

Original Paper

Machine Learning Approaches for Predicting Difficult Airway and First-Pass Success in the Emergency Department: Multicenter Prospective Observational Study

Syunsuke Yamanaka¹, MD, PhD; Tadahiro Goto², MD, MPH, PhD; Koji Morikawa³, PhD; Hiroko Watase⁴, MD, MPH; Hiroshi Okamoto⁵, MD, MPH; Yusuke Hagiwara⁶, MD, MPH; Kohei Hasegawa⁷, MD, MPH

¹Department of Emergency Medicine & General Internal Medicine, The University of Fukui, Fukui, Japan

²Department of Clinical Epidemiology & Health Economics, School of Public Health, The University of Tokyo, Tokyo, Japan

³Connect Inc, Tokyo, Japan

⁴Department of Surgery, University of Washington, Seattle, WA, United States

⁵Department of Intensive Care, St. Luke's International Hospital, Tokyo, Japan

⁶Department of Pediatric Emergency and Critical Care Medicine, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan

⁷Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:

Tadahiro Goto, MD, MPH, PhD

Department of Clinical Epidemiology & Health Economics

School of Public Health

The University of Tokyo

7-3-1 Hongo, Bunkyo-ku

Tokyo, 113-8655

Japan

Phone: 81 3 5841 1887

Email: tag695@mail.harvard.edu

Abstract

Background: There is still room for improvement in the modified LEMON (look, evaluate, Mallampati, obstruction, neck mobility) criteria for difficult airway prediction and no prediction tool for first-pass success in the emergency department (ED).

Objective: We applied modern machine learning approaches to predict difficult airways and first-pass success.

Methods: In a multicenter prospective study that enrolled consecutive patients who underwent tracheal intubation in 13 EDs, we developed 7 machine learning models (eg, random forest model) using routinely collected data (eg, demographics, initial airway assessment). The outcomes were difficult airway and first-pass success. Model performance was evaluated using c-statistics, calibration slopes, and association measures (eg, sensitivity) in the test set (randomly selected 20% of the data). Their performance was compared with the modified LEMON criteria for difficult airway success and a logistic regression model for first-pass success.

Results: Of 10,741 patients who underwent intubation, 543 patients (5.1%) had a difficult airway, and 7690 patients (71.6%) had first-pass success. In predicting a difficult airway, machine learning models—except for k-point nearest neighbor and multilayer perceptron—had higher discrimination ability than the modified LEMON criteria (all, $P \leq .001$). For example, the ensemble method had the highest c-statistic (0.74 vs 0.62 with the modified LEMON criteria; $P < .001$). Machine learning models—except k-point nearest neighbor and random forest models—had higher discrimination ability for first-pass success. In particular, the ensemble model had the highest c-statistic (0.81 vs 0.76 with the reference regression; $P < .001$).

Conclusions: Machine learning models demonstrated greater ability for predicting difficult airway and first-pass success in the ED.

(*Interact J Med Res* 2022;11(1):e28366) doi: [10.2196/28366](https://doi.org/10.2196/28366)

KEYWORDS

intubation; machine learning; difficult airway; first-pass success

Introduction

In the emergency department (ED), achieving successful tracheal intubation at the initial attempt (ie, first-pass success) is essential [1]. The literature has shown that repeated intubation attempts are associated with a higher rate of adverse events [2-4]. However, recent studies have also reported first-pass success rates of 74%-84% in the ED [5,6], suggesting that there are still occasions where repeated intubation attempts are required. To improve ED airway management, the development of effective risk stratification and prediction tools is instrumental.

A widely used prediction tool for difficult airway is the modified LEMON (look, evaluate, Mallampati, obstruction, neck mobility) criteria [7], which has been validated [8]. Although the criteria have good prediction ability (eg, sensitivity 86%, specificity 48% for direct laryngoscopy) [8], there remains room for improvement. Besides, no prediction tool accurately predicts first-pass success (or failure) in the ED. The recent advent of machine learning approaches has enabled clinicians and researchers to accurately predict various diseases and conditions, such as sepsis [9], acute asthma [10], and ED triage [11,12]. Compared with conventional prediction tools and regression approaches, modern machine learning approaches have several advantages, such as incorporating high-order, nonlinear interactions between predictors and mitigating overfitting [13]. Despite the clinical and research importance, no study has yet applied modern machine learning approaches to predict a difficult airway in advance of preparing for airway management or to predict first-pass success once the intubation strategy has been determined in the ED.

To address this significant knowledge gap in the literature, using data from a prospective, multicenter study of ED airway management, we aimed to develop machine learning models that accurately predict difficult airway and first-pass success and to compare their performance with conventional approaches.

Methods

Study Design, Setting, and Participants

This study analyzes data from a multicenter, prospective study of emergency airway management—the second Japanese Emergency Airway Network (JEAN-2) study. The details of the study design, setting, participants, methods of data measurement, and definitions of variables have been reported elsewhere [14]. In brief, the JEAN-2 study is a consortium of 13 academic and community EDs, including 10 level I and 3 level II equivalent trauma centers. These EDs are located in different geographic regions across Japan. The median ED census is 29,000 patients per year (range of 16,000 to 67,000 annual visits). These ED are affiliated with an emergency medicine residency training program. Attending physicians or resident physicians who are under the supervision of the attending physician perform intubations. In this observational study, patients were managed at the discretion of treating physicians. The institutional review board at each participating center approved the waiver of informed consent before data collection. This study used data from consecutive (both children and adults) patients who underwent ED management at one of

the participating EDs from January 1, 2010 through December 31, 2018. Patients who underwent surgical intubations at the first attempt were excluded.

Outcomes

The outcomes of interests were difficult airway and first-pass success. According to the American Society of Anesthesiologists (ASA) guidelines, a difficult airway was defined as multiple intubation attempts by emergency physicians or anesthesiologists according to the ASA guidelines [15]. First-pass success was defined as intubation success at the initial attempt of each encounter [16]. Intubation success was defined as the proper placement of a tracheal tube through the vocal cord, confirmed by the use of qualitative or end-tidal CO₂ monitoring [17]. An intubation attempt was defined as a single insertion of the laryngoscope past the teeth [18].

Predictors of Machine Learning Models

To develop machine learning models for the difficult airway outcome, we used the following variables that are routinely obtained in advance of the actual intubation attempt: patient demographics (age, sex, estimated height and body weight, BMI), components of the modified LEMON criteria, pre-intubation vital signs (pulse rate, systolic blood pressure, respiratory rate, oxygen saturation), and Glasgow coma scale. To develop models that predict the first-pass success outcome (once the intubation strategy has been determined), we used all available intubation-related information—in addition to the aforementioned predictors—such as type of day (weekend/weekday), medications, intubation methods, intubation devices, intubator's post-graduate year, and intubator's specialty.

Statistical Analysis

Summary statistics were used to describe the characteristics of patients and airway management. After performing imputations for missing continuous variables (most predictors had <10% missingness; [Multimedia Appendix 1](#)) using random forest [19], we conducted predictor preprocessing, such as one-hot encoding (ie, creation of dummy variables), normalization, and standardization. The nonlinear predictors included in the developed models were age, body weight, height, BMI, and pre-intubation vital signs. In the training set (80% random sample), for each outcome, we developed 7 machine learning models: (1) logistic regression model with elastic-net (penalized logistic regression) [20], (2) random forest [21], (3) gradient boosting decision tree [22], (4) multilayer perceptron [23], (5) k-point nearest neighbor [24], (6) XGBoost [25], and (7) ensemble model (ridge regression and the random forest with an equal weight) [26]. For the difficult airway outcome, the modified LEMON criteria model was used as the reference model. For the first-pass success outcome, a (nonpenalized) logistic regression model was used as the reference model. We performed stratified 5-fold cross-validation to determine the optimal hyperparameters with the highest c-statistic (ie, the area under the receiver operating characteristic [ROC] curve).

In the test set (the remaining 20% of the random sample), we measured the performance of reference and machine learning models. We estimated the c-statistic of each model and examined the following association measures: sensitivity,

specificity, positive and negative predictive values, and positive and negative likelihood ratios. The c-statistic is the probability that, given 2 individuals (one who experiences the outcome of interest and the other who does not), the model estimates a higher probability for the first patient than for the second [27]. We determined the threshold of perspective prediction (cut-off) results based on the ROC curve from the Youden method [28]. For the model with the highest c-statistic among the 7 machine learning models, we computed the variable importance—how strongly each of the predictors improved the c-statistic. We also examined calibration plots of the best-performing machine learning model for each of the outcomes. Data were analyzed using python (version 3.7.3) and R (version 3.6.2). Two-sided *P* values <.05 were considered statistically significant.

Results

Patient Characteristics

During the 108-month study period, the JEAN-2 study recorded data for 10,816 patients (capture rate, 96%) who underwent emergency airway management at one of the 13 participating EDs. Of these, 75 patients who underwent surgical intubation at their first attempt were excluded; the remaining 10,741 patients comprised the analytic cohort. The patient characteristics, details of airway management, and intubation outcomes are shown in Table 1. The median age was 71 (IQR 56–81) years, 2.8% (304/10,741) were children, and 38.0% (4079/10,741) were female. Overall, 5.1% (543/10,741) of patients had a difficult airway outcome, while 71.6% (7690/10,741) had first-pass success. An aborted intubation attempt occurred for 39 patients.

Table 1. Patient characteristics, airway management, and outcomes in 10,741 patients who underwent tracheal intubation in the emergency department.

Variables	Results
Age (years), median (IQR)	71 (56-81)
Children (< 18 years), n (%)	304 (2.8)
Female gender, n (%)	4079 (38.0)
Estimated height (cm), median (IQR)	160 (153-170)
Estimated body weight (kg), median (IQR)	60 (50-67)
BMI (kg/m ²), median (IQR)	22.0 (19.5-24.3)
Primary indication, n (%)	
Medical cardiac arrest	3785 (35.2)
Traumatic cardiac arrest	438 (4.1)
Medical indication	5440 (50.6)
Airway problem (eg, obstruction)	289 (2.7)
Breathing problem (eg, respiratory failure)	1673 (15.6)
Circulation problem (eg, shock)	1080 (10.1)
Altered mental status	2036 (19.0)
Others	360 (3.4)
Traumatic indication	1080 (10.1)
Modified LEMON^a criteria, n (%)	
Look externally	583 (5.0)
3-3-(2) rule	3620 (33.7)
Obstruction	774 (7.2)
Neck mobility	1101 (10.3)
Any criterion met in the modified LEMON criteria	4709 (43.8)
Intubation outcomes, n (%)	
Difficult airway	543 (5.1)
First-pass success	7690 (71.6)

^aLEMON: look, evaluate, Mallampati, obstruction, neck mobility.

Prediction Performance for Difficult Airway Outcomes

Table 2 summarizes the performance of the modified LEMON criteria (reference) and 7 machine learning models when predicting a difficult airway outcome in the ED. Compared with the modified LEMON criteria, the discrimination ability of machine learning models—except for the k-point nearest-neighbor model and multilayer perceptron model—were significantly greater ($P \leq .001$). Among the 7 machine learning models, the ensemble model had the highest c-statistic (0.74, 95% CI 0.67-0.79; Figure 1A), with a sensitivity of 0.67 (95% CI 0.65-0.69), specificity of 0.70 (95% CI 0.68-0.72), positive

predictive value of 0.09 (95% CI 0.08-0.11), and negative predictive value of 0.98 (95% CI 0.97-0.98). Compared with the modified LEMON criteria, which had a specificity of 0.57 (95% CI 0.56-0.58), all machine learning models had higher specificity, with the multilayer perceptron model achieving a specificity of 0.92 (95% CI 0.90-0.93). The calibration plot (Figure 2A)—which indicates how far the predicted risk from the ensemble model deviated from the actual risk—showed that the ensemble model overestimated the risk of the outcome, while there was a positive relationship between the predicted and actual risks, largely due to the class imbalance (ie, difficult airway outcome occurred only in 5% of the sample).

Table 2. Performance of 7 machine learning models and modified LEMON (look, evaluate, Mallampati, obstruction, neck mobility) criteria when predicting difficult airway outcome in the emergency department.

Models	C-statistic ^a (95% CI)	P value	Sensitivity (95% CI)	Specificity (95% CI)	PPV ^b (95% CI)	NPV ^c (95% CI)	PLR ^d (95% CI)	NLR ^e (95% CI)
Modified LEMON criteria (reference)	0.62 (0.60- 0.64)	Reference ^f	0.67 (0.66-0.68)	0.57 (0.56-0.58)	0.08 (0.07- 0.08)	0.97 (0.97- 0.97)	1.57 (1.48- 1.68)	0.57 (0.54- 0.61)
Penalized logistic regression	0.73 (0.68- 0.79)	<.001	0.66 (0.64-0.68)	0.68 (0.66-0.70)	0.09 (0.08- 0.10)	0.98 (0.97- 0.98)	2.05 (1.75- 2.40)	0.51 (0.43- 0.59)
Random forest	0.72 (0.67- 0.77)	<.001	0.58 (0.56-0.60)	0.74 (0.72-0.75)	0.09 (0.08- 0.11)	0.97 (0.97- 0.98)	3.84 (2.50- 5.90)	0.84 (0.55- 1.29)
Gradient boost	0.72 (0.66- 0.77)	.001	0.77 (0.75-0.79)	0.58 (0.56-0.60)	0.08 (0.07- 0.09)	0.98 (0.98- 0.99)	1.84 (1.63- 2.08)	0.39 (0.35- 0.44)
Multilayer percep- tron	0.57 (0.50- 0.63)	.14	0.19 (0.17-0.20)	0.92 (0.90-0.93)	0.09 (0.08- 0.11)	0.96 (0.95- 0.97)	2.24 (1.44- 3.48)	0.89 (0.57- 1.38)
K-point nearest neighbor	0.54 (0.49- 0.61)	.02	0.39 (0.36-0.41)	0.70 (0.68-0.72)	0.06 (0.05- 0.07)	0.96 (0.95- 0.97)	1.30 (1.00- 1.68)	0.87 (0.67- 1.14)
XGBoost	0.72 (0.67- 0.77)	<.001	0.69 (0.67-0.71)	0.60 (0.58-0.62)	0.07 (0.06- 0.09)	0.98 (0.97- 0.98)	1.70 (1.47- 1.97)	0.52 (0.45- 0.61)
Ensemble model ^g	0.74 (0.67- 0.79)	<.001	0.67 (0.65-0.69)	0.70 (0.68-0.72)	0.09 (0.08- 0.11)	0.98 (0.97- 0.98)	2.21 (1.89- 2.58)	0.48 (0.41- 0.56)

^aC-statistic in the modified LEMON was evaluated using 95% CIs.

^bPPV: positive predictive value.

^cNPV: negative predictive value.

^dPLR: positive likelihood ratio.

^eNLR: negative likelihood ratio.

^fComparison of the area under the curve of the reference model (modified LEMON) with that of each machine learning model using the DeLong test.

^gEnsemble prediction model using these machine learning models (that combined ≥ 2 models).

Figure 1. Discrimination ability of the ensemble model and logistic regression (reference) model in predicting intubation outcomes, including (A) difficult airway outcomes and (B) first-pass success outcomes. mLEMON: modified look, evaluate, Mallampati, obstruction, neck mobility model.

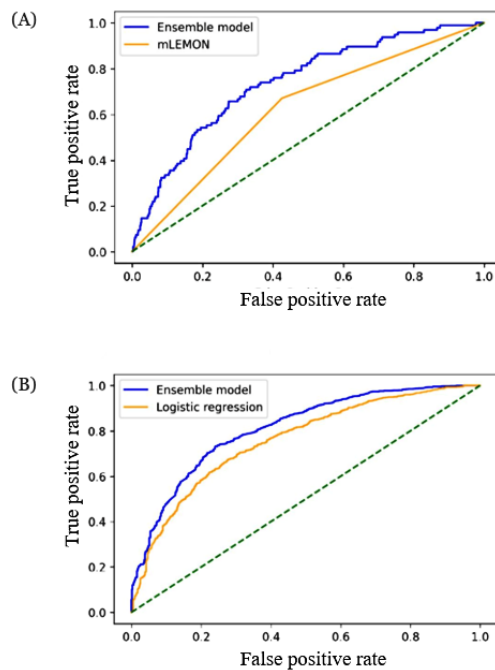
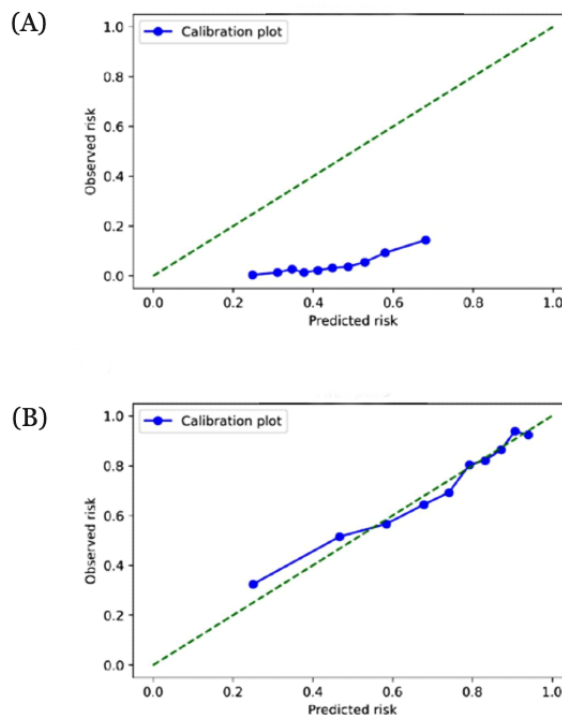


Figure 2. Calibration plots of ensemble models in predicting intubation outcomes, including (A) difficult airway outcomes and (B) first-pass success outcomes.



Prediction Performance for First-Pass Success Outcomes

Table 3 summarizes the performance of the reference model and 7 machine learning models when predicting the first-pass success outcome in the ED. Compared with the reference model, the discrimination ability of machine learning models—except for the random forest and k-point nearest neighbor models—was significantly higher (all $P < .05$). Among the 7 machine learning models, the ensemble model had the highest c-statistic (0.81,

95% CI 0.79-0.83; Figure 1B). Compared with the reference model, the ensemble model had a higher sensitivity (0.79, 95% CI 0.77-0.81) and specificity (0.67, 95% CI 0.65-0.69), with a PPV of 0.85 (95% CI 0.84-0.87) and NPV of 0.57 (95% CI 0.55-0.59). Compared with the reference model, which had a specificity of 0.36 (95% CI 0.34-0.38), most machine learning models had higher specificity, with the random forest model achieving a specificity of 0.70 (95% CI 0.68-0.72). In the calibration plot of the ensemble model (Figure 2B), the

model-predicted probability was well-matched with the observed probabilities.

Table 3. Performance of 7 machine learning models and reference model when predicting first-pass success outcome in the emergency department.

Models	C statistic (95% CI)	P value	Sensitivity (95% CI)	Specificity (95% CI)	PPV ^a (95% CI)	NPV ^b (95% CI)	PLR ^c (95% CI)	NLR ^d (95% CI)
Logistic regression (reference)	0.76 (0.74-0.78)	(Reference) ^e	0.91 (0.89-0.92)	0.36 (0.34-0.38)	0.78 (0.76-0.79)	0.61 (0.59-0.63)	1.42 (1.33-1.51)	0.26 (0.24-0.27)
Penalized logistic regression	0.81 (0.79-0.83)	.001	0.79 (0.77-0.80)	0.70 (0.68-0.72)	0.86 (0.85-0.88)	0.57 (0.55-0.59)	2.59 (2.29-2.92)	0.31 (0.27-0.35)
Random forest	0.78 (0.76-0.81)	.12	0.78 (0.76-0.79)	0.64 (0.62-0.66)	0.84 (0.83-0.86)	0.54 (0.52-0.56)	2.16 (1.94-2.41)	0.35 (0.31-0.39)
Gradient boost	0.80 (0.78-0.82)	.005	0.92 (0.91-0.94)	0.40 (0.38-0.43)	0.79 (0.77-0.81)	0.69 (0.67-0.71)	1.55 (1.45-1.66)	0.19 (0.17-0.20)
Multilayer perceptron	0.81 (0.79-0.83)	.002	0.92 (0.91-0.93)	0.44 (0.42-0.46)	0.80 (0.78-0.82)	0.69 (0.67-0.71)	1.64 (1.53-1.76)	0.18 (0.17-0.19)
K-point nearest neighbor	0.75 (0.73-0.77)	.60	0.98 (0.97-0.98)	0.18 (0.16-0.20)	0.74 (0.73-0.76)	0.78 (0.76-0.80)	1.19 (1.15-1.24)	0.12 (0.11-0.12)
XGBoost	0.81 (0.79-0.83)	<.001	0.94 (0.93-0.95)	0.38 (0.36-0.40)	0.79 (0.77-0.81)	0.73 (0.71-0.75)	1.53 (1.43-1.62)	0.15 (0.14-0.16)
Ensemble model ^f	0.81 (0.79-0.83)	<.001	0.79 (0.77-0.81)	0.67 (0.65-0.69)	0.85 (0.84-0.87)	0.57 (0.55-0.59)	2.39 (2.13-2.67)	0.31 (0.28-0.35)

^aPPV: positive predictive value.

^bNPV: negative predictive value.

^cPLR: positive likelihood ratio.

^dNLR: negative likelihood ratio.

^eComparison of the area under the curve of the reference model with that of each machine learning model using the DeLong test.

^fEnsemble prediction model using these machine learning models (that combined ≥ 2 models).

Variable Importance

Table 4 shows the variable importance of the best performance model (the ensemble model) for each outcome. For the difficult airway prediction, the most contributing predictor was age,

followed by any criterion met in the modified LEMON criteria and hyoid mental distance ≥ 3 fingers. For the first-pass success prediction, the best contributing predictor was the use of laryngeal pressure, followed by lifting force required for laryngeal deployment and Cormack grade of 3.

Table 4. Importance of each predictor of ensemble model when predicting difficult airway and first-pass success outcomes.

Predictors	Δ c-statistics ^a
Predictors for difficult airway outcome	
Age	0.093
Any modified LEMON ^b criterion met	0.093
Hyoid mental distance ≥ 3 fingers	0.091
Interincisor distance of 3 fingers	0.084
BMI	0.080
Body weight	0.073
>80 years old	0.070
Hyoid mental distance of 2 fingers	0.053
Airway obstruction	0.049
Neck mobility	0.048
Predictors for first-pass success outcome	
Use of laryngeal pressure	0.118
Lifting force required for laryngeal deployment	0.108
Cormack grade of 3	0.099
Any modified LEMON criterion met	0.094
Cormack grade of 1	0.094
Intubator's post-graduation year of 1 or 2	0.090
Neuromuscular blocking agent (rocuronium)	0.077
Rapid sequence intubation	0.076
Video laryngoscope (C-MAC)	0.075
Video Cormack grade of 1	0.074
Interincisor distance ≥ 3 fingers	0.079

^aThe variable importance of a predictor is agnostic regarding the direction of the association.

^bLEMON: look, evaluate, Mallampati, obstruction, neck mobility.

Discussion

Principal Findings

In this analysis of multicenter prospective data from 10,741 ED patients, we applied modern machine learning models to predict intubation-related outcomes in the ED. Specifically, compared with conventional approaches (ie, modified LEMON criteria and nonpenalized logistic regression model), most machine learning models demonstrated superior discrimination performance when predicting both difficult airway and first-pass success outcomes. Additionally, these machine learning models also achieved higher specificity when predicting these 2 outcomes. To the best of our knowledge, this is the first study that has investigated the performance of modern machine learning models when predicting clinically important intubation outcomes in the ED setting.

Consistent with our findings, the following has been reported as predictors for first-pass success in the ED: patient characteristics (eg, restricted mouth opening, restricted neck extension, and swollen tongue), high Cormack grade, intubators' characteristics (eg, clinical experience and working department),

the use of rapid-sequence intubation, and the use of video laryngoscope at the first attempt [6,29-31].

The importance of accurate prediction for difficult airways has been emphasized in ED airway management [8]. Although the modified LEMON criteria (and the LEMON criteria) have been validated as an indicator for difficult airways, their prediction ability is suboptimal for clinical use [7,8]. In the operating room setting, a couple of studies have reported a potential benefit of machine learning models for predicting difficult airways [32,33]. For example, in a single-center study of 80 patients, a deep learning approach using data from the patients' facial images had high discrimination ability for difficult airways—defined as multiple attempts by an intubator with at least 12 months of anesthesia experience, grade 3 or 4 laryngoscopic view, need for a second intubator, or nonelective use of an alternative airway device [32]. Our multicenter study—with a sample size that is many times larger than the prior studies on this topic—builds on these earlier reports and extends them by demonstrating that modern machine learning models outperform conventional approaches for predicting intubation outcomes in the ED.

The observed improvement in prediction ability by machine learning approaches may be explained by several reasons. First, the machine learning approaches account for high-order interactions between predictors and nonlinear relationships with an outcome, which traditional modeling approaches cannot address [34]. Second, the modified LEMON criteria may be too parsimonious (ie, the use of a limited number of predictors), while the applied machine learning models could use a larger number of predictors. Third, the modern machine learning approaches enabled us to minimize overfitting, such as lasso and ridge penalizations (ie, elastic net model and cross-validation). In addition to these strengths, modern machine learning models also are scalable for further improvement by integration with recently developed techniques such as image analysis of patients' faces and necks [32,35].

Although the machine learning models achieved a more significant predictive ability, their performance remained imperfect. This may be explained, at least partially, by the limited set of predictors (eg, lack of detailed information on the intubation competency and experience of each intubator) and data measurement errors. Additionally, one may surmise that the modified LEMON criteria are simpler and easier to use in the ED. Despite the known trade-off between parsimonious models and more complex models with a larger number of predictors, the use of modern machine learning models has advantages in the era of health information technology, including automated data entry through voice recognition, natural language processing, continuous sophistication of models through sequential extractions of electronic health records, and reinforcement learning [36,37]. Our findings and the recent advent of machine learning approaches collectively support cautious optimism that machine learning may enhance the clinician's ability—as assistive technology—to predict patient outcomes in the ED. The resulting accurate prediction of intubation outcomes has several important implications in airway practice in the ED. For example, early identification of difficult airways should help ED providers develop individualized and optimal management strategies and prepare for rescue airways [14,38]. Besides, the accurate estimation of the probability of first-pass success given the conditions (eg, the airway management strategies and intubator to be used) would not only increase the opportunity for clinical training (eg, which patient can be safely intubated by the intubator) but also improve patient safety.

To implement our developed machine learning models, a web-based application or integrated emergency department information system is needed. The rapid development of health information technology (eg, web-based artificial intelligence application with the model) enables us to implement the developed model into the real clinical setting. Furthermore, the current models can be used not only for practice but as an educational tool. For example, in simulation-based intubation training, supervisors can evaluate the trainee's intubation strategy by indicating the actual probability of difficult airway and first-pass success.

Limitations

Several potential limitations of this study should be noted. First, our data may be subject to self-reporting and measurement bias (eg, underreporting difficult airways). However, the study was conducted by investigators using a standardized protocol [6], which led to the high capture rate (96%) and low proportion of missingness in the predictors and outcomes (Multimedia Appendix 1). Second, we did not have detailed information on the procedural competency of each intubator, as this factor is also challenging to define and measure in real-world settings. To address this issue, we used years of experience and specialty, which are readily available in most ED settings, as a proxy for the competency. Third, machine learning models have a common limitation in the interpretability of models. Fourth, because of the small samples of children (2.8%), our model may not have optimal prediction ability in pediatric populations. Finally, our models may not be generalizable to other practice settings, although the study sample consisted of a geographically diverse patient across Japan.

Conclusions

In summary, based on the extensive multicenter, prospective data from 10,741 ED intubations, we developed modern machine learning models to predict clinically essential intubation outcomes. Using routinely available data as the predictors, we found that the machine learning models had a greater ability to predict difficult airways and first-pass success than conventional approaches. Although formal validation is required, this study lends support to the application of machine learning models for the prediction of intubation-related outcomes, which will, in turn, improve airway management practice and outcomes of critically ill patients in the ED.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Proportion of missingness in predictors used in machine learning models.

[\[DOCX File , 17 KB-Multimedia Appendix 1\]](#)

References

1. Goto T, Goto Y, Hagiwara Y, Okamoto H, Watase H, Hasegawa K. Advancing emergency airway management practice and research. *Acute Med Surg* 2019 Oct;6(4):336-351 [FREE Full text] [doi: [10.1002/ams2.428](https://doi.org/10.1002/ams2.428)] [Medline: [31592072](https://pubmed.ncbi.nlm.nih.gov/31592072/)]
2. Hasegawa K, Shigemitsu K, Hagiwara Y, Chiba T, Watase H, Brown CA, Japanese Emergency Medicine Research Alliance Investigators. Association between repeated intubation attempts and adverse events in emergency departments: an analysis

- of a multicenter prospective observational study. *Ann Emerg Med* 2012 Dec;60(6):749-754.e2. [doi: [10.1016/j.annemergmed.2012.04.005](https://doi.org/10.1016/j.annemergmed.2012.04.005)] [Medline: [22542734](https://pubmed.ncbi.nlm.nih.gov/22542734/)]
3. Walls RM, Brown CA, Bair AE, Pallin DJ, NEAR II Investigators. Emergency airway management: a multi-center report of 8937 emergency department intubations. *J Emerg Med* 2011 Oct;41(4):347-354. [doi: [10.1016/j.jemermed.2010.02.024](https://doi.org/10.1016/j.jemermed.2010.02.024)] [Medline: [20434289](https://pubmed.ncbi.nlm.nih.gov/20434289/)]
 4. Pallin DJ, Dwyer RC, Walls RM, Brown CA, NEAR III Investigators. Techniques and trends, success rates, and adverse events in emergency department pediatric intubations: a report from the National Emergency Airway Registry. *Ann Emerg Med* 2016 May;67(5):610-615.e1. [doi: [10.1016/j.annemergmed.2015.12.006](https://doi.org/10.1016/j.annemergmed.2015.12.006)] [Medline: [26921968](https://pubmed.ncbi.nlm.nih.gov/26921968/)]
 5. Park L, Zeng I, Brainard A. Systematic review and meta-analysis of first-pass success rates in emergency department intubation: Creating a benchmark for emergency airway care. *Emerg Med Australas* 2017 Feb;29(1):40-47. [doi: [10.1111/1742-6723.12704](https://doi.org/10.1111/1742-6723.12704)] [Medline: [27785883](https://pubmed.ncbi.nlm.nih.gov/27785883/)]
 6. Goto Y, Goto T, Hagiwara Y, Tsugawa Y, Watase H, Okamoto H, Japanese Emergency Medicine Network Investigators. Techniques and outcomes of emergency airway management in Japan: An analysis of two multicentre prospective observational studies, 2010-2016. *Resuscitation* 2017 May;114:14-20. [doi: [10.1016/j.resuscitation.2017.02.009](https://doi.org/10.1016/j.resuscitation.2017.02.009)] [Medline: [28219617](https://pubmed.ncbi.nlm.nih.gov/28219617/)]
 7. Reed MJ, Dunn MJG, McKeown DW. Can an airway assessment score predict difficulty at intubation in the emergency department? *Emerg Med J* 2005 Feb;22(2):99-102 [FREE Full text] [doi: [10.1136/emj.2003.008771](https://doi.org/10.1136/emj.2003.008771)] [Medline: [15662057](https://pubmed.ncbi.nlm.nih.gov/15662057/)]
 8. Hagiwara Y, Watase H, Okamoto H, Goto T, Hasegawa K, Japanese Emergency Medicine Network Investigators. Prospective validation of the modified LEMON criteria to predict difficult intubation in the ED. *Am J Emerg Med* 2015 Oct;33(10):1492-1496. [doi: [10.1016/j.ajem.2015.06.038](https://doi.org/10.1016/j.ajem.2015.06.038)] [Medline: [26166379](https://pubmed.ncbi.nlm.nih.gov/26166379/)]
 9. Islam MM, Nasrin T, Walther BA, Wu C, Yang H, Li Y. Prediction of sepsis patients using machine learning approach: A meta-analysis. *Comput Methods Programs Biomed* 2019 Mar;170:1-9. [doi: [10.1016/j.cmpb.2018.12.027](https://doi.org/10.1016/j.cmpb.2018.12.027)] [Medline: [30712598](https://pubmed.ncbi.nlm.nih.gov/30712598/)]
 10. Goto T, Camargo CA, Faridi MK, Yun BJ, Hasegawa K. Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. *Am J Emerg Med* 2018 Sep;36(9):1650-1654. [doi: [10.1016/j.ajem.2018.06.062](https://doi.org/10.1016/j.ajem.2018.06.062)] [Medline: [29970272](https://pubmed.ncbi.nlm.nih.gov/29970272/)]
 11. Goto T, Camargo CA, Faridi MK, Freishtat RJ, Hasegawa K. Machine learning-based prediction of clinical outcomes for children during emergency department triage. *JAMA Netw Open* 2019 Jan 04;2(1):e186937 [FREE Full text] [doi: [10.1001/jamanetworkopen.2018.6937](https://doi.org/10.1001/jamanetworkopen.2018.6937)] [Medline: [30646206](https://pubmed.ncbi.nlm.nih.gov/30646206/)]
 12. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care* 2019 Feb 22;23(1):64 [FREE Full text] [doi: [10.1186/s13054-019-2351-7](https://doi.org/10.1186/s13054-019-2351-7)] [Medline: [30795786](https://pubmed.ncbi.nlm.nih.gov/30795786/)]
 13. Kotsiantis S, Zaharakis I, Pintelas P. Supervised machine learning: A review of classification techniques. In: Maglogiannis I, Karpouzis K, Wallace BA, Soldatos J, editors. *Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*. Amsterdam, Netherlands: IOS Press; 2007:3-27.
 14. Goto T, Gibo K, Hagiwara Y, Morita H, Brown DF, Brown CA, Japanese Emergency Medicine Network Investigators. Multiple failed intubation attempts are associated with decreased success rates on the first rescue intubation in the emergency department: a retrospective analysis of multicentre observational data. *Scand J Trauma Resusc Emerg Med* 2015 Jan 16;23(1):5 [FREE Full text] [doi: [10.1186/s13049-014-0085-8](https://doi.org/10.1186/s13049-014-0085-8)] [Medline: [25700237](https://pubmed.ncbi.nlm.nih.gov/25700237/)]
 15. Apfelbaum J, Hagberg C, Caplan R, Blitt C, Connis R, Nickinovich D, American Society of Anesthesiologists Task Force on Management of the Difficult Airway. Practice guidelines for management of the difficult airway: an updated report by the American Society of Anesthesiologists Task Force on Management of the Difficult Airway. *Anesthesiology* 2013 Feb;118(2):251-270 [FREE Full text] [doi: [10.1097/ALN.0b013e31827773b2](https://doi.org/10.1097/ALN.0b013e31827773b2)] [Medline: [23364566](https://pubmed.ncbi.nlm.nih.gov/23364566/)]
 16. Goto T, Gibo K, Hagiwara Y, Okubo M, Brown DFM, Brown CA, et al. Factors associated with first-pass success in pediatric intubation in the emergency department. *West J Emerg Med* 2016 Mar;17(2):129-134 [FREE Full text] [doi: [10.5811/westjem.2016.1.28685](https://doi.org/10.5811/westjem.2016.1.28685)] [Medline: [26973736](https://pubmed.ncbi.nlm.nih.gov/26973736/)]
 17. Link MS, Berkow LC, Kudenchuk PJ, Halperin HR, Hess EP, Moitra VK, et al. Part 7: adult advanced cardiovascular life support. *Circulation* 2015 Nov 03;132(18_suppl_2):S444-S464. [doi: [10.1161/cir.0000000000000261](https://doi.org/10.1161/cir.0000000000000261)]
 18. Hasegawa K, Hagiwara Y, Chiba T, Watase H, Walls RM, Brown DF, Japanese Emergency Medicine Research Alliance. Emergency airway management in Japan: Interim analysis of a multi-center prospective observational study. *Resuscitation* 2012 Apr;83(4):428-433. [doi: [10.1016/j.resuscitation.2011.11.027](https://doi.org/10.1016/j.resuscitation.2011.11.027)] [Medline: [22155701](https://pubmed.ncbi.nlm.nih.gov/22155701/)]
 19. Stekhoven D. Nonparametric missing value imputation using random forest. 2012. URL: <https://ui.adsabs.harvard.edu/abs/2015ascl.soft05011S> [accessed 2021-01-08]
 20. Warton DI. Penalized normal likelihood and ridge regularization of correlation and covariance matrices. *Journal of the American Statistical Association* 2012 Jan 01;103(481):340-349. [doi: [10.1198/016214508000000021](https://doi.org/10.1198/016214508000000021)]
 21. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci* 2003 Nov 04;43(6):1947-1958. [doi: [10.1021/ci034160g](https://doi.org/10.1021/ci034160g)] [Medline: [14632445](https://pubmed.ncbi.nlm.nih.gov/14632445/)]

22. Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Front Neurobot* 2013;7:21 [FREE Full text] [doi: [10.3389/fnbot.2013.00021](https://doi.org/10.3389/fnbot.2013.00021)] [Medline: [24409142](https://pubmed.ncbi.nlm.nih.gov/24409142/)]
23. Pal S, Mitra S. Multilayer perceptron, fuzzy sets, and classification. *IEEE Trans Neural Netw* 1992;3(5):683-697. [doi: [10.1109/72.159058](https://doi.org/10.1109/72.159058)] [Medline: [18276468](https://pubmed.ncbi.nlm.nih.gov/18276468/)]
24. Bay S. Nearest neighbor classification from multiple feature subsets. *Intelligent Data Analysis* 1999 Sep;3(3):191-209. [doi: [10.1016/s1088-467x\(99\)00018-9](https://doi.org/10.1016/s1088-467x(99)00018-9)]
25. Chen T, He T, Benesty M, Khotilovich V, Tang Y. Xgboost: extreme gradient boosting. R package version 2015:04 [FREE Full text]
26. Ruta D, Gabrys B. Classifier selection for majority voting. *Information Fusion* 2005 Mar;6(1):63-81. [doi: [10.1016/j.inffus.2004.04.008](https://doi.org/10.1016/j.inffus.2004.04.008)]
27. Pencina MJ, D'Agostino RB. Evaluating discrimination of risk prediction models: the c statistic. *JAMA* 2015 Sep 08;314(10):1063-1064. [doi: [10.1001/jama.2015.11082](https://doi.org/10.1001/jama.2015.11082)] [Medline: [26348755](https://pubmed.ncbi.nlm.nih.gov/26348755/)]
28. Taner T, Antony J. The assessment of quality in medical diagnostic tests: a comparison of ROC/Youden and Taguchi methods. *Int J Health Care Qual Assur Inc Leadersh Health Serv* 2000;13(6-7):300-307. [doi: [10.1108/09526860010378744](https://doi.org/10.1108/09526860010378744)] [Medline: [11484648](https://pubmed.ncbi.nlm.nih.gov/11484648/)]
29. Sagarin MJ, Barton ED, Chng Y, Walls RM, National Emergency Airway Registry Investigators. Airway management by US and Canadian emergency medicine residents: a multicenter analysis of more than 6,000 endotracheal intubation attempts. *Ann Emerg Med* 2005 Oct;46(4):328-336. [doi: [10.1016/j.annemergmed.2005.01.009](https://doi.org/10.1016/j.annemergmed.2005.01.009)] [Medline: [16187466](https://pubmed.ncbi.nlm.nih.gov/16187466/)]
30. Jung W, Kim J. Factors associated with first-pass success of emergency endotracheal intubation. *Am J Emerg Med* 2020 Jan;38(1):109-113. [doi: [10.1016/j.ajem.2019.09.001](https://doi.org/10.1016/j.ajem.2019.09.001)] [Medline: [31843066](https://pubmed.ncbi.nlm.nih.gov/31843066/)]
31. Taboada M, Soto-Jove R, Mirón P, Martínez S, Rey R, Ferreira E, et al. Evaluation of the laryngoscopy view using the modified Cormack-Lehane scale during tracheal intubation in an intensive care unit. A prospective observational study. *Rev Esp Anesthesiol Reanim (Engl Ed)* 2019 May;66(5):250-258. [doi: [10.1016/j.redar.2019.01.004](https://doi.org/10.1016/j.redar.2019.01.004)] [Medline: [30862397](https://pubmed.ncbi.nlm.nih.gov/30862397/)]
32. Connor C, Segal S. Accurate classification of difficult intubation by computerized facial analysis. *Anesth Analg* 2011 Jan;112(1):84-93. [doi: [10.1213/ANE.0b013e31820098d6](https://doi.org/10.1213/ANE.0b013e31820098d6)] [Medline: [21081769](https://pubmed.ncbi.nlm.nih.gov/21081769/)]
33. Moustafa MA, El-Metainy S, Mahar K, Mahmoud Abdel-magied E. Defining difficult laryngoscopy findings by using multiple parameters: A machine learning approach. *Egyptian Journal of Anaesthesia* 2019 May 17;33(2):153-158. [doi: [10.1016/j.ejga.2017.02.002](https://doi.org/10.1016/j.ejga.2017.02.002)]
34. Kuhn M, Johnson K. Applied predictive modeling. New York, NY: Springer; 2013.
35. An G, Omodaka K, Tsuda S, Shiga Y, Takada N, Kikawa T, et al. Comparison of machine-learning classification models for glaucoma management. *J Healthc Eng* 2018;2018:6874765 [FREE Full text] [doi: [10.1155/2018/6874765](https://doi.org/10.1155/2018/6874765)] [Medline: [30018755](https://pubmed.ncbi.nlm.nih.gov/30018755/)]
36. Freeman MB. Method and apparatus for automated data entry. *Justia Patents*. 2001. URL: <https://patents.justia.com/patent/6299063> [accessed 2022-01-14]
37. Thanaki J. Python natural language processing. Birmingham, United Kingdom: Packt Publishing Ltd; 2017.
38. Lockey D, Crewdson K, Weaver A, Davies G. Observational study of the success rates of intubation and failed intubation airway rescue techniques in 7256 attempted intubations of trauma patients by pre-hospital physicians. *Br J Anaesth* 2014 Aug;113(2):220-225 [FREE Full text] [doi: [10.1093/bja/aeu227](https://doi.org/10.1093/bja/aeu227)] [Medline: [25038154](https://pubmed.ncbi.nlm.nih.gov/25038154/)]

Abbreviations

ASA: American Society of Anesthesiologists

ED: emergency department

JEAN-2: second Japanese Emergency Airway Network

ROC: receiver operating characteristic

Edited by A Mavragani; submitted 03.03.21; peer-reviewed by A Miller, A Beane, Y Katayama; comments to author 16.04.21; revised version received 07.06.21; accepted 06.12.21; published 25.01.22

Please cite as:

Yamanaka S, Goto T, Morikawa K, Watase H, Okamoto H, Hagiwara Y, Hasegawa K

Machine Learning Approaches for Predicting Difficult Airway and First-Pass Success in the Emergency Department: Multicenter Prospective Observational Study

Interact J Med Res 2022;11(1):e28366

URL: <https://www.i-jmr.org/2022/1/e28366>

doi: [10.2196/28366](https://doi.org/10.2196/28366)

PMID:

©Syunsuke Yamanaka, Tadahiro Goto, Koji Morikawa, Hiroko Watase, Hiroshi Okamoto, Yusuke Hagiwara, Kohei Hasegawa. Originally published in the Interactive Journal of Medical Research (<https://www.i-jmr.org/>), 25.01.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.i-jmr.org/>, as well as this copyright and license information must be included.