

Overview of Deep Learning in Gastrointestinal Endoscopy

Jun Ki Min, Min Seob Kwak, and Jae Myung Cha

Department of Internal Medicine, Kyung Hee University School of Medicine, Seoul, Korea

Artificial intelligence is likely to perform several roles currently performed by humans, and the adoption of artificial intelligence-based medicine in gastroenterology practice is expected in the near future. Medical image-based diagnoses, such as pathology, radiology, and endoscopy, are expected to be the first in the medical field to be affected by artificial intelligence. A convolutional neural network, a kind of deep-learning method with multilayer perceptrons designed to use minimal preprocessing, was recently reported as being highly beneficial in the field of endoscopy, including esophagogastroduodenoscopy, colonoscopy, and capsule endoscopy. A convolutional neural network-based diagnostic program was challenged to recognize anatomical locations in esophagogastroduodenoscopy images, *Helicobacter pylori* infection, and gastric cancer for esophagogastroduodenoscopy; to detect and classify colorectal polyps; to recognize celiac disease and hookworm; and to perform small intestine motility characterization of capsule endoscopy images. Artificial intelligence is expected to help endoscopists provide a more accurate diagnosis by automatically detecting and classifying lesions; therefore, it is essential that endoscopists focus on this novel technology. In this review, we describe the effects of artificial intelligence on gastroenterology with a special focus on automatic diagnosis, based on endoscopic findings. (Gut Liver 2019;13:388-393)

Key Words: Artificial intelligence; Convolutional neural network; Deep learning; Diagnosis, computer-assisted; Endoscopy

INTRODUCTION

Artificial intelligence (AI) can be defined by an intelligence demonstrated by machines in contrast to the natural intelligence displayed by humans and other animals.¹ AI is gener-

ally applied when a machine mimics “cognitive” functions that humans associate with other human minds, such as “learning” and “problem solving.”² The concept of AI first appeared at the Dartmouth Conference held in 1956 by McCarthy *et al.*³ AI pioneers dreamed about developing a super-complex computer with characteristics similar to human intelligence, for example, AI that thinks like human beings with human senses and thinking power. However, today’s AI is characterized by its ability to perform certain tasks, such as image classification or face recognition functions. The recent explosion of AI became possible by the introduction of graphic processing units that offer fast and robust parallel processing capabilities of computer and big data analysis. The three approaches to AI include symbolism (rule based, such as IBM Watson), connectionism (network- and connection-based, such as deep learning or artificial neural net), and Bayesian (based on the Bayesian theorem).⁴

In March 2016, “Baduk (board game Go) AI (AlphaGo)” and Korea’s professional Baduk artist (former world Go champion) had five confrontations, and the result was human defeat.⁵ This confrontation led to the realization that AI is a reality rather than the distant future story, as perceived in movies and novels. Today’s AI offerings struggle to understand natural human language, compete at the highest level in strategic game systems, such as chess or the board game Go, drive autonomous cars, and resolve content delivery network and military simulations. AI is likely to perform many roles performed by humans, although the adoption of AI-based medicine in gastroenterology practice is expected in the near future. Medical image-based diagnoses, such as pathology, radiology, and endoscopy, will be the first in the medical field to be affected by AI. The rapid advancement in AI technology requires physicians to be knowledgeable about AI to aid in the understanding of AI and how it might change and influence the medical field in the near future. Based on advanced deep learning (DL) technology, many start-ups in South Korea are developing businesses in the medical

Correspondence to: Jae Myung Cha

Department of Internal Medicine, Kyung Hee University Hospital at Gangdong, Kyung Hee University School of Medicine, 892 Dongnam-ro, Gangdong-gu, Seoul 05278, Korea

Tel: +82-2-440-6113, Fax: +82-2-440-6295, E-mail: drcha@khu.ac.kr

Received on August 29, 2018. Revised on September 22, 2018. Accepted on October 1, 2018. Published online January 11, 2019

pISSN 1976-2283 eISSN 2005-1212 <https://doi.org/10.5009/gnl18384>

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

image analysis area. The rapid growth of domestic start-ups in a short period of time is probably due to the abundance of data at the core of DL, which is among the strengths of Korea. Korea has been storing digital medical images for a long time and has the advantage of having large amounts of electrical medical records stored because it can receive medical services at a much lower cost than other countries.

In this review, the effects of AI on gastroenterology will be described with specific focus on the automatic diagnosis of endoscopic findings.

DEVELOPMENT OF MACHINE LEARNING

Physicians may be confused about the precise concept of AI, machine learning (ML), and DL (Fig. 1).⁶ In AI, computers can imitate human beings and show intelligence similar to that of human beings. Murphy⁷ defined ML as a set of methods that automatically detect data patterns and then use the detected patterns to predict future data or enable decision-making in uncertain conditions. Thus, it ultimately aims to learn how to do things by “learning” using massive amounts of data and algorithms rather than coding specific instructions directly into the software. The most representative characteristic of ML is that it is driven by data, and the decision process is accomplished with minimal human interventions. Today’s ML has achieved great results in some fields such as computer vision; however, it has the limitation that a certain amount of human instruction is needed in the process. The image recognition rate of ML is sufficient for commercialization, but it remains low in certain fields; this is why image recognition skills are still inferior compared to human capabilities.

DL is the process by which a computer collects, analyzes, and rapidly processes the data it needs without having to accept formal data when performing certain tasks. Therefore, DL is the way a machine responds to create a learning model based on vast amounts of data without receiving instructions from humans. DL is characterized by self-learning; once a training data-

set has been provided, the program can extract the key features and quantities without human indications by using a back-propagation algorithm and changing the internal parameters of each neural network layer.⁸ DL is a type of ML that is developed in an artificial neural network (ANN) and uses data input and output hierarchies similar to brain neurons to learn data. In fact, the ANN was inspired by the biological properties of the human brain, particularly neuronal connections. An ANN is a network in which several nodes (which is corresponding to human brain neurons) are connected, and various types of networks can be created depending on how these nodes are connected. Because ANN requires a great deal of computation power, DL was finally completed by acceleration of neural network computation speed from the development of supercomputers. DL has the potential to automatically detect lesions, suggest differential diagnoses, and compose preliminary medical reports. When ANN was introduced in 1950, it had many limitations, such as a vanishing gradient, overfitting problems, a lack of computing power, and the absence of big data to train the neural network.⁴ The deep neural network (DNN) is an ANN with several layers hidden between the input and output layers; this allows for the modeling of complex data with fewer nodes than a similar ANN. In 2012, DNN was implemented by Google and professor Andrew NG of Stanford University with more than a billion neural networks with 16,000 computers. In ImageNet Large Scale Visual Recognition Challenge 2012, Super Vision, which was proposed by Geoffrey Hinton, professor at the University of Toronto, used a different type of DNN and achieved 84% accuracy, 10% higher than that in the previous year. DL was used to develop computer-aided systems that can support physicians’ diagnoses, and previous reports reported a high level of performance for radiological diagnosis,⁹⁻¹⁴ skin cancer classification,¹⁵ diabetic retinopathy,^{16,17} and histopathology.¹⁸

CONVOLUTIONAL NEURAL NETWORK

Image recognition using DNN has been attempted in some medical fields since it is well suited for medical big data. Convolutional neural network (CNN) is a DNN based on the principle that the visual cortex of the human brain processes and recognizes images. The CNN contains multilayer perceptrons (artificial neurons) and is designed to use minimal preprocessing. CNN leverages the multiple network layers (consecutive convolutional layers followed by pooling layers) to extract the key features from an image and provide a final classification through the fully connected layers as the output.⁸ The convolutional layer, which extracts features from the input data, is composed of a filter that extracts the features and an activation function that converts the value of the filter to a non-linear value (Fig. 2). Because there are many features in the input value, multiple filters are used in the CNN. The combination of multiple filters that extract different features can be applied to

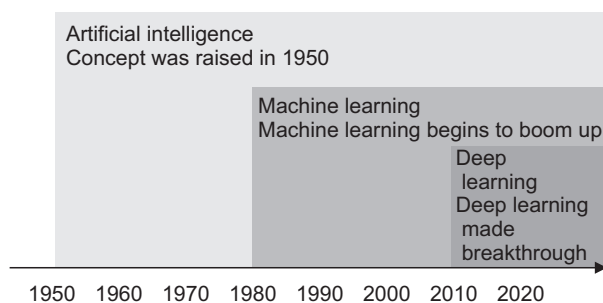


Fig. 1. Machine learning is a manner of achieving artificial intelligence, and deep learning is one of many machine learning methods. Adapted from NVIDIA. Available from: <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>.⁶

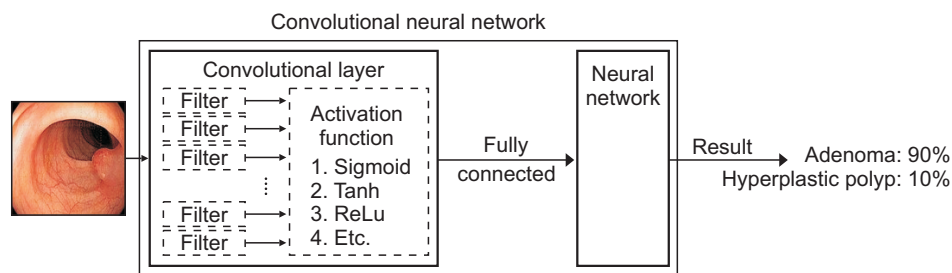


Fig. 2. A convolutional neural network is a multilayer construct of perceptrons designed to use minimal preprocessing. The convolutional layer is composed of a filter for extracting features and an activation function for converting the filter's value to a nonlinear value.

the CNN to determine the characteristics of the original data. The filter is automatically created after recognition of the features through learning from the learning data. Once the feature map has been extracted through the filters, the activation function is applied to make the quantitative value non-linear (yes or no value). Compared to other DL structures, CNN is a popular method for image recognition since it offers good performance in both video and audio applications. For example, CNN has the best performance for image classification in a large image repository, such as ImageNet.¹⁹ Furthermore, CNN is easier to train than other ANN techniques and has the advantage of using fewer parameters. The more data an ML algorithm has access to, the more effective it can be; however, medical images often have insufficient data because they involve privacy concerns. To overcome this, several data augmentation techniques have been proposed to reduce overfitting on models by increasing the amount of training data.

DL-BASED AUTOMATIC DETECTION AND CLASSIFICATION FOR ENDOSCOPY

CNN was recently reported to be highly beneficial in the field of endoscopy, including esophagogastroduodenoscopy (EGD), colonoscopy, and capsule endoscopy (CE). In EGD, a CNN-based diagnostic program was challenged to recognize the anatomical location in EGD images,⁸ *Helicobacter pylori* (HP) infection,^{20,21} and gastric cancer.²² Takiyama *et al.*⁸ reported that the trained CNN can recognize the anatomical location in EGD images. They constructed a CNN-based diagnostic program using 27,335 EGD images from 1,750 patients and validated their program's performance using 17,081 independent EGD images from 435 patients. In this study, the performance of the trained CNN to classify the anatomical location in EGD images provided an area under the curve (AUC) of 1.00 for the larynx and esophagus images and an AUC of 0.99 for the stomach and duodenum images. Furthermore, the trained CNN could recognize specific anatomical locations within the stomach, with an AUC of 0.99 for the upper, middle, and lower stomach. This study highlighted the significant potential for future application as a DL-aided EGD diagnostic system. Two Japanese studies^{20,21} suggested that a CNN-based diagnostic program could recognize HP gastritis with higher accuracy. Itoh *et al.*²⁰ used 179 EGD

images obtained from 139 patients (65 HP-positive and 74 HP-negative) for the development of the CNN; of them, 149 were used as training images and 30 as test images. The sensitivity and specificity for the CNN-based detection of HP infection were 86.7% and 86.7%, respectively, and the AUC was 0.956. Shichijo *et al.*²¹ used a dataset of 32,208 EGD images (HP-positive or -negative) for the first CNN and trained secondary CNN using images classified according to eight anatomical locations. In this study, separate test datasets (11,481 images) from 397 patients was evaluated independently by the CNN and 23 endoscopists. The sensitivity, specificity, and accuracy were 81.9%, 83.4%, and 83.1% for the first CNN and 88.9%, 87.4%, and 87.7% for the secondary CNN, respectively. The secondary CNN had significantly higher accuracy than the endoscopists (by 5.3%; 95% confidence interval, 0.3 to 10.2). Hirasawa *et al.*²² recently challenged the Single Shot MultiBox Detector-based diagnosis of gastric cancer using 13,584 EGD images of gastric cancer and an independent 2,296 test gastric images. The sensitivity and positive predictive value of the CNN for gastric cancer detection were 92.2% and 30.6%, respectively. Furthermore, 70 of the 71 lesions (98.6%) ≥ 6 mm as well as all invasive cancers were correctly detected. However, 69.4% of the false-positive lesions were gastritis with redness or an irregular mucosal surface. All missed lesions were superficially depressed and differentiated-type intramucosal cancers that were difficult for even experienced endoscopists to distinguish from gastritis. This study was very promising, but it had limitations since only high-quality endoscopic images were used from the same endoscope type and endoscopic video system.

For colonoscopy, a CNN-based diagnostic program was applied to detect and classify colorectal polyps.²³⁻²⁶ In 2017, Komeda *et al.*²³ reported on CNN-based diagnosis of colorectal polyps using 1,200 colonoscopy images and 10 additional video images of unlearned processes. The accuracy of the 10-fold cross-validation was 0.751, where the accuracy is the ratio of the number of correct answers to the number of all answers produced by the CNN. Byrne *et al.*²⁴ developed a deep CNN model for the real-time assessment of colonoscopic video images of colorectal polyps. To train and validate the model, they used only narrow-band image (NBI) video frames (split equally between relevant multi-classes) and unaltered videos from routine exams not specifically designed or adapted for CNN

classification. The model was tested on a separate series of 125 videos of 106 consecutively encountered diminutive polyps. The accuracy, sensitivity, specificity, negative predictive value, and positive predictive value of the CNN for the detection of adenoma were 94%, 98%, 83%, 97%, and 90%, respectively. Zhang *et al.*²⁵ reported automatic detection and classification of colorectal polyps by transferring low-level CNN features from the nonmedical domain. They used 1,104 non-polyp images taken under both white-light and NBI endoscopy and 826 NBI endoscopic polyp images (263 hyperplastic polyps and 563 adenomas). CNN-based diagnoses showed similar precision (87.3% vs 86.4%) but a higher recall rate (87.6% vs 77.0%) and accuracy (85.9% vs 74.3%) than endoscopist diagnoses. Yu *et al.*²⁶ proposed a novel offline and online three-dimensional (3D) DL integration framework by leveraging the fully 3D CNN for automated polyp detection. Compared with the hand-crafted features or two-dimensional CNN, the fully 3D CNN is capable of learning more representative spatiotemporal features from colonoscopy videos. They integrated offline and online learning to effectively reduce the number of false positives generated by the offline network and further improve detection performance using the dataset of the 2015 Medical Image Computing and Computer Assisted Intervention (MICCAI) Challenge on Polyp Detection.

In the field of CE, a CNN-based diagnostic program was challenged to recognize celiac disease,²⁷ hookworm,²⁸ and small intestine motility characterization.²⁹ A CNN-based diagnostic program using CE is difficult to develop because CE image quality is usually poor due to hardware and light limitations and low resolution (320×320 pixels) (Fig. 3). CE image quality is further limited by various orientations because of the free motion of the capsule and various extraneous matters such as bile, bubble, food, and fecal material. Zhou *et al.*²⁷ attempted to develop the automatic diagnosis of celiac disease using CE clips from six celiac disease patients and five controls preprocessed for training. The clips of CE from five additional celiac disease patients and five additional control patients were used for testing, and

the GoogLeNet achieved 100% sensitivity and specificity for the testing set. A t-test confirmed significant evaluation confidence to distinguish the celiac disease patients from controls. He *et al.*²⁸ designed a CNN-based program of two integrated networks (edge extraction and hookworm classification) to detect hookworm in CE images. The edge extraction network was used to produce the edge maps for tubular region detection since tubular structures are important clues for hookworm detection.²⁸ They achieved 84.6% sensitivity and approximately 10% to 15% improvement in sensitivity, specificity, and accuracy compared with the previous program. Wu *et al.*³⁰ demonstrated high sensitivity and accuracy for hookworm detection using one of the largest CE datasets (440,000 images of 11 patients). In this study, accuracy, sensitivity, and specificity reach around 78%. Seguí *et al.*²⁹ introduced a deep CNN system for small intestine motility characterization. They demonstrated the superiority of the learned features over alternative classifiers and a mean classification accuracy of 96% for six intestinal motility events, thereby outperforming the other classifiers by a large margin (a 14% relative performance increase).

Despite the promising results of a CNN-based diagnosis program for endoscopic images, its diagnostic ability is highly dependent on training data quality and amount. More training images might result in a more accurate diagnostic ability of the CNN; however, there may be legal or ethical issues surrounding the commercial use of high-quality endoscopy images. When a CNN-based system makes a diagnostic error, legal liability issues could be raised because it is difficult to explain the technical and logical methodology of the system.

LIMITATIONS AND FUTURE DIRECTION

It is not sufficiently verified whether it is possible to improve the medical performance, to reduce the medical cost, and to improve the satisfaction of the patient and medical staff using AI in the medical field. Furthermore, treatment results may vary depending on the indications and range of application, even

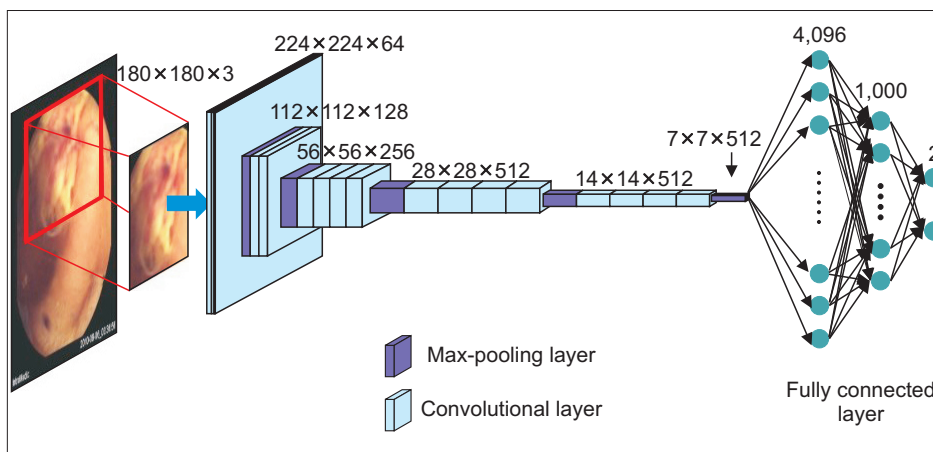


Fig. 3. Sample of a deep convolutional neural network-based ulcer detection algorithm using capsule endoscopy.

using the same AI. Therefore, it is necessary to demonstrate the clinical efficacy of AI, to develop doctor-friendly interface of AI, and to educate physicians to utilize AI. However, it is very difficult to demonstrate the efficacy of AI with clinical trials and to establish guidelines for applying AI. Furthermore, there is a problem that the number of medical staff who can educate AI for physicians is very limited. AI will help physicians in their care because the physician is only responsible for the results of the treatment using AI. In the interpretation of endoscopic images using AI, AI will help physician to reduce medical errors and to improve medical efficiency in the interpretation of endoscopic images. Therefore, physicians should seek best ways to provide better care to their patient with the help of AI. However, it is necessary to make a social consensus for medicolegal issues and to educate future physicians for AI, because AI can cause legal and ethical issues as future AI may exclude physician's decision for the interpretation of endoscopic images.

CONCLUSIONS

DL-based medical technology, such as the automatic detection of cancer or pathological endoscopic findings and more accurate decision-making, are currently or soon to be in use in our daily lives. The CNN-based automatic diagnosis of endoscopic findings is expected to be a mainstream DL technology in the next few decades. DL is expected to help endoscopists provide a more accurate diagnosis by automatically detecting and classifying endoscopic lesions. One of the most important factors for the development of DL for endoscopy would be the availability of large amounts of high-quality endoscopic images, as well as an increased understanding of the technology by the endoscopist. Therefore, it is essential that endoscopists focus on this novel technology.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGEMENTS

This work was supported by a grant from Kyung Hee University in 2018 (KHU-20181044).

ORCID

Jun Ki Min <https://orcid.org/0000-0003-0354-855X>
 Min Seob Kwak <https://orcid.org/0000-0002-8988-7423>
 Jae Myung Cha <https://orcid.org/0000-0001-9403-230X>

REFERENCES

1. Wikipedia. Artificial intelligence [Internet]. San Francisco: Wikipedia Foundation, Inc. [cited 2018 Jun 30]. Available from: https://en.wikipedia.org/wiki/Artificial_intelligence.
2. Russell SJ, Norvig P. Artificial intelligence: a modern approach. 3rd ed. Harlow: Pearson Education, 2009.
3. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI Mag* 2006;27:12-14.
4. Lee JG, Jun S, Cho YW, et al. Deep learning in medical imaging: general overview. *Korean J Radiol* 2017;18:570-584.
5. Bae J, Cha YJ, Lee H, et al. Social networks and inference about unknown events: a case of the match between Google's AlphaGo and Sedol Lee. *PLoS One* 2017;12:e0171472.
6. NVIDIA. What's the difference between artificial intelligence, machine learning, and deep learning? [Internet]. Santa Clara: NVIDIA; c2016 [cited 2018 Jul 29]. Available from: <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>.
7. Murphy KP. Machine learning: a probabilistic perspective. Cambridge: MIT Press, 2012.
8. Takiyama H, Ozawa T, Ishihara S, et al. Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks. *Sci Rep* 2018;8:7497.
9. Larson DB, Chen MC, Lungren MP, Halabi SS, Stence NV, Langlotz CP. Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. *Radiology* 2018;287:313-322.
10. Yasaka K, Akai H, Kunimatsu A, Abe O, Kiryu S. Liver fibrosis: deep convolutional neural network for staging by using gadoteric acid-enhanced hepatobiliary phase MR images. *Radiology* 2018;287:146-155.
11. Yasaka K, Akai H, Abe O, Kiryu S. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. *Radiology* 2018;286:887-896.
12. Cruz-Roa A, Gilmore H, Basavanahally A, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent. *Sci Rep* 2017;7:46450.
13. Trebeschi S, van Griethuysen JJ, Lambregts DM, et al. Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. *Sci Rep* 2017;7:5301.
14. Bibault JE, Giraud P, Burgun A. Big data and machine learning in radiation oncology: state of the art and future prospects. *Cancer Lett* 2016;382:110-117.
15. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115-118.
16. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy

- in retinal fundus photographs. *JAMA* 2016;316:2402-2410.
17. Ting DS, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017;318:2211-2223.
 18. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318:2199-2210.
 19. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJ, Bottou L, Weinberger KQ, eds. *Advances in Neural Information Processing Systems 25*. Red Hook: Curran Associates, Inc., 2012:1097-1105.
 20. Itoh T, Kawahira H, Nakashima H, Yata N. Deep learning analyzes *Helicobacter pylori* infection by upper gastrointestinal endoscopy images. *Endosc Int Open* 2018;6:E139-E144.
 21. Shichijo S, Nomura S, Aoyama K, et al. Application of convolutional neural networks in the diagnosis of *Helicobacter pylori* infection based on endoscopic images. *EBioMedicine* 2017;25:106-111.
 22. Hirasawa T, Aoyama K, Tanimoto T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer* 2018;21:653-660.
 23. Komeda Y, Handa H, Watanabe T, et al. Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: preliminary experience. *Oncology* 2017;93 Suppl 1:30-34.
 24. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019;68:94-100.
 25. Zhang R, Zheng Y, Mak TW, et al. Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. *IEEE J Biomed Health Inform* 2017;21:41-47.
 26. Yu L, Chen H, Dou Q, Qin J, Heng PA. Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos. *IEEE J Biomed Health Inform* 2017;21:65-75.
 27. Zhou T, Han G, Li BN, et al. Quantitative analysis of patients with celiac disease by video capsule endoscopy: a deep learning method. *Comput Biol Med* 2017;85:1-6.
 28. He JY, Wu X, Jiang YG, Peng Q, Jain R. Hookworm detection in wireless capsule endoscopy images with deep learning. *IEEE Trans Image Process* 2018;27:2379-2392.
 29. Seguí S, Drozdal M, Pascual G, et al. Generic feature learning for wireless capsule endoscopy analysis. *Comput Biol Med* 2016;79:163-172.
 30. Wu X, Chen H, Gan T, Chen J, Ngo CW, Peng Q. Automatic hookworm detection in wireless capsule endoscopy images. *IEEE Trans Med Imaging* 2016;35:1741-1752.