Machine Learning for Prediction of Successful Extubation of Mechanical Ventilated Patients in an Intensive Care Unit: A Retrospective Observational Study

Takanobu Otaguro, Hidenori Tanaka, Yutaka Igarashi, Takashi Tagami, Tomohiko Masuno, Shoji Yokobori, Hisashi Matsumoto, Hayato Ohwada and Hiroyuki Yokota

1Department of Emergency and Critical Care Medicine, Nippon Medical School, Tokyo, Japan
2Department of Industrial Administration, Tokyo University of Science, Chiba, Japan

Background: Ventilator weaning protocols are commonly implemented for patients receiving mechanical ventilation. However, despite such protocols, the rate of extubation failure remains high. This study analyzed the usefulness and accuracy of machine learning in predicting extubation success.

Methods: We retrospectively evaluated data from patients who underwent intubation for respiratory failure and received mechanical ventilation in an intensive care unit (ICU). Information on 57 features, including patient demographics, vital signs, laboratory data, and ventilator data, were extracted. Extubation failure was defined as re-intubation within 72 hours of extubation. For supervised learning, data were labeled as intubation-required or not. We used three learning algorithms (Random Forest, XGBoost, and LightGBM) to predict successful extubation. We also analyzed important features and evaluated the area under curve (AUC) and prediction metrics.

Results: Overall, 13 of the 117 included patients required re-intubation. LightGBM had the highest AUC (0.950), followed by XGBoost (0.946) and Random Forest (0.930). The accuracy, precision, and recall performance were 0.897, 0.910, and 0.909 for Random Forest; 0.910, 0.912, and 0.931 for XGBoost; and 0.927, 0.915, and 0.960 for LightGBM, respectively. The most important feature was duration of mechanical ventilation, followed by fraction of inspired oxygen, positive end-expiratory pressure, maximum and mean airway pressures, and Glasgow Coma Scale.

Conclusions: Machine learning predicted successful extubation of ICU patients on mechanical ventilation. LightGBM had the best overall performance. Duration of mechanical ventilation was the most important feature in all models. (J Nippon Med Sch 2021; 88: 408–417)

Key words: machine learning, mechanical ventilation, extubation failure, intensive care unit

Introduction

Mechanical ventilation is a life-saving modality for respiratory support of critically ill patients. In the United States, approximately 800,000 patients receive mechanical ventilation annually. Among patients receiving mechanical ventilation, weaning from ventilator support is one of the most important challenges in the intensive care unit (ICU). Extubation failure significantly increases the risk for adverse clinical events, length of ICU and hospital stay, and mortality. Unsuccessfully extubated patients are approximately seven times as likely to die as successfully extubated patients. Therefore, appropriate timing of extubation is an important issue for physicians.

Many ventilator weaning protocols have been developed and verified to improve extubation success rates. Using these protocols, physicians extubate patients with a higher probability for successful weaning, as indicated by clinical variables such as consciousness, vital signs, ar-
Predicting Successful Extubation

Interial blood gas findings, and ventilator settings. Compared with standard care, use of weaning protocols can reduce the duration of mechanical ventilation by 25%, weaning duration by 78%, and length of ICU stay by 10%. However, even when clinical practice for extubation adheres to the American Thoracic Surgery weaning protocol, extubation failure occurs in 10% to 15% of cases in the United States. The incidence of extubation failure has remained high despite the use of weaning protocols, and no significant improvement has been achieved in the last few decades. Accordingly, a new model is needed to improve the prediction accuracy.

Machine learning is a field of computational science that incorporates numerous factors to create systems that can learn from data in their environment and make predictions and take actions when confronted with a new situation. Machine learning might improve prediction of extubation success. Although numerous studies have investigated mechanical ventilation, few have used machine learning to predict the success of weaning from ventilatory support. Thus, this study analyzed the performance and accuracy of machine learning to predict extubation success.

Materials and Methods

Study Population

This study was approved by the Ethics Committee of the Nippon Medical School Hospital (30-06-949). The need for informed consent was waived. This single-center retrospective observational study was conducted from January 1, 2015 to December 31, 2018. Patients diagnosed with respiratory failure on admission, underwent intubation, and remained on mechanical ventilation for longer than 24 hours in the ICU were included in this study. Respiratory failure was defined as satisfying one of the following criteria: hypoxia with a partial pressure of arterial oxygen (PaO2) to a fraction of inspired oxygen (FiO2; P/F) ratio of ≤300 mm Hg; peripheral oxygen saturation (SpO2) of ≤90% when breathing room air; respiratory acidosis with a pH ≤7.32 and a partial pressure of arterial carbon dioxide (PaCO2) ≥45 mm Hg; and tachypnea with a respiratory rate of ≥30/min. The exclusion criteria were as follows: age <18 years, altered consciousness as the only indication for intubation, intubation for emergency surgery, death in facility, transportation to other hospitals with mechanical ventilation, and tracheostomy without attempted extubation.

Weaning Protocol

The need for extubation was determined by physicians using the weaning protocol developed by the joint committee of three academic societies in Japan, namely, the Japanese Society of Intensive Care Medicine, the Japanese Society of Respiratory Care Medicine, and the Japan Academy of Critical Care Nursing. The protocol is shown in Figure 1. Briefly, spontaneous awakening trials (SATs) are usually performed when patients are stable. A successful SAT is defined as absence of agitation and tachypnea after discontinuation of sedative agents. Then, spontaneous breathing trials (SBTs) are performed. Successful SBT is defined as meeting the following conditions under T-piece ventilation or pressure support ventilation: set to FiO2 ≤0.5 and a positive end-expiratory pressure (PEEP) ≤5 cm H2O from 30 minutes to 2 hours: (1) a respiratory rate of ≤30/min; (2) SpO2 of ≥94% or PaO2 of ≥70 mm Hg; (3) a heart rate of ≤140/min and no signs or symptoms of arrhythmia or myocardial ischemia; (4) no hypertensive urgency or emergency; and (5) no clinical signs or symptoms of respiratory distress (use of accessory muscles, seesaw breathing, severe dyspnea, anxiety, agitation). When SBT was successful, physicians performed extubation.

Data Collection

Data were collected from the electronic health record (Mirrel, Fukuda Denshi Co., Ltd., Tokyo, Japan). All data were anonymized and included patient demographics; vital signs per minute; laboratory values; ventilator data (per minute); diagnosis; clinical severity scores at ICU admission, such as Acute Physiology and Chronic Health Evaluation (APACHE) II score, Sequential Organ Failure Assessment (SOFA) score, and simplified acute physiology score (SAPS) II; and attending physician’s notes.

The following variables were analyzed: (1) demographic factors, including age, sex, height, body weight, and body mass index; (2) vital signs and other clinical factors per minute (systolic, diastolic, and mean blood pressure, heart rate, respiratory rate, body temperature [using a Foley catheter], SpO2, end-tidal carbon dioxide, the number of premature ventricular contraction) or at a predetermined interval (pupil diameters and Glasgow Coma Scale [GCS]); (3) arterial blood gas findings, including PaCO2, PaO2, HCO3−, base excess, sodium (Na+), potassium (K+), chloride (Cl−), calcium (Ca2+), anion gap, lactate, arterial oxygen saturation, carboxyhemoglobin, methemoglobin, pH, P/F ratio, and AaDO2; (4) laboratory results, including white blood cell count, red blood cell count, hematocrit, platelet count, prothrombin time-international normalized ratio, activated partial thromboplastin time (APTT), and levels of
Weaning protocols for mechanical ventilation

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Description</th>
</tr>
</thead>
</table>

SBT success criteria:
- Adequate oxygenation
- Stable hemodynamics
- Normal blood gas analysis
- No abnormal breathing
- No use of accessory respiratory muscles
- No use of respiratory assistance
- No severe carbon dioxide retention
- No severe acidosis
- No severe hypoxemia
- No severe hyperventilation

Before extubation:
- Oxygen saturation >98%
- Hemoglobin >8 g/dL
- Platelet count >100,000/mm³
- No clinical evidence of myocardial ischemia
- No severe electrolyte imbalance
- No severe acidosis
- No severe hypoxemia
- No severe hyperventilation

**Fig. 1** Ventilator weaning protocols developed by a joint committee in Japan
hemoglobin, fibrinogen, blood urea nitrogen (BUN), creatinine, total protein (TP), albumin, total bilirubin, aspartate transaminase (AST), alanine aminotransferase (ALT), lactic acid dehydrogenase (LDH), creatine phosphokinase (CPK), amylase, and C-reactive protein (CRP); (5) ventilator data including FiO2, maximum and mean airway pressure, PEEP, tidal volume, minute ventilation, and duration of mechanical ventilation.

Handling of Missing Values
Although vital signs were measured every minute, blood samples were measured, and consciousness was evaluated, at predetermined intervals or in relation to the patient’s condition. Some patients had missing values. Consciousness level was imputed by using the same values until the next observation. Arterial blood gas findings were imputed by using the same values as an hour before and after measurement.

Labeling
Extubation failure was defined as re-intubation within 72 hours after extubation. When high-flow nasal oxygen or noninvasive positive-pressure ventilation was required after extubation, it was still defined as successful extubation if re-intubation was not required within 72 hours after extubation. For patients who require mechanical ventilation, improvement in respiratory failure and weaning from ventilatory support takes several days. Patients were assumed to have required intubation and mechanical ventilation when they were intubated because of respiratory failure and required mechanical ventilation within 2 hours after intubation. They were considered as unsuccessfully extubated if they had required mechanical ventilation during the 3 hours before extubation and as successfully extubated if they had not required mechanical ventilation during the 3 hours before extubation (Fig. 2).

Statistical Analyses
Age and clinical severity score are reported as median and interquartile range. A receiver operating characteristic curve was drawn, and the area under the curve (AUC) was calculated. To predict successful extubation, we used five-fold cross validation to optimize evaluation metrics by the machine learning algorithms Random Forest, XGBoost, and LightGBM and analyzed important associated features. The Python language was used for coding the algorithm. We evaluated the prediction performance of all the methods in relation to accuracy, precision (positive predictive value), recall (sensitivity), and F1 score. In addition, we entered test data into the resulting algorithm and validated the rate of successful extubation per minute on trend graphs.

Results
Participant Characteristics and Collected Data
Data from 117 patients were included in this study (Ta-
ble 1). Two-thirds of the patients were men, and median patient age was 73 years (interquartile range [IQR], 59-84 years). Overall, 39 (33%) patients were diagnosed with pneumonia, 13 (11%) with trauma, and 10 (9%) with CO2 narcosis. The median APACHE II, SOFA, and SAPS II scores on admission were 22 (IQR, 19-25), 11 (IQR, 9-13), and 52 (IQR, 44-64), respectively. The median duration of mechanical ventilation was 5 days (IQR, 2-8 days) and the median duration of hospital stay was 16 days (IQR, 10-28 days). There were 13 patients who failed extubation; the indication for reintubation was respiratory failure in 7 (54%) patients, upper airway constriction in 3 (23%) patients, and aspiration in the other 3 (23%) patients. The characteristics of patients in the successful and failed extubation groups are shown in Table 2. The total number of collected values on 57 patient features was 12,268. Of these, 6,721 and 5,547 were labeled as intubation-required and not, respectively.

**Performance and Analysis of Feature Importance**

ROC curves were generated (Fig. 3), and the AUC values were 0.931 (95% confidence interval [CI]: 0.889-0.972), 0.947 (95% CI: 0.908-0.985), and 0.950 (95% CI: 0.909-0.992) on Random Forest, XGBoost, and LightGBM, respectively. The performance characteristics are shown in Table 3. Random Forest had an accuracy of 0.897, precision of 0.910, recall of 0.909, and F1 score of 0.909. XGBoost had an accuracy of 0.910, precision of 0.912, recall of 0.931, and F1 score of 0.921. LightGBM had an accuracy of 0.927, precision of 0.915, recall of 0.960, and F1 score of 0.937.

The results of analysis of feature importance are shown in Figure 4-6. The most important feature was duration of mechanical ventilation. Other important features, by descending order of importance, were maximum airway pressure, mean airway pressure, FiO2, pH, GCS, TP, albumin, and base excess (on Random Forest); CPK, mean blood pressure, PEEP, BUN, age, ALT, AST, and anion gap (on XGBoost); and age, PEEP, LDH, APTT, GCS, BUN, AaDO2 and CRP (on LightGBM).

**Trends in Extubation Success Rate**

Extubation success rate per minute was examined in this model, and trend graphs were drawn. All models showed that the probability of successful extubation increased over time. Random Forest and LightGBM showed low probabilities at the time of extubation and accurately predicted extubation failure. Figure 7 shows characteristic trends in extubation success rate for patients with successful and failed extubation.

**Discussion**

The present results suggest that machine learning can predict successful extubation of patients on mechanical ventilation in the ICU. When the predictive performance of the three models was evaluated, almost all metrics were sufficient to predict successful extubation. In this study, LightGBM had the highest AUC value among the three models. This is consistent with a previous study, which showed that, as compared with XGBoost, an artificial neural network, and support vector machine, LightGBM was the most effective model for predicting extubation. In the evaluation metrics, high precision decreases re-intubation, and high recall decreases unnecessary ventilator use and tracheostomy. LightGBM had the best precision, and this translates to a reduction in re-intubations due to extubation failure.
Predicting Successful Extubation

Table 2 Characteristics of successfully and unsuccessfully intubated patients

<table>
<thead>
<tr>
<th>Variables</th>
<th>Successful extubation</th>
<th>Failed extubation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=104</td>
<td>N=13</td>
</tr>
<tr>
<td>Age, median (IQR)</td>
<td>72 (58-83)</td>
<td>84 (79-88)</td>
</tr>
<tr>
<td>Male</td>
<td>69 (66%)</td>
<td>8 (62%)</td>
</tr>
<tr>
<td>BMI, median (IQR)</td>
<td>22 (19-25)</td>
<td>21 (18-24)</td>
</tr>
<tr>
<td>APACHE II score, median (IQR)</td>
<td>24 (21-28)</td>
<td>27 (23-30)</td>
</tr>
<tr>
<td>SOFA scores, median (IQR)</td>
<td>11 (9-13)</td>
<td>12 (11-14)</td>
</tr>
<tr>
<td>SAPS II score, median (IQR)</td>
<td>50 (44-63)</td>
<td>66 (55-73)</td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pneumonia</td>
<td>33 (32%)</td>
<td>6 (46%)</td>
</tr>
<tr>
<td>Trauma</td>
<td>13 (13%)</td>
<td>0</td>
</tr>
<tr>
<td>CO₂ narcosis</td>
<td>9 (9%)</td>
<td>1 (8%)</td>
</tr>
<tr>
<td>Intoxication</td>
<td>9 (9%)</td>
<td>0</td>
</tr>
<tr>
<td>Endocrine or metabolic disorder</td>
<td>6 (6%)</td>
<td>0</td>
</tr>
<tr>
<td>Sepsis</td>
<td>12 (12%)</td>
<td>4 (31%)</td>
</tr>
<tr>
<td>Cardiac failure</td>
<td>5 (5%)</td>
<td>0</td>
</tr>
<tr>
<td>Heat stroke or accidental hypothermia</td>
<td>4 (4%)</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>2 (2%)</td>
<td>1 (8%)</td>
</tr>
<tr>
<td>Burn</td>
<td>3 (3%)</td>
<td>0</td>
</tr>
<tr>
<td>Asthma</td>
<td>1 (1%)</td>
<td>0</td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>1 (1%)</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>6 (6%)b</td>
<td>0</td>
</tr>
<tr>
<td>Outcome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of mechanical ventilation in days, median (IQR)</td>
<td>5 (2-9)</td>
<td>7 (5-8)</td>
</tr>
<tr>
<td>Duration of hospitalization in days, median (IQR)</td>
<td>15 (10-27)</td>
<td>23 (18-29)</td>
</tr>
<tr>
<td>Hospital mortality rate</td>
<td>4 (4%)</td>
<td>5 (39%)</td>
</tr>
</tbody>
</table>

aData on sepsis due to pneumonia were excluded
bFive patients with water inhalation and one with asphyxia

Abbreviations: APACHE, Acute Physiology and Chronic Health Evaluation; IQR, interquartile range; SAPS, Simplified Acute Physiology Score; SOFA, Sequential Organ Failure Assessment

Fig. 3 ROC curves for the machine learning methods
Further, this study identified features associated with successful extubation. Duration of mechanical ventilation was the most important feature in all models, which is consistent with the findings of another study. Long-term intubation is a risk factor for extubation failure, and long-term mechanical ventilation is an independent risk factor for worse prognosis. Additionally, long-term mechanical ventilation is associated with pneumonia incidence, acute lung injury, and worse mortality. However, the present models could not determine whether short-term or long-term intubation was associated with extubation failure. The importance of other features varied among the models, but FiO2, PEEP, maximum and mean airway pressure (parallel to PEEP and pressure support), and GCS were identified as very important features, consistent with the weaning protocol. Advanced age had a strong influence on increasing the risk of extubation failure. Although albumin and CRP were not included in the weaning protocol, they reflect pathophysiology and were identified as important features in this study. It was unclear how some of the important features identified affect decision-making.
Predicting Successful Extubation

Linear models such as logistic regression can predict successful extubation and explain the extent to which a predictive variable is associated with an objective variable. While it is more difficult for machine learning to explain the predictive variables by intuition, it may improve prediction accuracy. Our results show that, as compared with the weaning protocol, machine learning appears to improve prediction of successful extubation. The extubation success rate can be shown per minute using this model, and trend graphs can be drawn (Fig. 7). The model successfully predicted extubation success or failure for cases that were not used to develop the model. The process for determining the possibility of successful extubation in machine learning is completely different from that used in the weaning protocol. In the weaning protocol, the need for extubation is determined by using defined factors. By contrast, in machine learning, numerous factors from digital health records can be used to determine the probability of successful extubation. Therefore, machine learning could be a useful clinical decision support tool for predicting successful extubation.

This study has several limitations. First, because only patients with respiratory failure were eligible, it is unclear whether these models can be used for patients who undergo intubation for other reasons, such as status epilepticus or severe traumatic brain injury. The applicability of the models for patients who underwent intubation without respiratory failure needs to be validated. Second, although chest X-ray images, cuff-leak testing, diaphragm ultrasonography, and fluid balance are helpful for predicting successful extubation, this study did not include features and modalities commonly used to predict successful extubation. Because the most suitable way to use machine learning with varied data types

Fig. 6 Ranking of feature importance in the LightGBM model

Fig. 7 Trends for successful extubation (a) and extubation failure (b). Blue indicates the Random Forest model; green, the XGBoost model; and orange, LightGBM.
is unclear, we did not include both images and numbers when performing machine learning. The inclusion of additional features and other modalities may increase prediction accuracy. Further, it might be difficult to use the present models to predict extubation failure due to post-extubation laryngeal edema, as we did not include a variable reflecting laryngeal edema. Validation is required in order to determine whether this model can predict extubation success with other datasets.

The results of this study suggest that machine learning can predict successful extubation of ICU patients on mechanical ventilation. LightGBM had the best overall performance. Although variables such as FiO2, PEEP, maximum and mean airway pressure, and GCS were included in the weaning protocol, the most important feature was duration of mechanical ventilation, which is not included in the weaning protocol.

Acknowledgement: This work was supported by JSPS KAKENHI Grant Number 20K17876.

Conflict of Interest: The authors declare no conflicts of interest.

References


Predicting Successful Extubation


(Received, August 10, 2020)

(Accepted, October 28, 2020)

(J-STAGE Advance Publication, March 9, 2021)