



Artificial Intelligence for Prediction of Difficult Airway in Anesthesia: A Systematic Review and Descriptive Synthesis

Marcell Dion Wibowo^{1*}

1. Department of Anesthesiology and Intensive Care, Prof. IGNG Ngoerah General Hospital - Faculty of Medicine, Udayana University, Denpasar, Indonesia

*penulis korespondensi

DOI : 10.55497/majanestcricar.v44i3.527

ABSTRACT

Introduction: Difficult airway remains an important cause of anesthesia-related morbidity because failure to identify high-risk airway before induction may lead to repeated laryngoscopy attempts, hypoxemia, aspiration, emergency front-of-neck access, cardiac arrest, or death. This review evaluated artificial intelligence (AI) models for difficult-airway prediction in anesthesia and airway management.

Methods: A systematic literature review with descriptive synthesis was performed using a PICOTS framework and PRISMA-based reporting. PubMed/MEDLINE, Embase, and Scopus were used for the primary evidence map, with supplementary updating through PubMed/MEDLINE, Google Scholar, and reference checking to 13 October 2026. Eligible studies were original human clinical studies using AI, machine learning, deep learning, or computer vision to predict difficult intubation, difficult laryngoscopy, difficult video laryngoscopy, difficult mask ventilation, or composite difficult airway. Studies predicting respiratory-failure intubation or tube size/depth without difficult-airway assessment were excluded.

Results: Nineteen eligible difficult-airway-related studies were included in the final descriptive synthesis. Sixteen studies reported extractable AUC or c-statistic values, while three studies were summarized narratively only. One estimate per study was summarized descriptively, with median discrimination of 0.847 (IQR 0.773-0.865; range 0.710-0.970). Calibration was reported in five studies and external validation in one study.

Conclusion: AI models show promising discrimination for difficult-airway prediction, but evidence remains limited by heterogeneous outcome definitions, incomplete calibration reporting, sparse external validation, and uncertain clinical impact. AI should currently be considered an adjunct to anesthesiologist judgment rather than a stand-alone diagnostic method.

Keywords: Anesthesia, artificial intelligence, deep learning, difficult airway, machine learning



Kecerdasan Buatan Untuk Prediksi Jalan Napas Sulit Pada Anestesi: Tinjauan Sistematis dan Sintesis Deskriptif

Marcell Dion Wibowo^{1*}

1. Departemen Anestesiologi dan Perawatan Intensif, RSUP IGNG Ngoerah - Fakultas Kedokteran, Universitas Udayana, Denpasar, Indonesia

*corresponding author

DOI : 10.55497/majanestcricar.v44i3.527

ABSTRAK

Pendahuluan: Jalan napas sulit masih menjadi penyebab penting morbiditas terkait anestesi karena kegagalan mengenali risiko sebelum induksi dapat menyebabkan laringoskopi berulang, hipoksemia, aspirasi, akses jalan napas darurat, henti jantung, atau kematian. Tinjauan ini mengevaluasi model kecerdasan buatan (AI) untuk prediksi jalan napas sulit pada anestesi dan manajemen jalan napas.

Metode: Tinjauan literatur sistematis dengan sintesis deskriptif dilakukan menggunakan kerangka PICOTS dan pelaporan berbasis PRISMA. PubMed/MEDLINE, Embase, dan Scopus digunakan untuk pemetaan bukti primer, dengan pembaruan melalui PubMed/MEDLINE, Google Scholar, dan penelusuran referensi hingga 13 Oktober 2026. Studi yang memenuhi kriteria adalah penelitian klinis manusia yang menggunakan AI, machine learning, deep learning, atau computer vision untuk memprediksi intubasi sulit, laringoskopi sulit, videolaringoskopi sulit, ventilasi masker sulit, atau jalan napas sulit komposit. Studi yang memprediksi intubasi akibat gagal napas atau ukuran/kedalaman pipa tanpa penilaian jalan napas sulit dikeluarkan.

Hasil: Sebanyak 19 studi terkait jalan napas sulit memenuhi kriteria sintesis deskriptif akhir. Enam belas studi melaporkan AUC atau c-statistic yang dapat diekstraksi, sedangkan tiga studi dirangkum secara naratif. Satu estimasi per studi dirangkum secara deskriptif dengan median diskriminasi 0,847 (IQR 0,773-0,865; rentang 0,710-0,970). Kalibrasi dilaporkan pada lima studi dan validasi eksternal pada satu studi.

Simpulan: Model AI menunjukkan diskriminasi menjanjikan untuk prediksi jalan napas sulit, namun bukti dibatasi oleh heterogenitas definisi luaran, pelaporan kalibrasi tidak lengkap, validasi eksternal jarang, dan dampak klinis belum pasti. AI sebaiknya diposisikan sebagai pendukung keputusan klinis, bukan pengganti penilaian ahli anestesi.

Kata kunci: Anestesi, jalan napas sulit, kecerdasan buatan, machine learning, tinjauan sistematis

INTRODUCTION

Airway management is a central responsibility of anesthesiology, critical care, and emergency medicine. Although most tracheal intubations are uncomplicated, the unanticipated difficult airway remains one of the most feared perioperative emergencies. Difficult airway may involve difficult mask ventilation, difficult supraglottic airway placement, difficult laryngoscopy, difficult tracheal intubation, failed intubation, or a combined difficult-airway scenario. The clinical consequences are time-sensitive and potentially catastrophic, especially after administration of anesthetic agents and neuromuscular blockade.^{1,2} Conventional preoperative assessment relies on bedside tests that are easy to perform but imperfect. Modified Mallampati classification, thyromental distance, inter-incisor distance, sternomental distance, mandibular protrusion, neck circumference, upper lip bite test, cervical mobility, and composite scores provide useful information but often perform poorly as stand-alone screening tools.³⁻⁵ In daily clinical practice, low-sensitivity screening may falsely reassure clinicians and fail to trigger preparation for awake intubation, video laryngoscopy, fiberoptic intubation, or advanced airway rescue equipment.⁶ Artificial intelligence (AI) and machine learning have been proposed to support airway risk stratification by integrating nonlinear associations across multidimensional data. Facial photographs, ultrasound, radiographs, acoustic signals, electronic health record variables, and multimodal data may capture anatomical or functional features that are not consistently quantified during routine bedside assessment.⁷ This systematic review and descriptive synthesis aimed to evaluate the predictive performance of AI models for difficult airway prediction, identify common input modalities and algorithms, summarize methodological quality, and assess readiness for clinical implementation.

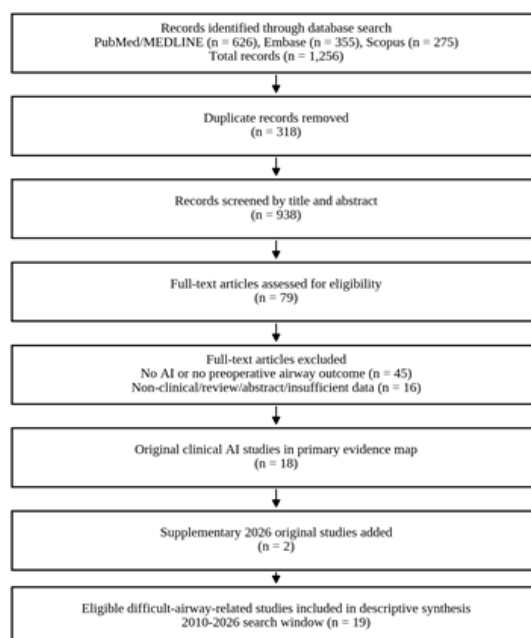
METHODS

This manuscript was revised as a systematic review with descriptive synthesis rather than a pooled diagnostic analysis. The review question was structured using the PICOTS framework and prediction-model review principles. Population:

human patients undergoing anesthesia or airway management. Index model: artificial intelligence (AI), machine learning, or deep learning prediction models. Comparator: conventional airway assessment, clinician judgment, or no comparator. Outcomes: difficult intubation, difficult laryngoscopy, difficult video laryngoscopy, difficult mask ventilation, failed intubation, or composite difficult airway. Timing: preoperative or pre-intubation prediction. Setting: operating room, intensive care unit, emergency department, or procedural sedation settings.⁸⁻¹¹ The primary evidence map was based on structured searches of PubMed/MEDLINE, Embase, and Scopus covering January 2010 to December 2025. The database counts used in the flow diagram were PubMed/MEDLINE (n = 626), Embase (n = 355), and Scopus (n = 275), yielding 1,256 records before duplicate removal. Because separate database export dates for the primary evidence map were not preserved, this is reported as a limitation. A supplementary update was last conducted on 13 October 2026 using PubMed/MEDLINE, Google Scholar, and reference checking to identify original 2026 studies not covered in the primary evidence map. The protocol was not registered in PROSPERO or another registry. Search terms combined airway concepts, anesthesia concepts, and AI concepts, including difficult airway, difficult intubation, difficult laryngoscopy, difficult mask ventilation, failed intubation, airway management, anesthesia, anesthesia, intubation, laryngoscopy, artificial intelligence, machine learning, deep learning, neural network, convolutional neural network, computer vision, random forest, support vector machine, XGBoost, LightGBM, and prediction model. Eligible studies were original human clinical studies that developed, validated, or evaluated an AI or machine-learning model for difficult-airway prediction and reported at least one quantitative performance measure. Reviews, editorials, protocols, manikin-only studies, simulation-only studies, non-human studies, studies without an AI component, and studies without extractable performance data were excluded. To align with the PICOTS, studies predicting respiratory-failure intubation, endotracheal tube size, bronchial tube size, or

tube depth without a difficult-airway outcome were excluded from the final synthesis. Extracted variables included reference number, author and year, country, design, clinical setting, sample size, airway outcome definition, input data type, algorithm, validation method, cohort from which the performance estimate was taken, AUC or c-statistic, sensitivity, specificity, accuracy, calibration, external validation, explainability, and risk-of-bias judgement. Where several models or thresholds were reported within a single study, one estimate per study was selected for the discrimination plot, prioritizing the authors' primary/test-set model or the

model most directly related to the difficult-airway outcome. Risk of bias was appraised using PROBAST+AI concepts, including participant selection, predictor definition, outcome measurement, sample size, missing data, class imbalance, model development, validation, calibration, applicability, and clinical usability.⁹ Because studies did not consistently report two-by-two diagnostic tables, common thresholds, or AUC standard errors, formal random-effects pooling was not performed. Reported AUC or c-statistic values were summarized descriptively using median, interquartile range, and range.



Protocol not registered. Last supplementary search: 13 October 2026. Three studies were summarized narratively only.

Figure 1. Flowchart of studies included in the systematic review and descriptive synthesis

RESULT

The database search identified 1,256 records from PubMed/MEDLINE, Embase, and Scopus. After removal of 318 duplicates, 938 records were screened by title and abstract, and 79 full-text articles were assessed for eligibility. Eighteen original clinical AI studies were identified for the primary evidence map.⁷ One study of intubation for respiratory failure was excluded because it did not assess difficult airway outcomes. Studies predicting only endotracheal tube size or insertion depth were excluded unless

they also reported a difficult-airway-related outcome. Two original studies published in 2026 were added through the supplementary update. The final descriptive synthesis included 19 eligible studies of difficult-airway-related outcomes; 16 reported extractable area under the curve (AUC) or c-statistic values, and 3 were summarized narratively only.¹²⁻³⁰ The eligible studies were predominantly single-center model-development or development-validation studies. Countries represented included Turkey, Japan, South Korea, China, Spain, the United

States, Iran, and multicenter emergency department cohorts. Clinical targets included difficult direct laryngoscopy, difficult tracheal intubation, difficult video laryngoscopy, difficult mask ventilation, Cormack-Lehane grade III/IV, poor glottic exposure, first-pass failure related to difficult airway, and composite difficult-airway endpoints. Algorithms included random forest, balanced random forest, support vector machine, XGBoost, LightGBM, gradient boosting, logistic regression with feature selection, k-nearest neighbors, convolutional neural networks, VGG16, EfficientNet, ResNet, and multimodal fusion models.¹²⁻³⁰ The most common input categories were facial photographs, clinical and anthropometric variables, ultrasound images, lateral cervical radiographs, three-dimensional facial scans, voice recordings, emergency airway variables, and electronic health record variables. Sixteen studies reported AUC or c-statistic values that were extractable for the discrimination plot. Three studies were retained for narrative synthesis but not plotted because AUC or c-statistic was not extractable or not directly comparable with the difficult-airway discrimination analysis. When one estimate per study was selected, reported discrimination ranged from 0.710 to 0.970, with a median of 0.847 and an interquartile range of 0.773 to 0.866. Calibration was reported in five studies, while external validation was reported in one study. These results indicate promising discrimination but limited evidence

for transportability and absolute risk accuracy. Sixteen studies provided an extractable AUC or c-statistic and were therefore included in the discrimination analysis shown in Figure 2. One estimate per study was used to avoid double-counting multiple models from the same cohort; when several models were reported, the final or clinically most representative model was selected. Reported discrimination ranged from 0.710 to 0.970, with a median of 0.847 and an interquartile range of 0.773-0.865. Seven studies reported AUC/c-statistic values of 0.85 or higher, four were between 0.80 and below 0.85, and five were below 0.80. This distribution suggests promising but heterogeneous performance across input modalities, populations, outcome definitions, and validation designs. The highest values were reported in studies using skeletal/radiographic features or selected image-based models, whereas lower values were observed in several photographic, emergency-airway, and video laryngoscopy-related models. These differences should not be interpreted as proof that one modality is superior because the studies used different populations, definitions of difficult airway, validation strategies, and model-selection procedures. Because diagnostic thresholds, two-by-two tables, and AUC standard errors were not consistently available, formal pooled diagnostic meta-analysis was not performed; the quantitative analysis is therefore descriptive rather than pooled.

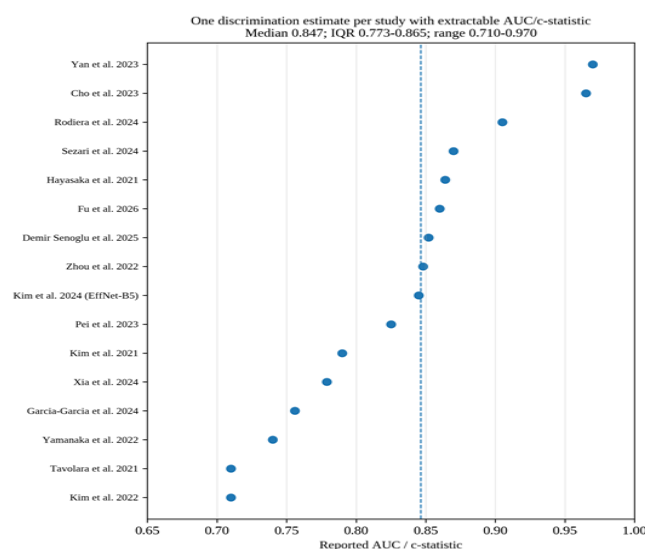


Figure 2. Reported discrimination of AI models in studies with extractable AUC or c-statistic

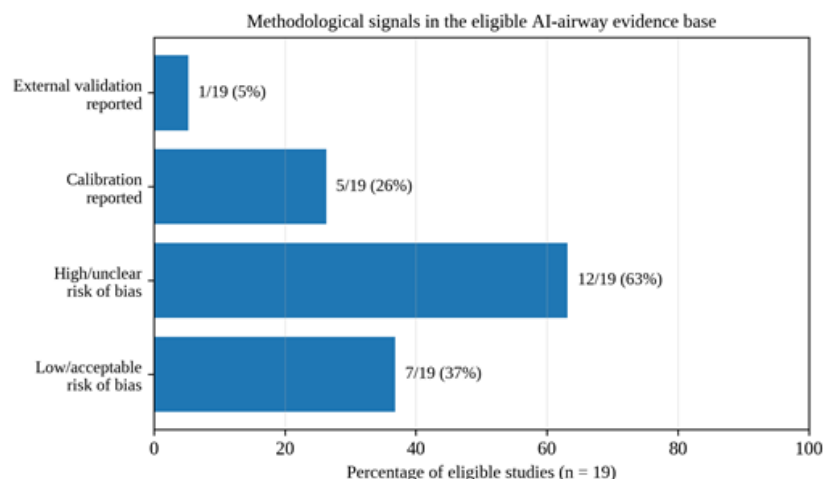


Figure 3. Methodological signals in the eligible AI-airway evidence base

Among the 19 eligible studies, calibration or calibration-related metrics were reported in five studies, and external validation was reported in only one study. High or unclear risk-of-bias signals were identified in 12 studies, whereas seven were judged to have low or acceptable

risk-of-bias signals based on the extracted design, validation, and reporting characteristics. This pattern indicates that the main weakness of the literature is not the absence of promising discrimination, but insufficient evidence of calibration, transportability, and clinical impact.

Table 1. Characteristics and outcome analysis of eligible AI-airway studies

Year	Study/ country	Design/ sample	Outcome definition	Input/ algorithm	Validation/ cohort for estimate	Main result
2021	Celik and Aydemir ¹⁶ / Turkey	Single-blind prospective observational; N=1,486	Difficult tracheal intubation using Cormack-Lehane III/IV classification	Clinical tests and anthropometric variables / random forest	Internal 10-fold cross-validation; total cohort used for model development	Sensitivity 92.85%; calibration not reported
2021	Hayasaka et al. ¹² / Japan	Single-center observational; N=202	Intubation difficulty / Cormack-Lehane-based classification	Face images / VGG16 transfer learning	Internal cross-validation with augmentation and oversampling	AUC 0.864; sensitivity 81.8%; specificity 83.3%
2021	Tavolara et al. ¹³ / United States	Single-center development; N=152	Difficult intubation	Frontal face images / CNN ensemble	Internal fold cross-validation with attention-based multiple-instance learning	AUC 0.710; calibration not reported
2021	Kim et al. ²⁰ / South Korea	Single-center development and validation; N=1,677	Difficult laryngoscopy defined as Cormack-Lehane III/IV	Neck circumference and thyromental height / balanced random forest	Internal 80/20 split; test cohort estimate	AUC 0.790; calibration not reported

Year	Study/ country	Design/ sample	Outcome definition	Input/ algorithm	Validation/ cohort for estimate	Main result
2022	Kim et al. ²¹ / South Korea	Single-center retrospective cohort; N=616	Difficult laryngoscopy defined as Cormack- Lehane III/IV	Clinical and airway measurements / LightGBM	Internal split; test cohort estimate	80/20 AUC 0.710; calibration not reported
2022	Zhou et al. ²⁸ / China	Single-center retrospective thyroid surgery cohort; N=500	Composite difficult airway endpoint	Clinical variables / gradient boosting	Internal split; test cohort estimate	70/30 AUC 0.848; calibration not reported
2022	Yamanaka et al. ²⁹ / Japan	Multicenter prospective ED cohort; N=10,741	Difficult airway and first-pass success during emergency intubation	Routinely collected ED airway variables / ensemble ML models	Internal random test set	80/20 Best difficult- airway c-statistic 0.740; calibration slope reported
2023	Cho et al. ¹⁷ / South Korea	Single-center development and validation; N=5,939	Difficult laryngoscopy defined as Cormack- Lehane III/IV	Lateral cervical radiographs / 4-layer CNN	Internal repeated random splits; test-set estimate	AUC 0.965; Brier score 0.023
2023	Pei et al. ²⁴ / China	Single-center development and validation; N=669	Difficult mask ventilation	3D facial scans / logistic regression plus PCA / ML	Internal cross- validation; validation estimate	AUC 0.825; calibration not reported
2023	Yan et al. ²⁷ / China	Training plus external validation; N=247 + 82	Difficult intubation using skeletal features in sleep apnea- hypopnea syndrome; Cormack- Lehane III/IV	Skeletal radiographic features / LASSO plus logistic regression	Temporal external validation; external validation estimate	AUC 0.970; calibration plot and Brier score reported
2024	Garcia-Garcia et al. ¹⁹ / Spain	Prospective cohort; N=623	Difficult tracheal intubation	Preoperative photographs and engineered morphometric features / XGBoost	Nested 10-fold cross-validation; cross-validated estimate	AUC 0.746- 0.766; calibration plot reported
2024	Kim et al. ²² / South Korea	Single-center prospective study; N=1,283	Difficult direct laryngoscopy	Minimal image analysis / EfficientNet-B5	Internal cross- validation with multitask learning	AUC 0.810- 0.880; calibration not reported
2024	Kim et al. ²³ / South Korea	Secondary analysis; sample from prior video- laryngoscopy cohort	Video laryngoscopic view / POGO- related glottic exposure	Clinical predictors / machine- learning models	Internal validation based on study cohort	Model performance reported for video- laryngoscopy view prediction; included as difficult- airway- related glottic exposure study

Year	Study/ country	Design/ sample	Outcome definition	Input/ algorithm	Validation/ cohort for estimate	Main result
2024	Xia et al. ²⁶ / China	Single-center development and validation; N=5,849	Difficult videolaryngoscopy	Facial image analysis / ResNet-18 plus LightGBM	Internal 60/20/20 split; test cohort estimate	AUC 0.779; Brier score reported
2024	Rodiera et al. ²⁵ / Spain	Multicenter development; N=313	Difficult laryngoscopy defined as Cormack-Lehane III/IV	Voice recordings / machine learning and logistic regression	Internal 70/30 split; test cohort estimate	AUC 0.900- 0.910; calibration not reported
2024	Sezari et al. ³⁰ / Iran	Cross-sectional two-hospital study; N=622	Difficult laryngoscopy based on Cormack- Lehane grade III; no grade IV cases observed	Clinical and airway predictors/KNN, SVM, random forest, decision tree, logistic regression	Internal 10-fold cross-validation with data balancing	Best AUC 0.870; accuracy 0.870; calibration not reported
2025	Demir Senoglu et al. ¹⁸ / Turkey	Single-center prospective development; N=329	Difficult intubation / Cormack-Lehane III/IV	Anthropometric and ultrasound measures / SVM	Internal 80/20 split plus cross- validation; SMOTE on training data	AUC 0.852; calibration not reported
2026	Fu et al. ¹⁴ / China	Single-center internal validation; N=903	Difficult direct laryngoscopy and videolaryngoscopy exposure	Neck ultrasound images / CNN ultrasound framework	Five-fold cross- validation plus internal test set	CL-AI AUC 0.860; VIDIAC-AI AUC 0.820; no external validation
2026	Koca et al. ¹⁵ (2026) / Turkey	Original clinical photographic study	Difficult tracheal intubation using three-class airway difficulty classification	Multi-angle smartphone photographs / CNN and EfficientNet	Internal model testing under multiple learning rates and batch sizes	Accuracy 88.64%, precision 88.01%, F1-score 87.28%; AUC not extracted

Abbreviations: AI, artificial intelligence; AUC, area under the receiver operating characteristic curve; CNN, convolutional neural network; ED, emergency department; KNN, k-nearest neighbors; LightGBM, light gradient boosting machine; LR, logistic regression; ML, machine learning; PCA, principal component analysis; POGO, percentage of glottic opening; SVM, support vector machine. Note: Celik and Aydemir, Kim et al. (J Pers Med 2024), and Koca et al. were retained in the descriptive synthesis but not included in Figure 2 because AUC/c-statistic was not extractable or not directly comparable.

DISCUSSION

This review demonstrates that AI-based difficult-airway prediction is a rapidly expanding field with promising but immature clinical evidence. Across eligible studies reporting AUC or c-statistic, most models showed moderate to excellent discrimination, with a median estimate of 0.847. These results suggest that AI models can capture clinically relevant risk patterns from photographs, ultrasound, radiographs, voice signals, three-dimensional facial morphology, and structured clinical variables. However, this review did not perform a direct pooled

comparison between AI and conventional airway tests; therefore, superiority over bedside assessment cannot be concluded from this synthesis alone. The main conceptual advantage of AI is its capacity to combine multiple measurements and model nonlinear anatomical relationships. Facial imaging models may capture craniofacial morphology and soft tissue distribution, ultrasound models may provide objective information about anterior neck soft tissue and laryngeal structures, radiographic or skeletal-feature models can quantify cervical and airway geometry, and voice-analysis

models may capture functional upper-airway dynamics.^{12-15,17,19,24-27} These modalities should be viewed as complementary rather than competitive, because difficult laryngoscopy, difficult intubation, and difficult mask ventilation represent related but different clinical problems. High discrimination alone does not establish clinical readiness. A model that performs well in one institution may perform poorly in another because of differences in patient ethnicity, obesity distribution, surgical population, camera distance, ultrasound equipment, anesthesiologist experience, airway device, and outcome definition. External validation was identified in only one study, which is insufficient for broad implementation. This limitation is important because airway management decisions are high-stakes: an overconfident false-negative prediction may lead to inadequate preparation, while a high false-positive rate may result in unnecessary resource use or airway manipulation. Calibration is another major gap. For clinical decision support, clinicians need to know whether a predicted 20% risk corresponds to approximately 20% observed risk. Calibration or calibration-related metrics were reported in only five studies. Without calibration, AI outputs may be useful for ranking risk but not for absolute decision thresholds such as whether to plan awake intubation, require senior anesthesiologist presence, prepare a flexible bronchoscope, or choose video laryngoscopy as the primary approach.^{8,9} Outcome heterogeneity also limits interpretation. Some studies defined difficult airway using Cormack-Lehane grade III/IV, while others used difficult intubation, difficult videolaryngoscopy, difficult mask ventilation, first-pass failure, POGO-related glottic exposure, or composite endpoints. These outcomes are related but not interchangeable. Difficult laryngoscopy does not always equal difficult intubation, especially when video laryngoscopy or external laryngeal manipulation is used. Conversely, difficult mask ventilation may occur despite an acceptable glottic view.^{2,24,26,29}

AI models should therefore be positioned as clinician-support tools rather than replacements for anesthesiologist judgment. The most realistic implementation pathway is integration into preoperative assessment,

where models can generate a risk probability, highlight relevant anatomical regions, and recommend preparation steps according to local difficult-airway algorithms. The ultimate clinical endpoint should not be AUC alone but reduction in unanticipated difficult airway, fewer repeated attempts, lower hypoxemia rates, earlier escalation, improved first-pass success, and better patient safety outcomes.^{2,8,9} This review has several limitations. First, the available literature is heterogeneous in population, outcome definition, input modality, algorithm, and validation method. Second, several studies did not report complete diagnostic two-by-two tables, common thresholds, or AUC standard errors, preventing formal pooled diagnostic modeling. Third, several high-performing models were internally validated only, which may overestimate performance. Fourth, the protocol was not prospectively registered. Fifth, because the supplementary update was targeted rather than a full replicated database export from all sources, future reviews should perform duplicate independent screening, save all database exports, and publish complete search strings. Future studies should be prospective, multicenter, externally validated, standardized in difficult-airway definitions, and reported according to TRIPOD+AI and PROBAST+AI principles.^{8,9}

CONCLUSION

Artificial intelligence models show promising performance for predicting difficult airway in anesthesia, with reported discrimination commonly in the moderate-to-excellent range. Image-based, ultrasound-based, radiographic, acoustic, emergency-airway, and multimodal approaches may complement conventional assessment by integrating anatomical and clinical variables. Nevertheless, the evidence remains limited by inconsistent definitions, incomplete reporting, limited calibration, sparse external validation, and uncertain clinical impact. AI-based difficult-airway prediction should currently be considered an adjunctive clinical decision-support strategy rather than a stand-alone diagnostic method. Prospective multicenter validation and outcome-impact studies are required before routine implementation.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

1. Cook TM, Woodall N, Frerk C. Major complications of airway management in the UK: results of the Fourth National Audit Project of the Royal College of Anaesthetists and the Difficult Airway Society. Part 1: anaesthesia. *Br J Anaesth*. 2011;106(5):617-31. doi:10.1093/bja/aer058
2. Apfelbaum JL, Hagberg CA, Connis RT, Abdelmalak BB, Agarkar M, Dutton RP, et al. 2022 American Society of Anesthesiologists practice guidelines for management of the difficult airway. *Anesthesiology*. 2022;136(1):31-81. doi:10.1097/ALN.0000000000004002
3. Shiga T, Wajima Z, Inoue T, Sakamoto A. Predicting difficult intubation in apparently normal patients: a meta-analysis of bedside screening test performance. *Anesthesiology*. 2005;103(2):429-37. doi:10.1097/00000542-200508000-00027
4. Roth D, Pace NL, Lee A, Hovhannisyan K, Warenits AM, Arrich J, et al. Airway physical examination tests for detection of difficult airway management in apparently normal adult patients. *Cochrane Database Syst Rev*. 2018;5:CD008874. doi: 10.1002/14651858.CD008874.pub2
5. Wang Z, Jin Y, Zheng Y, Chen H, Feng J, Sun J. Evaluation of preoperative difficult airway prediction methods for adult patients without obvious airway abnormalities: a systematic review and meta-analysis. *BMC Anesthesiol*. 2024;24:242. doi:10.1186/s12871-024-02627-1
6. Norskov AK, Rosenstock CV, Wetterslev J, Astrup G, Afshari A, Lundstrom LH. Diagnostic accuracy of anaesthesiologists' prediction of difficult airway management in daily clinical practice: a cohort study of 188064 patients registered in the Danish Anaesthesia Database. *Anaesthesia*. 2015;70:272-81. doi:10.1111/anae.12955
7. Shen Y, Wu Y, Qiu Y, Wu J. AI-based decision models for difficult airway assessment: from research innovation to clinical implementation - a narrative review. *Front Med (Lausanne)*. 2026;13:1818061. doi:10.3389/fmed.2026.1818061
8. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi:10.1136/bmj-2023-078378
9. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ*. 2025;388:e082505. doi:10.1136/bmj-2024-082505
10. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10):e1001744. doi:10.1371/journal.pmed.1001744
11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71
12. Hayasaka T, Kawano K, Kurihara K, Suzuki H, Nakane M, Kawamae K. Creation of an artificial intelligence model for intubation difficulty classification by deep learning using face images: an observational study. *J Intensive Care*. 2021;9:38. doi:10.1186/s40560-021-00551-x
13. Tavorara TE, Gurcan MN, Segal S, Niazi MKK. Identification of difficult to intubate patients from frontal face images using an ensemble of deep learning models. *Comput Biol Med*. 2021;136:104737. doi:10.1016/j.compbimed.2021.104737
14. Fu C, Luan C, Liu H, Wang W, Zhou X, Jia Y, et al. An ultrasound-based artificial intelligence framework for difficult airway prediction: a two-model, three-step decision framework. *PLoS One*. 2026;21(2):e0342339. doi:10.1371/journal.pone.0342339
15. Koca E, Kutlusoy S, Er MB, Koca T. Predicting difficult tracheal intubation using multi-angle

- photographic analysis with convolutional neural networks and EfficientNet. *Diagnostics* (Basel). 2026;16(7):1042. doi:10.3390/diagnostics16071042
16. Celik F, Aydemir E. Prediction of difficult tracheal intubation by artificial intelligence: a prospective observational study. *Duzce Med J*. 2021;23(1):47-54. doi:10.18678/dtfd.862467
 17. Cho HY, Lee K, Kong HJ, Yang HL, Jung CW, Park HP, et al. Deep-learning model associating lateral cervical radiographic features with Cormack-Lehane grade 3 or 4 glottic view. *Anaesthesia*. 2023;78:64-72. doi:10.1111/anae.15874
 18. Demir Senoglu G, Kara Timucin E, Kirisoglu S. Evaluation of anthropometric and ultrasonographic measurements with different machine learning methods in predicting difficult intubation: a prospective observational study. *BMC Anesthesiol*. 2025;25:496. doi:10.1186/s12871-025-03370-x
 19. Garcia-Garcia F, Lee DJ, Mendoza-Garces FJ, Garcia-Gutierrez S. Reliable prediction of difficult airway for tracheal intubation from patient preoperative photographs by machine learning methods. *Comput Methods Programs Biomed*. 2024;248:108118. doi:10.1016/j.cmpb.2024.108118
 20. 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;21:125. doi:10.1186/s12871-021-01343-4
 21. Kim JH, Choi JW, Kwon YS, Kang SS. Predictive model for difficult laryngoscopy using machine learning: retrospective cohort study. *Braz J Anesthesiol*. 2022;72:622-8. doi:10.1016/j.bjane.2021.06.016
 22. Kim JH, Jung HS, Lee SE, Hou JU, Kwon YS. Improving difficult direct laryngoscopy prediction using deep learning and minimal image analysis: a single-center prospective study. *Sci Rep*. 2024;14:14209. doi:10.1038/s41598-024-65060-x
 23. Kim JH, Han SW, Hwang SM, Lee JJ, Kwon YS. Machine learning predictions and identifying key predictors for safer intubation: a study on video laryngoscopy views. *J Pers Med*. 2024;14:902. doi:10.3390/jpm14090902
 24. Pei B, Jin C, Cao S, Ji N, Xia M, Jiang H. Geometric morphometrics and machine learning from three-dimensional facial scans for difficult mask ventilation prediction. *Front Med (Lausanne)*. 2023;10:1203023. doi:10.3389/fmed.2023.1203023
 25. Rodiera C, Fortuny H, Valls A, Borrás R, Ramirez 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:e70246. doi:10.1002/hsr2.70246
 26. Xia M, Jin C, Zheng Y, Wang J, Zhao M, Cao S, et al. Deep learning-based facial analysis for predicting difficult videolaryngoscopy: a feasibility study. *Anaesthesia*. 2024;79:399-409. doi:10.1111/anae.16194
 27. 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 Anesthesiol*. 2023;89:498-509. doi:10.23736/S0375-9393.22.16869-0
 28. Zhou CM, Wang Y, Xue Q, Yang JJ, Zhu Y. Predicting difficult airway intubation in thyroid surgery using multiple machine learning and deep learning algorithms. *Front Public Health*. 2022;10:937471. doi:10.3389/fpubh.2022.937471
 29. Yamanaka S, Goto T, Morikawa K, Watase H, Okamoto H, Hagiwara Y, et al. 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. doi:10.2196/28366
 30. Sezari P, Kohzadi Z, Dabbagh A, Jafari A, Khoshtinatani S, Mottaghi K, et al. Unravelling intubation challenges: a machine learning approach incorporating multiple predictive parameters. *BMC Anesthesiol*. 2024;24:453. doi:10.1186/s12871-024-02842-w