

Automated Detection of Laryngeal Carcinoma in Laryngoscopic Images from a Multicenter Database using a Convolutional Neural Network

Peikai Yan¹, Shaohua Li², Zhou Zhou³, Qian Liu⁴, Jiahui Wu¹, Qingyi Ren¹, Qiuhan Chen⁵, Zhipeng Chen⁶, Ze Chen⁷, Shao-hua Chen⁸, Austin Scholp⁹, Jack J Jiang⁹, Jing Kang¹, and Pingjiang Ge¹⁰

¹Guangdong Provincial People's Hospital

²Zhongshan Hospital of Traditional Chinese Medicine

³Shenzhen People's Hospital

⁴Shenzhen People's Hospital

⁵Zhaoqing Gaoyao People's Hospital

⁶The Second People's Hospital of Longgang District

⁷Gaozhou People's Hospital

⁸Guangdong general Hospital & Guangdong Academy of Medical Science

⁹University of Wisconsin-Madison School of Medicine and Public Health

¹⁰Guangdong General Hospital

September 28, 2021

Abstract

OBJECTIVE Little is known about the efficacy of using artificial intelligence to identify laryngeal carcinoma from images of vocal lesions taken in different hospitals with multiple laryngoscope systems. This multicenter study was aimed to establish an artificial intelligence system and provide a reliable auxiliary tool to screen for laryngeal carcinoma. **Study Design:** Multicentre case-control study **Setting:** Six tertiary care centers **Participants:** The laryngoscopy images were collected from 2179 patients with vocal lesions. **Outcome Measures:** An automatic detection system of laryngeal carcinoma was established based on Faster R-CNN, which was used to distinguish vocal malignant and benign lesions in 2179 laryngoscopy images acquired from 6 hospitals with 5 types of laryngoscopy systems. Pathology was the gold standard to identify malignant and benign vocal lesions. **Results:** Among 89 cases of the malignant group, the classifier was able to evaluate the laryngeal carcinoma in 66 patients (74.16%, sensitivity), while the classifier was able to assess the benign laryngeal lesion in 503 cases among 640 cases of the benign group (78.59%, specificity). Furthermore, the CNN-based classifier achieved an overall accuracy of 78.05% with a 95.63% negative prediction for the testing dataset. **Conclusion:** This automatic diagnostic system has the potential to assist clinical laryngeal carcinoma diagnosis, which may improve and standardize the diagnostic capacity of endoscopists using different laryngoscopes.

Automated Detection of Laryngeal Carcinoma in Laryngoscopic Images from a Multicenter Database using a Convolutional Neural Network

ABSTRACT

OBJECTIVE Little is known about the efficacy of using artificial intelligence to identify laryngeal carcinoma from images of vocal lesions taken in different hospitals with multiple laryngoscope systems. This

multicenter study was aimed to establish an artificial intelligence system and provide a reliable auxiliary tool to screen for laryngeal carcinoma.

Study Design: Multicentre case-control study

Setting: Six tertiary care centers

Participants: The laryngoscopy images were collected from 2179 patients with vocal lesions.

Outcome Measures: An automatic detection system of laryngeal carcinoma was established based on Faster R-CNN, which was used to distinguish vocal malignant and benign lesions in 2179 laryngoscopy images acquired from 6 hospitals with 5 types of laryngoscopy systems. Pathology was the gold standard to identify malignant and benign vocal lesions.

Results: Among 89 cases of the malignant group, the classifier was able to evaluate the laryngeal carcinoma in 66 patients (74.16%, sensitivity), while the classifier was able to assess the benign laryngeal lesion in 503 cases among 640 cases of the benign group (78.59%, specificity). Furthermore, the R-CNN-based classifier achieved an overall accuracy of 78.05% with a 95.63% negative prediction for the testing dataset.

Conclusion: This automatic diagnostic system has the potential to assist clinical laryngeal carcinoma diagnosis, which may improve and standardize the diagnostic capacity of endoscopists using different laryngoscopes.

Keywords: Artificial intelligence, Vocal fold lesions, Laryngoscope, Laryngeal Carcinoma, multicenter study

Key points:

- Artificial intelligence (AI) has achieved outstanding performance within medical imaging interpretation and triage tasks and has been successfully used to diagnose carcinoma.
- Higher sensitivity was achieved for the malignancy by autonomous classification of endoscopic images with artificial intelligence technology in a single center.
- It is still unknown whether different laryngoscopic systems would impact automatic recognition accuracy from multicenter clinics.
- An automatic classifier for laryngeal carcinoma based on Faster R-CNNs(region-based convolutional neural networks) was established.
- High accuracy was also obtained for multiple clinics and multiple laryngoscopic systems.

Introduction

Laryngeal carcinoma is one of the most common malignant tumors of the head and neck with incidence estimated to be more than 24,500 cases per year by 2030.¹ Survival outcomes in laryngeal cancer are affected by several aspects, including tumor stage,² subsites,³ age,^{2,4} treatment modality,^{2,5} comorbidities³, etc. An early-stage diagnosis is one of the most crucial factors to decrease the mortality rate and preserve both laryngeal anatomy and vocal function. The 5-year survival rates are 100% and 80% for patients with stage 0 and stage I laryngeal carcinoma, respectively, while it decreases to only 70% for patients with advanced-stage cancer.⁶

Currently, the optic laryngoscope is a routine method to diagnose laryngeal cancer, as well as identify the extent of invasion and provide accurate clinical staging. However, physicians still have difficulty distinguishing early-stage cancer from mucosal abnormalities.⁷ Thus, misdiagnoses and missed diagnoses are not rare while only using a laryngoscope. High levels of diagnostic inconsistency have been observed, even among experts.⁸

Today, artificial intelligence (AI) has achieved outstanding performance within medical imaging interpretation and triage tasks and has been successfully used to diagnose skin cancer,⁹ lung cancer,¹⁰ glioma,¹¹ and breast histopathology.¹² As a popular technique of deep-learning algorithms, region-based convolutional neural networks (R-CNNs) proposed an efficient method in object detection that utilizes a feature map from a convolutional neural network.¹³ Several recent approaches, including Fast R-CNN, Faster R-CNN, and Mask

R-CNN, were developed based on R-CNNs.¹⁴⁻¹⁶ In particular, Faster R-CNN is one of the first end-to-end two-stage detectors that has displayed remarkable efficiency and accuracy.¹⁷

Recently, two studies regarding the automatic recognition of laryngoscopy images based on the convolutional neural network have achieved promising results. In one previous study, the binary classifier distinguished benign and malignant-premalignant lesions with an overall accuracy of 93.0%.¹⁸ A sensitivity of 89% and a specificity of 99.33% were achieved for the malignancy by autonomous classification of endoscopic images with artificial intelligence technology.¹⁹ However, the studies mentioned above were both carried out in a single center. It is still unknown whether the source of laryngoscopic images, the resolution of the laryngoscopy images, and the different endoscopic systems would impact automatic recognition accuracy. Therefore, a multicenter clinical trial is essential to determine whether the AI technique could cope with complex situations in the real world.

In this study, a multicenter experiment of laryngeal carcinoma detection was carried out based on an autonomous endoscopic image classifier using the Faster R-CNN system. Our research established an artificial intelligence system for the detection of laryngeal carcinoma and evaluated the performance of this system, aiming to provide a reliable auxiliary tool to diagnose early-stage laryngeal carcinoma efficiently and help untrained technicians accomplish objective and accurate screening.

Methods

2.1 Data Collection and Study Population

All subjects were patients that underwent laryngoscopy prior to surgery and were diagnosed with malignant or benign laryngeal lesions via pathological diagnosis. All duplicate images, unclear images with low resolution, and images of unexposed vocal folds were excluded from this study. Twenty-six images were excluded from this study.

A total of 2179 laryngoscopy images, including 288 images of malignant lesions and 1891 images of benign lesions from 2179 patients were qualified for inclusion. Each image only displayed a patient's glottic lesion or tumor. The benign lesion images were composed of 998 vocal polyps, 169 vocal cysts, 276 nodules, 356 vocal leukoplakia, 80 vocal process granuloma, and 12 vocal papillomae. Of the 288 malignant lesions, there were 22 cases of T3 tumors, 62 of T2 tumors, and 204 of T1 tumors. The images were retrospectively collected in 6 hospitals with 5 endoscopic systems including First hospital (n=783, XION EndoSTROB D and Olympus ENF), Second hospital (n=622, Olympus ENF), Third hospital (n=531, Olympus VISER Pro and XION EndoSTROB D), Forth hospital (n=89, XION LCD), Fifth hospital (n=78, Olympus CV170), and Sixth hospital (n=76, Olympus ENF). All images were taken between January 1, 2015, and December 31, 2018. The related information is described in Table 1. All patients in this study underwent the white light laryngoscopic scan and were diagnosed by two experienced doctors. The images were distributed into malignant and benign groups by pathology.

2.2 Data Preparation

The dataset was randomly split into two parts: 1450 images for training and 729 images for testing. For training, the number of images in the malignant and benign groups was 199 and 1251, respectively. For the testing group, 89 were malignant and 640 were benign (Table 1). Two senior otolaryngologists used the labeling software Label Tool to interpret the laryngoscope images and label the exact region of the biopsy.

2.3 Outcome Measures

The automatic detection of laryngeal carcinoma was carried out using Faster R-CNN, which is one of the most popular two-stage detection networks used for many medical image detection situations. The architecture of Faster R-CNN generally consists of the region proposal network (RPN) and Fast R-CNN detector, as shown in Figure 1. RPN is a deep fully convolutional network using simultaneous object boundaries and object scores prediction at each detection for generating high-quality region proposals. It should be highlighted that RPNs are introduced to share convolutional layers with the Fast R-CNN detector to a more effective

detection network’s computation (Figure 1).¹⁵ Fast R-CNN includes region of interest (RoI) pooling, which produces the fixed-size feature maps from non-uniform inputs by doing max-pooling on the inputs, aiming to refine the proposals and process the sketch map.²⁰

Results

After the training procedure described above, the CNN-based classifier was used to detect laryngeal lesions. Comparisons were conducted of laryngoscopic images detected by the CNN-based classifier with that by the pathological results to determine the diagnostic potential of the classification system for the testing dataset.

Among 89 cases of the malignant group, the classifier was able to accurately evaluate the laryngeal carcinoma in 66 cases (74.16%, sensitivity). For the benign lesions, the classifier was able to accurately assess lesions in 503 cases of the 640 cases in the benign group (78.59%, specificity). Furthermore, the CNN-based classifier achieved an overall accuracy of 78.05% on the testing dataset. The results for the classifier are summarized in Table 2 and Table 3.

Discussion

For years, the laryngoscope has been recommended as an essential tool for assessing vocal fold lesions. However, laryngoscopes have limitations in making diagnoses owing to subtle differences in mucosal tissue. Distinguishing different mucosa can be difficult, especially for early-stage laryngeal carcinoma. Insufficient experience and excessive workload can also lead to missed diagnosis and misdiagnosis. Therefore, an assistive system for identifying laryngeal carcinoma is needed to improve and standardize clinicians’ diagnostic capacity.

The Faster R-CNN detection system has the potential to analyze the characteristics, extract the images’ commonalities, and then classify the data. In our study, a 78.05% accurate rate was achieved. Our results confirmed that the model we have established could offer valuable clinical assistance to preliminarily screen patients with laryngeal carcinoma and identify patients with benign lesions. This system would be able to help clinicians make more efficient and accurate diagnoses and enable standardized evaluation.

In this study, 2179 laryngoscope images taken were from different endoscopic systems between six hospitals. We did not impose special restrictions on equipment or images. Neither did we have restrictions on age or gender. Such processing maximized the natural diversity of different laryngoscopic appearances in various conditions to reflect whether AI technology could cope with complex situations in the real world.

Our results indicate that the automatic classifier has a sensitivity of 74.16%, a specificity of 78.59%, and 78.05% accuracy for detecting laryngeal carcinoma, which outperformed physicians in laryngeal carcinoma recognitions as demonstrated in several articles. In Ren’s study, twelve human experts only achieved an overall accuracy of 54% in malignancy detection in a set of 706 laryngoscopic testing images.¹⁹ Additionally, a recent meta-analysis showed that only 65% of the tested population was correctly identified as not having cancer by ENT doctors, even with videostroboscopy.⁸ Generally speaking, an automated system could help physicians make more confident determinations to exclude the malignant lesion, reducing the burden of endoscopists and patients’ waiting time, while it could raise red flags when a patient tests positive, reminding doctors to complete further clinical investigation to determine if the patient is truly positive.

However, two recent studies using the CNN-based technique obtained over 90% accuracy for malignancy lesions detection, which outperformed our classifier. A possible explanation for this might be that our laryngoscope images were from different endoscopic systems in several hospitals. The diverse laryngoscope equipment could attribute to complicated commonality extractions, leading to the limited quality of data classification. This situation would be a common problem for clinical application, which is an obstacle that needs to be overcome in future research. Additionally, increasing the training dataset size may be another approach to improving our classifier’s sensitivity and specificity.

Further research is necessary to include high-quality laryngoscope images with NBI from multiple centers. Since the sensitivity of NBI has been found to be superior to white light endoscopy,²¹ the CNN-based classifier

may provide more efficient and accurate information with NBI.

Conclusion

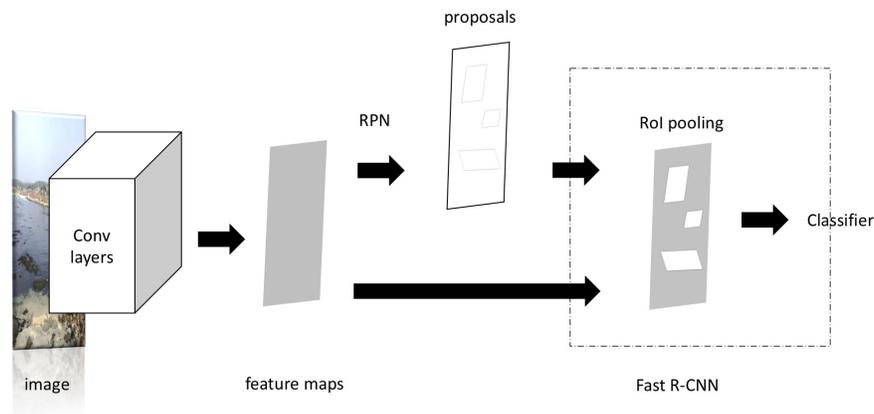
The fiber-optic laryngoscope has been a routine method to diagnose laryngeal lesions; however, physicians still have difficulty distinguishing early-stage cancer. In this study, an automatic classifier for laryngeal carcinoma based on Faster R-CNN was established for multiple clinics and multiple laryngoscopic systems. An accuracy of 78.05% was obtained. This diagnostic system could offer promising applicability to daily medical practice, which may improve clinical diagnosis quality and benefit patients.

REFERENCES

1. Gupta B, Johnson NW, Kumar N. Global Epidemiology of Head and Neck Cancers: A Continuing Challenge. *Oncology* 2016; 91:13-23.
2. Nguyen-Tan PF, Le QT, Quivey JM, et al. Treatment results and prognostic factors of advanced T3–4 laryngeal carcinoma: the University of California, San Francisco (UCSF) and Stanford University Hospital (SUH) experience. *Int J Radiat Oncol Biol Phys* 2001; 50:1172-1180.
3. Ramroth H, Schoeps A, Rudolph E et al. Factors predicting survival after diagnosis of laryngeal cancer. *Oral Oncol* 2011; 47:1154-1158.
4. Choi SH, Terrell JE, Fowler KE et al. Socioeconomic and Other Demographic Disparities Predicting Survival among Head and Neck Cancer Patients. *PloS one* 2016; 11:e0149886.
5. Harris BN, Bhuskute AA, Rao S, Farwell DG, Bewley AF. Primary surgery for advanced-stage laryngeal cancer: A stage and subsite-specific survival analysis. *Head & neck* 2016; 38:1380-1386.
6. Yang Y, Liu J, Song F, Zhang S. The clinical diagnostic value of target biopsy using narrow-band imaging endoscopy and accurate laryngeal carcinoma pathologic specimen acquisition. *Clinical otolaryngology* 2017; 42:38-45.
7. Stachler RJ, Francis DO, Schwartz SR, et al. Clinical Practice Guideline: Hoarseness (Dysphonia) (Update). *Otolaryngol Head Neck Surg* 2018; 158:S1-s42.
8. Mehlum CS, Rosenberg T, Groentved AM, Dyrvig AK, Godballe C. Can videostroboscopy predict early glottic cancer? A systematic review and meta-analysis. *Laryngoscope* 2016; 126:2079-2084.
9. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542:115-118.
10. van Ginneken B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiol Phys Technol* 2017; 10:23-32.
11. Ertosun MG, Rubin DL. Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of convolutional neural networks. *AMIA Annu Symp Proc* 2015; 2015:1899-1908.
12. Bejnordi BE, Zuidhof G, Balkenhol Met al. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. *J Med Imaging (Bellingham)* 2017; 4:044504.
13. Roh M-C, Lee J-y. Refining faster-RCNN for accurate object detection *2017 fifteenth IAPR international conference on machine vision applications (MVA)* : IEEE, 2017:514-517.
14. Girshick R. Fast r-cnn *Proceedings of the IEEE international conference on computer vision* , 2015:1440-1448.
15. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks *Advances in neural information processing systems* , 2015:91-99.

16. He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell* 2020; 42:386-397.
17. Cheng B, Wei Y, Shi H, Feris R, Xiong J, Huang T. Revisiting rcnn: On awakening the classification power of faster rcnn *Proceedings of the European conference on computer vision (ECCV)* , 2018:453-468.
18. Dunham ME, Kong KA, McWhorter AJ, Adkins LK. Optical Biopsy: Automated Classification of Airway Endoscopic Findings Using a Convolutional Neural Network. *Laryngoscope* 2020.
19. Ren J, Jing X, Wang Jet al. Automatic Recognition of Laryngoscopic Images Using a Deep-Learning Technique. *Laryngoscope* 2020.
20. Ma S, Huang Y, Che X, Gu R. Faster RCNN-based detection of cervical spinal cord injury and disc degeneration. *J Appl Clin Med Phys* 2020; 21:235-243.
21. Sun C, Han X, Li X, Zhang Y, Du X. Diagnostic Performance of Narrow Band Imaging for Laryngeal Cancer: A Systematic Review and Meta-analysis. *Otolaryngol Head Neck Surg* 2017; 156:589-597.

Figure 1 Schematic diagram of Fast R-CNN detector



Hosted file

TABLES1.docx available at <https://authorea.com/users/438359/articles/539581-automated-detection-of-laryngeal-carcinoma-in-laryngoscopic-images-from-a-multicenter-database-using-a-convolutional-neural-network>

Hosted file

TABLES 2.docx available at <https://authorea.com/users/438359/articles/539581-automated-detection-of-laryngeal-carcinoma-in-laryngoscopic-images-from-a-multicenter-database-using-a-convolutional-neural-network>

Hosted file

TABLES 3.docx available at <https://authorea.com/users/438359/articles/539581-automated-detection-of-laryngeal-carcinoma-in-laryngoscopic-images-from-a-multicenter-database-using-a-convolutional-neural-network>