Technology and Health Care 31 (2023) S313–S322
DOI 10.3233/THC-236027
IOS Press

Evaluation of deep learning methods for early gastric cancer detection using gastroscopic images

Xiufeng Su[a], Qingshan Liu[b], Xiaozhong Gao[a] and Liyong Ma[b,*]

[a] *Weihai Municipal Hospital, Cheeloo College of Medicine, Shandong University, Weihai, Shandong, China*

[b] *School of Information Science and Engineering, Harbin Institute of Technology, Weihai, Shandong, China*

Abstract.

BACKGROUND: A timely diagnosis of early gastric cancer (EGC) can greatly reduce the death rate of patients. However, the manual detection of EGC is a costly and low-accuracy task. The artificial intelligence (AI) method based on deep learning is considered as a potential method to detect EGC. AI methods have outperformed endoscopists in EGC detection, especially with the use of the different region convolutional neural network (RCNN) models recently reported. However, no studies compared the performances of different RCNN series models.

OBJECTIVE: This study aimed to compare the performances of different RCNN series models for EGC.

METHODS: Three typical RCNN models were used to detect gastric cancer using 3659 gastroscopic images, including 1434 images of EGC: Faster RCNN, Cascade RCNN, and Mask RCNN.

RESULTS: The models were evaluated in terms of specificity, accuracy, precision, recall, and AP. Fast RCNN, Cascade RCNN, and Mask RCNN had similar accuracy (0.935, 0.938, and 0.935). The specificity of Cascade RCNN was 0.946, which was slightly higher than 0.908 for Faster RCNN and 0.908 for Mask RCNN.

CONCLUSION: Faster RCNN and Mask RCNN place more emphasis on positive detection, and Cascade RCNN places more emphasis on negative detection. These methods based on deep learning were conducive to helping in early cancer diagnosis using endoscopic images.

Keywords: Deep learning, early gastric cancer, gastroscopic images, RCNN

1. Introduction

Cancer is the leading cause of death among people aged $\leqslant$ 70 years in nearly one-third of countries worldwide. Cancer is a huge obstacle to improving people's health and increasing life expectancy, and China has now become a veritable "cancer country". According to the global burden data released by the World Health Organization International Agency for Cancer Research in 2020, China ranks first in terms of the number of new cancer cases and deaths worldwide [1]. The stage of disease progression at diagnosis plays a crucial role in the treatment plan and the survival rate of patients. In China, the incidence

[*] Corresponding author: Liyong Ma, School of Information Science and Engineering, Harbin Institute of Technology, Weihai, Shandong, China. E-mail: maly@hitwh.edu.cn.

of gastric cancer ranks third among all cancers. However, nearly half of the patients with gastric cancer are diagnosed at an advanced stage [2]. The 5-year survival rate of patients diagnosed with advanced gastric cancer is less than 30% even after surgical treatment, while the 5-year survival rate of patients with early gastric cancer (EGC) can reach more than 90% after timely treatment [3,4]. Detecting EGC is extremely important for diagnosis and treatment. Currently, gastroscopy has been widely used in the diagnosis of gastric cancer. However, the clinical symptoms of EGC are not obvious, and hence missed diagnosis and misdiagnosis are easy to occur if not diagnosed by experienced endoscopists. Artificial intelligence (AI) also shows potential application ability in diagnosing gastric cancer. During the examination, the endoscopist saves images of various anatomical locations in the stomach as image files. The identified early cancer and its location are displayed by applying AI technology to these archived images. AI has been used to detect gastric cancer, predict the depth of gastric cancer, and distinguish gastric cancer from gastric ulcer [5–8].

When the field of vision of the endoscopic image is small, the method of image classification can be used to detect gastric cancer. The Inception-v3 model was used to classify noncancerous lesions and EGCs in one study [9]; the sensitivity, specificity, and accuracy were 91.18%, 90.64%, and 90.91%, respectively. A modified U-Net model was employed in another study [10] to classify different cancers and lesions, and the accuracy was 94.1%. The EGC detection performance of VGG16 and ResNet50 models compared in a previous study [11] showed an accuracy of 84.2% and 73.4%, respectively. The developed AI tool Tango showed a higher sensitivity of 84.7% in EGC detection compared with the 65.8% obtained by specialists [12]. However, the network model of this tool was not revealed.

More generally, it is necessary to not only judge the presence of EGC in the endoscopic image but also mark its position with a rectangular box in the endoscopic image. This requires object detection methods, not just classification methods. YOLOv3 was the most widely used deep learning object detection model used for EGC diagnosis. The results were compared with those of senior and junior endoscopists in a previous study [13]. The diagnostic accuracy of AI and senior and junior endoscopists reached 93.2%, 85.9%, and 79.5%, respectively. The YOLOv3 model was further validated in another study [14]. Another commonly used object detection model, Mask RCNN, was also used for EGC detection, and its results were compared with those generated by endoscopists [15]. AI achieved a sensitivity of 93.0%, which was significantly higher than the 80.20% obtained by experts.

Object detection methods based on deep learning are mainly divided into two categories: one-stage methods and two-stage methods. The two-step method first determines the candidate region of the target and then completes the object detection. RCNN is a typical model of this two-step method. The one-step method accomplishes these tasks simultaneously, and its typical model is YOLO. The two-stage method has high detection accuracy and a good detection effect of multiscale targets. It is generally more suitable for medical diagnosis that does not need detection speed but emphasizes on detection sensitivity and specificity [16,17]. As the most classical two-stage methods, RCNN series models have been applied in various kinds of gastric diagnosis: for example, Mask RCNN model for ulcer detection [18], Cascade RCNN model for EGC detection by endoscopic examination combined with ultrasound imaging [19], and Faster RCNN for gastric lesion detection by capsule endoscopy [20].

A meta-analysis [21] showed that AI-based EGC detection was superior to that achieved by human experts. Previous studies reported that the RCNN series models successfully detected EGC. However, no comparative study explored the performances of these different RCNN models in detecting EGC. The main contribution of this study was to compare and analyze the EGC detection performance of different RCNN models. A total of 3659 gastroscopic images were collected, including 1434 images of EGC. Faster RCNN, Mask RCNN, and Cascade RCNN methods were used for EGC detection, and their performances were compared.

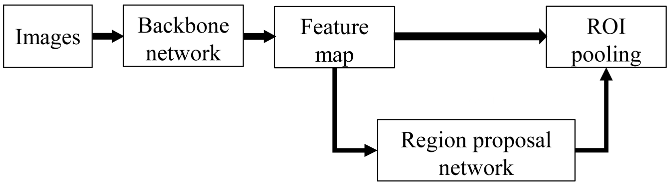

Fig. 1. Network architecture of Faster RCNN.

2. Materials and method

2.1. Imaging data

A total of 2108 patients (1003 male and 1105 female, with an average age of 56.3 years, ranging from 38 to 75 years) with EGC were enrolled at Shandong Gastroscope Quality Control Center from June 2020 to August 2022. A total of 3659 gastroscopic images were collected, including 1434 images of EGC. The study was approved by the ethics committee of the Weihai Municipal Hospital (2020027), which was in accordance with the principles of the Declaration of Helsinki. All patients had signed an informed consent. This study was registered in the Chinese Clinical Trial Registry (ChiCTR2100052273).

2.2. Deep learning detection methods

Three deep learning-based methods were employed in this study: Faster RCNN, Mask RCNN, and Cascade RCNN.

2.2.1. Faster RCNN

Faster RCNN is the most classical object detection algorithm in the two-stage algorithm, and its network architecture is shown in Fig 1.

The Faster RCNN network was divided into three parts: backbone, region proposal network (RPN), and region of interest (ROI) pooling. In the training process, the backbone network generates a feature map, and RPN generates a large number of anchors, which are fine-tuned to a certain extent. Then, the fine-tuned anchors are sent to the proposal layer in RPN to screen out 2000 anchors. Finally, the RPN output and feature map are entered into RoI pooling, which are re-shaped into the two fully connected layers for classification and regression.

During the training process, the loss of Faster RCNN consisted of four parts: RPN Head classification loss and regression loss, and RoI head classification loss and regression loss; the final sum is the total loss. The loss functions for classification and regression are usually cross-entropy and L1 loss.

2.2.2. Cascade RCNN

The network architecture of Cascade RCNN has many similarities with Faster RCNN. However, Cascade RCNN does a key job in improving detection performance. The strategy is changed for selecting intersection over union (IoU) during training.

Experiments showed a mismatch problem in object detection, that is, the detection was best only when the IoU values of the Train and Inference processes were close. However, simply increasing the threshold led to more serious overfitting problems. The Cascade RCNN algorithm proposed a Cascade multistage architecture to solve this problem, indicating that each stage had different IoU thresholds, such as 0.5, 0.6, and 0.7). Its multistage architecture diagram is shown in Fig. 2.

In this way, the IoU was incrementally increased in a cascading manner. After the feature extraction

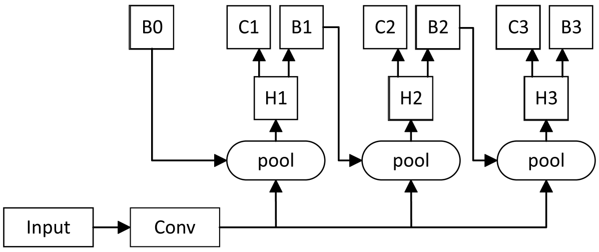

Fig. 2. Multistage architecture of Cascade RCNN.

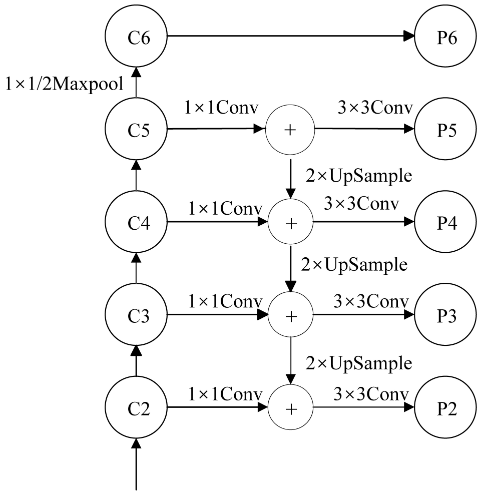

Fig. 3. Architecture of the FPN network for Mask RCNN.

network generated the feature map, Cascade RCNN used different IoU thresholds to train multiple cascade detectors so as to improve the accuracy.

2.2.3. *Mask RCNN*

The Mask RCNN algorithm is also an extension of the Faster RCNN architecture. Its main task is to add an image segmentation branch and combine object detection and semantic segmentation to achieve the effect of instance segmentation. However, Mask RCNN also showed good results in object detection. This is because, unlike Faster RCNN, which uses the VGG network as the backbone, Mask RCNN uses ResNet + FPN as the backbone, which improves feature extraction ability. Meanwhile, in the RoI process, RoI pooling is replaced by RoIAlign without rounding operations to solve the region mismatch problem. The loss of Mask RCNN during training consists of three parts: RPN Head loss, RoI Head loss, and Mask loss for semantic segmentation.

The feature pyramid network (FPN) improves the feature extraction ability and solves the multiscale problem in object detection. The architecture of the FPN network is shown in Fig. 3.

The FPN architecture is made up of three paths. The bottom-up path is the path from which the backbone neural network extracts features. The top-down path is the semantic transfer path, which enables the low-level to obtain the high-level semantic information through upsampling the feature map. The lateral connection path is used to reach the feature fusion between different levels of features. Using FPN, we can make independent predictions at different feature layers, thus improving the detection performance, especially the detection performance of small objects.

Table 1
Model architecture

Architecture	Faster RCNN	Cascade RCNN	Mask RCNN
Backbone	ResNet50	RasNet50+FPN	RasNet50+FPN
RoI	RoI Pooling	RoI Align	RoI Align
BBoxHead	Two FC Head	Cascade Two FC Head	Two FC Head
Cascade IoU	–	[0.5, 0.6, 0.7]	–

Table 2
Hypothetical diagnosis results

		Diagnose	
		Positive	Negative
Reality	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

2.3. Experimental methods

A total of 3659 gastroscopy images were collected, including 2225 normal gastroscopy images (negative) and 1434 EGC images. The images in the experimental data set came from different gastroscopes, and the resolution of these images was different. After image preprocessing, these images were adjusted to the same size. We uniformly divided the data set as follows to visually compare the detection ability and effect of the three methods. We first randomly divided 1434 positive images into train/val dataset (1291 images) and test dataset (143 images) in a 9:1 ratio. An appropriate amount of untargeted data could increase the training accuracy, and too much untargeted data affected the convergence speed of the model. Hence, we chose to randomly select 20% (445 images) from 2225 negative images into the train/val dataset after comprehensive consideration. A total of 133 images were selected from the remaining negative images and imported into the test dataset. In this way, we had the train/val dataset with 1736 images and the test dataset with 276 images. The experiment was repeated five times to avoid the error caused by a single experiment. That is, we randomly divided the data five times by the same method and obtained five different datasets for experiments.

We used three basic architectures in the training process: Faster RCNN, Mask RCNN, and Cascade RCNN. The architecture composition of each architecture is shown in Table 1, FC is full connection layer.

We input the same training dataset into different models during the training process. After training, the test dataset in the same group was input into the model to obtain and analyze the prediction results.

We uniformly set the training rounds at 12 epochs to better compare the results of the three methods. Hyper-parameters were optimized for every model. After experimental comparison, the momentum optimizers were adopted to achieve the best performance indicators. The training and testing processes were carried out on the Ubuntu20.04 operation system with one graphics processing unit (GPU) of GeForce RTX 2080Ti.

The specificity, accuracy, precision, recall, and average precision (AP) were employed as indicators to evaluate different deep learning-based detection methods.

Specificity is an important indicator in medical diagnosis. The diagnosis results and the actual situation are shown in Table 2.

We calculated the specificity as

$$\textit{Specificity} = TN/(TN + FP) \tag{1}$$

Accuracy was calculated as

$$\textit{Accuracy} = (TP + TN)/(TP + TN + FP + FN) \tag{2}$$

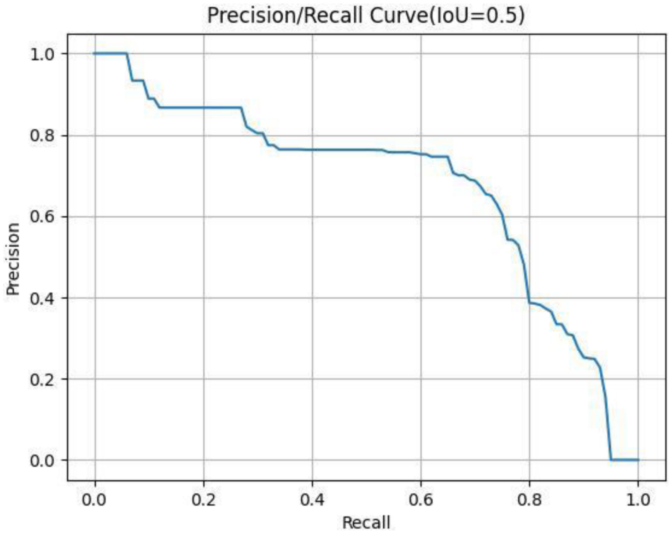

Fig. 4. Typical P-R curve.

Precision and recall were calculated as

$$Precision = TP/(TP + FP) \tag{3}$$

$$Recall = TP/(TP + FN) \tag{4}$$

In this study, when IoU between the detected box and the true box was greater than 0.5, we considered it as TP. The number of all detected boxes was (TP + FP) and the number of all true boxes (lesions) was (TP + FN). We used precision and recall as the evaluation metrics. Precision and recall are contradictory. Therefore, we needed an indicator to measure the model's ability to balance them. AP is the most common evaluation indicator used in Microsoft common objects in context (COCO) datasets. We plotted the precision-recall curve when IoU took a certain value, such as 0.5, as shown in Fig. 4.

The horizontal and vertical coordinates of the images were Recall (R) and Precision (P). Then, the area under the P-R curve was the AP corresponding to this IoU value. AP reflected the comprehensive performance of the model and the ability to detect and locate the targets of the model. It was also the most recognized evaluation indicator in the official COCO competition. In this study, we chose the AP value as the evaluation indicator when the IoU threshold was 0.5.

3. Results and discussion

We arranged the experimental results based on the aforementioned evaluation indicators, as shown in Table 3 and Fig. 5.

Mask RCNN was used for EGC detection with white light endoscope images in [16]. The accuracy and specificity were 90.25% and 89.01% respectively. When Mask RCNN method was used in this study, the accuracy was 93.5% and the specificity was 90.08%. This result was similar to that in [16]. As one-stage method for object detection, YOLOv3 was used for EGC detection in [14], and the accuracy of 85.15% and 86.02% were reported on two data sets. RCNN is two-stage method with high detection accuracy and good detection effect of EGC targets. The accuracy of RCNN in this paper was higher than 93.5%, which also verified that the two-stage method was better than the one-stage method in detecting EGC.

Table 3
Experimental results of different models

Model	Dataset	Precision	Recall	AP (IoU = 0.5)	Specificity	Accuracy
Faster RCNN	*Group 1*	0.425	0.792	0.672	0.932	0.953
	Group 2	0.363	0.759	0.638	0.940	0.938
	Group 3	0.437	0.874	0.746	0.910	0.938
	Group 4	0.429	0.842	0.690	0.917	0.942
	Group 5	0.423	0.822	0.706	0.842	0.902
	Average	*0.415*	*0.818*	*0.690*	*0.908*	*0.935*
Cascade RCNN	*Group 1*	0.719	0.764	0.740	0.977	0.960
	Group 2	0.697	0.793	0.780	0.940	0.920
	Group 3	0.689	0.776	0.789	0.902	0.902
	Group 4	0.669	0.801	0.733	0.962	0.960
	Group 5	0.617	0.685	0.691	0.947	0.946
	Average	*0.678*	*0.764*	*0.747*	*0.946*	*0.938*
Mask RCNN	*Group 1*	0.588	0.792	0.727	0.932	0.949
	Group 2	0.584	0.814	0.737	0.940	0.949
	Group 3	0.587	0.825	0.800	0.910	0.931
	Group 4	0.596	0.829	0.738	0.910	0.942
	Group 5	0.569	0.795	0.733	0.850	0.902
	Average	*0.585*	*0.811*	*0.747*	*0.908*	*0.935*

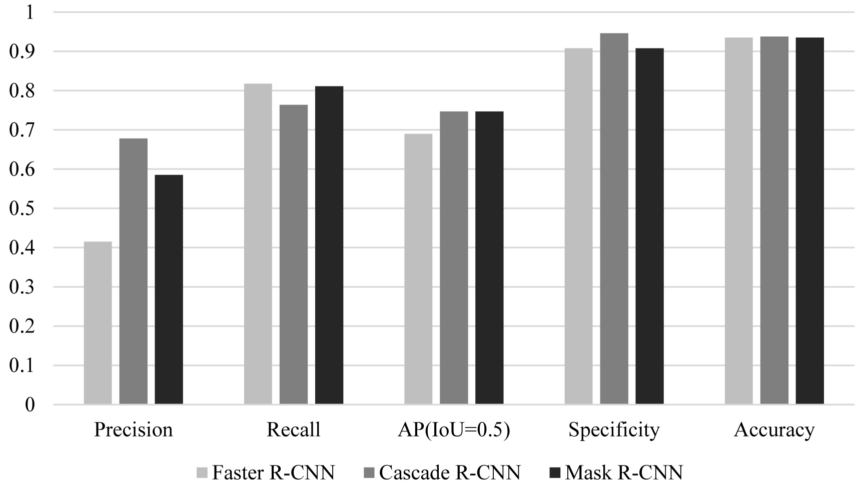

Fig. 5. Performance comparison of different models.

The precision of the Faster RCNN model was lower than that of the Cascade RCNN and Mask RCNN models, but the recall was higher than that of other models. This was because the Faster RCNN model put more emphasis on whether the target could be recognized rather than the accuracy of the recognition. Although the IoU thresholds of the non-maximum suppression (NMS) of the three models were all set to 0.5, significant differences in actual detection were found between the three models. We chose a representative set of images as shown in Fig. 6.

The top left picture is the real box of a test picture, the top right picture is the Cascade RCNN result, the bottom left picture is the Faster RCNN result, and the bottom right picture is the Mask RCNN result. Despite Cascade RCNN and Mask RCNN successfully detecting the lesions, the IoU value of the detection box and the real box was less than 0.5. Therefore, these detections were categorized as missed detections in the statistics. However, Faster RCNN often produced multiple target boxes near a target. In

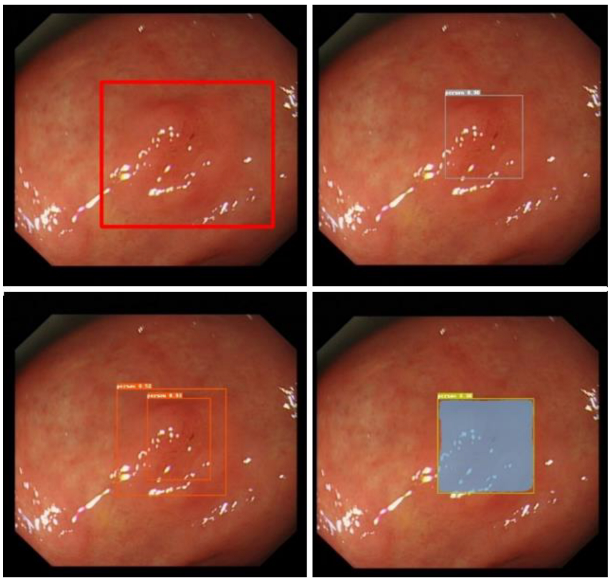

Fig. 6. Comparison of model detection results.

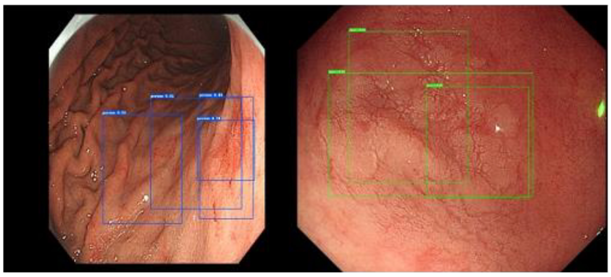

Fig. 7. Multitarget box examples of Faster RCNN.

this example, the larger detection box in the Faster RCNN prediction result successfully matched the real box. Figure 7 shows the multitarget box phenomenon in the Faster RCNN model detection results.

Considering the comprehensive performance of the model (AP), the comprehensive performance of the Cascade RCNN and Mask RCNN models was better than that of Faster RCNN, which was largely attributed to the FPN feature extraction network.

Despite differences in comprehensive performance, the performances of the three models were similar when applied to gastroscopy. Faster RCNN and Mask RCNN put more emphasis on positive detection, and Cascade RCNN puts more emphasis on negative detection.

Precision and recall are at odds. In fact, specificity and recall are also contradictory to some extent. We found that Faster RCNN showed a multitarget box phenomenon, but some positive pictures were observed where Cascade RCNN and Mask RCNN did not detect lesions but Faster RCNN detected lesions. This was because Faster RCNN had low specificity, which enabled it to detect some lesions that were dissimilar to the actual lesions, thus improving the recall rate. Faster RCNN obtained a lower AP, and in object detection, we generally believed that the effect of the model was not good. However, when applied to medical diagnosis, it achieved similar or even better results (more lesions were detected) compared with the other models. We believed that better models did not necessarily lead to better results

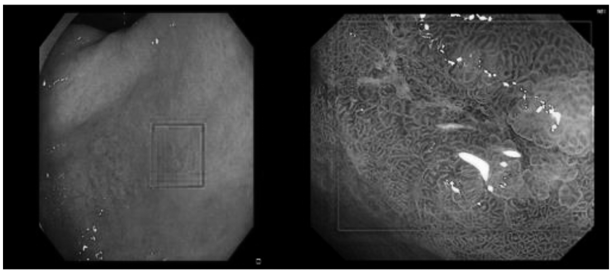

Fig. 8. Multiscale phenomenon in the dataset.

in the field of medical diagnosis, and this is what we needed to pay attention to when selecting models.

From the dataset perspective, it was difficult to deal with the obvious multiscale phenomenon in the dataset, besides the variety of colors and shapes of our EGC dataset, as shown in Fig. 8.

The lesion was extremely small in some images, whereas the entire picture was visible in some images. These objects with different scales, colors, and shapes were no small challenges for model training and testing. Solving these problems is also the direction for further improvement in our model.

4. Conclusion

Faster RCNN, Cascade RCNN, and Mask RCNN models used for detecting EGC were compared in this study. The results were reported and discussed. These methods based on deep learning were conducive to helping in early cancer diagnosis using endoscopic images. However, the performance of these methods could not meet the needs of clinical practice. On the one hand, more advanced AI methods should be proposed for early cancer detection. On the other hand, a combination of more advanced imaging technology and AI methods needs to be developed. For example, the combination of optical and ultrasonic images, as well as photoacoustic images and AI, may promote the development of early cancer detection technology.

Acknowledgments

This study was funded by the Shandong Province Medical and Health Scientific and Technological Development Plan Project (2018WSA10024) and the Shandong Province Natural Science Foundation (ZR2022MH047).

Conflict of interest

None to report.

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA-A Cancer Journal for Clinicians. 2021; 71(3): 209-249.

[2] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA-A Cancer Journal for Clinicians. 2019; 69(1): 7-34.

[3] Zeng HM, Ran XH, An L, Zheng RS, Zhang SW, Ji JS, et al. Disparities in stage at diagnosis for five common cancers in China: A multicentre hospital-based observational study. Lancet Public Health. 2021; 6(12): E877-E887.

[4] Suzuki H, Oda I, Abe S, Sekiguchi M, Mori G, Nonaka S, Yoshinaga S, Saito Y. High rate of 5-year survival among patients with early gastric cancer undergoing curative endoscopic submucosal dissection. Gastric Cancer. 2016; 19(1): 198-205.

[5] Hideo S, Tokai Y, Toshiyuki Y, Tomohiro T. Artificial intelligence for cancer detection of the upper gastrointestinal tract. Digestive Endoscopy. 2021; 33: 254-262.

[6] Ma L, Ma C, Liu Y, Wang X. Thyroid Diagnosis from SPECT Images Using Convolutional Neural Network with Optimization. Computational Intelligence and Neuroscience. 2019; 6212759.

[7] Sun M, Ma L, Su X, Gao X, Liu Z, Ma L. Channel separation-based network for the automatic anatomical site recognition using endoscopic images. Biomedical Signal Processing and Control. 2022; 71: 103167.

[8] Yu C, Helwig EJ. Artificial intelligence in gastric cancer: A translational narrative review. Annals of Translational Medicine. 2021; 9(3): 269.

[9] Li L, Chen YS, Shen Z, Zhang XQ, Sang JZ, Ding Y, et al. Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. Gastric Cancer. 2020; 23: 126-132.

[10] Qiu WG, Xie J, Shen Y, Xu J, Liang J. Endoscopic image recognition method of gastric cancer based on deep learning model. Expert Systems. 2021; 39: e12758.

[11] Ling TS, Wu LL, Fu YW, Xu QW, An P, Zhang J, et al. A deep learning-based system for identifying differentiation status and delineating the margins of early gastric cancer in magnifying narrow-band imaging endoscopy. Endoscopy. 2021; 53: 469-477.

[12] Ishioka M, Osawa H, Hirasawa T, Kawachi H, Nakano K, Fukushima N, et al. Performance of an artificial intelligence-based diagnostic support tool for early gastric cancers: Retrospective study. Digestive Endoscopy. 2022. doi: 10.1111/den.14455.

[13] Tang D, Ni M, Zheng C, Ding X, Zhang N, Yang T, et al. A deep learning-based model improves diagnosis of early gastric cancer under narrow band imaging endoscopy. Surgical Endoscopy. 2022; 36: 7800-7810.

[14] Yao ZD, Jin T, Mao BN, Lu B, Zhang YF, Li SS, Chen WC. Construction and multicenter diagnostic verification of intelligent recognition system for endoscopic images from early gastric cancer based on YOLO-V3 algorithm. Frontiers in Oncology. 2022; 12: 815951.

[15] Jing J, Zhang QQ, Dong B, Ma T, Mei XC, Wang X, et al. Automatic detection of early gastric cancer in endoscopy based on mask region-based convolutional neural networks (Mask RCNN). Frontiers in Oncology. 2022; 12. doi: 10.3389/fonc.2022.927868.

[16] Yang RX, Yu YY. Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis. 2021; 11. doi: 10.3389/fonc.2021.638182.

[17] Kaur A, Singh Y, Neeru N, Kaur L, Singh A. A Survey on Deep learning approaches to medical images and a systematic look up into real-time object detection. Archives of Computational Methods in Engineering. 2022; 29: 2071-2111.

[18] Khan MA, Khan MA, Ahmed F, Mittal M, Goyal LM, Hemanth DJ, Satapathy SC. Gastrointestinal diseases segmentation and classification based on duo-deep architectures. Pattern Recognition Letters. 2020; 131: 193-204.

[19] Wang L, Song H, Wang M, Wang H, Ge R, Shen Y, Yu YL. Utilization of ultrasonic image characteristics combined with endoscopic detection on the basis of artificial intelligence algorithm in diagnosis of early upper gastrointestinal cancer. Journal of Healthcare Engineering. 2021; 2773022.

[20] Xia J, Xia T, Pan J, Gao F, Wang S, Qian YY, et al. Use of artificial intelligence for detection of gastric lesions by magnetically controlled capsule endoscopy. Gastrointestinal Endoscopy. 2021; 93(1): 133.

[21] Jiang KL, Jiang XT, Pan JL, Wen Y, Huang YC, Weng SH, et al. Current Evidence and future perspective of accuracy of artificial intelligence application for early gastric cancer diagnosis with endoscopy: A systematic and meta-analysis. Frontiers in Medicine. 2021; 8: 629080.