Egyptian Journal of Anaesthesia 33 (2017) 153–158

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Egyptian Journal of Anaesthesia

journal homepage: www.sciencedirect.com

Research article

Defining difficult laryngoscopy findings by using multiple parameters: A machine learning approach

Moustafa Abdelaziz Moustafa [a,*], Shahira El-Metainy [a], Khaled Mahar [b], Essam Mahmoud Abdel-magied [c]

[a] *Alexandria Faculty of Medicine, Alexandria, Egypt*
[b] *Arab Academy for Science and Technology, Alexandria, Egypt*
[c] *Faculty of Commerce, Alexandria, Egypt*

ARTICLE INFO

Article history:
Received 27 April 2016
Revised 14 February 2017
Accepted 22 February 2017
Available online 6 March 2017

Keywords:
Difficult
Laryngoscopy
Machine learning

ABSTRACT

Background: Preoperative identification of patients whose trachea will be difficult to intubate would decrease the rate of anesthesia related adverse respiratory events. Each test for airway examination may predict a separate aspect of airway. A computer-based approach is tested in this study to precisely evaluate difficult laryngoscopy.

Aim of the work: Aim of the work was to evaluate the efficacy and accuracy of a multiparameter computer-based system for prediction of difficult laryngoscopy.

Study design: 50 Adult patients presenting for non emergency surgery at Alexandria main university hospital from February 2015 to Feruary 2016 with unanticipated difficult laryngoscopy were assessed postoperatively according to selected nine airway parameters. The same was done for their matched 50 controls after full recovery from general anesthesia. All data were entered into an information–based computer system where they were converted into numerical variables.

All data have been processed and analyzed using the Microsoft visual studio 2008 (C#.net) and WEKA (Waikato Environment for Knowledge Analysis) machine learning algorithms. Classification was done using J48 algorithm based on a decision tree and a "Weighter" filter was used to allow one to specify a numeric attribute to be used as an instance weight.

Results: Processed data have been designed as a software termed **"Alex Difficult Laryngoscopy Software" (ADLS)**. Positive predictive value was 76%, Negative predictive value was 76%, Matthews correlation coefficient was 0.52 and area under the ROC curve was 0.79.

Conclusion: **"Alex Difficult Laryngoscopy Software" (ADLS)** is a machine learning program for prediction of difficult laryngoscopy. New cases can be entered to the training set thus improving the accuracy of the software.

1. Background

Safe airway management is crucial in anesthesiology and resuscitation. Clearly, preoperative prediction of patients in whom laryngoscopy is difficult would decrease the rate of anesthesia related adverse respiratory events [1]. This requires a precise preoperative airway assessment. But still there is a controversy which tests and anatomical landmarks are the best predictors [2]. Difficult laryngoscopy is described in 1.5–3% of patients [3].

The ability to predict difficult laryngoscopy permits anesthesiologists to take precautions to decrease the risk [4]. Each test for airway examination may predict a separate aspect of the airway so, it is crucial to examine multiple parameters for proper airway assessment. In the present research, nine different parameters were taken into consideration in a computer-based approach to evaluate difficult laryngoscopy.

Data mining algorithms have many implementations. WEKA is one of such implementations and it is an open source software developed at the University of Waikato in New Zealand [5]. Prediction is one of those implementations that can be used based on training the program on input and output data [6]. This criterion has been used in the present study to allow the program predict difficult laryngoscopy according to input of cases and their parameters measurements.

Peer review under responsibility of Egyptian Society of Anesthesiologists.

* Corresponding author.
E-mail address: m.3abdelaziz@hotmail.com (M.A. Moustafa).

http://dx.doi.org/10.1016/j.egja.2017.02.002
1110-1849/© 2017 Publishing services by Elsevier B.V. on behalf of Egyptian Society of Anesthesiologists.

2. Aim of the work

Aim of the work was to evaluate the efficacy and accuracy of a multiparameter computer-based system to define difficult laryngoscopy.

3. Ethical approval

The study was approved by the local ethical committee of Alexandria Faculty of Medicine. A written informed consent was obtained from all participants.

4. Patients selection

Adult patients presenting for non emergency surgery at Alexandria main university hospital from February 2015 to February 2016 with unanticipated difficult laryngoscopy excluding major external facial or neck abnormalities, laryngeal abnormalities or tumors. Patients' age was ranging from 18 to 60 years. Unanticipated difficult laryngoscopy was identified by an experienced anesthesiologist (at least 5 years of experience) and defined as a view of the larynx grades 2B-4 according to Cook's modification [7] of Cormack-Lehane's classification [8].

5. Study design

Postoperatively, adult patients who had unanticipated difficult laryngoscopy over the study period were 50 and were closely matched demographically according to age and gender to 50 control patients chosen such that no difficulties were encountered during laryngoscopy or tracheal intubation.

Airway assessment was done postoperatively for the difficult laryngoscopy patients and their matched controls after full recovery from general anesthesia by a single blinded anesthesia consultant and focused on airway assessment which included the following parameters:

(1) Body mass index (BMI) = Weight in kilograms/Height in meters^2: Grade 0 = < 18.5
Grade 1 = 18.5–24.9
Grade 2 = 25–29.9
Grade 3 = 30 or greater
(2) Neck circumference: measured using a flexible tape at the level of the cricoid cartilage while patients were in the sitting position with the head and neck in the neutral posture. Grade 0 = <44 cm and Grade 1 = >44 cm
(3) Mandibular length. Grade 1 > 9 cm and Grade 2 < 9 cm
(4) Interincisor distance [9]: the patient is asked to open his/her mouth as wide as possible and the distance between upper and lower incisors was measured by a small ruler in the midline. Grade 1 = >4 cm and Grade 2 = <4 cm
(5) Thyromental distance [5]: measured by a small pocket ruler with the head fully extended and the mouth closed. Grade 1 = >6.5 cm and Grade 2 = <6.5 cm
(6) Sternomental distance [5]: measured by a ruler with the head fully extended and the mouth closed. Grade 1 = >13.5 cm and Grade 2 = <13.5 cm
(7) Modified Mallampati score [10]: It was performed with the patient in the sitting position and the neck held in neutral position and the tongue fully protruded without phonation.
Grade 0: Tip of the epiglottis is seen.
Grade 1: Tonsils, pillars and soft palate are clearly visible.
Grade 2: The uvula, pillars and upper pole are visible.
Grade 3: Only part of the soft palate is visible.
Grade 4: Only the hard palate is visible.
(8) Upper lip bite test (ULBT) [11]: The ULBT class was determined according to the following criteria:
Grade 1: Lower incisures can bite the upper lip above the vermilion line.
Grade 2: Lower incisors can bite the upper lip below the vermilion line.
Grade 3: Lower incisors cannot bite the upper lip
(9) Atlanto-occipital joint extension [12]: The patient was asked to hold head erect, facing directly to the front, then was asked to extend the head maximally and the examiner estimates the angle traversed by the occlusion surface of upper teeth. Any reduction in extension was expressed in grades: Grade I: $\geq 35°$, Grade II: 22–34°, Grade III: 12–21° and Grade IV: $\leq 11°$.

All data were entered into an information–based computer system where they were converted into numerical variables. Scores were classified into different categories that were supposed to be correlated with the difficulty of the laryngoscopic view.

6. Data analysis

The analysis aim is to classify data into two classes: Easy, which combines classes1 and 2A and Difficult, which combines classes 2B, 3A, 3B and 4. Each record was assessed using nine airway parameters that have been used for classification. For this purpose, data mining or machine learning algorithms were applied. Data mining is an interdisciplinary field which involves databases, statistics, and machine learning for finding patterns and consistency in sets of data [5].

The present study classifies data by using J48 Decision tree algorithm and illustrates it as follows. During the training phase of the program, it selects the parameter which correlates highly with the occurrence of difficult laryngoscopy. All patients who have difficult laryngoscopy are categorized in this area. While, other patients who are not difficult are put in another node of the tree and another parameter is started to be taken into consideration which is having a less correlation with difficult laryngoscopy and so on to complete the whole parameters tree.

Once the decision tree is complete, the order of attribute selection obtained by the tree is followed in each new case and hence prediction can be done according to the data base in the program which is designed in the form of a decision tree.

6.1. Proposed solution

All data have been processed and analyzed using the following two software programs:

- Microsoft visual studio 2008 (C#.net) [13]

  IF-THEN-ELSE rules

- WEKA (Waikato Environment for Knowledge Analysis) machine learning algorithms [14].

In the present study, a system was proposed that performs processing and analysis of data through the following steps:

Step1: Extracting Patient classification

Through airway parameters grades window (Fig. 1), and using IF-THEN-ELSE rules, the system in this step automatically determines whether laryngoscopy is expected to be difficult or easy. This is done after insertion of the parameters values for each patient, then

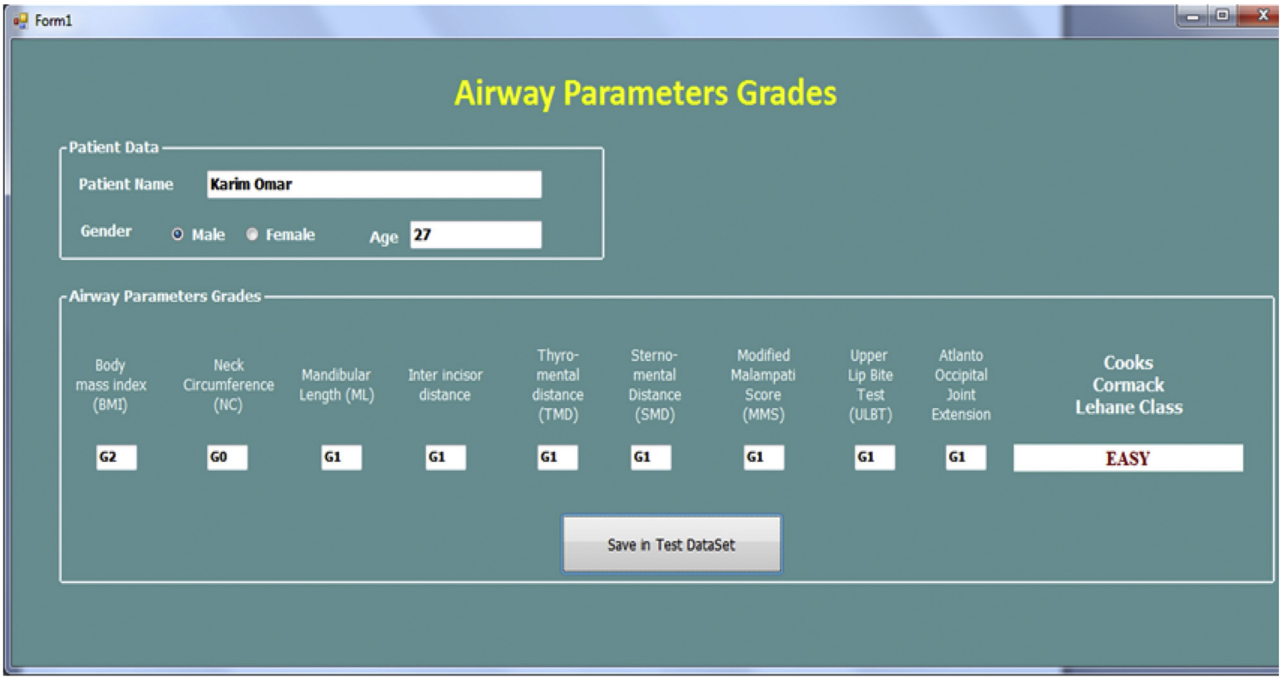

Fig. 1. Case data entry window. Case data are entered and the system expects whether it's difficult or not.

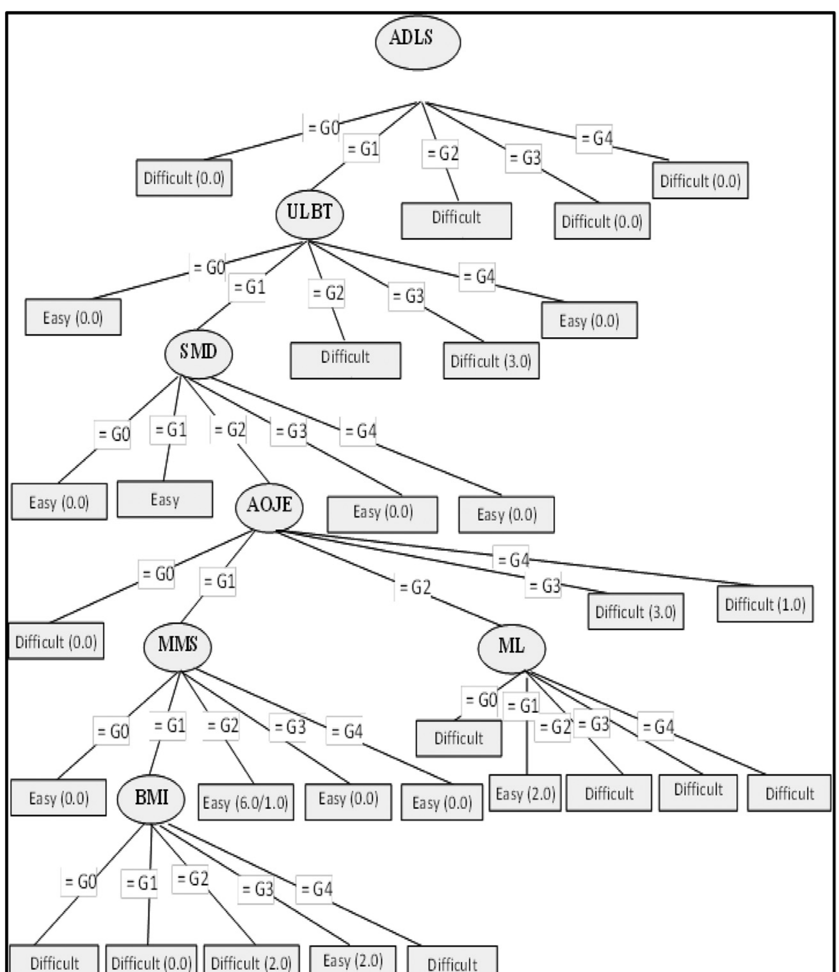

Fig. 2. Decision tree. Each parameter is a node of a tree and the node ends when expectation is easy. ADLS: Alex Difficult Laryngoscopy Software. ULBT: Upper Lip Bite Test. SMD: Sterno-Mental Distance. AOJE: Atlanto-Occipital Joint Extension. MMS: Modified Mallampati Score. ML: Mandibular Length. BMI: Body Mass Index. G: Grade.

the record is saved in the dataset file to be confirmed by J48 classifier:

- WEKA application is opened and the case is loaded into WEKA training set.
- J48 classifier algorithm is selected.
- supplied test set is chosen from test options menu then classification of supplied data is started

Step 2: Make Classification using Cross-validation Test

- Cross-validation with [10folds] is selected from test option and test is started to classify data.
- Decision tree can be visualized from the result list.

7. Results

The decision tree (Fig. 2) has been extracted from the cross validation test using J48 algorithm from which the confusion matrix (Fig. 3)a specific table layout that allows visualization of the performance of an algorithm- has been extrapolated.

Confusion matrix shows the following results (Fig. 4):

- **Sensitivity** = true positives (TP)/true positives + false negatives (FN)

Sensitivity or true positive rate (TPR) = 38/(38 + 12) = 76%

- **Specificity** = true negatives (TN)/true negatives + false positives (FP)

Specificity (SPC) or True Negative Rate (TNR) = 38/(38 + 12) = 76%

- **False positive rate (α)** = 1 – specificity = FP/(FP + TN) = 12/(12 + 38) = 24%
- **False negative rate (β)** = 1 – sensitivity = FN/(TP + FN) = 12/(38 + 12) = 24%
- **Precision or positive predictive value (PPV)** = TP/TP + FP = 38/(38 + 12) = 76%
- **Negative predictive value (NPV)** = TN/TN + FN = 38/38 + 12 = 76%
- **Fall-out or false positive rate (FPR)** = FP/FP + TN = 1 – SPC = 24%
- **False discovery rate (FDR)** = FP/TP + FP = 1 – PPV = 1 – 76 = 24%
- **Accuracy (ACC)** = (TP + TN)/(Positives P + Negatives N) = (38 + 38)/(50 + 50) = 76%
- **F1 score**, is the harmonic mean of precision and sensitivity

F1 = 2TP/(2TP + FP + FN) = 2 × 38/(2 × 38 + 12 + 12) = 76%

- **Matthews correlation coefficient (MCC)** =

$$\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

$$\frac{38 * 38 - 12 * 12}{\sqrt{(38 + 12)(38 + 12)(38 + 12)(38 + 12)}} = 0.52$$

Area under the ROC curve (Fig. 5) was 0.79.

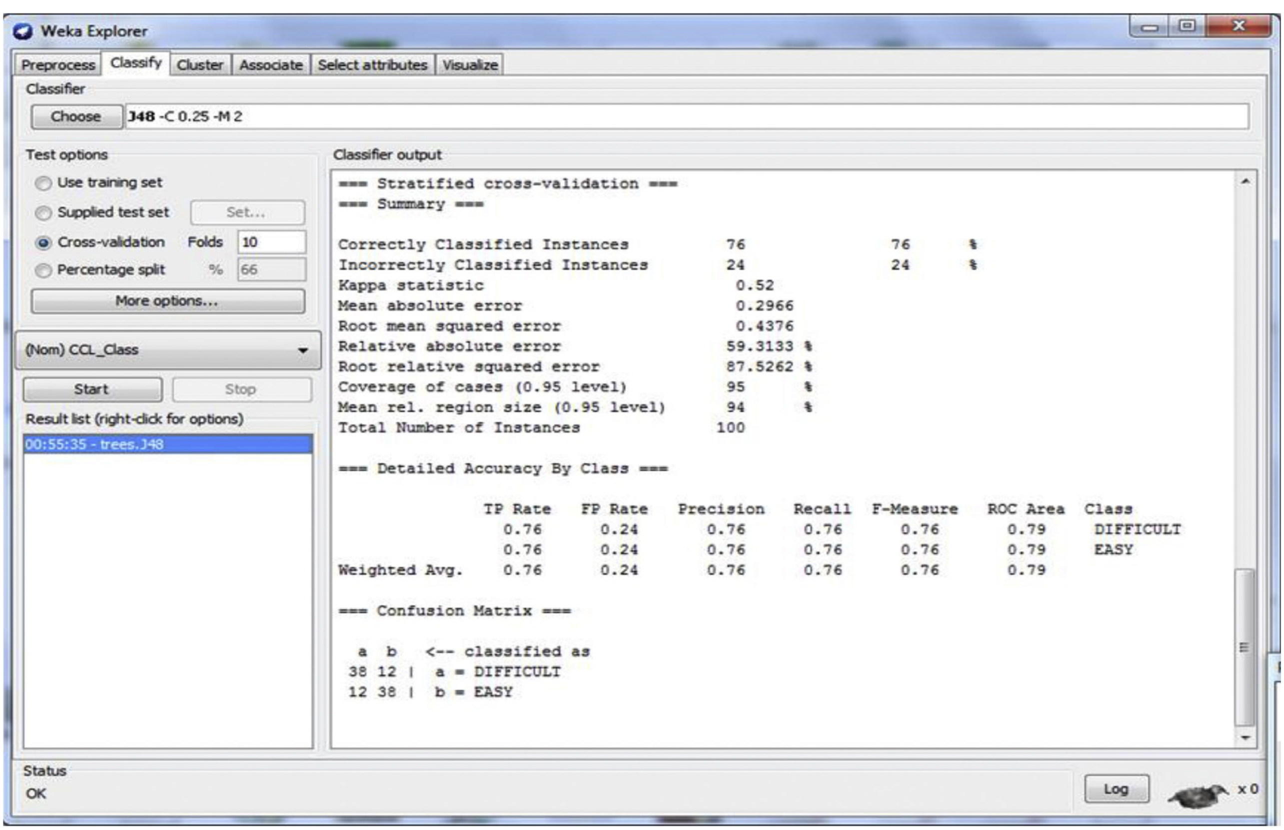

Fig. 3. Classifier accuracy and extracted confusion matrix. Using the cross validation test, performance of the program can be extracted and details of the operation can be displayed using the confusion matrix.

		Condition		
		Condition positive	Condition negative	
Test outcome	Test outcome positive	True positive (TP) = 38	False positive (FP) = 12	Precision = Σ True positive / Σ Test outcome positive
	Test outcome negative	False negative (FN) = 12	True negative (TN) = 38	Negative predictive value = Σ True negative / Σ Test outcome negative
		Sensitivity = Σ True positive = TP / (TP + FN) = 38 / (38 + 12) = 76%	Specificity = Σ True negative = TN / (FP + TN) = 38 / (12 + 38) = 76%	Accuracy
		Σ Condition positive	Σ Condition negative	

Fig. 4. Confusion matrix. Confusion matrix after its extraction from J48 classifier showing the results which can be extracted from it.

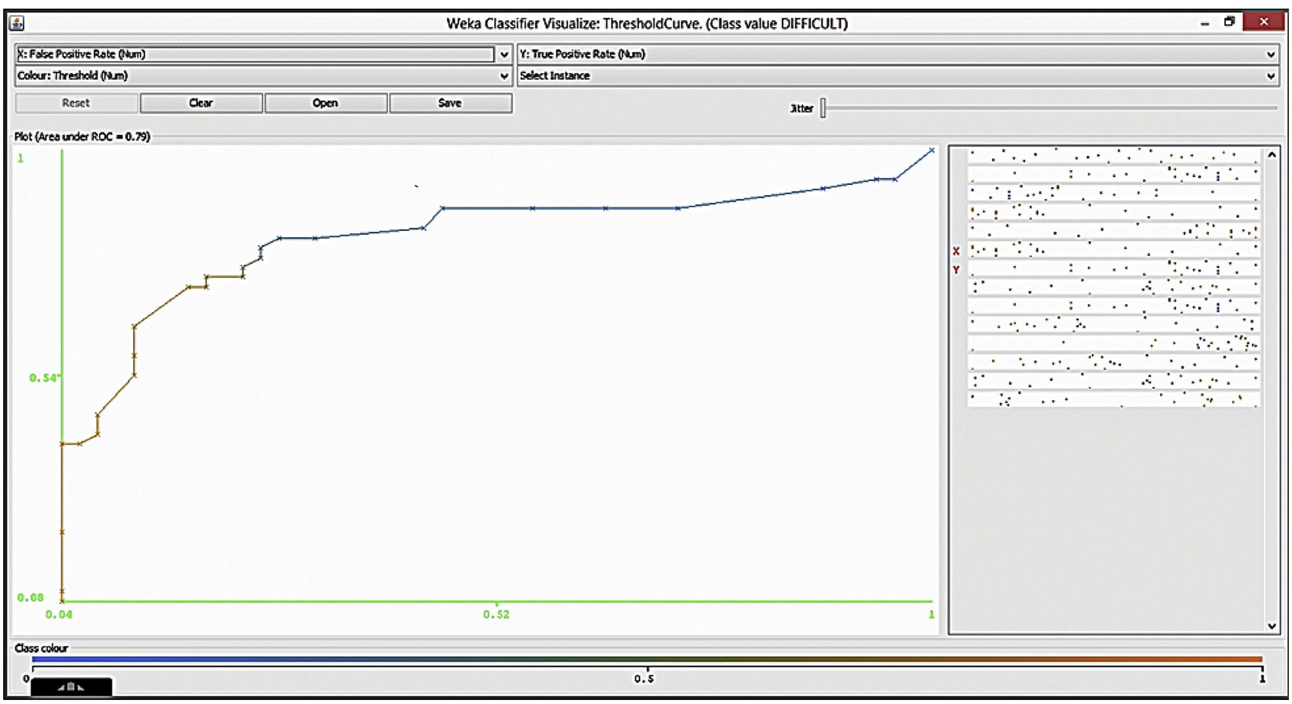

Fig. 5. ROC curve. The area under the curve is 0.79.

8. Discussion

Prediction of difficult laryngoscopy of a future case using the Microsoft Visual Studio and the WEKA engine based on a training set of the study cases has been termed **"Alex Difficult Laryngoscopy Software (ADLS)"**.

The use of electronic health records (EHRs) can serve as a component of a larger potential framework for real-time data-driven clinical decision support, or "adaptive decision support" [15]. The medical field is ripe for applications of artificial intelligence that can learn over time to predict optimal treatments and minimize side effects [16].

Langeron et al have previously assessed computer-assisted models that allow interactions between different variables for accurate prediction of difficult tracheal intubation. Their goal was to evaluate the precise mechanisms involved in the prediction of difficult tracheal intubation and they observed poor predictive value of the following variables; mouth opening, thyromental distance, Mallampati class and body mass index [17].

Several studies tried to combine two or more airway parameters to predict every patient scheduled for general anesthesia using endotracheal intubation. However, those studies used simple logistic models which in turn largely simplified the complex relationship between continuous variables and outcome. Combination of the Mallampati test and thyromental distance yielded low sensitivity. Patients with a 5% pretest probability of difficult intubation were shown to have a 34% risk of difficult laryngoscopy [1]. The simplified predictive intubation difficulty score (SPIDS) was investigated by L'Hermite et al. [18]. It was a manual score testing five airway parameters and found that the threshold for an optimal predictive level of the SPIDS was above 10. The sensitivity and specificity of the SPIDS above 10 for predicting difficult intubation were, respectively, 65% and 76% given by the cross-validation method. One of the methodological biases in this study was the calculation of the intubation difficulty score. LEMON airway assessment method incorporated a score with a maximum of 10 points calculated by assigning 1 point for each LEMON criterion including the external look, incisor distance, hyoid-mental distance, thyroid-to-mouth distance, Mallampati score and the presence of obstruction and neck mobility [19]. Arne and colleagues [20] produced a scoring system based on multifactorial analysis. It included the presence or absence of overt airway pathology such that the sensitivity and specificity levels of this system were above 90%. An M-TAC score developed by Ambesh et al had a sensitivity of 96% and specificity of 86% with a positive predictive value of 44% and a very low false negative value of 2%. Analysis of the receiver operating characteristic (ROC) curve for predicting difficult laryngoscopy revealed an area under the curve of 0.83 [21].

The present study tried to build up a computer based system that combines several airway parameters. This system using the J48 algorithm was trained to predict difficult laryngoscopy according to the training set in the form of two groups of cases that have been confirmed to be either easy or difficult for laryngoscopy. The sensitivity and specificity of the system as extracted from the confusion matrix was 76%. F1 score was 76% which represents a measure of a test's accuracy considering both the precision and the recall of the test. Matthews correlation coefficient (MCC) was 0.52 which is used in machine learning as a measure of the quality of binary (two-class) classifications. It takes into account true and false positives and negatives and ranges from -1 to $+1$ according to the power of correlation. Analysis of the (ROC) curve for predicting difficult laryngoscopy revealed an area under the curve of 0.79 (Fig. 5).

Some limitations in the present study should be considered. Patients belong only to adult population. Also, some of the selected airway parameters seemed to be poorly correlated with accurate prediction of difficult laryngoscopy.

It's concluded that the current system has achieved its goals with reasonable results. It is computer based not subjected to bias of manually giving each factor a score which may over or underestimate the true predictive value of each factor. Future refinement of the accuracy and correlation of the software is predicted as every new case can be entered into the training set thus improving the performance of the system. Development of the system is going on to add and remove airway parameters such that they correlate with more accurate prediction of difficult laryngoscopy. As the training set is enriched, the classification can be further subdivided into more grades according to the modified Cormack and Lehane's classification thus narrowing the prediction scope of the system.

Authors' contributions

Moustafa MA: Data collection, Writing 1st draft of the paper.
El-Metainy S: Idea of the paper, Data collection.
Mahar K: Data entry, Data analysis.
Mahmoud E: Results description, Data analysis.

Conflicts of interest

Authors declare that they no conflicts of interest.

Funds

Self fund.

References

[1] 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;103:429–37.
[2] Turkan S, Ates Y, Cuhruk H, Tekdemir I. Should we reevaluate the variables for predicting the difficult airway in anesthesiology? Anesth Analg 2002;94:1340–4.
[3] Langeron O, Masso E, Huraux C, Guggiari M, Bianchi A, Coriat P. Prediction of difficult mask ventilation. Anesthesiology 2000;92:1229–36.
[4] Practice guidelines for management of the difficult airway: An updated report by the American Society of Anesthesiologists task force on management of the difficult airway. Anesthesiology 2003;98:1269–77.
[5] Frank E, Hall MA and Ian H. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.
[6] Carvalho D. A hybrid decision tree/genetic algorithm method for data mining. Inf Sci 2004;163:13–35.
[7] Cook TM. A new practical classification of laryngeal view. Anaesthesia 2000;55:274–9.
[8] Larsen R. Anestezie. 2nd ed. Prague: Grada publishing; 2004.
[9] Merah NA, Wong DT, Foulkes-Crabbe DJ. Modified Mallampati test, thyromental distance and inter-incisor gap are the best predictors of difficult laryngoscopy in West Africans. Can J Anaesth 2005;52:291–6.
[10] Mashour GA, Sandberg WS. Craniocervical extension improves the specificity and predictive value of the Mallampati airway evaluation. Anesth Analg 2006;103:1256–9.
[11] Khan ZH, Kashfi A, Ebrahimkhani E. A comparison of the upper lip bite test (a simple new technique) with modified Mallampati classification in predicting difficulty in endotracheal intubation: a prospective blinded study. Anesth Analg 2003;96:595–9.
[12] Bellhouse CP, Dore C. Criteria for estimating likelihood of difficulty of endotracheal intubation with the Macintosh laryngoscope. Anaesth Intensive Care 1988;16:329.
[13] Snell M. Microsoft visual studio. 2nd ed. SAMS Publishing; 2012.
[14] Ian Witten H, Frank E, Hall Mark A. Data Mining: practical machine learning tools and techniques. 3rd ed. San Francisco: Morgan Kaufmann; 2011.
[15] Bennett CC, Doub TW, Bragg AD, Luellen J, Van Regenmorter C, Lockman J, et al. Data mining session-based patient reported outcomes (PROs) in a mental health setting: toward data-driven clinical decision support and personalized treatment. In: IEEE health informatics, imaging, and systems biology conference; 2012, May, 11; San Jose, CA, 2011. p. 229–36.
[16] Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, et al. The coming of age of artificial intelligence in medicine. Artif Intell Med 2009;46(1):5–17.
[17] Langeron O, Cuvillon, Ibanez-Esteve C, Lenfant F, Riou B, Le Manach Y. Prediction of difficult tracheal intubation: time for a paradigm change. Anesthesiology 2012;117(6):1223–33.
[18] L'Hermite J, Nouvellon E, Cuvillon P, Fabbro-Peray P, Langeron O, Ripart J. The simplified predictive intubation difficulty score: a new weighted score for difficult airway assessment. Eur J Anaesthesiol 2009;26(12):1003–9.
[19] Reed MJ, Dunn MJG, McKeown DW. Can an airway assessment score predict difficulty at intubation in the emergency department? Emerg Med J 2005;22:99–102.
[20] Arne J, Descoins P, Bresard D, Aries J, Fuseiardi J. A new clinical score to predict difficult intubation. Br J Anaesth 1993;70(suppl):A1.
[21] Ambesh SP, Singh N, Rao PB, Gupta D, Singh PK, Singh U. A combination of the modified Mallampati score, thyromental distance, anatomical abnormality, and cervical mobility (M-TAC) predicts difficult laryngoscopy better than Mallampati classification. Acta Anaesthesiol Taiwanica 2013;51(2):58–62.