

RESEARCH

Open Access



Identification of difficult laryngoscopy using an optimized hybrid architecture

XiaoXiao Liu^{1,2†} , Colin Flanagan^{2†}, Gang Li^{3†}, Yiming Lei^{4*}, Liaoyuan Zeng⁵, Jingchao Fang⁶, Xiangyang Guo⁷, Sean McGrath^{2*} and Yongzheng Han^{7*}

Abstract

Background Identification of difficult laryngoscopy is a frequent demand in cervical spondylosis clinical surgery. This work aims to develop a hybrid architecture for identifying difficult laryngoscopy based on new indexes.

Methods Initially, two new indexes for identifying difficult laryngoscopy are proposed, and their efficacy for predicting difficult laryngoscopy is compared to that of two conventional indexes. Second, a hybrid adaptive architecture with convolutional layers, spatial extraction, and a vision transformer is proposed for predicting difficult laryngoscopy. The proposed adaptive hybrid architecture is then optimized by determining the optimal location for extracting spatial information.

Results The test accuracy of four indexes using simple model is 0.8320. The test accuracy of optimized hybrid architecture using four indexes is 0.8482.

Conclusion The newly proposed two indexes, the angle between the lower margins of the second and sixth cervical spines and the vertical direction, are validated to be effective for recognizing difficult laryngoscopy. In addition, the optimized hybrid architecture employing four indexes demonstrates improved efficacy in detecting difficult laryngoscopy.

Trial registration Ethics permission for this research was obtained from the Medical Scientific Research Ethics Committee of Peking University Third Hospital (IRB00006761-2015021) on 30 March 2015. A well-informed agreement has been received from all participants. Patients were enrolled in this research at the Chinese Clinical Trial Registry (<http://www.chictr.org.cn>, identifier: ChiCTR-ROC-16008598) on 6 June 2016.

Keywords Difficult laryngoscopy, Hybrid architecture, Atlantooccipital gap, Cervical spondylosis, Radiological variables

[†]XiaoXiao Liu, Colin Flanagan and Gang Li contributed equally to this work; they share the first authorship.

*Correspondence:

Yiming Lei

leiym@pku.edu.cn

Sean McGrath

sean.mcgrath@ul.ie

Yongzheng Han

hanyongzheng@bjmu.edu.cn

¹ College of Mathematics and Information Science, Hebei University,

Baoding, China

² Electronic and Computer Engineering, University of Limerick, Limerick,

Ireland

³ Department of General Surgery (GL), Peking University Third Hospital, Beijing, China

⁴ Ministry of Education Engineering Research Centre on Mobile Digital Hospital Systems, School of Electronics, Peking University, Beijing, China

⁵ School of Communications, University of Electronic Science and Technology of China, Chengdu, China

⁶ Department of Radiology (JCF), Peking University Third Hospital, Beijing, China

⁷ Department of Anaesthesiology, Peking University Third Hospital, Beijing, China



Introduction

Difficulty in airway management is one of the leading causes of anaesthesia-related cardiac arrest, death, and brain damage in patients undergoing surgery [1]. Unanticipated difficult airway has always been an issue even for experienced anaesthesiologists [2]. Adequate evaluation of airway structures prior to surgery is essential for safe and effective tracheal intubation, which is a prerequisite for safe anaesthesia, minimising distress, and reducing the need for additional sedation for reintubation [3]. Clinically, many factors are associated with difficult laryngoscopy, including head-neck trauma [4], airway abnormalities [5], gastroesophageal reflux disease [6], difficulty in opening mouth [7], impaired cervical mobility [8]. To achieve optimal laryngoscopic views, the oral, pharyngeal and laryngeal axes need to guarantee closest match, creating feasibility for anaesthesiologists to expose the glottis in patients with cervical spondylosis [7]. However, gold standards and guidelines have not been laid out for difficult laryngoscopy detection. The Mallampati classification predicts intubation ease based on oral cavity visibility, however, despite suggested modifications, adding neck mobility and mandibular space, it has limitations due to subjectivity and an inadequate assessment of airway problems.

Recent technological advancements in artificial intelligence (AI) algorithms, computer hardware, and large medical imaging datasets have enabled computer scientists and healthcare researchers to collaborate closely to improve airway management [9] and laryngoscopy prediction [10–12]. As a potent subfield of AI, deep learning has the potential to analyse large medical databases in parallel [13] via its multiple computational nodes and to identify potential intubation hazards via its hidden layers [14]. As one of the popular deep learning applications, vision transformers employing self-attention structures are able to extract global information, recognise patterns, and capture long-distance relationships from images, enabling efficient imaging detection, such as difficult laryngoscopy identification.

This study aims to identify difficult laryngoscopy using a novel attention-based AI model on a large preoperative X-ray dataset, as well as to investigate the relationship between image-measured indicators and difficult laryngoscopy. Two new indicators will be introduced to reflect the range of motion of the upper and lower cervical spine in relation to difficult laryngoscopy, and their efficacy will be evaluated by comparing them to two published indicators. Combining convolutional neural networks, spatial extraction, and vision transformer structures, this study will introduce a novel AI architecture for the identification of difficult laryngoscopy. In order to enhance the performance of the hybrid AI model, an attention-based

spatial extraction structure will be implemented and its optimal placement will be determined. For the first time, the performance of difficult laryngoscopy prediction using a single index and using multiple indexes will be validated uniformly. The ranking of the four indicators will also be provided.

Material and methods

Datasets and samples

This prospective cohort study recruited patients undergoing elective cervical spine surgery under general anaesthesia during the period June 2016 to December 2021. The following criteria were included: (1) Age range of 20 to 70 years, (2) Psychiatric health, and (3) Intact radiological and medical records. Rule-out conditions were as below. (1) Airway neoplasm or foreign objects (tumours of the larynx, pharynx, tongue, floor of the mouth, or cysts involving the mandible or medial neck), (2) Serious cervical vertebral trauma, (3) Cervical instability, (4) Unstable physical condition (ASA IV or V), and (5) Anticipated difficulty with facemask ventilations (previous surgical intubation difficulties, surgery, neck radiotherapy). Ethics permission for this research was obtained from the Medical Scientific Research Ethics Committee of Peking University Third Hospital (IRB00006761-2015021) on 30 March 2015. A well-informed agreement has been received from all participants. Patients were enrolled in this research at the Chinese Clinical Trial Registry (<http://www.chictr.org.cn>; identifier: ChiCTR-ROC-16008598) on 6 June 2016.

Routine preoperative monitoring of non-invasive blood pressure, heart rate, pulse oximetry, and electrocardiography was performed. Sufentani (0.3 $\mu\text{g}/\text{kg}$) and propofol (2 mg/kg) were administered to induce anaesthesia. In unconscious patients, neuromuscular blockade was induced by rocuronium (0.6 mg/kg). The difficulty of laryngoscopy was ascertained by the single advanced anaesthesiologist using the Cormack-Lehane scales with the Macintosh laryngoscope for all participants in the olfactory position (Table 1) [15]. The anaesthesiologist was not engaged in perioperative radiographic evaluations. Those with grade III or IV views were allocated to the difficult laryngoscopy category, and those who had grade I or II views were allocated to the simple

Table 1 The Cormack Lehane (C-L) scale

Classification	Description
Class I	Vocal cords were completely visible
Class II	The arytenoids were visible
Class III	Only the epiglottis was visible
Class IV	The epiglottis was not visible

laryngoscopy category. Patients who were unsuccessful with the Macintosh laryngoscope were addressed in accordance with the Difficult Airway Society 2015 guidelines [16]. No patients were involved in the radiological data measurements nor were they involved in developing plans for the design and accomplishment of the present study. None of the patients was asked to advise on the interpretation. The radiology staff was blinded to the examination. The results will be disseminated to investigators and patients through this publication.

A previous study showed that the incidence of difficult laryngoscopy was 24% [17]. In our preliminary study, the incidence of difficult laryngoscopy was 18.6%. It was estimated that the sample sizes of 610 would achieve to detect a difference in indicators between the difficult and easy laryngoscopy groups ($\alpha = 0.05$ and $\beta = 0.1$), and in consideration of 10% dropout rate, 671 patients were enrolled in this study.

Data pre-processing

Patients’ clinical and radiological information was obtained from their medical records and the image archiving and communication system (PACS). Pre-processing is a crucial step in medical image classification, especially in the processing of large datasets. AI techniques for medical images typically rely on supervised learning, utilizing datasets containing data points (e.g., images) and labels (e.g., object classes) [13]. Pre-processing including data segmentation, labeling, data

enhancement [18], and data balancing was mainly performed on histograms and labelled images in this study (Fig. 1). Feature extraction is vital for image classification. Medical image segmentation is one of the most promising methods in medical image analysis, which identifies pixels of organs or lesions from backgrounds such as X-ray images, providing critical morphologic and spatial information of these images [19]. In this study, hybrid segmentation methods were utilized to extract characteristics and enhance the imaging detection performance, including grayscale conversion, binary transformation, skeleton extraction, central axis transformation, gradient extraction, and K-means method (Fig. 2). In (a), images were segmented using the watershed algorithm.

- (A) The original image.
- (B) The original image after binarization. The threshold ranged from 90 to 255, and the pixel value 255 became 1 after binarization.
- (C) The gradient of the image obtained after noise filtering, using gradients below 10 as the starting gradient points.
- (D) The gradient and marker information are used to generate a gradient-based watershed map.

In (b), images were segmented using the K-means clustering algorithm.

- (A) The original image.

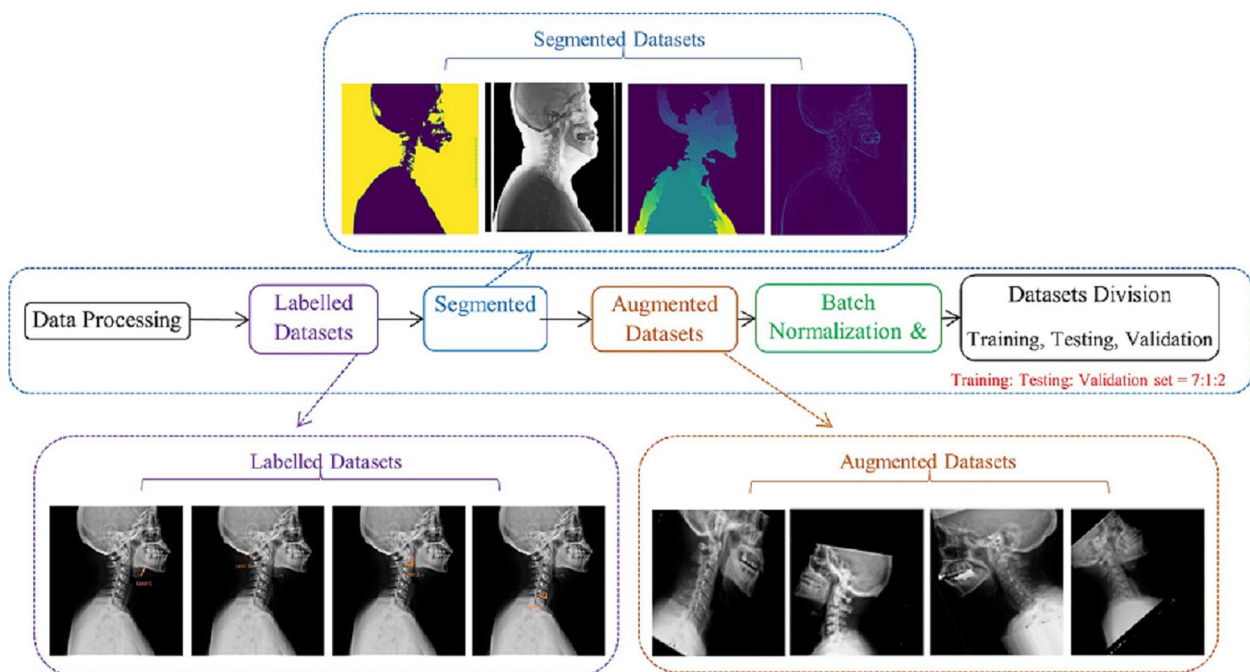


Fig. 1 Dataset processing, including data labelling, data segmentation and data augmentation

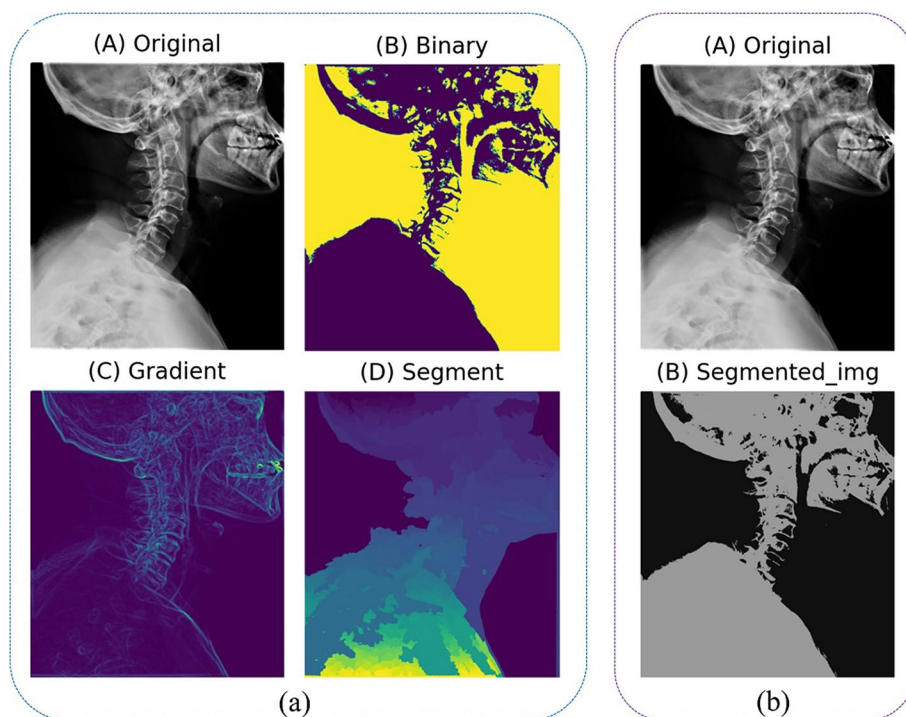


Fig. 2 Hybrid segmentation

(B) This image was obtained using K-means clustering. CNNs have achieved leading-edge capability in numerous biomedical image categorization assignments [20] including diverse modalities [21]. They require a richly tagged database, where the category of each of the pixels or voxels is known, to direct the processing of the database. Yet, collecting intensively tagged biomedical pictures is challenging, as tagged medical databases demand field-specific information and the pixel level annotations are potentially time consuming [22].

Data enhancement is an efficient method for increasing the number and diversity of datasets via stochastic moulding [23]; In the imaging field, commonly available augmentation techniques include image resizing, rescaling, and vertical rotation. In this research, rotational variation, breadth variation, altitude variation, stochastic clipping, scaling variation, and horizontal flipping were implemented to execute data augmentation. After data augmentation, each image was resized to 64 x 64 pixels for the CNN models and 180 x 180 pixels for the transfer learning models.

Data segmentation is a useful technique for removing superfluous features from difficult laryngoscopy images; however, poor contrast and imprecise brightness distributions in laryngoscopy imaging may result in unreliable

segmentation [24]. Histogram equilibrium plays a crucial role in image quality enhancement [25], and it was utilised to mitigate such issues in this study [26]. Category weights are utilized during model training to evaluate the loss function. This manipulation enables the target model to 'give more consideration' to samples from representative categories. The weights are based on the proportion of difficult and simple laryngoscopy.

Proposed indexes

This study presents two new indicators that reflect the range of motion of the upper and lower cervical spine in relation to difficult laryngoscopy and compares them to two previously published indicators to determine their efficacy. Figure 3 depicts the two new indicators, Label-3 and Label-4, in addition to the two previously published indicators, Label-1 and Label-2. The first new indicator, Label-3, is the angle between the lower margin of the second cervical spine and the vertical direction. This measurement is of utmost importance because it reflects the degree of anterior laryngeal displacement, which is known to be associated with difficult laryngoscopy [27]. A smaller angle in this measurement indicates a higher likelihood of difficult laryngoscopy. The angle between the lower margin of the sixth cervical spine and the vertical direction is the another new indicator, Label-4. This measurement is crucial because it reflects the degree of

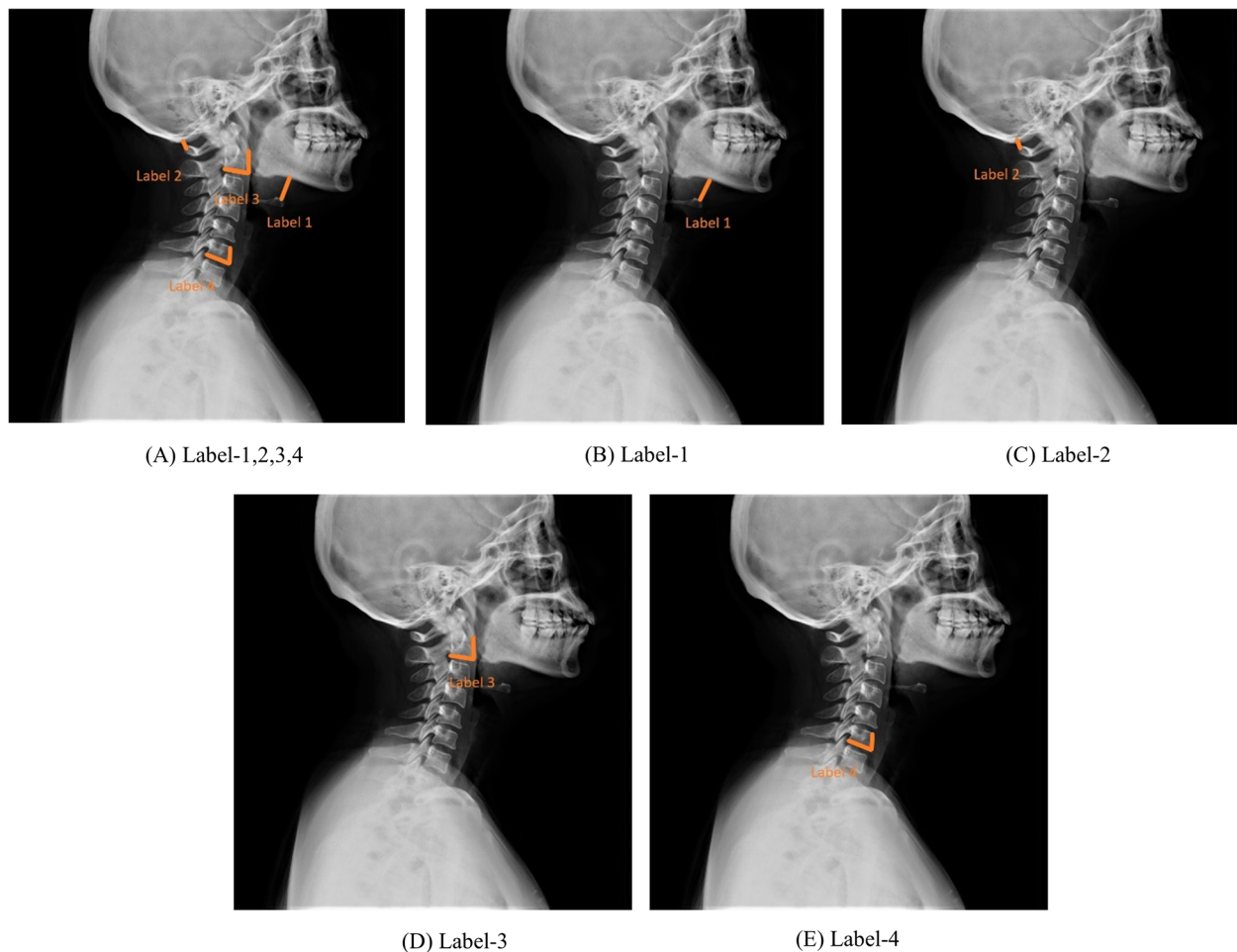


Fig. 3 Labelled laryngoscope imaging

cervical flexion, which can also contribute to difficult laryngoscopy. The likelihood of encountering difficult laryngoscopy increases as the angle of this measurement increases. Label-1: Vertical distance from the highest point of the hyoid bone to the mandibular body [28]. Label-2: Atlanto-occipital gap [29]. Label-3: The angle between the lower margin of the second cervical spine and the vertical direction, indicating the range of motion of the upper cervical spine [27]. Label-4: The angle between the lower margin of the sixth cervical spine and the vertical direction, indicating the range of motion of the lower cervical spine.

Each of the four indications pertains to cervical spine information; therefore, they are all intrinsically linked to the identification of a difficult laryngoscopy. No comparison of the four indicators was performed in previous studies. The work evaluated and contrasted the efficacy of four indicators and ranked them in order of efficacy. The neutral position was used for all cervical spinal x-rays.

Classical architecture

Various advances in artificial intelligence (AI) are rapidly sweeping the medical imaging field. They have the ability to correctly interpret external data, draw experience and lessons from it, and adapt flexibly to achieve particular objectives [30]. The key deep learning technique leveraged in these tasks is the convolutional neural network (CNN), a type of deep learning algorithm that hardcodes translational invariance, which is a key feature of image data. CNNs have achieved extraordinary success in medical video classification and detection [31], medical image regression [32], medical image classification [33], medical image segmentation, or image registration tasks. This work applied a six-layer CNN model to identify difficult laryngoscopy. The implementation of maxpooling structures and sigmoid functions came after each CNN layer. At the end of the model, a dense layer was applied.

Many deep learning and data mining algorithms assume that the training data and the future real data

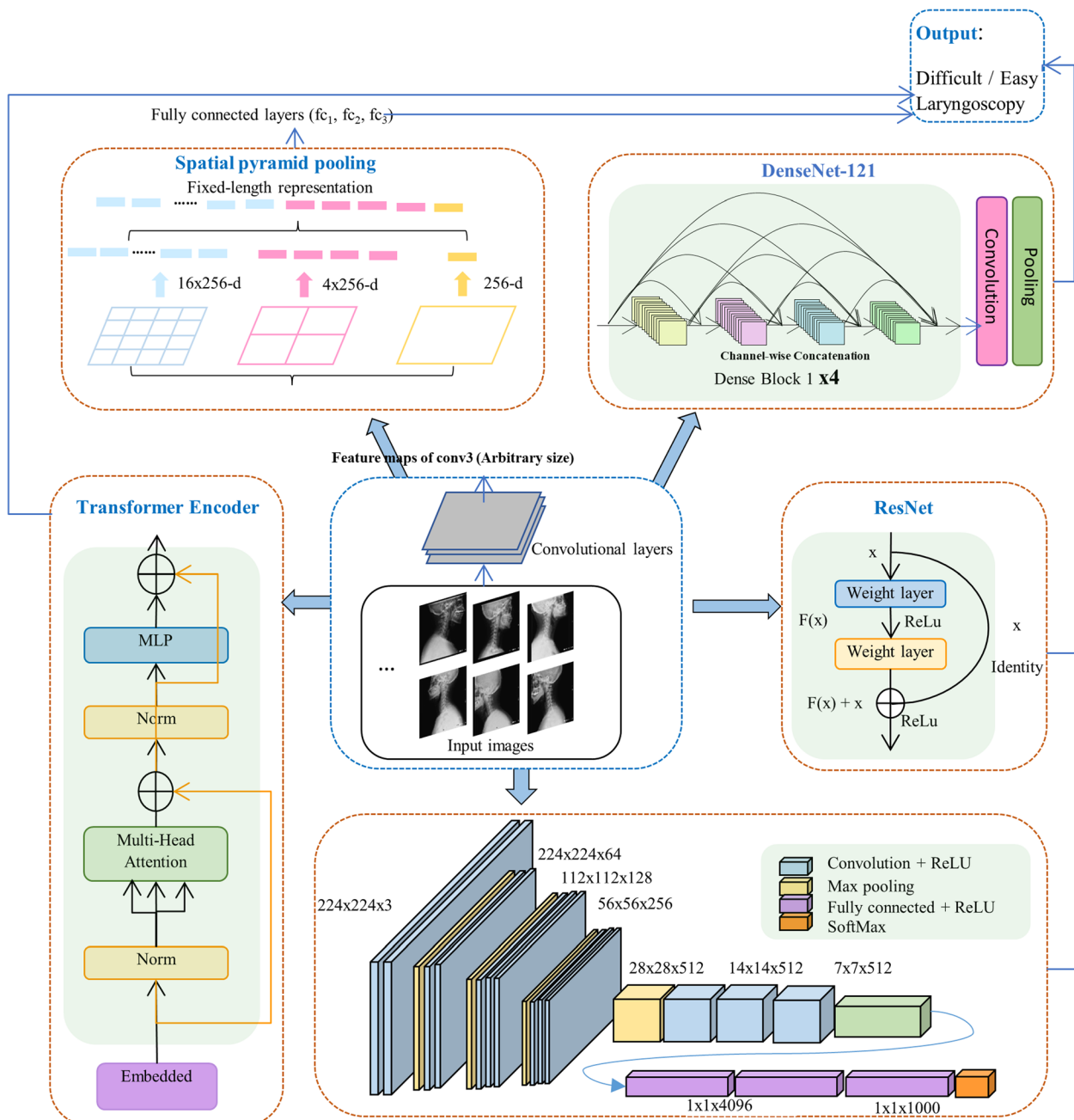


Fig. 4 Adaptive multi model structure

must have the same characteristics and distribution. However, this assumption may not hold true in many real-world applications. Transfer learning models, in contrast, use different domains, tasks, and distributions for training and testing [34]. Pre-training on widely accepted large datasets, such as ImageNet or COCO, can improve the ability of artificial intelligence models to generalize to new medical imaging datasets. This work applied three different transfer learning model: DenseNet-121, ResNet-50, and VGG-16.

Proposed hybrid structure

This work proposes a hybrid artificial intelligence (AI) architecture for the analysis of laryngoscopy images that integrates Convolutional Neural Networks (CNN), spatial extraction, and Vision Transformers (ViT) with attention mechanism (MSCNN) (Fig. 4). The proposed architecture is made up of a CNN for low-level feature extraction, such as lines and blocks, a spatial extraction structure (STN) for capturing multi-scale information, and a ViT for high-level representation learning, such

as cervical spine curvature. Specifically, vision transformers are designed to capture long-range dependencies and global context by calculating loss using positional and patch features from the transformer encoder layers, which has the potential to enhance the efficacy of difficult laryngoscopy prediction; the spatial extraction component can divide the input image into multiple regions and independently process each region, allowing the model to concentrate on relevant areas and increase its localization accuracy.

CNNs are capable of extraction and classification of features. However, they may not be able to capture long-range dependencies and contextual information in the images, which can limit their performance on more complex tasks, such as identification of difficult laryngoscopy imaging. The hybrid model proposed has the potential to enhance performance by leveraging the benefits of its AI structures, specifically by incorporating contextual information from the entire image and learning discriminative features from each structure. Four additional structures, CNN, DenseNet-121, ResNet-50, and VGG-16, were utilized to validate the MSCNN.

The application of large CNN with many layers on a dataset will lead to over-fitting, that is, the models perform well on the training set, but with poor generalization ability. Therefore, it cannot predict on the unknown samples. This causes the CNN model to identify difficult laryngoscopy database of patients in the dataset used, but it cannot be generalized to identify whether other patient images are difficult. In this study, the method of learning employed by the model is adaptive, and its hyperparameters degrade automatically. If the loss value stays the same within ten steps, the early stop will be performed, and the learning rate will decrease. The resolution ratio of the square images of the training and validation sets was reduced to 64 x 64 pixels. The learning rate hyperparameter automatically decreases. Specifically, If the loss function does not decrease over three runs, the model will adapt and reduce the learning rate to 0.6 times, contributing to slower run times and more efficient models, and the model with the lowest loss will be selected. Vision Transformer has evolved into one of the most contemporary and predominant architectures in medical imaging. Transformer is a deep neural network based on a self-attention mechanism that facilitates substantially large receptive fields. It can capture global context with respect to CNN with local receptive fields [35]. Multi-head attention transformer, ViTBase16, was applied to capture global context of input images. Visual transfer models were utilized to improve the accuracy of

difficult laryngoscopy detection. It first segments the imported picture to patches and casts the flattened ones into a feature space, which is processed by its encoder to generate the resulting classified outputs (Fig. 5).

Optimization of proposed hybrid structure

AI is emerging as a formidable implement in the analysis of biomedical imaging, with deep learning employed to categorize airway prediction imaging for advanced detection of difficult laryngoscopy. Standard convolutional actions in deep learning do not expressly consider space-related interaction; therefore, space interactivity is considered as it has the potential to improve the efficiency of difficult laryngoscopy categorization by effectively preserving the spatial information of various scales prior to processing, resulting in robust information extraction. In this research, several granular spatial interaction structures were employed to eliminate the constant size limitation of this model. The spatial extraction structure was added on top of the final convolution layer, with the combination of three max-pooling structures positioned after the second, fourth, and sixth convolutional layers. The spatial extraction structure aggregates characteristics and produces a constant-length outcome, which is then sent to a fully concatenated layer (or another classifier). In other words, in order to avoid clipping or twisting at the beginning, certain “aggregation” is carried out at a higher stage of the model hierarchy (between convolution and fully connected hierarchies) [36] (Fig. 6) shows the structure of a Fine-grained Spatial Interaction. Here, 256 is the filter number of the last convolutional layer. In this study, three visual transfer learning methods (ResNet-50, DenseNet-121, and VGG-16) were used for feature extraction from the laryngoscopy images. All the transfer learning models were pretrained on the ImageNet dataset. The hidden deep layers used in the deep transfer learning map input data to indexes to analyse hidden patterns in complicated data.

Label-1,2,3,4 was used by all of the AI models in this study to validate their performance. Additionally, the following five different label image data were employed on DenseNet-121, CNN+SPP, and Hybrid Model, respectively, in order to completely verify the importance of various variables in discriminating difficult laryngoscopy: Label-1,2,3,4, Label-1, Label-2, Label-3, and Label-4.

Statistical analysis

This research presents a comprehensive approach to binary classification using CNN, ResNet-50, VGG-16, DenseNet-121, CNN+SPP, and a hybrid model that incorporates both CNN, SPP and Vision Transformer (ViT). It compared various aspects, including deep

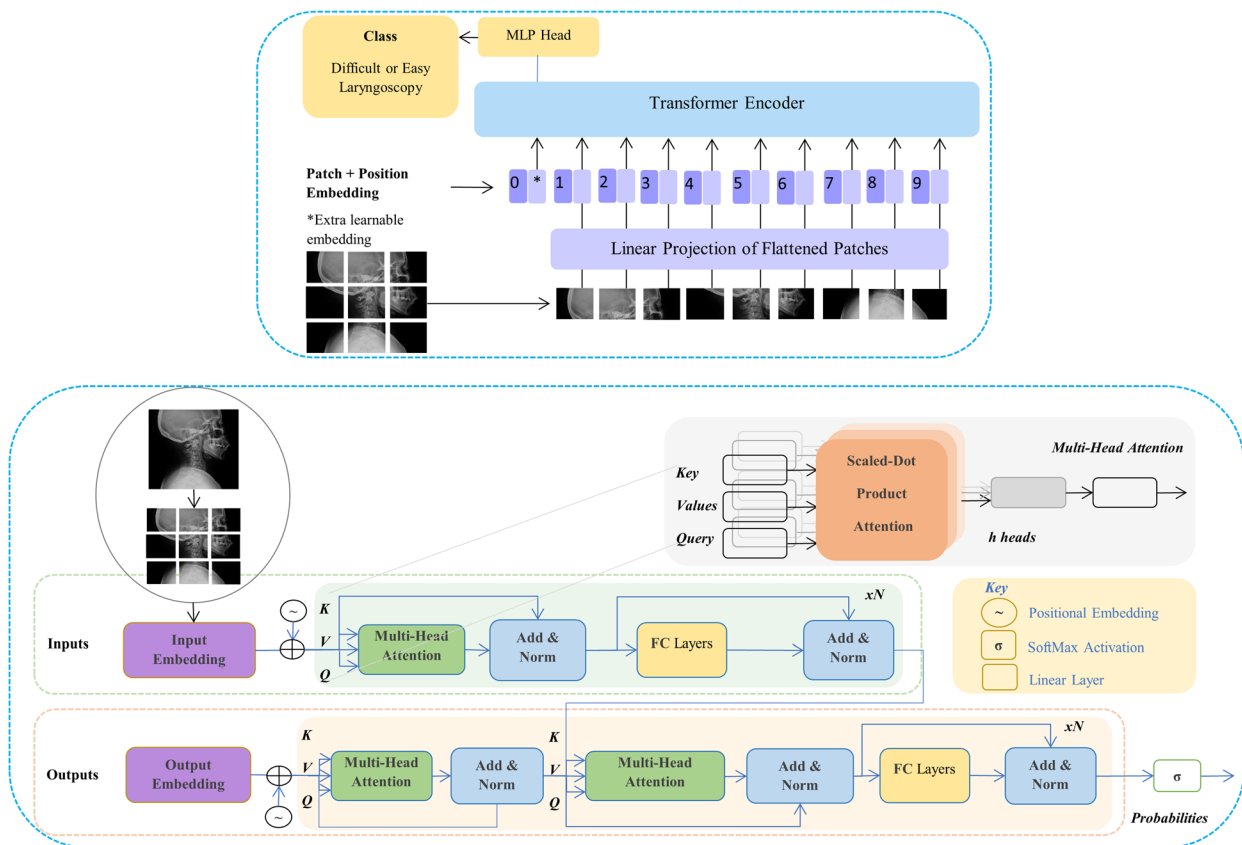


Fig. 5 Visual multi head self attention transformer

learning techniques, evaluation metrics, Python implementation, and descriptive components such as loss functions, activation functions, optimizers, and model architecture.

This study leverages Python, a versatile programming language, and popular deep learning library, TensorFlow. It provides the necessary tools for model construction, training, and evaluation. For binary classification, this research employed binary cross-entropy. ReLU (Rectified Linear Unit) and sigmoid are utilized in this study, they are crucial for introducing non-linearity into the model. They enable the network to learn complex patterns and make predictions. Adam and SGD (Stochastic Gradient Descent) are used to update model parameters during training. The selection of an optimizer influences the convergence speed and final model performance.

Model performance is assessed using various evaluation metrics, including test accuracy and the average precision score. Test accuracy is a commonly used evaluation metric for classification models. It measures the proportion of correctly classified instances in the test dataset. In binary classification, it calculates the ratio of true positives (correctly predicted positive instances) and true negatives

(correctly predicted negative instances) to the total number of instances in the test set. The average precision score was calculated in this research. It quantifies the area under the precision-recall curve (PR curve), which plots precision against recall for different classification thresholds. It provides a single value that summarizes the model’s ability to make precise positive predictions while considering all possible classification thresholds. The higher the average precision score, the better the model’s performance. To compute the average precision score, you can use libraries like Scikit-Learn in Python.

This research implements the binary classification model using Python code, leveraging deep learning libraries for model construction and training. Visualization libraries like Matplotlib and Seaborn are employed for result visualization.

Results

Trial setup

A total of 671 patients preoperative cervical spine X-ray images are collected and used in this study, including 548 easy and 123 difficult laryngoscopy patients. The processor uses GPUs. The optimal model is preserved by

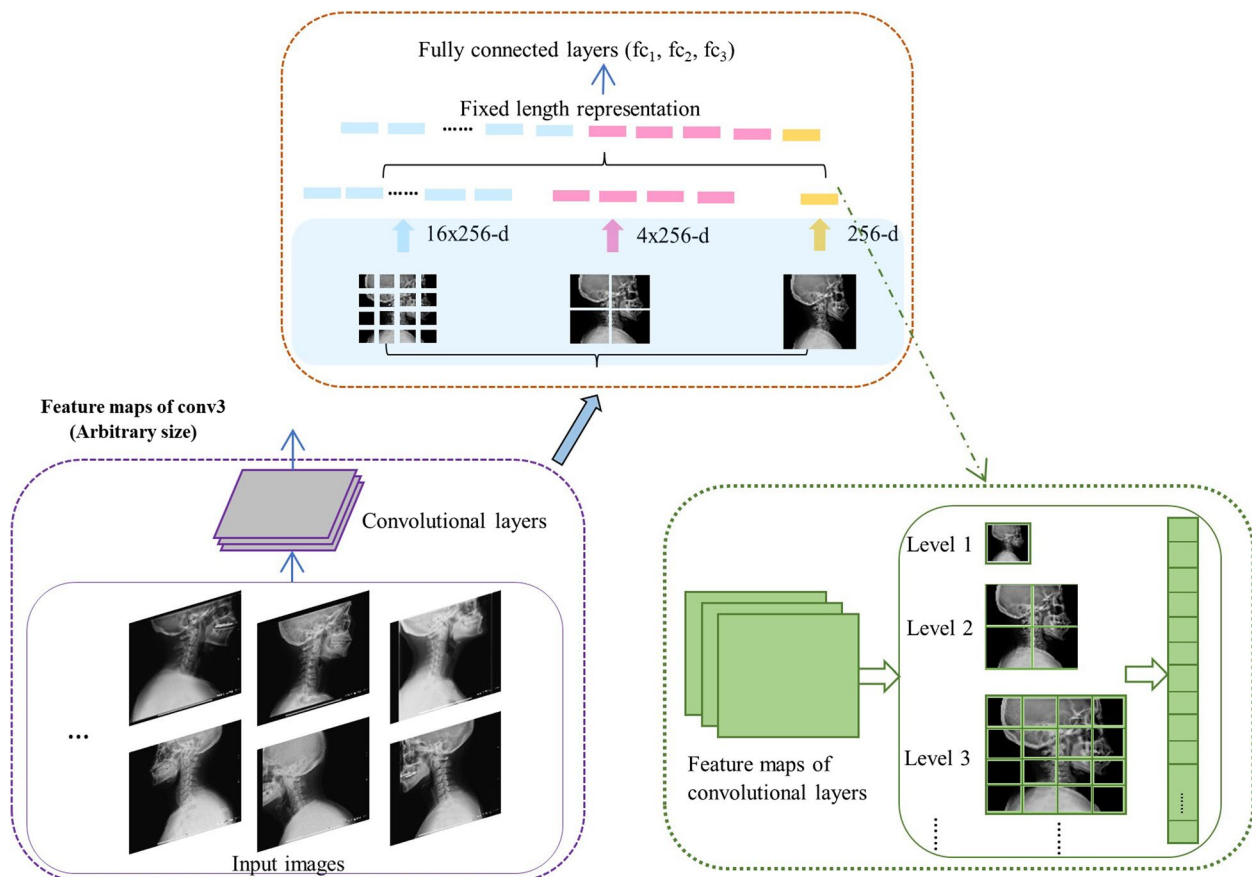


Fig. 6 A network structure with a Fine-grained Spatial Interaction

keeping the model checkpoint with the least loss value. The dataset is shuffled and divided into training, test, and validation sets in a ratio of 7:1:2. The resolution of each image of the training, test and validation sets is reduced to 224 x 224 pixels. The dimension of each image is 700 x 700 with a bit depth of 8.

Accuracy of classical structures using proposed indexes and classical indexes

Figure 7 depicts CNN and three transfer learning models (DenseNet-121, VGG-16, and ResNet-50) without labelling. These results represent the mean of ten iterations of the model on the laryngoscopy imaging dataset. According to the results, DenseNet-121 outperformed ResNet-50 and VGG-16. DenseNet outperforms VGG on the ImageNet dataset; this study supports this finding using a difficult laryngoscopy image dataset.

Accuracy of optimized hybrid structures using proposed indexes and classical indexes

The optimized hybrid structure is combined with convolutional layers, spatial extraction, and vision transformer

structures. Four indicators were applied collectively and separately on this optimized hybrid structure. Label-1,2,3,4 indicates that all four indicators were simultaneously applied to the model, whereas Label-1 indicates that only the first indicator was used to identify difficult laryngoscopy, as do Label-2, Label-3 and Label-4. The efficacy of these indicators in identifying difficult laryngoscopy was evaluated using the X-ray dataset, and a ranking order was established.

Figure 7 shows the results of CNN, DenseNet-121, and vision transformer applying segmentation. No indicators were labelled on the imaging dataset for difficult laryngoscopy, when applying segmentation. Vision Transformer represents the optimized hybrid structure combined with vision transformer structure in Figs. 7 and 8. The hybrid model achieved the best accuracy of 0.8125 before segmentation. Figure 9 depicts the results of comparing four indicators collectively and separately, as well as simple and hybrid model combined with Vision Transformer. The prediction accuracy of the four indicators were compared separately (0.8309 vs. 0.8320 vs 0.8318 vs 0.8320). The rank of the four indicators was proposed,

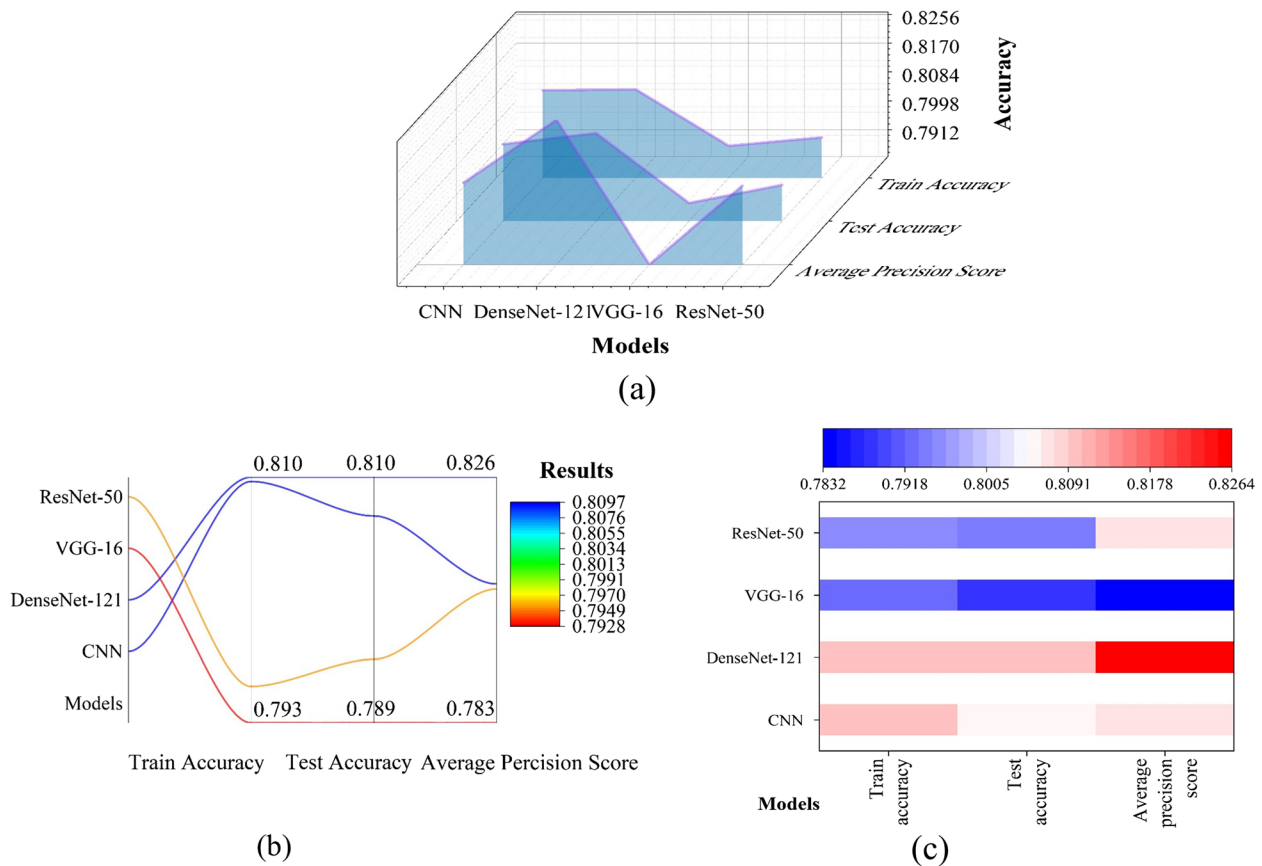


Fig. 7 The test accuracy results of CNN, DenseNet, VGG and ResNet

and Label-2 was the best. Under the combined effect of the four indicators, MSCNN provided the highest level of efficacy, which was 0.8482. Table 2 showed the results of extracting spatial information from different locations. The best place to extract spatial information is to place the spatial extraction structure after each convolutional layer and before batch normalization (Fig. 10). The results showed that the addition of the four indicators improved the performance of difficult laryngoscopy identification.

Data are presented as mean ± standard deviation or number (proportion, %).

Presently, the popular bedside physical predictors contain inter-incisor gap, thyromental distance and modified Mallampati test. Table 3 displays the performance of the modified Mallampati test in predicting difficult laryngoscopy.

Discussion

Proposed indexes

The sufficient prediction of difficult laryngoscopy is considered amongst pivotal priorities for anaesthesiologists in day-to-day work. Nevertheless, the underlying reasons for difficult airways are complicated and there is an

absence of gold standards concerning difficult laryngoscopy. Not much research has been done on applying peri-operative X-ray databases to differentiate participants with difficult laryngoscopy. This study introduced two new indicators for identification of difficult laryngoscopy: Label-3, the angle between the lower margin of the second cervical spine and the vertical direction, denoting the extent of movement at the higher cervical vertebrae, and Label-4, the angle between the lower margin of the sixth cervical spine and the vertical direction, denoting the extent of movement of the lower cervical vertebrae.

Label-1, the vertical distance from the greatest point of the hyoid bone to the mandibular body, reflects the location of the epiglottis, which has been described by Naguib [37] and Chou [38] among earlier researches. Naguib noticed the absence of a distinction in Label-1 in the difficult laryngoscopy and easy laryngoscopy categories. Nevertheless, Chou discovered Label-1 was lengthier for the difficult laryngoscopy category compared to the easier laryngoscopy category. The test accuracy for Label-1 was 0.8309, suggesting a favourable forecast success. Consistent with the study of Chou HC, Label-1 plays an essential role and produces significant results in

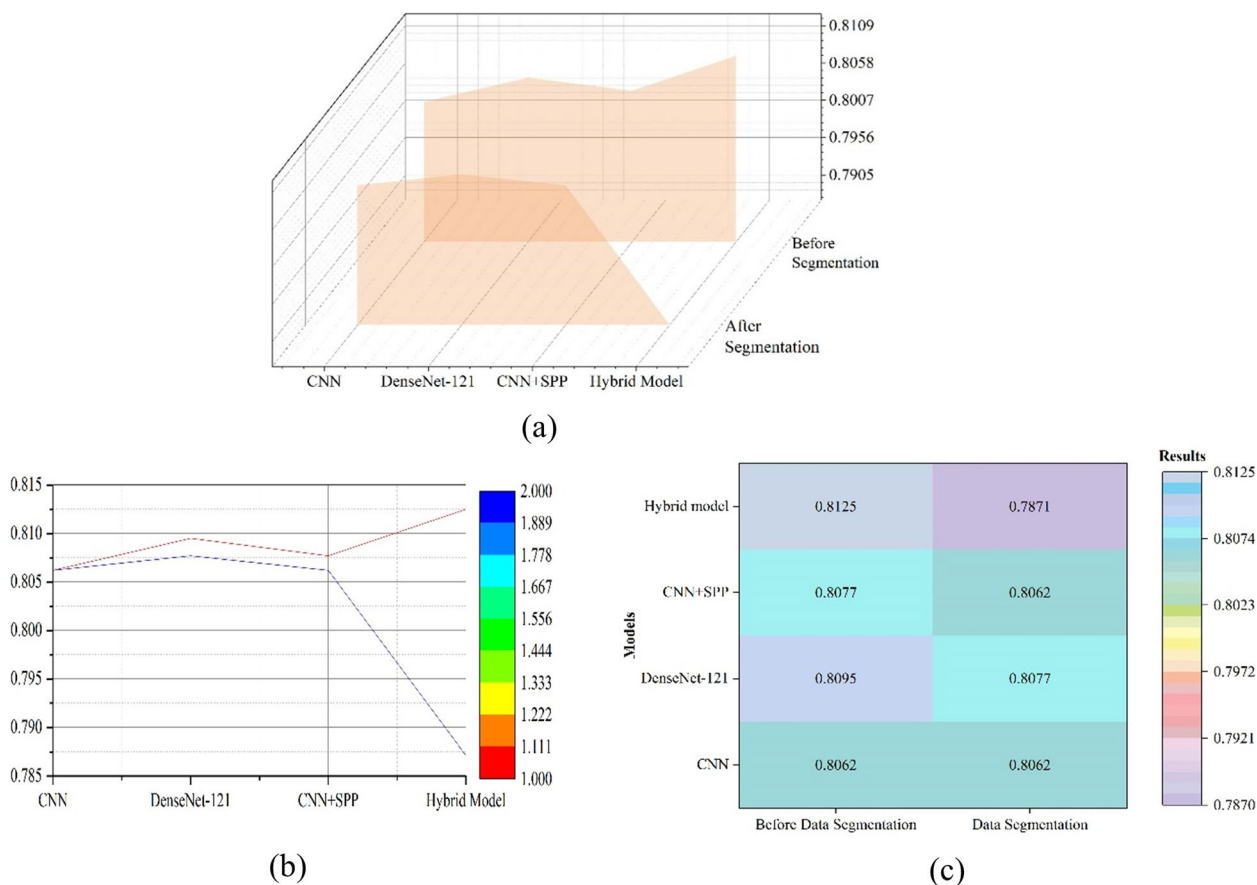


Fig. 8 The outcomes of combining CNN, DenseNet-121, and the hybrid model applying segmentation

predicting difficult laryngoscopy. Horton et al. [39] have proved that the space between the mandible and hyoid is always about 50% of the space between the mandible and the glottis. The large distance between the body of the mandible and the peak of the hyoid bone suggests that the vocal fenestra is profound. In such situations, it was extremely challenging for the anaesthesiologist to reveal the vocal folds due to the presence of tissues in front of the vocal cords. Label-2 is the space separating the occipital bone and the first cervical vertebrae in neutral-positioned intubation participants. Individuals with atlantooccipital damage are at greater risk than normal patients for experiencing difficulties during laryngoscopy [40]. Label-2 is related to the atlantooccipital complex as well as to mandibular prominence. Higher prevalence of distressed airways in those suffering from atlantooccipital composite lesions [40]. The Label-2 has a testing accuracy of 0.8320, indicating a premium level of performance. In addition, the smaller Label-2 length appears to mirror, to some extent, a reduction in movement scope and a slight atlantooccipital joint union. The atlantooccipital was markedly significant in detecting difficult

laryngoscopy among the Macintosh laryngoscopy and assistant technical groups during the research.

Proposed hybrid structure

The proposed MSCNN yielded substantial results. Using MSCNN on a large dataset, the first unified verification of the performance of difficult laryngoscopy identification with a single metric and a combined multimeric was performed. In the study, the predictive accuracies of the four indicators Label-1, Label-2, Label-3, and Label-4 were meaningful (0.8309 vs. 0.8320 vs 0.8318 vs 0.8320). When the four indexes were applied simultaneously, the testing accuracy was 0.8482. The performance of the four indexes, both individually and collectively, was statistically significant. The results showed that the hybrid model reached remarkable levels of performance, surpassing other statistical methods [41] and has the ability to predict difficult laryngoscopy. The usage of four indicators improved the performance of difficult laryngoscopy identification. Jointly and separately, the outcomes of four indicators are comparable, suggesting that clinicians have the clinical discretion to select the most

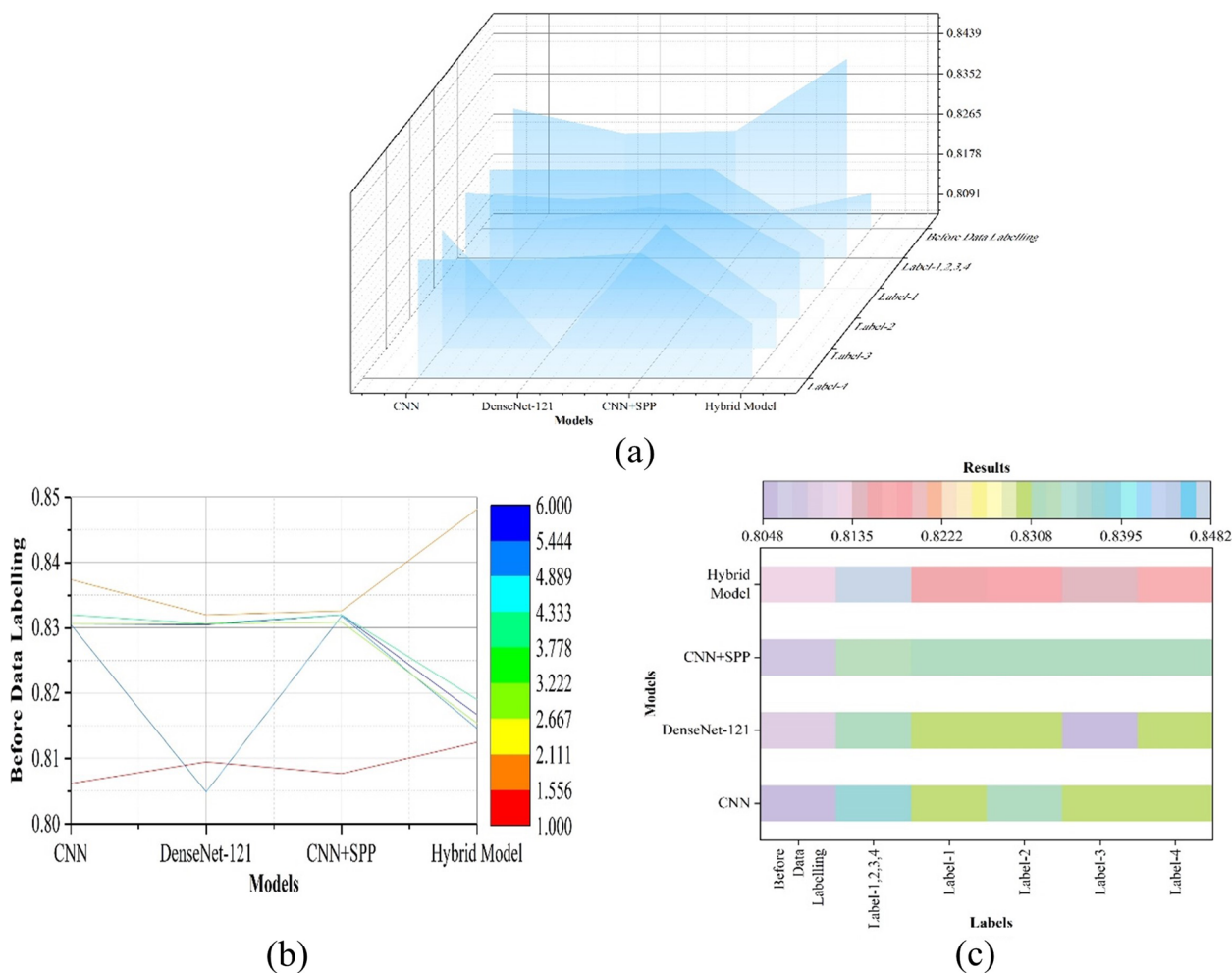


Fig. 9 The outcomes of CNN and visual transfer applications using four indicators collectively and individually

Table 2 The results of extracting spatial information from different locations. SPP is spatial pyramid pooling

Models	Location	Test accuracy	Average precision score
CNN+SPP_1	Location 1: After three convolutional layers	82.93%	82.93%
CNN+SPP_2	Location 2: After each convolutional layer, before Batch Normalization	83.33%	83.74%
CNN+SPP_3	Location 3: After each convolutional layer, after Batch Normalization	83.26%	83.20%

suitable measures for predicting problematic laryngoscopy in clinical practise. The findings demonstrate that it is reasonable to consider substituting alternative indexes when some labelling information in an image dataset is obscured or difficult to annotate due to pose or other factors.

A ranking of the significance of four determinants was presented to assist the anaesthesiologist in recognizing difficult laryngoscopy. As a result, label-2, the atlantooccipital, was discovered to be the most reliable indicator in

determining the difficult laryngoscopy outcomes of the research; it behaved marginally better than label-4, the inferior border of the sixth cervical vertebrae in relation to the perpendicular angle, denoting the extent of movement of the inferior cervical vertebrae; followed by label-1, perpendicular to the mandible from the peak of the hyoid bone, then label-3, the corner of the inferior border of the second cervical vertebrae in relation to the perpendicular, showing the extent of movement of the superior cervical spine.

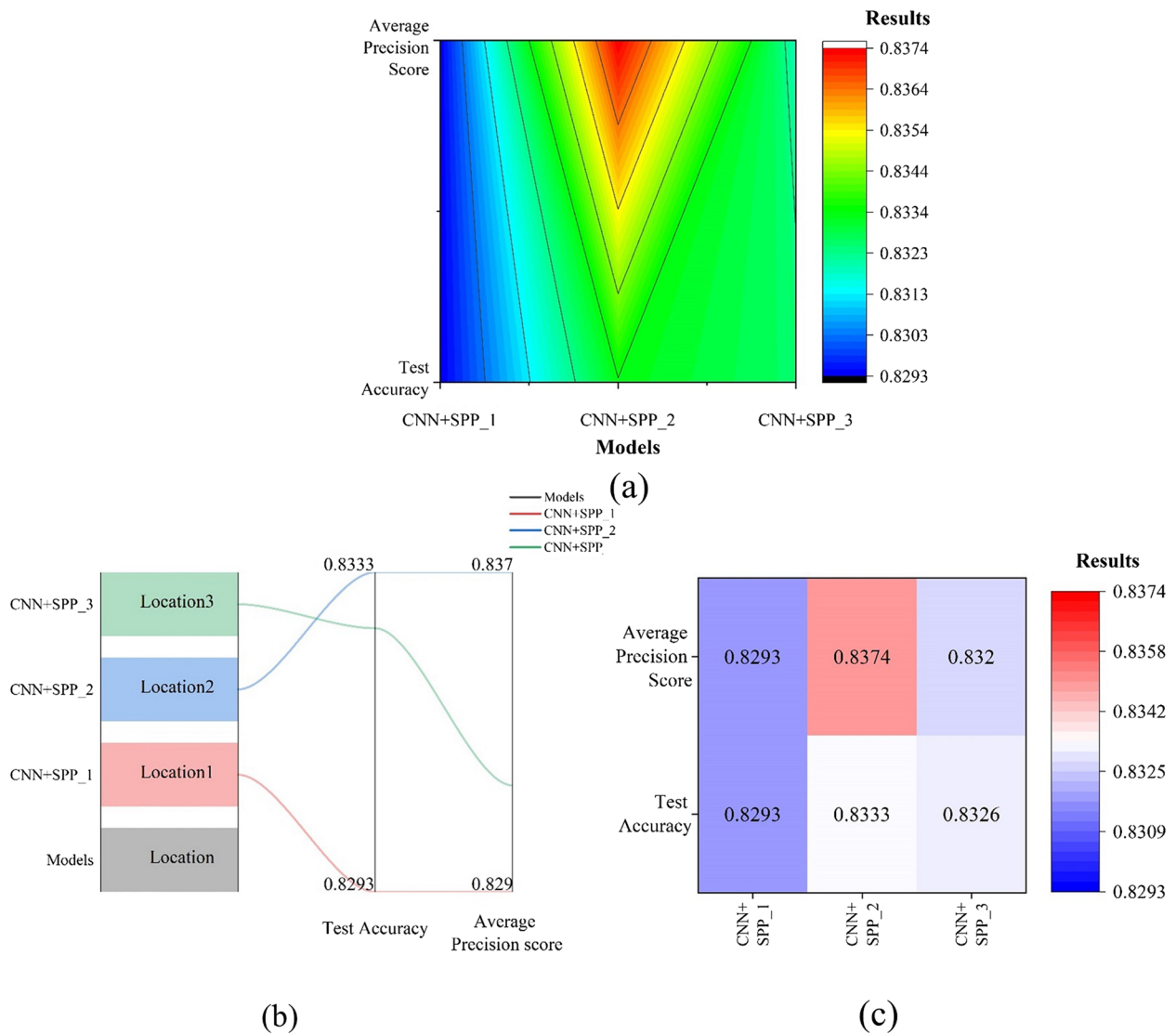


Fig. 10 The results of extracting spatial information from different locations. SPP is spatial pyramid pooling

Table 3 Clinical predictors of the easy and difficult laryngoscopy groups and their values for predicting difficult laryngoscopy

Items	Easy laryngoscopy group (n = 548)	Difficult laryngoscopy group (n = 123)	P values	AUC
IIG (cm)	4.4 ± 0.6	4.1 ± 0.6	< 0.001	0.659
TMD (cm)	7.9 ± 1.5	7.6 ± 1.7	0.028	0.560
MMT [class I II/class III IV]	371(67.7%)/177(32.3%)	64(52.0%)/59(48.0%)	0.003	0.596

^aAbbreviations: IIG inter-incisor gap, TMD thyromental distance, MMT modified Mallampati test, AUC area under the curve

The Mallampati test’s categorization outcome is 0.5960, according to Table 3, this study’s 0.8482 outcome utilizing the MSCNN model is far better than the Mallampati test. The comparison between the simple models and the hybrid model demonstrated that the

proposed model enhanced the performance of difficult laryngoscopy identification, and DenseNet-121 outperformed other transferred structures, demonstrating its adaptability to the laryngoscopy database. ImageNet is the pre-trained dataset for all of these transfer

learning structures, and it differs significantly from the difficult laryngoscopy databases. For the spatial extraction structure, it is determined that the optimal location for extracting spatial information is to position the spatial-extraction structure after each convolution phase and before batch normalisation. Future emphasis will be placed on pre-training transfer learning structures on open-source medical datasets such as Medmnist [42]. Further work will also concentrate on the comparison of a wider variety of metrics related to the difficult laryngoscopy categorization, both separately and holistically.

Conclusion

Causes of difficult laryngoscopy vary widely and there is no gold standard. The first unified validation of difficult laryngoscopy decision-making under both singular and combined multi-indicators is presented. This study demonstrated the reliability and efficacy of two new indicators related to upper and lower cervical motion for identifying difficult laryngoscopy. The efficacy of predicting difficult laryngoscopy was improved by combining two new and two established indicators. The classification of the four indicators independently revealed that the atlantooccipital clearance was slightly better than the other indicators. The MSCNN method outperforms all the other methods in difficult laryngoscopy prediction most of the time. The proposed MSCNN supported by deep learning, spatial extraction and vision transformer structures, enables effective and dependable predictions of difficult laryngoscopy.

Acknowledgements

We would like to thank the Customer Solution Centers, Dell Technologies for providing access to the servers for network training.

Authors' contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication. Xiaoxiao Liu conducted the study and wrote the main manuscript. Sean McGrath acquired and analysed data and references and revised the manuscript. Colin Flanagan gave algorithms guidance and revised the manuscript. Liaoyuan Zeng examined the data and algorithmic logic. Jingchao Fang helped measure the radiologic parameters. Gang Li analysed medical data. Xiangyang Guo gave clinical anesthesia guidance. Yiming Lei and Yongzheng Han conceptualized this work and revised it.

Funding

This study was funded by grants from Wu Jieping Medical Foundation (320.6750.2023-08-5), Key Clinical Projects of Peking University Third Hospital (BYSYZD2021013), Innovation and Transformation Project, Peking University Third Hospital (BYSYZHKC2022103), Clinical Medicine Plus X - Young Scholars Project from Peking University (PKU2022LCXQ031).

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request. The data that support the findings of this study are available from Peking University Third Hospital but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Peking University Third Hospital.

Declarations

Ethics approval and consent to participate

All the procedures were followed in accordance with the relevant guidelines (Ethics approval from Peking University Third Hospital), specifically, ethics permission for this research was obtained from the Medical Scientific Research Ethics Committee of Peking University Third Hospital (IRB00006761-2015021) on 30 March 2015. All participants provide informed consent to participate in the study. Patients were enrolled in this research at the Chinese Clinical Trial Registry (<http://www.chictr.org.cn>; identifier: ChiCTR-ROC-16008598) on 6 June 2016.

Consent for publication

All participants have provided their authorization and consent for the manuscript's publication in case any personally identifiable information or images of the participants are present.

Competing interests

The authors declare no competing interests.

Received: 8 May 2023 Accepted: 1 December 2023

Published online: 04 January 2024

References

- Huang AS, Hajduk J, Rim C, Coffield S, Jagannathan N. Focused review on management of the difficult paediatric airway. *Indian J Anaesth.* 2019;63(6):428.
- Mundotiya S, Palta S, Saroa R, Chhabra S, et al. C-MAC® videolaryngoscope in difficult paediatric airway: Need to update our perspective. *Airway.* 2020;3(3):144.
- Zhong G, Downey RG. Airway topicalisation via direct injection of local anaesthetic into the lumen of high flow oxygenation devices. *Anaesth Intensive Care.* 2020;48(5):409–10.
- Lee J, Kim JS, Kang S, Shin YS, Doo AR. Prediction of difficult airway management in traumatic cervical spine injury: influence of retropharyngeal space extension. *Ther Clin Risk Manag.* 2019;15:669.
- Hews J, El-Boghdady K, Ahmad I. Difficult airway management for the anaesthetist. *Br J Hosp Med.* 2019;80(8):432–40.
- Nasr VG, Abdallah C. Gastroesophageal reflux disease causing a difficult airway. *J Clin Anesth.* 2010;22(5):389–90.
- Han Yz, Tian Y, Xu M, Ni C, Li M, Wang J, et al. Neck circumference to inter-incisor gap ratio: a new predictor of difficult laryngoscopy in cervical spondylosis patients. *BMC Anesthesiol.* 2017;17(1):1–6.
- Han Y, Fang J, Zhang H, Xu M, Guo X. Anterior neck soft tissue thickness for airway evaluation measured by MRI in patients with cervical spondylosis: prospective cohort study. *BMJ Open.* 2019;9(5):e029987.
- Biro P, Hofmann P, Gage D, Boehler Q, Chautems C, Braun J, et al. Automated tracheal intubation in an airway manikin using a robotic endoscope: a proof of concept study. *Anaesthesia.* 2020;75(7):881–6.
- Boehler Q, Gage DS, Hofmann P, Gehring A, Chautems C, Spahn DR, et al. REALITI: a robotic endoscope automated via laryngeal imaging for tracheal intubation. *IEEE Trans Med Robot Bionics.* 2020;2(2):157–64.
- Kim HJ, Min NH, Lee JS, Lee W, Kim DH. Anterior neck soft tissue measurements on computed tomography to predict difficult laryngoscopy: a retrospective study. *Sci Rep.* 2021;11(1):1–8.
- Ock J, Gwon E, Kim Dh, Kim Sh, Kim N. Patient-specific and hyper-realistic phantom for an intubation simulator with a replaceable difficult airway of a toddler using 3D printing. *Sci Rep.* 2020;10(1):1–12.
- Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, et al. Deep learning-enabled medical computer vision. *NPJ Digit Med.* 2021;4(1):1–9.
- Ji C, Ni Q, Chen W. Diagnostic accuracy of radiology (CT, X-ray, US) for predicting difficult intubation in adults: A meta-analysis. *J Clin Anesth.* 2018;45:79–87.
- Krage R, Van Rijn C, Van Groeningen D, Loer S, Schwarte L, Schober P. Cormack-Lehane classification revisited. *Br J Anaesth.* 2010;105(2):220–7.
- Frerk C, Mitchell VS, McNarry AF, Mendonca C, Bhagrath R, Patel A, et al. Difficult Airway Society 2015 guidelines for management of unanticipated difficult intubation in adults. *Br J Anaesth.* 2015;115(6):827–48.

17. Sharma D, Prabhakar H, Bithal PK, Ali Z, Singh GP, Rath GP, et al. Predicting difficult laryngoscopy in acromegaly: a comparison of upper lip bite test with modified Mallampati classification. *J Neurosurg Anesthesiol*. 2010;22(2):138–43.
18. Cubuk ED, Zoph B, Mane D, Vasudevan V, Le QV. Autoaugment: Learning augmentation policies from data. arXiv preprint arXiv:1805.09501. 2018.
19. Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. *J Digit Imaging*. 2019;32(4):582–96.
20. Liu F, Tian Y, Chen Y, Liu Y, Belagiannis V, Carneiro G. ACPL: Anti-Curriculum Pseudo-Labeling for Semi-Supervised Medical Image Classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. p. 20697–20706.
21. Huang SC, Shen L, Lungren MP, Yeung S. Gloria: A multimodal global-local representation learning framework for label-efficient medical image recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. p. 3942–3951.
22. Hu X, Zeng D, Xu X, Shi Y. Semi-supervised contrastive learning for label-efficient medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2021. p. 481–490.
23. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;25.
24. Kuo CFJ, Lai WS, Barman J, Liu SC. Quantitative laryngoscopy with computer-aided diagnostic system for laryngeal lesions. *Sci Rep*. 2021;11(1):1–9.
25. Acharya UK, Kumar S. Genetic algorithm based adaptive histogram equalization (GAAHE) technique for medical image enhancement. *Optik*. 2021;230:166273.
26. Kharel N, Alsadoon A, Prasad P, Elchouemi A. Early diagnosis of breast cancer using contrast limited adaptive histogram equalization (CLAHE) and Morphology methods. In: 2017 8th International Conference on Information and Communication Systems (ICICS). IEEE; 2017. p. 120–124.
27. Zhou Y, Han Y, Li Z, Zhao Y, Yang N, Liu T, et al. Preoperative X-ray C2C6AR is applicable for prediction of difficult laryngoscopy in patients with cervical spondylosis. *BMC Anesthesiol*. 2021;21(1):1–7.
28. Han Y, Tian Y, Zhang H, Zhao Y, Xu M, Guo X. Radiologic indicators for prediction of difficult laryngoscopy in patients with cervical spondylosis. *Acta Anaesthesiol Scand*. 2018;62(4):474–82.
29. Liu B, Song Y, Liu K, Zhou F, Ji H, Tian Y, et al. Radiological indicators to predict the application of assistant intubation techniques for patients undergoing cervical surgery. *BMC Anesthesiol*. 2020;20(1):1–8.
30. Kaplan A, Haenlein M. Siri, Siri, in my hand: who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Bus Horiz*. 2019;62(1):15–25.
31. Loey M, Smarandache F, M Khalifa NE. Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning. *Symmetry*. 2020;12(4):651.
32. Fiorentino MC, Moccia S, Capparuccini M, Giamberini S, Frontoni E. A regression framework to head-circumference delineation from US fetal images. *Comput Methods Prog Biomed*. 2021;198:105771.
33. Castiglioni I, Ippolito D, Interlenghi M, Monti CB, Salvatore C, Schiaffino S, et al. Artificial intelligence applied on chest X-ray can aid in the diagnosis of COVID-19 infection: a first experience from Lombardy. *MedRxiv: Italy*; 2020.
34. Yoo JY, Kang SY, Park JS, Cho YJ, Park SY, Yoon HI, et al. Deep learning for anatomical interpretation of video bronchoscopy images. *Sci Rep*. 2021;11(1):1–11.
35. Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, et al. Transformers in medical imaging: a survey. arXiv preprint arXiv:2201.09873. 2022.
36. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intel*. 2015;37(9):1904–16.
37. Naguib M, Malabarey T, AlSattli RA, Al Damegh S, Samarkandi AH. Predictive models for difficult laryngoscopy and intubation. A clinical, radiologic and three-dimensional computer imaging study. *Can J Anesth*. 1999;46:748–59.
38. Chou HC, Wu TL. Mandibulohyoid distance in difficult laryngoscopy. *Br J Anaesth*. 1993;71(3):335–9.
39. Horton W, Fahy L, Charters P. Disposition of cervical vertebrae, atlanto-axial joint, hyoid and mandible during X-ray laryngoscopy. *Br J Anaesth*. 1989;63(4):435–8.
40. Calder I, Calder J, Crockard H. Difficult direct laryngoscopy in patients with cervical spine disease. *Anaesthesia*. 1995;50(9):756–63.
41. Moustafa MA, El-Metainy S, Mahar K, Mahmoud Abdel-magied E. Defining difficult laryngoscopy findings by using multiple parameters: a machine learning approach. *Egypt J Anaesth*. 2017;33(2):153–8.
42. Yang J, Shi R, Ni B. Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE; 2021. p. 191–5.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

