AI/ML to Predict Extubation Failure

A.P. Michelson[1], S.C. Yu[2], A. Gupta[2], K. Betthauser[3], A.M. Lai[2], M.H. Kollef[1], P.R.O. Payne[2].
[1]Division of Pulmonary and Critical Care, Washington University School of Medicine, St. Louis, MO. [2]Institute for Informatics, Washington University School of Medicine, St. Louis, MO.
[3]Department of Pharmacy, Barnes-Jewish Hospital, St. Louis, MO

Introduction

Endotracheal (ET) intubation with mechanical ventilation (MV) is a life-saving procedure that provides respiratory support during critical illness.[1]

It is required in up to **40% of all intensive care unit (ICU) patients** and is responsible for **12% of all hospital costs.**[2,3]

Prolonged periods of MV are associated with worse outcomes and early extubation remains a paramount treatment goal.[4,5]

Reintubation occurs in approximately **15% of patients** after ET tube removal.[6]

Extubation failure is associated with a **4.7-fold longer ICU stay** and at least a **30% higher risk of inpatient mortality**. [6-8]

Machine learning approaches to predict successful extubation have not been well studied.[9-11]

Objectives

Develop a novel machine learning approach to **predict extubation failure within 48 hours of ET tube removal.**

Methods

Study Population: Data was extracted from the MIMIC-III dataset and included all patients ≥18 years of age who were intubated for at least 6 hours.[12] Model features were generated from demographics, vital signs, labs, and ventilation-related documentation.

MV: Initiation and duration of MV was not explicitly provided in the dataset and had to extrapolated, by categorizing all respiratory elements, such positive end expiratory pressure (PEEP), into three categories: 1) definite intubation; 2) definite extubation; and 3) ambiguous. Ambiguous was needed as some charted elements could be seen with non-invasive positive pressure ventilation modalities, such as FiO2.

Extubation failure: Defined as reintubation ≤48 hours after extubation.

Analysis: Data were analyzed in Python using jupyter notebook. XGBoost was trained with 4-fold cross-validation.[13,14] Group comparisons were assessed using the t- and χ^2-tests where appropriate.

Results

Population Characteristics

Variable	Total	Extubation Success	Extubation Failure	
Number of Events	7,401	6,943	458	
Age	65.3 (53.6 - 76.8)	65.5 (53.7 - 76.9)	63.7 (51.7 - 74.7)	*
Gender (Female)	3,140 (42.4%)	2,947 (42.4%)	193 (42.1%)	
Race (White)	5,376 (72.6%)	5,067 (73.0%)	309 (67.5%)	*
Race (Hispanic)	285 (3.9%)	269 (3.9%)	16 (3.5%)	
Race (Black)	626 (8.5%)	563 (8.1%)	63 (13.8%)	*
Race (Asian)	186 (2.5%)	170 (2.4%)	16 (3.5%)	
Race (Other/Unknown)	928 (12.5%)	874 (12.6%)	54 (11.8%)	
Body Mass Index	27.9 (24.3 - 32.5)	27.9 (24.4 - 32.6)	26.9 (23.8 - 31.9)	*
Intubation Duration (hours)	24.6 (14.0 - 65.5)	24.4 (14.0 - 65.6)	27.7 (15.5 - 62.6)	
Time-to-Reintubation (hours)	N/A	N/A	17.0 (10.4 - 28.5)	

Table 1. Baseline population characteristics. *P <.05

Model Performance

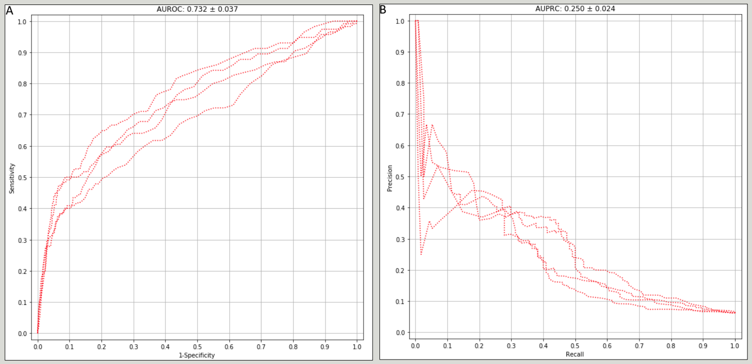

Figure 1. XGBoost model performance metrics. A. Area under the receiver operating characteristic curve (AUROC). Mean 4-fold AUROC 0.732. **B.** Area under the precision-recall curve (AUPRC). Mean 4-fold AUPRC 0.25.

Feature Importance

Feature	Cross validation fold 1	2	3	4	Mean	Standard deviation
PEEP set (std)	0.059	0.076	0.005	0.026	0.042	0.028
Minute volume (std)	0.031	0.019	0.030	0.004	0.021	0.011
Minute volume (first)	0.013	0.022	0.011	0.012	0.015	0.004
Mean airway pressure (std)	0.010	0.004	0.023	0.019	0.014	0.007
Peak inspiratory pressure (std)	0.004	0.004	0.038	0.003	0.012	0.015
GCS motor (1st percentile)	0.007	0.011	0.009	0.008	0.009	0.001
Arterial systolic blood pressure (last)	0.009	0.011	0.004	0.005	0.007	0.003
Tidal volume (observed) (std)	0.011	0.008	0.004	0.004	0.007	0.003
Tidal volume (observed) (25th percentile)	0.005	0.018	0.003	0.000	0.007	0.007
SpO2 (delta)	0.007	0.009	0.004	0.004	0.006	0.002

Table 2. Cross validation feature importance. The top 10 features, sorted by mean feature importance across folds. Feature importance instability can be observed in some features including the topmost feature (PEEP set std) which was underused in cross validation fold 3. Abbreviations: PEEP = positive end expiratory pressure; insp = inspiratory; GCS = Glasgow coma scale; SBP = systolic blood pressure; SpO2 = cutaneous oxygen saturation (%); std = standard deviation.

Results

Model Calibration

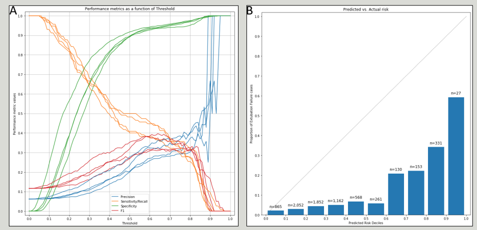

Figure 2. Model Calibration. A. Performance metrics for all cross-validations as a function of threshold. The harmonic mean of precision and recall peaks at a threshold value of approximately 0.65. **B.** Risk calibration plot. For each decile of extubation failure risk the proportion of actual extubation failures are shown.

Conclusions

Extubation failure occurred in 6% of this population and was more likely in patients who are **slightly younger** with a **lower BMI**.

The median **time to extubation failure was 17 hours.**

The **most predictive feature** was **standard deviation of set PEEP**.

Extrapolation of MV initiation and termination times from the electronic health record **remains a challenge** and **more advanced heuristics** and **new approaches** to documentation are **required for accurate prediction**.

References

1. Fan E, Del Sorbo L, Goligher EC, et al. An Official American Thoracic Society/European Society of Intensive Care Medicine/Society of Critical Care Medicine Clinical Practice Guideline: Mechanical Ventilation in Adult Patients with Acute Respiratory Distress Syndrome. *Am J Respir Crit Care Med*. 2017;195(9):1253-1263. doi:10.1164/rccm.201703-0548ST.
2. Wunsch H, Wagner J, Herlim M, Chong DH, Kramer AA, Halpern SD. ICU Occupancy and Mechanical Ventilator Use in the United States*. *Crit Care Med*. 2013;41(12):2712-2719. doi:10.1097/CCM.0b013e318298a139.
3. Wunsch H, Linde-Zwirble WT, Angus DC, Hartman ME, Milbrandt EB, Kahn JM. The epidemiology of mechanical ventilation use in the United States*. *Crit Care Med*. 2010;38(10):1947-1953. doi:10.1097/CCM.0b013e3181ef4460.
4. Colice GL, D TASP, S BDM. Laryngeal Complications of Prolonged Intubation*. *Chest*. 2015;96(4):877-884. doi:10.1378/chest.96.4.877.
5. Rothaar RC, Epstein SK. Extubation failure: magnitude of the problem, impact on outcomes, and prevention. *Cur Opin Crit Care*. 2003;9(1):59-66.
6. Scott K Epstein MD F, MD RLC, MD JBW. Effect of Failed Extubation on the Outcome of Mechanical Ventilation*. *Chest*. 2016;112(1):186-192. doi:10.1378/chest.112.1.186.
7. Thille AW, Harrois A, Schortgen F, Brun-Buisson C, Brochard L. Outcomes of extubation failure in medical intensive care unit patients*. *Crit Care Med*. 2011;39(12):2612-2618. doi:10.1097/CCM.0b013e3182282a5a.
8. Yan Feng MD P, Yaw Amoateng-Adjepong MD P, MD DK, MD CG, Constantine A Manthous MD F. Age, Duration of Mechanical Ventilation, and Outcomes of Patients Who Are Critically Ill. *Chest*. 2009;136(3):759-764. doi:10.1378/chest.09-0515.
9. Kwong MT, Colopy GW, Weber AM, Ercole A, Bergmann JHM. The efficacy and effectiveness of machine learning for weaning in mechanically ventilated patients at the intensive care unit: a systematic review. *Biodes Manuf*. 2018;2(1):31-40. doi:10.1007/s42242-018-0030-1.
10. Hsieh M-H, Hsieh M-J, Chen C-M, Hsieh C-C, Chao C-M, Lai C-C. An Artificial Neural Network Model for Predicting Successful Extubation in Intensive Care Units. *JCM*. 2018;7(9):240–10. doi:10.3390/jcm7090240.
11. Chen T, Xu J, Ying H, et al. Prediction of Extubation Failure for Intensive Care Unit Patients Using Light Gradient Boosting Machine. *IEEE Access*. 2019;7:150960-150968. doi:10.1109/ACCESS.2019.2946980.
12. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;*3:160035 doi*: 10.1038/sdata.2016.35. Available at: http://www.nature.com/articles/sdata201635.
13. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proc 22nd ACM SIGKDD Int Conf Know Disc Data Mining*.2016;785-794. doi:10.1145/2939672.2939785.
14. skopt module. Sci-Kit Learn. https://scikit-optimize.github.io. Accessed December 18, 2019.

Future Directions

Develop a more advanced heuristic to determine MV status and include more features, such as comorbidities and volume status.

Washington University School of Medicine in St. Louis

Institute for Informatics (I^2)