Archives of Anesthesiology and Critical Care (Mar-Apr 2026); 12(2): 181-186.

Available online at http://aacc.tums.ac.ir

Review Article

Machine Learning Revolution in Predicting Difficult Intubation: A Systematic Review

Parisa Moradimajd[1],Alireza Babajani[2], Fatemeh Mehdipour[3], Mahdi Nazari[3]*

[1]*Department of Anesthesia, Faculty of Allied Medical Sciences, Iran University of Medical Sciences, Tehran, Iran.*

[2]*Department of Anesthesiology, School of Allied Medical Sciences, Alborz University of Medical Sciences, Karaj, Iran.*

[3]*Department of Anesthesia, School of Allied Medical Sciences, Iran University of Medical Sciences, Tehran, Iran.*

ARTICLE INFO

Article history:
Received 11 July 2025
Revised 01 August 2025
Accepted 22 August 2025

Keywords:
Difficult airway;
Machine learning;
Intubation prediction;
Laryngoscopy;
Artificial intelligence

ABSTRACT

Background: The presence of a difficult airway (DA) remains a major concern in anesthesia, contributing significantly to patient complications and adverse outcomes. Traditional clinical assessments often fall short in accurately predicting difficult intubation. With the advancement of artificial intelligence, machine learning (ML) has emerged as a promising approach for enhancing airway risk prediction. This systematic review aimed to evaluate current studies that utilize machine learning models for predicting difficult laryngoscopy and intubation and to assess the features, algorithms, and predictive performance of these models.

Methods: Following PRISMA guidelines, a comprehensive search was conducted in seven databases (PubMed, Scopus, Web of Science, Science Direct, Wiley, SID, and Google Scholar) to identify relevant original articles published between 2000 and July 2025. Studies using ML models to predict difficult intubation based on clinical, morphological, or acoustic features were included. A total of nine eligible studies were reviewed.

Results: Various ML algorithms, including KNN, SVM, Random Forest, XGBoost, and decision trees (J48), were applied across studies. Feature inputs ranged from traditional clinical parameters (e.g., Mallampati score, neck circumference) to advanced modalities such as voice analysis and facial image processing. Reported model performance (AUC) ranged from 0.71 to 0.924, indicating generally high predictive accuracy. Models incorporating non-traditional data (e.g., acoustic or imaging features) tended to perform better.

Conclusion: ML-based models show strong potential in improving the prediction of difficult airways and can serve as supportive tools in preoperative assessment. However, standardization of input features, external validation, and enhanced model interpretability are essential for successful clinical implementation.

Introduction

The term Difficult Airway (DA) refers to situations where an anesthesiologist faces problems with mask ventilation, intubation, or both" [1-2]. Even with improvements in anesthetic practice and technology, the presence of DA during intubation continues to result in significant complications and contributes to almost one-third of anesthesia-related mortality [3].

Traditional assessment of a patient's airway is conducted through two main methods: medical history and clinical examinations at the bedside. Medical history plays a key role in identifying potential difficulties in the airway, as certain diseases and specific conditions are

The authors declare no conflicts of interest.
*Corresponding author.
E-mail address: mahdinazari200079@gmail.com
DOI: 10.18502/aacc.v12i2.20958

clearly associated with difficult airways. For example, congenital disorders affecting the shape of the face or mouth, rheumatoid arthritis, acromegaly, a history of radiation therapy in the head and neck regions, and obstructive sleep apnea syndrome are among these conditions. Subsequently, the physician's clinical examinations include assessing the patient's facial and jaw characteristics, such as the degree of mouth opening, the status of the anterior teeth, the Mallampati classification, and the upper lip bite test (ULBT). Additionally, the physician performs simpler anatomical measurements such as the hyomental distance, sternomental distance, distance between the anterior teeth, and neck circumference to better evaluate the condition [3-4].

Machine learning, as part of artificial intelligence, focuses on designing algorithms and statistical models. This technology allows computers to learn from their experiences and improve their performance rather than relying solely on explicit programming. The origins of machine learning date back to the 1950s when early neural networks and perceptrons were developed. With technological advancements and the increasing availability of data, this field has experienced significant progress, and machine learning algorithms are now widely used across various industries and applications [5]. Currently, machine learning is extensively utilized in airway assessment and the prediction of difficult intubation [6-8].

The research questions that guided this review are as follows:

- What machine learning (ML) approaches have been used in various studies to predict laryngoscopy and difficult intubation?
- What morphological features or clinical variables have been used as predictors in these studies?
- How accurate and sensitive are the different models in predicting difficult intubation?

Methods

This systematic review was carried out in accordance with the PRISMA reporting guidelines (Figure 1).

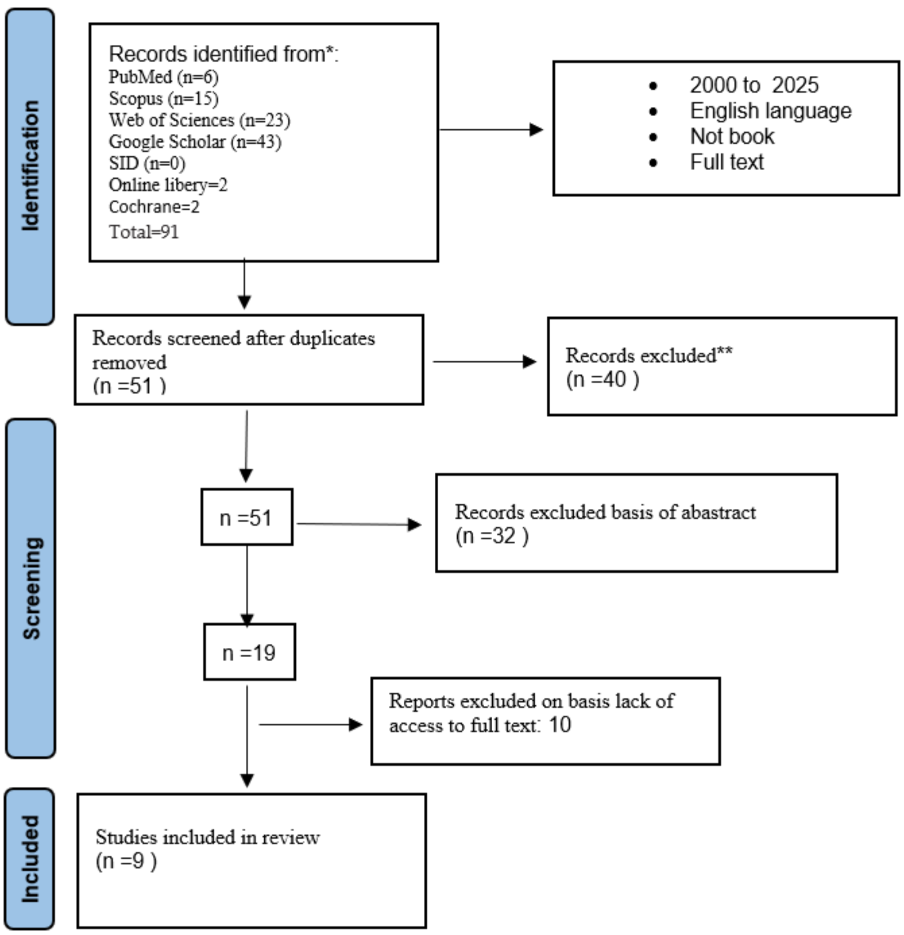

Figure 1- Flowchart of article selection process PRISMA 2020

A comprehensive search was performed across seven databases, including PubMed, Cochrane Library, Web of Science, ScienceDirect, Wiley, SID, and Google Scholar. To identify the appropriate keywords, the MeSH (Medical Subject Headings) framework was applied, and terms were selected collaboratively by two specialists in artificial intelligence and anesthesia. Relevant terms were then searched in the databases. Data extraction was independently performed by two researchers. The applied search strategy was: ((Machine Learning[Title]) OR (Transfer Learning[Title])) AND (Intratracheal OR Intratracheal Intubation* OR Intubation, Endotracheal OR Endotracheal Intubations [Title/Abstract]) (Box 1). Based on this strategy, eligible studies were identified and screened.**

Inclusion criteria for article selection were as follows:

1. Original research papers published in the study domain.
2. Studies conducted between 2000 and July 2025.
3. Publications available in English or Persian relevant to the research scope.

Exclusion criteria included:

1- Studies without accessible full text.
2- Research with unclear or inconclusive findings that did not directly address the study objectives.
3- Publications in the form of letters to the editor, posters, or articles from non-peer-reviewed journals.

Box 1- Search strategy

Keyword relating to Machine learning Machine OR Transfer Learning OR Learning, Transfer
Keyword relating to Intubation Intratracheal OR Intratracheal Intubation* OR Intubation, Endotracheal OR Endotracheal Intubations

Results

Initially, the identified set of articles, totaling 91, was organized in Endnote X16. After removing duplicate entries, the count decreased to 51 distinct articles. Then, an evaluation based on titles and abstracts was conducted, resulting in the removal of 32 articles. Authors of articles without full-text access were contacted, and extensive efforts were made to obtain the full text. Ultimately, low-quality articles, along with those without full-text access and conference presentations, were excluded from the study. In the end, the scope of the study was limited to 9 selected articles (Figure 1, Table 1).

Table 1- Summary of Selected Studies Applying Machine Learning Models for Difficult Airway Prediction

Title	Year/Author /Country	Input information Algorithm	Results
Voice Analysis as a Method for Preoperatively Predicting a Difficult Airway Based on Machine Learning Algorithms: Original Research Report [9]	2024 C. Rodiera Spain	Collection of clinical features of patients and traditional predictive tests, recording the vowel sounds "A, E, I, O, U" in normal, bent, and stretched positions.	The Cormack grade was assessed, and the data were analyzed using KNIME, resulting in the creation of multiple models based on demographic and acoustic data. - Top Models:Two models for predicting difficult airways performed best. - Case Analysis: These models exclusively focused on analyzing Cormack grades I and IV. - Model 1: - Includes: demographic data, vowel sound "A" in all positions, voice harmonics - AUC: 0.91 - Model 2: - Includes: demographic data, vowel sound "O" in normal positions, acoustic parameters - AUC: 0.90 Models that focused on analyzing all Cormack grades (I, II, III, IV) performed less effectively.
Unravelling intubation challenges: a machine learning approach incorporating multiple predictive parameters [6]	2024 P. Sezari Iran	Collecting patients' clinical characteristics and traditional predictive tests, recording Cormack-Lehan score at the time of intubation	Among the algorithms, KNN performed better than the other algorithms. AUC: 0.87

Reliable prediction of difficult airway for tracheal intubation from patient preoperative photographs by machine learning methods. Computer Methods and Programs in Biomedicine [10]	2024 Fernando García Spain	Morphological features of preoperative images: - 59 engineered features including: distances, areas, angles, ratios Demographic variables	The best pipeline method was XGB, which had the lowest number of false negatives at the optimal Bayesian decision threshold. AUC: 0.716
Development of A Machine Learning Model for Predicting Unanticipated Difficult Tracheal Intubation. Journal of Anesthesia and Translational Medicine [5]	2022 B. Wang China	Collection of clinical characteristics of patients and traditional predictive tests	XGBoost is an effective machine learning model for predicting unexpected DTI. AUC: 0.924
Predictive model for difficult laryngoscopy using machine learning: retrospective cohort study [11]	2022 J. H. Kim South Korea	Age, Mallampati grade, body mass index (BMI), sternomental distance, and neck circumference.	The best performance was achieved using the lightweight gradient boosting machine algorithm with Mallampati score, age, and sternomental distance as the predictive model parameters. AUC: 0.71
A prediction model for difficult intubation using skeletal features in patients affected by apnea-hypopnea syndrome [12]	2022 S. Yan China	Collecting clinical characteristics of patients and traditional predictive tests, OSA severity	Based on the results of LASSO regression, age and four skeletal features (sternomental distance, maximum mandibular prominence, mentohyoid distance, and neck hypokinesia grade) were included in the final model.
Development and validation of a difficult laryngoscopy prediction model using machine learning of neck circumference and thyromental height [13]	2021 J. H. Kim South Korea	Age, sex, height, weight, body mass index, neck circumference, and thyromental distance	Among the algorithms, random forest performed better than other algorithms. AUC: 0.79
Defining difficult laryngoscopy findings by using multiple parameters: A machine learning approach [8]	2017 M. A. Moustafa Egypt	Collecting clinical characteristics of patients and traditional predictive tests	The collected data were processed using Microsoft Visual Studio software and WEKA machine learning algorithms. - The data were classified using the J48 algorithm, which is based on decision trees, ultimately resulting in the creation of the "Alexander Difficult Laryngoscopy Software" (ADLS).
SVM-based decision support system for clinic aided tracheal intubation predication with multiple features [14]	2009 Q. Yan China	Collecting clinical characteristics of patients and traditional predictive tests	The results indicate that the SVM-based decision support system can provide an excellent practical outlook in supporting clinical diagnosis by fully considering multiple features of airway physical examination. AUC: 0.905

Discussion

The systematic review of nine studies demonstrates the growing potential of machine learning (ML) in predicting difficult laryngoscopy and intubation. Various ML models, including K-nearest neighbors (KNN), decision trees (J48), random forests, support vector machines (SVM), and gradient boosting algorithms such as

XGBoost, have shown promising diagnostic performance using clinical, demographic, and morphological data. Among the reviewed studies, those that incorporated non-traditional features—such as acoustic data or preoperative facial images—achieved higher accuracy in prediction. For instance, Rodiera et al. (2024) utilized patient voice recordings while pronouncing vowels to analyze Cormack-Lehane grades, achieving impressive AUC values of 0.91 and 0.90 in their models [9]. Similarly, García-García et al. (2024) used preoperative facial image analysis with geometric features, resulting in an AUC of 0.716 [10].

In contrast, models based solely on traditional clinical parameters such as age, Mallampati score, BMI, sternomental distance, and neck circumference showed moderate predictive performance [11,13]. This highlights the importance of integrating advanced data types with conventional clinical information to improve model outcomes.

However, direct comparison across studies remains difficult due to methodological inconsistencies, such as variations in dataset sizes, validation approaches (e.g., train/test split vs. cross-validation), and lack of external validation. Although some studies reported high performance—such as Wang et al. with an AUC of 0.924 and Yan et al. with an AUC of 0.905 [5,14]—generalizability remains uncertain without external testing on independent populations.

Another major concern is the risk of overfitting, particularly in studies with limited or imbalanced datasets. Most studies lacked external validation cohorts, which weakens the applicability of these models in real clinical environments. Additionally, few studies addressed data bias, such as underrepresentation of specific demographic groups, which could impact model fairness and safety.

A critical barrier to clinical implementation is the interpretability of the models. While high-performing models like XGBoost or SVM offer robust predictions, they often act as “black boxes” and may not be easily accepted by clinicians without insight into how decisions are made [5-6]. Striking a balance between predictive accuracy and model transparency is therefore essential, especially in high-risk clinical scenarios like airway management.

Future research should focus on standardizing feature definitions and airway difficulty criteria to enhance comparability and reproducibility. Multicenter studies with diverse and sufficiently large datasets are necessary to improve generalizability. Incorporating model interpretability techniques, such as SHAP values, can foster clinician trust. Moreover, evaluating models prospectively in real-time clinical settings will help ensure their robustness and clinical applicability.

Conclusion

In summary, the use of machine learning in predicting difficult laryngoscopy and intubation can contribute to improved clinical outcomes. However, further research is needed in this area to optimize models and enhance prediction accuracy. Additionally, examining the impact of various variables on model performance can aid in the development of better clinical tools.

Acknowledgment

The authors thank the Department of Anesthesia at Iran University of Medical Sciences for their support and guidance during this study.

References

[1] Sammut C, Webb GI. Encyclopedia of machine learning: Springer Science & Business Media; 2011.

[2] Crawley S, Dalton A. Predicting the difficult airway. BJA Educ. 2015;15(5):253-7.

[3] Peterson GN, Domino KB, Caplan RA, Posner KL, Lee LA, Cheney FW. Management of the difficult airway: a closed claims analysis. Anesthesiology. 2005;103(1):33-9.

[4] Heidegger T. Management of the difficult airway. N Engl J Med. 2021;384(19):1836-47.

[5] Wang B, Yao W, Wang M, Xu J, Huang D, Lu X, et al. Development of A Machine Learning Model for Predicting Unanticipated Difficult Tracheal Intubation. J Anesth Transl Med. 2022;1(1):26-31.

[6] Sezari P, Kohzadi Z, Dabbagh A, Jafari A, Khoshtinatan S, Mottaghi K, et al. Unravelling intubation challenges: a machine learning approach incorporating multiple predictive parameters. BMC Anesthesiol. 2024;24(1):453.

[7] Sreekantha D, Glen RRC, Prajna M, editors. Prediction of Difficulties in Airway Management using Machine Learning Algorithms. 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC); 2018: IEEE.

[8] Moustafa MA, El-Metainy S, Mahar K, Mahmoud Abdel-magied E. Defining difficult laryngoscopy findings by using multiple parameters: A machine learning approach. Egypt J Anaesth. 2017;33(2):153-8.

[9] Rodiera C, Fortuny H, Valls A, Borras R, Ramírez C, Ros B, et al. Voice Analysis as a Method for Preoperatively Predicting a Difficult Airway Based on Machine Learning Algorithms: Original Research Report. Health Sci Rep. 2024;7(12):e70246.

[10] García-García F, Lee D-J, Mendoza-Garcés FJ, García-Gutiérrez S. Reliable prediction of difficult airway for tracheal intubation from patient preoperative photographs by machine learning

methods. Comput Methods Programs Biomed. 2024;248:108118.
[11] Kim JH, Choi JW, Kwon YS, Kang SS. Predictive model for difficult laryngoscopy using machine learning: retrospective cohort study. Braz J Anesthesiol. 2022;72(5):622-8.
[12] Yan S, Guo M, Gao Z, Wu H, Liu X, Yin G, et al. A prediction model for difficult intubation using skeletal features in patients affected by apnea-hypopnea syndrome. Minerva Anestesiol. 2022.
[13] Kim JH, Kim H, Jang JS, Hwang SM, Lim SY, Lee JJ, et al. Development and validation of a difficult laryngoscopy prediction model using machine learning of neck circumference and thyromental height. BMC Anesthesiol. 2021;21:1-7.
[14] Yan Q, Yan H, Han F, Wei X, Zhu T. SVM-based decision support system for clinic aided tracheal intubation predication with multiple features. Expert Syst Appl. 2009;36(3):6588-92.