67567_Auto_Edited - IT check.docx

*Retrospective Study*

**Application of endoscopic ultrasonography for detecting the esophageal lesion based on convolutional neural network**

Liu GS *et al*. Application of EUS for detecting esophageal lesion based on CNN

Gao-Shuang Liu, Peiyun Huang, Min-Li Wen, Shuai-Shuai Zhuang, Jie Hua, Xiaopu He

**Abstract**

BACKGROUND

A convolutional neural network (CNN) is a deep learning algorithm based on the principle of human brain visual cortex processing and image recognition.

AIM

This study aimed to automatically identify the invasion depth and origin of esophageal lesion based on a CNN.

METHODS

A total of 1670 white-light images were used to train and validate the CNN system. The method proposed in this paper included the following two parts: (1) location module, an object detection network, locating the classified main image feature regions of the image for subsequent classification tasks; (2) classification module, a traditional classification convolution neural network, classifying the images cut out by the object detection network.

RESULTS

The CNN system proposed in this study achieved an overall accuracy of 82.49%, sensitivity of 80.23%, and specificity of 90.56%. In this study, after follow-up pathology, 726 patients were compared for endoscopic pathology. The misdiagnosis rate of endoscopic diagnosis in the lesion invasion range was approximately 9.5%; 41 patients showed no lesion invasion to the muscularis propria, but 36 of them pathologically showed invasion to the superficial muscularis propria. The patients with invasion of the tunica adventitia were all treated by surgery with an accuracy rate of 100%. For the examination of submucosal lesions, the accuracy of endoscopic ultrasonography (EUS) was approximately 99.3%. Results of this study showed that EUS had a high accuracy rate for the origin of submucosal lesions, whereas the misdiagnosis rate was slightly high in the evaluation of the invasion scope of lesions. Misdiagnosis could be due to

different operating and diagnostic levels of endoscopists, unclear ultrasound probes, and unclear lesions.

CONCLUSION

This study is the first to recognize esophageal EUS images through deep learning, which can automatically identify the invasion depth and lesion origin of submucosal tumors and classify such tumors, thereby achieving good accuracy. In future studies, this method can provide guidance and help to clinical endoscopists.

**Key Words:** Endoscopic ultrasonography; Convolutional neural network; Esophageal lesion; Automatically; Classify; Identify

Liu GS, Huang P, Wen ML, Zhuang SS, Hua J, He X. Application of endoscopic ultrasonography for detecting the esophageal lesion based on convolutional neural network. *World J Gastroenterol* 2021; In press

**Core Tip:** CNNs with self-learning abilities are an effective method in medical image classification, segmentation, and detection. Endoscopic ultrasonography plays an important role in the diagnosis and treatment of esophageal lesions. However, its operation and lesion identification are more difficult than ordinary endoscopy. Automatic identification technology is of great significance to its development.

**INTRODUCTION**

Endoscopic ultrasonography (EUS) is a mature and widely used invasive examination technology. It is a combination of ultrasound and endoscopy, and the micro high-frequency ultrasound probe is placed on the top of the endoscope to directly observe gastrointestinal mucosal lesions when the endoscope enters the body. It can be used for preoperative staging of digestive tract tumors [1], estimating the invasion depth and range [2], assessing the nature of submucosal lesions of the digestive tract wall [3–4],

and diagnosing bilio-pancreatic system tumors [5–6]. The esophagus is the narrowest part of the digestive tract, and its diameter is not uniform. EUS can be used for high-precision evaluation of the esophagus. At present, common lesions of the esophagus include submucosal tumors (including leiomyoma, stromal tumor, and lipoma), precancerous lesions, and esophageal cancer.

China has a high incidence of esophageal cancer, and the current 5-year survival rate of esophageal cancer is 15%–25% [7–10]. A 13.5-year follow-up study has shown that 74% of patients with high-grade intraepithelial neoplasia (HGIN) have developed esophageal cancer [11]. Surgical and endoscopic resection (ER) has been approved as the standard treatment for esophageal HGIN, and endoscopic therapy has become a widely used method because of its safety, efficiency, and low risk of complications [12–13]. Compared with surgical treatment, endoscopic therapy can reduce complications and preserve esophageal function [14]. EUS plays an important role in the evaluation of esophageal lesions and in the determination of the need for surgery.

With the popularization of endoscopic ultrasonography in clinic, several endoscopic physicians are needed. However, endoscopic ultrasonography is different from other ultrasounds in that ultrasonic examination is performed on the basis of digestive endoscopy; therefore, the examiners should be professional endoscopic physicians with ultrasonic training. In addition, with the increase of chronic diseases, ageing populations, and dwindling resources, we[199491] need to shift to models that can intelligently extract, analyze, interpret, and understand increasingly complex data. The rapidly evolving field of machine learning is a key to this paradigm shift. Artificial intelligence (AI) is a mathematical predictive technology that automatically learns and recognizes data patterns. Deep learning is an AI algorithm and an advanced machine learning method using neural networks [15]. This method aims to build a neural network that mimics brain's mechanisms for interpreting data, such as images, sounds, and text. Deep learning combines low-level features to form abstract high-level representation attribute categories or features and discover distributed feature representation of data. It is often used in AI algorithms, and it has been applied to medical diagnosis [16–19].

At present, relevant studies on deep learning in ultrasound have been found, including the classification of thyroid nodules [20–21], the diagnosis of liver diseases [22], the identification of benign and malignant breast masses [23], the imaging interpretation of Carotid ARTERY IMT (FEATURED by its length) [24], and the prediction of the pancreatic ductal papillary mucinous tumor [25]. However, the research on the identification and classification of esophageal lesions under endoscopic ultrasonography has not yet been found.

This study aimed to construct a framework of deep learning network and study the application of deep learning in esophageal EUS in identifying the origin of submucosal lesions and defining the scope of esophageal lesions.


## MATERIALS AND METHODS

The method proposed in this paper included the following two parts: (1) location module, an object detection network, locating the main image feature regions of the image for subsequent classification tasks; (2) classification module, a traditional classification convolution neural network, classifying the images cut out by the object detection network.

### 2.1 Location module

For gastric ultrasound images, identifying the classification of the whole image was not necessary because the introduction of background information can cause irrelevant features and affect the determination of the model. Therefore, the core area of the image was located, which can classify the image. In this paper, RetinaNet object detection network was used.

RetinaNet network is a one-stage object detection method. It uses the focal loss cost function in the classification branch, which addresses the imbalance between positive and negative samples in one-stage object detection method.

Resnet50 network was used in the feature extraction part of RetinaNet network, and the FPN structure was added, that is, a top-down structure and lateral connection were adopted to detect the object of different scales. Finally, the features obtained were

used to predict the classification branch and regression branch. The classification branch was retained; thus, the multitask training can improve the accuracy of the model, and it can be used to compare the experiment with the classification network (Figure 1)

For object detection anchor setting, this paper adopted the default anchor setting methods, with anchor sizes of 32, 64, 128, 256, and 512 and aspect ratios of 1:2, 2:1, and 1:1.

## 2.2 Classification module

The object detection module cut out the core image block of the image and resize it to 224 x 224. Image augmentation artificially created training images through different ways of processing or combination of multiple processing, such as random rotation, shifts, shear, and flips, to address over-fitting and improve the generalization of the network (Table 3).

The classification convolution neural network used the VGGNet network. Compared with the previous network, the VGGNet network used small convolution and deeper network layers, which achieved better classification accuracy. The classification network was pre-trained on the ImageNet dataset. The last three fully connected layers of network were removed, and then two fully connected layers were added. The number of neurons was 256 and 5 (5 is the classification number of the dataset used in this paper, Figure 2).

## 3. Experiment

During the experiment, GPU configuration was a NVIDIA Geforce GTX1070ti. All the codes were written in Python 3.6, and keras was used as the deep learning library.

In the object detection experiment, smooth L1 cost function was used for regression, and focal loss cost function was used for classification. Adam was used as the optimizer, and the learning rate, momentum, and weight decay were set to 1e-5, 0.9, and 0.0005, respectively. The short side of the input image was rescaled to 800, and the long side was rescaled in equal proportion, with the maximum of 1333.

In the classification experiment, the cross-entropy loss function was used. The optimizer used SGD and set the initial learning rate to 1e-3. When the loss did not decrease, the learning rate was multiplied by a factor of 0.5 until it reached 1e-6. The batch size was set to 16.

Evaluation criteria. The quantitative and qualitative results of the proposed method were provided. A confusion matrix is a table that is often used to describe the performance of a classification model. Each row of the confusion matrix represents the instances of an actual class, and each column represents the instances of a predicted class. The diagonal values represent the number of correct classifications for each class, and the others represent the confusion number between every two classes. As shown in Figure 3, the accuracy of using the classification network is improved compared with the classification branch of the object detection network, which is directly used for classification prediction. The classification accuracy in muscularis propria, muscularis mucosae, and muscularis propria increased significantly.[199491]

The classification performance was measured using sensitivity (SENS), specificity (SPEC), and accuracy (ACC) to clarify the evaluation metrics. Sensitivity measures how much a classifier can recognize positive examples. Specificity measures how much a classifier can recognize negative examples. Accuracy is the ratio of correct decisions made by a classifier. The sensitivity, specificity, and accuracy were defined by Eqs. (1)–(3):

, (1)

, (2)

, (3)

where TP is the amount of positive elements predicted as positive; TN is the amount of negative elements predicted as negative; FP is the amount of negative elements predicted as positive; FN is the amount of positive elements predicted as negative.

As shown in Table 4, the classification network outperforms the classification branch of the object detection network in SENS, SPEC, and accuracy.

Class activation maps (CAM) can be used to obtain the discriminative image regions using a convolutional neural network and identify a specific class in the image. Therefore, a CAM can identify the regions in the image that were relevant to this class. The results of CAM are shown in Figure 4.

## RESULTS

A total of 1115 patients were included in this analysis, including 694 males and 421 females. The age ranged from 24 to 88, with the majority of patients aged over 45. In this study, patients who underwent surgery or endoscopic treatment were followed up for pathology, and 726 of them were compared for ultrasound diagnosis and endoscopic pathology. In the diagnosis of endoscopic ultrasonography, lesions invading the muscularis propria were not detected in 41 patients, and the misdiagnosis rate was approximately 9.5%. In addition, 36 patients showed pathologic invasion of the superficial muscularis propria. All patients with serosal invasion were treated by surgery, and the accuracy rate was 100%. For the detection of submucosal lesions by EUS, the accuracy of EUS was approximately 99.3% (Table 5).

Table 6 shows the ACC, SENS, and SPEC of each category in this study. The ACC of invasion of the serosa was 98.9%, and that of invasion of muscularis propria and muscularis mucosa and submucosa was 73.4% and 77.11%, respectively. In submucosal lesions, the correct identification rate was 78.64% and 84.4%. The image originating from the muscularis mucosa was easily confused with the image invading the muscularis mucosa and submucosa.

## DISCUSSION

Traditionally, the treatment of benign and malignant esophageal lesions is mostly surgery or chemoradiotherapy, but the surgical trauma and sequelae are relatively large. With the development of gastrointestinal endoscopy in recent years, some lesions

can be treated under endoscopy, such as ER, particularly submucosal dissection, which has become the standard treatment for esophageal mucosal cancer [27–29]. In Japan, ER of esophageal cancer refers to the tumor located in the mucosal epithelium (T1A-EP is equal to Tis in situ in the TNM stage [30]) or the proper mucosa (T1A-LPM) because the incidence of lymph node metastasis is low (≤5%) [31]. The recommended diagnostic methods are endoscopic ultrasonography [32], extended endoscopy, and narrow-band imaging (ME-NBI) [33].

Endoscopic ultrasonography plays an important role in the diagnosis and treatment of benign and malignant esophageal lesions. Xu *et al* [3] compared postoperative histological diagnosis with preoperative EUS diagnosis in patients with esophageal leiomyoma, and the diagnostic accuracy of EUS was 88.6%. No treatment-related complications occurred in patients undergoing ER, and no recurrence was found in follow-up examination. In addition, Rana *et al* [34] found that small-probe EUS could predict the response to endoscopic dilation by describing the involvement degree of esophageal benign stenosis and measuring the thickness of the wall. Therefore, EUS plays an important role in the diagnosis of digestive tract diseases. In this study, after follow-up pathology, 726 of the patients were compared with endoscopic pathology. The misdiagnosis rate of endoscopic diagnosis in the lesion invasion range was approximately 9.5%; 41 patients did not show lesion invasion to the muscularis propria, but 36 of them pathologically showed invasion to the superficial muscularis propria. The patients with invasion of the tunica adventitia were all treated by surgery with an accuracy rate of 100%. For the examination of submucosal lesions, the accuracy of EUS was approximately 99.3%. The results of this study showed that endoscopic ultrasonography had a high accuracy rate for the origin of submucosal lesions, and the misdiagnosis rate was slightly higher in the evaluation of the invasion scope of lesions. Misdiagnosis might be due to the different operating and diagnostic levels of endoscopists, unclear ultrasound probes, and unclear lesions. However, when we underestimated the depth of infiltration, the disadvantage became smaller because ER combined with additional chemoradiotherapy and surgical treatment could achieve

good therapeutic effects [35–36]. However, the use of EUS is difficult for non-specialist endoscopist. Therefore, a more advanced measurement method is necessary to help non-specialists achieve diagnostic quality comparable to that of specialists without special techniques.

Deep learning is a new technique to evaluate image information objectively. In recent years, with the application of deep learning in various medical fields, remarkable progress has been made. Artificial intelligence is also becoming popular in ultrasonic examination. Kuwahara *et al* [5] predicted benign and malignant IPMNs by studying endoscopic ultrasonography of intraductal papillary mucinous tumors of the pancreas (IPMNs). The results showed that the SENS, SPEC, and ACC of AI malignant probability were 95.7%, 92.6%, and 94.0%, respectively, which were higher than human diagnosis. Although few studies have been conducted on AI in endoscopic ultrasonography, the diagnosis based on deep learning has shown its advantages. In this study, we found that deep learning algorithm could be used to clear the invasion scope and origin of lesions in endoscopic ultrasound images. In addition, VGGNet was used to perform classification tasks, and multiple superimposed filters were used to increase the nonlinearity of the whole function and reduce the number of parameters. The overall ACC, SENS, and SPEC were 82.49%, 80.23%, and 90.56%, respectively. However, in this study, the images of lesions invading the mucosal musculus and submucosa were easily confused with the images of lesions from the mucosal musculus, which may be caused by the lack of evident stratification caused by tube wall compression or by the excessive noise in the ultrasound images caused by small lesions. No study has been conducted on the application of deep learning-based endoscopic ultrasonography in the esophagus or gastrointestinal tract, and the network framework applicable to endoscopic ultrasonography needs further screening and testing.

In addition, this study has some limitations. First, different probes or operators in endoscopic ultrasound images lead to uneven image quality levels, which may explain the low accuracy in this study. Second, in this study, although the endoscopic ultrasonography probe could see seven layers of tube wall, 1–3 Layers were classified

into one category because the resolution of 1–3 Layers was not clear in most images. Finally, the number of endoscopic ultrasound images was small, and all images were obtained from one center. Therefore, in the following study, we will obtain multi-center samples to increase the sample size and improve the image quality.

## CONCLUSION

This study is the first to recognize esophageal endoscopic ultrasonography images through deep learning, which can automatically identify the lesion invasion depth and lesion source of submucosal tumors and classify such tumors, thereby achieving good accuracy. In future studies, this method can provide guidance and help to clinical endoscopists.

# 67567_Auto_Edited - IT check.docx

ORIGINALITY REPORT

# 8%

SIMILARITY INDEX

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | journals.lww.com<br>Internet | 40 words — 1% |
| 2 | Masanobu Nakajima, Hiroto Muroi, Haruka Yokoyama, Maiko Kikuchi, Satoru Yamaguchi, Kinro Sasaki, Hiroyuki Kato. " F-Fluorodeoxyglucose positron emission tomography can be used to determine the indication for endoscopic resection of superficial esophageal cancer ", Cancer Medicine, 2018<br>Crossref | 35 words — 1% |
| 3 | Wen-shu Chen, Xiao-ling Zheng, Long Jin, Xiao-jie Pan, Ming-fan Ye. "Novel Diagnosis and Treatment of Esophageal Granular Cell Tumor: Report of 14 Cases and Review of the Literature", The Annals of Thoracic Surgery, 2014<br>Crossref | 26 words — 1% |
| 4 | Huai-Ming Sang, Jiu-Liang Cao, Muhammad Djaleel Soyfoo, Wei-Ming Zhang, Jian-Xia Jiang, Shun-Fu Xu. "Endoscopic and Histopathology Characteristics in Patients with Esophageal High-Grade Intraepithelial Neoplasia", Digestive Surgery, 2019<br>Crossref | 24 words — 1% |
| 5 | ai291537555.wordpress.com<br>Internet | 23 words — 1% |

6   Altug Akay, Henry Hess. "Deep Learning: Current and Emerging Applications in Medicine and Technology", IEEE Journal of Biomedical and Health Informatics, 2019
Crossref
16 words — 1%

7   atm.amegroups.com
Internet
16 words — 1%

8   insights.ovid.com
Internet
15 words — 1%

9   Wei Tao, Yuxi Cheng, Xiao-Yu Liu, Bin Zhang, Chao Yuan, Dong Peng, Wei Zhang. "A Simple Predictive Index of the Abdominal Shape for Postoperative Complications After Laparoscopy-Assisted Distal Gastrectomy for Gastric Cancer", Research Square Platform LLC, 2021
Crossref Posted Content
14 words — < 1%

10  Ahmad Khaskiye, Didier Renaud. "Posthatching changes in levels and molecular forms of acetylcholinesterase in slow and fast muscles of the chicken: Effects of denervation and direct electrical stimulation", Differentiation, 1988
Crossref
12 words — < 1%

EXCLUDE QUOTES          ON                    EXCLUDE MATCHES          < 12 WORDS
EXCLUDE BIBLIOGRAPHY    ON