Artificial Intelligence in Gastroenterology
MINIREVIEWS

60  Artificial intelligence: A new budding star in gastric cancer

71  Artificial intelligence in gastrointestinal cancer: Recent advances and future perspectives
    Kudou M, Kosuga T, Otsuji E
ABOUT COVER

Editorial Board Member of Artificial Intelligence in Gastroenterology, Dr. Matteo Donadon is a General Surgeon and Assistant Professor of Surgery at Humanitas University, Humanitas Clinical and Research Center. After receiving his MD from the University of Milan, where he also completed his Residency in General Surgery, he earned his PhD in Hepatobiliary Oncology at the University of Sacred Heart. He went on to carry out a Research Fellowship at the Department of Surgical Oncology, University of Texas MD Anderson Cancer Center. Currently, he works as a Senior Surgeon, performing major hepatobiliary procedures. His clinical and research activities remain dedicated to liver and biliary surgery, including translational research. He has authored/coauthored more than 130 articles that are published in peer-reviewed journals, as well as copious abstracts and presentations at national and international meetings. (L-Editor: Filipodia)

AIMS AND SCOPE

The primary aim of Artificial Intelligence in Gastroenterology (AIG, Artif Intell Gastroenterol) is to provide scholars and readers from various fields of artificial intelligence in gastroenterology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

AIG mainly publishes articles reporting research results obtained in the field of artificial intelligence in gastroenterology and covering a wide range of topics, including artificial intelligence in gastrointestinal cancer, liver cancer, pancreatic cancer, hepatitis B, hepatitis C, nonalcoholic fatty liver disease, inflammatory bowel disease, irritable bowel syndrome, and Helicobacter pylori infection.

INDEXING/ABSTRACTING

There is currently no indexing.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: Yu-Jie Ma; Production Department Director: Xiang Li; Editorial Office Director: Jin-Lei Wang.
Artificial intelligence in gastrointestinal cancer: Recent advances and future perspectives

Michihiro Kudou, Toshiyuki Kosuga, Eigo Otsuji

Abstract

Artificial intelligence (AI) using machine or deep learning algorithms is attracting increasing attention because of its more accurate image recognition ability and prediction performance than human-aid analyses. The application of AI models to gastrointestinal (GI) clinical oncology has been investigated for the past decade. AI has the capacity to automatically detect and diagnose GI tumors with similar diagnostic accuracy to expert clinicians. AI may also predict malignant potential, such as tumor histology, metastasis, patient survival, resistance to cancer treatments and the molecular biology of tumors, through image analyses of radiological or pathological imaging data using complex deep learning models beyond human cognition. The introduction of AI-assisted diagnostic systems into clinical settings is expected in the near future. However, limitations associated with the evaluation of GI tumors by AI models have yet to be resolved. Recent studies on AI-assisted diagnostic models of gastric and colorectal cancers in the endoscopic, pathological, and radiological fields were herein reviewed. The limitations and future perspectives for the application of AI systems in clinical settings have also been discussed. With the establishment of a multidisciplinary team containing AI experts in each medical institution and prospective studies, AI-assisted medical systems will become a promising tool for GI cancer.

Key Words: Artificial intelligence; Gastric cancer; Colorectal cancer; Endoscopy; Pathology; Radiology
Core Tip: Artificial intelligence (AI) is attracting increasing attention because of its more accurate image recognition ability and prediction performance than human-aid analyses. The application of AI models to gastrointestinal clinical oncology has been investigated, and the findings obtained indicate its capacity for automatic diagnoses with similar accuracy to expert clinicians and the prediction of malignant potential. However, limitations in the evaluation of gastrointestinal tumors by current AI models have yet to be resolved. The limitations of and future perspectives for the application of AI-assisted systems to clinical settings have been discussed herein.

INTRODUCTION

Recent advances in diagnostic technology and treatment strategies for gastrointestinal cancer have improved clinical outcomes. Even with the development of novel imaging modalities with high accuracy and resolution, image reading, and novel biomarkers, such as the genetic screening of tumors, circulating tumor DNA, and micro RNA, the diversity and quantity of data on tumor malignant potential is beyond the limits of human interpretation[1-4]. Therefore, the establishment of more accurate diagnostic methods with high objectivity using computer-aided diagnosis systems (CAD), such as technologies involving artificial intelligence (AI), is needed in clinical settings[5-11].

AI is defined by the intelligence of machines in contrast to the natural intelligence of humans. It is generally applied when a machine mimics the cognitive functions of humans, such as learning and problem solving[12]. The concept of AI was initially advocated in 1956 by McCarthy et al[13], and the development of machines with the ability to think like humans with intelligence was anticipated. However, machines or computer programs that function as classifiers or detectors, such as image classification and recognition and the prediction of characteristics in populations, are currently regarded as AI.

Recent AI technologies were developed due to technical advances in machine learning and deep neural network algorithms[14-17]. Convolutional neural networks (CNN) are one of the deep neural networks that are useful for image analyses. Algorithms using CNN models have been applied to many research fields in gastrointestinal cancer, such as the automatic endoscopic detection of tumors, the automatic diagnosis of cancer in pathological specimens, and image analyses of radiological modalities[18-20]. In endoscopic research, CNN are trained using thousands of endoscopic images to detect tumors, differentiate between benign and malignant tumors, and predict tumor invasion depth[21-23]. In recent years, a real-time CAD endoscopic system was developed using trained CNN. In the area of pathology, deep learning has been performed using non-cancerous and cancer images to automatically identify and segment the cytoplasm, nucleus, and stromal cells. CNN and machine learning models with image analyses, such as a texture analysis, were subsequently built to identify cancerous regions or diagnose cancer[24]. In the field of radiology, a CAD system of image modalities, such as X-ray, computed tomography (CT), and magnetic resonance images (MRI), was developed using a deep learning model constructed using cancer and non-cancer images to recognize anatomy and detect and segment tumors[25]. The malignant potential of tumors has been analyzed using a radiomics approach, which aims to quantitatively assess tumor heterogeneity by an analysis of medical images through the deep or machine learning of histograms, textures, and shapes[26-30]. AI models of gastrointestinal cancer are summarized in Figure 1.

AI with strong analytical power has attracted the attention of many researchers; therefore, the number of studies on diagnostic AI systems in gastrointestinal cancer has rapidly increased in the past decade. We herein investigate recent advances and future perspectives through a review of the literature.

In this minireview, the bibliographic search was performed using the database MEDLINE (through PubMed) for identifying studies published on AI technology in
Figure 1 Clinical research using artificial intelligence in gastrointestinal cancer. Deep learning based on convolutional neural networks showing the input layer with raw data of the image, such as endoscopic, pathological, and radiological images, the hidden layer with a series of convolutions computed for each layer and the classification of the image, the prediction of malignant potentials, and the segmentation of tumor in the output layer.

APPLICATION OF AI TO ENDOSCOPY IN GASTROINTESTINAL CANCER

Previous studies on the endoscopic diagnosis of gastric cancer (GC) and colorectal cancer (CRC) using AI between 2016 and 2020 were summarized in Tables 1 and 2.

Gastric cancer

The purposes of the studies reviewed on AI for GC were (1) tumor detection; (2) the diagnosis of malignancy; (3) real-time detection; and (4) the prediction of tumor invasion depth. The basic method of these studies was as follows: Endoscopic images of GC, gastritis, and non-cancerous mucosae, which were diagnosed pathologically or by an expert endoscopist, were captured and CNN was subsequently trained using these images. Diagnostic and detection accuracy were then assessed using the constructed CNN models.

Yoon et al. attempted to develop CNN models with the ability to detect early GC and predict invasion depth. The areas under the curves of receiver operating characteristic curves (AUC) for early GC detection and depth prediction were 0.981 and 0.851, respectively. Moreover, the diagnostic accuracy of invasion depth was lower for undifferentiated GC than for differentiated GC. Zhu et al. also trained a CNN model to predict the invasion depth of GC. The AUC, positive predictive value (PPV), and negative predictive value (NPV) of their model were 0.94, 89.6%, and 88.9%, respectively. The CNN-CAD system achieved significantly higher accuracy and specificity than a human endoscopist. Li et al. also developed CNN models for the detection of GC with high diagnostic accuracy (sensitivity: 91.1%, specificity: 90.6%, and PPV: 90.9%). Hirasawa et al. reported that CNN models exhibited difficulties distinguishing between differentiated-type intramucosal cancers with a diameter of 6 mm or less and gastritis. Ishioka et al. examined the detection accuracy of a real-time endoscopic diagnosis of GC using CNN models that they had constructed; the detection rate of GC using these models was 94.1%. CNN identified the region of GC that had been difficult to distinguish from background gastritis, even by experienced endoscopists. Luo et al. developed a gastrointestinal AI diagnostic system (GRAIDs) and compared its diagnostic accuracy with that of expert and trainee endoscopists. PPV was 0.814 for GRAIDs, 0.932 for the expert endoscopist, and 0.824 for the trainee endoscopist, while NPV was 0.978 for GRAIDs, 0.980 for the expert endoscopist, and
Table 1 Previous studies on upper endoscopy of gastric cancer using artificial intelligence

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Targets</th>
<th>Sample sizes</th>
<th>Inputs</th>
<th>Tasks</th>
<th>Analysis method</th>
<th>Diagnostic performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yoon et al[28]</td>
<td>GC (ESD/surgery)</td>
<td>800 cases</td>
<td>GC/non-GC images in close-up and distant views</td>
<td>Detection and invasion depth prediction</td>
<td>CNN</td>
<td>AUC: detection, 0.981; depth, 0.851</td>
</tr>
<tr>
<td>Zhu et al[29]</td>
<td>GC</td>
<td>993 images</td>
<td>GC images</td>
<td>Diagnosis of invasion depth</td>
<td>CNN</td>
<td>Sensitivity: 76.4%, PPV: 89.6%</td>
</tr>
<tr>
<td>Li et al[30]</td>
<td>GC and healthy</td>
<td>386 GC and 1702 NC images</td>
<td>NBI images</td>
<td>Diagnosis of GC</td>
<td>CNN</td>
<td>Sensitivity: 91.1%, PPV: 90.6%</td>
</tr>
<tr>
<td>Hirasawa et al[31]</td>
<td>GC</td>
<td>13584 training and 2296 test images</td>
<td>GC images</td>
<td>Diagnosis of GC</td>
<td>CNN</td>
<td>Sensitivity: 92.2%, PPV: 30.6%</td>
</tr>
<tr>
<td>Ishioka et al[32]</td>
<td>EGC</td>
<td>62 cases</td>
<td>Real-time images</td>
<td>Detection</td>
<td>CNN</td>
<td>Detection rate: 94.1%</td>
</tr>
<tr>
<td>Luo et al[33]</td>
<td>GC</td>
<td>1036496 images</td>
<td>GC images</td>
<td>Detection</td>
<td>CNN</td>
<td>PPV: 0.814, NPV:0.978</td>
</tr>
<tr>
<td>Horii et al[34]</td>
<td>GC and gastritis</td>
<td>1492 GC and 1078 gastritis images</td>
<td>NBI images</td>
<td>Detection</td>
<td>CNN</td>
<td>Sensitivity: 95.4%, PPV: 82.3%</td>
</tr>
</tbody>
</table>

GC: Gastric cancer; CNN: Convolutional neural network; AUC: Area under the curve; PPV: Positive predictive value; NC: Non-cancer; NBI: Narrow-band image; EGC: Early gastric cancer.

Table 2 Previous studies on colonoscopy using artificial intelligence

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Targets</th>
<th>Sample sizes</th>
<th>Inputs</th>
<th>Tasks</th>
<th>Analysis method</th>
<th>Diagnostic performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akbari et al[35]</td>
<td>Screening endoscopy</td>
<td>300 polyp images</td>
<td>Polyp images</td>
<td>Auto segmentation of polyps</td>
<td>CNN</td>
<td>Accuracy: 0.977, Sensitivity: 74.8%</td>
</tr>
<tr>
<td>Jin et al[36]</td>
<td>Screening endoscopy</td>
<td>Training: 2150 polyps, test: 300 polyps</td>
<td>NBI images</td>
<td>Differentiation of adenoma and hyperplastic polyps</td>
<td>CNN</td>
<td>The model reduced the time of endoscopy and increased accuracy by novice endoscopists</td>
</tr>
<tr>
<td>Urban et al[37]</td>
<td>Screening endoscopy</td>
<td>8641 polyp images and 20 colonoscopy videos</td>
<td>Polyp images</td>
<td>Detection of polyps</td>
<td>CNN</td>
<td>AUC: 0.991, Accuracy: 96.4%</td>
</tr>
<tr>
<td>Yamada et al[38]</td>
<td>Screening endoscopy</td>
<td>4840 images, 77 colonoscopy videos</td>
<td>Real-time images</td>
<td>Differentiation of the early signs of CRC</td>
<td>CNN</td>
<td>Sensitivity: 97.3%, Specificity: 99.0%</td>
</tr>
</tbody>
</table>

CNN: Convolutional neural network; NBI: Narrow-band image; AUC: Area under the curve.

0.904 for the trainee endoscopist. These findings demonstrated that the diagnostic accuracy of GRAIDs for the detection of GC was similar to that of the expert endoscopist and superior to that of the trainee endoscopist. CNN models of narrow-band imaging (NBI) for GC have been reported, with sensitivity and PPV of 91.1-95.4% and 82.3-90.6%, respectively[34].

Colorectal cancer

The purposes of the studies reviewed on AI for CRC were (1) the segmentation and detection of polyps; and (2) the diagnosis of polyp pathology. In the development of efficient automatic diagnostic models, models need to automatically segment polyps and extract their features. Akbari et al[35] attempted to construct CNN models of colonoscopy for automatic segmentation and feature extraction. The accuracy, specificity, and sensitivity of the model for automatic segmentation were 0.977, 0.993, and 0.758, respectively. An ideal CAD system of colonoscopy needs to have the ability to predict the pathological diagnosis of an automatically detected tumor and subsequently recommend appropriate treatment strategies for lesions. Jin et al[36] reported a CNN model for predicting the pathological diagnosis of small lesions (≤ 5 mm) using NBI data from colonoscopy. The accuracy, sensitivity, specificity, PPV, and NPV of their model for predicting the pathological diagnosis of polyps, adenoma vs hyperplasia were 86.7%, 83.3%, 91.7%, 93.8%, and 78.6%, respectively. On the other hand, the accuracies of polyp diagnoses by novices, experts, and NBI-trained expert endoscopists were 73.8%, 83.8%, and 87.6%, respectively. Using CNN-processed
results, overall accuracy by novice endoscopists significantly increased to 85.6%. A real-time diagnostic system in colonoscopy was developed using CNN models. Urban et al. constructed CNN models to identify polyps, which were subsequently adapted to colonoscopy videos, and these models exhibited the ability to detect either type of polyp equally well and identify polyps with an ROC value of 0.991 and accuracy of 96.4%. Yamada et al. applied their CNN model, which was developed to detect early signs of CRC, to colonoscopic videos. The sensitivity and specificity of their AI system for detecting the regions of CRC were 97.3% and 99.0%, respectively, while the sensitivity and specificity of endoscopists were 87.4% and 96.4%; respectively. Therefore, the AI system may be used to alert endoscopists in real-time to overlooked abnormalities, such as non-polypoid polyps, during colonoscopy, thereby increasing the early detection of this disease.

APPLICATIONS OF AI TO THE PATHOLOGICAL DIAGNOSIS OF GASTROINTESTINAL CANCER

Previous studies on the pathological diagnosis of GC and CRC using AI between 2016 and 2020 are summarized in Tables 3 and 4. An automatic pathological diagnosis of gastrointestinal cancer generally involves the following processes: (1) Automatic segmentation: Distinguishing various structures, such as the cytoplasm, nuclei, and stoma, and the recognition of atypia; (2) The diagnosis and grading of carcinoma; (3) The diagnosis of malignant potential, such as invasion depth and lymphovascular invasion; and (4) The prediction of survival. Therefore, previous studies aimed to develop a CAD system with the ability to perform these processes.

Gastric cancer
Qu et al. attempted to develop CNN models for (1) and (2), proposed a novel stepwise fine-tuning-based deep learning scheme for gastric pathology image classification, and established a novel protocol to further boost the performance of state-of-the-art deep neural networks and overcome the insufficiency of well-annotated data. In their proposed two-stage method, CNN was initially trained using tissue-wise data on the background, epithelium, and stoma as well as cell-wise data on nuclei and the cytoplasm, and was then tuned using well-annotated data from benign or malignant data sets. The diagnostic accuracy of their constructed two-stage CNN models was higher than that of one-stage models. Yoshida et al. attempted to develop CNN models for (1) and (2) with the ability to automatically segment malignant regions in full-slide images of biopsy samples and subsequently diagnose histological classifications through a nuclear analysis at high magnification. In negative biopsy specimens, the concordance rate between their AI system and expert pathologists was 90.6%; however, the concordance rate for positive biopsy specimens was less than 50%. Mori et al. trained CNN models for (3) to discriminate the tumor invasion depth of gastric signet-ring cell carcinoma. Their models exhibited the ability to diagnose intramucosal or advanced histological characteristics with an accuracy of 85%, sensitivity of 90%, specificity of 81%, and AUC of 0.91. The prediction of survival in GC patients using the deep learning method has also been examined. Jiang et al. investigated the efficacy of deep learning models for (4) using a support vector machine (SVM). They classified GC patients into two groups using SVM based on patient characteristics and immunohistochemistry (IHC) data on the following immunomarkers: CD3, CD8, CD45RO, CD45RA, CD57, CD68, CD66b, and CD34. The findings obtained revealed that the classifier of SVM was a stronger prognostic factor than the TNM stage or CA19-9.

Colorectal cancer
Numerous studies on the pathology of CRC using AI were reported compared to GC, and are classified as follows.

Studies on AI models for automatic segmentation: Van Eycke et al. and Graham et al. developed CNN models to segment the glandular epithelium. The F1 values of these models ranged between 0.9 and 0.912. Abdelsamea et al. developed tumor parcellation and quantification (TuPaQ), which is a tool for refining biomarker analyses through the rapid and automated segmentation of the tumor epithelium. Tissue microarray (TMA) cores from CRC were manually annotated and analyzed to provide the ground truth, epithelial or non-epithelial tissue. CNN (TuPaQ) was trained using these data. The accuracy, sensitivity, and specificity of TuPaQ were
Table 3 Previous studies on the pathology of gastric cancer using artificial intelligence

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Targets</th>
<th>Sample size</th>
<th>Input</th>
<th>Task</th>
<th>Analysis method</th>
<th>Diagnostic performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qu et al[39]</td>
<td>GC</td>
<td>15000 images</td>
<td>Pathological images</td>
<td>Evaluation of stepwise methods</td>
<td>CNN</td>
<td>AUC: 0.828-0.920</td>
</tr>
<tr>
<td>Yoshida et al[40]</td>
<td>GC</td>
<td>3062 biopsy samples</td>
<td>Pathological images stained by H&amp;E</td>
<td>Automatic segmentation, diagnosis of carcinoma</td>
<td>CNN</td>
<td>Sensitivity: 89.5%, specificity: 50.7%</td>
</tr>
<tr>
<td>Mori et al[41]</td>
<td>GC (surgery)</td>
<td>516 images from 10 GC cases</td>
<td>Pathological images stained by H&amp;E</td>
<td>Diagnosis of invasion depth in signet cell carcinoma</td>
<td>CNN</td>
<td>Sensitivity: 90%, Specificity: 81%</td>
</tr>
<tr>
<td>Jiang et al[42]</td>
<td>GC (surgery)</td>
<td>786 cases IHC (CD3, CD8, CD45RO, CD45RA, CD57, CD68, CD66b, and CD34)</td>
<td>Prediction of survival</td>
<td>SVM</td>
<td>The immunomarker SVM was useful for predicting survival</td>
<td></td>
</tr>
</tbody>
</table>

GC: Gastric cancer; AUC: Area under the curve; H&E: Hematoxylin eosin staining; CNN: Convolutional neural network; IHC: Immunohistochemistry; SVM: Support vector machine.

0.939, 0.779, and 0.946, respectively. Yan et al[46] examined the diagnostic accuracy of their AI models for the classification, segmentation, and visualization of large-scale tissue histopathology images. The accuracies of their models ranged between 81.3 and 93.2%. Haj-Hassan et al[47] attempted to develop CNN models for the automatic segmentation of benign hyperplasia, intra-epithelial neoplasms, and carcinoma, and the findings obtained showed that the models segmented tumors with a high accuracy of 99.1%.

**Diagnosis and grading of carcinoma:** Rathore et al[48] reported deep learning models for cancer detection and grading. The features of CRC biopsy samples were extracted based on pink-colored connecting tissues, purple-colored nuclei, and white-colored epithelial cells and lumina. The extracted features, particularly white-colored epithelial cells and lumina, were classified using SVM and classification performance was subsequently assessed. The accuracies of cancer detection and grading by their model were 95.4 and 93.4%, respectively. Yang et al[49] proposed a combination of SVM and color histograms to classify pathological images. The AUC of the model for diagnosing carcinoma was 0.891. Chaddad et al[50] reported that the classification of images using a texture analysis effectively diagnosed carcinoma (accuracy: 98.9%). Yoshida et al[51] showed that a CAD system using a previously described CNN model for GC was useful for diagnosing adenoma and carcinoma (undetected rate of carcinoma and adenoma: 0.93% and 0.9-9.9%, respectively).

**Diagnosis of malignant potential:** Takamatsu et al[52] reported the prediction of lymph node metastasis using a machine learning analysis of morphological parameters (such as shape and roundness) in cytokeratin-stained T1 CRC images. The AUC of the model was 0.94. The automatic evaluation of tumor budding in IHC with CNN and machine learning was previously performed[53]. Models were constructed to assess tumor budding using TMA on pan-cytokeratin-stained tumors, and the $R^2$ value of the correlation of the models with manual counting for the diagnosis of tumor budding was 0.86.

**Prediction of survival:** Bychkov et al[54] proposed AI models for the automatic prediction of survival in CRC patients using the TMA of CRC pathological images. The automatic detection of tumors was initially achieved using CNN; CNN cases were subsequently classified by a recurrent neural network. Predicted survival by their model correlated with actual clinical outcomes. Kather et al[55] reported automatic models for discriminating structures in tissue samples and then predicting survival. Their models predicted the survival of CRC more accurately than the TNM stage or manual evaluations of cancer-associated fibroblasts. Moreover, survival prediction SVM models using immunomarkers evaluated by IHC, such as CD3 and CD8, have been developed[56], and the classifier correlated with patient survival.
<table>
<thead>
<tr>
<th>Ref.</th>
<th>Targets</th>
<th>Sample size</th>
<th>Input</th>
<th>Task</th>
<th>Analysis method</th>
<th>Diagnostic performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Eycke et al [43]</td>
<td>CRC</td>
<td>CRC</td>
<td>H&amp;E staining, IHC image</td>
<td>Segmentation of the glandular epithelium</td>
<td>TMA, CNN</td>
<td>F1 value: 0.912</td>
</tr>
<tr>
<td>Graham et al [44]</td>
<td>CRC</td>
<td>333 samples</td>
<td>H&amp;E staining</td>
<td>Differentiation of the tumor epithelium</td>
<td>CNN</td>
<td>F1 values: 0.90</td>
</tr>
<tr>
<td>Abdelsamea et al [45]</td>
<td>CRC</td>
<td>333 samples</td>
<td>H&amp;E staining, IHC (CD3)</td>
<td>Differentiation of the tumor epithelium</td>
<td>TMA, CNN</td>
<td>Accuracy: 0.93-0.94</td>
</tr>
<tr>
<td>Yan et al [46]</td>
<td>CRC</td>
<td>180 samples</td>
<td>H&amp;E staining</td>
<td>Tumor classification, segmentation of tumors</td>
<td>CNN</td>
<td>Accuracy: Classification, 97.8%; segmentation, 84%</td>
</tr>
<tr>
<td>Haj-Hassan et al [47]</td>
<td>CRC</td>
<td>Multispectral images</td>
<td>Segmentation of carcinoma</td>
<td>CNN</td>
<td>Accuracy: 99.1%</td>
<td></td>
</tr>
<tr>
<td>Rathore et al [48]</td>
<td>CRC</td>
<td>Biopsy samples</td>
<td>H&amp;E staining</td>
<td>Detection and grading of tumors</td>
<td>Texture and morphology patterns, SVM</td>
<td>Recognition rate: Detection, 95.4%; grading: 93.4%</td>
</tr>
<tr>
<td>Yang et al [49]</td>
<td>CRC</td>
<td>1328 samples</td>
<td>H&amp;E staining</td>
<td>Diagnosis of benign tumors, neoplasms, and carcinoma</td>
<td>SVM, histogram, texture</td>
<td>AUC: 0.852</td>
</tr>
<tr>
<td>Chaddad et al [50]</td>
<td>CRC</td>
<td>30 cases</td>
<td>H&amp;E staining</td>
<td>Diagnosis of carcinoma, adenoma, and benign tumors</td>
<td>Automatic segmentation, texture</td>
<td>Accuracy: 98.9%</td>
</tr>
<tr>
<td>Yoshida et al [51]</td>
<td>CRC</td>
<td>180 cases</td>
<td>H&amp;E staining</td>
<td>Diagnosis of benign tumors, neoplasms, and carcinoma</td>
<td>CNN, automatic analysis of structure</td>
<td>Undetected rate of carcinoma and adenoma: 0-9.3% and 0-9.9%, respectively</td>
</tr>
<tr>
<td>Takamatsu et al [52]</td>
<td>CRC surgery</td>
<td>397 samples</td>
<td>H&amp;E staining</td>
<td>Prediction of lymph node metastasis</td>
<td>LR, shape analysis</td>
<td>AUC: 0.94</td>
</tr>
<tr>
<td>Weis et al [53]</td>
<td>CRC</td>
<td>596 cases</td>
<td>IHC (AE1/AE3)</td>
<td>Automatic evaluation of tumor budding</td>
<td>TMA, CNN</td>
<td>Correlation; R2 value: 0.86</td>
</tr>
<tr>
<td>Bychkov et al [54]</td>
<td>CRC surgery</td>
<td>420 cases</td>
<td>H&amp;E staining</td>
<td>Prediction of survival</td>
<td>TMA, CNN</td>
<td>Good biomarker for predicting survival</td>
</tr>
<tr>
<td>Kather et al [55]</td>
<td>CRC</td>
<td>973 slides</td>
<td>H&amp;E staining</td>
<td>Prediction of survival</td>
<td>Stronah pattern, CNN</td>
<td>Good biomarker for predicting survival</td>
</tr>
<tr>
<td>Reichling et al [56]</td>
<td>CRC surgery</td>
<td>1018 cases</td>
<td>HE, IHC (CD3, CD8)</td>
<td>Prediction of survival</td>
<td>RF, monogram</td>
<td>Good biomarker for predicting survival</td>
</tr>
</tbody>
</table>

CRC: Colorectal cancer; H&E: Hematoxylin eosin staining; IHC: Immunohistochemistry; TMA: Tissue microarray; CNN: Convolutional neural network; SVM: Support vector machine; AUC: Area under the curve; LR: Linear regression.

**APPLICATIONS OF AI TO A RADIOLOGICAL DIAGNOSIS OF GASTROINTESTINAL CANCER**

Previous studies on the radiological diagnosis of GC and CRC using AI between 2016 and 2020 were summarized in Tables 5 and 6.

**Gastric cancer**

Regarding GC, many researchers have attempted to develop AI models using (1) a radiomics approach; or (2) CNN models predicted malignant potential, such as survival, lymph node metastasis, and post-operative recurrence, through analyses of the radiological image features of GC.

**Radiomics approach:** Li et al [57] developed a survival prediction model involving a general radiomics analysis of CT. The region of interest was manually drawn along the margin of the tumor on CT images, and radiological features were extracted. After manual image segmentation, the heterogeneity of the extracted feature was quantified using an image analysis, such as texture and histogram analyses. Analyzed cases were then classified based on the risk score (R-signature) evaluated using the least absolute shrinkage and selection operator method. The performance of a radiomics nomogram, including factors correlating with survival, was then evaluated. The findings obtained showed that the R-signature correlated with the survival of GC patients. Furthermore, the prediction of survival by the radiomics monogram including the R-signature was...
Table 5 Previous studies on the radiological diagnosis of gastric cancer using radiomics or artificial intelligence

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Targets</th>
<th>Sample size</th>
<th>Input</th>
<th>Task</th>
<th>Analysis method</th>
<th>Diagnostic performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li et al [57]</td>
<td>GC, radical surgery</td>
<td>181 cases</td>
<td>Primary tumor, preoperative CT</td>
<td>Prediction of survival</td>
<td>Manual segmentation, radiomics, Nomograms</td>
<td>The TNM stage and radiomics signature were good biomarkers</td>
</tr>
<tr>
<td>Zhang et al [64]</td>
<td>GC, radical surgery</td>
<td>669 cases</td>
<td>Primary tumor, preoperative CT</td>
<td>Prediction of early recurrence</td>
<td>Manual segmentation, radiomics, Nomograms</td>
<td>AUC: 0.806–0.831</td>
</tr>
<tr>
<td>Li et al [58]</td>
<td>GC, radical surgery</td>
<td>204 cases</td>
<td>Primary tumor, preoperative dual-energy CT</td>
<td>Pre-operative diagnosis of LNM</td>
<td>Manual segmentation, radiomics, Nomogram</td>
<td>AUC: 0.82–0.84</td>
</tr>
<tr>
<td>Li et al [59]</td>
<td>GC, radical surgery</td>
<td>554 cases</td>
<td>Primary tumor, preoperative CT</td>
<td>Prediction of a pathological status, survival</td>
<td>Semi-automatic segmentation, radiomics</td>
<td>AUC for prediction of the pathological status: 0.77, the TNM stage and radiomics signature were good biomarkers</td>
</tr>
<tr>
<td>Wang et al [60]</td>
<td>GC, radical surgery</td>
<td>187 cases</td>
<td>Primary tumor, preoperative dynamic CT</td>
<td>Pre-operative prediction of intestinal-type GC</td>
<td>Manual segmentation, radiomics, Nomograms</td>
<td>AUC: 0.904</td>
</tr>
<tr>
<td>Jiang et al [61]</td>
<td>GC, surgery</td>
<td>214 cases</td>
<td>Primary tumor, preoperative PET-CT</td>
<td>Prediction of survival</td>
<td>Manual segmentation, radiomics, Nomograms</td>
<td>C-index: DFS, 0.800; OS, 0.786</td>
</tr>
<tr>
<td>Chen et al [62]</td>
<td>GC, surgery</td>
<td>146 cases</td>
<td>Primary tumor, preoperative MRI</td>
<td>Pre-operative diagnosis of lymph node metastasis</td>
<td>Manual segmentation, radiomics analysis</td>
<td>AUC: 0.878</td>
</tr>
<tr>
<td>Gao et al [63]</td>
<td>GC, surgery</td>
<td>627 cases, 17340 images</td>
<td>Lymph nodes, preoperative CT</td>
<td>Pre-operative diagnosis of lymph node metastasis</td>
<td>Manual segmentation, deep learning</td>
<td>AUC: 0.9541</td>
</tr>
<tr>
<td>Huang et al [64]</td>
<td>GC, surgery</td>
<td>Primary tumor, preoperative CT</td>
<td>Pre-operative diagnosis of peritoneal metastasis</td>
<td>Manual segmentation, CNN</td>
<td>Ongoing, retrospective cross-sectional study</td>
<td></td>
</tr>
</tbody>
</table>

GC: Gastric cancer; CT: Computed tomography; AUC: Area under the curve; LNM: Lymph node metastasis; DFS: Disease-free survival; MRI: Magnetic resonance imaging; CNN: Convolutional neural network.

more accurate than that by normal nomograms (T and N stages and differentiation). Previous studies investigated the prediction of malignant potential using a radiomics approach. Zhang et al [65] evaluated the diagnostic accuracy of CT radiomics models for predicting post-operative recurrence in GC patients, and the AUC of the models were 0.806-0.831. Li et al [66] reported CT radiomics models for predicting lymph node metastasis, with an AUC of 0.82-0.84. Li et al [67] also developed CT radiomic models with the ability to predict the pathological status and survival with high accuracy. Wang et al [68] analyzed primary tumors on CT images of the arterial phase, portal phase, and delay phase for the discrimination of intestinal-type GC by a radiomics approach. The AUC of their model was 0.904. Jiang et al [69] described a radiomics model of PET-CT for predicting survival. The C-indexes of this model for overall survival and disease-free survival were 0.786 and 0.800, respectively. A radiomics analysis of MRI for GC has also been conducted. Chen et al [70] examined the heterogeneity of primary tumors on MRI using a radiomics approach, and showed that the model was useful for predicting the N stage.

**CNN model:** Gao et al [71] developed a CNN model of CT for predicting lymph node metastasis. Radiologists initially labeled upper abdominal-enhanced CT images of metastatic lymph nodes. CNN models were then constructed using the labeled image data, and the AUC of the model was 0.954. Huang et al [72] described a protocol for predicting peritoneal metastasis using CNN models, and this research is ongoing.

**Colorectal cancer**

Treatment strategies for lower rectal cancer (LRC) have recently been attracting increasing attention because of the difficulties associated with achieving curative treatment. Therefore, many researchers have targeted LRC patients for the development of AI models for radiological diagnoses. The aims of a recent AI study on CRC were (1) the automatic detection or segmentation of primary tumors; (2) the
prediction of treatment responses; and (3) the prediction of malignant potential.

**Automatic detection or segmentation of primary tumors**: Trebeschi et al\(^\text{[64]}\) reported a CNN model for the automatic segmentation of primary tumors on MRI. CNN models were trained using T2-weighted images (T2WI) and diffusion-weighted images with primary tumor labeling by expert radiologists. The CNN model showed high segmentation accuracy, with a dice similarity coefficient (DSC) of 0.68-0.70. The AUC of the resulting probability maps was 0.99. Two CNN models were also developed for the automatic segmentation of primary tumors on T2WIs, with DSC of 0.82 and 0.74, respectively\(^\text{[67,68]}\). Men et al\(^\text{[69]}\) attempted to develop CNN models for automatic segmentation on CT images with an application to the delineation of the clinical target volume (CTV) and surrounding organs for radiotherapy. The mean DSC values of the models were 87.7% for the CTV, 93.4% for the bladder, 92.1% for the left femoral head, 92.3% for the right femoral head, 65.3% for the intestines, and 61.8% for the colon.

**Prediction of treatment responses**: Shayesteh et al\(^\text{[74]}\) reported radiomics models predicting treatment responses to neo-adjuvant chemoradiotherapy. Primary tumors on MRI T2WI were manually segmented and an image analysis of the data, shape, texture as well as a histogram analysis were performed. The relationship between the pathological features and treatment responses to CRT was assessed by a machine learning approach, which revealed that the AUC and accuracy of the model were 95 and 90%, respectively. Shi et al\(^\text{[70]}\) and Ferrari et al\(^\text{[71]}\) also described the efficacy of radiomics models for predicting CRT responses using pre-treatment, mid-radiation, and post-treatment MRI (AUC for predicting a complete response (CR): 0.83 and 0.86, respectively). Bibault et al\(^\text{[72]}\) compared the diagnostic accuracy of several models, Cox’s regression, CNN, and SVM for predicting CR in pre-operative CRT using CT data. CNN exhibited the ability to predict CR with the highest accuracy (80%). A radiomics model for predicting chemotherapeutic responses has also been reported. Derclle et al\(^\text{[73]}\) demonstrated that their radiomic model using CT images successfully predicted sensitivity to anti-EGFR therapy (AUC: 0.80).
Prediction of malignant potential: Ding et al. developed AI models to predict lymphatic node metastasis using pre-operative MRI. CNN models were constructed using MRI lymph node images manually labeled by radiologists. They compared the diagnostic accuracy of CNN and a radiologist for predicting lymph node metastasis. As a result, CNN was more accurate than radiologists in identifying pelvic metastatic lymph nodes. A model for predicting gene profiles was also reported. These research methods are generally called radiogenomics. Taguchi et al. showed that a machine learning model using a texture analysis of CT images and SUV values of PET-CT predicted KRAS mutations with high accuracy (AUC: 0.82).

CURRENT ISSUES AND FUTURE PERSPECTIVES

AI research for endoscopy

The majority of studies previously reported that a CAD system using AI for endoscopy had the ability to diagnose gastrointestinal tumors with high accuracy; however, there were many limitations. Researchers were more likely to use high-quality endoscopic images to construct AI models, which cannot always be acquired in clinical settings. Furthermore, outcome indicators for clinical applications have not yet been defined. Therefore, parameters to assess the functional performance of AI models need to be established. In addition, the majority of studies have been retrospective in nature using still images from non-clinical settings. These conditions do not mimic real-time clinical settings, in which endoscopists often encounter difficult-to-analyze images in daily practice. Moreover, it currently remains unclear whether AI models will enhance medical performance, reduce medical costs, and increase the satisfaction of patients and medical staff in clinical settings. Another limitation is that many clinicians and clinical researchers do not have sufficient knowledge to understand AI systems; therefore, non-AI experts as well as medical journal reviewers may encounter difficulties when assessing research on AI and its applications. Furthermore, the number of medical staff with the skill to educate physicians on AI is very limited.

Nevertheless, once these limitations are resolved, CAD systems using AI will markedly improve diagnostic quality in endoscopic examinations. CAD systems for endoscopy are expected to serve as a second observer during real-time endoscopy, facilitating the detection of more neoplasms by endoscopists. Some CAD systems may also provide “optical biopsies” to differentiate the types of colon polyps. Therefore, CAD systems have a promising future in the effective training of junior endoscopists as assistant observers.

AI research for pathology

Previous studies reported that AI models distinguish structures in tissues and detect cancerous regions with high accuracy. Furthermore, survival may be predicted using image analyses by AI. However, there are also a number of limitations in research. AI models are educated using pathological images of cancer tissue labeled by pathologists. However, interobserver disagreement in pathological diagnoses commonly occurs between pathologists. Therefore, the quality of teaching data varied in each study. Furthermore, the majority of AI models were constructed using a small cohort. It might be possibility non-reproducible laboratory-specific machine learning methods. In addition, the clinical use of AI models requires a digital slide scanner, image storage, maintenance contracts, image analysis software, and IT support systems, which may be expensive in clinical settings. Moreover, many pathologists and technicians do not have sufficient knowledge to understand AI systems. Therefore, the recruitment of AI experts to introduce AI systems into clinical settings is needed for education and the adjustment of systems to different clinical settings.

Despite these limitations, whole-slide scanning using AI models, such as the TMA method, is advantageous for pathologists and clinicians. This method may be a second observer in the prevention of false diagnoses by pathologists and the teaching of trainees. Furthermore, the heterogeneities of cancer tissue cannot be precisely evaluated by the human eyes of pathologists. Therefore, the assessment of cancer tissue using AI models is a novel research method beyond human cognition that is expected to predict proteomics, genomics, and the molecular signaling pathways of tumors as precision medicine by cancer genome sequencing.
AI research for radiology

Previous studies reported the efficacy of automatic segmentation or diagnosis in solid malignant tumors. However, difficulties are associated with automatic segmentation by AI models in the field of gastrointestinal cancer because of large individual differences in imaging features of the gastrointestinal tract, except for the rectum. The radiomics approach represents an attractive method for detecting malignant potential and imaging biomarkers for precision medicine through image analyses of intratumor heterogeneity. However, a number of limitations need to be considered. The manual or semi-automatic segmentation of tumors is generally needed in the radiomics approach. Interobserver variability in manual segmentation often occurs in this process, resulting in the poor reproducibility of data by the radiomics model. Furthermore, previous studies demonstrated that radiomic features may be affected by a number of parameters, such as the scanning equipment, image pre-processing, acquisition protocols, image reconstruction algorithms, and delineation. In addition, although researchers of radiology or AI experts are knowledgeable about radiomics and AI models, they often cannot target the clinical task that needs to be improved for clinicians or patients in clinical settings. However, clinicians are not sufficiently aware of AI, and few reviewers of scientific literature on clinical medicine often are developing AI models or are able to judge research involving AI. Therefore, a multidisciplinary team needs to be introduced into research and medical teams to promote AI-supported medicine.

Despite these limitations, radiomic models for the image diagnosis or prediction of malignancy have the potential to support clinical teams for more accurate and rapid diagnoses. These models may increase patient satisfaction levels for homogenized diagnostic accuracy. Moreover, radiogenomics may have a major impact on precision medicine. Non-invasive assessments of the entire tumor tissue may be possible, without having to rely on a single biopsy to represent all cancer lesions within a patient. As further information becomes available on these imaging markers, the characteristics of cancers will be elucidated in more detail. Therefore, the radiomics approach will enhance the treatment effects of molecular biological approaches for oncological precision medicine.

DISCUSSION

AI will be an important component of diagnostic methods to diagnosis patient disease, determine most appropriate treatments, and predict prognosis and drug resistance. A lot of research methods have been developed with the aims and found to have varying levels of performance. For clinical use of disease diagnosis, AI seems valuable for use in endoscopy, where it could increase detection of benign polyp and malignant tumor. Meanwhile, AI may be useful to analysis intratumor heterogeneity of radiological and pathological images in order to predict malignant potentials, such as the prognosis of patients and therapeutic effects. Our minireview covered only articles listed in MEDLINE, and might have missed some literatures in medical image analysis journals and computer science. Despite of the limitation, AI has become an important part of clinical cancer research in recent years.

There is no turning back for the development of AI in gastrointestinal cancer, and future implications are large. However, some limitations that require caution should be recognized. Most studies were performed using low-quality datasets from pre-clinical studies. Furthermore, AI algorithms are often considered to be black-box models. The difficulty in understanding the process of AI decision may prevent physicians from finding the potential confounding factors. Ethical challenge is one of the problems to be considered. In the present AI system, AI is not aware of the human preferences or legal liabilities. Therefore, medical staff will have to make decisions for patients according to their preferences, environment, and ethics. AI will not completely replace doctors, and computer technology and medical staff will always have to work together. However, the diagnostic accuracy of AI systems has markedly increased and may detect novel biomarkers that cannot be identified by the human eye or in human-aid analyses. AI systems will be introduced into general hospitals in the near future under the management of multidisciplinary teams consisting of medical staff and AI experts.
CONCLUSION

We reviewed the recent published literatures on AI in gastrointestinal cancer, suggesting that AI may be used to accurately diagnose clinical images, identify new therapeutic targets, and process clinical data from large patient datasets. Although the physicians must recognize the limitations of AI diagnostic system, AI-assisted medical systems will become a promising tool for gastrointestinal cancer.

REFERENCES


AIG

https://www.wjgnet.com

83

November 28, 2020

Volume 1

Issue 4

Kudou M et al. AI in GI cancer


Kudou M et al. AI in GI cancer

10.1111.his.13838


