



Artificial Intelligence in Respiratory Endoscopy

Zhicheng Cao, Bin Ye, Yong Zhou, and Bin Liu

Contents

1	Introduction	2
2	Application of Artificial Intelligence to Nasopharyngolaryngoscopy (NPL)	3
2.1	Quality Control Issues with NPL and Current AI Applications	3
2.2	Categorization of Vocal Cord Abnormalities	3
2.3	Identification of Laryngeal and Pharyngeal Malignancies	4
2.4	Identification of Nasal and Nasopharyngeal Lesions	4
2.5	Future Research Directions for AI Application in NPL	5
3	Application of Artificial Intelligence to Bronchoscopy	5
3.1	Blind Spot Monitoring During Bronchoscopic Examination	5
3.2	Automatic Diagnosis of Bronchial Lesions	6
3.3	Enhancement of Bronchoscopic Images	6
4	Deep Learning–Based Methods for Respiratory Endoscopic AI	8
4.1	Deep Learning Theory: The Basics	8
4.2	Branch Recognition Methods for Bronchoscopic Quality Control	8
4.3	Lesion Detection Methods for Automatic Bronchoscopic Diagnosis	11
5	Conclusion	14
	References	15

Abstract

Bronchoscopy plays an important role in the diagnosis and treatment of pulmonary diseases which can directly observe the internal conditions of the trachea and bronchi. However, the accuracy of its examination results is greatly affected by the experience, habits, and skill levels of endoscopists, resulting in unstable and even mistaken examination results, i.e., there is a lack of standardization in bronchoscopy.

Z. Cao
School of Life Science and Technology, Xidian University, Xi'an, China

B. Ye
Ruijin Hospital, Shanghai Jiaotong University School of Medicine,
Shanghai, China

Y. Zhou
Xi'an Chest Hospital, Xi'an, China

B. Liu (✉)
Shanghai EndoVista Information Technology Co., Ltd., Shanghai,
China

Meanwhile, different pulmonary endoscopists and physicians may interpret the same lesion differently, leading to inconsistent diagnostic results and corresponding treatment. Similar issues can be found in the area of nasopharyngolaryngoscopy (NPL). Artificial intelligence–assisted bronchoscopic and NPL examination, diagnosis, and treatment has become an alternative to the traditional man-powered medical paradigm. This chapter discusses current innovations and breakthroughs in the area of respiratory endoscopic artificial intelligence. Topics such as blind spot monitoring during bronchoscopic examination, bronchoscopic lesion detection, enhancement of bronchoscopic images, quality control for NPL, categorization, and identification of malignancies at different NPL parts (such as nose, vocal cord, throat, and pharynx) are mentioned. Several deep learning models and algorithms are also studied to show how respiratory endoscopic artificial intelligence is implemented and realized.

Keywords

Artificial intelligence · Bronchoscopy · Nasopharyngolaryngoscopy · Quality control · Lesion detection · Image quality · Deep learning · Neural networks

1 Introduction

The artificial intelligence (AI)-assisted diagnosis of electronic bronchoscopy is an emerging medical practice that utilizes advanced AI technology to assist physicians in bronchoscopy examination and diagnosis. This technology combines advanced computer algorithms of image processing, pattern recognition, deep learning, and big data to automatically interpret and analyze bronchoscopy imagery, helping endoscopists and physicians to diagnose various respiratory system diseases more accurately and efficiently. Specifically, systems of AI-assisted bronchoscopy can preprocess bronchoscopy images; extract key features such as lesion morphology, color, and texture; and then use AI algorithms to automatically analyze and recognize these features. Continuous refinement through comparison and training with extensive clinical data enhances the diagnostic accuracy and consistency of these algorithms, ultimately providing endoscopists with more accurate and reliable diagnostic results.

Presently, AI-assisted diagnosis has witnessed significant growth across various medical domains, owing to its accessibility and affordability. Notably, its application in gastrointestinal endoscopy is experiencing a dramatic upsurge. Initially integrated into computer-aided detection (CAD) to discern the characteristics of tumor and non-tumor colon polyps, AI-assisted endoscopic technology has spurred the rapid evolution of gastrointestinal endoscopic systems. These systems, integrated with computer-aided diagnosis (CADx) and computer-aided detection (CADE), aim to enhance endoscopy diagnostic capabilities, to make the endoscopic workflow more efficient, and to facilitate a more precise risk stratification for common gastrointestinal ailments such as gastrointestinal bleeding and tumors.

The application of AI endoscopic diagnosis within the respiratory system predominantly centers around chest X-rays, chest computed tomography, and lung function assessments for patients with respiratory conditions such as pulmonary nodules, chronic obstructive pulmonary diseases, and interstitial lung diseases. Additionally, research is underway in the domain of mechanical ventilation and the diagnosis of bronchial asthma. For example, as for the application of AI in evaluating pulmonary nodules, Chinese medical experts have proposed the expert-robot multidisciplinary team (MDT), which combines human and computer MDT. Its purpose is to interact human experts with AI systems of pulmonary nodule

assessment to provide personalized diagnostic plans for patients. Such a practice, on the one hand, can fully take advantage of the AI system that experts cannot match, such as accurately assessing the 3D longest diameter, volume, and density of lung nodules. On the other hand, it can also avoid problems of AI such as false positives and false negatives as much as possible.

More specifically, for bronchoscopy—a procedure globally undergone by the millions each year—the development and utility of AI-assisted bronchoscopic diagnosis stand as a crucial element of the diagnosis of respiratory diseases. High-quality endoscopic examination serves to fortify diagnostic precision in these cases. With the help of AI-assisted bronchoscopy technology, endoscopists can evaluate lung nodule experience, enable human-machine interaction during bronchoscopy examination, and integrate different lung imaging modalities. Moreover, this technology enables real-time monitoring during bronchoscopy examinations, provides automatic prompts for missed areas, and standardizes the bronchoscopy operation process, which has advantages such as a high level of automation, low cost, non-objectiveness, and stability, as well as avoiding human fatigue. In a word, it helps achieve precise quality control. Last but not least, AI-assisted bronchoscopic technology can also enable the automatic detection of lesions in the airway, the determination of benign and malignant lesions, the identification of lesions invading the surrounding boundaries and blood vessels, the measurement of lesion invasion depth, the measurement of the length and diameter of airway stenosis sites, the analysis and localization of peripheral lung lesions, and so on so forth.

Nasopharyngolaryngoscopy (NPL) is a daily-used tool in the practice of otolaryngologists, which is heavily relied upon for examinations. It plays a crucial role in diagnosing and evaluating the treatment of diseases of the nasal cavity, oral cavity, pharynx, and larynx, particularly within the context of head and neck cancers (HNCs). Since the 1980s, the advent of NPL has significantly transformed the clinical diagnosis and treatment of otolaryngological diseases by providing clinicians with clear images of the larynx and pharynx. Entering the twenty-first century, the era of digital and 3D imaging has already come, which further enhanced the clarity and three-dimensionality of NPL images.

However, several challenges persist in the area of NPL endoscopy. These include incomplete examination coverage, failure to detect and diagnose lesions, and poor image quality. As a result, more than often, repeated examinations or constant expert supervision are needed. In a word, the current use of NPLs in examinations is significantly influenced by the subjective proficiency of the examining endoscopists. Moreover, there is a glaring lack of third-party supervision and quality control, indicating room for improvement.

2 Application of Artificial Intelligence to Nasopharyngolaryngoscopy (NPL)

Compared to gastrointestinal (GI) endoscopy which is subject to stringent quality control and management (as instructed by international guidelines and consensus as well as expert groups), NPL operations are generally perceived as relatively simple and straightforward. On one hand, there is a lack of specialized endoscopic associations for regulation and standardization. On the other hand, there is a deficiency in dedicated endoscopic personnel, operational quality control, and standardized training. In the medical systems of many countries (such as China), there are no designated NPL endoscopists, further complicating endoscopic quality control. A comprehensive NPL examination requires exposing the entire anatomical site, obtaining a clear view, and identifying abnormal lesions. Due to the unique anatomy of the laryngopharyngeal regions, the examination process demands attentiveness from the operating NPL endoscopist and cooperation from the patient to fully expose relevant areas. As such, blind spots remain the primary reason for missed diagnoses in NPL.

The positioning and cooperation of patients directly impact the NPL results. For instance, examining various regions requires specific patient actions: Nasopharynx examination needs nasal breathing, epiglottis examination requires tongue extension, vocal cord examination demands inhalation or phonation, and piriform fossa examination calls for phonation or air inflation. Unsatisfactory exposure of the piriform fossa and retropharyngeal area often results in missed diagnoses. Better exposure of the entire hypopharynx can be achieved through head rotation, the Valsalva maneuver, and the Killian position. However, there is a lack of effective methods for supervising and evaluating the quality of NPL examinations—it cannot be quantitatively assessed whether the NPL endoscopists have thoroughly examined each anatomical area. Given high outpatient volumes in otolaryngology head and neck surgery clinics, it is quite important to consider how high quality and accuracy in NPL can be achieved within a limited timeframe.

2.1 Quality Control Issues with NPL and Current AI Applications

Compared to the rapid development of AI-assisted gastrointestinal (GI) endoscopy, research on AI-assisted NPL is relatively lagging behind, currently still in the laboratory stage, and has not yet been put into large-scale clinical applications. The current research on artificial intelligence electronic nasopharyngoscopy mainly focuses on the identification and diagnosis of anatomical parts, vocal cord lesions, malignant tumors of the throat, and nasal cavity lesions.

Optimizing the accuracy of AI models for NPL relies heavily on high-quality and accurate annotations during the training phase. Unlike the GI endoscopy where the anatomical features throughout different parts of the GI are quite similar to each other, the anatomical features are significantly different at different parts of the NPL, such as the nasal cavity, oral cavity, larynx, and pharynx. Thus, AI models are able to discern the various positions within NPL more accurately than they can within GI endoscopy, yielding consistent and superior classification rates. However, there are no established international guidelines or consensus about which standard anatomical sites should be covered during NPL procedures.

In 2023, Zhu et al. identified 20 key anatomical landmarks based on six anatomical divisions of the nasal cavity, oral cavity, pharynx, nasopharynx, oropharynx, and hypopharynx. They formulated an AI-driven Intelligent Nasopharyngoscope Localization Monitoring Assistant (ILMA), achieving a classification accuracy of 97.60% through the process of training and validation utilizing endoscopic images. In the same year, Nakajo et al. assessed pharyngeal and laryngeal images captured during upper GI endoscopy with NBI or blue laser imaging (BLI). Their study mainly focused on the oral cavity, oropharynx, hypopharynx, and larynx, categorizing these into 15 distinct anatomical sites. Their model achieved a classification accuracy of 93.3% for pharyngeal and laryngeal images and incorporated a functionality of blind spot identification during examinations.

2.2 Categorization of Vocal Cord Abnormalities

The classification of vocal cord abnormalities in images has risen as a primary area of interest in the AI applications to laryngoscopy. Deep learning (DL) models have been developed utilizing vocal cord traits such as hue, texture, and geometry. These DL models employ laryngoscopic images of the vocal cords for training and validation, facilitating objective and effective identification of both normal and abnormal vocal cord conditions, thereby diminishing diagnostic discrepancies among endoscopists.

In 2020, Ren et al. introduced a CNN-based AI classification model that reached recognition accuracies of 98% for vocal cord nodules, 91% for polyps, 91% for leukoplakia, and 90% for malignant tumors, exceeding the performance of laryngologists at all levels. In 2021, Kuo et al. established an objective AI-assisted diagnostic system specifically designed to detect laryngeal lesions. By employing contour features of vocal cord images captured from laryngoscope videos, this system selects clear images, thereby enhancing detection performance. Additionally, it automatically segments the vocal cords and performs a quantitative analysis of their

hue and geometric features. It achieved a classification accuracy of 100% for healthy vocal cords and leukoplakia and 93.15%, 95.16%, and 96.42% accuracy for polyps, cysts, and tumor lesions, respectively. Also in 2021, Cho et al. became the first group to compile pathological diagnoses often misdiagnosed as vocal cord polyps by laryngologists, including Reinke's edema (18.4%), vocal cord cysts (13.8%), nodules (10.5%), papilloma (9.6%), and leukoplakia (6.2%). The AI model's ability to diagnose Reinke's edema and vocal cord nodules was found to be comparable to the top-performing laryngologists.

In 2022, Zhao et al. utilized a deep convolutional neural network (DCNN)-based diagnostic system to distinguish among four types of vocal cord lesions: normal, polyps, keratosis, and cancer. The overall accuracy of this system was 80.23%. It excelled in diagnosing three out of four types of lesions (except keratosis), surpassing clinical otolaryngologists in terms of accuracy and sensitivity while requiring significantly less time (6.47 s vs. 138.5 min). Moreover, when classifying vocal cord lesions into nonurgent (normal, polyps) and urgent (keratosis, cancer) categories, the system demonstrated strong classification performance with an overall accuracy of 93.9%. In 2023, Chen et al. applied the 3D VOSNet algorithm to analyze video laryngoscopy, performing segmentation of the vocal cords and glottis. This approach computed six parameters that include the glottal area, vocal cord area, vocal cord length, vocal cord length deviation, and vocal cord symmetry. Such metrics offer an objective assessment of vocal cord conditions, assisting laryngologists in making reliable diagnoses, particularly for conditions such as vocal cord paralysis.

2.3 Identification of Laryngeal and Pharyngeal Malignancies

Identifying malignant tumors constitutes a pivotal objective for NPLs. HNCs rank as the ninth leading cause of cancer-related mortality, with the incidence rising in recent years [1]. Since these cancers directly affect patients' ability to swallow, speak, and breathe, the importance of early diagnosis and minimally invasive treatment for improving survival rates and quality of life cannot be overstated. AI has garnered considerable recognition for its successes in identifying malignancies of the larynx and pharynx, greatly enhancing the sensitivity and accuracy of tumor detection.

In 2019, Xiong et al. developed and trained a DL model for assisting in laryngoscopic diagnoses, achieving an accuracy of 86.7% and a sensitivity of 73.1% in recognizing laryngeal cancer and precancerous lesions. This performance provided significant advantages for early diagnosis of laryngeal cancer compared to specialists with 10 to 20 years of experience. In 2020 and 2021, Atsuko et al. and Mitsuhiro

et al., respectively, designed DL models to recognize images of pharyngeal cancer obtained from upper GI endoscopy using WLI, NBI, and BLI. These models achieved recognition accuracies of 100% and 92%, although the number of patient cases included in the validation assessment was relatively small, comprising 40 and 25 cases, respectively. In 2023, Li et al. conducted a large-scale, multicenter study on the Laryngopharyngeal AI Diagnostic System (LPAIDS), which they developed to recognize laryngopharyngeal cancer images in WLI and NBI. They evaluated its effectiveness using real-time laryngoscopy videos and found that the system demonstrated superior diagnostic performance with impressive accuracy (0.949–0.984) and sensitivity (0.901–0.986). In 2024, Sampieri et al. developed the DL model SegMENT-Plus to identify and delineate the superficial boundaries of laryngeal cancer in laryngoscopic images of WLI and NBI. This model performed comparably to otolaryngology residents and was evaluated using five real surgical laryngoscopy videos, showcasing its potential to assist in surgical decision-making and margin delineation.

However, research on DL models for endoscopic recognition of oropharyngeal cancer remains in the early stages. In 2021, Paderno et al. used a fully CNN model to analyze NBI endoscopic videos of oral (43 cases) and oropharyngeal cancers (35 cases). The model automatically segmented tumors with Dice similarity coefficients (DSC) of 0.6559 and 0.7603, respectively, demonstrating the potential of AI in the analysis and segmentation of endoscopic images of oral and oropharyngeal cancers.

2.4 Identification of Nasal and Nasopharyngeal Lesions

Nasal polyps and inverted papilloma are common benign lesions in the nasal cavity; the former represent inflammatory changes, while the latter constitute benign tumors. The strategies for medication, surgery, and prognosis between the two differ significantly. However, distinguishing between them under nasopharyngoscopy can present a substantial challenge, particularly for less experienced otolaryngologists. To address this issue, Girdler and colleagues devised a CNN-based diagnostic model in 2021 that reliably identified nasal polyps and inverted papilloma in endoscopic images, performing comparably to an endoscopy specialist with 7 years of experience (74.2% vs. 74.9%). In the following year, Ay et al. presented a reliable CNN-based diagnostic system for nasal polyps, which automatically recognized these lesions in endoscopic images with an impressive accuracy rate of 98.3%.

Nasopharyngeal carcinoma (NPC) is a common malignant tumor in the head and neck region, particularly prevalent in Southeast Asia, where it has the highest incidence rate of 5.5

per 100,000 individuals. In 2022, China reported 51,010 new cases of NPC, representing 42.4% of the global total and establishing China as the country with the most newly diagnosed patients. In 2018, Li et al. designed a DL-based detection model for nasopharyngeal malignancies (eNPM-DM), which achieved a diagnostic accuracy of 88.7%. This model surpassed NPC specialists, having over 5 years of experience, in a prospective test (88.0% vs. 80.5%). Furthermore, the model was capable of automatically segmenting malignant lesions in endoscopic images, as evidenced by an average DSC of 0.75. In 2023, Shi and colleagues introduced a deep weakly semi-supervised framework, Point SEGTR, for segmenting NPC lesions, thereby contributing to the iterative development of NPC segmentation algorithms for endoscopy. In 2024, Wang et al. utilized an DCNN-based AI system to automatically diagnose NPC using WLI and NBI images from nasopharyngoscopy. Their system achieved accuracies of 92.0% (WLI) and 97.5% (NBI), respectively, with sensitivity and recall rates surpassing those of junior otolaryngologists. These preliminary studies highlight the potential of AI models in enhancing the detection capabilities and efficiency of endoscopic examinations of nasal and nasopharyngeal lesions.

2.5 Future Research Directions for AI Application in NPL

Firstly, quality control of NPL is still one of the most important future research directions for AI applications in NPL. As an integral diagnostic aid for otolaryngologists, the accuracy and reliability of NPL are paramount. Achieving this requires stringent quality control and supervision during the examination process, which is a topic that needs far more attention. Also, the examination process should promote standardization and uniformity, taking into account aspects such as patient positioning, physician placement, coverage of anatomical sites, selection of images, and the drafting of reports. Furthermore, the exposure of key NPL spots such as the pharyngeal opening of the Eustachian tube, epiglottic vallecula, piriform sinus, and retrocricoid area is crucial and necessitates patient cooperation with specific maneuvers.

Secondly, continuous improvement of current AI algorithms is also a crucial problem. Existing AI diagnostic models have shown promising laboratory results in NPL, improving the detection of lesions; however, mature products have not yet emerged. This suggests that AI-assisted diagnostic systems for NPL currently lack robust, multicenter randomized controlled clinical trials. Future research should concentrate on refining algorithmic models to strengthen autonomous learning of images, enhancing the accuracy, reliability, convenience, and adaptability of AI models.

Finally, hoarseness or changes in voice are frequent complaints prompting clinical NPL examinations; thus, assessing vocal cord lesions and alterations in voice quality is a valued potential functionality of NPL AI systems. Nevertheless, the current market provides no NPL products capable of recording patients' voices. Future improvements could include incorporating modules for voice and vocal cord movement analysis into AI-assisted NPLs (Fig. 1).

3 Application of Artificial Intelligence to Bronchoscopy

3.1 Blind Spot Monitoring During Bronchoscopic Examination

The blind spot monitoring system of bronchoscopic examination can be realized using deep learning algorithms and models such as deep convolutional neural network (DCNN) and deep reinforcement learning (DRL). Such a real-time examination quality control system can be used to monitor blind spots and suspicious lesions and optimize the efficiency of bronchoscopy examination.

When designing the blind spot monitoring system, the trachea and bronchi images are categorized into 31 classes according to the 31 anatomical branches, whereby normal anatomical structure images are utilized for automated recognition, image capture, and quality control to avoid potential missed lesions caused by time constraints. Simultaneously, routine examination checkpoints are supervised to prevent missed bronchoscopic examinations. The monitoring system functions as a teaching tool for beginners by providing reminders for the timing and coverage of bronchoscopy examinations, facilitating an understanding of the bronchial anatomical structure, and identifying missed areas. The system will display a red box prompt on the screen when certain anatomical points are missed or when images are obscured, such as in cases where an anterior segment of the upper right lobe is not reached during bronchoscopy or in cases of unclear exposure. In addition, the time-lapse of the bronchoscopic examination is also recorded and displayed so that the endoscopist or her/his supervisor can evaluate if the examination is too fast or too slow. For example, a time-lapse shorter than 2 min may suggest the likelihood of missing spots.

Figure 2 shows a commercial blind spot monitoring (i.e., quality control system) developed by a Chinese company, Shanghai EndoVista Information Technology, Inc. The left panel displays coverage of each bronchus on the bronchial tree (green denotes coverage, while gray denotes non-coverage). The right panel displays the real-time imaging of the bronchoscope. The top left part of the screen displays the

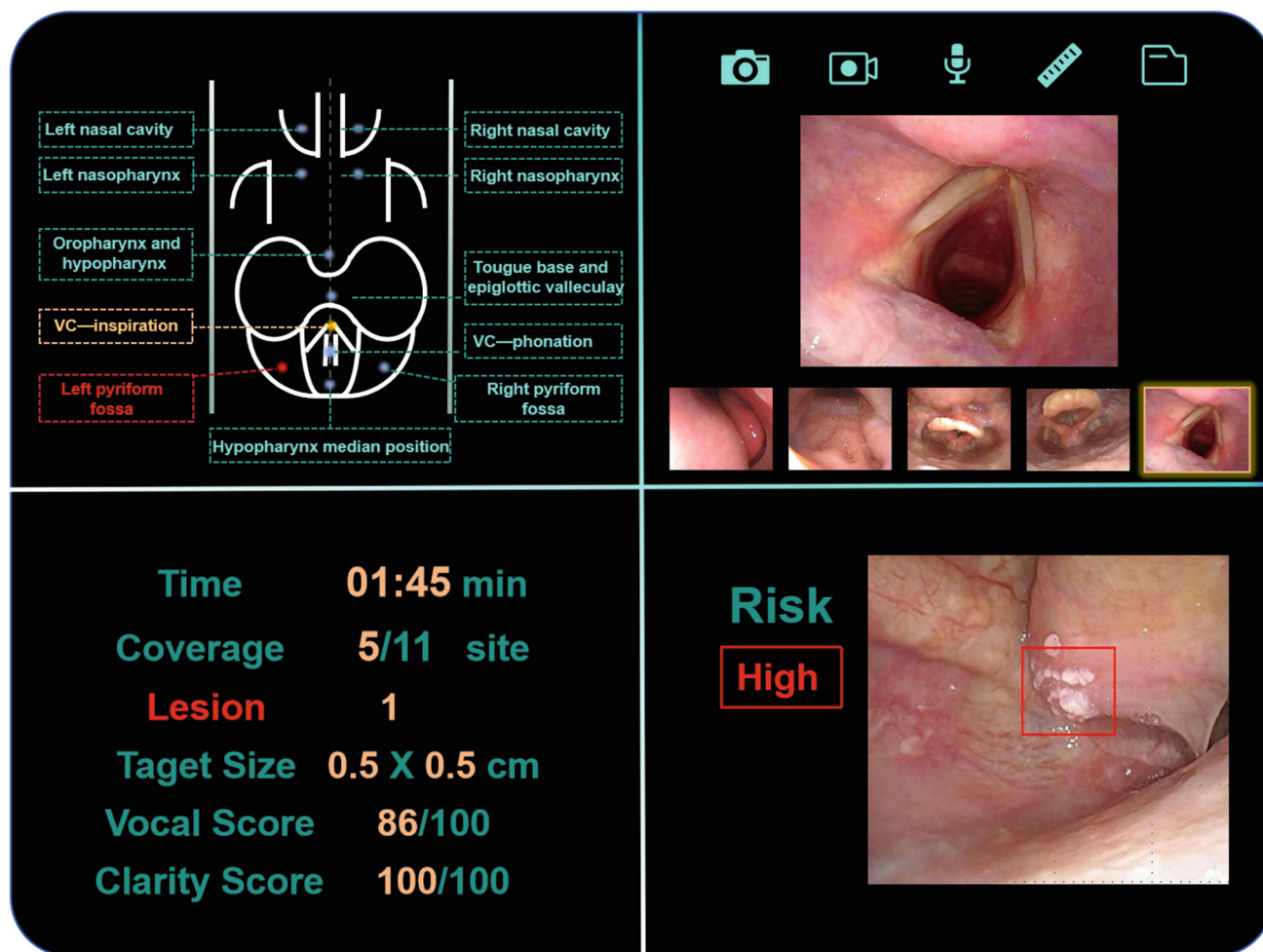


Fig. 1 A functional schematic for an AI-enhanced NPL system. The first part (the upper left of the figure) shows a diagram of 11 anatomical sites, where the colored sites denote the ones covered and recognized during NPL examination; the second part (the upper right of the figure) shows real-time recording capabilities for images, videos, voice, and

dimensions; the third part (the lower left of the figure) shows operational parameters such as examination duration, coverage, image clarity, and the number and size of lesions; the fourth part (the lower right of the figure) shows risk assessment for targeted lesions

examination time-lapse and coverage summary (ratio of coverage and current bronchus).

3.2 Automatic Diagnosis of Bronchial Lesions

In addition to blind spot monitoring during the bronchoscopic examination, AI-assisted bronchoscopic diagnostic system can also identify and locate bronchial lesions. With a thorough understanding of diverse pathological lesion characteristics, AI algorithms which further leverage the potential of deep learning and data mining are able to identify and categorize pathologically confirmed lesions, differentiate between benign and malignant airway conditions, and categorize bronchial tuberculosis using extensive clinical data.

In accordance with the diagnosis guidelines for tracheal and bronchial tuberculosis, ultimately aiding clinical

practitioners in the precise and prompt identification of airway tuberculosis, it leads us to earlier treatment interventions, and enabling real-time lesion diagnosis (highlighted by a red box prompt, as shown in Fig. 3). Furthermore, by utilizing artificial intelligence, the bronchoscopic diagnostic system can discern lesion size, infiltration depth, area, surrounding boundaries, and diameter and length of the narrowed lumen, which enables the idea of dynamic and personalized surgical plans and treatment.

3.3 Enhancement of Bronchoscopic Images

It is also another important problem that the AI-assisted bronchoscopic monitoring system can enhance the images of bronchial mucosa during the examination. Currently, during bronchial examinations, due to the rapid movement of the

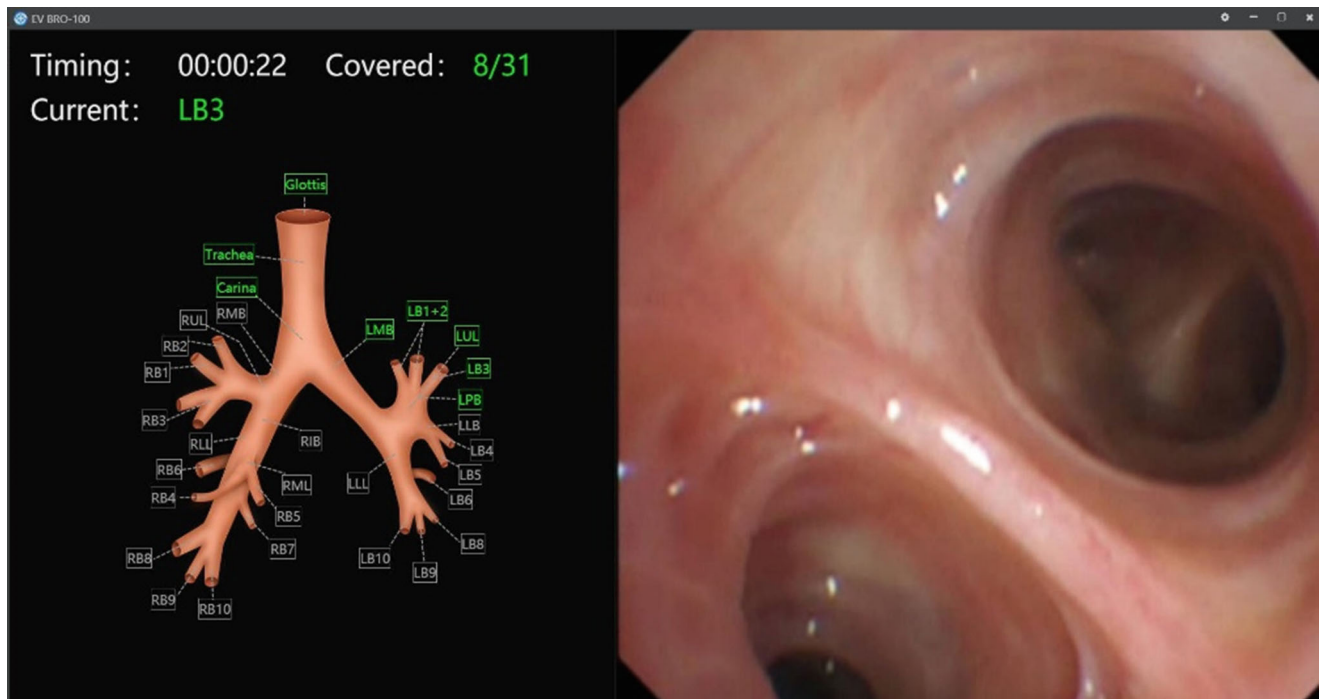


Fig. 2 The GUI of a blind spot monitoring system of bronchoscopic examination, which monitors and displays the 31 bronchial sites in real time on a computer screen



Fig. 3 An example of the automatic lesion detection result by an AI-assisted bronchoscopic diagnosis system

bronchoscope at certain times—especially in areas that are easily overlooked or missed—the quality of the stored bronchoscopic images can be very low. This issue of low image

quality causes a problem when bronchoscopists or physicians conduct a follow-up check or re-examination. Such a low image quality is also difficult for the bronchoscopic AI

algorithms, and mistakes in the recognition of bronchial parts or detection of lesions could happen, leading to a less reliable bronchoscopic AI system.

The countermeasures of the issue mentioned above are to either assess the image quality and discard the image if it is below a certain quality threshold or to conduct a process of image enhancement on the low-quality bronchoscopy images originally stored. The first countermeasure is more straightforward and easier to implement. Corresponding methods for image assessment are designed using the information and metrics of the image such as illuminance, contrast, noise, blurring level, and color.

The second countermeasure is, on the other hand, more complex. Corresponding methods for image enhancement include Gamma transform, histogram equalization, and Retinex algorithm, for image contrast enhancement, filter-based image denoising methods for image denoising, and blind deconvolution for image deblurring. More recently, deep learning-based methods have become the mainstream for image enhancement. For example, DnCNN has been shown to surpass traditional denoising methods. DeblurGAN and RetinexNet have demonstrated success for the tasks of image deblurring and image contrast enhancement, respectively. Figure 4 illustrates the comparison between the original blurry bronchoscopic image and its deblurred result by DeblurGAN.

4 Deep Learning–Based Methods for Respiratory Endoscopic AI

4.1 Deep Learning Theory: The Basics

Coined in the 1950s, the term artificial intelligence refers to man-made intelligence and functions that resemble human beings. With the development of various theories and techniques of AI—especially the advent of deep learning theory in the early twentieth century—AI has made vast advancements, and numerous applications have emerged in many aspects of our society and life. How to combine AI with medicine (including respiratory endoscopy) is currently a hot topic.

As a special case of neural network, deep learning (i.e., deep neural network, DNN) involves significantly more layers—especially the hidden layers—than traditional non-deep neural networks. Such a deeper structure enables better feature extraction and representation and usually results in higher model performance. Nonetheless, as the network goes deeper, training of the network has become a serious problem, where the large size of parameters and limited training data and computing power lead to issues such as local optimum, long training time, and gradient vanishing. These issues have been overcome by a collection

of large-sized datasets, the invention of more powerful GPU computing hardware, and newer network structure designs of ingenuity such as local connectivity, parameter sharing, and convolutional layers (Fig. 5).

Consider a multilayer feedforward neural network, as shown in Fig. 6; let L_l denote all the neurons in the l -th layer, and $x_j^{(l)}$ represent the input of the j -th node in layer l . $u_j^{(l)}$ is the net input, and the output is $y_j^{(l)}$. For the convenience of representation, the input, net input, and output are written in vector form, which are $\mathbf{x}^{(l)}$, $\mathbf{u}^{(l)}$, and $\mathbf{y}^{(l)}$, respectively. Assume that the weight from the i node of the $(l-1)$ -th layer to the j -th node of the l -th layer is $w_{ji}^{(l)}$, and the corresponding matrix form is $\mathbf{W}^{(l)}$; the bias of the j -th node in the l -th layer is $b_j^{(l)}$, and all biases in the l -th layer are $\mathbf{b}^{(l)}$.

Then, the input-output relationship of any node in the network is as follows:

$$\begin{cases} y_j^{(l)} = \sigma(u_j^{(l)}) \\ u_j^{(l)} = \sum_{i \in L_{l-1}} w_{ji}^{(l)} x_i^{(l-1)} + b_j^{(l)} = \sum_{i \in L_{l-1}} y_i^{(l-1)} + b_j^{(l)} \end{cases} \quad (1)$$

Expressed in the matrix form, the input-output relationship of any layer in the network can be written as

$$\mathbf{y}^{(l)} = \sigma(\mathbf{u}^{(l)}) = \sigma(\mathbf{W}^{(l)} \mathbf{y}^{(l-1)} + \mathbf{b}^{(l)}) \quad (2)$$

Followingly, by defining a loss function $()$ and training the network using sample data with known labels, the weights \mathbf{W} and \mathbf{b} can be obtained using an optimization method such as gradient descent.

$$\mathbf{W}^*, \mathbf{b}^* = \arg \min_{\mathbf{W}, \mathbf{b}} L_{(\mathbf{W}, \mathbf{b})}(\mathbf{y}, \hat{\mathbf{y}}) \quad (3)$$

where \mathbf{y} and $\hat{\mathbf{y}}$ are the predictive values and the actual label values.

4.2 Branch Recognition Methods for Bronchoscopic Quality Control

Quality control and management are crucial for bronchoscopy which aims to continuously improve the quality of endoscopic diagnosis and treatment through a series of measures. Through quality control, operational risks can be reduced, diagnostic and treatment accuracy can be improved, and safer and more efficient medical services can be provided to patients. Measures of bronchoscopic quality control include but are not limited to standardizing operating procedures, personnel training, and establishing strict quality inspection standards.

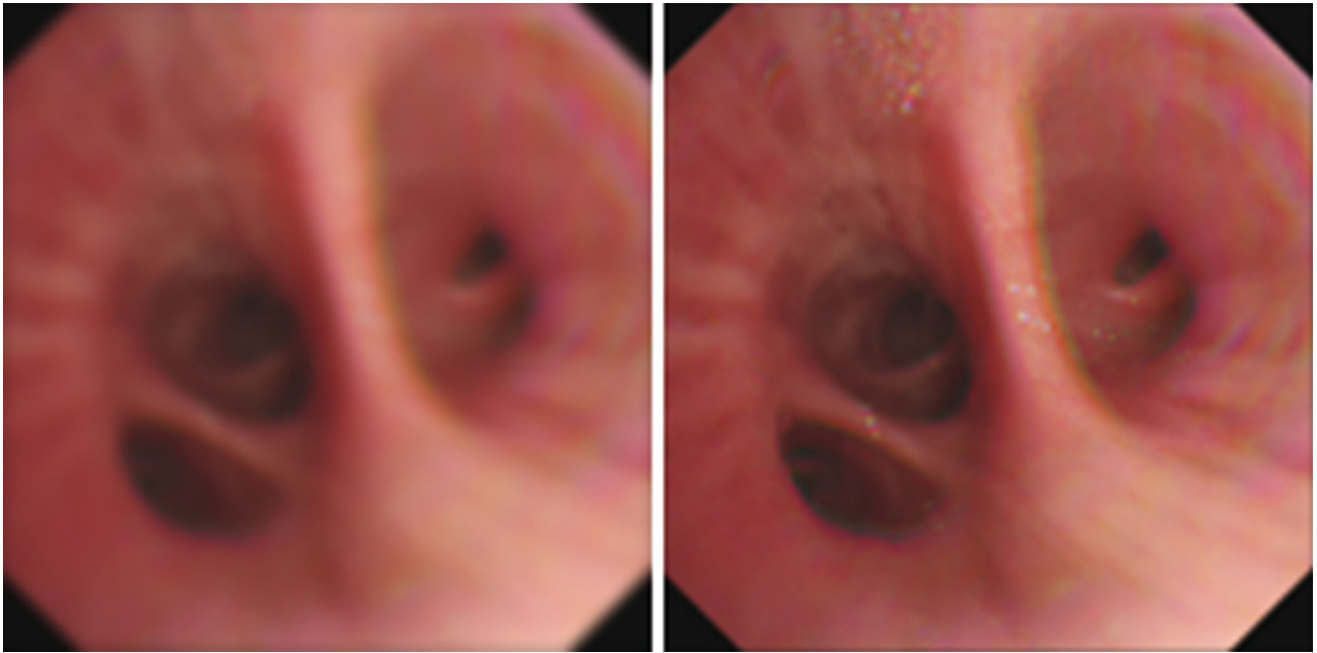


Fig. 4 Comparison between the original blurry bronchoscopic image (left) and its deblurred result by DeblurGAN (right)

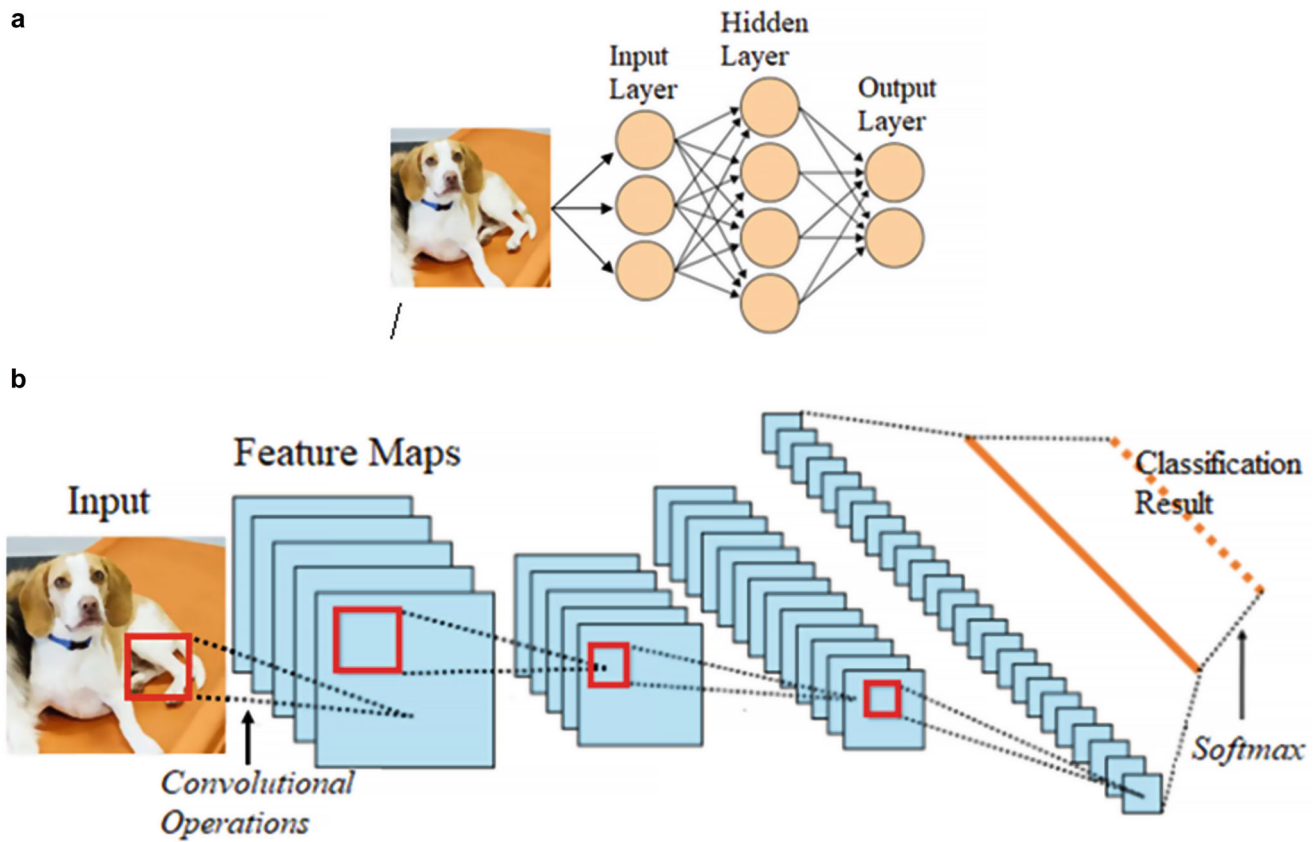


Fig. 5 Comparison between the structures of non-deep neural network and deep neural network. (a) traditional non-deep neural network. (b) deep neural network

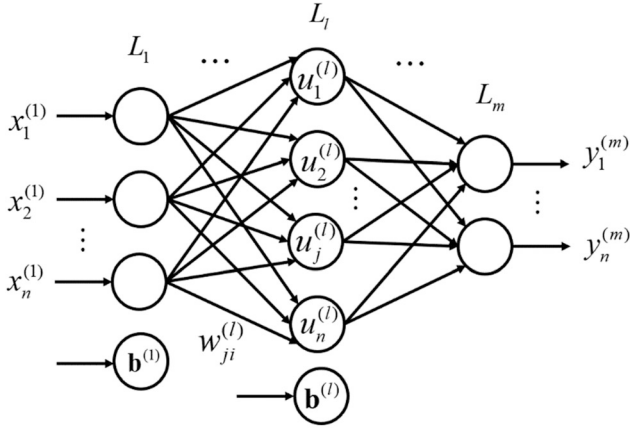


Fig. 6 The structure of the feedforward neural network

However, how to ensure and implement quality control has long been a trick problem since a strict implementation of quality control requires the supervision of the operating bronchoscopists by other bronchoscopic experts in charge. Such a practice consumes a large amount of manpower and makes things even worse for hospitals that already have limited medical resources. Moreover, bronchoscopists are prone to fatigue and errors during prolonged work hours, making supervision of the examination process unreliable. Therefore, the automation and intelligentization of bronchoscopic quality control is a topic of research which has attracted more and more attention from academia and industry.

With the rapid advancement of artificial intelligence technology in recent years—especially driven by deep learning—the goal of automatic and intelligent quality control for bronchoscopic examination is no longer imaginative and out of reach. This book chapter discusses several deep learning models for bronchoscopic quality control and presents the algorithm's performance.

In essence, bronchoscopic quality control is an object classification problem, in which the computer needs to determine which bronchial site/part the current bronchoscopic image covers. In other words, it is a branch recognition problem. For a bronchoscopic quality control AI system that considers three levels of the bronchial tree, the quality control problem is a 31-class classification problem (see Fig. 7). It should be noted that more bronchial levels may also be considered, but currently, a level higher than the third level is hard to implement, since not only is such a bronchoscope very expensive but also it is hard for a bronchoscopist to operate.

Traditionally, object classification algorithms rely on manually designed operators to extract object features such as contours, textures, and colors and then use classifiers (e.g., SVM, random forest, AdaBoost) for subsequent

classification. Since 2014, object classification models based on deep learning (i.e., deep neural networks) have emerged and replaced these traditional algorithms due to their convenience and high performance. These deep neural network models do not rely on the manual designing of an explicit feature extractor and are usually more robust in performance. Some of the most successful and representative deep learning models for object classification are VGG, GoogLeNet, and ResNet.

Proposed by the Visual Geometry Group at the University of Oxford in 2014, VGG is one of the most fundamental and popular CNN models for object classification. The work of VGG proved that increasing the network depth has a positive effect (to a certain extent) on the network performance. Compared with a preceding network model, AlexNet, VGG uses stacked small convolution kernels of 3×3 rather than the larger ones in AlexNet (e.g., 11×11 , 7×7 , and 5×5). It is proven that the superposition of two 3×3 convolution kernels with a stride of 1 has a receptive field equivalent to that of a 5×5 convolution kernel, which means VGG can achieve a higher level of network nonlinearity (as is beneficial for extraction of more complex features) with a relatively smaller number of parameters. The structure of the VGG model is shown in Fig. 8.

Unlike VGG blocks that only use fixed-size kernels of 3×3 , GoogLeNet combines convolution kernels of different sizes to increase the network's adaptability of multi-scales. Such a design of the multi-scale block is named Inception (see Fig. 9). The entire GoogLeNet model consists of nine Inception blocks, a max pooling layer between blocks, and finally a global average pooling layer and a fully connected layer. Under the same computational load, features that are more robust can be extracted by GoogLeNet, thereby improving the classification performance.

ResNet is also another representative classification model proposed in 2015. As shown in Fig. 10, the core idea of ResNet is to involve the residual blocks, which use identity mapping in every two convolutional layers to reduce a series of problems caused by deep networks, such as gradient vanishing. The proposal of residual networks enables deeper networks to be trained.

To verify the performance of the aforementioned three deep learning models, a dataset containing the training data and testing data as well as the corresponding labels of bronchoscopic images needs to be provided. One dataset available via inquiry is the dataset collected by Shanghai Chest Hospital. In total, 12 videos of complete bronchoscopy operations were collected, all covering all 31 bronchial locations as mentioned previously. Each video was segmented into image frames. A typical image was selected for each location of each video, and thus, a dataset of 372 images was formed for clinical evaluation. The labels are generated by four senior expert physicians with greater than 5 years of experience in

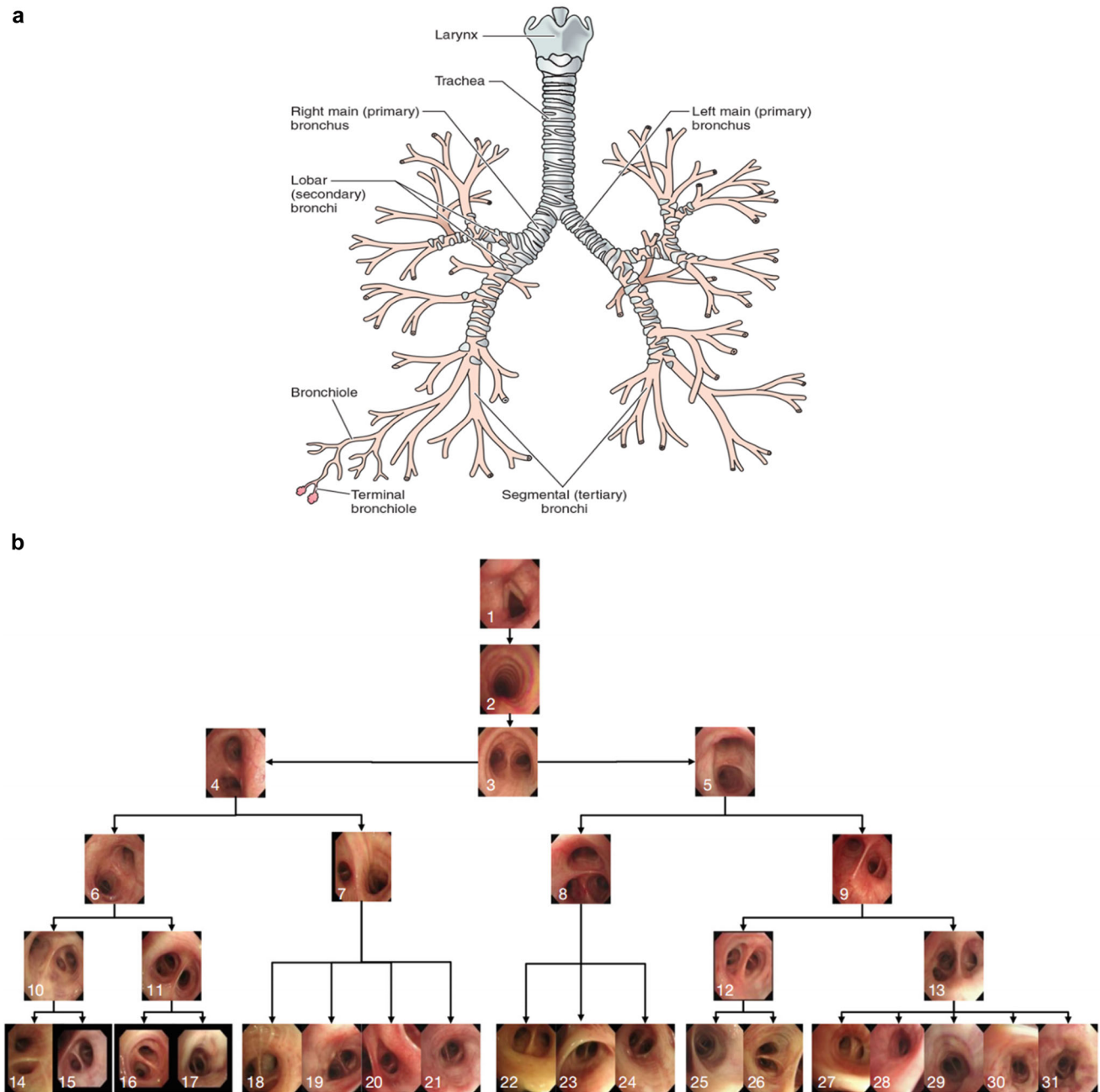


Fig. 7 Illustration of the bronchi and the 31-class classification problem. (a) bronchial anatomy. (b) the 31-class classification problem

interventional pulmonology and four junior physicians with less than 1 year of experience in interventional pulmonology.

The experimental results of the three DNN models, i.e., VGG, GoogLeNet, and ResNet, are shown in Table 1, where the classification performance in terms of accuracy (ACC) for the three deep learning models are compared. The results show that AI-assisted is indeed possible, considering that a considerable portion of bronchoscopists have a recognition accuracy lower than 80% with naked eyes.

4.3 Lesion Detection Methods for Automatic Bronchoscopic Diagnosis

During bronchial examination, the detection of lesions such as a tumor, tuberculosis, and inflammation is crucial to the diagnosis of respiratory diseases. However, lesion detection by manpower, i.e., bronchoscopists, is a task that is not only tedious and exhausting but also subjective which inevitably leads to diagnostic errors. Introducing AI to lesion detection

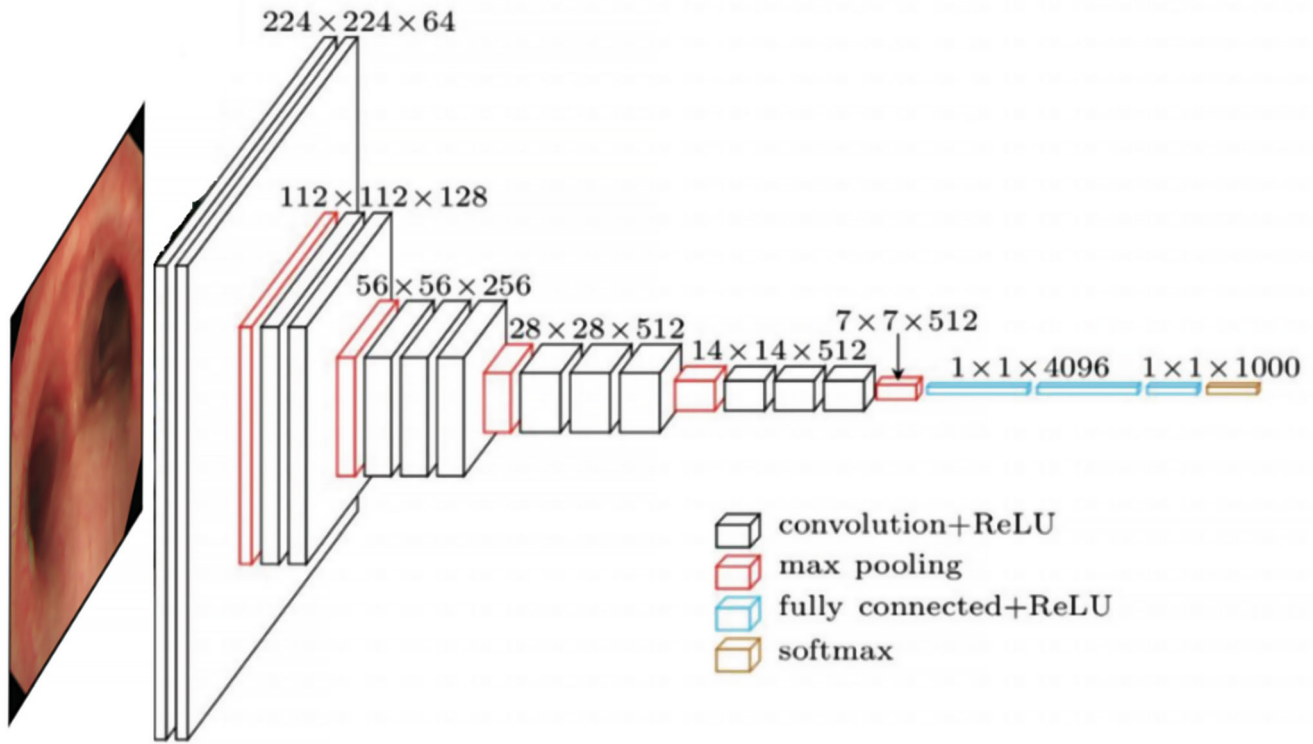
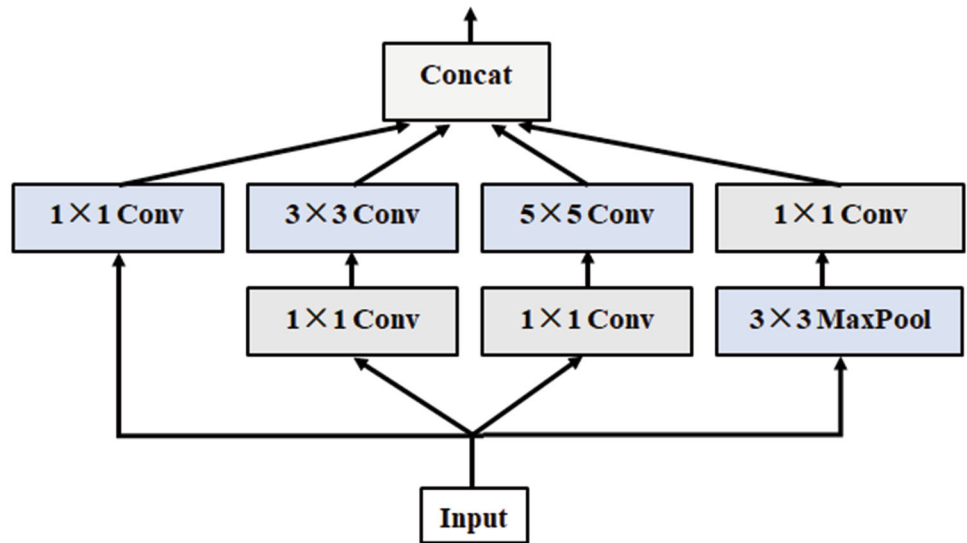


Fig. 8 The structure of the VGG model

Fig. 9 The building block of Inception within the GoogLeNet model



during the bronchial examination is a good alternative or supplementary solution.

Lesion detection is essentially an object detection problem which is a core topic in the field of computer vision. It aims to enable computers to recognize and locate objects in images. Specifically, it not only needs to identify which objects are in the image but also determine their positions in the image, i.e.,

the outputs of lesion detection should be represented in the form of bounding boxes with category labels.

The core idea behind object detection (and lesion detection) is to parse visual information that can be understood by the computer. The issues that object detection needs to address include (but are not limited to) object classification, position estimation, size changes, occlusion, background

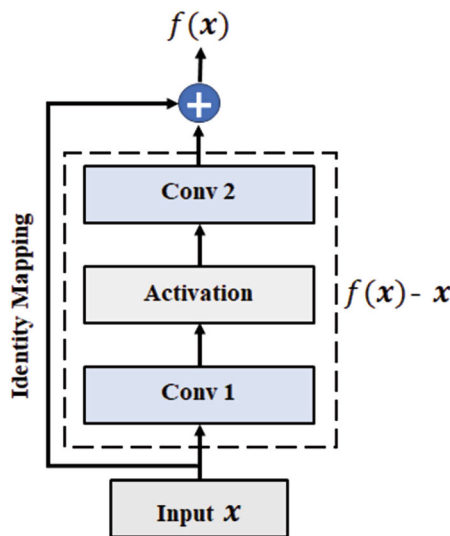


Fig. 10 The residual block of the ResNet model

Table 1 Classification performance in terms of accuracy (ACC) for the DNN models, VGG, GoogLeNet, and ResNet

AI model	Accuracy (%)
VGG	87.22
GoogLeNet	70.34
ResNet	88.16

interference, and real-time processing capabilities. This involves techniques such as image processing, feature extraction, pattern recognition, and machine learning.

At present, deep learning-based methods are the mainstream solutions in the field of object detection, which are usually divided into two categories: (A) Two-stage detection methods: This category of methods first extracts candidate object regions from the image and then performs detailed classification and bounding box fine-tuning on these regions. Representative models include the RCNN series (e.g., Fast R-CNN, Faster R-CNN) and region-based fully convolutional networks (such as Mask R-CNN). (B) Single-stage detection method: This type of method directly predicts the class and position of objects on the image. Single-stage methods are usually faster but may be slightly less accurate than two-stage methods. Representative models include SSD, RetinaNet, and YOLO. Most recently, transformer-based models have also been introduced for object detection, such as the DETR (Detection Transformer) model, which marks a new development direction for object detection technology.

This book chapter demonstrates how to utilize the YOLO model (more specifically, YOLO X) for the task of lesion detection. The overall network structure of YOLO is illustrated in Fig. 11. In general, YOLOX consists of four main modules: the input, the backbone, the neck, and the predictor. The input image is a color image in the R, G, and B channels,

with a fixed resolution of 416×416 . The backbone employs the module of CSP-DarkNet, which is an improved version of DarkNet and is inspired by CSPNet.

More specifically, the backbone module is built using four basic types of units: CBS, Focus, CSP, and SPP. The CBS unit is a cascaded connection of convolution, batch normalization, and SiLU activation. The SPP unit is nothing but a CBS unit processed by the spatial pyramid pooling (SPP) technique and then followed by another CBS unit. The CSP unit comprises paralleled CBS units of different numbers that are concatenated and then followed by an additional CBS unit. Lastly, the Focus unit first samples the feature maps and is then followed by another CBS unit.

The neck of YOLO is built using the feature pyramid network (FPN) and the path aggregation network (PAN) techniques. Such a form of neck can fuse features and information at different scales. During prediction, the advanced technique of “Decoupled Head” is used which separates classification and localization into paralleled branches, yielding a higher detection performance as well as faster convergence of the model. It should be noted that an IoU branch is also added to the regression branch.

As for data augmentation, strategies such as Mosaic and MixUp can be utilized to boost the model’s performance. The anchor-free technique is also involved by reducing the predictions for each location from 3 to 1 and making them directly predict four values, i.e., the height and width and the two offsets of the predicted box. Finally, the technique of multi-positives is also utilized which assigns a central area of 3×3 (rather than the center alone) as positives for each object. The multi-positive technique alleviates the issue of positive/negative sample imbalance during training. A simplified version of the advanced label assignment technique of OTA, i.e., SimOTA, is used which treats the assigning procedure as an optimal transport problem and takes a dynamic top-k strategy to obtain an approximate solution.

After the design of the detection model, the loss function needs to be specially designed accordingly. The loss function of YOLOX includes three parts: the category prediction loss (*cls*), the bounding box regression loss (*reg*), and the target existence probability loss (*obj*). Firstly, the category prediction loss calculates the difference between the predicted category and the actual category. It uses binary cross entropy loss (BCE with logits loss) to measure the predictive ability of the model for the target category. Secondly, the boundary box regression loss calculates the difference in the Intersection over Union (IoU) between the predicted and the actual boundary boxes. Lastly, the target existence probability loss calculates the difference between the predicted target existence probability and the true existence probability of the model. Similar to the case of category prediction loss, binary cross entropy loss is also used, ensuring that the model can accurately determine whether the target exists in the given image region.

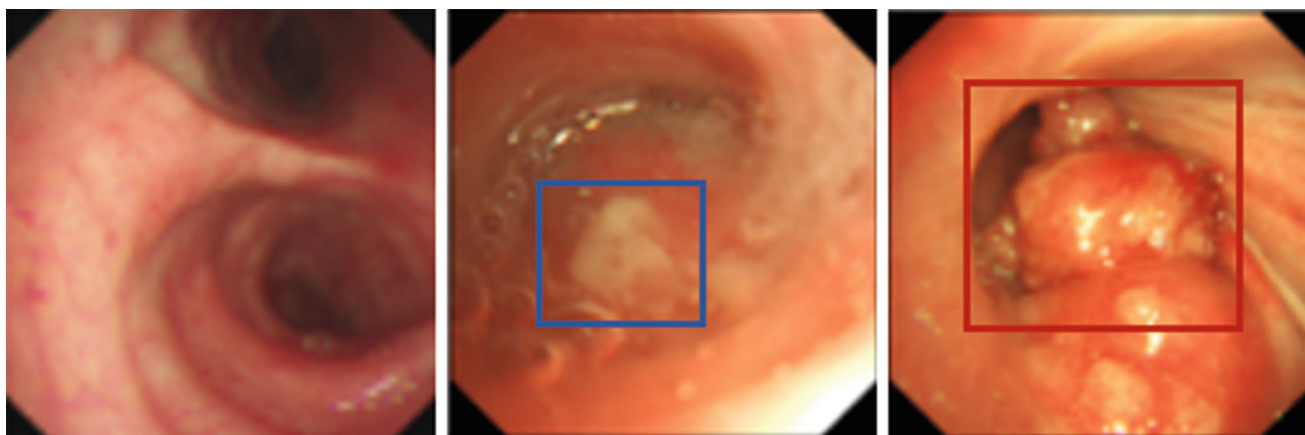


Fig. 13 Lesion detection for automatic bronchoscopic diagnosis. From left to right: normal bronchus (no detection box), bronchus with a benign lesion (blue box), and bronchus with a malignant lesion (red box)

diagnostic technology of bronchoscopy and nasopharyngolaryngoscopy can not only greatly improve the work efficiency and accuracy of respiratory and NPL specialists, but also reduce the occurrence of misdiagnosis and medical malpractice, and ultimately provide patients with safer and more effective medical services. It can be foreseen that with continuous development and improvement of AI technology, AI-assisted bronchoscopy and nasopharyngolaryngoscopy will play an increasingly important role in the diagnosis and treatment of respiratory system diseases and otolaryngological diseases.

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

1. Ali S. Where do we stand in AI for endoscopic image analysis? Deciphering gaps and future directions [J]. *NPJ Digit Med*. 2022;5(1):184.
2. Yoo JY, Kang SY, Park JS, Cho YJ, Park SY, Yoon HI, Park SJ, Jeong HG, Kim T. Deep learning for anatomical interpretation of video bronchoscopy images. *Sci Rep*. 2021;11(1):23765.
3. Li Y, Zheng X, Xie F, Ye L, Bignami E, Tandon YK, Rodríguez M, Gu Y, Sun J. Development and validation of the artificial intelligence (AI)-based diagnostic model for bronchial lumen identification. *Transl Lung Cancer Res*. 2022;11(11):2261–74.
4. Chen C, Herth FJ, Zuo Y, Li H, Liang X, Chen Y, Ren J, Jian W, Zhong C, Li S. Distinguishing bronchoscopically observed anatomical positions of airway under by convolutional neural network. *Ther Adv Chronic Dis*. 2023;14:20406223231181495.
5. Matava C, Pankiv E, Raisbeck S, Caldeira M, Alam F. A convolutional neural network for real time classification, identification, and labelling of vocal cord and tracheal using laryngoscopy and bronchoscopy video. *J Med Syst*. 2020;44(2):44.
6. Cold KM, Xie S, Nielsen AO, Clementsen PF, Konge L. Artificial intelligence improves novices' bronchoscopy performance: a randomized controlled trial in a simulated setting. *Chest*. 2024;165(2):405–13.
7. Li Y, Gu W, Yue H, Lei G, Guo W, Wen Y, Tang H, Luo X, Tu W, Ye J, Hong R, Cai Q, Gu Q, Liu T, Miao B, Wang R, Ren J, Lei W. Real-time detection of laryngopharyngeal cancer using an artificial intelligence-assisted system with multimodal data. *J Transl Med*. 2023;21(1):698.
8. Xiong H, Lin P, Yu JG, Ye J, Xiao L, Tao Y, Jiang Z, Lin W, Liu M, Xu J, Hu W, Lu Y, Liu H, Li Y, Zheng Y, Yang H. Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images. *EBioMedicine*. 2019;48:92–9.
9. Mekov E, Miravittles M, Petkov R. Artificial intelligence and machine learning in respiratory medicine. *Expert Rev Respir Med*. 2020;14(6):559–64.
10. Kaplan A, Cao H, FitzGerald JM, Iannotti N, Yang E, Kocks JWH, Kostikas K, Price D, Reddel HK, Tsiligianni I, Vogelmeier CF, Pfister P, Mastoridis P. Artificial intelligence/machine learning in respiratory medicine and potential role in asthma and COPD diagnosis. *J Allergy Clin Immunol Pract*. 2021;9(6):2255–61.
11. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25(1):30–6.
12. Thiébaud R, Thiessard F, Section Editors for the IMIA Yearbook Section on Public Health and Epidemiology Informatics. Artificial intelligence in public health and epidemiology. *Yearb Med Inform*. 2018;27(1):207–10.
13. Antão J, de Mast J, Marques A, Franssen FME, Spruit MA, Deng Q. Demystification of artificial intelligence for respiratory clinicians managing patients with obstructive lung diseases. *Expert Rev Respir Med*. 2023;17(12):1207–19.
14. Weimin D. Diagnosis and treatment guidelines for tracheal and bronchial tuberculosis (trial) [J]. *Chin J Tuberc Respir Dis*. 2012;35(08):581–7.
15. China IoT Expert Group for Evaluation and Management of Pulmonary Nodules. Internet of Things (IoT)-assisted evaluation and management of lung nodules in China: expert consensus [J]. *Int J Respir*. 2022;42(1):5–12.
16. Dawei Y, Lin T, Powell Charles A, et al. Difficult to classify pulmonary nodules [J]. *Int J Respir*. 2022;42(1):1–4.
17. Chinese Medical Association of Respiratory Disease Branch. ACCP. Chinese expert consensus on the diagnosis and treatment of pulmonary nodules. *Chin J Tuberc Respir Dis*. 2024;47(8):716–29.
18. Verikas A, Uloza V, Bacauskiene M, Gelzinis A, Kelertas E. Advances in laryngeal imaging. *Eur Arch Otorrinolaringol*. 2009;266(10):1509–20.
19. Schade G, Fleischer S, Müller F, Kutta H, Hess M. Influence of the head position in laryngoscopy. *HNO*. 2004;52(10):921–6.
20. Ni XG, Cheng RR, Lai SQ, et al. Novel laryngoscopic strategies to improve evaluation of the site and extent of primary hypopharyngeal tumours. *J Laryngol Otol*. 2013;127(9):882–9.
21. Emura F, Sharma P, Arantes V, et al. Principles and practice to facilitate complete photodocumentation of the upper gastrointestinal

- tract: World Endoscopy Organization position statement. *Dig Endosc.* 2020;32(2):168–79.
22. Takiyama H, Ozawa T, Ishihara S, et al. Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks. *Sci Rep.* 2018;8(1):7497.
 23. Wu L, Zhang J, Zhou W, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. *Gut.* 2019;68(12):2161–9. <https://doi.org/10.1136/gutjnl-2018-317366>.
 24. Chang YY, Yen HH, Li PC, et al. Upper endoscopy photo-documentation quality evaluation with novel deep learning system. *Dig Endosc.* 2022;34(5):994–1001.
 25. Yuan P, Bai R, Yan Y, et al. Subjective and objective quality assessment of gastrointestinal endoscopy images: from manual operation to artificial intelligence. *Front Neurosci.* 2022;16:1118087.
 26. Zhou J, Wu L, Wan X, et al. A novel artificial intelligence system for the assessment of bowel preparation (with video). *Gastrointest Endosc.* 2020;91(2):428–435.e2.
 27. Shaukat A, Rector TS, Church TR, et al. Longer withdrawal time is associated with a reduced incidence of interval cancer after screening colonoscopy. *Gastroenterology.* 2015;149(4):952–7.
 28. Gong D, Wu L, Zhang J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol Hepatol.* 2020;5(4):352–61.
 29. Zhu JQ, Wang ML, Li Y, et al. Convolutional neural network based anatomical site identification for laryngoscopy quality control: a multicenter study. *Am J Otolaryngol.* 2023;44(2):103695.
 30. Nakajo K, Ninomiya Y, Kondo H, et al. Anatomical classification of pharyngeal and laryngeal endoscopic images using artificial intelligence. *Head Neck.* 2023;45(6):1549–57.
 31. Ren J, Jing X, Wang J, et al. Automatic recognition of laryngoscopic images using a deep-learning technique. *Laryngoscope.* 2020;130(11):E686–93.
 32. Kuo CFJ, Lai WS, Barman J, et al. Quantitative laryngoscopy with computer-aided diagnostic system for laryngeal lesions. *Sci Rep.* 2021;11(1):10147.
 33. Cho WK, Lee YJ, Joo HA, et al. Diagnostic accuracies of laryngeal diseases using a convolutional neural network-based image classification system. *Laryngoscope.* 2021;131(11):2558–66.
 34. Zhao Q, He Y, Wu Y, et al. Vocal cord lesions classification based on deep convolutional neural network and transfer learning. *Med Phys.* 2022;49(1):432–42.
 35. Chen IM, Yeh PY, Hsieh YC, et al. 3D VOSNet: segmentation of endoscopic images of the larynx with subsequent generation of indicators. *Heliyon.* 2023;9(3):e14242.
 36. Xiong H, Lin P, Yu JG, et al. Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images. *EBioMedicine.* 2019;48:92–9.
 37. Tamashiro A, Yoshio T, Ishiyama A, et al. Artificial intelligence-based detection of pharyngeal cancer using convolutional neural networks. *Dig Endosc.* 2020;32(7):1057–65.
 38. Kono M, Ishihara R, Kato Y, et al. Diagnosis of pharyngeal cancer on endoscopic video images by mask region-based convolutional neural network. *Dig Endosc.* 2021;33(4):569–76.
 39. Li Y, Gu W, Yue H, et al. Real-time detection of laryngopharyngeal cancer using an artificial intelligence-assisted system with multimodal data. *J Transl Med.* 2023;21(1):698.
 40. Sampieri C, Azam MA, Ioppi A, et al. Real-time laryngeal cancer boundaries delineation on white light and narrow-band imaging laryngoscopy with deep learning. *Laryngoscope.* Published online January 4, 2024;134:2826.
 41. Paderno A, Piazza C, Del Bon F, et al. Deep learning for automatic segmentation of oral and oropharyngeal cancer using narrow band imaging: preliminary experience in a clinical perspective. *Front Oncol.* 2021;11:626602.
 42. Girdler B, Moon H, Bae MR, Ryu SS, Bae J, Yu MS. Feasibility of a deep learning-based algorithm for automated detection and classification of nasal polyps and inverted papillomas on nasal endoscopic images. *Int Forum Allergy Rhinol.* 2021;11(12):1637–46.
 43. Ay B, Turker C, Emre E, Ay K, Aydin G. Automated classification of nasal polyps in endoscopy video-frames using handcrafted and CNN features. *Comput Biol Med.* 2022;147:105725.
 44. Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. *Lancet.* 2019;394(10192):64–80.
 45. Bray F, Laversanne M, Sung H, et al. Global Cancer Statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* Published online April 4, 2024.
 46. Li C, Jing B, Ke L, et al. Development and validation of an endoscopic images-based deep learning model for detection with nasopharyngeal malignancies. *Cancer Commun (Lond).* 2018;38(1):59.
 47. Shi Y, Wang H, Ji H, et al. A deep weakly semi-supervised framework for endoscopic lesion segmentation. *Med Image Anal.* 2023;90:102973.
 48. Wang S-X, Yi L, Zhu J-Q, et al. The detection of nasopharyngeal carcinomas using a neural network based on nasopharyngoscopic images. *Laryngoscope.* 2024;134(1):127.
 49. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks [J]. *Adv Neural Inf Proces Syst.* 2012;25:1097–105.
 50. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. *arXiv preprint arXiv:1409.1556.* 2014.
 51. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016. p. 770–8.
 52. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2015. p. 1–9.
 53. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features [C]. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1; IEEE, 2001.* p. 1–11.
 54. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2014. p. 580–7.
 55. Ren S. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *arXiv preprint arXiv:1506.01497.* 2015.
 56. He K, Gkioxari G, Dollár P, et al. Mask R-CNN [C]. In: *Proceedings of the IEEE International Conference on Computer Vision*; 2017. p. 2961–9.
 57. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C]. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016. p. 779–88.
 58. Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector [C]. In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, vol. 2016.* Springer International Publishing. p. 21–37.
 59. Ge Z. Yolox: exceeding yolo series in 2021 [J]. *arXiv preprint arXiv:2107.08430.* 2021.
 60. Yoo JY, Kang SY, Park JS, et al. Deep learning for anatomical interpretation of video bronchoscopy images[J]. *Sci Rep.* 2021;11(1):23765.
 61. Li Y, Zheng X, Xie F, et al. Development and validation of the artificial intelligence (AI)-based diagnostic model for bronchial lumen identification [J]. *Transl Lung Cancer Res.* 2022;11(11):2261.
 62. Cold KM, Xie S, Nielsen AO, et al. Artificial intelligence improves novices' bronchoscopy performance: a randomized controlled trial in a simulated setting[J]. *Chest.* 2024;165(2):405–13.
 63. Chen C, Herth FJF, Zuo Y, et al. Distinguishing bronchoscopically observed anatomical positions of airway under by convolutional neural network [J]. *Ther Adv Chronic Dis.* 2023;14:20406223231181495.