFT window length N of 100 s. To obtain sample-by-sample estimates of HR, it was necessary to compute the STFT overlapping by $N-1$ samples. To improve computational efficiency, we decimated the ECG signal to 10 Hz prior to STFT computation, and the result was interpolated back with cubic splines to obtain the HR signal.

F. Instantaneous Power Estimates

The instantaneous power estimate of all metrics described above was also computed. To this end, each of the continuous metrics was squared and averaged over a symmetric, two-sided window of length N_{ma}. The correlation between the respiratory and heart rates (C_{RH}) was estimated by averaging the product $f_{max}*h_{max}$ over the same window. Short and long term estimates were obtained by setting N_{ma} to 1 and 5 min. respectively.

III. Machine Learning Approach

A. Features

The following features generated by the signal processing step were used for classification:

1. The seven features produced by AUREA (f_{max}, m^{rc}, m^{ab}, r^+, p^{rc}, p^{ab}, ϕ), during the second minute of he Phase II recording, the SBT trial.
2. The power and variability of the respiratory frequency estimate (f_{max}), estimated from: (i) the second minute of the Phase II recording, and (ii) from minutes 40-45 of Phase I.
3. The average heart rate (h_{max}) and its inter-quartile range measured over (i) the second minute of Phase II and (ii) the last 20 minutes of Phase I.
4. The power of the correlation between the heart rate and respiratory frequency C_{RH}, measured as above.

B. Classification Tools

For classification, we used Support Vector Machines (SVMs), a state-of-art classification algorithm. SVMs build a decision boundary between classes based on a batch of instances, so as to maximize the distance between the boundary and these instances. This makes the algorithm robust to noise. We used Gaussian kernels, which allow for a non-linear decision boundary. We used a publically available Matlab package for SVM learning [10]. SVMs allow controlling the trade-off between generalization and the classification of the training data using a parameter C, which specifies a tolerance threshold for examples to be misclassified. Based on small preliminary tests, we set this parameter to 10, and varied the width of the Gaussian kernel, W. We note that W and C have opposite effects, and many pairs of values yield similar results

For comparison, we also classified the data using logistic regression using all the features as inputs. The classifier provides a probability for belonging to the class of interest. The probability threshold on the failure class was varied to obtain the ROC curve

C. Classification Methods

There were 3,000 samples of the AUREA features for each baby. For each sample, we generated a separate instance for training the classifier. We used binary classification, with the classes corresponding to success or failure in the extubation. During testing, we evaluated performance per baby rather than per instance, as is usual in machine learning, this type of evaluation is more meaningful for the application. A baby was classified as belonging to a class if more than half of the instances generated for that baby were in that class. Although ties are possible in theory, they never occurred in the data

To assess the performance of the machine learning approach, we performed five-fold cross-validation, splitting the data into 5 subsets. The data for each baby belonged to exactly one subset. We repeatedly picked a subset on which to test, and trained the classifier on the remaining data. Thus, data from the same baby was never used in both training and testing

IV. Empirical Results

Data from two babies who succeeded extubation and one who failed were excluded because of missing data segments. The remaining data set had 42 babies who succeeded extubation and 11 who failed.

Fig. 1 presents an ROC curve obtained by varying W, the width of the Gaussian kernel in the SVM algorithm while holding the value of the C parameter constant at 10. The ROC curve for a logistic regression fit (using the same folds) is included for comparison. The SVM results were significantly better demonstrating that the non-linear decision boundaries allowed by SVM improved the classification performance.

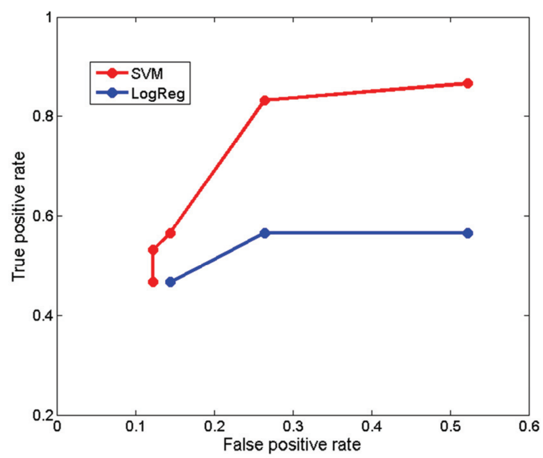

Figure 1: ROC Curves for Classification Using SVM and Logistic Regression.

Table I presents the accuracy obtained during training and testing, broken down by class, for the optimal value of W, shown in Figure 1. These results correspond to the point from the ROC curve with the best trade-off between sensitivity and specificity.

TABLE I. ACCURACY OF IDENTIFICATION

	Failure class	*Success class*
Training Accuracy	85.4%	89.7%
Testing Accuracy	83.2%	73.6%

The high accuracy within the failure class demonstrates that babies who will fail extubation can be identified with high precision. However, the lower accuracy within the success class indicates that the number of false positives is still somewhat high. This means that babies that actually succeeded were sometimes mistakenly predicted to fail.

The difference between the training and testing accuracy within the success class should also be noted since this suggests that there may be a problem with over-fitting. More work is needed to eliminate features which are not truly predictive.

V. Conclusions & Discussion

The accuracy with which the Failure Class subjects were classified is very encouraging. Note that all of the infants in the study had been judged to be ready for extubation on the basis of the best currently available clinical judgments. This suggests that the addition of our classification results to current clinical measures would make it possible to reduce the extubation failure rate by more than 80%, that is from the current 25% to 5%.

On the other hand, the accuracy of classification within the Success Class, would also result in a delay of extubation of some infants who might otherwise have been successfully extubated. The question of whether this trade off is acceptable depends on the relative clinical cost, measured in terms of health outcome and health care dollars, of extubation failure versus delayed extubation. It should be noted that it is possible to take the relative costs into account when selecting the best trade off between sensitivity and selectivity. How to do this is a difficult clinical question that is beyond the scope of the present paper

References

[1] M. C. Walsh, B. H. Morris, L. A. Wrage, B. R. Vohr, W. K. Poole, J. E. Tyson, L. L. Wright, R. A. Ehrenkranz, B. J. Stoll, and A. A. Fanaroff, "Extremely low birthweight neonates with protracted ventilation: mortality and 18-month neurodevelopmental outcomes," J Pediatr, vol. 146, pp. 798-804, Jun 2005.

[2] F. Hermeto, M. N. Bottino, K. Vaillancourt, and G. M. Sant'Anna, "Implementation of a respiratory therapist-driven protocol for neonatal ventilation: impact on the premature population," Pediatrics, vol. 123, pp. e907-16, May 2009.

[3] S. D. Baisch, W. B. Wheeler, S. C. Kurachek, and D. N. Cornfield, "Extubation failure in pediatric intensive care incidence and outcomes," Pediatr Crit Care Med, vol. 6, pp. 312-8, May 2005.

[4] Kaczmarek J, Kamlin O, Morley C, Davis Peter, and Sant'Anna GM, "Respiratory Variability Prior to Extubation in Very Low Birth Weight Infants. ," Arch. Dis. Child. Fetal & Neonatal Ed, vol. ., in press., 2012.

[5] S. C. J. Kaczmarek, C. Marchica, M. Dwaihy, L. Grundy, and GM Sant'Anna., "Heart Rate Variability and Extubation Readiness in Extreme Preterm Infants: A Prospective, Observational Study " in Pediatric Academic Society Annual Meeting,, Boston, USA, 2012.

[6] J. Kaczmarek, S. Chawla, C. Marchica, M. Dwaihy, L. Grundy, and G. Sant'Anna., "Respiratory Variability and Extubation Readiness in Extreme Preterm Infants: A Prospective, Observational Study " presented at the Pediatric Academic Society Annual Meeting, Boston, USA 2012.

[7] C. A. Robles-Rubio, J. Kaczmarek, K. A. Brown, R. E. Kearney, and G. M. Sant'anna, "Prediction of Extubation Readiness in Extreme Preterm Infants using Automated Analysis of the Respiratory Pattern," in Pediatric Academic Society Annual Meeting, Boston, USA 2012, p. (Accepted for presentation).

[8] C. A. Robles-Rubio, K. A. Brown, and R. E. Kearney, "Automated Unsupervised Respiratory Event Analysis," in Proc. 33rd IEEE Ann. Int. Conf. Eng. Med. Biol. Soc., Boston, USA, 2011, pp. 3201-3204.

[9] A. Aoude, R. E. Kearney, K. A. Brown, H. Galiana, and C. A. Robles-Rubio, "Automated Off-Line Respiratory Event Detection for the Study of Postoperative Apnea in Infants," IEEE Trans. Biomed. Eng., vol. 58, pp. 1724-1733, 2011.

[10] S. R. Gunn. Support Vector Machines for Classification and Regression". Technical report, University of Southampton, 1998. . Available: http://www.isis.ecs.soton.ac.uk/resources/svminfo/