REVIEW

2434 Role of modern radiotherapy in managing patients with hepatocellular carcinoma
Chen LC, Lin HY, Hung SK, Chiou WY, Lee MS

2458 Open reading frame 3 protein of hepatitis E virus: Multi-function protein with endless potential
Yang YL, Nan YC

2474 Breakthroughs and challenges in the management of pediatric viral hepatitis
Nicastro E, Norsa L, Di Giorgio A, Indolfi G, D'Antiga L

MINIREVIEWS

2495 Pancreatitis after endoscopic retrograde cholangiopancreatography: A narrative review
Ribeiro IB, do Monte Junior ES, Miranda Neto AA, Proença IM, de Moura DTH, Minata MK, Ide E, dos Santos MEL, Luz GO, Matuguma SE, Cheng S, Baracat R, de Moura EGH

2507 RON in hepatobiliary and pancreatic cancers: Pathogenesis and potential therapeutic targets
Chen SL, Wang GP, Shi DR, Yao SH, Chen KD, Yao HP

2521 Evolving role of endoscopy in inflammatory bowel disease: Going beyond diagnosis
Núñez FP, Krugliak Cleveland N, Quera R, Rubin DT

2531 Deep learning for diagnosis of precancerous lesions in upper gastrointestinal endoscopy: A review
Yan T, Wong PK, Qin YY

2545 State of machine and deep learning in histopathological applications in digestive diseases
Kobayashi S, Saltz JH, Yang VW

2576 COVID-19 in normal, diseased and transplanted liver
Signorello A, Lenci I, Milana M, Grassi G, Baiocchi L

ORIGINAL ARTICLE

Basic Study

2586 Upregulation of long noncoding RNA W42 promotes tumor development by binding with D8N1 in hepatocellular carcinoma

Retrospective Cohort Study

2603 Understanding celiac disease monitoring patterns and outcomes after diagnosis: A multinational, retrospective chart review study
### Contents

**World Journal of Gastroenterology**  
Weekly Volume 27 Number 20 May 28, 2021

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2615</td>
<td>Development and validation of a prognostic model for patients with hepatorenal syndrome: A retrospective cohort study</td>
<td>Sheng XY, Lin FY, Wu J, Cao HC</td>
</tr>
<tr>
<td></td>
<td><strong>Observational Study</strong></td>
<td></td>
</tr>
<tr>
<td>2630</td>
<td>Inflammatory bowel disease in Tuzla Canton, Bosnia-Herzegovina: A prospective 10-year follow-up</td>
<td>Tulumović E, Salkić N, Tulumović D</td>
</tr>
<tr>
<td></td>
<td><strong>META-ANALYSIS</strong></td>
<td></td>
</tr>
<tr>
<td>2643</td>
<td>Association between oral contraceptive use and pancreatic cancer risk: A systematic review and meta-analysis</td>
<td>Ilic M, Milicic B, Ilic I</td>
</tr>
<tr>
<td></td>
<td><strong>CASE REPORT</strong></td>
<td></td>
</tr>
<tr>
<td>2657</td>
<td>Cyclophosphamide-associated enteritis presenting with severe protein-losing enteropathy in granulomatosis with polyangiitis: A case report</td>
<td>Sato H, Shirai T, Fujii H, Ishii T, Harigae H</td>
</tr>
</tbody>
</table>
ABOUT COVER

Editorial Board Member of *World Journal of Gastroenterology*, Fernando J Castro, MD, AGAF, FACG, Gastroenterology Training Program Director, Cleveland Clinic Florida, 2950 Cleveland Clinic Blvd, Weston, FL 33331, United States. castrof@ccf.org

AIMS AND SCOPE

The primary aim of *World Journal of Gastroenterology* (WJG, *World J Gastroenterol*) is to provide scholars and readers from various fields of gastroenterology and hepatology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online. WJG mainly publishes articles reporting research results and findings obtained in the field of gastroenterology and hepatology and covering a wide range of topics including gastroenterology, hepatology, gastrointestinal endoscopy, gastrointestinal surgery, gastrointestinal oncology, and pediatric gastroenterology.

INDEXING/ABSTRACTING

The WJG is now indexed in Current Contents®/Clinical Medicine, Science Citation Index Expanded (also known as SciSearch®), Journal Citation Reports®, Index Medicus, MEDLINE, PubMed, PubMed Central, and Scopus. The 2020 edition of Journal Citation Report® cites the 2019 impact factor (IF) for WJG as 3.665; IF without journal self cites: 3.534; 5-year IF: 4.048; Ranking: 35 among 88 journals in gastroenterology and hepatology; and Quartile category: Q2. The WJG’s CiteScore for 2019 is 7.1 and Scopus CiteScore rank 2019: Gastroenterology is 17/137.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: Ji-Hong Liu; Production Department Director: Yun-Xiaojuan Wu; Editorial Office Director: Ze-Mao Gong.

NAME OF JOURNAL

*World Journal of Gastroenterology*

ISSN

ISSN 1007-9327 (print) ISSN 2219-2840 (online)

LAUNCH DATE

October 1, 1995

FREQUENCY

Weekly

EDITORS-IN-CHIEF

Andrzej S Tarnawski, Subrata Ghosh

EDITORIAL BOARD MEMBERS

http://www.wjgnet.com/1007-9327/editorialboard.htm

PUBLICATION DATE

May 28, 2021

COPYRIGHT

© 2021 Baishideng Publishing Group Inc

INSTRUCTIONS TO AUTHORS

https://www.wjgnet.com/bpg/gerinfo/204

GUIDELINES FOR ETHICS DOCUMENTS

https://www.wjgnet.com/bpg/GerInfo/287

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH

https://www.wjgnet.com/bpg/gerinfo/240

PUBLICATION ETHICS

https://www.wjgnet.com/bpg/GerInfo/288

PUBLICATION MISCONDUCT

https://www.wjgnet.com/bpg/gerinfo/208

ARTICLE PROCESSING CHARGE

https://www.wjgnet.com/bpg/gerinfo/242

STEPS FOR SUBMITTING MANUSCRIPTS

https://www.wjgnet.com/bpg/GerInfo/239

ONLINE SUBMISSION

https://www.f6publishing.com

© 2021 Baishideng Publishing Group Inc. All rights reserved. 7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA

E-mail: bpgoffice@wjgnet.com  https://www.wjgnet.com
Deep learning for diagnosis of precancerous lesions in upper gastrointestinal endoscopy: A review

Tao Yan, Pak Kin Wong, Ye-Ying Qin

ORCID number: Tao Yan 0000-0002-8929-015X; Pak Kin Wong 0000-0002-7623-6904; Ye-Ying Qin 0000-0002-7779-1045.

Author contributions: Wong PK and Yan T contributed to concept design and drafted the manuscript; Yan T and Qin YY collected the data; All the authors have approved the final version of the manuscript.

Supported by The Science and Technology Development Fund, Macau SAR, No. 0021/2019/A.

Conflict-of-interest statement: All authors declare no conflict of interest.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/Licenses/by-nc/4.0/

Manuscript source: Invited

Abstract

Upper gastrointestinal (GI) cancers are the leading cause of cancer-related deaths worldwide. Early identification of precancerous lesions has been shown to minimize the incidence of GI cancers and substantiate the vital role of screening endoscopy. However, unlike GI cancers, precancerous lesions in the upper GI tract can be subtle and difficult to detect. Artificial intelligence techniques, especially deep learning algorithms with convolutional neural networks, might help endoscopists identify the precancerous lesions and reduce interobserver variability. In this review, a systematic literature search was undertaken of the Web of Science, PubMed, Cochrane Library and Embase, with an emphasis on the deep learning-based diagnosis of precancerous lesions in the upper GI tract. The status of deep learning algorithms in upper GI precancerous lesions has been systematically summarized. The challenges and recommendations targeting this field are comprehensively analyzed for future research.

Key Words: Artificial intelligence; Deep learning; Convolutional neural network; Precancerous lesions; Endoscopy

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.
Lower gastrointestinal (GI) cancers, mainly including gastric cancer (8.2% of total cancer deaths) and esophageal cancer (5.3% of total cancer deaths), are the leading cause of cancer-related deaths worldwide[1]. Previous studies have shown that upper GI cancers always go through the stages of precancerous lesions, which can be defined as common conditions associated with a higher risk of developing cancers over time[2-4]. The detection of the precancerous lesions before cancer occurs could significantly reduce morbidity and mortality rates[5,6]. Currently, the main approach for the diagnosis of disorders or issues in the upper GI tract is endoscopy[7,8]. Compared with GI cancers, which usually show typical morphological characteristics, the precancerous lesions often appear in flat mucosa and exhibit few morphological changes. Manual screening through endoscopy is labor-intensive, time-consuming and relies heavily on clinical experience. Computer-assisted diagnosis based on artificial intelligence (AI) can overcome these dilemmas.

Over the past few decades, AI techniques such as machine learning (ML) and deep learning (DL) have been widely used in endoscopic imaging to improve the diagnostic accuracy and efficiency of various GI lesions[9-13]. The exact definition of AI, ML and DL can be misunderstood by physicians. AI, ML and DL are overlapping disciplines (Figure 1). AI is a hierarchy that encompasses ML and DL; it describes a computerized solution to address the issues of human cognitive defined by McCarthy in 1956[14]. ML is a subset of AI in which algorithms can execute complex tasks, but it needs handcrafted feature extraction. ML originated around the 1980s and focuses on patterns and inference[15]. DL is a subset of ML and became feasible in the 2010s; it is focused specifically on deep neural networks. A convolutional neural network (CNN) is the primary DL algorithm for image processing[16,17].

The diagnostic process of an AI model is similar to the human brain. We take our previous research as an example to illustrate the diagnostic process of ML, DL and human experts. When an input image with gastric intestinal metaplasia (GIM), a precancerous lesion of gastric cancer, is fed into the ML system, it usually needs a manual feature extraction step, while the handcrafted features are unable to discern slight variations in the endoscopic image (Figure 2)[15]. Unlike conventional ML algorithms, CNNs can automatically learn representative features from the endoscopic images[17]. When we apply a CNN model to detect GIM, it performs better than conventional ML models and is comparable to experienced endoscopists[18]. For a broad variety of image processing activities in endoscopy, CNNs also show excellent results, and some CNN-based algorithms have been used in clinical practice[19-21]. However, DL, especially CNN, has some limitations. First, DL requires a lot of data and easily leads to overfitting. Second, the diagnostic accuracy of DL relies on the training data, but the clinical data of different types of diseases are always imbalanced, which easily causes diagnosis bias. In addition, a DL model is complex and requires a huge calculation, so most researchers can only use the ready-made model.

Despite the above limitations, DL-based AI systems are revolutionizing GI endoscopy. While there are several surveys on DL for GI cancers[9-13], no specific review on the application of DL in the endoscopic diagnosis of precancerous lesions is available in the literature. Therefore, the performance of DL on gastroenterology is summarized in this review, with an emphasis on the automatic diagnosis of precancerous lesions in the upper GI tract. GI cancers are out of the scope of this review. Specifically, we review the status of intelligent diagnoses of esophageal and gastric precancerous lesions. The challenges and recommendations based on the findings of the review are comprehensively analyzed to advance the field.
DL IN ENDOSCOPIC DETECTION OF PRECANCEROUS LESIONS IN ESOPHAGEAL MUCOSA

Esophageal cancer is the eighth most prevalent form of cancer and the sixth most lethal cancer globally[1]. There are two major subtypes of esophageal cancer: Esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC)[22]. Esophageal squamous dysplasia (ESD) is believed to be the precancerous
lesion of ESCC[23-25], and Barrett’s esophagus (BE) is the identifiable precancerous lesion associated with EAC[4,26]. Endoscopic surveillance is recommended by GI societies to enable early detection of the two precancerous lesions of esophageal cancer[25,27].

However, the current endoscopic diagnosis methods for patients with BE, such as random 4-quadrant biopsy, laser-induced endomicroscopy, image enhanced endoscopy, etc., have disadvantages concerning the learning curve, cost, interobserver variability and time-consuming problems[28]. The current standard for identifying ESD is Lugol’s chromoendoscopy, but it shows poor specificity[29]. Besides, iodine staining often presents a risk of allergic reactions. To overcome these challenges, DL-based AI systems have been established to help endoscopists identify ESD and BE.

**DL in ESD**

Low-grade and high-grade intraepithelial neoplasms, collectively referred to as ESD, are deemed as precancerous lesions of ESCC. Early and accurate detection of ESD is essential but also full of challenges[23-25]. DL is reliably able to depict ESD in real-time upper endoscopy. Cai et al.[30] designed a novel computer-assisted diagnosis system to localize and identify early ESCC, including low-grade and high-grade intraepithelial neoplasia, through real-time white light endoscopy (WLE). The system achieved a sensitivity, specificity and accuracy of 97.8%, 85.4% and 91.4%, respectively. They also demonstrated that when referring to the results of the system, the overall diagnostic capability of the endoscopist has been increased. This research paved the way for the real-time diagnosis of ESD and ESCC. Following this work, Guo et al.[31] applied 6473 narrow band (NB) images to train a real-time automated computer-assisted diagnosis system to support non-experts in the detection of ESD and ESCC. The system serves as a “second observer” in an endoscopic examination and achieves a sensitivity of 98.04% and specificity of 95.30% on NB images. The per-frame sensitivity was 96.10% for magnifying narrow band imaging (M-NBI) videos and 60.80% for non-M-NBI videos. The per lesion sensitivity was 100% in M-NBI videos.

**DL in BE**

BE is a disorder in which the lining of the esophagus is damaged by gastric acid. The critical purpose of endoscopic Barrett’s surveillance is early detection of BE-related dysplasia[4,26-28]. Recently, there have been many studies on the DL-based diagnosis of BE, and we review some representative studies. de Groof et al.[32] performed one of the first pilot studies to assess the performance of a DL-based system during live endoscopic procedures of patients with or without BE dysplasia. The system demonstrated 90% accuracy, 91% sensitivity and 89% specificity in a per-level analysis. Following up this work, they improved this system using stepwise transfer learning and five independent endoscopy data sets. The enhanced system obtained higher accuracy than non-expert endoscopists and with comparable delineation performance[33]. Furthermore, their team also demonstrated the feasibility of a DL-based system for tissue characterization of NBI endoscopy in BE, and the system achieved a promising diagnostic accuracy[34].

Hashimoto et al.[35] borrowed from the Inception-ResNet-v2 algorithm to develop a model for real-time classification of early esophageal neoplasia in BE, and they also applied YOLO-v2 to draw localization boxes around regions classified as dysplasia. For detection of neoplasia, the system achieved a sensitivity of 96.4%, specificity of 94.2% and accuracy of 95.4%. Hussein et al.[36] built a CNN model to diagnose dysplastic BE mucosa with a sensitivity of 88.3% and specificity of 80.0%. The results preliminarily indicated that the diagnostic performance of the CNN model was close to that of experienced endoscopists. Ebigbo et al.[37] exploited the use of a CNN-based system to classify and segment cancer in BE. The system achieved an accuracy of 89.9% in 14 patients with neoplastic BE.

DL has also achieved excellent results in distinguishing multiple types of esophageal lesions, including BE. Liu et al.[38] explored the use of a CNN model to distinguish esophageal cancers from BE. The model was trained and evaluated on 1272 images captured by WLE. After pre-processing and data augmentation, the average sensitivity, specificity and accuracy of the CNN model were 94.23%, 94.67% and 85.83%, respectively. Wu et al.[39] developed a CNN-based framework named ELNet for automatic esophageal lesion (i.e. EAC, BE and inflammation) classification and segmentation, the ELNet achieved a classification sensitivity of 90.34%, specificity of 97.18% and accuracy of 96.28%. The segmentation sensitivity, specificity and accuracy were 80.18%, 96.55% and 94.62%, respectively. A similar study was proposed by Ghatwary et al.[40], who applied a CNN algorithm to detect BE, EAC and ESCC from...
endoscopic videos and obtained a high sensitivity of 93.7% and a high F-measure of 93.2%.

The studies exploring the creation of DL algorithms for the diagnosis of precancerous lesions in esophageal mucosa are summarized in Table 1.

---

**DL IN ENDOSCOPIC DETECTION OF PRECANCEROUS LESIONS IN GASTRIC MUCOSA**

Gastric cancer is the fifth most prevalent form of cancer and the third most lethal cancer globally[1]. Even though the prevalence of gastric cancer has declined during the last few decades, gastric cancer remains a significant clinical problem, especially in developing countries. This is because most patients are diagnosed in late stages with poor prognosis and restricted therapeutic choices[41]. The pathogenesis of gastric cancer involves a series of events starting with Helicobacter pylori-induced (H. pylori-induced) chronic inflammation, progressing towards atrophic gastritis, GIM, dysplasia and eventually gastric cancer[42]. Patients with the precancerous lesions (e.g., H. pylori-induced chronic inflammation, atrophic gastritis, GIM and dysplasia) are at considerable risk of gastric cancer[3,6,43]. It has been argued that the detection of such precancerous lesions may significantly reduce the incidence of gastric cancer. However, endoscopic examination is difficult to identify these precancerous lesions, and the diagnostic result also has high interobserver variability due to their subtle morphological changes in the mucosa and lack of experienced endoscopists[44,45]. Currently, many researchers are trying to use DL-based methods to detect gastric precancerous lesions; here, we review these studies in detail.

**DL in H. pylori infection**

Most of the gastric precancerous lesions are correlated with long-term infections with H. pylori[46]. Shichijo et al[47] performed one of the pioneering studies to apply CNNs in the diagnosis of H. pylori infection. The CNNs were built on GoogLeNet and trained on 32208 WLE images. One of their CNN models has higher accuracy than endoscopists. The study showed the feasibility of using CNN to diagnose H. pylori from endoscopic images. After this study, Itoh et al[48] developed a CNN model to detect H. pylori infection in WLE images and showed a sensitivity of 86.7% and specificity of 96.7% in the test dataset. A similar model was developed by Zheng et al[49] to evaluate H. pylori infection status, and the per-patient sensitivity, specificity and accuracy of the model were 91.6%, 98.6% and 93.8%, respectively. Besides WLE, blue laser imaging-bright and linked color imaging systems were prospectively applied by Nakashima et al[50] to collect endoscopic images. With these images, they fine-tuned a pre-trained GoogLeNet to predict H. pylori infection status. As compared with linked color imaging, the model achieved the highest sensitivity (96.7%) and specificity (86.7%) when using blue laser imaging-bright. Nakashima et al[51] also did a single center prospective study to build a CNN model to identify the status of H. pylori in uninfected, currently infected and post-eradication patients. The area under the receiver operating characteristic curve for the uninfected, currently infected and post-eradication categories was 0.90, 0.82 and 0.77, respectively.

**DL in atrophic gastritis**

Atrophic gastritis is a form of chronic inflammation of the gastric mucosa; accurate endoscopic diagnosis is difficult[52]. Guimarães et al[53] reported the application of CNN to detect atrophic gastritis; the system achieved an accuracy of 93% and performed better than expert endoscopists. Recently, another CNN-based system for detecting atrophic gastritis was reported by Zhang et al[54]. The CNN model was trained and tested on a dataset containing 3042 images with atrophic gastritis and 2428 without atrophic gastritis. The diagnostic accuracy, sensitivity and specificity of the model were 94.2%, 94.5% and 94.0%, respectively, which were better than those of the experts. More recently, Horiuchi et al[55] explored the diagnostic ability of the CNN model to distinguish early gastric cancer and gastritis through M-NBI; the 22-layer CNN was built on GoogleNet and pretrained using 2570 endoscopic images, and the sensitivity, specificity and accuracy on 258 images were 95.4%, 71.0% and 85.3%, respectively. Except for high sensitivity, the CNN model also showed an overall test speed of 0.02 s per image, which was faster than human experts.
**Table 1 Summary of studies using deep learning for detection of esophageal precancerous lesions**

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Year</th>
<th>Imaging</th>
<th>Study design</th>
<th>Study aim</th>
<th>DL model</th>
<th>Dataset</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gao et al[31]</td>
<td>2020</td>
<td>NBI, M-NBI</td>
<td>Retrospective</td>
<td>Detection of precancerous lesions and early ESCC</td>
<td>SegNet</td>
<td>13144 images and 168865 video frames</td>
<td>Sensitivity: 96.10% for M-NBI videos, 60.80% for non-M-NBI videos, 98.04% for images. Specificity: 99.00% for non-M-NBI/M-NBI videos, 95.30% for images</td>
</tr>
<tr>
<td>Struyvenberg et al[34]</td>
<td>2021</td>
<td>NBI</td>
<td>Retrospective</td>
<td>Detection of Barrett’s neoplasia</td>
<td>ResNet/U-Ne</td>
<td>2677 images</td>
<td>Sensitivity: 88%. Specificity: 78%. Accuracy: 84%</td>
</tr>
<tr>
<td>Hashimoto et al[35]</td>
<td>2020</td>
<td>WLE, NBI</td>
<td>Retrospective</td>
<td>Recognition of early neoplasia in BE</td>
<td>Inception-ResNet-v2, YOLO-v2</td>
<td>2290 images</td>
<td>Sensitivity: 96.4%. Specificity: 94.2%. Accuracy: 95.4%</td>
</tr>
<tr>
<td>Ebigbo et al[37]</td>
<td>2020</td>
<td>WLE</td>
<td>Retrospective</td>
<td>Diagnosis of early EAC in BE</td>
<td>DeepLab V.3+, Resnet101</td>
<td>191 images</td>
<td>Sensitivity: 83.7%. Specificity: 100%. Accuracy: 89.9%</td>
</tr>
<tr>
<td>Liu et al[38]</td>
<td>2020</td>
<td>WLE</td>
<td>Retrospective</td>
<td>Detection of esophageal cancer from precancerous lesions</td>
<td>Inception-ResNet</td>
<td>1272 images</td>
<td>Sensitivity: 94.23%. Specificity: 94.67%. Accuracy: 95.83%</td>
</tr>
<tr>
<td>Wu et al[39]</td>
<td>2021</td>
<td>WLE</td>
<td>Retrospective</td>
<td>Automatic classification and segmentation for esophageal lesions</td>
<td>ELNet</td>
<td>1051 images</td>
<td>Classification sensitivity: 90.34%. Classification specificity: 97.18%. Classification accuracy: 96.28%. Segmentation sensitivity: 80.18%, Segmentation specificity: 96.55%, Segmentation accuracy: 94.62</td>
</tr>
<tr>
<td>Ghatwary et al[40]</td>
<td>2021</td>
<td>WLE</td>
<td>Retrospective</td>
<td>Detection of esophageal abnormalities from endoscopic videos</td>
<td>DenseConvLstm, Faster R-CNN</td>
<td>42425 video frames</td>
<td>Sensitivity: 93.7%. F-measure: 93.2%</td>
</tr>
</tbody>
</table>

BE: Barrett’s esophagus; DL: Deep learning; EAC: Esophageal adenocarcinoma; ESCC: Esophageal squamous cell carcinoma; M-NBI: Magnifying narrow band imaging; NBI: Narrow band imaging; WLE: White light endoscopy.

**DL in GIM**

GIM is the replacement of gastric-type mucinous epithelial cells with intestinal-type cells, which is a precancerous lesion with a worldwide prevalence of 25%[56]. The morphological characteristics of GIM are subtle and difficult to observe, so the manual diagnosis of GIM is full of challenges. Wang et al[57] reported the first instance of an AI system for localizing and identifying GIM from WLE images. The system achieved a high classification accuracy and a satisfactory segmentation result. A recent study developed a CNN-based diagnosis system that can detect atrophic gastritis and GIM from WLE images[58], and the detection sensitivity and specificity for atrophic gastritis were 87.2% and 91.1%, respectively. For detection of GIM, the system also achieved a sensitivity of 90.3% and a specificity of 93.7%. Recently, our team also developed a novel DL-based diagnostic system for detection of GIM in endoscopic images[18]. The difference from the previous research is that our system is composed of three independent CNNs, which can identify GIM from either NBI or M-NBI. The per-patient sensitivity, specificity and accuracy of the system were 91.9%, 86.0% and 88.8%, respectively. The diagnostic performance showed no significant differences as compared with human experts. Our research showed that the integration of NBI and M-NBI into the DL system could achieve satisfactory diagnostic performance for GIM.
DL in gastric dysplasia

Gastric dysplasia is the penultimate step of gastric carcinogenesis, and accurate diagnosis of this lesion remains controversial[59]. To accurately classify advanced low-grade dysplasia, high-grade dysplasia, early gastric cancer and gastric cancer, Cho et al[60] established three CNN models based on 5017 endoscopic images. They found that the Inception-Resnet-v2 model performed the best, while it showed lower five-class accuracy compared with the endoscopists (76.4% vs 87.6%). Inoue et al[61] constructed a detection system using the Single-Shot Multibox Detector, which can automatically detect duodenal adenomas and high-grade dysplasia from WLE or NBI. The system detected 94.7% adenomas and 100% high-grade dysplasia on a dataset containing 1080 endoscopic images within only 31 s. Although most of the AI-assisted system can achieve high accuracy on endoscopic diagnosis, no study has investigated the role of AI in the training of junior endoscopists. To evaluate the role of AI in the training of junior endoscopists in predicting histology of endoscopic gastric lesions, including dysplasia, Lui et al[62] designed and validated a CNN classifier based on 3000 NB images. The classifier achieved an overall accuracy of 91.0%, sensitivity of 97.1% and specificity of 85.9%, which was superior to all junior endoscopists. They also demonstrated that with the feedback from the CNN classifier, the learning curve of junior endoscopists was improved in predicting histology of gastric lesions.

The studies exploring the creation of DL algorithms for the diagnosis of precancerous lesions in gastric mucosa are summarized in Table 2.

CHALLENGES AND RECOMMENDATIONS

AI has gained much attention in recent years. In the field of GI endoscopy, DL is also a promising innovation in the identification and characterization of lesions[9-13]. Many successful studies have focused on GI cancers. Accurate detection of precancerous lesions such as ESD, BE, H. pylori-induced chronic inflammation, atrophic gastritis, GIM and gastric dysplasia can greatly reduce the incidence of cancers and require less cancer treatment. DL-assisted detection of these precancerous lesions has increasingly emerged in the last 5 yrs. To perform a systematic review of the status of DL for diagnosis of precancerous lesions of the upper GI tract, we conducted a comprehensive search for all original publications on this target between January 1, 2017 and December 30, 2020. A variety of published papers has verified the outstanding performance of DL-assisted systems, several challenges remain from the viewpoint of physicians and algorithm engineers. The challenges and our recommendations on future research directions are outlined below.

Prospective studies and clinical verification

The current literature reveals that most studies were designed in a retrospective manner with a strong probability of bias. In these retrospective studies, researchers tended to collect high-quality endoscopic images that showed typical characteristics of the detected lesions from a single medical center, while they excluded common low-quality images. This kind of selection bias may jeopardize the precision and lead to lower generalization of the DL models. Thus, data collected from multicenter studies with uninformative frames are necessary to build robust DL models, and prospective studies are needed to properly verify the accuracy of AI in clinical practice.

Handling of overfitting

Overfitting means an AI model performs well on the training set but has high error on unseen data. The deep CNN architectures usually contain several convolutional layers and fully connected layers, which produce millions of parameters that easily lead to strong overfitting[16,17]. Training these parameters needs large-scale well-annotated data. However, well-annotated data are costly and hard to obtain in the clinical community. Possible solutions for overcoming the lack of well-annotated data to avoid overfitting mainly include data augmentation[63], transfer learning[17,64], semi-supervised learning[65] and data synthesis using generative adversarial networks[66].

Data augmentation is a common method to train CNNs to reduce overfitting[63]. According to current literature, almost all studies use data augmentation. Data augmentation is performed by using several image transformations such as random image rotation, flipping, shifting, scaling and their combinations are shown in Figure 3. Transfer learning involves transfer knowledge learned from a large source domain to a target domain[17,64]. This technique is usually performed by initializing the CNN using the weights pretrained on ImageNet dataset. As there are many
Prospective design was used in the field of medical image synthesis; the synthetic images can be used to provide auxiliary training to improve the performance of the CNN models [6,65,67,68]. Generative adversarial networks are widely used in the field of medical image synthesis; the synthetic images can be used to provide auxiliary training to improve the performance of the CNN models [66]. de Souza et al. [69] introduced generative adversarial networks to generate high-quality endoscopic images for BE, and these synthetic images were used to provide auxiliary training. The detection results suggested that with the help of these synthetic images the CNN model outperformed the ones over the original datasets. Although the above techniques were used to reduce overfitting to a certain extent, the single use of one type did not guarantee resolution of the overfitting problem. Integration of these techniques is a promising strategy, and it was verified in our research [18,70].

### Table 2 Summary of studies using deep learning for detection of gastric precancerous lesions

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Year</th>
<th>Imaging</th>
<th>Study design</th>
<th>Study aim</th>
<th>DL model</th>
<th>Dataset</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shibuchi et al. [47]</td>
<td>2017</td>
<td>WLE</td>
<td>Retrospective</td>
<td>Diagnosis of H. pylori infection</td>
<td>GoogLeNet</td>
<td>43689 images</td>
<td>Sensitivity: 88.9%; Specificity: 87.4%; Accuracy: 87.7%</td>
</tr>
<tr>
<td>Itoh et al. [48]</td>
<td>2018</td>
<td>WLE</td>
<td>Retrospective</td>
<td>Analysis of H. pylori infection</td>
<td>GoogLeNet</td>
<td>179 images</td>
<td>Sensitivity: 86.7%; Specificity: 86.7%</td>
</tr>
<tr>
<td>Zheng et al. [49]</td>
<td>2019</td>
<td>WLE</td>
<td>Retrospective</td>
<td>Evaluation of H. pylori infection status</td>
<td>ResNet-50</td>
<td>15484 images</td>
<td>Sensitivity: 91.6%; Specificity: 98.6%; Accuracy: 93.8%</td>
</tr>
<tr>
<td>Nakashima et al. [50]</td>
<td>2018</td>
<td>BLI-bright, LCI</td>
<td>Prospective</td>
<td>Prediction of H. pylori infection status</td>
<td>GoogLeNet</td>
<td>666 images</td>
<td>Sensitivity: 96.7%; Specificity: 86.7%</td>
</tr>
<tr>
<td>Nakashima et al. [51]</td>
<td>2020</td>
<td>WLE, LCI</td>
<td>Prospective</td>
<td>Diagnosis of H. pylori infection</td>
<td>--</td>
<td>13127 images</td>
<td>For currently infected patients, the sensitivity and specificity are 62.5% and 92.5%, respectively</td>
</tr>
<tr>
<td>Guimarães et al. [53]</td>
<td>2020</td>
<td>WLE</td>
<td>Retrospective</td>
<td>Diagnosis of atrophic gastritis</td>
<td>VGG16</td>
<td>270 images</td>
<td>Accuracy: 93%</td>
</tr>
<tr>
<td>Zhang et al. [54]</td>
<td>2020</td>
<td>WLE</td>
<td>Retrospective</td>
<td>Diagnosis of atrophic gastritis</td>
<td>DenseNet121</td>
<td>5470 images</td>
<td>Sensitivity: 94.5%; Specificity: 94.0%; Accuracy: 94.2%</td>
</tr>
<tr>
<td>Horiuchi et al. [55]</td>
<td>2020</td>
<td>M-NBI</td>
<td>Retrospective</td>
<td>Differentiation between early gastric cancer and gastritis</td>
<td>GoogLeNet</td>
<td>2826 images</td>
<td>Sensitivity: 95.4%; Specificity: 71.0%; Accuracy: 85.3%</td>
</tr>
<tr>
<td>Wang et al. [57]</td>
<td>2019</td>
<td>WLE</td>
<td>Retrospective</td>
<td>Localization and identification of GIM</td>
<td>DeepLab V.3+</td>
<td>200 images</td>
<td>Accuracy: 89.51%</td>
</tr>
<tr>
<td>Zheng et al. [58]</td>
<td>2020</td>
<td>WLE</td>
<td>Retrospective</td>
<td>Detection of atrophic gastritis and GIM</td>
<td>ResNet-50</td>
<td>3759 images</td>
<td>Sensitivity for atrophic gastritis: 87.2%; Specificity for atrophic gastritis: 91.1%; Sensitivity for GIM: 90.3%; Specificity for GIM: 93.7%</td>
</tr>
<tr>
<td>Yan et al. [18]</td>
<td>2020</td>
<td>NBI, M-NBI</td>
<td>Retrospective</td>
<td>Diagnosis of GIM</td>
<td>EfficientNetB4</td>
<td>2357 images</td>
<td>Sensitivity: 91.9%; Specificity: 86.0%; Accuracy: 88.8%</td>
</tr>
<tr>
<td>Cho et al. [60]</td>
<td>2019</td>
<td>WLE</td>
<td>Prospective</td>
<td>Classification of multiclass gastric neoplasms</td>
<td>Inception-Resnet-v2</td>
<td>5217 images</td>
<td>Accuracy: 84.6%</td>
</tr>
<tr>
<td>Iinoue et al. [61]</td>
<td>2020</td>
<td>WLE, NBI</td>
<td>Retrospective</td>
<td>Detection of duodenal adenomas and high-grade dysplasias</td>
<td>Single-Shot Multibox Detector</td>
<td>1511 images</td>
<td>For high-grade dysplasia, the sensitivity and specificity are all 100%</td>
</tr>
<tr>
<td>Lui et al. [62]</td>
<td>2020</td>
<td>NBI</td>
<td>Retrospective</td>
<td>Classification of gastric lesions</td>
<td>ResNet</td>
<td>3000 images</td>
<td>Sensitivity: 97.1%; Specificity: 85.9%; Accuracy: 91.0%</td>
</tr>
</tbody>
</table>

**Note:** BLI-bright: Blue laser imaging-bright; DL: Deep learning; GIM: Gastric intestinal metaplasia; H. pylori: Helicobacter pylori; LCI: Linked color imaging; M-NBI: Magnifying narrow band imaging; NBI: Narrow band imaging; WLE: White light endoscopy.
Data augmentation for a typical magnifying narrow band image for training a convolutional neural network model. This is performed by using a variety of image transformations and their combinations. A: Original image; B: Flip horizontal and random rotation; C: Flip vertical and magnification; D: Random rotation and shift; E: Flip horizontal, minification and shift; F: Flip vertical, rotation and shift.

Improvement on interpretability
Lack of interpretability (i.e. the “black box” nature), which is the nature of DL technology, is another gap between studies and clinical applications in the field of precancerous lesion detection from endoscopic images. The black box nature means that the decision-making process by the DL model is not clearly demonstrated, which may reduce the willingness of doctors to use it. Although attention maps can help explain the dominant areas by highlighting them, they are constrained in that they do not thoroughly explain how the algorithm comes to its final decision\[71\]. The attention maps are displayed as heat maps overlaid upon the original images, where warmer colors mean higher contributions to the decision making, which usually correspond to lesions. However, the attention maps also have some defects such as inaccurate display of lesions as shown in Figure 4, where the attention maps only cover partial areas associated with BE and GIM. This is the inherent shortcoming of attention maps. Therefore, understanding the mechanism used by the DL model for prediction is a hot research topic. The network dissection\[72\], an empirical method to identify the semantics of individual hidden nodes in the DL model, may be a feasible solution to improve interpretability.

Network design
In this review, we analyzed the DL model used in the detection of precancerous lesions in the upper GI tract. The literature shows that almost all the DL-based AI systems are developed based on state-of-the-art CNN architectures. These CNNs can only handle a single task, such as GoogLeNet for disease classification\[47,48,50,55\], YOLO for lesion identification\[36\] and SegNet for lesion segmentation\[31\]. Few networks can handle multiple tasks simultaneously. Thus, networks must be designed for multitask learning, which is valuable in clinical applications. Networks designed for handling high-resolution images can help detect micropatterns, which is profitable for small precancerous lesions. Moreover, attempts should be guided to exploit the use of videos rather than images to minimize the processing time and keep DL algorithms working at almost real-time level. Therefore, as suggested by Mori et al\[73\] and Thakkar et al\[74\], the AI systems may be treated as an extra pair of eyes to prevent the absence of subtle lesions.
CONCLUSION

Upper GI cancers are a major cause of cancer-related deaths worldwide. Early detection of precancerous lesions could significantly reduce cancer incidence. Upper GI endoscopy is a gold standard procedure for identifying precancerous lesions in the upper GI tract. DL-based endoscopic systems can provide an easier, faster and more reliable endoscopic method. We have conducted a thorough review of detection of precancerous lesions of the upper GI tract using DL approaches since 2017. This is the first review on the DL-based diagnosis of precancerous lesions of the upper GI tract. The status, challenges and recommendations summarized in this review can provide guidance for intelligent diagnosis of other GI tract diseases, which can help engineers develop perfect AI products to assist clinical decision making. Despite the success of DL algorithms in upper GI endoscopy, prospective studies and clinical validation are still needed. Creation of large public databases, adoption of comprehensive overfitting prevention strategies and application of more advanced interpretable methods and networks are also necessary to encourage clinical application of AI for medical diagnosis.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. I Cheong Choi, Dr. Hon Ho Yu and Dr. Mo Fong Li from the Department of Gastroenterology, Kiang Wu Hospital, Macau for their advice on this manuscript.

REFERENCES

Yan T et al. Deep learning for precancerous lesions

10.1016/j.bpg.2015.03.011


Yan T et al. Deep learning for precancerous lesions


Yan T et al. Deep learning for precancerous lesions

10.1016/j.compbioimed.2020.104029


