Unravelling intubation challenges: a machine learning approach incorporating multiple predictive parameters

Parisa Sezari[1], Zeinab Kohzadi[2*] , Ali Dabbagh[3], Alireza Jafari[1], Saba Khoshtinatan[1], Kamran Mottaghi[3], Zahra Kohzadi[4] and Shahabedin Rahmatizadeh[2]

Abstract

Background To protect patients during anesthesia, difficult airway management is a serious issue that needs to be carefully planned for and carried out. Machine learning prediction tools have recently become increasingly common in medicine, frequently surpassing more established techniques. This study aims to utilize machine learning techniques on predictive parameters for challenging airway management.

Methods This study was cross-sectional. The Shahid Beheshti University of Medical Sciences in Iran's Loghman Hakim and Shahid Labbafinezhad hospitals provided 622 records in total for analysis. Using the forest of trees approach and feature importance, important features were chosen. The Synthetic Minority Oversampling Technique (SMOTE) and repeated edited nearest neighbor under-sampling were used to balance the data. Using Python and 10-fold cross-validation, seven machine learning algorithms were assessed: Logistic Regression, Support Vector Machines (SVM), Random Forest (INFORMATION-GAIN and GINI-INDEX), Decision Tree, and K-Nearest Neighbors (KNN). Metrics like F-measure, AUC, Recall, Accuracy, Specificity, and Precision were used to evaluate the performance of the model.

Results Twenty-four important features were chosen from the original 32 features. The under-sampling strategy produced better results than SMOTE. Among the algorithms, KNN (Euclidean, Minkowski) had better performance than other algorithms. The highest values for accuracy, precision, recall, F-measure, and AUC were obtained at 0.87, 0.88, 0.82, 0.85, and 0.87, respectively.

Conclusion Algorithms for machine learning provide insightful information for anticipating challenging airway management. By making it possible to forecast airway difficulties more accurately, these techniques can potentially improve clinical practice and patient outcomes.

Keywords Difficult airway, Airway management, Intratracheal intubation, Anesthesia, Machine learning algorithms, Artificial intelligence

*Correspondence:
Zeinab Kohzadi
z.kohzadi96@gmail.com
Full list of author information is available at the end of the article

Sezari *et al. BMC Anesthesiology*   (2024) 24:453

Page 2 of 14

Introduction

A difficult airway scenario occurs when a healthcare provider with adequate airway management skills faces challenges with laryngoscopy, intubation of the trachea, and more especially, ventilating and oxygenating the lungs. These challenges may be predicted based on the patient's anatomic or pathologic characteristics or may arise unexpectedly while the procedure is being performed. Additionally, resistance to tube placement may occur in some patients [1]. Unexpectedly difficult airway situations increase the risk of brain damage and death [2]. It follows to reduce these risks that preoperative identification of patients at risk for airway difficulties is warranted [3, 4]. Nevertheless, respiratory distress prediction remains a challenging task, and existing predictive accuracy is still quite low [5]. In addition, existing comprehensive clinical evaluation approaches for difficult airways, including the modified LEMON criteria and the Simplified Airway Risk Index (SARI) model, cannot achieve the desired prediction performance in predicting airway management difficulty [6, 7]. Difficult airway management requires a structured procedure, careful evaluation, and a focused effort to maintain oxygenation and minimize complications [8]. Artificial intelligence (AI) encompasses a diverse array of technologies, principles, and methodologies aimed at fostering the development of systems that can replicate and enhance human cognitive abilities. Algorithms are central to AI's functionality, which drives reasoning, problem-solving, and analytical tasks [9]. AI rapidly transforms various aspects of our lives, and its impact on the medical field is particularly profound. As AI technologies continue to evolve, their potential to revolutionize healthcare delivery is becoming increasingly evident [10]. Machine learning (ML) is a subfield of AI that enables systems to learn from data without being explicitly programmed. ML algorithms can identify patterns, predict, and adapt to new information [11–13]. Some research has addressed difficult airway prediction and management using machine learning (Table 1).

These studies show that Machine learning algorithms hold significant promise for predicting airway management challenges and the need for intubation, yet challenges such as selecting more comprehensive predictive features, enhancing data quality, and conducting more rigorous model evaluations persist.

Despite its clinical and research significance, no studies have utilized machine learning approaches to predict a difficult airway from this study's collated and localized data. Therefore, a data balancing technique was applied to this dataset, and significant features were selected. Then, machine learning algorithms were utilized to predict difficult airway management.

Methods

Study design, setting, and participants

This cross-sectional study assessed 622 adult patients scheduled for elective surgery and referred to Loghman Hakim and Labbafinezhad Hospitals, Tehran, Iran. Inclusion criteria: (1) Patient scheduled for elective surgery requiring general anesthesia and tracheal intubation, (2) The patient should not have any contraindications to the administration of neuromuscular blocking agents, (3) Age greater than 18 years, (4) Ability to open the mouth with inter incisor distance greater than 2 centimeters (for laryngoscope placement), (5) absence of recent cervical spine trauma (< 2 months), (6) No contraindications to anesthesia induction before airway placement. Exclusion criteria: (1) The development of new-onset acute neck pain or injury between the clinic visit and surgery, (2) Patient's consent withdrawal from the study at any time. Demographic information and a history of medical comorbidities were preoperatively collected from the patient using a structured questionnaire (Appendix). Airway assessments were performed and documented in standardized forms according to the predefined variables. A metal inclinometer and a mobile inclinometer app (Goniometer Records, Orthopedic Research Group Initiative, Indian Orthopedic Research Group, Last updated September 2018) are the two primary tools for measuring the cervical range of motion angles. Anesthesia induction was tailored to the patient's condition following a complete pre-operative assessment. Ventilation was initiated using an ambu bag, and the bag-mask ventilation score (Grade I: easy to ventilate, Grade II: requiring nasopharyngeal or oral airway to ventilate, Grade III: difficult to ventilate or require two providers, Grade IV: unable to ventilate using all the above maneuvers. Grades III and IV are defined as difficult mask ventilation) [19] was documented. Subsequently, laryngoscopy was performed using a Macintosh laryngoscope with an appropriately sized blade based on the patient's anatomy. The Cormack-Lehane score was recorded based on the observations and reports of the intubating personnel. In the event of a difficult airway at any stage, the necessary equipment or personnel were utilized based on the patient's condition. Data were entered daily into an Excel spreadsheet by an anesthesia assistant. These data were then analyzed using the Python programming language, and machine learning algorithms were applied and evaluated.

Sezari *et al. BMC Anesthesiology* (2024) 24:453

Page 3 of 14

Table 1 Related work

Row	Title	Methodology	Research Variables	Evaluation Results
1	Predictive Machine Learning Algorithms in Anticipating Problems with Airway Management [14]	The study aims at the use of artificial intelligence (ML) methods to foresee airway management obstacles. The methods used are both supervised and unsupervised machine learning. Some relevant models include Decision Trees, Random Forests, SVM, and Multi-layer Perceptron Neural Networks. The models are employed to tackle classification issues. For instance, the training of algorithms with clinical data via supervised learning models to predict outcomes, such as difficult airway (DA), difficult mask ventilation (DMV), and difficult intubation (DI) using input features like BMI, inter-incisor distance, Mallampati score, and neck extension limitation are the algorithms' tasks.	The set of features the algorithms were trained on includes the physical parameters such as BMI (> 30 kg/m²), inter-incisor distance (IID) (< 2 cm), Modified Mallampati (MMP) scores (Grade 3 and 4 indicate anticipated DI), thyromental distance (TMD) (< 6.5 cm), restricted neck extension, receded mandible, and poor submandibular compliance. These variables serve the purpose of predicting airway difficulties in clinical assessments.	The ways to evaluate the models are accuracy, sensitivity, specificity, and precision, and the study shows that many models such as GBM and XGBM usually do better than simpler models because of their error reduction and their prediction-improving feature by sequential learning. Finally, GBM and Logistic Regression were the best models in terms of offering both high accuracy and good discrimination. Nonetheless, the model selection is the primary healthcare concern and the amount of training data.
2	Development of a machine learning algorithm to predict intubation among hospitalized patients with COVID-19 [15].	The purpose of this study was to create a machine learning algorithm that would forecast intubation in COVID-19 hospitalized patients. Using data from 4,087 patients admitted between February and April 2020, a retrospective cohort design was employed. The method turned patient demographic, vital, and lab data into time-series data for model training using a random forest classifier. The model updated every 12 h and generated forecasts based on 24-hour windows.	Physical parameters like BMI (> 30 kg/m2), inter-incisor distance (IID) (< 2 cm), Modified Mallampati (MMP) scores (Grades 3 and 4 indicate anticipated DI), thyromental distance (TMD) (< 6.5 cm), restricted neck extension, receding mandible, and poor submandibular compliance are among the features that the algorithms were trained on. In clinical examinations, these indicators are used to predict airway issues. Among the variables were: 1. Vitals: pulse, oxygen saturation, systolic and diastolic blood pressure, and respiratory rate. 2. Laboratory data: levels of D-Dimer, creatinine, C-reactive protein, platelet count, white blood cell count, and arterial O2 and CO2. 3. Comorbidities include diabetes, chronic obstructive lung disease, liver disease, renal disease, and hypertension.	The performance of the model was assessed by the study using the following metrics: 1. Area Under the Curve (AUC): Evaluate how well the model distinguishes between positive and negative instances. 2. Area Under the Precision-Recall Curve (AUPRC): This measures how well the model can detect real positives while maintaining sensitivity. 3. Kaplan-Meier Survival Analysis: To assess the rates of intubation-free patients who were informed by the model vs. those who were not. The model's AUC is 0.84, whereas the ROX index is 0.64. AUPRC: The 0.30 score of the model beat the 0.13 score of the ROX index. The model-identified patients had a noticeably increased risk of intubation during their hospital stay.
3	Predicting the need for intubation in the first 24 h after critical care admission using machine learning approaches [16].	Within 24 h after ICU admission, the study creates a machine-learning model to forecast the necessity for intubation. With information from two sizable databases, MIMIC-III and eICU-CRD, containing more than 17,000 critically ill patients, it employs a retrospective cohort design. Two machine learning models, Random Forest (RF) and Logistic Regression (LR), were used to complete the prediction challenge. Autoencoders (AEs) were used to impute missing data, and models trained on 60% of the available data were tested on the 40% that remained.	The following are the main variables in the model: Demographics: Medical specialization, age, and gender. Vital signs include heart rate, respiration rate, systolic and diastolic blood pressure, oxygen saturation (SpO2), and Glasgow Coma Scale (GCS). Laboratory values: HCO3, PaO2, and PaCO2 are examples of blood gas parameters. Interventions include oxygen therapy and the use of vasopressors.	The following measures were used to assess the models' performance: The model's ability to distinguish between patients who are intubated and those who are not is measured by the area under the receiver operating curve, or AUC. Sensitivity: The model's capacity to accurately identify intubation-required patients. Specificity: The capacity to accurately determine which patients do not require intubation. The likelihood that genuine predictions will be made for both intubated and non-intubated instances is shown by Positive and Negative Predictive Values (PPV, NPV). With an AUC of 0.86 as opposed to 0.77 for Logistic Regression, the Random Forest model outperformed LR. The Random Forest model's sensitivity is 0.88, while its specificity is 0.66. Over the whole range of intubation risk projections, the Random Forest model showed good calibration.

Sezari *et al. BMC Anesthesiology* (2024) 24:453

Page 4 of 14

Table 1 (continued)

Row	Title	Methodology	Research Variables	Evaluation Results
4	Machine Learning Approaches for Predicting Difficult Airway and First-Pass Success in the Emergency Department: Multicenter Prospective Observational Study [17].	The research, a multicenter prospective observational study, was carried out in 13 Japanese emergency departments (EDs). 10,741 patients who had tracheal intubations between January 2010 and December 2018 were included in the dataset. Seven machine learning models, such as XGBoost, gradient boosting, and random forest, were created utilizing regularly gathered information on patient demographics and vital signs before intubation. The capacity of these algorithms to forecast two outcomes—difficult airway and first-pass intubation success—was used to assess their effectiveness.	The following are independent variables (predictors): Glasgow Coma Scale, pre-intubation vital signs (pulse rate, systolic blood pressure, respiratory rate, oxygen saturation), patient demographics (age, sex, BMI, etc.), elements of the modified LEMON criteria, and intubation-related factors (medications, intubation techniques, devices used). Dependent factors, or results: A difficult airway needs to be intubated more than once. First-pass success is the state in which intubation goes well on the first try.	With the exception of the k-point closest neighbor and multilayer perceptron, machine learning models fared better than the modified LEMON criterion for predicting problematic airways; the ensemble model had the highest c-statistic, at 0.74. With the exception of random forest and k-point closest neighbor, machine learning models surpassed the logistic regression reference model in first-pass success prediction (the ensemble model had the highest c-statistic of 0.81). In both scenarios, the ensemble model outperformed conventional techniques in terms of sensitivity and specificity.
5	Machine learning for the prediction of preclinical airway management in injured patients: a Registry-based trial [18].	A retrospective study was carried out using a registry-based dataset of adult trauma patients in Germany who received emergency medical care between 2018 and 2020. Random Forest and Naïve Bayes machine learning algorithms were employed to forecast the necessity of preclinical airway control. There were 25,556 patients in all, and 1,451 of them needed breathing assistance. Principal component analysis (PCA) was utilized to pick preprocessed attributes from the dataset that were the subject of the investigation. Important clinical factors that were taken into account for model training were auscultation, damage patterns, oxygen therapy, and shock index.	The following are examples of independent variables or features: thoracic drainage, oxygen therapy, noninvasive breathing, injury patterns, vital signs, and the usage of specific drugs, such as catecholamines. Dependent variable (result): Preclinical airway care is necessary. The Glasgow Coma Scale (GCS), starting heart rate, systolic blood pressure, and respiration rate are additional significant characteristics.	Regarding performance, the Random Forest (RF) model outperformed the Naïve Bayes (NB) model. AUC-ROC (area under the receiver operating characteristic) for RF was 0.96, and its overall accuracy (97.8%) was higher than that of NB. Additionally, the RF model was better at predicting the requirement for airway care than the NB model, as evidenced by its greater positive predictive value (PPV) of 0.85 compared to 0.46. When it came to the precision-recall area (0.83), the RF model outperformed the NB model (0.66).

Sezari *et al. BMC Anesthesiology*   (2024) 24:453

Page 5 of 14

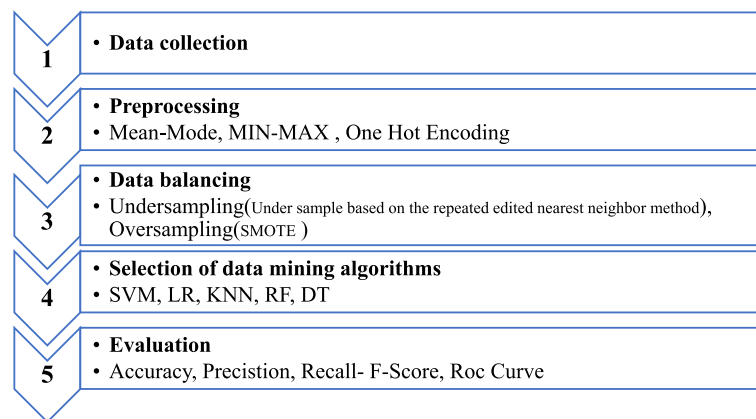**Fig. 1** Shows the step-by-step breakdown of the methodology

Data preprocessing

First, features that had the same value in most records were removed. Then the issue of missing data in continuous features was addressed using imputation by mean. Continuous features were then normalized using the MIN-MAX formula, while categorical features were converted to binary representations using One-Hot Encoding. This concluded the data preprocessing stage.

Data balancing

Before data balancing, the 10-fold cross-validation method separated the training and testing data. Due to

Table 2 Features that have a discrete value

Features	Feature importance	Mean	Std	Description
Age(year)	0.062004	43.098	14.08	-
BMI (kg/m^2)	0.041490	27.698	5.83	Body Mass Index
Mouth Opening(cm)	0.068357	4.376	0.77	The linear distance between the edge of the upper tooth (gum) and the edge of the lower tooth on the same side.
Neck Circumference(cm)	0.086946	37.405	3.63	The measurement of Neck Circumference (NC) entails assessing the circumference of the neck in its neutral position, just below the Adam's apple.
Neck Length(cm)	0.023407	11.165	1.33	This parameter is determined by measuring the distance from the external occipital protuberance to the seventh cervical vertebra in the neutral position of the neck. A measurement of less than 7 cm is correlated with an increased probability of encountering a difficult airway.
Thyromental Distance(cm)	0.030534	9.361	1.49	It is determined by measuring the linear distance between the thyroid notch and the mentum in the condition of full neck extension.
Sternomental Distance(cm)	0.106998	16.648	2.11	The linear distance between the upper border of the manubrium of the sternum and the mentum in the condition of full neck extension.
Rhinionmentum Distance(cm)	0.023414	6.795	0.87	The linear distance between the mentum and the base of the nasal septum.
Neck extension degree (clinometry)	0.044165	43.511	11.75	The angle between the horizon line and the neck in the state of maximum neck extension.
Anterior neck Flexion degree (clinometry)	0.039345	71.645	13.11	The angle between the horizon line and the neck in the state of maximum flexion of the neck.
left lateral neck flexion degree (clinometry)	0.032863	45.463	10.50	The angle between the midline and the maximum bending of the neck to the left.
Right lateral neck flexion degree (clinometry)	0.032591	47.778	11.02	The angle between the midline and the maximum bending of the neck to the right.
Neck extension degree (application)	0.037774	41.502	13.13	The angle between the horizon line and the neck in the state of maximum neck extension.
Anterior neck Flexion degree(application)	0.053250	73.685	14.56	The angle between the horizon line and the neck in the state of maximum flexion of the neck.
left lateral neck flexion degree (application)	0.025654	40.704	12.10	The angle between the midline and the maximum bending of the neck to the left.
Right lateral neck flexion degree (application)	0.030172	53.357	19.62	The angle between the midline and the maximum bending of the neck to the right.
AASI	0.031471	0.622	0.14	AcromioAxillo Suprasternal Notch Index.
Hyomental Distance(cm)	0.045787	5.814	0.98	The linear distance between the hyoid bone and the mentum in the condition of full neck extension.
HMDn	0.025921	4.152	0.69	The ratio of hyomental distance in neutral head position
HMDR	0.027409	1.418	0.25	The ratio of hyomental distance in neutral head position (HMDn) to hyomental distance in full extension of the head (HMDe).

Sezari *et al. BMC Anesthesiology*   (2024) 24:453

Page 6 of 14

the imbalanced nature of the data, various balancing techniques were employed in this step. This phase was carried out in two distinct steps:

> Step 1: The SMOTE algorithm was applied to the training data in each fold, augmenting the records of the minority class to match the size of the majority class records. This procedure resulted in a balanced training dataset. The prepared training data was then utilized for the subsequent phase of the experiment.
> Step 2: This step employed the Repeated Edited Nearest Neighbor (RENN) under-sampling technique to address the data imbalance. This method involves reducing the majority class data points to match the size of the minority class. Following data balancing, the hold-out method was utilized to divide the balanced dataset into training (70%) and testing (30%) sets. Feature importance was assessed using a forest of trees.

Selecting data mining algorithms

In this phase, various classification algorithms, including Logistic Regression, SVM (RBF), SVM (POLY), Random Forest (INFORMATION-GAIN), Random Forest (GINI-INDEX), Decision Tree, and KNN (Euclidean, Manhattan, Cosine, Minkowski), were employed for data mining on both the balanced datasets obtained from step 1 and step 2. Below is a brief description of each algorithm:

Logistic regression is a statistics technique that is utilized to model the relationship between a binary outcome variable and one or more predictor variables. In logistic regression, the estimate of the probability of the outcome variable is based on the values of the predictor variables [20]. SVM stands out as one of the most efficacious algorithms in machine learning, especially for classification issues. SVMs can do this by constructing an optimal hyperplane that effectively partitions the data points that belong to different classes. This is done by maximizing the margin between the classes, which ultimately leads to a model with a superior ability to generalize [21].

Decision trees provide a tree-like hierarchical structure for use in machine learning models. This structure features inner nodes representing data attributes, branches representing decision rules based on those attributes, and terminal leaf nodes representing the predicted outcome or class label [22]. Random Forest is an ensemble learning technique that uses decision trees as its base learners. It is operated by constructing a large number of decision trees. Their predictions are then aggregated to improve the final classification or regression output. The core principle of Random Forest is to generate a collection of decision trees during the training phase, with each tree producing a distinct prediction. In the predicting phase, the Random Forest algorithm achieves the result by combining the predictions of all the constituent decision trees [23]. KNN is a common nonparametric machine learning algorithm that can be applied to classification and regression problems. Its function is the identification of the k-nearest data points (neighbors) of a new data point. Then, based on the majority vote (classification) or the average (regression) of the labels or values of its K-nearest neighbors, the class label or value of the new data point is predicted [24].

Algorithm evaluation

In this phase, the performance of the algorithms was assessed using a suite of evaluation metrics, including

Table 3 Features that have a continuous value

Features	Feature importance	Frequency N (%)				Description
		Class 1	Class 2	Class 3	Class 4	
Loose front teeth	0.018674	113(18)				Class 1: present loose tooth/teeth.
MASK Ventilation Score	0.034665	372(60)	132(22)	112(18)		Grade I: easy to ventilate, Grade II: requires naso-pharyngeal or oral airway to ventilate, Grade III: difficult to ventilate or requires two providers, Grade IV: unable to ventilate using all the above maneuvers. Grades III and IV are defined as difficult mask ventilation)
Mallampati Score	0.030622	190(31)	220(35)	190(31)	22(3)	This classification system comprises four classes. To assess it, the patient is requested to open their mouth and protrude their tongue silently, allowing for observation of the pharyngeal structures. The classification is as follows: 1: Soft palate, pharynx, entire uvula, and pillars visible, 2: Soft palate, pharynx, and part of the uvula visible, 3: Soft palate and base of the uvula visible, 4: Only the hard palate visible.
Snoring	0.027833	165(26)				History of snoring during sleep declared by the patient

Sezari *et al. BMC Anesthesiology*   (2024) 24:453

Page 7 of 14

accuracy, specificity, precision, recall, F-measure, and the area under the ROC curve (AUC). These metrics provided comprehensive insights into the performance of each algorithm. We define the evaluation indices as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{F} - \text{measure} = \frac{(2 \text{ x Precision x Recall})}{(\text{Precision} + \text{Recall})} \tag{5}$$

In Eq. (1) to (4), TP, TN, FP, and FN denote the True-Positive, True-Negative, False-Positive, and False-Negative, respectively. The result of Eq. (5) is calculated using the results of Eq. (3) and Eq. (4). Figure 1 shows the step-by-step breakdown of the methodology employed in this study.
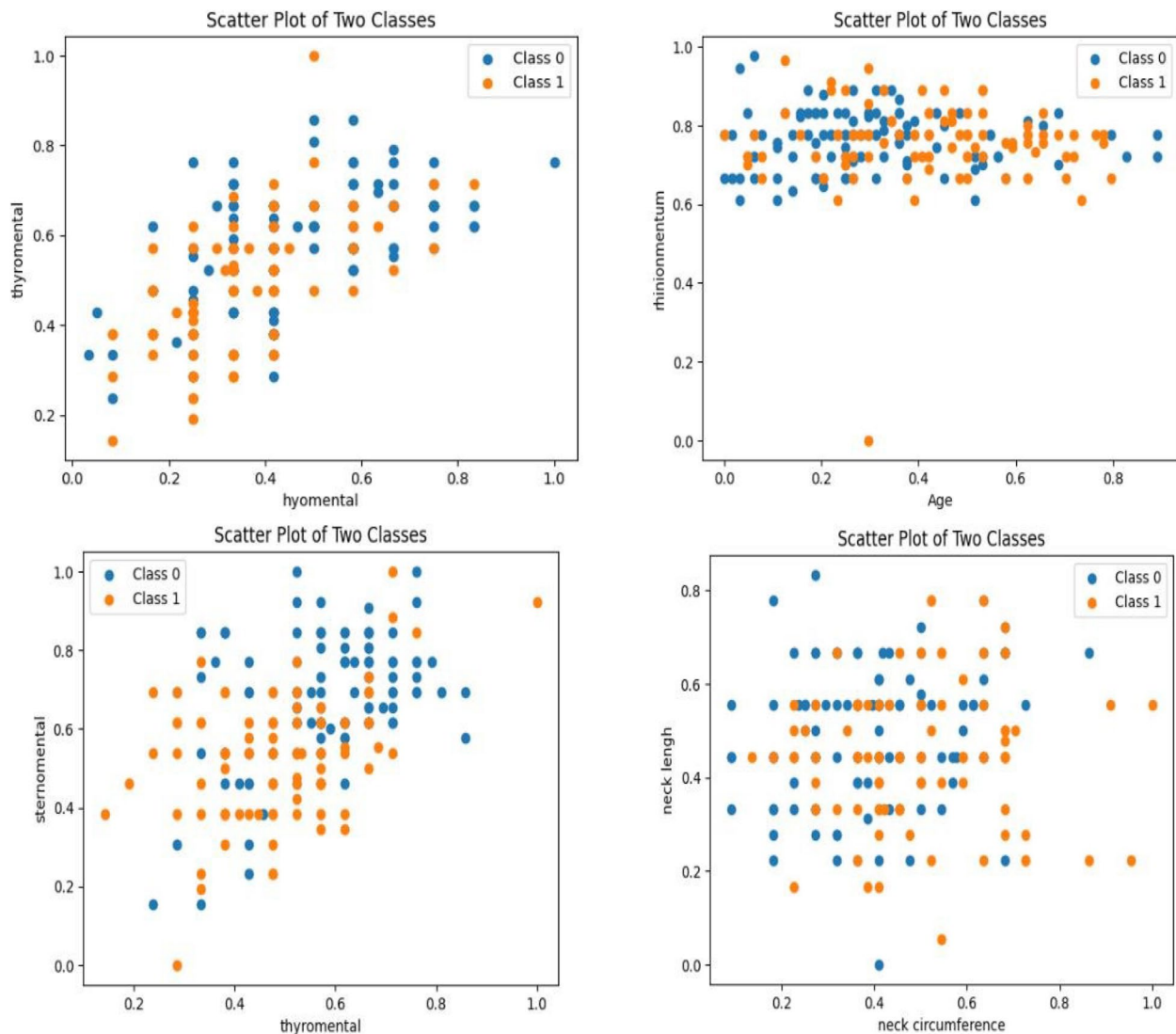

Fig. 2 Scatter plot of selected variables concerning two classes after balancing (under-sampling based on the repeated edited nearest neighbor method)
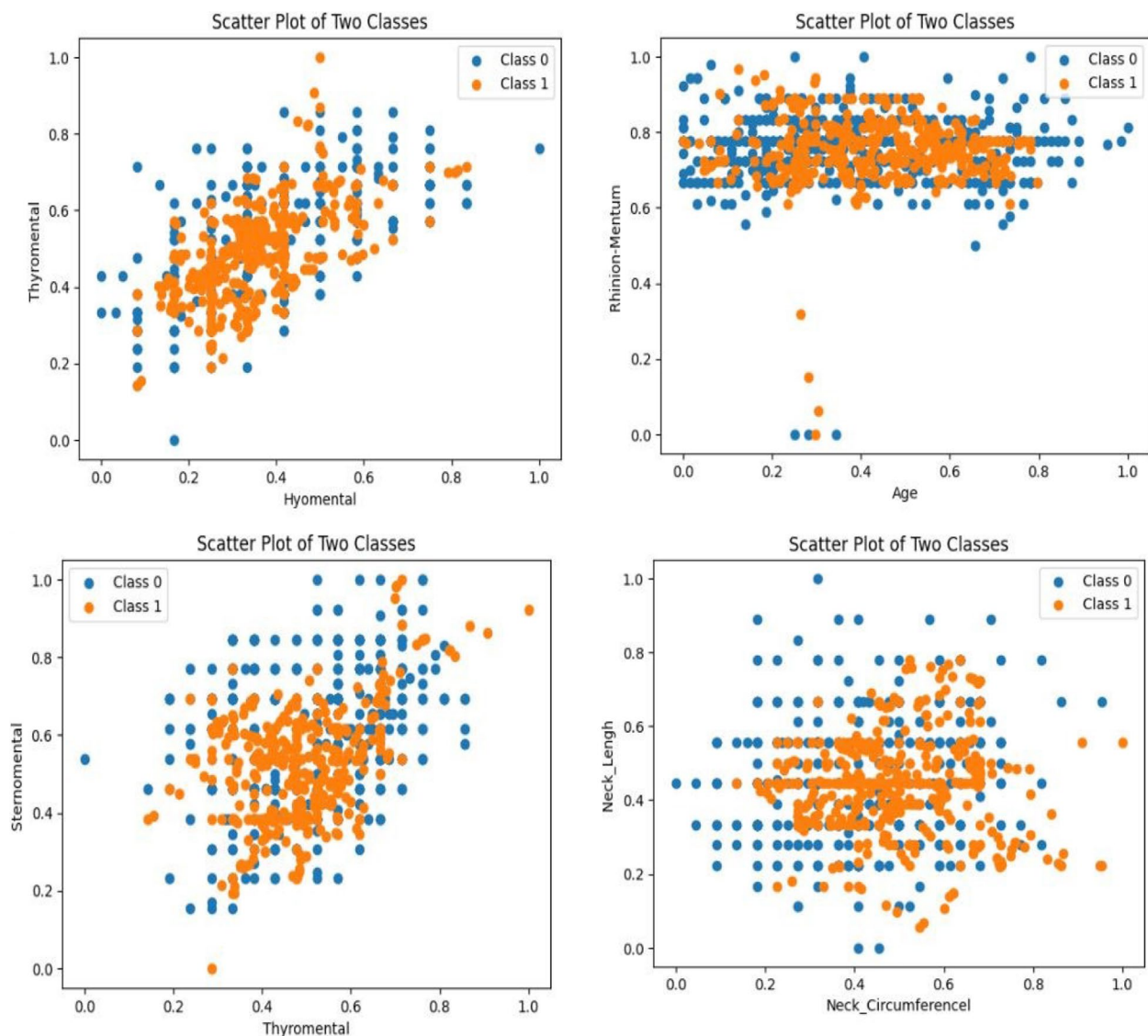
Sezari *et al. BMC Anesthesiology*   (2024) 24:453

Page 8 of 14

Fig. 3 Scatter plot of selected variables for two classes before balancing

Results

A total of 622 patient records were collected. The mean age of the patients was 3.09 ± 14.08 years. Out of 32 features in the collected data, 24 important features were identified. 20 features had discrete values and 4 features were continuous. Sternomental Distance was the most important (0.10) among the features. Table 2 shows the features with a discrete value, the importance, mean, and standard deviation. Table 3 shows the continuous features, their importance, and their classes.

Cormack-Lehane score was the target class. This classification system is based on the visualization of structures during laryngoscopy and is divided into four grades: grade 1: Complete visualization of the laryngeal entrance, grade 2: Partial visualization of the posterior aspect of the laryngeal entrance, grade 3: Only the epiglottis is visible, grade 4: No part of the epiglottis or larynx is visible. If Cormack-Lehane grade 3 is observed, it indicates a difficult laryngoscopy (class 2). Otherwise, it is labeled as Class 1. Class 1 included 531 patients and class 2 included 91 patients. In this study, there were no patients with grade 4.

Figures 2 and 3 illustrate scatter plots depicting the relationships between two variables and their corresponding class labels, both before and after data balancing. These scatter plots provide visual insights into the distribution of data points and the potential existence of class imbalances.

Sezari *et al. BMC Anesthesiology*   (2024) 24:453

Page 9 of 14

To evaluate the data's normality, Q-Q plots were employed in addition to scatter plots. These charts provide a comparison between the observed data's quantiles and a theoretical normal distribution. The data points in the Q-Q plots deviated from the expected diagonal line, indicating deviations from normalcy prior to balancing. On the other hand, the Q-Q plots improved once the data was balanced, with the data points aligning closer to the diagonal, indicating a more normal distribution of the data. Because many machine learning models, including some utilized in this study, perform better when the data is closer to a normal distribution, this enhanced normality is significant.

The data balancing approach, as demonstrated by the scatter plots and Q-Q plots, not only resolved the problem of class imbalance but also enhanced the dataset's statistical qualities, which enabled the machine learning models to produce predictions that were more precise and trustworthy. Finally, the robustness of the study's conclusions was enhanced by these visual aids, which offered important proof that the preprocessing procedures were successful. Figures 4 and 5 present Q-Q plots to assess the normality of the data both before and after balancing.

Table 4 shows a comparative analysis of various machine learning algorithms evaluated after applying two data balancing techniques: SMOTE and the Under-sampling Based on the Repeated Edited Nearest Neighbors (RENN). The KNN algorithms with K=3 generally outperformed those with K=5. Minkowski and Euclidean metrics with K=3 yielded the highest accuracy (0.87). SVM (POLY and RBF) achieved respectable accuracies
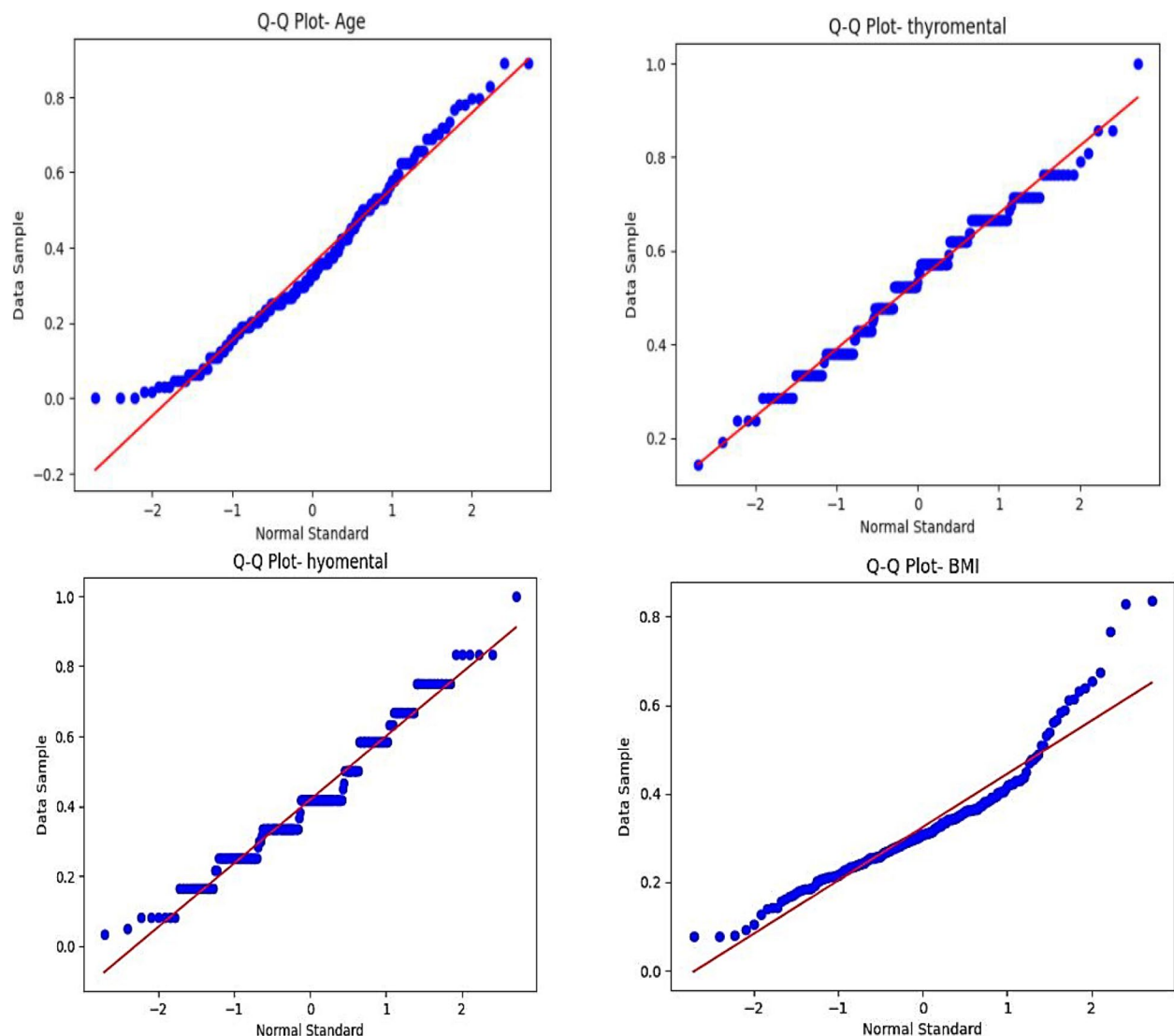

Fig. 4 Normality of selected variables after balancing (under-sampling based on the repeated edited nearest neighbor method)

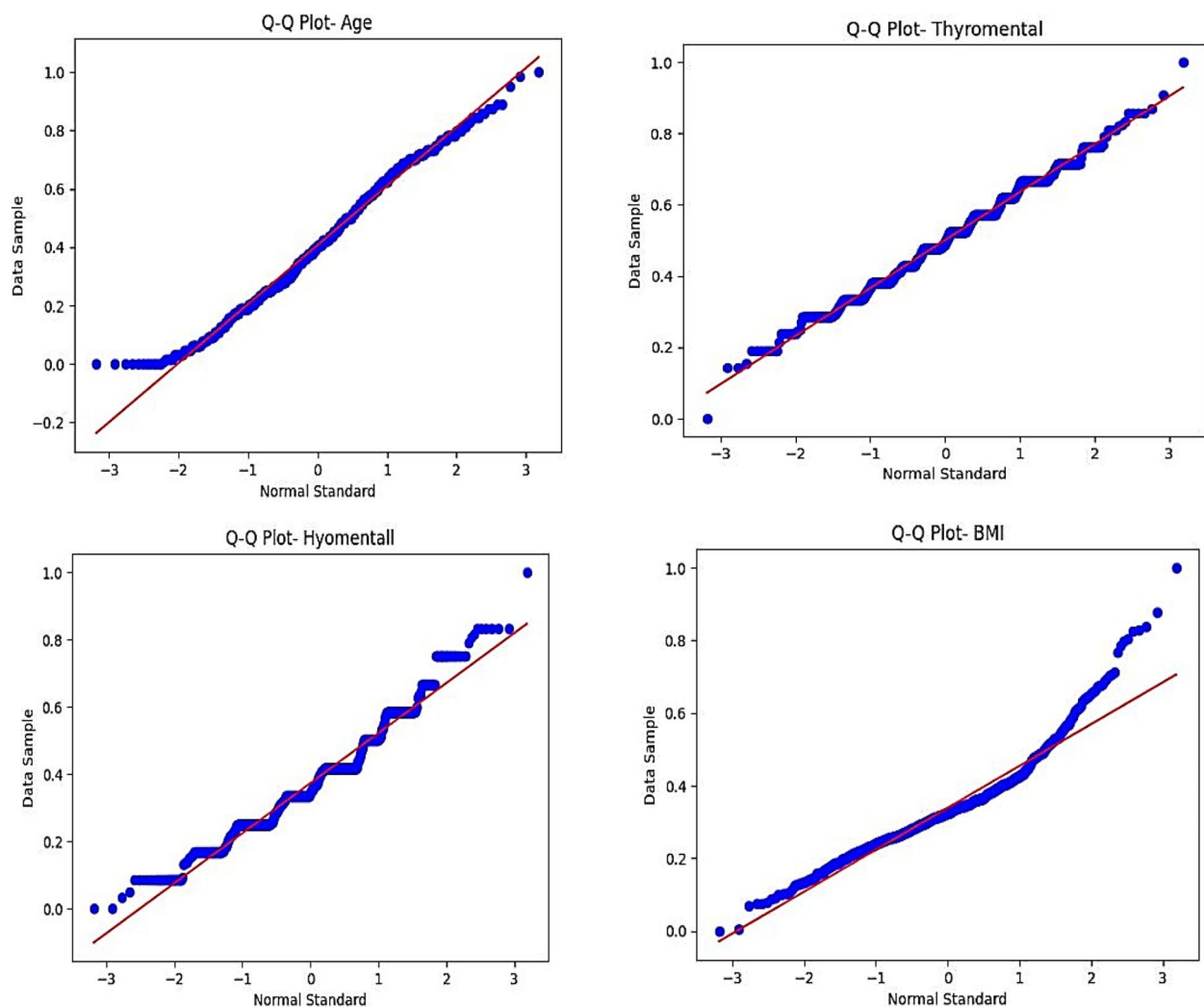Sezari *et al. BMC Anesthesiology*   (2024) 24:453

Page 10 of 14

Fig. 5 Normality of selected variables before balancing

of 0.82 and 0.80, respectively. The Decision Tree achieved a moderate accuracy of 0.77. Random Forest had slightly better accuracy when using Information Gain (0.76) than the Gini index (0.74). Logistic Regression, with an accuracy of 0.71, demonstrated the weakest performance compared to the other models.

Algorithms in the RENN method generally outperformed SMOTE across all metrics. When using the SMOTE technique, only Random Forest was able to produce the same level of accuracy as RENN.

Figure 6 shows the ROC curves for different classification algorithms when the data has been balanced using the Under-sampling Based on the RENN Method. The best performance was achieved when the number of neighbors was optimally set to K = 3 and the Minkowski metric was used, which allowed greater flexibility in calculating the distance. The SVM models with RBF and Polynomial kernels showed acceptable

performance and the Polynomial kernel was slightly better. Moderate performance was seen by the Random Forest (Information-Gain and Gini-Index). The Decision Tree, with an AUC of 0.77, outperformed Logistic Regression but still demonstrated lower accuracy and discriminative power compared to other models.

According to the findings presented in Table 4; Fig. 6, the KNN algorithm with three neighbors and Euclidean, Minkowski, and Cosine distance metrics outperformed the other algorithms.

Discussion

This study aims to determine how machine learning algorithms can be used to predict outcomes in challenging airway management, a critical area of anesthesia. Several algorithms, including KNN, Random Forest, Logistic Regression, and SVM were trained using data from 622 patient records gathered from two hospitals in Iran. 24

Sezari *et al. BMC Anesthesiology* (2024) 24:453

Page 11 of 14

Table 4 Performance of different classification algorithms after data balancing

Metric	Balanced Data- Under sample based on the repeated edited nearest neighbor method				Balanced Data- SMOTE			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Euclidean (K = 5)	0.84	0.87	0.75	0.80	0.68	0.17	0.37	0.22
Manhattan (K = 5)	0.80	0.83	0.71	0.76	0.67	0.17	0.39	0.23
Cosine (K = 5)	0.80	0.86	0.67	0.76	0.66	0.17	0.38	0.22
Minkowski (K = 5)	0.84	0.87	0.75	0.80	0.65	0.18	0.45	0.25
Euclidean (K = 3)	0.87	0.88	0.82	0.85	0.73	0.13	0.16	0.13
Manhattan (K = 3)	0.82	0.84	0.75	0.79	0.70	0.12	0.17	0.13
Cosine (K = 3)	0.85	0.88	0.78	0.83	0.68	0.12	0.20	0.14
Minkowski (K = 3)	0.87	0.88	0.82	0.85	0.67	0.12	0.20	0.14
Logistic Regression	0.71	0.70	0.60	0.65	0.66	0.17	0.40	0.23
Svm (RBF)	0.80	0.78	0.78	0.78	0.66	0.15	0.32	0.20
Svm (POLY)	0.82	0.84	0.75	0.79	0.66	0.16	0.35	0.21
Random Forest (INFORMATION-GAIN)	0.76	0.78	0.64	0.70	0.76	0.13	0.14	0.12
Random Forest (GINI-INDEX)	0.74	0.77	0.60	0.68	0.73	0.07	0.16	0.10
Decision Tree	0.77	0.79	0.67	0.73	0.65	0.15	0.33	0.21

significant features were found when the data was pre-processed and balanced, with Sternomental Distance being the most significant. The KNN method yielded the best accuracy of 0.87, especially when Euclidean and Minkowski metrics were used. These findings imply that machine learning can dramatically improve prediction accuracy in clinical settings, improving patient outcomes and enhancing the efficacy of airway control techniques.

Zhou's research showed that machine learning models, particularly gradient boosting algorithms, were very successful in predicting difficult airway intubations, which is similar to the current study. With an AUC of more than 0.80 and a flawless precision score, the Gradient Boosting Machine (GBM) proved to be the most effective algorithm in their investigation [25]. The utilization of physical metrics like BMI and Sternomental Distance and the excellent performance of ML models like GBM were constant across both investigations, although KNN was not the best model in Zhou's investigation. Wang et al. showed that the performance of machine learning algorithms, namely Naïve Bayes and Random Forest, was best for predicting difficult tracheal intubation (DTI) and difficult laryngoscopy (DL) with AUC values of 0.95 and 0.90 respectively. Linear models such as Logistic Regression and Decision Tree are less accurate and sensitive to predicting intubation difficulties than these models. The results of the present study are consistent with these findings, especially for the superior performance of complex models over simpler ones [26]. Hayasaka et al.

found that Convolutional Neural Networks (CNN) can predict intubation difficulty with high accuracy (AUC = 0.86) using facial images [27]. The closeness of the AUC values between the CNN model and the KNN model in this study indicates how machine learning techniques can be applied broadly to a variety of methods, such as facial image analysis and measurements of the neck circumference and mouth opening. Xia et al. trained a neural network using facial images and the Light Gradient Boosting Machine algorithm, achieving a commendable Area Under the Curve (AUC) of 0.77 for predicting intubation difficulty [28]. Although this AUC was slightly lower than in the current study, probably because the feature sets and models were different, it nevertheless shows how consistently machine learning models outperform conventional techniques in terms of prediction accuracy. Cuendet et al.'s study, utilizing Random Forest algorithms, yielded an AUC of 0.77, which while promising, was lower than the AUC of the KNN model employed in the current study [29]. Nonetheless, to improve prediction accuracy, both research stresses how crucial it is to combine various features. Although the ensemble aspect of Random Forest contributed to its good performance, KNN's use of distance-based metrics appeared to provide better results in this instance.

In summary, machine learning techniques show promising results for predicting difficult airway management. The choice of physical parameters used to train the algorithm impacts the accuracy of these
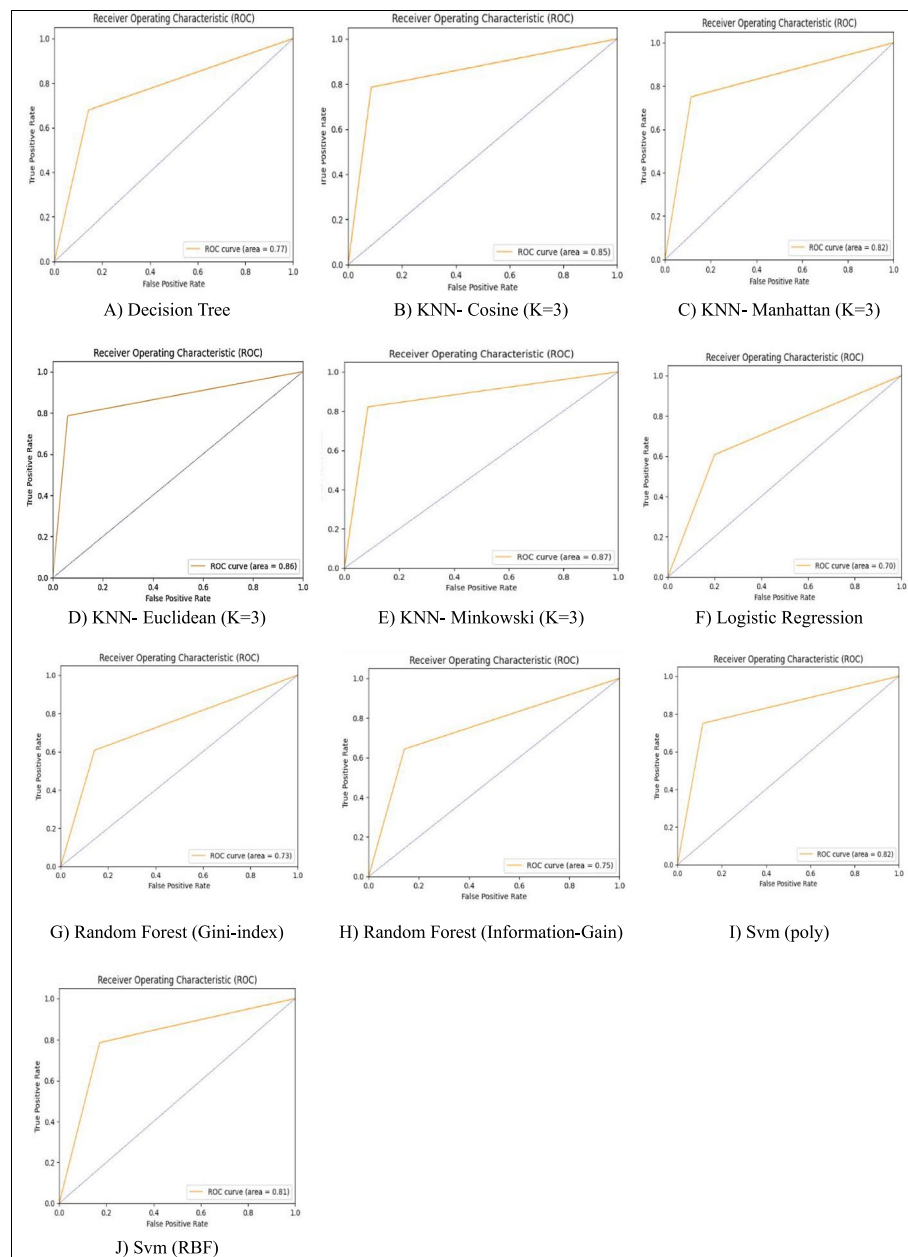
Sezari *et al. BMC Anesthesiology*   (2024) 24:453

Page 12 of 14

Fig. 6 ROC curves of different classification algorithms after balancing (under-sampling based on the repeated edited nearest neighbor method)

algorithms. With the right choice of parameters, machine learning algorithms can significantly impact clinical practice by improving the prediction of difficult airways, which can also lead to better patient outcomes.

Limitations and future research directions

Using only one type of anesthesia technique makes the results non-generalizable to other anesthesia methods. The model could become more accurate in many situations if we consider other more diverse techniques.

Furthermore, cases of emergency surgery were not included in this study. However, this may limit the model's applicability in emergencies since emergencies need faster diagnoses and better models. Variables used may not always be available in clinical settings or may not be easy to measure. Such a limitation would restrict the use of the model in general environments or emergencies. The exclusion of patients of different age groups (children or elderly) may make the findings unreliable for some age groups (e.g. children, elderly). The different physiological

Sezari *et al. BMC Anesthesiology*   (2024) 24:453

Page 13 of 14

characteristics in these groups may require specific algorithms. Given that the data was collected by two specialists, there is a potential for human error, which could affect the accuracy of the data and, ultimately, the model's accuracy. It is recommended that in future research, additional features, including physiological and biological characteristics (such as cardiovascular or respiratory health status), be added to the model. This could enhance prediction accuracy in various clinical conditions.

Incorporating imaging data, such as radiology images or CT scans of the airway, could assist machine vision algorithms in providing more accurate predictions in more complex conditions.

Examining the impact of specific characteristics of different age groups (children, elderly) or particular medical conditions (such as patients with chronic respiratory diseases) could help improve the model's performance in special conditions. Using deep learning algorithms and more complex neural networks could help better simulate complex airway patterns and improve prediction accuracy.

Conclusions

These studies show that machine learning, particularly KNN models, can provide accurate predictions in clinical contexts, assisting anesthesiologists in anticipating and better managing problematic airways. These algorithms are useful tools for enhancing patient outcomes and safety because of their capacity to examine several complex variables at once. The study concludes that machine learning has a lot of promise to improve preoperative planning by helping doctors identify patients who are more likely to experience difficulties in the airway and, as a result, reduce unfavorable outcomes following surgery. Applying such predictive models may result in more accurate and customized airway management techniques in practice, which would eventually lower the risk of problems and raise the standard of anesthesiology treatment as a whole.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12871-024-02842-w.

> Supplementary Material 1.

Declarations

Ethics approval and consent to participate
This article is extracted from a study approved by the Anesthesiology Research Center of Shahid Beheshti University of Medical Sciences (Ethics code: IR.SBMU.RETECH.REC. 1402.794). All methods of the present study were performed by the relevant guidelines and regulations of the Ethical Committee of Shahid Beheshti University of Medical Sciences. Participation was voluntary, the consent was verbal, but all participants responded via email or text message to approve their participation. Informed consent was obtained from all the participants. Participants had the right to withdraw from the study at any time without prejudice.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Anesthesiology, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [2]Department of Health Information Technology and Management, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, 1th floor, No 21, Darband St., Tajrish sq., Tehran, Iran. [3]Department of Anesthesiology, School of Medicine, Anesthesiology Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [4]Ilam County Health Center, Ilam University of Medical Sciences, Ilam, Iran.

References
1. Apfelbaum JL, Hagberg CA, Connis RT, Abdelmalak BB, Agarkar M, Dutton RP, et al. American society of anesthesiologists practice guidelines for management of the difficult airway. Anesthesiology. 2022;136:31–81. Available from: https://pubmed.ncbi.nlm.nih.gov/34762729/. Cited 2024 Apr 20.
2. 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. BJA: Br J Anaes. 2011;106:617–31. Available from: https://doi.org/10.1093/bja/aer058. Cited 2024 Apr 20.
3. Heidegger T. Management of the Difficult Airway. Longo DL, editor. New England Journal of Medicine. 2021;384:1836–47. https://www.nejm.org/doi/full/10.1056/NEJMra1916801.
4. Chrimes N, Bradley WPL, Gatward JJ, Weatherall AD. Human factors and the 'next generation' airway trolley. Anaesthesia. 2019;74:427–33. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/anae.14543. Cited 2024 Apr 20.
5. Nørskov AK, Rosenstock CV, Wetterslev J, Astrup G, Afshari A, Lundstrøm LH. Diagnostic accuracy of anaesthesiologists' prediction of difficult airway management in daily clinical practice: a cohort study of 188 064 patients registered in the Danish Anaesthesia Database. Anaesthesia. 2015;70:272–81. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/anae.12955. Cited 2024 Apr 21.
6. Hagiwara Y, Watase H, Okamoto H, Goto T, Hasegawa K. Prospective validation of the modified LEMON criteria to predict difficult intubation in the ED. Am J Emerg Med. 2015;33:1492–6 https://www.sciencedirect.com/science/article/abs/pii/S0735675715005173.

Sezari *et al. BMC Anesthesiology*   (2024) 24:453

Page 14 of 14

7. Nørskov AK, Wetterslev J, Rosenstock CV, Afshari A, Astrup G, Jakobsen JC, et al. Effects of using the simplified airway risk index vs usual airway assessment on unanticipated difficult tracheal intubation - a cluster randomized trial with 64,273 participants. BJA: Br J Anaes. 2016;116:680–9. Available from: https://doi.org/10.1093/bja/aew057. Cited 2024 Apr 21.
8. Rosenblatt WH, Yanez ND. A decision tree approach to airway management pathways in the 2022 difficult airway algorithm of the american society of anesthesiologists. Anesth Analg. 2022;134:910 Available from: /pmc/articles/PMC8986631/. Cited 2024 Apr 21.
9. Obermeyer Z, Emanuel EJ. Predicting the future — big data, machine learning, and clinical medicine. New Engl J Med. 2016;375:1216–9. Available from: https://www.nejm.org/doi/full/10.1056/NEJMp1606181. Cited 2024 Apr 20.
10. Lecun Y, Bengio Y, Hinton G, Deep learning. Deep learning. Nature. 2015;521:7553. Available from: https://www.nature.com/articles/nature14539. Cited 2024 Apr 20.
11. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science. 2015;349:255–60. Available from: https://pubmed.ncbi.nlm.nih.gov/26185243/. Cited 2024 Apr 20.
12. Kohzadi Z, Nickfarjam AM, Shokrizadeh Arani L, Kohzadi Z, Mahdian M. A comprehensive evaluation of ensemble learning methods and decision trees for predicting trauma patient discharge status using real-world data. Arch Trauma Res. 2023;12:137–49. Available from: https://archtrauma.kaums.ac.ir/article_181135.html. Cited 2024 Nov 23.
13. Demir F, Akbulut Y, Taşcı B, Demir K. Improving brain tumor classification performance with an effective approach based on new deep learning model named 3ACL from 3D MRI data. Biomed Signal Process Control. 2023;81:104424.
14. Senthilnathan M, Kundra P. Predictive machine learning algorithms in anticipating problems with airway management. Airway. 2023;6:4–9. Available from: https://journals.lww.com/arwy/fulltext/2023/06010/predictive_machine_learning_algorithms_in.2.aspx. Cited 2024 Oct 19.
15. Arvind V, Kim JS, Cho BH, Geng E, Cho SK. Development of a machine learning algorithm to predict intubation among hospitalized patients with COVID-19. J Crit Care. 2021;62:25–30.
16. Siu BMK, Kwak GH, Ling L, Hui P. Predicting the need for intubation in the first 24 h after critical care admission using machine learning approaches. Sci Rep. 2020;10:1–8. Available from: https://www.nature.com/articles/s41598-020-77893-3. Cited 2024 Oct 19.
17. 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. 2022;11:e28366. Available from: http://www.ncbi.nlm.nih.gov/pubmed/35076398. Cited 2024 Oct 19.
18. Luckscheiter A, Zink W, Lohs T, Eisenberger J, Thiel M, Viergutz T. Machine learning for the prediction of preclinical airway management in injured patients: a registry-based trial. Clin Exp Emerg Med. 2022;9:304. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC9834832/. Cited 2024 Oct 19.
19. Khan M, Siddiqui AS, Raza SA, Samad K. Incidence and predictors of difficult Mask Ventilation in High-Risk Adult Population scheduled for elective surgery. A Prospective Observational Study; 2022.
20. Fernandes AAT, Filho DBF, da Rocha EC, da Silva Nascimento W. Read this paper if you want to learn logistic regression. Revista De Sociologia E Política. 2020;28:11–1919.
21. Rizwan A, Iqbal N, Ahmad R, Kim DH. WR-SVM model based on the margin radius approach for solving the minimum enclosing ball problem in support vector machine classification. Appl Sci. 2021;11(10):4657.
22. Song YY, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry. 2015;27:130. Available from: /pmc/articles/PMC4466856/. Cited 2024 Apr 19.
23. Rigatti SJ. Random forest. J Insur Med. 2017;47:31–9. Available from: https://doi.org/10.17849/insm-47-01-31-39.1. Cited 2024 Apr 19.
24. Uddin S, Haque I, Lu H, Moni MA, Gide E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Sci Rep. 2022;12:6256. Available from: /pmc/articles/PMC9012855/. Cited 2024 Apr 19.
25. 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. Available from: https://pubmed.ncbi.nlm.nih.gov/36033770/. Cited 2024 Apr 19.
26. Wang B, Li X, Xu J, Wang B, Wang M, Lu X, et al. Comparison of Machine Learning Models for Difficult Airway. J Anesth Translational Med. 2023;2:21–8.
27. 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 (convolutional neural network) using face images: an observational study. J Intensive Care. 2021;9:1–14. Available from: https://pubmed.ncbi.nlm.nih.gov/33952341/. Cited 2024 Apr 19.
28. 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. Available from: https://pubmed.ncbi.nlm.nih.gov/38093485/. Cited 2024 Apr 19.
29. Cuendet GL, Schoettker P, Yüce A, Sorci M, Gao H, Perruchoud C, et al. Facial image analysis for fully automatic prediction of difficult endotracheal intubation. IEEE Trans Biomed Eng. 2016;63:328–9. Available from: https://pubmed.ncbi.nlm.nih.gov/26186767/. Cited 2024 Apr 19.

Publisher's Note