의학석사 학위논문

Deep Learning Model for Predicting Difficult Laryngoscopy in Thyroid Surgery Patients Based on a Cervical Spine Lateral X-Ray Image

경추 측면 X선 영상을 기반으로 한 갑상선 수술 환자에서 어려운 후두경 예측 딥러닝 모델

2022년 4월

서울대학교 대학원
의학과 마취통증의학전공
조 혜 연

Deep Learning Model for Predicting Difficult Laryngoscopy in Thyroid Surgery Patients Based on a Cervical Spine Lateral X-Ray Image

지도 교수 정철우

이 논문을 의학석사 학위논문으로 제출함
2022년 4월

서울대학교 대학원
의학과 마취통증의학전공
조혜연

조혜연의 의학석사 학위논문을 인준함
2022년 8월

위 원 장 ____________________ (인)
부위원장 ____________________ (인)
위　　원 ____________________ (인)

초 록

예상하지 못한 어려운 후두경은 심각한 기도관련 합병증과 연관되어 있다. 본 연구는 후향적으로 수집된 갑상선 수술을 받은 총 14,135명 환자의 경추 측면 X선을 통해 어려운 후두경 (Cormack−Lehane 등급 3−4)를 예측하는 딥러닝 모델을 개발 및 검증하였다. 개발 모델의 성능은 기존의 6개의 딥러닝 모델과 비교하였다. 개발 모델에서 어려운 후두경 예측의 민감도는 95.6%, 특이도 91.2%를 나타냈다. Area Under ROC curve의 경우 개발 모델에서 0.972(0.955~0.988), 기존 모델의 경우 각각 VGG−Net: 0.842, ResNet: 0.841, Xception: 0.863, ResNext: 0.825, DenseNet: 0.889, SENet: 0.875를 나타냈다. 어려운 후두경과 관련된 해부학적 특징을 설명하기 위해 클래스 활성화 맵(Class Activation Map)을 사용하였다. 클래스 활성화 맵에서 설골, 인두 및 경추 주변이 강조되었다. 본 연구를 통해 개발된 딥러닝 모델은 경추 측면 X선 영상을 이용한 어려운 후두경 예측에 높은 성능을 보였다.

주요어 : 기관내 삽관, 기도 평가, 딥러닝, 어려운 후두경, 인공 지능
학 번 : 2020−24809

Contents

Tables

Figures

Supplementary Materials

1. Introduction

Patients undergoing general anesthesia often require endotracheal intubation via laryngoscopy. Unanticipated difficult laryngoscopy is associated with serious airway-related complications, such as brain damage, cardiopulmonary arrest, or death [1]. The incidence of difficult laryngoscopy is known to be 6% [2], therefore, predicting difficult laryngoscopy is important for patient safety. Clinical airway evaluation and image-based indicators have been used to predict difficult laryngoscopy. Although clinical predictors, such as the modified Mallampati classification, thyromental distance, inter-incisor gap, and the upper lip bite test, can be used for airway evaluation before laryngoscopy, they require patient cooperation and have limitations of low sensitivity and large inter-assessor variability [3,4].

Image-based indicators, such as tongue size, the distance from the hyoid bone to the mandibular body, the angle between the hyoid bone, thyroid cartilage, and arytenoid cartilage, have been proposed as predictors of difficult laryngoscopy [5–7]. Previous studies have shown that image-based indicators have advantages of high predictive power and lower inter-assessor variability [6,7]. However, because image-based indicators usually rely on features extracted manually from images, the time and effort for feature extraction have limited their clinical application.

Recent advances in machine learning techniques have been widely adapted to many tasks in the field of medicine. In particular, deep learning techniques, such as convolutional neural networks (CNN), have shown excellent performance in the task of automatic interpretation of radiological images [8]. Moreover, the features extracted by the CNN models can be visualized using explainable artificial intelligence methods, such as the class activation map. Therefore, using these techniques, we can not only develop an accurate prediction model without labor-intensive feature extraction, but also understand more about the anatomical structures associated with the outcome.

In this study, we aimed to develop and validate a CNN-based deep learning model that can predict difficult laryngoscopy based on a cervical spine lateral X-ray image. Our hypothesis was that an accurate deep learning model for predicting difficult laryngoscopy based on a cervical spine lateral X-ray image can be developed, and insights into appropriate indicators for airway evaluation can be obtained by analyzing the features identified by the model.

2. Materials and Methods

This single-center retrospective study was approved by the Institutional Review Board of Seoul National University Hospital (No.1706-071-859). The requirement of obtaining written informed consent was waived owing to the retrospective nature of the study design.

2.1 Inclusion and Exclusion Criteria

Patients who received thyroid surgery under general anesthesia at Seoul National University Hospital between November 2004 and November 2020 were eligible for this study. Patients with the following features were excluded: 1) age <18 years; 2) no cervical spine lateral X-ray image obtained within six months before surgery; 3) no record of Cormack–Lehane grade on the anesthesia record; 4) use of other intubating devices such as lighted stylet from the first intubation attempt; 5) intubation into a tracheostomy site (Figure 1). In these patients, only data from the first surgery during the study period were used for the training and test datasets.

2.2 Anesthesia Management

In our institution, tracheal intubation for thyroid surgery was typically performed as follows. Without premedication, general anesthesia was induced with propofol (1–2 mg/kg), remifentanil (effect-site concentration of 3 ng/mL), and rocuronium (0.6–1.0 mg/kg). After adequate muscle relaxation, tracheal intubation was performed by an anesthesiologist. A reinforced tube with an internal diameter of 7.5 mm was used in male patients, with a Macintosh blade size 4, and a reinforced tube with an internal diameter of 7.0 mm was used in female patients, with a blade size 3.

2.3 Data Collection and Preprocessing

Patient demographic data, such as age, sex, height, and weight, were collected from the electronic medical records. The American Society of Anesthesiologists (ASA) physical status classification, use of other airway devices during intubation, and the Cormack–Lehane grade were extracted from the anesthesia records. The easy laryngoscopy was defined as a combination of the Cormack–Lehane grades 1-2 and the difficult laryngoscopy was defined as grades 3-4. The cervical spine lateral X-ray images, taken in standing and neutral positions, were extracted from the Picture Archiving and Communication System workstation (INFINITT PACS Version 5.0.0, INFINITT Healthcare, Seoul, Korea) in the DICOM format of 16-bit images. The images were resized from 4095 × 2047 to 256 × 256 pixels. To normalize the ranges of pixel values between images, all pixel values were subtracted by the mean value of the image and divided by the standard deviation. Therefore, the cervical spine X-ray pixel values were standardized with a mean of 0 and standard deviations of 1.

2.4 Model Building

A CNN-based deep learning model with a convolutional layer, pooling layer, self-attention layer, and final fully connected layer was developed to predict difficult laryngoscopy (Figure 2). This model used the input of preprocessed cervical spine lateral X-ray image and outputs the difficulty of a laryngoscopy. Our model was composed of three main components: the convolutional path, attention path, and classifier. The convolutional path (the upper path in Figure 2) was motivated by the down-sampling architecture of the fully convolutional network [9], and input images were abstracted and reduced in dimension with five convolution and pooling layers.

The attention path (the lower path in Figure 2) was designed by adding the class activation map attention module to VGG-Net [10] without pooling operations. The input images of the lower path were abstracted without dimensionality reduction, and normalized feature maps from each layer were extracted and used for class activation map (CAM) attention. The results from the two paths were merged by global averaging pooling and addition and fed into the classifier, which contained two fully connected layers and a SoftMax layer. In the end, our model has a structure in which a self-attention pathway is added to the CNN pathway to create the CAM that visualizes discriminative image regions classified by laryngoscope difficulty from cervical spine lateral X-ray images.

We initialized the weights of the parameters of the CNN models using a Gaussian initializer. All of the CNN models were trained with the Adam optimizer with a loss function of balanced-binary cross-entropy loss and a learning rate of 0.001. The learning rate was halved every 30 epochs. Each model was trained using 300 epochs and a batch size of 30. More detailed explanations of our model are provided in Appendix A.

For each training epoch, the CAM [11] was derived between the last layer and the input layer of the classifier and used for CAM attention. The CAM attention module, our original proposal for X-ray image interpretation, takes the input of normalized feature maps from the lower path and CAM images from the classifier. It computes the cosine similarity between the normalized feature maps of each layer and CAM images and updates the parameters of the lower path.

The training was performed using our custom-written program, prepared in Python 3.7, using TensorFlow 1.14 on a GPU server with 64-core (Intel Xeon Gold 6226R CPU @ 2.90 GHz) and 8 Nvidia GTX Titan XP.

2.5 Model Validation

For time-dependent hold-out validation, the data of the last three years were classified as the test dataset, and the remaining data were classified as the training dataset. Random down-sampling for the major class was performed to solve issues with an imbalance in the data. Therefore, only one-third (4,044 cases) of the images were used for training in the easy laryngoscopy group of the training dataset.

The performance of our model was evaluated with a test dataset and was compared with that of six other well-known convolutional neural network architectures: VGG-Net [10], ResNet [12], Xception [13], ResNext [14], DenseNet [15], and SENet [16].

The CAM was generated in each case of the test dataset to assess whether our CNN model made reasonable predictions. The CAM, as mentioned above, can visualize an attribution map in which the lesions in the X-ray image that are relevant for prediction are highlighted.

2.6 Sensitivity Analysis

A sensitivity analysis was performed to evaluate whether the value predicted from the cervical spine lateral X-ray image taken in the first operation had a high predictive performance in the next operation in the patient. Among the test datasets, patients who underwent tracheal intubation for the second operation during the study period were included in the sensitivity analysis.

2.7 Statistical Analysis

Baseline characteristics of patients were compared between the difficult and easy laryngoscopy groups for the datasets. The t-test or Mann–Whitney test was used for group comparisons of continuous variables, and the chi-square test or Fisher’s exact test was used for comparisons of categorical variables, as appropriate.

We defined the classification threshold as 0.5. If the probability was ≥0.5, it was predicted as the difficult laryngoscope group, and if the probability was <0.5, it was predicted as the easy laryngoscope group. The receiver operating characteristic analysis was performed and the area under the receiver operating characteristic curve (AUC) was used to evaluate the diagnostic value of each convolutional neural network architecture. The optimal cut-off point was determined by maximizing the sum of the sensitivity and specificity. Sensitivity, specificity, positive-predictive value (PPV), negative-predictive value (NPV), F1-score, and balanced accuracy were assessed to compare the performance of the CNN models. Balanced accuracy (mean of sensitivity and specificity) and F1 score, which is the harmonic mean of sensitivity (also called recall) and PPV (also called precision), were used for the performance comparison for imbalanced data.

Data are expressed as mean (standard deviation) for normally distributed continuous variables, median (interquartile range) for non-normally distributed variables, and number (percent) or the number of each group for categorical variables. Statistical analysis was performed using SPSS 25 (IBM Corp., Armonk, NY, USA), R software (version 3.6.1; R Development Core Team, Vienna, Austria), and Python 3.7.0 (Python Software Foundation, Wilmington, DE, USA). In all analyses, statistical significance was set at $P < 0.05$.

3. Results

3.1 Dataset Construction

Data from 14,135 patients undergoing thyroid surgery were included in the study. Among these patients, 1,687 (11.9%) patients, the data of the last three years, were assigned to the test dataset. The data of the remaining participants were assigned to the training dataset. The incidence of difficult laryngoscopies, defined as Cormack–Lehane grades 3–4, was 1.7% in the training dataset and 2.7% in the test dataset (P=0.005, Figure 1). The demographic data of the patients in this study are shown in Table 1 and 2. The difficult laryngoscopy group showed higher body mass index (23.7 vs 25.1 kg/m2, P < 0.019) in the test dataset (Table 2).

3.2 Performance of the Models

The performances of the trained models are listed in Table 2. The AUC and 95% confidence interval for the VGG-Net [10], ResNet [12], Xception [13], ResNext [14], DenseNet [15], SENet [16], and our model were 0.842 (0.786–0.898), 0.841 (0.789–0.893), 0.863 (0.816–0.911), 0.825 (0.762–0.889), 0.889 (0.848–0.931), 0.889 (0.848–0.931), 0.875 (0.848–0.927), and 0.972 (0.955–0.988), respectively (Table 3, Figure 3). Our model showed a larger AUC, compared with other models (P<0.001). In addition, the sensitivity, specificity, PPV, NPV, F1-score, and balanced accuracy are shown in Table 2. Our model showed a sensitivity of 95.6%, specificity of 91.2%, PPV of 22.9%, NPV of 99.9%, F1-score of 36.9, and balanced accuracy of 93.4%.

The CAM of difficult laryngoscopies are shown in Figure 4. The CAM demonstrated clear differences around the hyoid bone, pharynx, and cervical spine (Figure 4).

3.3 Sensitivity Analysis

A sensitivity analysis was performed on 83 patients who underwent two surgeries during the study period. The median interval between the first and the second laryngoscopic view was 150 days. In a total of 83 patients, the incidence of difficult laryngoscopy (Cormack–Lehane grades 3-4) in the first operation and following operation were 2.4% (n = 2) and 2.4% (n = 2), respectively. Among the 81 patients with easy laryngoscopy in the first operation, 80 patients (98.8%) had easy laryngoscopy in the following operation. Among the 2 patients with difficult laryngoscopy in the first operation, one patient had difficult laryngoscopy in the next operation. Both patients who had a change in the difficulty of laryngoscopy in the first operation and the following operation were predicted as difficult laryngoscopy by our model (Figure 5).

The performance of our model was retained in the sensitivity analysis. The AUC and 95% confidence interval for the VGG-Net, ResNet, Xception, ResNext, DenseNet, SENet, and our model were 0.432 (0.077–0.787), 0.620 (0.203–1.000), 0.920 (0.836–1.000), 0.889 (0.781–0.997), 0.676 (0.247–1.000), 0.664 (0.237–1.000), and 0.969 (0.927–1.000), respectively (see table, Supplementary Digital Content 1, listing results of sensitivity analyses).

4. Discussion

In this study, we developed a novel CNN-based deep learning model for predicting difficult laryngoscopy using a cervical spine lateral X-ray image. To our best knowledge, this is the first study to predict difficult laryngoscopy based on a cervical spine lateral X-ray image with visualization of the lesion associated with difficult laryngoscopy. The results showed that the deep learning-based model has excellent performance and reliability for predicting difficult laryngoscopy.

In previous meta-analyses and prospective studies, clinical predictors for difficult laryngoscopy have a sensitivity, specificity, and AUC of 35%, 91%, and 0.75 for the modified Mallampati classification, 68%, 77%, and 0.72 for the inter-incisor distance, 77%, 89%, and 0.83 for the upper lip bite test, 22%, and 74%, 82%, and 0.78 for thyromental distance, respectively [2,17,18]. These clinical predictors are generally less sensitive because they only evaluate a portion of factors associated with difficult laryngoscopy. Although combining multiple clinical predictors can increase the predictive power, they require more time for clinical examinations and need the patient's cooperation [19].

The radiographic images, such as X-ray, computed tomography, magnetic resonance, and ultrasound images, can be used to evaluate the patient's airway with higher sensitivity. In a previous study using a cervical spine lateral X-ray image, 100% sensitivity and specificity were reported using the angles between the upper airway structures such as the hyoid, epiglottis, and arytenoid [5]. In another study using the location of the vocal cord on magnetic resonance images, the sensitivity and specificity were 60% and 96%, respectively [6]. In other studies using tongue thickness and the hyo-mental distances measured by ultrasound, the AUCs were 0.9320 and 0.758 [21], respectively. However, although radiographic predictors have higher predictive power and are more objective than clinical predictors, they also require clinicians to extract features manually from the images [4–6,22].

Previous studies using machine learning algorithms to predict difficult laryngoscopy have also reported high predictive power [23–25], but they all required indicators that should be evaluated by clinicians. On the other hand, our method does not require any parameters evaluated by a human. Therefore, our model can be implemented on the Picture Archiving and Communication System for automated evaluation of the airways. Nevertheless, the performance of our model was higher than that of previous reports. The reason may be that our model can automatically extract many features from an image and use them together to predict the outcome.

The PPV of our model was 22.9%, which can be explained by the low incidence (2.5% in the test set) of difficult laryngoscopy. However, the sensitivity of our model is 95.6%, which is significantly higher than that of other CNN models (VGG 80.0%, ResNet 80.0%, Xception 80.0%, ResNext 77.8%, DenseNet 82.2%, SENet 80.0%), as well as the aforementioned clinical predictors and radiologic indicators. In addition, in the sensitivity analysis, two patients had changes in the difficulty of laryngoscopy (easy ↔ difficult) in the next operation, and both patients were classified as difficult laryngoscopy in our model. Since unanticipated difficult laryngoscopy is critical, our model will have an advantage in being used as a screening tool for difficult laryngoscopy.

In the CAM analysis, important areas for difficult laryngoscopies were mainly around the hyoid bone, pharynx, and cervical spine (Figure 4). These areas have already been reported as main anatomical landmarks associated with difficult laryngoscopy in other studies. In previous studies, a short distance from the skin to the hyoid bone and a long distance from the mandible to the hyoid bone were associated with difficult laryngoscopy [4,22,26]. If the hyoid bone is positioned caudally, a greater portion of the tongue is present in the hypopharynx, which interferes with tongue displacement, causing difficulty in direct laryngoscopy [26]. Pharyngeal space was also related to a large tongue in previous studies [20,22]. The atlanto-occipital gap and spinous processes associated with cervical spine mobility [27] were also reported as radiographic indicators of the difficult laryngoscopy. Compared with previous studies using the distance, size, and angle of the specific anatomical structures [4–6,22], our model has the advantage of being able to automatically evaluate the entire anatomical structures visible on the image.

Although our model showed a high predictive performance, radiation exposure is an inevitable drawback in our approach. However, in the results of the sensitivity analysis, the performance of our model was maintained with the X-ray image taken in the past. This result can be interpreted as the anatomical structures associated with difficult laryngoscopy do not change in many cases. A previous study also showed that a “history of a difficult laryngoscopy” was a highly specific predictor of difficult laryngoscopy [28]. Therefore, we expect that patients who have undergone a cervical spine lateral X-ray at least once may not need to undergo another X-ray to use our model unless there are significant changes in their cervical structures.

There were several limitations to our study. First, this study included only patients who underwent thyroid surgery with routine cervical spine X-rays taken before surgery. Therefore, our results might not be applicable to other patient groups or imaging modalities. Second, because our dataset only includes patients who underwent a direct laryngoscopy, it might not be applicable to other intubation methods, such as video laryngoscopy. Third, due to several missing values in the clinical airway evaluations of the hospital's anesthesia records, we could not compare the performance of clinical predictors to our predictive model. Fourth, our study used the Cormack–Lehane grade recorded on the anesthesia records. Although external laryngeal manipulation should not be applied according to the definition of the Cormack–Lehane grade [29,30], some practitioners rate the Cormack–Lehane grade as the best laryngeal view obtained with external laryngeal pressure. This can potentially result in an underestimation of the incidence of difficult laryngoscopy in our results. Therefore, a prospective study is needed, including pre-laryngoscopy evaluation of clinical predictors and controlled Cormack-Lehane grade evaluation in patients undergoing various other surgeries. Fifth, in this study, the CAM classified three areas associated with difficult laryngoscopy. However, there were few cases classified as difficult, and we could not determine which regions were more associated with difficult laryngoscopy. Further studies with a large sample size are needed to develop a CNN-based deep learning model that automatically and quantitatively scores major areas related to difficult laryngoscopy in cervical spine X-ray images. Lastly, this study was conducted with time-dependent hold-out validation, with consideration of the importance to verify the predictive power of the deep learning model from recent data. However, in this validation, the date variability of cervical X-ray and Cormack-Lehane grade are not considered, which is a limitation of this study. In a further study, the performance of our model can be confirmed through multiple random selections, including date variability.

5. Conclusions

In conclusion, our deep learning model for predicting difficult laryngoscopy based on a cervical spine lateral X-ray image showed excellent predictive performance. Our study also identified the hyoid bone, pharynx, and cervical spine as important areas in the class activation map for predicting difficult laryngoscopies. If future prospective validation studies confirm our results in other patient groups, this approach can be helpful in improving patient safety and preventing airway-related complications through objective and accurate airway evaluation.

Tables

Table 1. Patient characteristics and laryngoscopic parameters in dataset. In total, 14,135 eligible patients at the input data level, patients in the last three years were classified as the test dataset, and the remaining patients were classified as the training dataset.

	Training set (2004–2017, *n* = 12,448)	Test set (2018–2020, *n* = 1,687)	*P*–value
Sex (male)	2,419 (19.4)	361 (21.4)	0.061
Age	48.0 (39.0–57.0)	50.0 (39.7–59.0)	<0.001
Weight	59.5 (53.8–67.7)	61.1 (54.7–70.4)	<0.001
Height	159.7 (155.1–165.0)	160.5 (156.0–166.1)	<0.001
BMI	23.5 (21.5–25.8)	23.8(21.6–26.5)	<0.001
ASA			
1	6,934 (55.7)	503 (29.8)	<0.001
2	5,232 (42.0)	1,094 (64.8)	<0.001
≥3	282 (2.3)	90 (5.3)	<0.001
Cormack-Lehane grade			0.036
1	11,429 (91.8)	1,534 (90.9)	
2	811 (6.5)	108 (6.4)	
3	201 (1.6)	43 (2.5)	
4	7 (0.1)	2 (0.1)	

Data are expressed as median (interquartile range) and number (percent). BMI, Body Mass Index; ASA, American Society of Anesthesiologists.

Table 2. Patient characteristics and laryngoscopic parameters in test set. The easy laryngoscopy was defined as a combination of the Cormack–Lehane grades 1–2 and the difficult laryngoscopy was defined as a combination of grades 3–4.

	Easy laryngoscopy (*n* = 1,642)	Difficult laryngoscopy (*n* = 45)	*P*–value
Sex (male)	349 (21.3)	12 (26.7)	0.491
Age	50.0 (39.1–59.0)	54.0 (46.0–61.1)	0.076
Weight	61.0 (54.6–70.2)	65.5 (55.8–74.9)	0.051
Height	160.4(156.0–166.1)	161.5 (157.1–165.5)	0.530
BMI	23.7(21.6–26.4)	25.1 (22.6–28.3)	0.019
ASA			
1	498 (30.3)	5 (11.1)	0.009
2	1,056 (64.3)	38 (84.4)	0.008
≥3	88 (5.4)	2 (4.4)	0.999
Cormack-Lehane grade			<0.001
1	15.4 (93.4)	0 (0.0)	
2	108 (6.6)	0 (0.0)	
3	0 (0.0)	43 (95.6)	
4	0 (0.0)	2 (4.4)	

Data are expressed as median (interquartile range) and number (percent). BMI, Body Mass Index; ASA, American Society of Anesthesiologists.

Table 3. The performance of our model and other convolutional neural network architectures. The performance of each model is evaluated for the prediction of difficult laryngoscopy through the cervical spine lateral X-ray.

Model	Sensitivity Recall, TPR	Specificity TNR	Precision, PPV	NPV	F1 score	Balanced accuracy	AUC	95% CI
VGG	80.0	75.0	8.1	99.3	14.7	77.5	0.842	0.786 – 0.898
ResNet	80.0	76.0	8.4	99.3	15.2	78.0	0.841	0.789 – 0.893
Xception	80.0	77.1	8.7	99.3	15.8	78.6	0.863	0.816 – 0.911
ResNext	77.8	78.6	9.1	99.2	16.2	78.2	0.825	0.762 – 0.889
DenseNet	82.2	83.7	12.2	99.4	21.2	83.0	0.889	0.848 – 0.931
SENet	80.0	83.4	11.7	99.4	20.4	81.7	0.875	0.848 – 0.927
Ours	**95.6**	**91.2**	**22.9**	**99.9**	**36.9**	**93.4**	**0.972**	**0.955 – 0.988**

Data are expressed as number (percent). TPR, true positive rate; TNR, true negative rate; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve; CI, Confidence intervals

Figure Legends

Figure 1. Flow diagram of this study. C-spine, cervical spine.

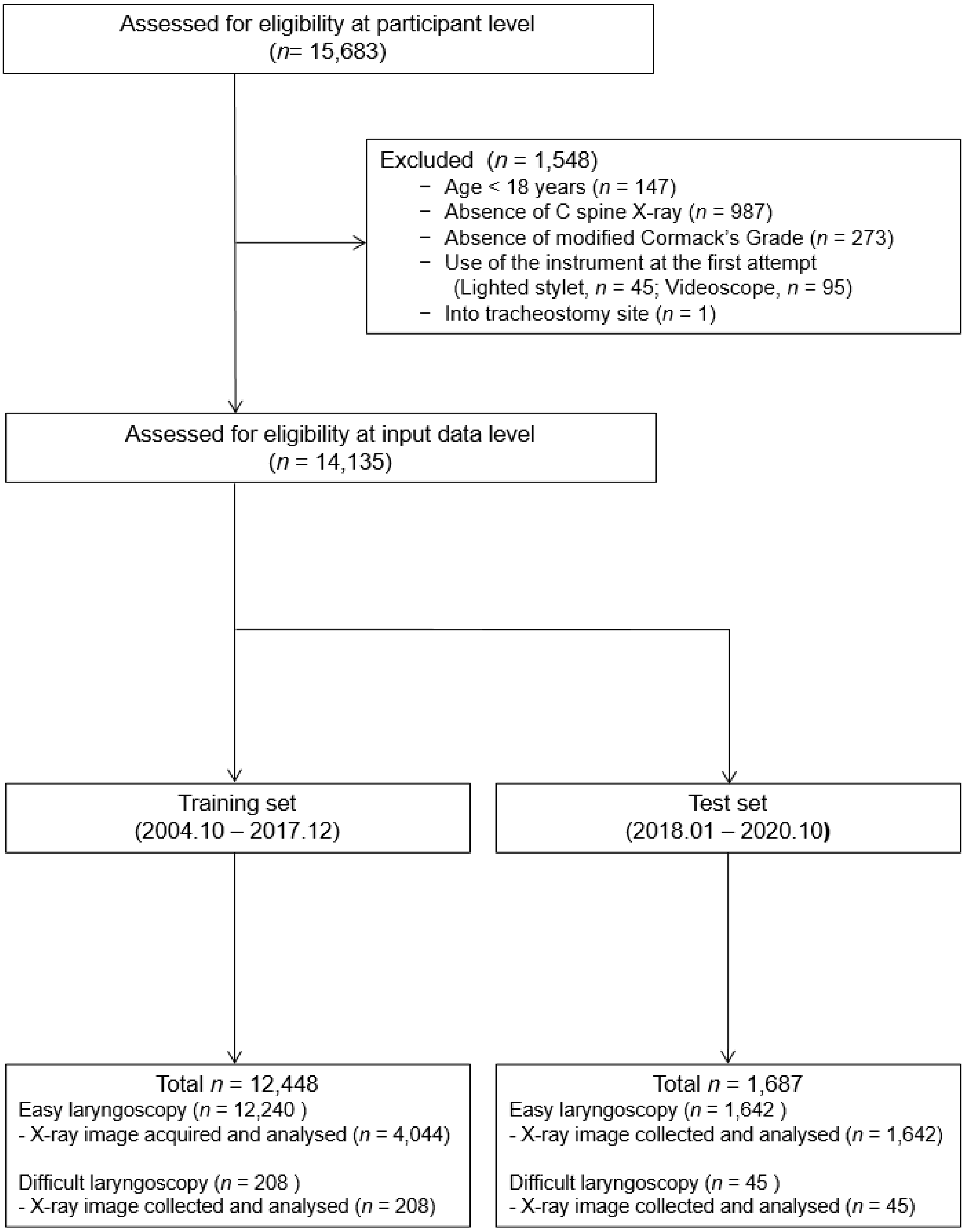

Assessed for eligibility at participant level
(*n*= 15,683)
Excluded (*n* = 1,548)
− Age < 18 years (*n* = 147)
− Absence of C spine X-ray (*n* = 987)
− Absence of modified Cormack's Grade (*n* = 273)
− Use of the instrument at the first attempt
(Lighted stylet, *n* = 45; Videoscope, *n* = 95)
− Into tracheostomy site (*n* = 1)
Assessed for eligibility at input data level
(*n* = 14,135)
Training set
(2004.10 – 2017.12)
Test set
(2018.01 – 2020.10)
Total *n* = 12,448
Easy laryngoscopy (*n* = 12,240)
- X-ray image acquired and analysed (*n* = 4,044)
Difficult laryngoscopy (*n* = 208)
- X-ray image collected and analysed (*n* = 208)
Total *n* = 1,687
Easy laryngoscopy (*n* = 1,642)
- X-ray image collected and analysed (*n* = 1,642)
Difficult laryngoscopy (*n* = 45)
- X-ray image collected and analysed (*n* = 45)

Figure 2. Convolutional neural network-based deep learning model with a convolutional layer, pooling layer, self-attention layer, and final fully connected layer to predict difficult laryngoscopy. The preprocessed cervical spine lateral X-ray image is used as the input and the model outputs the difficulty of laryngoscopy along with a class activation map that visualizes discriminative image regions.

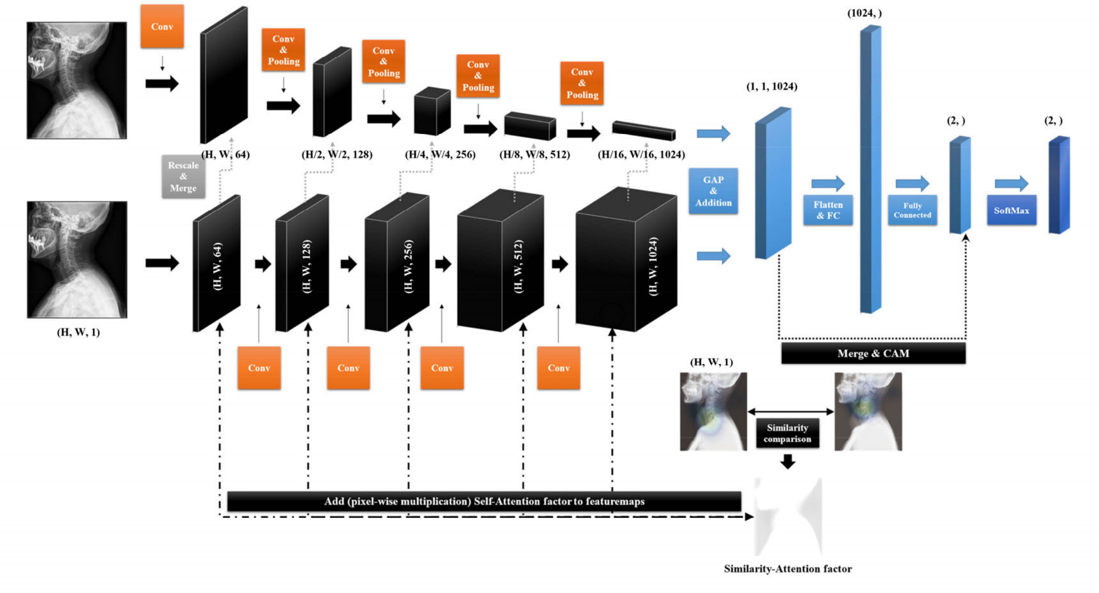


Figure 3. Receiver operating characteristic curves of each model.

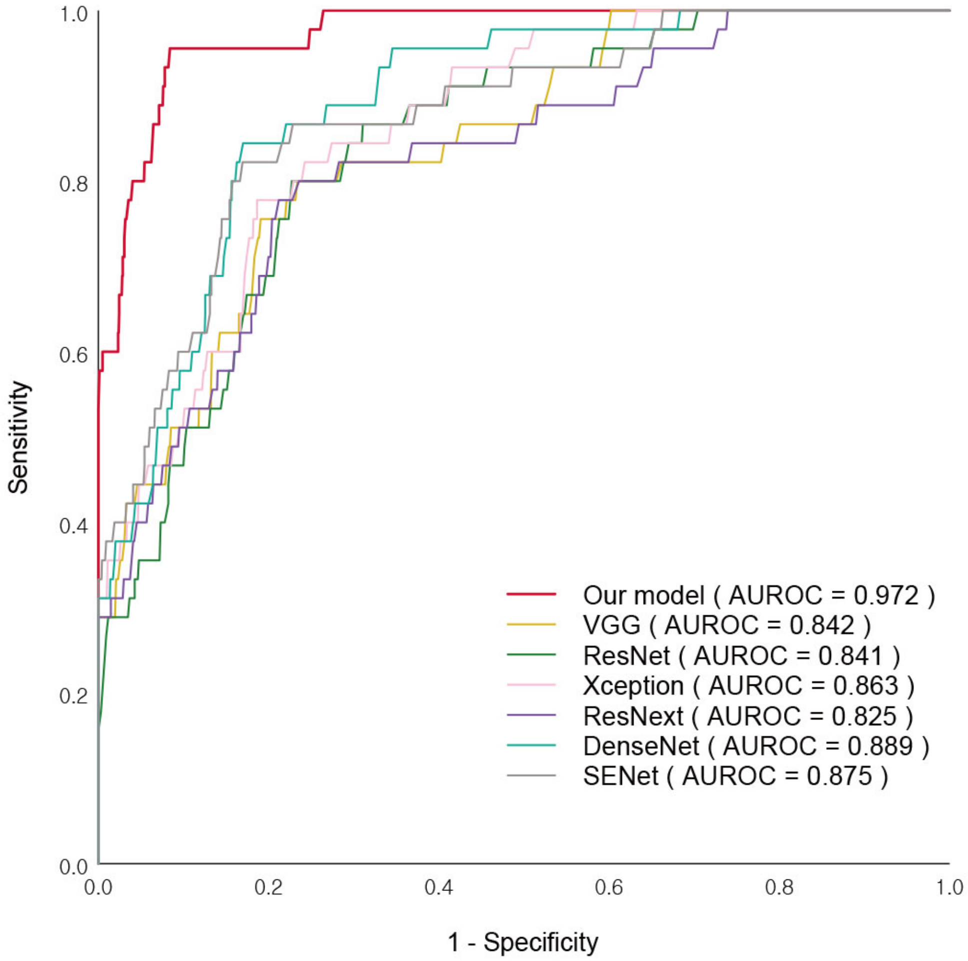


Figure 4. The examples of the class activation map in each group by three parts. Darker colors indicate the highlight class-specific image regions. (**a**) Anatomical landmarks of a cervical X ray image (**b**) The class activation map of the easy laryngoscopy group. (**c**) The hyoid bone highlighted in difficult laryngoscopy. (**d**) The pharynx highlighted in difficult laryngoscopy. (**e**) The cervical spine highlighted in difficult laryngoscopy.

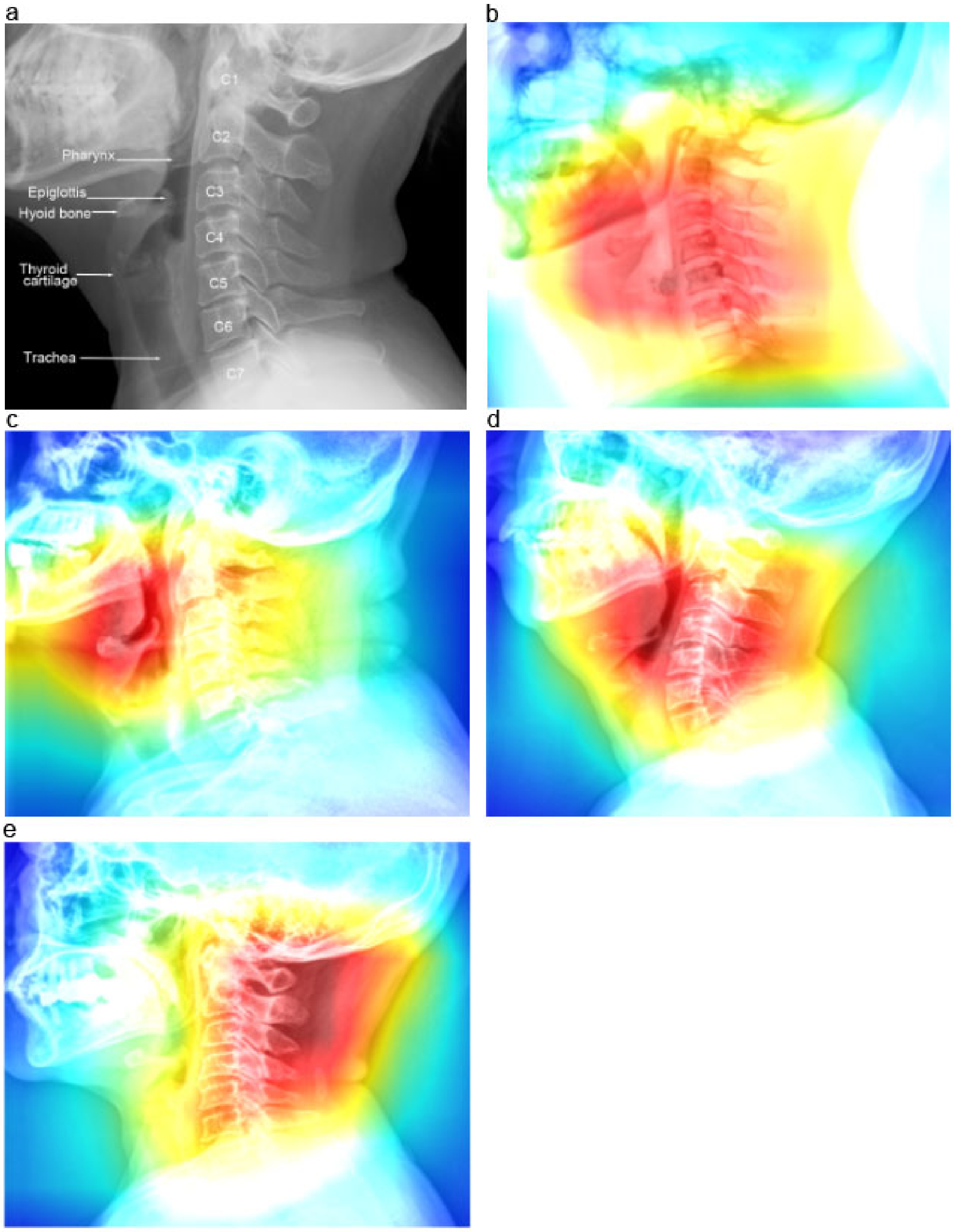
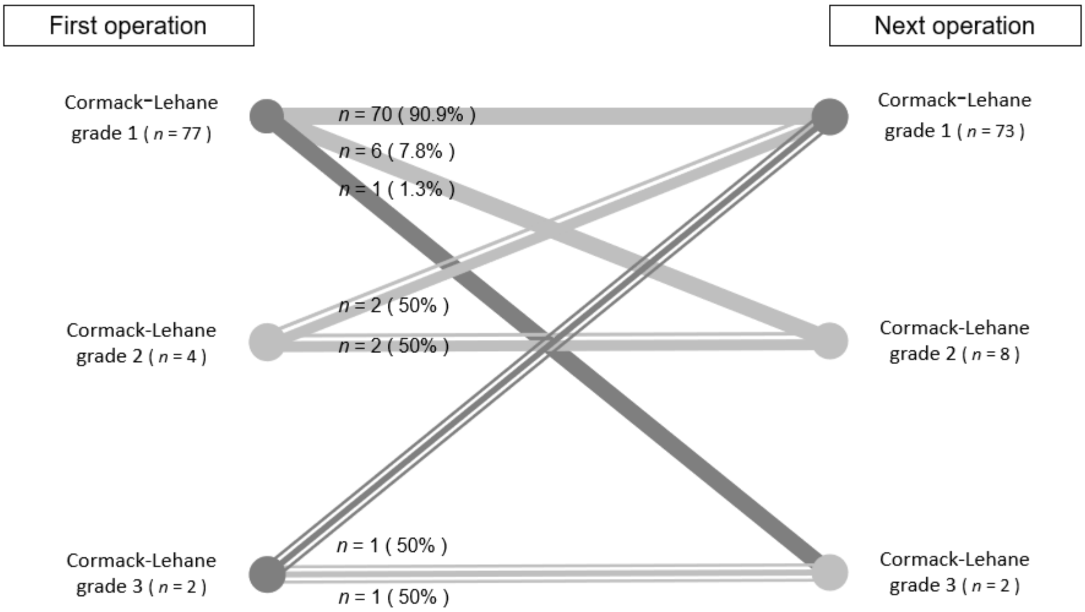


Figure 5. The changing Cormack–Lehane grade of laryngoscopy between the first operation and the next operation in the sensitivity analysis. The difficulty of laryngoscopy (easy ↔ difficult) was changed in 2 out of 83 patients in the following operation.


First operation
Next operation
Cormack-Lehane grade 1 (n = 77)
Cormack-Lehane grade 2 (n = 4)
Cormack-Lehane grade 3 (n = 2)
n = 70 (90.9%)
n = 6 (7.8%)
n = 1 (1.3%)
n = 2 (50%)
n = 2 (50%)
n = 1 (50%)
n = 1 (50%)
Cormack-Lehane grade 1 (n = 73)
Cormack-Lehane grade 2 (n = 8)
Cormack-Lehane grade 3 (n = 2)

Initial prediction of difficulty in the two patients who had a change in difficulty of laryngoscopy in each model

	Our model	VGG	ResNet	Xception	ResNext	DenseNet	SENet
Cormack-Lehane grade from 1 to 3	Difficult	Easy	Difficult	Difficult	Difficult	Easy	Easy
Cormack-Lehane grade from 3 to 1	Difficult	Difficult	Difficult	Difficult	Difficult	Easy	Difficult

Supplementary Digital Content 1. The performance of models in the sensitivity analysis in 83 patients from the test set undergoing two surgeries during the study period.

Model	Sensitivity, Recall, TPR	Specificity, TNR	AUC	95% confidence intervals
VGG	0.0	86.4	0.432	0.077–0.787
ResNet	50.0	74.1	0.620	0.203–1.000
Xception	100.0	84.0	0.920	0.836–1.000
ResNext	100.0	77.8	0.889	0.781–0.997
DenseNet	50.0	85.2	0.676	0.247–1.000
SENet	50.0	82.7	0.664	0.237–1.000
Ours	**100.0**	**93.8**	**0.969**	**0.927–1.000**

The median time interval between cervical spine X-ray imaging and intubation was 150 days. TPR, true positive rate; TNR, true negative rate; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve.

Appendix A-1

The class activation map has been reported previously.[14] In essence, it attempts to identify which part of the image is more important than other parts by multiplying and adding the weights from the fully connected layer with the previous feature-map during the image classification task. If there are objects in an image that are classified as a specific class, the weights of the feature map containing the object related to the target class of the network will be larger than the weights of other feature maps. When these values are weights summed, the object related to the target class is highlighted.

Let $f_k(x,y)$ indicate the point of (x,y) in the kth feature map and let the higher values of the points indicate highlighted activation. In the deep learning network, the global average pooling follows the fully connected layer, and the kth feature map is denoted as $F_k = \sum_{x,y} f_k(x,y)$. At this point, the class score for a specific class (c) is calculated as $S_c = \sum_k w_k^c F_k$, where w_k^c is the activation weight of the global average pooling value of the kth feature map for *class c*, which is a trainable parameter in the fully connected layer. That is, $S_c = \sum_k w_k^c \sum_{x,y} f_k(x,y) = \sum_{x,y} \sum_k w_k^c f_k(x,y)$ can be obtained. Here, let $M_c = \sum_k w_k^c f_k(x,y)$. Then, $S_c - \sum_{x,y} M_c$. Note that M_c directly represents the score of the (x,y) coordinates for class c, and M_c is the class activation map for class c. Because the calculated class activation map is smaller in size than the input image, the class activation map is resized to the input size by upsampling. Then, by using a heatmap of the class activation map, it is possible to identify which part affects classification.

Appendix A-2

Our model was designed based on a fully convolutional network and VGG-Net, with the addition of a self-attention module. The baseline model extracts feature maps from the convolution, pooling, and fully connected layers by a single path. In contrast, our network contains two paths that extract different features individually. The upper path extracts general features, such as the fully convolutional network or VGG-Net model, while the other extracts highlighted feature maps using the attention module. The two paths are combined using a pixel-wise addition operation, which penetrates the fully connected operations. The SoftMax operation, which is the last layer of our network, generates the probability for each class.

Each layer of our network is a three-dimensional tensor with the shape (h, w, c), where h, w, and c are the height, width, and channel, respectively. In the upper path, the layers are reconstructed to the feature maps by every convolutional operation, and the size of each layer is halved by pooling operations. Here, the convolution operation includes scaled exponential linear unit (SeLU) activation to add nonlinearity and batch normalization.

The lower path of our network is constructed based on the VGG-Net, but without a pooling operation. Instead, each feature map is mutated by the self-attention module by adding a similarity-attention factor. A detailed description of the self-attention module is provided below:

Let $F_i(I^{(n)})$ be the *i*th feature map whose shapes are (H, W, C) in the lower path, using the *n*th input image ($I^{(n)}$), and let $F_i(I^{(n)})_c$ be each feature of $F_i(I^{(n)})$ in channel c. Then, the mean features of all channels of $F_i(I^{(n)})$ is calculated as follows:

$$\frac{1}{c}\sum_{c} F_i(I^{(n)})$$

Here, we define the matrix $S(F_i(I^{(n)})) = norm\left(\frac{1}{c}\sum(F_i(I^{(n)}))\right)$, where $norm$ is a min–max normalization. The normalized feature of $F_i(I^{(n)})$, which is $S(F_i(I^{(n)}))$, implicates the integrated insight (attention) on the localization of an input image for each layer $F_i(I^{(n)})$. Here, the shape of $I(X)$ is $(H, W, 1)$. Moreover, the class activation map, denoted as $cam(I^{(n)})$, whose shape is $(H, W, 1)$ is generated in the fully connected operation and SoftMax operation. Note that each element of $S(F_i(I^{(n)}))$and $cam(I^{(n)})$ is normalized to the range of [0, 1].

In the self-similarity module, the similarity between $S(F_i(I^{(n)}))$ and $cam(I^{(n)})$ is calculated for the individual *i*th feature in the lower layer. Here, the similarity indicates the cosine similarity (pixel-wise outer product). That is, the following equation is utilized to calculate the similarity-attention factor ($sf_i(I^{(n)})$):

$$sf_i(I^{(n)})_{h,w} = \frac{S\left(F_i(I^{(n)})\right)_{h,w} \times cam(I^{(n)})_{h,w}}{\left(S(F_i(I^{(n)}))_{h,w}\right)^2 + \left(cam(I^{(n)})_{h,w}\right)^2 - \left(F_i(I^{(n)})\right)_{h,w} \times cam(I^{(n)})_{h,w}}$$

The similarity-attention factor is designed to localize the attention of the individual layer to the same location as that of class activation map. Therefore, the similarity factor is multiplied by the feature maps, and the feature map for each layer is updated as follows:

$$F_l\left(I^{(n)}\right) \leftarrow F_l\left(I^{(n)}\right) * norm\left(sf_l\left(I^{(n)}\right)\right)$$

Because the similarity factor has an element that is normalized to the range of [0, 1], the regions that are similar to those of class activation map are emphasized, and the regions with low similarity are de-emphasized.

Therefore, in the lower path, the forced and localized features are extracted. In addition, to utilize the highlighted features in the upper path, the highlighted features are resized and added to compensate for the features that are extracted in the lower path.

Our model was designed by adding a self-attention module to existing basic models for classification tasks. The module extracts feature maps from the convolution operation without a pooling operation to maintain the dimensions of the input image.

The modules of localization of attention are embedded in the developed convolutional neural network-based model to integrate the laryngoscopy regions that most affect the classification results predicted by a deep learning network. In general, only fully optimized deep learning networks generate accurate predictions with integrated insight (attention) on the impact region in X-ray images when classification tasks are conducted. In contrast, our attention layers, and embedded modules force localization of laryngoscopy regions from the early training steps. Early localization of integrated attention to the laryngoscopy regions improves the directionality of the gradients of trainable variables of a deep learning network in the training step. The attention module-embedded ChestXNet-based developed networks were utilized to classify the difficulty of a laryngoscopy.

Appendix A-3.

To verify the feasibility of using a convolutional neural network to determine the level of difficulty of a laryngoscopy, the default values of the hyper-parameters were utilized. For instance, the Adam optimizer is used to train deep learning networks for the experiments. β_1 and β_2 were 0.9 and 0.99, respectively, and ϵ was 1e-7. In addition, the learning rate for the optimizer was 0.001, and it was halved after every 30 epochs. The sizes of all convolution filters were 3×3 and the channels for the deep learning networks were 64, 128, 256, 512, and 1024 for each deeper step. Furthermore, the stride of the convolution filters was two. Mirror padding was applied to all the deep learning models. The initialization of deep learning models was performed using a Gaussian distribution with a mean value of 0 and a standard deviation of 1. The batch normalization technique was applied after every convolution operation and activation function, which involved SeLU activation. Cross-validation was applied for the optimization of deep learning networks, and the ratio of the number of training images, validation images, and test images was set at 6:1:1 because the number of images for optimizing deep learning networks was insufficient.

References

1. Cook TM, MacDougall-Davis SR: Complications and failure of airway management. Br J Anaesth 2012; 109 Supplement 1:i68–85
2. Richard M. Cooper; Preparation for and Management of “Failed” Laryngoscopy and/or Intubation. Anesthesiology 2019; 130:833–849.
3. Lundstrøm LH, Vester-Andersen M, Møller AM, Charuluxananan S, L’Hermite J, Wetterslev J, Danish Anaesthesia Database: Poor prognostic value of the modified Mallampati score: a meta-analysis involving 177 088 patients. Br J Anaesth 2011; 107:659–667
4. De Cassai A, Boscolo A, Rose K, Carron M, Navalesi P: Predictive parameters of difficult intubation in thyroid surgery: a meta-analysis. Minerva Anestesiol 2020; 86:317–326
5. Han YZ, Tian Y, Zhang H, Zhao YQ, Xu M, Guo XY: Radiologic indicators for prediction of difficult laryngoscopy in patients with cervical spondylosis. Acta Anaesthesiol Scand 2018; 62:474–482
6. Kamalipour H, Bagheri M, Kamali K, Taleie A, Yarmohammadi H: Lateral neck radiography for prediction of difficult orotracheal intubation. Eur J Anaesthesiol 2005; 22:689–693
7. Münster T, Hoffmann M, Schlaffer S, Ihmsen H, Schmitt H, Tzabazis A: Anatomical location of the vocal cords in relation to cervical vertebrae: A new predictor of difficult laryngoscopy? Eur J Anaesthesiol 2016; 33:257–262
8. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, Ashrafian H, Darzi A: Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. Npj Digit Med 2021; 4:65
9. Long J, Shelhamer E, Darrell T: Fully convolutional networks for semantic segmentation, IEEE, 2015
10. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. ArXiv Pre-Print Serv, 2015. Available at: April 10. https://arxiv.org/abs/1409.1556
11. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A: Learning deep features for discriminative localization, IEEE, 2016
12. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition, IEEE, 2016. http://doi.org/10.1109/cvpr.2016.90.
13. Xception CF: Deep learning with depthwise separable convolutions, IEEE, 2017. http://doi.org/10.1109/cvpr.2017.195.
14. Xie S, Girshick R, Dollar P, Tu Z, He K: Aggregated residual transformations for deep neural networks, IEEE, 2017. http://doi.org/10.1109/cvpr.2017.634.

15. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ: Densely connected convolutional networks, In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp 4700–4708
16. Hu J, Shen L, Albanie S, Sun G, Wu E: Squeeze-and-Excitation Networks. ArXiv Pre-Print Serv, 2019. Available at: May 16. https://arxiv.org/abs/1709.01507
17. Khan ZH, Mohammadi M, Rasouli MR, Farrokhnia F, Khan RH: The diagnostic value of the upper lip bite test combined with Sternomental distance, Thyromental distance, and interincisor distance for prediction of easy laryngoscopy and intubation: A prospective study. Anesth Analg 2009; 109:822–824
18. 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; 103:429–437
19. Iohom G, Ronayne M, Cunningham AJ: Prediction of difficult tracheal intubation. Eur J Anaesthesiol 2003; 20:31–36
20. Xu L, Dai S, Sun L, Shen J, Lv C, Chen X: Evaluation of 2 ultrasonic indicators as predictors of difficult laryngoscopy in pregnant women: A prospective, double blinded study. Medicine (Baltimore) 2020; 99:e18305
21. Andruszkiewicz P, Wojtczak J, Sobczyk D, Stach O, Kowalik I: Effectiveness and validity of sonographic upper airway evaluation to predict difficult laryngoscopy. J Ultrasound Med 2016; 35:2243–2252
22. Lee HC, Kim MK, Kim YH, Park HP: Radiographic predictors of difficult laryngoscopy in acromegaly patients. J Neurosurg Anesthesiol 2019; 31:50–56
23. Aguilar K, Alférez GH, Aguilar C: Detection of difficult airway using deep learning. Mach Vis Appl 2020; 31
24. Kim JH, Kim H, Jang JS, Hwang SM, Lim SY, Lee JJ, Kwon YS: Development and validation of a difficult laryngoscopy prediction model using machine learning of neck circumference and thyromental height. BMC Anesthesiol 2021; 21:125
25. Connor CW, Segal S: Accurate classification of difficult intubation by computerized facial analysis. Anesth Analg 2011; 112:84–93
26. Chou HC, Wu TL: Mandibulohyoid distance in difficult laryngoscopy. Br J Anaesth 1993; 71:335–339
27. Lopez AJ, Scheer JK, Leibl KE, Smith ZA, Dlouhy BJ, Dahdaleh NS: Anatomy and biomechanics of the craniovertebral junction. Neurosurg Focus 2015; 38:E2
28. Arné J, Descoins P, Fusciardi J, Ingrand P, Ferrier B, Boudigues D, Ariès J: Preoperative assessment for difficult intubation in general and ENT surgery: predictive value of a clinical multivariate risk index. Br J Anaesth 1998; 80:140–146

29. Cormack RS, Lehane J: Difficult tracheal intubation in obstetrics. Anaesthesia 1984; 39:1105–1111

30. Krage R, Van Rijn C, Van Groeningen D, Loer SA, Schwarte LA, Schober P: Cormack–Lehane classification revisited. Br J Anaesth 2010; 105:220–227

Abstract

Deep Learning Model for Predicting Difficult Laryngoscopy in Thyroid Surgery Patients Based on a Cervical Spine Lateral X-Ray Image

Hye-Yeon Cho
Department of Anesthesiology and Pain Medicine
The Graduate School
Seoul National University

An unanticipated difficult laryngoscopy is associated with serious airway-related complications. We here developed and validated a deep learning-based model that predicts a difficult laryngoscopy (Cormack–Lehane grade 3–4) from a cervical spine lateral X-ray using data from 14,135 patients undergoing thyroid surgery. The performance of our model was compared with six representative deep learning architectures. A class activation map was created to elucidate the anatomical features associated with difficult laryngoscopy. Our model showed 95.6% sensitivity and 91.2% specificity for predicting difficult laryngoscopy. The area under the receiver operating characteristic curve of our model was 0.972 (0.955–0.988), which was higher than that of other models (VGG-Net: 0.842, ResNet: 0.841, Xception: 0.863, ResNext: 0.825, DenseNet: 0.889, and SENet: 0.875, all $P < 0.001$). The class activation map demonstrated clear differences around the hyoid bone, pharynx, and cervical spine. The model showed excellent performance for predicting difficult laryngoscopy using a cervical spine lateral X-ray image.

Keywords: airway evaluation, artificial intelligence, deep learning, difficult laryngoscopy, intubation, intratracheal
Student Number: 2020-24809

감사의 글

석사과정을 마치고 학위 논문을 제출하게 되었습니다.

먼저 바쁘신 가운데에도 더 나은 논문이 될 수 있도록 학위 심사에 많은 지도를 해주신 정철우 선생님, 김진태 선생님, 이규언 선생님께 감사 드립니다.

본 논문이 완성되기까지 2년의 시간 동안 따뜻한 가르침과 많은 도움을 주신 이형철 선생님께 머리 숙여 감사의 말씀을 전합니다. 선생님의 지도를 통해 연구의 즐거움을 알게 되었습니다. 또한 연구를 함께 해주신 융합의학과 공현중 선생님, 모델 개발에 힘써주신 이경수 연구원님 감사드립니다. 데이터 수집에 큰 도움을 주신 심다연 연구원님께도 감사의 마음 전합니다.

오늘을 발판 삼아 더 나은 연구자가 될 수 있도록 정진하겠습니다.

마지막으로 면학에 전념할 수 있게 응원해주고 도와주신 부모님과 남편에게 마음을 전합니다.