

OPEN

The Contribution of Chest X-Ray to Predict Extubation Failure in Mechanically Ventilated Patients Using Machine Learning-Based Algorithms

OBJECTIVES: To evaluate the contribution of a preextubation chest X-ray (CXR) to identify the risk of extubation failure in mechanically ventilated patients.

DESIGN: Retrospective cohort study.

SETTINGS: ICUs in a tertiary center (the Medical Information Mart for Intensive Care IV database).

PATIENTS: Patients greater than or equal to 18 years old who were mechanically ventilated and extubated after a spontaneous breathing trial.

INTERVENTIONS: None.

MEASUREMENTS AND MAIN RESULTS: Among 1,066 mechanically ventilated patients, 132 patients (12%) experienced extubation failure, defined as reintubation or death within 48 hours of extubation. To predict extubation failure, we developed the following models based on deep learning (EfficientNet) and machine learning (LightGBM) with the training data: 1) model using only the rapid-shallow breathing index (RSBI), 2) model using RSBI and CXR, 3) model using all candidate clinical predictors (i.e., patient demographics, vital signs, laboratory values, and ventilator settings) other than CXR, and 4) model using all candidate clinical predictors with CXR. We compared the predictive abilities between models with the test data to investigate the predictive contribution of CXR. The predictive ability of the model using CXR as well as RSBI was not significantly higher than that of the model using only RSBI (c-statistics, 0.56 vs 0.56; $p = 0.95$). The predictive ability of the model using clinical predictors with CXR was not significantly higher than that of the model using all clinical predictors other than CXR (c-statistics, 0.71 vs 0.70; $p = 0.12$). Based on SHapley Additive exPlanations values to interpret the model using all clinical predictors with CXR, CXR was less likely to contribute to the predictive ability than other predictors (e.g., duration of mechanical ventilation, inability to follow commands, and heart rate).

CONCLUSIONS: Adding CXR to a set of other clinical predictors in our prediction model did not significantly improve the predictive ability of extubation failure in mechanically ventilated patients.

KEY WORDS: chest X-ray; extubation; intubation; machine learning; mechanical ventilation; reintubation

Identifying the optimal timing of extubation in mechanically ventilated patients in ICUs is of great interest to intensivists. Both long-term mechanical ventilation with delayed extubation and reintubation as a result of premature extubation are known to be associated with higher mortality due to secondary medical complications, such as ventilator-associated pneumonia or muscle weakness (1–4). Thus, protocols for ventilator weaning and extubation

Kiyoyasu Fukuchi¹
Itsuki Osawa, MD²
Shunya Satake, MD³
Honoka Ito⁴
Junichiro Shibata¹
Eisuke Dohi, MD⁵
Daisuke Kasugai, MD⁶
Yoshihisa Miyamoto, MD, PhD⁷
Hiroyuki Ohbe, MD⁸
Mitsuhiro Tamoto, RN, MS⁹
Naoki Yamada, MD¹⁰
Keisuke Yoshikawa¹¹
Tadahiro Goto, MD, MPH, PhD^{8,12}

Copyright © 2022 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of the Society of Critical Care Medicine. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/CCE.0000000000000718

have been developed to reduce the total duration of mechanical ventilation and minimize the risk of extubation failure (5–8).

Protocol-based ventilator weaning and extubation may reduce the risk of extubation failure (6–8); however, patients eligible for the protocol of a previous study still experienced postextubation respiratory failure (9.7%) and reintubation (5.2%) (9). A previous study indicated that among clinical parameters used for ventilator weaning and extubation, a tolerance for spontaneous breathing trial (SBT), factors affecting airway competence, and a stable cardiovascular system are key predictors of extubation failure (5). Given that chest X-rays (CXRs) have a crucial role in ICUs for evaluating the severity of cardiopulmonary illness and the complications caused by indwelling devices (10), preextubation CXR findings are thought to be beneficial to more accurately assess the risk of extubation failure. As exemplified by the study by Rackley et al (11) showing that 40% of patients subsequently diagnosed with acute lung injury had bilateral infiltrates on their CXRs 12 hours prior to their diagnoses, preextubation CXR findings could prognosticate future postextubation respiratory failure. However, as a routine CXR is not recommended for a preextubation assessment as per expert opinion (12), there have been no studies that evaluated the usefulness of preextubation CXR for the risk assessment of extubation failure. In this context, we developed machine learning-based models that predict extubation failure using clinical parameters with and without CXR available at the time of extubation. By comparing the prediction performance of each model, we aimed to evaluate the predictive ability of a preextubation CXR in mechanically ventilated patients.

MATERIALS AND METHODS

Study Design and Patients

This is a retrospective cohort study using the Medical Information Mart for Intensive Care IV (MIMIC-IV) dataset Version 1.0 (13) and the Medical Information Mart for Intensive Care Chest X-ray (MIMIC-CXR) dataset Version 2.0.0 (14). MIMIC-IV is an extensive, publicly available database consisting of deidentified health-related data from over 60,000 patients admitted to the ICUs of the Beth Israel Deaconess Medical Center in the United States from 2008 to

2019. MIMIC-CXR is a large publicly available dataset of CXRs in DICOM format. Approval by the research ethics committee was not needed for this study because MIMIC-IV and MIMIC-CXR are deidentified according to the HIPAA Safe Harbor provision, and only credentialed authors who signed and conformed to the specified data use agreement accessed and analyzed the data (13). Because of this, the TXP Medical Ethical Review Board waived the requirement for the ethical approval statement and informed consent (TXPREC-008) on January 20, 2022.

Study Participants

The inclusion and exclusion criteria are shown in **Figure 1**. We identified patients greater than or equal to 18 years who were mechanically ventilated and extubated in the ICU. Our analysis was limited to the first intubation and extubation episode during the first ICU stay of each patient regardless of the number of ICU stays during the hospitalization. We excluded patients using the following criteria: patients without a record of completing the SBT; patients whose rapid-shallow breathing index (RSBI) could not be calculated due to missing data on either tidal volume (TV) or respiratory rate (RR) measured by ventilators (15, 16); patients without CXRs taken within 48 hours before extubation; and patients palliatively extubated (i.e., patients for whom vasopressors [epinephrine, norepinephrine, dopamine, and dobutamine] were stopped within 6 hr before extubation and died within 24 hr after extubation) (17).

Measurements and Candidate Predictors

The following clinical variables were extracted from the MIMIC-IV database as candidate predictor input to prediction models (18–22): patient demographics (age, sex, body mass index, Charlson Comorbidity Index, the maximum Sequential Organ Failure Assessment score in first 24 hr of ICU stay, and duration of mechanical ventilation); average vital signs during the last 6 hours before extubation (heart rate, mean arterial blood pressure, body temperature, Glasgow Coma Scale [GCS], and history of vasopressor administration); average values of arterial blood gas during the last 6 hours before extubation (pH, PaO₂/Fio₂ ratio, arterial CO₂ pressure, and bicarbonate); the latest laboratory tests in the last 48 hours before extubation (WBC,

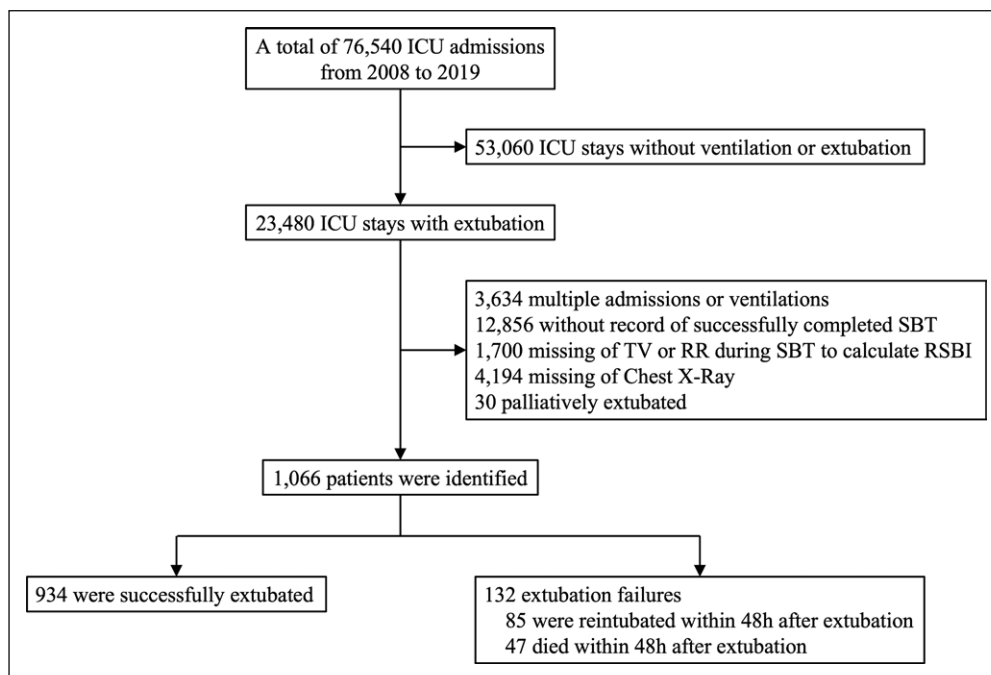


Figure 1. Flow diagram of study participants' selection for analysis. Among a total of 76,540 ICU admissions from 2008 through 2019, we identified 1,066 mechanically ventilated patients. RR = respiratory rate, RSBI = Rapid-Shallow Breathing Index, SBT = spontaneous breathing trial, TV = tidal volume.

hemoglobin, platelet, prothrombin time-international normalized ratio, activated partial thromboplastin time, sodium, potassium, chloride, inorganic phosphorus, lactate, glucose, blood urea nitrogen, and creatinine); average ventilator parameters during the last 6 hours before extubation (peak inspiratory pressure, positive end-expiratory pressure, pressure support, TV, RR, and F_{IO_2}); the RSBI during the SBT calculated using the average value of TV and RR within 1 hour from the SBT initiation; and the latest CXR image in the last 48 hours before extubation.

Outcomes and Variables

The predicted outcome was a failure of extubation, defined as reintubation or death within 48 hours of extubation (21). If there was no record of reintubation or death within that interval after extubation, the extubation was considered successful.

Statistical Analysis

After identifying patients and extracting variables from the database, we examined and imputed missing values and outliers using MissForest (23), a nonparametric machine learning-based method for imputation. The

valid ranges of variables and proportions of missing values are shown in **Supplemental Table 1** (<http://links.lww.com/CCX/B10>).

We used summary statistics to delineate the characteristics of the extracted and imputed data. For categorical variables, the proportions were calculated by dividing the number of events by the total number of patients. For continuous variables, median and interquartile ranges were calculated.

The entire architecture of the models is shown in **Supplemental Figure 1** (<http://links.lww.com/CCX/B10>). To predict the

probability of extubation failure using the data available at the time of extubation, we applied two machine learning and deep learning models. First, to extract features from the CXR images, we used image preprocessing (24) (e.g., randomly rotating, cropping, or horizontally flipping images) and EfficientNet (25), which is a convolutional neural network (CNN)-based model known to achieve much better accuracy and efficiency with fewer hyperparameters than other CNN models (e.g., VGG [26], InceptionNet [27], ResNet [28], DenseNet [29], or VisionTransformer [30]) in image classification. To avoid model overfitting given the insufficient sample size of our study, we selected the EfficientNet. We trained and tuned EfficientNet models and predicted probabilities of extubation failure using nested stratified five-fold cross validation (31) to make predictions of extubation failure on images not used during training or tuning. The predicted probabilities were treated as extracted features from CXR images. Second, they were concatenated with other candidate clinical predictors. Third, concatenated features were input to LightGBM, a high-performance machine learning algorithm based on gradient boosting and decision trees, which ensembles predictions of many decision trees and outputs the probability

of extubation failure (32). We used LightGBM-based models with probability calibration (33) to improve each model's prediction ability for extubation failure. We trained and tuned the LightGBM-based models and predicted probabilities of extubation failure using nested stratified five-fold cross validation in the same way as the EfficientNet models. Hyperparameters were optimized using the five-fold cross validation method for area under the receiver-operating-characteristic curve (AUROC) by Bayesian optimization, which is suitable for efficiently optimizing models (34). We compared the predictive abilities of an EfficientNet-based image classification model with only CXR and four LightGBM-based models using a set of different variables: 1) model using only RSBI, 2) model using RSBI and CXR, 3) model using all candidate clinical predictors without CXR, and 4) model using all candidate clinical predictors with CXR.

As for performance measures, we calculated the AUROC, the area under the precision-recall (PR) curve, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) using the predicted probabilities for the test data. The test by DeLong et al (35) was used to test the significance of AUROC scores between the feature combinations input into the LightGBM-based models. Cutoffs to calculate sensitivity, specificity, PPV, and NPV were selected based on the Youden index (36). We also plotted calibration curves (33) of the prediction models. To consider the different weights of different misclassification types with a direct clinical interpretation (e.g., trade-offs between underestimation and overestimation), we performed a decision curve analysis (DCA) (37). DCA is a method to evaluate clinical usefulness of prediction models (38). For each model trained with each combination of the features, the decision curve was plotted based on the net benefit of each model. Finally, we performed SHapley Additive exPlanations (SHAP) (39), a game-theoretic approach for explaining the influence of input variables on the prediction of a machine learning model by approximating the Shapley value. The SHAP value of each variable was estimated using the LightGBM model with all the variables, and a summary plot was created for the variables in the descending order of the SHAP value.

A p value of less than 0.05 was considered statistically significant. We performed all analyses with Google BigQuery, Python (Version 3.7.9) and R (Version 4.1.1).

RESULTS

Participants' Characteristics

There were a total of 76,540 admissions to the ICU from 2008 to 2019. We identified 1,066 patients who met the inclusion criteria. Among them, we identified 132 extubation failures (12.4%), consisting of 85 reintubations and 47 deaths within 48 hours after extubation. We excluded 30 patients who were possibly palliatively extubated.

Patients' characteristics are shown in **Supplemental Table 2** (<http://links.lww.com/CCX/B10>). Compared with patients with successful extubations, patients with extubation failure were more likely to have been ventilated longer (2.8 vs 2.2 d; $p < 0.01$), less likely to obey commands in assessments of GCS (71 patients [54%] vs 711 patients [76%]; $p < 0.01$), more likely to receive vasopressors (34 patients [26%] vs 158 patients [17%]; $p < 0.01$), and have higher RRs (20 vs 18/min; $p < 0.01$), whereas there were no clinically significant differences in other characteristics.

Performance of the Developed Models

The diagnostic performance and receiver-operating-characteristic, PR, and calibration curves of models are shown in **Table 1**, **Figure 2**, and **Supplemental Figure 2** (<http://links.lww.com/CCX/B10>). The EfficientNet-based image classification model using only a CXR had a similar diagnostic performance (AUROC, 0.55 [95% CI, 0.49–0.60]) to the LightGBM model using only RSBI (AUROC, 0.56 [95% CI, 0.51–0.62]).

The addition of CXR to RSBI did not improve the prediction ability (AUROC, 0.56 [95% CI, 0.51–0.62] for the model using only RSBI vs 0.56 [95% CI, 0.52–0.62] for the model using CXR as well as RSBI; $p = 0.95$). The prediction model using other clinical predictors had a higher discrimination ability than the model using only RSBI. However, the addition of CXR did not improve the predictive ability (AUROC, 0.70 [95% CI, 0.65–0.75] for the model using clinical predictors with CXR vs 0.71 [95% CI, 0.66–0.76] for the model using clinical predictors without CXR; $p = 0.12$). As for sensitivity, specificity, PPV, and NPV, there were no clinically significant differences between the LightGBM-based models with and without CXR.

TABLE 1.
Diagnostic Performance of the EfficientNet-Based Model and Four LightGBM-Based Models to Predict Extubation Failure

Models	Area Under the Receiver-Operating-Characteristic Curve (95% CI)	p^a	Area Under the Precision-Recall Curve (95% CI)	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
CXR	0.55 (0.49–0.60)	0.95	0.14 (0.00–0.16)	0.66	0.44	0.14	0.90
RSBI	0.56 (0.51–0.62)		0.15 (0.10–0.19)	0.48	0.65	0.16	0.90
RSBI + CXR	0.56 (0.51–0.62)	0.12	0.14 (0.10–0.18)	0.61	0.57	0.17	0.91
RSBI + other variables ^b	0.70 (0.65–0.75)		0.25 (0.13–0.33)	0.62	0.69	0.22	0.93
RSBI + CXR + other variables ^b	0.71 (0.66–0.76)		0.24 (0.13–0.34)	0.66	0.66	0.21	0.93

CXR = chest X-ray, RSBI = Rapid-Shallow Breathing Index.

^aWe compared the area under the curve between each LightGBM-based model with and without CXR using the DeLong test.

^bWe used all clinical parameters listed in Supplemental Table 2 (<http://links.lww.com/CCX/B10>).

Decision Curve Analysis

Decision curves of prediction models are shown in **Figure 2**. The X-axis indicates threshold probabilities for extubation failure. If the probability predicted by the model is higher than the threshold, the extubation is considered to have failed. The Y-axis indicates the net benefit over a specified range of threshold probabilities of extubation failure, calculated across the range of threshold probabilities by using the following formula: sensitivity \times prevalence – (1 – specificity) \times

(1 – prevalence) \times threshold/(1 – threshold), where prevalence is the frequency of extubation failure, calculated across the range of threshold probabilities. The unit of the net benefit is true-positive. For example, if the difference in net benefits between two models is 0.1, it means one model can accurately identify one additional patient positive for a diagnostic test out of every 10 patients in the target population than the other model. There were no clinically significant differences in the net benefit between the LightGBM-based model using only RSBI and the model using RSBI and

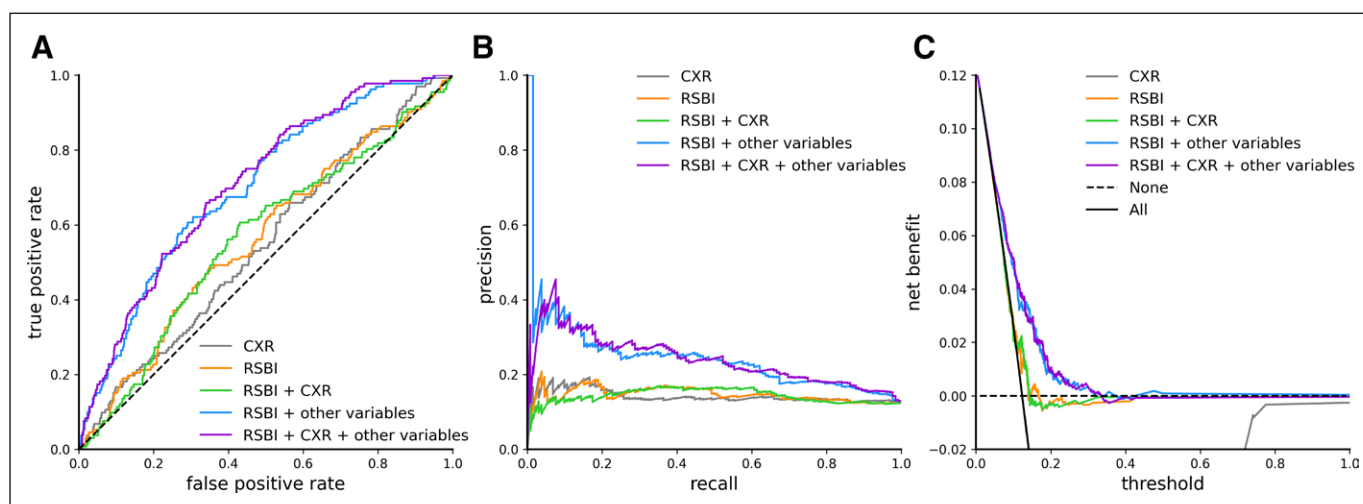


Figure 2. Curves to evaluate predictive abilities of the EfficientNet-based model and four LightGBM-based models. **A**, Receiver-operating-characteristic curves. The corresponding values of the area under the receiver-operating-characteristic curve for each model (i.e., the c-statistics) are presented in Table 1. **B**, Precision-recall curves. The corresponding values of the area under the precision-recall curve for each model are presented in Table 1. **C**, Decision curve analysis. The X-axis indicates threshold probabilities for extubation failure. The Y-axis indicates net benefit. The unit of the net benefit is true positive. CXR = chest X-ray, RSBI = Rapid-Shallow Breathing Index.

CXR. Similarly, we did not find clinically meaningful differences in the net benefit between the models using all candidate clinical predictors with and without CXR. The models using other candidate clinical predictors had the higher net benefit than treating all or no patients, with threshold probability ranging from about 0.1–0.5.

SHapley Additive exPlanations

SHAP values with prediction models using all variables are shown in **Figure 3**. CXR was not likely to have an impact on the prediction model for extubation failure compared with other clinical predictors (e.g., inability to follow commands, use of vasopressors, and higher RRs during SBT).

DISCUSSION

Using data on 1,066 mechanically ventilated patients in the ICU from the MIMIC-IV database, we investigated whether CXR data could improve the performance of machine learning-based prediction models for extubation failure. Our study showed that adding CXR to a set of other candidate clinical predictors did not significantly enhance the predictive ability for extubation failure in mechanically ventilated patients. In addition, as the SHAP values of our prediction model suggest, CXR may contribute less to the predictive performance of extubation failure than other predictors such as inability to follow commands (motor response of GCS < 6), use of vasopressors, and higher RRs during SBT.

Identifying the risk of extubation failure is clinically important, and RSBI was developed as a simple assessment method for extubation failure (40). In a previous meta-analysis (41), the sensitivity and specificity of RSBI in predicting extubation failure were higher than those of our study (sensitivity, 0.58 vs 0.48; and specificity, 0.83 vs 0.65). This is possibly because we included patients at a relatively low risk of extubation failure. RSBI of more than 100 is known to be a risk factor for extubation failure, but only 0.2% of patients in our study had RSBI of more than 100. Furthermore, a prediction model using LightGBM in a previous study had a higher discrimination ability than our model (AUROC, 0.81 vs 0.71) (21). We believe that this is possibly because the study did not exclude patients who were palliatively extubated and instead classified them as extubation failure. Given that patients who

were palliatively extubated are more likely to be critically ill than those who were successfully extubated, it may have been easier to predict extubation failure in such severely ill patients.

Some intensivists may routinely order CXRs to evaluate the severity of cardiopulmonary diseases and the complications caused by indwelling devices (10). In a web-based Delphi study from France, Hejblum et al (12) reported that CXRs should be considered routinely right after specific procedures and during mechanical ventilation for respiratory failure, whereas a routine CXR was not recommended for a preextubation assessment. However, there has been no qualitative evidence on the utility of a preextubation CXR beyond expert consensus. Given that the lung infiltrates might precede future postextubation respiratory failure in some patients (11), we hypothesized that a routine CXR right before extubation might be justified from a risk-benefit perspective to evaluate potential preexisting lung infiltrates. However, our findings suggested that the utility of a preextubation CXR to the prediction of extubation failure was limited. Thus, to make the most of the information extracted from CXRs to assess the risk of extubation failure, it may be necessary to follow longitudinal changes in CXRs over time.

This study applied comprehensive methods to incorporate images and other features into machine learning-based models, tune hyperparameters, train the models efficiently, and evaluate performance of those models. Our approach can be applied within many frameworks for machine learning (e.g., PyTorch [42], TensorFlow [43], or Scikit-learn [44]). We used EfficientNet to incorporate images into machine learning-based models, but other object detection models (e.g., YOLO [45], SSD [46], or DETR [47]) or segmentation models (e.g., U-Net [48], PSPNet [49], or DeepLab [50]) can efficiently extract useful features from a cross-sectional preextubation CXR to predict extubation failure. However, since there were no annotations of CXRs in the MIMIC-CXR dataset, we could not use these models in this study. Hyperparameter tuning or optimization is an essential aspect of machine learning to achieve the desired metric values. We used Bayesian optimization to search for a better combination of hyperparameters with a smaller number of trials than the conventional grid search (51). Nested cross validation for hyperparameter optimization and

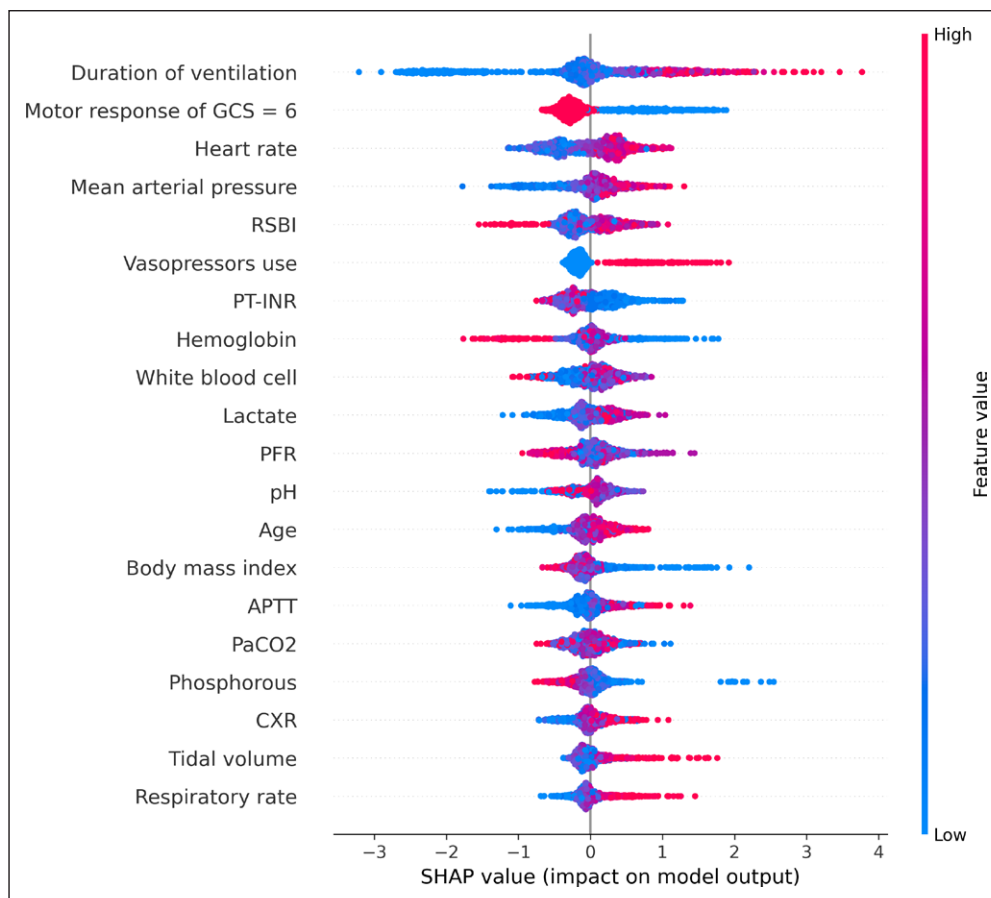


Figure 3. SHapley Additive exPlanations (SHAP) summary plot of top 20 variables of LightGBM-based model using all candidate clinical predictors with chest X-ray (CXR). The *horizontal axis* represents SHAP values, and a *dot* indicates the attribution of each variable at a feature value from the data sample. The *color of a dot* indicates the absolute value of each variable (e.g., *red dots* represent higher feature values, and *blue dots* represent lower feature values). The higher the SHAP value, the higher the possibility of extubation failure. The *vertical axis* represents all variables input to the prediction models, which are sorted based on the impact on the prediction models, which was calculated by averages of absolute SHAP values across all data. APTT = activated partial thromboplastin time, GCS = Glasgow Coma Scale, PFR = $\text{PaO}_2/\text{FiO}_2$ ratio, PT-INR = prothrombin time-international normalized ratio, RSBI = Rapid-Shallow Breathing Index.

model evaluation can also contribute to efficiently evaluating developed models. An ROC curve and a PR curve are standard methods to describe how accurately each model can predict outcomes. DCA is also an approach to evaluate the clinical utility of the models. SHAP is a game-theoretic method to explain the impact of each predictor on the predictive output. These exhaustive and flexible approaches will accelerate future research on clinical prediction models.

Our study has several limitations. First, there are several sources of selection bias in the exclusion criteria (e.g., exclusion of patients with missing data on SBT records, TV and RR values during the SBT and CXRs, and patients who were possibly palliatively

extubated). We acknowledge that our findings may lack external validity and cannot be extrapolated to the patients excluded from our analysis. However, included patients were likely more ill than those excluded from our analysis (e.g., patients considered clinically at low risk of extubation failure without SBT), and thus, our study is clinically meaningful in that it investigated whether CXR is useful for identifying patients at high risk of extubation failure. Furthermore, given that CXRs are more likely to augment prediction in the included (likely to be sicker) patients than those in the excluded (likely to be less sick) patients and that our findings showed no improvement in the predictive performance in the high-risk population included in the analysis, we believe that our results have the potential to be extrapolated to all mechanically ventilated

patients. Second, our data from the MIMIC-IV database included missing data (0.1–45% of data were missing depending on the variable), which could be a potential source of bias. However, we prioritized using several clinically important variables even though the missing proportion was high (e.g., 45% of the data on MAP was missing) and attempted to minimize the risk of bias by the random forest imputation method, a rigorous and widely used technique for the imputation of missing data (52, 53). Third, due to the retrospective nature of this study, we could not consider in our models some clinically important predictors of extubation failure (e.g., cough strength, sputum volume, and degree of disuse syndrome) because these data were

unavailable in the MIMIC-IV dataset (9, 54). Fourth, our study was based on data from a single center in the United States, and our findings may not be generalizable to other hospital settings. However, the utility of a preextubation CXR for predicting extubation failure might not vary across hospital settings because CXRs simply reflect the patients' physiologic condition. Finally, although our findings showed that CXR before extubation might not be useful in predicting extubation failure, this is possibly because of technical issues with the EfficientNet-based image classification model or incorporation of the extracted features from CXRs into the LightGBM-based prediction models. The image classification model alone did not have sufficient performance, but it should be noted that the predictive performance of CXR for extubation failure was as good as that of RSBI, which is widely used to estimate the risk of extubation failure in the ICU. In addition, a CXR before extubation may still be useful for estimating the risk of extubation failure by observing longitudinal changes in cardiopulmonary conditions through repeated CXRs of a patient. Whether the changes in CXRs over time could be a good predictor for extubation failure was not verified in this study, and further studies on the utility of longitudinal CXRs are warranted.

CONCLUSIONS

Using the data on mechanically ventilated patients from the MIMIC-IV database, we found no significant qualitative differences in the predictive performance of extubation failure between the machine learning-based prediction models using clinical parameters with and without CXR. A cross-sectional preextubation CXR may not contribute to the prediction of extubation failure.

ACKNOWLEDGMENT

We thank Prof. Satoru Hashimoto (Kyoto Prefectural University of Medicine, Japan) and the International Mentoring Team of SG Healthcare AI Datathon & Expo 2020 for their administrative support for our research.

- 3 Department of Medicine, The University of Tokyo Hospital, Tokyo, Japan.
- 4 Nursing Course, College of Nursing, School of Medicine and Health Sciences, University of Tsukuba, Ibaraki, Japan.
- 5 Department of Neuroscience of Disease, Brain Research Institute, Niigata University, Niigata, Japan.
- 6 Department of Emergency and Critical Care Medicine, Nagoya University Graduate School of Medicine, Aichi, Japan.
- 7 National Cancer Center, Institute for Cancer Control, Tokyo, Japan.
- 8 Department of Clinical Epidemiology and Health Economics, School of Public Health, The University of Tokyo, Tokyo, Japan.
- 9 Department of Medical Statistics, Graduate School of Medicine, Osaka City University, Osaka, Japan.
- 10 Department of Emergency Medicine, University of Fukui Hospital, Fukui, Japan.
- 11 DOWELL Co., Ltd., Tokyo, Japan.
- 12 TXP Medical Co. Ltd., Tokyo, Japan.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (<http://journals.lww.com/ccejjournal>).

Mr. Fukuchi and Dr. Osawa contributed equally as cofirst authors.

Drs. Osawa and Goto conceived and designed the study. All authors interpreted the data, critically revised the article for important intellectual content, and approved the final article. Mr. Fukuchi, Dr. Satake, and Ms. Ito performed the statistical analyses. Mr. Fukuchi, Dr. Osawa, and Dr. Goto drafted the initial article. Dr. Dohi, Dr. Kasugai, Dr. Miyamoto, Dr. Ohbe, Mr. Tamoto, Dr. Yamada, Mr. Yoshikawa, and Dr. Goto supervised the study. Mr. Fukuchi, Dr. Osawa, and Dr. Goto are the guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

The authors have disclosed that they do not have any potential conflicts of interest.

Address requests for reprints to: Itsuki Osawa, MD, Department of Emergency and Critical Care Medicine, The University of Tokyo Hospital, 7-3-1, Hongo, Bunkyo-ku, Tokyo 1138655, Japan. E-mail: ioosawa-ky@umin.ac.jp

Given that the clinical data used for this research were obtained from the Medical Information Mart for Intensive Care IV and Medical Information Mart for Intensive Care Chest X-ray databases, publicly available and deidentified database based on the HIPAA Safe Harbor provision, the TXP Medical Ethical Review Board waived the requirement for the ethical approval statement and informed consent (TXPREC-008).

The Medical Information Mart for Intensive Care IV (MIMIC-IV) dataset Version 1.0 is an extensive, publicly available database consisting of deidentified health-related data. Information on how to access the MIMIC-IV database may be found at <https://physionet.org/content/mimiciv/1.0/>. The Medical Information Mart for Intensive Care Chest X-ray (MIMIC-CXR) dataset is a large publicly available dataset of CXRs in DICOM format with free-text

- 1 Faculty of Medicine, The University of Tokyo, Tokyo, Japan.
- 2 Department of Emergency and Critical Care Medicine, The University of Tokyo Hospital, Tokyo, Japan.

radiology reports. Information on how to access the MIMIC-CXR database may be found at <https://physionet.org/content/mimic-cxr/2.0.0/>.

REFERENCES

- Epstein SK, Ciubotaru RL, Wong JB: Effect of failed extubation on the outcome of mechanical ventilation. *Chest* 1997; 112:186–192
- Seymour CW, Martinez A, Christie JD, et al: The outcome of extubation failure in a community hospital intensive care unit: A cohort study. *Crit Care* 2004; 8:R322–R327
- Frutos-Vivar F, Esteban A, Apezteguia C, et al: Outcome of reintubated patients after scheduled extubation. *J Crit Care* 2011; 26:502–509
- Thille AW, Harrois A, Schortgen F, et al: Outcomes of extubation failure in medical intensive care unit patients. *Crit Care Med* 2011; 39:2612–2618
- Khamiees M, Raju P, DeGirolamo A, et al: Predictors of extubation outcome in patients who have successfully completed a spontaneous breathing trial. *Chest* 2001; 120:1262–1270
- Namen AM, Ely EW, Tatter SB, et al: Predictors of successful extubation in neurosurgical patients. *Am J Respir Crit Care Med* 2001; 163:658–664
- Krishnan JA, Moore D, Robeson C, et al: A prospective, controlled trial of a protocol-based strategy to discontinue mechanical ventilation. *Am J Respir Crit Care Med* 2004; 169:673–678
- Navalesi P, Frigerio P, Moretti MP, et al: Rate of reintubation in mechanically ventilated neurosurgical and neurologic patients: Evaluation of a systematic approach to weaning and extubation. *Crit Care Med* 2008; 36:2986–2992
- Nitta K, Okamoto K, Imamura H, et al: A comprehensive protocol for ventilator weaning and extubation: A prospective observational study. *J Intensive Care* 2019; 7:50
- Ganapathy A, Adhikari NK, Spiegelman J, et al: Routine chest x-rays in intensive care units: A systematic review and meta-analysis. *Crit Care* 2012; 16:R68
- Rackley CR, Levitt JE, Zhuo H, et al: Clinical evidence of early acute lung injury often precedes the diagnosis of ALI. *J Intensive Care Med* 2013; 28:241–246
- Hejblum G, Ios V, Vibert JF, et al: A web-based Delphi study on the indications of chest radiographs for patients in ICUs. *Chest* 2008; 133:1107–1112
- Johnson A, Bulgarelli L, Pollard T, et al: MIMIC-IV (Version 1.0). PhysioNet. 2021. Available at: <https://doi.org/10.13026/s6n6-xd98>. Accessed May 11, 2022
- Johnson A, Pollard T, Mark R, et al: MIMIC-CXR Database (Version 2.0.0). PhysioNet. 2019. Available at: <https://doi.org/10.13026/C2JT1Q>. Accessed May 11, 2022
- Yang KL, Tobin MJ: A prospective study of indexes predicting the outcome of trials of weaning from mechanical ventilation. *N Engl J Med* 1991; 324:1445–1450
- Frutos-Vivar F, Ferguson ND, Esteban A, et al: Risk factors for extubation failure in patients following a successful spontaneous breathing trial. *Chest* 2006; 130:1664–1671
- Komorowski M, Celi LA, Badawi O, et al: The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med* 2018; 24:1716–1720
- Saugel B, Rakette P, Hapfelmeier A, et al: Prediction of extubation failure in medical intensive care unit patients. *J Crit Care* 2012; 27:571–577
- Hsieh MH, Hsieh MJ, Chen CM, et al: Comparison of machine learning models for the prediction of mortality of patients with unplanned extubation in intensive care units. *Sci Rep* 2018; 8:17116
- Hsieh MH, Hsieh MJ, Chen CM, et al: An artificial neural network model for predicting successful extubation in intensive care units. *J Clin Med* 2018; 7:240
- 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
- Hsieh MH, Hsieh MJ, Cheng AC, et al: Predicting weaning difficulty for planned extubation patients with an artificial neural network. *Medicine (Baltimore)* 2019; 98:e17392
- Stekhoven DJ, Bühlmann P: MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012; 28:112–118
- Marcel S, Rodriguez Y: Torchvision the machine-vision package of torch. *In: Proceedings of the International Conference on Multimedia*. 2010, pp 1485
- Tan M, Le QV: Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv*. Preprint posted online May 26, 2019. Available at: <https://arxiv.org/abs/1905.11946>. Accessed May 11, 2022
- Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. *arXiv*. Preprint posted online September 4, 2014. Available at: <http://arxiv.org/abs/1409.1556>
- Szegedy C, Liu W, Jia Y, et al: Going deeper with convolutions. *arXiv*. Preprint posted online September 17, 2014. Available at: <http://arxiv.org/abs/1409.4842>
- He K, Zhang X, Ren S, et al: Deep residual learning for image recognition. *arXiv*. Preprint posted online December 10, 2015. Available at: <http://arxiv.org/abs/1512.03385>
- Huang G, Liu Z, van der Maaten L, et al: Densely connected convolutional networks. *arXiv*. Preprint posted online January 28, 2018. Available at: <http://arxiv.org/abs/1608.06993>
- Dosovitskiy A, Beyer L, Kolesnikov A, et al: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv*. Preprint posted online October 22, 2020. Available at: <http://arxiv.org/abs/2010.11929>
- Stone M: Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc Series B Methodol* 1974; 36:111–133
- Ke G, Meng Q, Finley T, et al: LightGBM: A highly efficient gradient boosting decision tree. *In: Conference Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017)*. 2017. Available at: <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>
- Niculescu-Mizil A, Caruana R: Predicting good probabilities with supervised learning. *In: ICML'05 Proceedings of the 22nd International Conference on Machine Learning*. 2005, pp 625–632

34. Akiba T, Sano S, Yanase T, et al: Optuna: A Next-generation Hyperparameter Optimization Framework. *arXiv*. Preprint posted online July 25, 2019. Available at <http://arxiv.org/abs/1907.10902>
35. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 1988; 44:837–845
36. Youden WJ: Index for rating diagnostic tests. *Cancer* 1950; 3:32–35
37. Vickers AJ, Elkin EB: Decision curve analysis: A novel method for evaluating prediction models. *Med Decis Making* 2006; 26:565–574
38. Vickers AJ, van Calster B, Steyerberg EW: A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res* 2019; 3:18
39. Lundberg SM, Lee S-I: A unified approach to interpreting model predictions. In: Conference Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017). 2017. Available at <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>. Accessed May 11, 2022
40. Burns KEA, Raptis S, Nisenbaum R, et al: International practice variation in weaning critically ill adults from invasive mechanical ventilation. *Ann Am Thorac Soc* 2018; 15:494–502
41. Trivedi V, Chaudhuri D, Jinah R, et al: The usefulness of the rapid shallow breathing index in predicting successful extubation: A systematic review and meta-analysis. *Chest* 2022; 161:97–111
42. Paszke A, Gross S, Massa F, et al: PyTorch: An imperative style, high-performance deep learning library. *arXiv*. Preprint posted online December 3, 2019. Available at: <http://arxiv.org/abs/1912.01703>
43. Abadi M, Barham P, Chen J, et al: TensorFlow: A system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16). 2016, pp 265–283. Available at: <https://research.google/pubs/pub45381>
44. Pedregosa F, Varoquaux G, Gramfort A, et al: Scikit-learn: Machine learning in python. *J Mach Learn Res* 2011; 12:2825–2830
45. Redmon J, Divvala S, Girshick R, et al: You only look once: Unified, real-time object detection. *arXiv*. Preprint posted online June 8, 2015. Available at <http://arxiv.org/abs/1506.02640>
46. Liu W, Anguelov D, Erhan D, et al: SSD: Single shot MultiBox detector. *arXiv*. Preprint posted online December 29, 2016. Available at: <https://arxiv.org/abs/1512.02325>
47. Carion N, Massa F, Synnaeve G, et al: End-to-end object detection with transformers. *arXiv*. Preprint posted online May 26, 2020. Available at: <http://arxiv.org/abs/2005.12872>
48. Ronneberger O, Fischer P, Brox T: U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv*. Preprint posted online May 18, 2015. Available at: <http://arxiv.org/abs/1505.04597>
49. Chen LC, Papandreou G, Kokkinos I, et al: DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *arXiv*. Preprint posted online June 2, 2016. Available at: <http://arxiv.org/abs/1606.00915>
50. Zhao H, Shi J, Qi X, et al: Pyramid scene parsing network. *arXiv*. Preprint posted online December 4, 2016. Available at: <http://arxiv.org/abs/1612.01105>
51. Bergstra J, Bengio Y: Random Search for hyper-parameter optimization. *J Mach Learn Res* 2012; 13:281–305
52. Shah AD, Bartlett JW, Carpenter J, et al: Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. *Am J Epidemiol* 2014; 179:764–774
53. Osawa I, Goto T, Yamamoto Y, et al: Machine-learning-based prediction models for high-need high-cost patients using nationwide clinical and claims data. *NPJ Digit Med* 2020; 3:148
54. Epstein SK: Putting it all together to predict extubation outcome. *Intensive Care Med* 2004; 30:1255–1257