ISSN 2644-3236 (online)

Artificial Intelligence in *Gastroenterology*

Artif Intell Gastroenterol 2022 February 28; 3(1): 1-27

Published by Baishideng Publishing Group Inc

A I G

Artificial Intelligence in Gastroenterology

Contents

Bimonthly Volume 3 Number 1 February 28, 2022

EDITORIAL

MINIREVIEWS

Artificial Intelligence in Gastroenterology

Bimonthly Volume 3 Number 1 February 28, 2022

Contents

ABOUT COVER

Editorial Board Member of *Artificial Intelligence in Gastroenterology*, Natalia V Baryshnikova, MD, PhD, Associate Professor, Research Scientist, Senior Lecturer, Department of Internal Diseases, First Pavlov State Medical University, St-Petersburg 195257, Russia. baryshnikova_nv@mail.ru

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: *Rui-Rui Wu*; **Production Department Director:** *Xiang Li*; **Editorial Office Director:** *Jin-Lei Wang*.

NAME OF JOURNAL
Artificial Intelligence in Gastroenterology

ISSN
ISSN 2644-3236 (online)

LAUNCH DATE
July 28, 2020

FREQUENCY
Bimonthly

EDITORS-IN-CHIEF
Rajvinder Singh, Ferruccio Bonino

EDITORIAL BOARD MEMBERS
https://www.wjgnet.com/2644-3236/editorialboard.htm

PUBLICATION DATE
February 28, 2022

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

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 February 28; 3(1): 1-12

DOI: 10.35712/aig.v3.i1.1

ISSN 2644-3236 (online)

EDITORIAL

Gastrointestinal disorders in children with autism: Could artificial intelligence help?

Mohammed Al-Biltagi, Nermin Kamal Saeed, Samara Qaraghuli

Specialty type: Pediatrics

Provenance and peer review: Invited article; externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): 0
Grade D (Fair): D
Grade E (Poor): 0

P-Reviewer: Morozov S, Zhang X

Received: December 26, 2021
Peer-review started: December 26, 2021
First decision: February 10, 2022
Revised: February 12, 2022
Accepted: February 20, 2022
Article in press: February 20, 2022
Published online: February 28, 2022

Mohammed Al-Biltagi, Department of Pediatrics, Faculty of Medicine, Tanta University, Tanta 31511, Alghrabia, Egypt

Mohammed Al-Biltagi, Department of Pediatrics, University Medical Center, King Abdulla Medical City, Arabian Gulf University, Dr Sulaiman Al Habib Medical Group, Manama 26671, Manama, Bahrain

Nermin Kamal Saeed, Medical Microbiology Section, Pathology Department, Salmaniya Medical Complex, Ministry of Health, Kingdom of Bahrain, Manama 12, Manama, Bahrain

Nermin Kamal Saeed, Microbiology Section, Pathology Department, Irish Royal College of Surgeon, Bahrain, Busaiteen 15503, Muharraq, Bahrain

Samara Qaraghuli, Department of Pharmacognosy and Medicinal Plant, Faculty of Pharmacy, Al-Mustansiriya University, Baghdad 14022, Baghdad, Iraq

Corresponding author: Mohammed Al-Biltagi, MBChB, MD, MSc, PhD, Chairman, Professor, Department of Pediatrics, Faculty of Medicine, Tanta University, Al Bahra street, Tanta 31511, Alghrabia, Egypt. mbelrem@hotmail.com

Abstract

Autism is one of the pervasive neurodevelopmental disorders usually associated with many medical comorbidities. Gastrointestinal (GI) disorders are pervasive in children, with a 46%-84% prevalence rate. Children with Autism have an increased frequency of diarrhea, nausea and/or vomiting, gastroesophageal reflux and/or disease, abdominal pain, chronic flatulence due to various factors as food allergies, gastrointestinal dysmotility, irritable bowel syndrome (IBS), and inflammatory bowel diseases (IBD). These GI disorders have a significant negative impact on both the child and his/her family. Artificial intelligence (AI) could help diagnose and manage Autism by improving children's communication, social, and emotional skills for a long time. AI is an effective method to enhance early detection of GI disorders, including GI bleeding, gastroesophageal reflux disease, Coeliac disease, food allergies, IBS, IBD, and rectal polyps. AI can also help personalize the diet for children with Autism by microbiome modification. It can help to provide modified gluten without initiating an immune response. However, AI has many obstacles in treating digestive diseases, especially in children with Autism. We need to do more studies and adopt specific algorithms for children with Autism. In this article, we will highlight the role of AI in helping children with gastrointestinal disorders, with particular emphasis on children

with Autism.

Key Words: Autism; Gastrointestinal disorders; Artificial Intelligence; Children

Core Tip: Autism is a prevalent childhood neurodevelopmental condition. Gastrointestinal (GI) disorders are pervasive in children, with a 46%-84% prevalence rate. The presence of GI can negatively impair children's management and education. Artificial intelligence (AI) could help diagnose and manage autism by improving children's communication, social, and emotional skills for a long time. AI is an effective method to enhance early detection and management of GI disorders, including GI bleeding, gastroesophageal reflux disease, Coeliac disease, food allergies, irritable bowel syndrome, inflammatory bowel diseases, and rectal polyps. However, we still have some obstacles to increasing the benefit of AI in medicine, particularly in children with autism.

Citation: Al-Biltagi M, Saeed NK, Qaraghuli S. Gastrointestinal disorders in children with autism: Could artificial intelligence help? *Artif Intell Gastroenterol* 2022; 3(1): 1-12
URL: https://www.wjgnet.com/2644-3236/full/v3/i1/1.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i1.1

INTRODUCTION

Since its description for the first time by Leo Kanner in 1943, the rate of autism has been on the rise and steadily increasing[1]. Autism is a neurodevelopmental condition. Autism, Asperger's disorder, pervasive developmental disorder-not otherwise specified, form the autism spectrum disorders (ASD). At the same time, autism spectrum disorders, together with Rett's disorder, childhood disintegrative disorder, and the overactive disorder accompanied with mental retardation and stereotyped movements, are a part of the pervasive developmental disorders[2]. The prevalence of autism varies from one country to another depending on the racial differences and the diagnostic facilities available, with an average of 1% worldwide. The autism incidence in the United States of America may reach up to 1/110, increasing to 1/64 in the United Kingdom[3]. In other parts of the world, the prevalence of autism may be underestimated. For example, in Bahrain, the prevalence of autism is 1/1000, with possible underestimation because of missed diagnosis and no official recording in some cases. Autism is also 4-5 times more common in boys than girls. Autism shows a wide range of prevalence according to the race, being more common in non-Hispanic white children, less in Hispanic and African American/black children, with wide variability in Asian/Pacific Residents[4].

The genesis of autism is still unclear. Nevertheless, we can assert that autism development is due to the complex interaction of several genetic, biological, advanced parental age, environmental, immunological, and psychosocial factors[5]. Recently, genetic studies discovered a wide variety of genetic mutations in most patients. These mutations do not necessarily follow the same pattern with a wide range of variability. However, these mutations can ultimately induce brain changes and inflammation [6]. This neuroinflammation can also occur in utero through defective placenta augmented by the immaturity of the blood-brain barrier of the fetus and the newly delivered baby. This neuroinflammation can be triggered either as a part of the maternal immune response to infection during pregnancy, premature delivery, as a part of postnatal encephalitis, or exposure to a toxic environment[1]. We still need to have more knowledge to understand the different causes and their effects on patients with autism.

The clinical presentation of autism is heterogeneous, formed mainly from a constellation of social, cognitive, motor, and perceptual symptoms, which usually appear before three years of age. Children with autism have a diverse range of behaviors, communications, interactions, and learning ways from most other children. The abnormal social communication and interaction skills are manifested by poor eye contact, a stern facial expression such as happiness or sadness, lack of interest with others, and lack of interest in playing or interacting with others. They also have restricted interests manifested by playing with the same toys the same way every time, getting upset with changing routine or minor changes, and focusing on certain parts of the toys or the body with obsessive interest. Additionally, they have repetitive or stereotyped behaviors such as constantly repeating words or phrases (*i.e.*, echolalia), flapping hands, body rocking, or spinning self in circles). They also suffer from delayed language, movement, and cognitive or learning skills[7,8].

MEDICAL COMORBIDITIES

Besides the classic manifestations of autism, the affected patients may suffer the presence of many other medical comorbidities that are more common in people with ASD than in the general population. The presence of these comorbidities is one of the reasons for the significant increase of early mortality in patients with ASD, with death rates 3-10 times higher than the general population. These comorbidities may increase the risk of death in patients with autism and could affect their quality of life, impair proper diagnosis, interfere with their compromised learning capacity, and impair their ability to retain the acquired learning skills. Early recognition of these comorbidities helps improve the quality of life for both children and their families[9]. These comorbidities may include but are not limited to genetic, inborn errors of metabolism, congenital anomalies of the nervous system, neurologic disorders such as epilepsy and neuroinflammation, gastrointestinal (GI) disorders, and allergic disorders[10]. However, diagnosis of these comorbidities is not easily accessible due to communication impairments, occasional ambiguity of the symptoms, changes of the symptoms over time, and mimicking some of the classic symptoms of autism. A lack of available diagnostic instruments to screen these disorders further augments these difficulties[11].

GASTROINTESTINAL DISORDERS IN CHILDREN WITH AUTISM

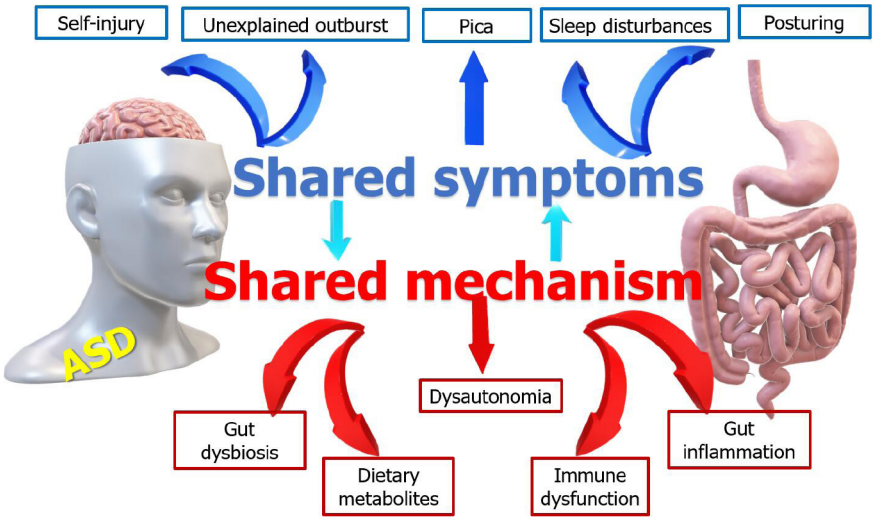
Children with autism have a high prevalence of GI disorders occurring in 46%-84% of them. The interaction between autism and gastrointestinal disorders is shown in Figure 1. Unfortunately, many of these children cannot effectively communicate their symptoms or discomfort to their doctors. Chronic constipation occurs in about 50%. They have a restricted diet with low fibers, abnormal bowel training, increased intestinal transit time, and a high incidence of hypothyroidism, increasing the frequency of constipation. Diarrhea is three times more common in children with autism than the control due to increased prevalence of food sensitivities, gut dysbiosis, immune dysfunction, and the increased infection rate due to increased incidence of pica and abnormal child behavior[12,13]. They have also increased frequency of nausea and/or vomiting, gastroesophageal reflux and/or disease, and chronic flatulence due to various factors such as food allergies gastrointestinal dysmotility[14]. Abdominal pain is also frequent in children with autism which results from simple, functional disorders such as irritable bowel syndrome, or organic causes such as food allergies, food intolerance, parasitic infestations due to pica, colitis, ulcers, or inflammatory bowel diseases[15].

Food allergies occur in about one-quarter of children with autism compared to 5%-8% in the general Pediatric population[9,16]. The link between autism and Celiac diseases (CD) is debatable. However, some high-quality studies proved this link even in the absence of GI symptoms[17]. Given that children with autism are more prone to suffer from atopy & food allergies, possible non-coeliac gluten sensitivity (NCGS) or wheat sensitivity in those children needs to be considered, especially when irritable bowel syndrome symptoms are present[18]. Physicians should consider the possibility of NCGS in some patients with ASD, especially those presenting with atopic diseases, migraines, and mood and anxiety disorders. Therefore, investigating CD and non-coeliac gluten sensitivity even in the absence of typical GI symptoms could yield good results for children with autism[19].

Children with ASD are more liable to have various feeding disorders; behavioral, sensory-based, or medically related feeding problems. The behavioral feeding disorders may include aversive eating behaviors (such as food refusal, frequent choking or gagging, the expulsion of the food without a medical reason), and frequent Pica habits. The sensory-based feeding problems include restrictive or selective eating and textural refusal of specific foods, usually involving larger textures. The medically-related feeding disorder may affect oesophageal and swallowing disorders and motor delays[20]. Almost two-thirds of children with autism eat less than 20 types of foods and accept fewer foods from the primary food groups than typically developing children[21]. This high prevalence of feeding problems in ASD may be related to their propensity to concentrate on details, their fear of novelty, their way of perseveration, and impulsivity. The associated sensory impairments and the deficits in social compliance of children with autism augment their feeding disorders. These feeding behaviors could also be aggravated by specific biological food intolerance and parental anxiety, reinforcing negative feeding patterns[22]. These feeding disorders have tremendous effects on both children and their families. They may increase the risk of child abuse and the occurrence of specific nutritional deficiencies, but Weight and height are usually not affected. They also increase parental anxiety and stress, ending with child abuse[23].

IMPORTANCE AND DIFFICULTIES IN DIAGNOSING GI DISORDERS IN CHILDREN WITH AUTISM

It is essential to check for the presence of gastrointestinal disorders in children with autism, as they can



DOI: 10.35712/aig.v3.i1.1

Figure 1 The interaction between autism and gastrointestinal disorders.

cause deterioration of autistic behaviors. For example, Abdominal pain related to especially reflux esophagitis and disaccharide malabsorption can cause irritation and discomfort to children with autism, which may contribute to the aggravation of their behavioral problems. It also could interfere with their learning abilities[24]. Meanwhile, gastrointestinal pain can cause behaviors that might be misdiagnosed as a behavior problem instead of a medical issue. For example, posturing, self-Injury, and/or outbursts without apparent cause could result from gastroesophageal reflux or esophagitis. The symptoms of GI dysfunction could induce sleep disturbances, which further aggravate the autistic manifestations[25].

Primary lactase deficiency that does not cause intestinal inflammation or injury is common in children with autism and may contribute to abdominal discomfort, pain, and observed aberrant behavior. Clinicians should screen for constipation and diarrhea or underwear soiling in children with autism who have prominent rigid–compulsive symptoms. If the GI disorder is recognized and medical treatment is effective, the behavioral problem may improve. When abdominal pain or discomfort is not alleviated, failure of psychotropic medications is more likely to occur. At the same time, these medications may even aggravate the problem if they have adverse gastrointestinal effects[9,26].

There is much evidence that modulation of the gut microbiota may be a manageable strategy for developing innovative therapies for complex CNS disorders, including autism[27]. The strong positive correlation of the gastrointestinal symptoms with the severity of autism indicates that children more severely affected by autism are likely to have severe gastrointestinal symptoms. Healthcare professionals should consider the possibility of gastrointestinal dysfunction in children with ASD, especially in those presenting with strange posturing or movements, sleep disorders, food intolerances, and aggressive or self-injurious behaviors[28]. The symptoms of GI dysfunction are associated with sleep disorders and food intolerance. Thus, it is essential to consider such association when evaluating and treating these comorbidities.

WHY IS IT DIFFICULT TO DIAGNOSE GI DISORDERS IN CHILDREN WITH AUTISM?

It is not always easy to detect GI manifestations in ASD. Children with autism have impaired communication skills, and many of them are nonverbal and cannot adequately express their pain, discomfort, or complaint through speech. Even those who can communicate verbally cannot adequately describe their symptoms. In addition, the symptoms of GI disorders may be missed as one of the classic symptoms and behavior commonly observed in children with autism. For example, toe-walking may be one of the typical stereotyped motor manifestations of autism to reduce feet overstimulation. It could also occur due to abdominal pain or loaded rectum or bladder. At the same time, GI disorders may present in atypical ways[29]. For example, suppose the child has abdominal pain or discomfort. In that case, he/she may touch his/her abdomen in a stereotyped way so that it can be easily missed with other stereotyped behaviors.

Moreover, GI disorders may present with non-GI manifestations. For example, sleep problems could be the manifestations of chronic GI disorders. It can be missed as being attributed to autism. The rate of sheep disorders increases from 30% of children with autism without GI disorders to reach 50% in the presence of GI disorders[30]. Children with autism may have hypersensitivity to various stimuli. On the other hand, they could occasionally have pain hyposensitivity with a high-pain threshold, affecting their

symptoms[31]. Unfortunately, no clinical practice guidelines exist to diagnose the presence of GI disorders in patients with ASD.

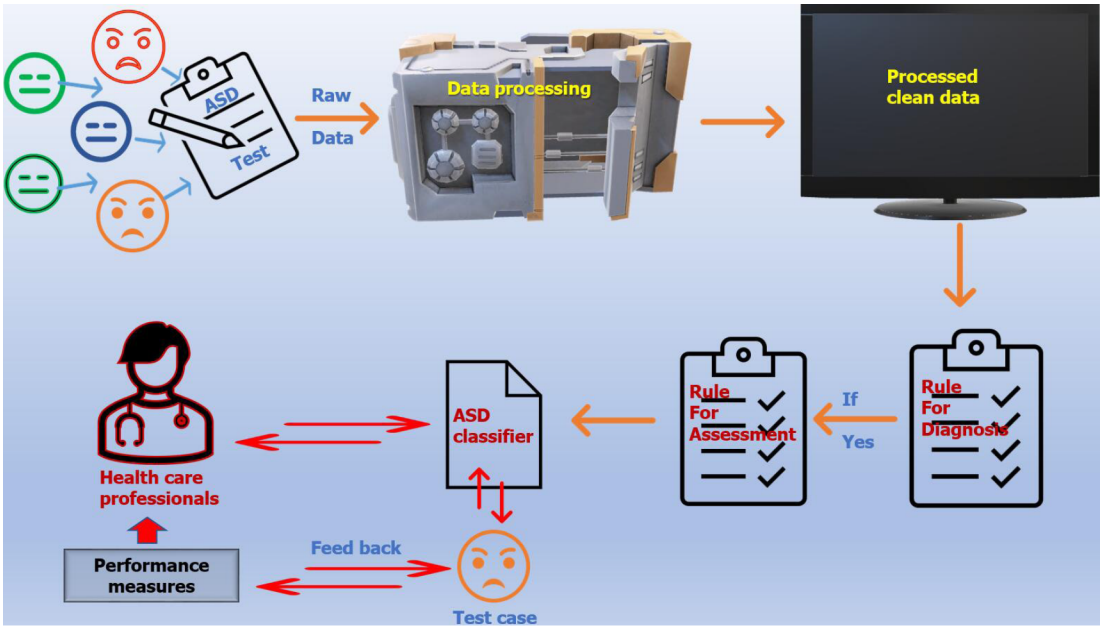
ROLE OF ARTIFICIAL INTELLIGENCE IN CHILDREN WITH AUTISM

Artificial intelligence (AI) enables a computer or computer-operated robot to perform tasks that humans usually do because they require human intelligence and judgment. The extensive application of artificial intelligence in various areas of life, including health, has begun to bear fruit. Whether we acknowledge it or not, artificial intelligence is inevitable and has a significant role in almost every aspect of our lives. The most important feature of AI is its ability to learn from its interaction, with the interaction-learning-interaction cycle. So, through pre-programmed flexible adaptation, AI can accurately interpret supplied external data, use these data to learn, and reuse the achieved learning to reach specific goals and duties. Machine learning is a part of AI and computer science that focuses on using data and algorithms to imitate how humans learn and gradually improve accuracy (Figure 2).

Deep learning, together with supervised and unsupervised learning, is a sub-class of machine learning that combines certain approaches that use specific algorithms to process and interpret data quicker, simpler, and more precisely[32]. In supervised learning, AI uses a computer algorithm to analyze predefined data to train and learn and then accurately names the new, hidden data. In unsupervised learning, the computer learns from massive, unlabelled led data and recognizes similarities and commonalities. Personalized medicine is an example of unsupervised medicine. The computers analyze the medical history, the result of the neck ultrasound or other radiology procedure, and the laboratory results for a patient with thyroid cancer to provide new perceptions for the treatment and the prognosis[33]. Medical sciences have greatly benefited from artificial intelligence, whether in diagnosing diseases, inventing appropriate medicines and treatments, or improving communication between doctors and patients[34]. There is much promise that AI will help improve healthcare services in many ways, including patient diagnosis, patient outcome, and drug invention, and assist the physician assistant and provide a better and more patient-tailored experience. This hope is driven by some of the emerging successful AI applications in healthcare[35].

It was a dream that AI could help diagnose and manage autism by improving children's communication, social, and emotional skills for a long time. However, this dream starts to convert into reality despite not being the norm yet. Diagnosis of autism is subjective. Consequently, it becomes a real challenge in many situations. Parents and physicians may miss children with mild symptoms, while the more severe cases can simulate many other developmental disorders. Diagnosis of autism can be achieved using machine learning to provide a rapid, simple, and easy technique to provide for autism early diagnosis[36]. Machine learning can also help in improving the efficacy of behavioral health screening. The addition of machine learning techniques to complement the conventional methods in diagnosing autism helps fasten the diagnosis and reduce its cost[37]. An example of machine learning recognizes abnormal behavior using video monitor and artificial intelligence analysis of the body movement and behavior in children to detect early children with autism. Alcañiz Raya *et al*[38] used machine-learning techniques to detect stereotyped and repetitive behaviors biomarkers, characteristics for autism. They used a depth sensor camera to track the body movements of the examined children. Consecutively, they exposed the children to different visual, auditory, and olfactory stimuli. They found that children with ASD had more significant body movements than typically developed children, especially in the head, trunk, and feet and for visual, followed by visual-auditive, and lastly for visual-auditive-olfactory stimuli.

An exciting study by Rahman *et al*[39] aimed to study the ability of machine learning to predict the risk of autism during the neonatal period. They combined the machine learning techniques with electronic medical records using parental sociodemographic information and medical histories and the prescribed medications data to create features to train various machine learning algorithms. They succeeded in capturing early-life features that increase the risk of ASD. They were also able to uncover previously unknown features linked with increased ASD risk. An additional exciting study used fetal ultrasound features by a computer program to predict the child's autism from the first day after birth. The fetal features included the baby's head and stomach size, thighs length, and the time of acquiring a vertex presentation in preparation for delivery. The program also used the peri-labor data such as heart rate and body temperature and followed the children up to 6 years. Then the program can independently recognize the associations between different fetal characteristics and outcomes[40]. Artificial intelligence can obtain data on a large scale from all over the world, then re-study it and extract data used to increase the accuracy of autism diagnosis. Many applications are used in diagnosing autism. Artificial intelligence can use the data collected by these applications and process it, so we get accurate results that represent helpful diagnostic tools for their application in different parts of the world. It is also possible to determine the criteria of autism for each race according to its culture and customs. One of the widely used applications to diagnose autism worldwide was created by Dr. Fadi Fayez and Dr. Reza Shahamiri (Nelson Marlborough Institute of Technology, New Zealand). It uses ten questions for the four age groups, from toddlerhood, childhood, and adolescence to adulthood. They



DOI: 10.35712/aig.v3.i1.1

Figure 2 Machine learning and Artificial intelligence in Autism.

used 70% of the data to identify the presence of autism and 30% to ensure that AI has appropriately learned the autistic features[41].

AI is promising in treating children with autism, despite currently being costly. Robots can train and interact with the advantage of showing different facial expressions, proper social interaction, and response to different social cues with unlimited patience and the ability to repeat the cues in the same manner, unlimited times without variation. Some robots can show social expression by changing their eye color, raising their arms, or changing their voice tone[42]. Some children with autism have a better response to the robot than a human therapist. Some robots can incorporate data about individual children using video, audio, and measurements of vital signs such as heart rate and temperature and presence or absence of skin sweat to personalize their response to the child's behaviors[43]. Despite being promising and effective, robotic intervention needs more wide-scale research, especially cost-effectiveness.

ROLE OF AI IN DIAGNOSIS OF GI DISORDERS

Diagnostic and therapeutic endoscopies have provided significant help in managing Pediatric gastrointestinal disorders for decades. Endoscopy with small bowel sampling is the gold standard to diagnose coeliac disease. Endoscopy also provides excellent assistance in diagnosing various GI diseases such as gastro-oesophageal reflux disease, eosinophilic oesophagitis, and inflammatory bowel disease. It is also used to stop GI bleeding, insert a gastrostomy tube, dilate a stricture, and remove a polyp. The recent marvelous endoscopic field achievements helped us reach previously non-reachable areas of the mid-small intestine using the wireless capsule video-endoscopy[44]. There was a broad jump in the endoscopic industry from the white light to the blue light endoscopy and recently endocytoscopy and endomicroscopy. These recent modalities helped visualize the mucosal structure at the cellular level with adequate histopathology determination. It helps gather a vast amount of data that needs many hours of interpretation by a highly experienced physician[45]. AI helps process these vast amounts of data and allows rapid and precious interpretation.

Gastroesophageal reflux disease (GERD) is a principal reason for abnormal behaviors in children with autism. Upper gastrointestinal endoscopy is one of the preferred modalities to diagnose and detect complications of GERD, including Barret esophagitis, by evaluating the oesophageal mucosa. It also can rule out other possible causes of the child's symptoms, such as eosinophilic esophagitis[46]. AI helps improve the mucosal images' quality and detect their exact anatomical location. Changes in the oesophageal mucosa such as Micro-erosions, Changes in intrapapillary capillary loops, and increased vascularity are landmarks for GERD detected by narrow-band imaging with the help of AI model using convolutional neural networks (CNNs)[47].

Meanwhile, Takiyama *et al*[48] used CNNs to precisely recognize the anatomical location of esophagogastroduodenoscopy images. Pace *et al*[49] developed an artificial neural networks (ANN) model that can predict the presence of GERD without the need for invasive diagnostic techniques in

patients with GERD symptoms. We hope that AI will help classify patients with GERD provide personalized therapeutic approaches[50]. However, the ability of ANN to expect GERD diagnosis depending on the symptoms still requires more verifications in different clinical settings. Considering that patients with autism may have other ways of expressing GER- or GERD-related symptoms, they may need specific and additional protocols to be applied with AI.

Coeliac disease is a common but underdiagnosed autoimmune disorder affecting 1/100 people worldwide with a relatively higher incidence in children with autism. The presence of villous atrophic histology in duodenal mucosal biopsy samples obtained by endoscopy is the gold standard for diagnosis. Endoscopy can also detect unsuspected cases of coeliac disease by meticulous analysis of the small bowel mucosa and identification of subtle findings of villous atrophy. However, it needs multiple biopsies not to miss the lesion as it is patchy[51]. New endoscopic techniques such as the modified immersion technique under traditional white-light or narrow-band imaging significantly improve the visual confirmation of coeliac disease during endoscopy[52,53]. Video capsule endoscopy is reasonably well-tolerated and safe in children. It can identify mucosal lesions in the bowel, especially in the small intestine, with the risk of radiation or sedation[54]. Video capsule imaging could also help identify the coeliac disease[55]. Augmentation techniques using AI can help augment the obtained original mucosal images to avoid the effects of conditions that could affect the quality of images, such as the rotation of the endoscope or the effects of distant viewpoint from the mucosal wall changes. However, a patency capsule test should be done before video capsule endoscopy, especially in infants and young children [56]. Recent AI modalities using deep learning techniques such as convolutional neural network (CNN), Bayesian inference, or support vector machines are innovative computer technology that can aid computerized coeliac disease diagnosis[57]. Foers *et al*[58] used machine learning methods to classify intestinal T-cell receptor repertoires to detect patients with coeliac disease irrespective of their dietary gluten status.

Pediatric colonoscopy needs a high experience not to miss lesions and detect colonic lesions as early as possible. The significant progress in developing computerized vision during gastrointestinal endoscopy allowed the gathering and annotation of high-quality video information. The addition of AI to real-time endoscopy significantly improves the automated detection of colonic or rectal polyps, such as in juvenile polyp or familial adenomatous polyposis. As mentioned before, children with autism have an increased risk to develop inflammatory bowel disease (IBD). This increased risk is due to multifactorial pathogenesis, including an overactive immune system and disturbances of the brain-gut-microbiota axis[59]. A massive flow of data about IBD is currently available using electronic medical records, genetic analysis, and imaging modalities. Analysis, interpretation, and integration of these data with the help of AI can aid to build models that can predict the risk of IBD and increase its detection accuracy[60].

Ozawa *et al*[61] succeeded to develop a neural network trained on colonoscopy images from patients with ulcerative colitis. With the help of a computer-assisted system, this network was able to identify the normal mucosa, mucosal healing states, mucosa on remission, and mucosa in severe degrees of inflammation with high sensitivity and specificity. These findings will help the physicians personalize the treatment according to the patients' conditions. In children, Mossotto *et al*[62] classified Crohn's and ulcerative colitis activity at diagnosis, using machine learning with integrating both endoscopic and histologic imaging. They were able to subtype the patients using this model with an accuracy reaching 80%, which significantly improves the diagnostic accuracy and permits a good option for targeted therapy. Dhaliwal *et al*[63] developed an algorithm using Random Forest Supervised and Unsupervised Machine learning in children to identify features that could help discriminate between ulcerative colitis and colonic Crohn's disease. They have a correct classification in 98% and 95% of children with ulcerative colitis and colonic Crohn disease, respectively.

Video capsule endoscopy is a safe, non-invasive procedure that can help diagnose IBD, especially for the patchy intestinal lesion of Crohn's disease. Nemeth *et al*[64] examined the accuracy and safety of video capsule endoscopy in 154 children and adolescents with suspected or established Crohn's disease. They found that video capsule endoscopy was safe and able to confirm the diagnosis of Crohn's disease with a significant impact on clinical management. However, interpretation of images obtained, and diagnosis based on *Video capsule endoscopy* is reader-dependant. As a result of the human concentration limitation, the lesion miss rate in capsule endoscopy ranges between 0.5% to 19% depending on the nature of the lesion[65]. AI can improve the accuracy of capsule endoscopy diagnosis by identifying distinct lesions and areas of interest with ease. However, there are many limitations to providing reliable classifications due to insufficient accuracy[66]. A convolutional neural network is used to analyze the large number of images obtained by capsule endoscopy to overcome these limitations. The convolutional neural network can differentiate normal intestinal mucosa, ulcers, erosion, polyps, and even worms with high accuracy reaching up to 96%[67,68].

Li *et al*[69] developed an AI system to automatically distinguish colorectal cancer early signs during colonoscopy with high sensitivity and specificity. Aguilar *et al*[70] found that transabdominal ultrasound augmented with a preceding AI model allows precise, fast, and non-invasive diagnosis of Buried bumper syndrome, complicating percutaneous endoscopic gastrostomy in children. Urban *et al* [71] successfully identified and removed rectal polyp using a deep neural network with a real-time accuracy rate of 96.4%. They used convolutional neural networks with an ordinary desktop machine

with a contemporary graphics processing unit. They concluded that their trained model could identify and locate polyps in real-time with high accuracy. However, caution should be taken when using convolutional neural networks as we cannot generalize the results of these studies to other situations, and we do not know the exact effect of using convolutional neural networks on the endoscopists inspection behavior with overreliance on the technology. In addition, the direct and the indirect cost of these technologies and their acceptance to be a part of the diagnostic tools by the physician should also be considered[72].

ROLE OF AI IN MANAGING GI DISORDERS IN CHILDREN WITH AUTISM

Children with autism have a three-fold increase in the risk of gastrointestinal disorders than the typically developed children. However, the rate of parental reporting of these disorders is less in children with autism than in the typically developed ones. The wide varieties of GI manifestations in children with autism are related to the general heterogeneity of autism disorder and the underlying neurobiological mechanisms and disturbances of the neurotransmitters in both brain and gut[73]. Some GI symptoms are evident as diarrhea and constipation, while the others may vague and challenging to be recognized and can be missed as behavioral changes. Artificial intelligence can help detect and classify autism early, even in young infants.

Meanwhile, AI can help detect and classify gastrointestinal disorders in children. Regrettably, according to the best of our knowledge, there are no currently available specific AI models to detect gastrointestinal disorders in children with autism. AI models designed to detect GI disorders in patients with autism should consider the differences in the symptomatology from the typically developed children and should design algorithms to detect these disorders.

Fascinating use of AI in children with autism is the use of 'SMART TOILET' to monitor bowel health and to help to detect irritable bowel syndrome and inflammatory bowel disease. An artificial intelligence tool with a camera and microcomputer are attached to the traditional toilet to help evaluate patients' stool, including form, defecation time, urination, and the presence or absence of blood[74]. This' SMART TOILET' will significantly help manage toilet disorders, common in children with autism, as they cannot correctly report their bowel habits, dysfunction, or defecatory disorders. The use of microbiota transfer therapy (MTT) showed significant potential in alleviating the symptoms associated with GI complications and reducing the severity of behavioral symptoms in children with autism[75]. Children with autism who had MTT also showed changes in their plasma metabolite profile to be nearly similar to the typically developing peers[76]. Qureshi *et al*[77] examined the differences in gut microbial metabolites between children with autism and GI disorders *vs* the typically developing children without GI disorders and determined the effects of gut MTT on the fecal metabolites of the group with autism. They used machine learning to create 5-metabolite fecal models for classification, which showed significant changes before and after MTT. The developed multivariate metabolite models showed the potential of fecal metabolite panels to effectively categorize children with autism from the typically developed. Similar machine learning models can diagnose children with autism using their gut microbiome data compared to subjects with and without autism.

About 10% of children with autism are on a special diet. Despite no diet specific for autism, children with autism are frequently put on a gluten-free, casein-free diet. However, children should not start a special diet except when it is evidence-based. Dietary management can help alleviate many of the functional gastrointestinal symptoms in patients with irritable bowel syndrome, which is relatively common in children with autism. One of these dietary managements is restoring the imbalance in gut microbiota. Karakan *et al*[78] studied the efficacy of artificial intelligence-based personalized modification of the dietary microbiome in patients with IBS. They designed an AI-based diet to optimize a personalized nutritional strategy using an algorithm about the individual gut microbiome features. According to the IBS index score, they developed the algorithm to design the diet. The algorithm assessing an IBS index score used the microbiome composition to design the optimized diets based on the microbiome modification to match the observed with the healthy scores.

Gluten sensitivity is common in children with autism. AI can help produce allergen-free gluten in plants with high gluten content, such as wheat and corn. This new gluten retains its unique beneficial quality regarding texture, taste, and nutritional value without the ability to stimulate the autoimmune response and cascade of gluten sensitivity or coeliac disease. Another way to overcome coeliac disease and gluten sensitivity is to create an oral enzyme able to degrade the ingested gluten. The proposed enzyme should be stable and active in both stomach and duodenum, rapidly neutralize the gluten-peptides that can activate T-cell, and be safe to be ingested by humans. Many enzymes, including cysteine proteases, prolyl endopeptidases, and subtilisin's, could split the non-digestible gluten peptides *in vivo* and vitro. AI can help develop new techniques like enteric coating to protect the enzyme or genetic modification, increasing its production and enhancing its stability in the GI tract[79].

LIMITATION FOR THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence has many obstacles in treating digestive diseases, especially in children with autism. Among these obstacles are the ethical aspects and the confidence of the medical staff in the mechanisms of artificial intelligence. The development of artificial intelligence also requires a robust infrastructure with enhanced patient confidentiality controls. Also, committees must be established to control the work of artificial intelligence to avoid the inappropriate and unethical use of artificial intelligence. Another significant limitation is the difference in symptoms of GI disorders in children with autism than the typically developed children. Children with autism may need a minimum level of communication abilities and cognitive function to use AI-directed models. When building an algorithm, it should be tailored to children with autism.

CONCLUSION

Autism is a neurodevelopmental condition with multiple comorbidities. Besides the classic manifestations of autism, the affected patients may suffer the presence of many other medical comorbidities that are more common in people with ASD than in the general population. Children with autism have a high prevalence of GI disorders occurring in 46%-84% of them with a bilateral mutual pathway between autism and GI disorders. Children with autism have an increased frequency of diarrhea, nausea and/or vomiting, gastroesophageal reflux and/or disease, abdominal pain, chronic flatulence due to various factors as food allergies, gastrointestinal dysmotility, IBS, and IBD. AI could help diagnose and manage autism by improving children's communication, social, and emotional skills for a long time. AI is an effective method to enhance early detection of GI disorders, including GI bleeding, gastroesophageal reflux disease, Coeliac disease, food allergies, IBS, IBD, and rectal polyps. AI can also help personalize the diet for children with autism by microbiome modification. AI can help to provide modified gluten without the ability to initiate an immune response. However, AI has many obstacles in treating digestive diseases, especially in children with autism. There is a need to do more studies and adapt specific algorithms for children with autism.

ACKNOWLEDGEMENTS

We thank the anonymous referees for their valuable suggestions.

FOOTNOTES

Author contributions: Al-Biltagi M, Saeed NK, and Qaraghuli S wrote and revised the manuscript.

Conflict-of-interest statement: The authors declare that they have no conflict of interest.

Country/Territory of origin: Egypt

ORCID number: Mohammed Al-Biltagi 0000-0002-7761-9536; Nermin Kamal Saeed 0000-0001-7875-8207; Samara Qaraghuli 0000-0002-6909-2658.

S-Editor: Liu JH
L-Editor: A
P-Editor: Liu JH

REFERENCES

1 **Ratajczak HV**. Theoretical aspects of autism: causes--a review. *J Immunotoxicol* 2011; **8**: 68-79 [PMID: 21299355 DOI: 10.3109/1547691X.2010.545086]
2 **McPartland J**, Volkmar FR. Autism and related disorders. *Handb Clin Neurol* 2012; **106**: 407-418 [PMID: 22608634 DOI: 10.1016/B978-0-444-52002-9.00023-1]

3 **Baird G**, Simonoff E, Pickles A, Chandler S, Loucas T, Meldrum D, Charman T. Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet* 2006; **368**: 210-215 [PMID: 16844490 DOI: 10.1016/S0140-6736(06)69041-7]

4 **Becerra TA**, von Ehrenstein OS, Heck JE, Olsen J, Arah OA, Jeste SS, Rodriguez M, Ritz B. Autism spectrum disorders and race, ethnicity, and nativity: a population-based study. *Pediatrics* 2014; **134**: e63-e71 [PMID: 24958588 DOI: 10.1542/peds.2013-3928]

5 **Bahmani M**, Sarrafchi A, Shirzad H, Rafieian-Kopaei M. Autism: Pathophysiology and Promising Herbal Remedies. *Curr Pharm Des* 2016; **22**: 277-285 [PMID: 26561063 DOI: 10.2174/1381612822666151112151529]

6 **Famitafreshi H**, Karimian M. Overview of the Recent Advances in Pathophysiology and Treatment for Autism. *CNS Neurol Disord Drug Targets* 2018; **17**: 590-594 [PMID: 29984672 DOI: 10.2174/1871527317666180706141654]

7 **Hyman SL**, Levy SE, Myers SM; COUNCIL ON CHILDREN WITH DISABILITIES, SECTION ON DEVELOPMENTAL AND BEHAVIORAL PEDIATRICS. Identification, Evaluation, and Management of Children With Autism Spectrum Disorder. *Pediatrics* 2020; **145** [PMID: 31843864 DOI: 10.1542/peds.2019-3447]

8 **Lin D**. Commentary on "The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer." Chakravarty D, Sboner A, Nair SS, Giannopoulou E, Li R, Hennig S, Mosquera JM, Pauwels J, Park K, Kossai M, MacDonald TY, Fontugne J, Erho N, Vergara IA, Ghadessi M, Davicioni E, Jenkins RB, Palanisamy N, Chen Z, Nakagawa S, Hirose T, Bander NH, Beltran H, Fox AH, Elemento O, Rubin MA, University of Washington-Urology, Seattle, WA. Nat Commun 2014; 5:5383. *Urol Oncol* 2016; **34**: 522 [PMID: 27814882 DOI: 10.1016/j.urolonc.2016.02.007]

9 **Al-Beltagi M**. Autism medical comorbidities. *World J Clin Pediatr* 2021; **10**: 15-28 [PMID: 33972922 DOI: 10.5409/wjcp.v10.i3.15]

10 **Brondino N**, Fusar-Poli L, Miceli E, Di Stefano M, Damiani S, Rocchetti M, Politi P. Prevalence of Medical Comorbidities in Adults with Autism Spectrum Disorder. *J Gen Intern Med* 2019; **34**: 1992-1994 [PMID: 31144278 DOI: 10.1007/s11606-019-05071-x]

11 **Casanova MF**, Frye RE, Gillberg C, Casanova EL. Editorial: Comorbidity and Autism Spectrum Disorder. *Front Psychiatry* 2020; **11**: 617395 [PMID: 33329163 DOI: 10.3389/fpsyt.2020.617395]

12 **McElhanon BO**, McCracken C, Karpen S, Sharp WG. Gastrointestinal symptoms in autism spectrum disorder: a meta-analysis. *Pediatrics* 2014; **133**: 872-883 [PMID: 24777214 DOI: 10.1542/peds.2013-3995]

13 **Wang LW**, Tancredi DJ, Thomas DW. The prevalence of gastrointestinal problems in children across the United States with autism spectrum disorders from families with multiple affected members. *J Dev Behav Pediatr* 2011; **32**: 351-360 [PMID: 21555957 DOI: 10.1097/DBP.0b013e31821bd06a]

14 **James DM**, Kozol RA, Kajiwara Y, Wahl AL, Storrs EC, Buxbaum JD, Klein M, Moshiree B, Dallman JE. Intestinal dysmotility in a zebrafish (*Danio rerio*) *shank3a;shank3b* mutant model of autism. *Mol Autism* 2019; **10**: 3 [PMID: 30733854 DOI: 10.1186/s13229-018-0250-4]

15 **Luna RA**, Oezguen N, Balderas M, Venkatachalam A, Runge JK, Versalovic J, Veenstra-VanderWeele J, Anderson GM, Savidge T, Williams KC. Distinct Microbiome-Neuroimmune Signatures Correlate With Functional Abdominal Pain in Children With Autism Spectrum Disorder. *Cell Mol Gastroenterol Hepatol* 2017; **3**: 218-230 [PMID: 28275689 DOI: 10.1016/j.jcmgh.2016.11.008]

16 **Tan Y**, Thomas S, Lee BK. Parent-reported prevalence of food allergies in children with autism spectrum disorder: National health interview survey, 2011-2015. *Autism Res* 2019; **12**: 802-805 [PMID: 30964233 DOI: 10.1002/aur.2106]

17 **Quan J**, Panaccione N, Jeong J, Underwood FE, Coward S, Windsor JW, Ronksley PE, Gidrewicz D, deBruyn J, Turner JM, Lebwohl B, Kaplan GG, King JA. Association Between Celiac Disease and Autism Spectrum Disorder: A Systematic Review. *J Pediatr Gastroenterol Nutr* 2021; **72**: 704-711 [PMID: 33847288 DOI: 10.1097/MPG.0000000000003051]

18 **Buie T**. The relationship of autism and gluten. *Clin Ther* 2013; **35**: 578-583 [PMID: 23688532 DOI: 10.1016/j.clinthera.2013.04.011]

19 **Navarro E**, Araya M. [Non-celiac gluten sensitivity: Another condition that responds to gluten]. *Rev Med Chil* 2015; **143**: 619-626 [PMID: 26203574 DOI: 10.4067/S0034-98872015000500010]

20 **Barnhill K**, Tami A, Schutte C, Hewitson L, Olive ML. Targeted Nutritional and Behavioral Feeding Intervention for a Child with Autism Spectrum Disorder. *Case Rep Psychiatry* 2016; **2016**: 1420549 [PMID: 27051550 DOI: 10.1155/2016/1420549]

21 **Bandini LG**, Curtin C, Phillips S, Anderson SE, Maslin M, Must A. Changes in Food Selectivity in Children with Autism Spectrum Disorder. *J Autism Dev Disord* 2017; **47**: 439-446 [PMID: 27866350 DOI: 10.1007/s10803-016-2963-6]

22 **Chistol LT**, Bandini LG, Must A, Phillips S, Cermak SA, Curtin C. Sensory Sensitivity and Food Selectivity in Children with Autism Spectrum Disorder. *J Autism Dev Disord* 2018; **48**: 583-591 [PMID: 29116421 DOI: 10.1007/s10803-017-3340-9]

23 **Baraskewich J**, von Ranson KM, McCrimmon A, McMorris CA. Feeding and eating problems in children and adolescents with autism: A scoping review. *Autism* 2021; **25**: 1505-1519 [PMID: 33653157 DOI: 10.1177/1362361321995631]

24 **Greydanus DE**, Gregoire-Bottex MM, Merrick J. Gastrointestinal dysfunction and autism: caution with misdiagnoses as many mysteries remain to be unraveled! *Int J Adolesc Med Health* 2016; **29** [PMID: 27977400 DOI: 10.1515/ijamh-2016-0127]

25 **Madra M**, Ringel R, Margolis KG. Gastrointestinal Issues and Autism Spectrum Disorder. *Psychiatr Clin North Am* 2021; **44**: 69-81 [PMID: 33526238 DOI: 10.1016/j.psc.2020.11.006]

26 **Ferguson BJ**, Marler S, Altstein LL, Lee EB, Akers J, Sohl K, McLaughlin A, Hartnett K, Kille B, Mazurek M, Macklin EA, McDonnell E, Barstow M, Bauman ML, Margolis KG, Veenstra-VanderWeele J, Beversdorf DQ. Psychophysiological Associations with Gastrointestinal Symptomatology in Autism Spectrum Disorder. *Autism Res* 2017; **10**: 276-288 [PMID: 27321113 DOI: 10.1002/aur.1646]

27 **Fattorusso A**, Di Genova L, Dell'Isola GB, Mencaroni E, Esposito S. Autism Spectrum Disorders and the Gut Microbiota. *Nutrients* 2019; **11** [PMID: 30823414 DOI: 10.3390/nu11030521]

28 **Adams JB**, Johansen LJ, Powell LD, Quig D, Rubin RA. Gastrointestinal flora and gastrointestinal status in children with autism--comparisons to typical children and correlation with autism severity. *BMC Gastroenterol* 2011; **11**: 22 [PMID:

21410934 DOI: 10.1186/1471-230X-11-22]
29 **Wasilewska J**, Klukowski M. Gastrointestinal symptoms and autism spectrum disorder: links and risks - a possible new overlap syndrome. *Pediatric Health Med Ther* 2015; 6: 153-166 [PMID: 29388597 DOI: 10.2147/PHMT.S85717]
30 **Carmassi C**, Palagini L, Caruso D, Masci I, Nobili L, Vita A, Dell'Osso L. Systematic Review of Sleep Disturbances and Circadian Sleep Desynchronization in Autism Spectrum Disorder: Toward an Integrative Model of a Self-Reinforcing Loop. *Front Psychiatry* 2019; **10**: 366 [PMID: 31244687 DOI: 10.3389/fpsyt.2019.00366]
31 **Allely CS**. Pain sensitivity and observer perception of pain in individuals with autistic spectrum disorder. *Scientific World Journal* 2013; **2013**: 916178 [PMID: 23843740 DOI: 10.1155/2013/916178]
32 **Deo RC**. Machine Learning in Medicine. *Circulation* 2015; **132**: 1920-1930 [PMID: 26572668 DOI: 10.1161/CIRCULATIONAHA.115.001593]
33 **Alzubaidi L**, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 2021; **8**: 53 [PMID: 33816053 DOI: 10.1186/s40537-021-00444-8]
34 **Basu K**, Sinha R, Ong A, Basu T. Artificial Intelligence: How is It Changing Medical Sciences and Its Future? *Indian J Dermatol* 2020; **65**: 365-370 [PMID: 33165420 DOI: 10.4103/ijd.IJD_421_20]
35 **Ahuja AS**. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ* 2019; **7**: e7702 [PMID: 31592346 DOI: 10.7717/peerj.7702]
36 **Abbas H**, Garberson F, Glover E, Wall DP. Machine learning approach for early detection of autism by combining questionnaire and home video screening. *J Am Med Inform Assoc* 2018; **25**: 1000-1007 [PMID: 29741630 DOI: 10.1093/jamia/ocy039]
37 **Vakadkar K**, Purkayastha D, Krishnan D. Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques. *SN Comput Sci* 2021; **2**: 386 [PMID: 34316724 DOI: 10.1007/s42979-021-00776-5]
38 **Alcañiz Raya M**, Marín-Morales J, Minissi ME, Teruel Garcia G, Abad L, Chicchi Giglioli IA. Machine Learning and Virtual Reality on Body Movements' Behaviors to Classify Children with Autism Spectrum Disorder. *J Clin Med* 2020; **9** [PMID: 32357517 DOI: 10.3390/jcm9051260]
39 **Rahman R**, Kodesh A, Levine SZ, Sandin S, Reichenberg A, Schlessinger A. Identification of newborns at risk for autism using electronic medical records and machine learning. *Eur Psychiatry* 2020; **63**: e22 [PMID: 32100657 DOI: 10.1192/j.eurpsy.2020.17]
40 **Caly H**, Rabiei H, Coste-Mazeau P, Hantz S, Alain S, Eyraud JL, Chianea T, Caly C, Makowski D, Hadjikhani N, Lemonnier E, Ben-Ari Y. Machine learning analysis of pregnancy data enables early identification of a subpopulation of newborns with ASD. *Sci Rep* 2021; **11**: 6877 [PMID: 33767300 DOI: 10.1038/s41598-021-86320-0]
41 **Thabtah F**. Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. *Inform Health Soc Care* 2019; **44**: 278-297 [PMID: 29436887 DOI: 10.1080/17538157.2017.1399132]
42 **DiPietro J**, Kelemen A, Liang Y, Sik-Lanyi C. Computer- and Robot-Assisted Therapies to Aid Social and Intellectual Functioning of Children with Autism Spectrum Disorder. *Medicina (Kaunas)* 2019; **55** [PMID: 31387274 DOI: 10.3390/medicina55080440]
43 **Szymona B**, Maciejewski M, Karpiński R, Jonak K, Radzikowska-Büchner E, Niderla K, Prokopiak A. Robot-Assisted Autism Therapy (RAAT). Criteria and Types of Experiments Using Anthropomorphic and Zoomorphic Robots. Review of the Research. *Sensors (Basel)* 2021; **21** [PMID: 34071829 DOI: 10.3390/s21113720]
44 **Belsha D**, Bremner R, Thomson M. Indications for gastrointestinal endoscopy in childhood. *Arch Dis Child* 2016; **101**: 1153-1160 [PMID: 27246069 DOI: 10.1136/archdischild-2014-306043]
45 **Goetz M**, Watson A, Kiesslich R. Confocal laser endomicroscopy in gastrointestinal diseases. *J Biophotonics* 2011; **4**: 498-508 [PMID: 21567975 DOI: 10.1002/jbio.201100022]
46 **Sharma VK**. Role of endoscopy in GERD. *Gastroenterol Clin North Am* 2014; **43**: 39-46 [PMID: 24503358 DOI: 10.1016/j.gtc.2013.12.003]
47 **Fock KM**, Teo EK, Ang TL, Tan JY, Law NM. The utility of narrow band imaging in improving the endoscopic diagnosis of gastroesophageal reflux disease. *Clin Gastroenterol Hepatol* 2009; **7**: 54-59 [PMID: 18852068 DOI: 10.1016/j.cgh.2008.08.030]
48 **Takiyama H**, Ozawa T, Ishihara S, Fujishiro M, Shichijo S, Nomura S, Miura M, Tada T. Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks. *Sci Rep* 2018; **8**: 7497 [PMID: 29760397 DOI: 10.1038/s41598-018-25842-6]
49 **Pace F**, Buscema M, Dominici P, Intraligi M, Baldi F, Cestari R, Passaretti S, Bianchi Porro G, Grossi E. Artificial neural networks are able to recognize gastro-oesophageal reflux disease patients solely on the basis of clinical data. *Eur J Gastroenterol Hepatol* 2005; **17**: 605-610 [PMID: 15879721 DOI: 10.1097/00042737-200506000-00003]
50 **Savarino V**, Dulbecco P. Can artificial neural networks be beneficial in diagnosing gastro-oesophageal reflux disease? *Eur J Gastroenterol Hepatol* 2005; **17**: 599-601 [PMID: 15879719 DOI: 10.1097/00042737-200506000-00001]
51 **Ediger TR**, Hill ID. Celiac disease. *Pediatr Rev* 2014; **35**: 409-15; quiz 416 [PMID: 25274968 DOI: 10.1542/pir.35-10-409]
52 **Gasbarrini A**, Ojetti V, Cuoco L, Cammarota G, Migneco A, Armuzzi A, Pola P, Gasbarrini G. Lack of endoscopic visualization of intestinal villi with the "immersion technique" in overt atrophic celiac disease. *Gastrointest Endosc* 2003; **57**: 348-351 [PMID: 12612514 DOI: 10.1067/mge.2003.116]
53 **Valitutti F**, Oliva S, Iorfida D, Aloi M, Gatti S, Trovato CM, Montuori M, Tiberti A, Cucchiara S, Di Nardo G. Narrow band imaging combined with water immersion technique in the diagnosis of celiac disease. *Dig Liver Dis* 2014; **46**: 1099-1102 [PMID: 25224697 DOI: 10.1016/j.dld.2014.08.039]
54 **Wu J**, Huang Z, Wang Y, Tang Z, Lai L, Xue A, Huang Y. Clinical features of capsule endoscopy in 825 children: A single-center, retrospective cohort study. *Medicine (Baltimore)* 2020; **99**: e22864 [PMID: 33120825 DOI: 10.1097/MD.0000000000022864]
55 **Ciaccio EJ**, Tennyson CA, Bhagat G, Lewis SK, Green PH. Transformation of videocapsule images to detect small bowel mucosal differences in celiac versus control patients. *Comput Methods Programs Biomed* 2012; **108**: 28-37 [PMID:

22284703 DOI: 10.1016/j.cmpb.2011.12.008]
56 **Wimmer G**, Uhl A, Vecsei A. Evaluation of domain specific data augmentation techniques for the classification of celiac disease using endoscopic imagery. 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), 2017; 1-6 [DOI: 10.1109/mmsp.2017.8122221]
57 **Molder A**, Balaban DV, Jinga M, Molder CC. Current Evidence on Computer-Aided Diagnosis of Celiac Disease: Systematic Review. *Front Pharmacol* 2020; **11**: 341 [PMID: 32372947 DOI: 10.3389/fphar.2020.00341]
58 **Foers AD**, Shoukat MS, Welsh OE, Donovan K, Petry R, Evans SC, FitzPatrick ME, Collins N, Klenerman P, Fowler A, Soilleux EJ. Classification of intestinal T-cell receptor repertoires using machine learning methods can identify patients with coeliac disease regardless of dietary gluten status. *J Pathol* 2021; **253**: 279-291 [PMID: 33225446 DOI: 10.1002/path.5592]
59 **Lee M**, Krishnamurthy J, Susi A, Sullivan C, Gorman GH, Hisle-Gorman E, Erdie-Lalena CR, Nylund CM. Association of Autism Spectrum Disorders and Inflammatory Bowel Disease. *J Autism Dev Disord* 2018; **48**: 1523-1529 [PMID: 29170940 DOI: 10.1007/s10803-017-3409-5]
60 **Gubatan J**, Levitte S, Patel A, Balabanis T, Wei MT, Sinha SR. Artificial intelligence applications in inflammatory bowel disease: Emerging technologies and future directions. *World J Gastroenterol* 2021; **27**: 1920-1935 [PMID: 34007130 DOI: 10.3748/wjg.v27.i17.1920]
61 **Ozawa T**, Ishihara S, Fujishiro M, Saito H, Kumagai Y, Shichijo S, Aoyama K, Tada T. Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis. *Gastrointest Endosc* 2019; **89**: 416-421.e1 [PMID: 30367878 DOI: 10.1016/j.gie.2018.10.020]
62 **Mossotto E**, Ashton JJ, Coelho T, Beattie RM, MacArthur BD, Ennis S. Classification of Paediatric Inflammatory Bowel Disease using Machine Learning. *Sci Rep* 2017; **7**: 2427 [PMID: 28546534 DOI: 10.1038/s41598-017-02606-2]
63 **Dhaliwal J**, Erdman L, Drysdale E, Rinawi F, Muir J, Walters TD, Siddiqui I, Griffiths AM, Church PC. Accurate Classification of Pediatric Colonic Inflammatory Bowel Disease Subtype Using a Random Forest Machine Learning Classifier. *J Pediatr Gastroenterol Nutr* 2021; **72**: 262-269 [PMID: 33003163 DOI: 10.1097/MPG.0000000000002956]
64 **Nemeth A**, Agardh D, Wurm Johansson G, Thorlacius H, Toth E. Video capsule endoscopy in pediatric patients with Crohn's disease: a single-center experience of 180 procedures. *Therap Adv Gastroenterol* 2018; **11**: 1756284818758929 [PMID: 29531578 DOI: 10.1177/1756284818758929]
65 **Lewis BS**, Eisen GM, Friedman S. A pooled analysis to evaluate results of capsule endoscopy trials. *Endoscopy* 2005; **37**: 960-965 [PMID: 16189768 DOI: 10.1055/s-2005-870353]
66 **Iakovidis DK**, Koulaouzidis A. Software for enhanced video capsule endoscopy: challenges for essential progress. *Nat Rev Gastroenterol Hepatol* 2015; **12**: 172-186 [PMID: 25688052 DOI: 10.1038/nrgastro.2015.13]
67 **Leenhardt R**, Li C, Le Mouel JP, Rahmi G, Saurin JC, Cholet F, Boureille A, Amiot X, Delvaux M, Duburque C, Leandri C, Gérard R, Lecleire S, Mesli F, Nion-Larmurier I, Romain O, Sacher-Huvelin S, Simon-Shane C, Vanbiervliet G, Marteau P, Histace A, Dray X. CAD-CAP: a 25,000-image database serving the development of artificial intelligence for capsule endoscopy. *Endosc Int Open* 2020; **8**: E415-E420 [PMID: 32118115 DOI: 10.1055/a-1035-9088]
68 **Fan S**, Xu L, Fan Y, Wei K, Li L. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. *Phys Med Biol* 2018; **63**: 165001 [PMID: 30033931 DOI: 10.1088/1361-6560/aad51c]
69 **Yamada M**, Saito Y, Imaoka H, Saiko M, Yamada S, Kondo H, Takamaru H, Sakamoto T, Sese J, Kuchiba A, Shibata T, Hamamoto R. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Sci Rep* 2019; **9**: 14465 [PMID: 31594962 DOI: 10.1038/s41598-019-50567-5]
70 **Aguilar C**, Regensburger AP, Knieling F, Wagner AL, Siebenlist G, Woelfle J, Koehler H, Hoerning A, Jüngert J. Pediatric Buried Bumper Syndrome: Diagnostic Validity of Transabdominal Ultrasound and Artificial Intelligence. *Ultraschall Med* 2021 [PMID: 34034349 DOI: 10.1055/a-1471-3039]
71 **Urban G**, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W, Baldi P. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. *Gastroenterology* 2018; **155**: 1069-1078.e8 [PMID: 29928897 DOI: 10.1053/j.gastro.2018.06.037]
72 **Radaelli F**, Paggi S. Artificial intelligence and the endoscopist's skill and proficiency for polyp detection: no winner one without the other! *Transl Gastroenterol Hepatol* 2021; **6**: 7 [PMID: 33409401 DOI: 10.21037/tgh.2019.01.08]
73 **Rose S**, Bennuri SC, Murray KF, Buie T, Winter H, Frye RE. Mitochondrial dysfunction in the gastrointestinal mucosa of children with autism: A blinded case-control study. *PLoS One* 2017; **12**: e0186377 [PMID: 29028817 DOI: 10.1371/journal.pone.0186377]
74 **Wang XJ**, Camilleri M. A smart toilet for personalized health monitoring. *Nat Rev Gastroenterol Hepatol* 2020; **17**: 453-454 [PMID: 32483351 DOI: 10.1038/s41575-020-0320-x]
75 **Kang DW**, Adams JB, Gregory AC, Borody T, Chittick L, Fasano A, Khoruts A, Geis E, Maldonado J, McDonough-Means S, Pollard EL, Roux S, Sadowsky MJ, Lipson KS, Sullivan MB, Caporaso JG, Krajmalnik-Brown R. Microbiota Transfer Therapy alters gut ecosystem and improves gastrointestinal and autism symptoms: an open-label study. *Microbiome* 2017; **5**: 10 [PMID: 28122648 DOI: 10.1186/s40168-016-0225-7]
76 **Adams JB**, Borody TJ, Kang DW, Khoruts A, Krajmalnik-Brown R, Sadowsky MJ. Microbiota transplant therapy and autism: lessons for the clinic. *Expert Rev Gastroenterol Hepatol* 2019; **13**: 1033-1037 [PMID: 31665947 DOI: 10.1080/17474124.2019.1687293]
77 **Qureshi F**, Adams J, Hanagan K, Kang DW, Krajmalnik-Brown R, Hahn J. Multivariate Analysis of Fecal Metabolites from Children with Autism Spectrum Disorder and Gastrointestinal Symptoms before and after Microbiota Transfer Therapy. *J Pers Med* 2020; **10** [PMID: 33023268 DOI: 10.3390/jpm10040152]
78 **Karakan T**, Gundogdu, A, Alagözlü A, Ekmen A, Ozgul S, Hora M, Beyazgul D, Nalbantoglu UO. Artificial Intelligence based personalized diet: A pilot clinical study for IBS. *medRxiv* 2021 [DOI: 10.1101/2021.02.23.21251434]
79 **Wei G**, Helmerhorst EJ, Darwish G, Blumenkranz G, Schuppan D. Gluten Degrading Enzymes for Treatment of Celiac Disease. *Nutrients* 2020; **12** [PMID: 32679754 DOI: 10.3390/nu12072095]

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 February 28; 3(1): 13-20

DOI: 10.35712/aig.v3.i1.13

ISSN 2644-3236 (online)

MINIREVIEWS

Current advancements in application of artificial intelligence in clinical decision-making by gastroenterologists in gastrointestinal bleeding

Hasan Maulahela, Nagita Gianty Annisa

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): 0
Grade D (Fair): D, D
Grade E (Poor): 0

P-Reviewer: Treeprasertsuk S, Wang RG

Received: November 21, 2021
Peer-review started: November 21, 2021
First decision: January 9, 2022
Revised: January 24, 2022
Accepted: February 23, 2022
Article in press: February 23, 2022
Published online: February 28, 2022

Hasan Maulahela, Department of Internal Medicine, Gastroenterology Division, Faculty of Medicine University of Indonesia - Cipto Mangunkusumo General Central National Hospital, Jakarta 10430, Indonesia

Nagita Gianty Annisa, Faculty of Medicine, University of Indonesia, Jakarta 10430, Indonesia

Corresponding author: Hasan Maulahela, MD, Assistant Professor, Department of Internal Medicine, Gastroenterology Division, Faculty of Medicine University of Indonesia - Cipto Mangunkusumo General Central National Hospital, Jl. Pangeran Diponegoro No. 71 RW.5, Kenari, Kec. Senen, Jakarta 10430, Indonesia. hasan.maulahela@yahoo.com

Abstract

Artificial Intelligence (AI) is a type of intelligence that comes from machines or computer systems that mimics human cognitive function. Recently, AI has been utilized in medicine and helped clinicians make clinical decisions. In gastroenterology, AI has assisted colon polyp detection, optical biopsy, and diagnosis of *Helicobacter pylori* infection. AI also has a broad role in the clinical prediction and management of gastrointestinal bleeding. Machine learning can determine the clinical risk of upper and lower gastrointestinal bleeding. AI can assist the management of gastrointestinal bleeding by identifying high-risk patients who might need urgent endoscopic treatment or blood transfusion, determining bleeding stigmata during endoscopy, and predicting recurrence of gastrointestinal bleeding. The present review will discuss the role of AI in the clinical prediction and management of gastrointestinal bleeding, primarily on how it could assist gastroenterologists in their clinical decision-making compared to conventional methods. This review will also discuss challenges in implementing AI in routine practice.

Key Words: Gastrointestinal bleeding; Artificial intelligence; Machine learning; Artificial neural networks; Clinical decision making

Core Tip: Gastrointestinal bleeding is a common problem in the emergency department. Quick and appropriate clinical decision is needed in the management of gastrointestinal bleeding. Artificial intelligence, namely machine learning and deep learning, can utilize electronic health record data to provide insights which might help clinicians, especially gastroenterologists, in the management of gastrointestinal bleeding. The present review will discuss the roles of artificial intelligence in clinical prediction and management of gastrointestinal bleeding, and compare them to conventional methods. This review will also discuss challenges in the implementation of artificial intelligence in routine practice.

Citation: Maulahela H, Annisa NG. Current advancements in application of artificial intelligence in clinical decision-making by gastroenterologists in gastrointestinal bleeding. *Artif Intell Gastroenterol* 2022; 3(1): 13-20
URL: https://www.wjgnet.com/2644-3236/full/v3/i1/13.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i1.13

INTRODUCTION

Artificial intelligence (AI) simulates human intelligence processes and cognitive function using machines or computer systems. Several terminologies need to be understood before talking about AI. Machine learning (ML) is a technique of AI in which a computer or a system can learn to improve its function using experience and data without explicit instruction. There are several machine learning methods, for example, CNN (convolutional neural network), that can perform image analysis. ANN (artificial neural network) consists of a hidden-layered connection between input and output. Meanwhile, deep learning is a class of machine learning which extracts higher-level information progressively using multiple layers of neural networks[1]. AI has transformed information technology by making it possible to analyse large-scale data within a short time[2].

Recently, AI has been utilized in medicine. AI has a broad role in medicine, from guiding treatment decisions using electronic health record data to assisting in performing surgeries and intelligent prostheses for people with disabilities[3]. In gastroenterology, AI has assisted in diagnosing and treating gastrointestinal (GI) diseases. AI also has roles in small intestinal endoscopy and endoscopic ultrasound, especially in evaluating and diagnosing lesions[4].

This review aims to discuss the roles of AI in GI bleeding, especially in clinical decision-making for gastroenterologists. More specifically, this review will discuss the advancements in the application of AI in clinical prediction and management of upper and lower GI bleeding and its limitations and future challenges.

ARTIFICIAL INTELLIGENCE IN CLINICAL PREDICTION OF UPPER GASTROINTESTINAL BLEEDING

Several scoring systems or risk models have been developed to predict the clinical risk of GI bleeding. In patients using antithrombotic medications, these risk models include HAS-BLED (hypertension, abnormal kidney and liver function, stroke, bleeding, labile international normalized ratio, elder age, and drug or alcohol use), ATRIA (anticoagulation and risk factors in atrial fibrillation), ORBIT (Outcomes Registry for Better Informed Treatment of Atrial Fibrillation), and HEMORR2HAGES (hepatic or kidney disease, ethanol abuse, malignancy, older age, reduced platelet count or function, rebleeding, hypertension, anemia, genetic factors, excessive fall risk, and stroke)[5-7]. Among these models, HAS-BLED has the best performance to predict major bleeding events[8].

Compared to the previous risk models, the prediction model using machine learning is hypothesized to have better performance since it can utilize more extensive and updated data sets. Herrin *et al*[9] tested three machine learning algorithms: Regularized Cox regression (RegCox), random survival forests, and extreme gradient boosting (XGBoost) on adult patients who were prescribed antithrombotic drugs (vitamin K antagonists, direct oral anticoagulants (DOACs), and/or thienopyridine antiplatelet agents) to predict the probability of GI bleeding at 6 and 12 mo. The data were obtained from medical and pharmacy claims data of 300000 patients. They also compared the performance of the machine learning algorithms to the HAS-BLED risk model.

In that study, all machine learning algorithms performed superiorly to HAS-BLED score in predicting GI bleeding at 6 and 12 mo. HAS-BLED score achieved an area under the curve (AUC) of 0.61 [95% confidence interval (CI): 0.59-0.62] for 6-mo GI bleeding risk and AUC of 0.60 (95%CI: 0.59-0.61) for 12-mo GI bleeding risk. Meanwhile, RegCox, the most superior algorithm from the three machine learning algorithms, had an AUC of 0.68 (95%CI: 0.66-0.70) for 6-mo GI bleeding risk and AUC of 0.67 (95%CI: 0.65-0.69) for 12-mo GI bleeding risk. HAS-BLED and the three machine learning algorithms obtained a

similar sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). However, all of them had an AUC less than 0.70, which is the conventional threshold for acceptable performance[9].

HAS-BLED score was derived to predict major bleeding events from patients treated with warfarin [10]. However, recently, antiplatelet agents and DOACs are more commonly used. Even though clinical extrapolation to calculate the risk of GI bleeding in patients taking antithrombotics is common, there are still concerns regarding the accuracy of HAS-BLED in predicting bleeding events in patients taking other types of anticoagulants or antiplatelets. Capodanno *et al*[11] found that HAS-BLED score could not predict major bleeding events in patients undergoing PCI (percutaneous coronary intervention) without artrial fibrillation who were discharged with dual antiplatelets. Although not specifically developed to predict GI bleeding events, several scoring systems have been developed for predicting bleeding events in patients taking dual antiplatelet therapy, such as CRUSADE, ACUITY, and PRECISE-DAPT. However, each scoring system has different accuracies in predicting short-term and long-term bleeding complications. For example, CRUSADE and ACUITY are better in predicting short-term complications, while PRECISE-DAPT is better in predicting long-term bleeding events[12].

Machine learning algorithms that utilize real-time data, such as RegCox, should better predict GI bleeding than the scoring systems mentioned above. Moreover, machine learning algorithms can provide time-to-event outcomes that can be used in the prediction of both short-term and long-term GI bleeding events. Herrin *et al*[9] used data sets from insurance claims and could not provide actual clinical values, which might contribute to low AUCs in their study. Data sets from electronic health record data that contain laboratory values and endoscopic reports might result in a better accuracy for clinical prediction of GI bleeding.

In patients presenting with upper GI bleeding, especially in the emergency department, it is important to stratify a patient's risk and predict mortality outcomes and the need for transfusion and other hemostatic interventions. Scoring systems such as the Glasgow-Blatchford score (GBS), Rockall score, and AIMS65 predict pre-endoscopic risk in patients with acute upper GI bleeding based on clinical, hemodynamic, and initial laboratory variables. Shung *et al*[13] conducted a systematic review that included 14 studies with 30 assessments of ML models. The median AUC for mortality, interventions, or rebleeding outcomes for ML models was 0.84. AUCs were higher in studies using ANNs than other models. They found that ML performed better than clinical risk scores for mortality in upper GI bleeding.

Recently, Shung *et al*[14] validated a machine learning model for upper GI bleeding that predicted composite outcomes of the need for hospital-based interventions (red blood cell transfusion, endoscopic hemostatic intervention, or surgery) and 30-d all-cause mortality. The chosen ML model was the XGBoost model. Different from previous studies, this study did not collect data from insurance records but through medical data that was directly entered by a nurse, physician, or medical student.

The ML model obtained an AUC of 0.91 (95%CI: 0.90-0.93) in the internal validation group, and an AUC of 0.90 (95%CI: 0.87-0.93) in the external validation group. The model performed better than GBS (AUC = 0.87, 95%CI: 0.84-0.91; $P = 0.004$), admission Rockall (AUC = 0.65, 95%CI: 0.60-0.71; $P < 0.001$), and AIMS65 (AUC = 0.64, 95%CI: 0.59-0.69; $P < 0.001$)[14].

ML models could perform better than scoring systems in risk stratification in patients with upper GI bleeding because they could extract patterns from raw data and increase accuracy with additional data and experience. Moreover, ML models could analyze more complex and heterogeneous data.

ARTIFICIAL INTELLIGENCE IN CLINICAL PREDICTION OF LOWER INTESTINAL BLEEDING

AI also has roles in the clinical prediction of lower intestinal bleeding. In 2017, Loftus *et al*[15] conducted a study that compared ANN and a regression-based model to predict the severity of lower GI bleeding and the need for surgical intervention.

Loftus *et al*[15] performed the analysis retrospectively on 147 adult patients who underwent endoscopy, angiography, or surgery for acute lower intestinal bleeding. The regression-based model used was the Strate prediction rule. The ANN for prediction of severe bleeding incorporated six variables present on admission: Systolic blood pressure; hemoglobin; outpatient prescription of aspirin 325 mg daily; Charlson comorbidity index; base deficit ≥ 5 mEq/L; and international normalized ratio ≥ 1.5. Meanwhile, the ANN for prediction of the need for surgery combined three predictors from severe bleeding ANN with two additional variables, hemoglobin nadir and the occurrence of a 20% decrease in haematocrit[15].

The Strate risk factors in the study correlated significantly with severe bleeding ($r = 0.29$, $P < 0.001$). However, the Strate model was less accurate in predicting severe lower intestinal bleeding than the ANN [area under the receiver operating characteristic curve (AUROC) 0.66 (95%CI: 0.57-0.75) *vs* 0.98 (95%CI: 0.95-1.00)]. The ANN for predicting the need for surgical intervention also had good performance with an AUROC of 0.95 (95%CI: 0.90-1.00). ANN could perform better than the regression-based model because this program could incorporate intricate associations among variables into an

algorithm, similar to nonlinear statistical processing[15].

Ayaru *et al*[16] analyzed non-endoscopic variables from patients with acute lower GI bleeding in the emergency department for internal and external validation of the gradient boosting (GB) model. GB is a supervised machine learning algorithm used in regression and classification tasks with multiple simple learning algorithms used jointly to obtain better predictive performance. Their study compared GB model with BLEED classification, Strate prediction rule, and conventional multiple logistic regression in predicting severe bleeding, the need for therapeutic intervention, and recurrent bleeding in patients with acute lower GI bleeding.

Ayaru *et al*[16] found that the GB model performed better than other scoring systems with an accuracy of 88% for recurrent bleeding and therapeutic intervention and 78% for the need for therapeutic intervention. Meanwhile, conventional multiple logistic regression had an accuracy of 74% in predicting recurrent bleeding and the need for therapeutic intervention and an accuracy of 62% in predicting severe bleeding. BLEED classification and Strate prediction rule also performed more poorly than the GB model.

In their study, the GB model could provide variables contributing to the risk of severe acute lower GI bleeding and the contribution percentage. The variables and their contribution are platelet count (13.4%), activated partial thromboplastin time (13.0%), haematocrit (12.4%), urea (10.9%), creatinine (9.7%), prothrombin time (8.9%), diastolic blood pressure (6.8%), heart rate (4.1%), systolic blood pressure (3.9%), and alcohol abuse (3.9%)[16].

Both studies by Loftus *et al*[15] and Ayaru *et al*[16] found that AI performed better than scoring systems in predicting lower GI bleeding. Even though they used different algorithms, ANN and GB model both could perform better than other regression-based models and scoring systems. Moreover, the algorithms could provide variables contributing to the risk of bleeding and the need for therapeutic intervention. However, both studies were limited by their retrospective design. More prospective studies need to be conducted to determine the accuracy of ML models in lower GI bleeding prediction. More studies, including different AI algorithms, also need to be conducted to determine the better algorithm for predicting GI bleeding.

ARTIFICIAL INTELLIGENCE IN MANAGEMENT OF UPPER AND LOWER GASTROINTESTINAL BLEEDING

AI has a broad role in the management of GI bleeding, starting from patient's admission, during endoscopy, to patient's care post-endoscopy or surgery. In patient admission and during pre-endoscopy, AI, especially machine learning, can be used in the risk stratification of patients with GI bleeding. Machine learning can also be used to determine whether the patient needs urgent endoscopy, blood transfusion, or surgical intervention, or if the patient can be safely observed and discharged from the emergency room[17].

Early identification of patients with high-risk GI bleeding is important and can reduce mortality and morbidity. To identify low-risk patients, a GBS score of 0 or 1 can be used to determine whether the patient can be safely discharged from the emergency room (sensitivity 98.6%, specificity 34.6%). However, GBS and other scoring systems such as Rockall and AIMS65 still perform poorly in predicting high-risk patients needing endoscopic treatment or surgical intervention[18].

Shung *et al*[19] developed multiple natural language processing (NLP)-based approaches to identify patients with acute GI bleeding in the emergency room. They used electronic health record-based phenotyping algorithms and compared the performance with the Systematized Nomenclature of Medicine, a standard method to identify patients' conditions. They found that the NLP-based approach performed better than the Systematized Nomenclature of Medicine [PPV 85% (95%CI: 83%-87%) *vs* 69% (95%CI: 66%-72%); $P < 0.001$] in identifying patients with acute GI bleeding.

Seo *et al*[20] developed four machine learning algorithms to predict adverse events and hemodynamic instability in patients with initially stable non-variceal upper GI bleeding. The four machine learning algorithms were logistic regression with regularization, random forest classifier (RF), GB classifier, and voting classifier (VC). The adverse events analyzed included hypotension, mortality, and rebleeding within 7 d. The algorithms were compared with the standard scoring system GBS and Rockall scores. Among the machine learning algorithms, the RF model showed the best performance in predicting mortality (AUC: RF 0.917 *vs* GBS 0.710), while the VC model had the highest accuracies in predicting hypotension (AUC: VC 0.757 *vs* GBS 0.668) and rebleeding within 7 d (AUC: VC 0.733 *vs* GBS 0.694).

In the intensive care unit (ICU), Deshmukh *et al*[21] developed a machine learning model to calculate mortality risk in patients admitted with GI bleeding. They compared the model with the APACHE IVa risk score and found that the model performed better in classifying low-risk patients [AUC: 0.85 (95%CI: 0.80-0.90) *vs* 0.80 (95%CI: 0.73-0.86)]. The model achieved a sensitivity of 100% and specificity of 27%, compared with APACHE IVa risk score with a sensitivity of 100% and specificity of 4%.

Levi *et al*[22] also developed a machine learning algorithm to predict the need for blood transfusion in ICU patients with GI bleeding. Existing scoring systems such as GBS and Rockall score focus on predicting mortality and the need for intervention. They do not assist in determining the level of

monitoring needed for hospitalized patients. Moreover, these scoring systems were validated only for upper GI bleeding. Levi *et al*[22] trained the algorithm on different data sets: MIMIC-III (Medical Information Mart for Intensive Care-III); eICU-CRD (eICU Collaborative Research Database v.2.0); or both. All models performed well with an AUROC > 0.80. A similar study by Shung *et al*[23] also found that a long short-term memory model, a type of Recurrent Neural Network, performed better than a regression-based model (AUROC: 0.65 *vs* 0.56; $P < 0.001$) in determining high-risk GI bleeding patients requiring red blood cell transfusion in the ICU.

In patients with acute lower GI bleeding, Das *et al*[24] constructed ANN and multiple logistic regression models to predict the outcomes of intervention for control of hemorrhage, recurrent bleeding, and death. The models classify patients with lower GI bleeding as high-risk and low-risk patients. The study found that ANN was significantly better than BLEED (accuracy for predicting death 87% *vs* 21%; for recurrent bleeding 89% *vs* 41%; and for intervention 96% *vs* 46%) in internal validation. ANN was also better than multiple logistic regression models in predicting the three outcomes in the external validation (for death 97% *vs* 70%; for recurrent bleeding 93% *vs* 73%; and for intervention 94% *vs* 70%).

Shung *et al*[23], Seo *et al*[20], Deshmukh *et al*[21], and Das *et al*[24] showed that machine learning models could be used in risk stratification for patients with acute upper and lower GI bleeding. More advanced interventions, such as endoscopic or surgical intervention, could be considered in high-risk patients. Therefore, AI could help emergency physicians and gastroenterologists decide patients who might need urgent endoscopic or surgical intervention and help prepare the necessary interventions earlier. Meanwhile, Levi *et al*[22] showed that AI could help determine which patients need tighter monitoring. Many patients with GI bleeding admitted to ICU stop bleeding and do not require further intervention. In hospitals with limited ICU capacities, AI might help determine patients with GI bleeding who may or may not require ICU-level care.

All studies mentioned above used electronic health record data to train the models, making the results readily applicable for the hospital setting. These studies used different machine learning models. Interestingly, Seo *et al*[20] found that different models had different accuracies in determining the risk of different outcomes. Choosing the appropriate machine learning algorithm or model is essential to achieve the highest accuracy. However, there are still not many studies that compare the accuracies between different machine learning models.

During endoscopy, AI might help identify endoscopic characteristics of hemorrhage, such as determining the Forrest classification of peptic ulcer, which will help determine the management needed for the patient. Yen *et al*[25] compared the performance of deep learning with expert and novice endoscopists. They retrieved endoscopic still images of 1694 patients with peptic ulcer bleeding. Four deep learning models were pre-trained with ImageNet. In the end, the Mobile Net V2 model was chosen with the most optimum performance and compared with expert and novice endoscopists. For the 3-class categories, the sensitivity and specificity were 94.83% and 92.36%, respectively. Meanwhile, for the 4-class categories, the sensitivity and specificity were 95.40% and 92.70%, respectively. The deep learning model also had a higher interobserver agreement with expert endoscopists compared to novice endoscopists.

Gastric ulcer is a common medical condition, with a yearly incidence of more than 5 in 1000 adults. However, gastric ulcer also has a risk to develop into gastric cancer. The malignancy rate in endoscopically diagnosed gastric ulcers ranges from 2.4% to 21%. Therefore, early detection of malignant ulcers is important for further treatment and a better prognosis. Several studies have developed AI algorithms to differentiate between malignant and benign gastric ulcers. For example, Klang *et al*[26] developed a CNN model with an AUC of 0.91 (95%CI: 0.85-0.96) with a sensitivity of 92% and specificity of 75%. Similar studies were also conducted by Namikawa *et al*[27], Yoon *et al*[28], and Wu *et al*[29] using the CNN model to differentiate gastric ulcers and early gastric cancers with satisfying performances.

AI also aids in the diagnosis of *Helicobacter pylori* (*H. pylori*) infection. Itoh *et al*[30] developed a CNN model to diagnose *H. pylori* infection, using 149 training images and 30 test images from upper GI endoscopy images. The sensitivity of CNN for detection of *H. pylori* infection was 86.7%, while the specificity was 86.7%, with an AUC of 0.956. Mohan *et al*[31] conducted a systematic review consisting of five studies using CNN for detection of *H. pylori* infection. Images used for the diagnosis were from a combination of white-light, blue laser imaging, and linked color imaging. The pooled accuracy of AI for detecting *H. pylori* infection was 87.1% (95%CI: 81.8-91.1) with a sensitivity of 86.3% and specificity of 87.1%. Meanwhile, endoscopists achieved an accuracy of 82.9% (95%CI: 76.7-87.7), with a sensitivity of 79.6% and specificity of 83.8%.

AI also aids the detection of small bowel bleeding using wireless capsule endoscopy. Le Berre *et al*[32] reviewed 12 studies using various AI classifiers such as color spectrum transformation, MLP (multilayer perceptron network), SVM (support vector machine, a type of machine learning model), joint diagonalization, PCA (principal component analysis), and CNN. The sensitivity from various studies ranged from 87.8% to 100%, while the specificity ranged from 85.8% to 99.9%. The highest accuracy of 99.6% was obtained in a study by Xiao *et al*[33] using deep CNN and 10000 images (8200 training and 1800 test images).

After management in the hospital, AI can be used in identifying the risk of recurrent bleeding in patients with GI bleeding. Wong *et al*[34] developed a machine learning model to predict recurrent bleeding. The model was built based on six parameters (age, baseline haemoglobin, presence of gastric

ulcer, GI diseases, malignancies, and infections). The model identified patients with recurrent ulcer bleeding within 1 year with an AUROC of 0.775 and overall accuracy of 84.3%.

CONCLUSIONS AND FUTURE CHALLENGES

As discussed above, AI, especially machine learning and deep learning, has broad roles in clinical prediction and management of GI bleeding by utilizing data that could help clinicians in their decision-making. Even though AI can utilize a large set of electronic health record data, they might not be able to utilize several important data such as patient's behavior or endoscopic images, which might not be stated in electronic health records or stored in different servers[35].

Since machine learning outcomes depend on the data set, the outcome might not be replicable in other centers. For example, factors that influence the risk of GI bleeding might be different in different centers with different data sets using the same AI algorithm. The data set used for the algorithm training could influence the algorithm's performance. Hence, it is crucial to have a high-quality data set that is well-integrated with the AI system before establishing an AI system[35]. Once established, the integrated electronic health record and AI algorithm system could be copied to be used by different centers.

Adopting AI also has several barriers, especially in developing countries, such as insufficient technological infrastructure and difficulty integrating AI in the routine workflow. Adequate data warehouses, secure analytic platforms, and informatics and machine learning experts must be employed. Some clinicians might be reluctant to substitute clinical judgment with computational analysis. It is important to ensure the healthcare providers' trust before implementing the tool. A contingency plan concerning patients' safety should be established if the algorithm makes an error. Legal framework regarding clinical decision-making by AI and its responsibility is currently unavailable[35,36].

An issue related to the safety of AI is the "black-box" algorithms. Black box AI is any AI system whose inputs and operations are not visible to its users. Many machine learning models are considered a black box, and it is difficult to understand how the algorithm arrived at its conclusion, even for those who trained it. Clinicians who use the algorithm might not realize whether a clinical decision suggested by an AI model is wrong because they do not know how the model arrived at the conclusion. Moreover, AI is still prone to biases. A diagnosis or prognostic algorithm trained with data from mostly Caucasian patients, for example, might not be as accurate for Black or Asian patients. An algorithm developed in high resource settings might not recommend accurate or fair treatment in settings with more limited resources[37].

The black box algorithms also raise legal concerns. It is still unclear if it could be considered medical malpractice when a clinician gives a wrong treatment recommended by a black-box algorithm because they could not review the basis of recommendation. Lawsuits might also be brought to the hospitals that implement the AI algorithm or even to the technology companies that develop the algorithm[37]. Currently, it is recommended to use AI to support a clinical decision that has been already made instead of using AI to create a new clinical decision.

Another ethical concern regarding the use of AI in medicine is patients' privacy. Personal health condition is one of the most legally protected forms of data. Meanwhile, AI is usually provided by start-ups or private technological companies. Previous cases of data breaches or technological companies monetizing their customers' personal information are concerns that need to be addressed. Companies need to provide technical safeguards to maintain data privacy to prevent breaches. Patients should be informed of data uses, and patients should give their consent before their data is used[38].

To prevent misuse of patients' medical information, legal frameworks need to be updated to suit the rapid improvement of AI. Health Insurance Portability and Accountability Act (HIPAA) privacy rule is the United States national standard for protecting individual medical records and other individual health information. An example of a loophole in the regulation is if a genetic company sells their data to pharmaceutical or insurance firms, the HIPAA privacy rule could not apply because DNA information is not legally counted as healthcare[39]. Therefore, regulations concerning patients' privacy and safety need to be revisited and updated to catch up with the improvement of technology. Strict legal penalties should be implemented for those who break the regulations.

FOOTNOTES

Author contributions: Maulahela H proposed the idea of the manuscript, and wrote and edited the manuscript; Annisa NG performed data accusation and manuscript writing.

Conflict-of-interest statement: The authors declare that they have no conflict of interest to disclose.

Country/Territory of origin: Indonesia

ORCID number: Hasan Maulahela 0000-0002-0396-4433; Nagita Gianty Annisa 0000-0001-9708-0454.

S-Editor: Liu JH
L-Editor: Wang TQ
P-Editor: Liu JH

REFERENCES

1 **Yang YJ**, Bang CS. Application of artificial intelligence in gastroenterology. *World J Gastroenterol* 2019; **25**: 1666-1683 [PMID: 31011253 DOI: 10.3748/wjg.v25.i14.1666]
2 **Parasher G**, Wong M, Rawat M. Evolving role of artificial intelligence in gastrointestinal endoscopy. *World J Gastroenterol* 2020; **26**: 7287-7298 [PMID: 33362384 DOI: 10.3748/wjg.v26.i46.7287]
3 **Amisha**, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *J Family Med Prim Care* 2019; **8**: 2328-2331 [PMID: 31463251 DOI: 10.4103/jfmpc.jfmpc_440_19]
4 **Alagappan M**, Brown JRG, Mori Y, Berzin TM. Artificial intelligence in gastrointestinal endoscopy: The future is almost here. *World J Gastrointest Endosc* 2018; **10**: 239-249 [PMID: 30364792 DOI: 10.4253/wjge.v10.i10.239]
5 **Pisters R**, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. *Chest* 2010; **138**: 1093-1100 [PMID: 20299623 DOI: 10.1378/chest.10-0134]
6 **Apostolakis S**, Lane DA, Guo Y, Buller H, Lip GY. Performance of the HEMORR(2)HAGES, ATRIA, and HAS-BLED bleeding risk-prediction scores in patients with atrial fibrillation undergoing anticoagulation: the AMADEUS (evaluating the use of SR34006 compared to warfarin or acenocoumarol in patients with atrial fibrillation) study. *J Am Coll Cardiol* 2012; **60**: 861-867 [PMID: 22858389 DOI: 10.1016/j.jacc.2012.06.019]
7 **Senoo K**, Proietti M, Lane DA, Lip GY. Evaluation of the HAS-BLED, ATRIA, and ORBIT Bleeding Risk Scores in Patients with Atrial Fibrillation Taking Warfarin. *Am J Med* 2016; **129**: 600-607 [PMID: 26482233 DOI: 10.1016/j.amjmed.2015.10.001]
8 **Roldán V**, Marín F, Fernández H, Manzano-Fernandez S, Gallego P, Valdés M, Vicente V, Lip GYH. Predictive value of the HAS-BLED and ATRIA bleeding scores for the risk of serious bleeding in a "real-world" population with atrial fibrillation receiving anticoagulant therapy. *Chest* 2013; **143**: 179-184 [PMID: 22722228 DOI: 10.1378/chest.12-0608]
9 **Herrin J**, Abraham NS, Yao X, Noseworthy PA, Inselman J, Shah ND, Ngufor C. Comparative Effectiveness of Machine Learning Approaches for Predicting Gastrointestinal Bleeds in Patients Receiving Antithrombotic Treatment. *JAMA Netw Open* 2021; **4**: e2110703 [PMID: 34019087 DOI: 10.1001/jamanetworkopen.2021.10703]
10 **Faye A**, Hung K, Cheng K, Parikh N, Knotts R, Freedberg D, Lebwohl B. 626 HAS-BLED Scores Underestimate Gastrointestinal Bleeding Risk Among Those With *H. pylori*. *Am J Gastroenterol* 2019; **114**: S364-S364. [DOI: 10.14309/01.ajg.0000592040.17612.c1]
11 **Capodanno D**, Rossini R, Musumeci G, Lettieri C, Senni M, Valsecchi O, Angiolillo DJ, Lip GY. Predictive accuracy of CHA2DS2-VASc and HAS-BLED scores in patients without atrial fibrillation undergoing percutaneous coronary intervention and discharged on dual antiplatelet therapy. *Int J Cardiol* 2015; **199**: 319-325 [PMID: 26241637 DOI: 10.1016/j.ijcard.2015.07.064]
12 **Choi SY**, Kim MH, Cho YR, Sung Park J, Min Lee K, Park TH, Yun SC. Performance of PRECISE-DAPT Score for Predicting Bleeding Complication During Dual Antiplatelet Therapy. *Circ Cardiovasc Interv* 2018; **11**: e006837 [PMID: 30545256 DOI: 10.1161/CIRCINTERVENTIONS.118.006837]
13 **Shung D**, Simonov M, Gentry M, Au B, Laine L. Machine Learning to Predict Outcomes in Patients with Acute Gastrointestinal Bleeding: A Systematic Review. *Dig Dis Sci* 2019; **64**: 2078-2087 [PMID: 31055722 DOI: 10.1007/s10620-019-05645-z]
14 **Shung DL**, Au B, Taylor RA, Tay JK, Laursen SB, Stanley AJ, Dalton HR, Ngu J, Schultz M, Laine L. Validation of a Machine Learning Model That Outperforms Clinical Risk Scoring Systems for Upper Gastrointestinal Bleeding. *Gastroenterology* 2020; **158**: 160-167 [PMID: 31562847 DOI: 10.1053/j.gastro.2019.09.009]
15 **Loftus TJ**, Brakenridge SC, Croft CA, Smith RS, Efron PA, Moore FA, Mohr AM, Jordan JR. Neural network prediction of severe lower intestinal bleeding and the need for surgical intervention. *J Surg Res* 2017; **212**: 42-47 [PMID: 28550920 DOI: 10.1016/j.jss.2016.12.032]
16 **Ayaru L**, Ypsilantis PP, Nanapragasam A, Choi RC, Thillanathan A, Min-Ho L, Montana G. Prediction of Outcome in Acute Lower Gastrointestinal Bleeding Using Gradient Boosting. *PLoS One* 2015; **10**: e0132485 [PMID: 26172121 DOI: 10.1371/journal.pone.0132485]
17 **Yen HH**, Wu PY, Chen MF, Lin WC, Tsai CL, Lin KP. Current Status and Future Perspective of Artificial Intelligence in the Management of Peptic Ulcer Bleeding: A Review of Recent Literature. *J Clin Med* 2021; **10** [PMID: 34441823 DOI: 10.3390/jcm10163527]
18 **Stanley AJ**, Laine L, Dalton HR, Ngu JH, Schultz M, Abazi R, Zakko L, Thornton S, Wilkinson K, Khor CJ, Murray IA, Laursen SB; International Gastrointestinal Bleeding Consortium. Comparison of risk scoring systems for patients presenting with upper gastrointestinal bleeding: international multicentre prospective study. *BMJ* 2017; **356**: i6432 [PMID:

28053181 DOI: 10.1136/bmj.i6432]
19 **Shung D**, Tsay C, Laine L, Chang D, Li F, Thomas P, Partridge C, Simonov M, Hsiao A, Tay JK, Taylor A. Early identification of patients with acute gastrointestinal bleeding using natural language processing and decision rules. *J Gastroenterol Hepatol* 2021; **36**: 1590-1597 [PMID: 33105045 DOI: 10.1111/jgh.15313]
20 **Seo DW**, Yi H, Park B, Kim YJ, Jung DH, Woo I, Sohn CH, Ko BS, Kim N, Kim WY. Prediction of Adverse Events in Stable Non-Variceal Gastrointestinal Bleeding Using Machine Learning. *J Clin Med* 2020; **9** [PMID: 32796647 DOI: 10.3390/jcm9082603]
21 **Deshmukh F**, Merchant SS. Explainable Machine Learning Model for Predicting GI Bleed Mortality in the Intensive Care Unit. *Am J Gastroenterol* 2020; **115**: 1657-1668 [PMID: 32341266 DOI: 10.14309/ajg.0000000000000632]
22 **Levi R**, Carli F, Arévalo AR, Altinel Y, Stein DJ, Naldini MM, Grassi F, Zanoni A, Finkelstein S, Vieira SM, Sousa J, Barbieri R, Celi LA. Artificial intelligence-based prediction of transfusion in the intensive care unit in patients with gastrointestinal bleeding. *BMJ Health Care Inform* 2021; **28** [PMID: 33455913 DOI: 10.1136/bmjhci-2020-100245]
23 **Shung D**, Huang J, Castro E, Tay JK, Simonov M, Laine L, Batra R, Krishnaswamy S. Neural network predicts need for red blood cell transfusion for patients with acute gastrointestinal bleeding admitted to the intensive care unit. *Sci Rep* 2021; **11**: 8827 [PMID: 33893364 DOI: 10.1038/s41598-021-88226-3]
24 **Das A**, Ben-Menachem T, Cooper GS, Chak A, Sivak MV Jr, Gonet JA, Wong RC. Prediction of outcome in acute lower-gastrointestinal haemorrhage based on an artificial neural network: internal and external validation of a predictive model. *Lancet* 2003; **362**: 1261-1266 [PMID: 14575969 DOI: 10.1016/S0140-6736(03)14568-0]
25 **Yen HH**, Wu PY, Chen MF, Lin WC, Tsai CL, Lin KP. Current Status and Future Perspective of Artificial Intelligence in the Management of Peptic Ulcer Bleeding: A Review of Recent Literature. *J Clin Med* 2021; **10** [PMID: 34441823 DOI: 10.3390/jcm10163527]
26 **Klang E**, Barash Y, Levartovsky A, Barkin Lederer N, Lahat A. Differentiation Between Malignant and Benign Endoscopic Images of Gastric Ulcers Using Deep Learning. *Clin Exp Gastroenterol* 2021; **14**: 155-162 [PMID: 33981151 DOI: 10.2147/CEG.S292857]
27 **Namikawa K**, Hirasawa T, Nakano K, Ikenoyama Y, Ishioka M, Shiroma S, Tokai Y, Yoshimizu S, Horiuchi Y, Ishiyama A, Yoshio T, Tsuchida T, Fujisaki J, Tada T. Artificial intelligence-based diagnostic system classifying gastric cancers and ulcers: comparison between the original and newly developed systems. *Endoscopy* 2020; **52**: 1077-1083 [PMID: 32503056 DOI: 10.1055/a-1194-8771]
28 **Yoon HJ**, Kim S, Kim JH, Keum JS, Oh SI, Jo J, Chun J, Youn YH, Park H, Kwon IG, Choi SH, Noh SH. A Lesion-Based Convolutional Neural Network Improves Endoscopic Detection and Depth Prediction of Early Gastric Cancer. *J Clin Med* 2019; **8** [PMID: 31454949 DOI: 10.3390/jcm8091310]
29 **Wu L**, Wang J, He X, Zhu Y, Jiang X, Chen Y, Wang Y, Huang L, Shang R, Dong Z, Chen B, Tao X, Wu Q, Yu H. Deep learning system compared with expert endoscopists in predicting early gastric cancer and its invasion depth and differentiation status (with videos). *Gastrointest Endosc* 2022; **95**: 92-104.e3 [PMID: 34245752 DOI: 10.1016/j.gie.2021.06.033]
30 **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 [PMID: 29399610 DOI: 10.1055/s-0043-120830]
31 **Mohan BP**, Khan SR, Kassab LL, Ponnada S, Mohy-Ud-Din N, Chandan S, Dulai PS, Kochhar GS. Convolutional neural networks in the computer-aided diagnosis of *Helicobacter pylori* infection and non-causal comparison to physician endoscopists: a systematic review with meta-analysis. *Ann Gastroenterol* 2021; **34**: 20-25 [PMID: 33414617 DOI: 10.20524/aog.2020.0542]
32 **Le Berre C**, Sandborn WJ, Aridhi S, Devignes MD, Fournier L, Smaïl-Tabbone M, Danese S, Peyrin-Biroulet L. Application of Artificial Intelligence to Gastroenterology and Hepatology. *Gastroenterology* 2020; **158**: 76-94.e2 [PMID: 31593701 DOI: 10.1053/j.gastro.2019.08.058]
33 **Xiao Jia**, Meng MQ. A deep convolutional neural network for bleeding detection in Wireless Capsule Endoscopy images. *Annu Int Conf IEEE Eng Med Biol Soc* 2016; **2016**: 639-642 [PMID: 28268409 DOI: 10.1109/EMBC.2016.7590783]
34 **Wong GL**, Ma AJ, Deng H, Ching JY, Wong VW, Tse YK, Yip TC, Lau LH, Liu HH, Leung CM, Tsang SW, Chan CW, Lau JY, Yuen PC, Chan FK. Machine learning model to predict recurrent ulcer bleeding in patients with history of idiopathic gastroduodenal ulcer bleeding. *Aliment Pharmacol Ther* 2019; **49**: 912-918 [PMID: 30761584 DOI: 10.1111/apt.15145]
35 **Sheikhtaheri A**, Sadoughi F, Hashemi Dehaghi Z. Developing and using expert systems and neural networks in medicine: a review on benefits and challenges. *J Med Syst* 2014; **38**: 110 [PMID: 25027017 DOI: 10.1007/s10916-014-0110-5]
36 **Shung DL**. Advancing care for acute gastrointestinal bleeding using artificial intelligence. *J Gastroenterol Hepatol* 2021; **36**: 273-278 [PMID: 33624892 DOI: 10.1111/jgh.15372]
37 **Gerke S**, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. In: Artificial Intelligence in Healthcare [Internet]. Elsevier; 2020 [cited 2021 Dec 11]. 295–336. Available from: https://Linkinghub.elsevier.com/retrieve/pii/B9780128184387000125
38 **Murdoch B**. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics* 2021; **22**: 122 [PMID: 34525993 DOI: 10.1186/s12910-021-00687-3]
39 **Ahuja AS**. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ* 2019; **7**: e7702 [PMID: 31592346 DOI: 10.7717/peerj.7702]

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 February 28; 3(1): 21-27

DOI: 10.35712/aig.v3.i1.21

ISSN 2644-3236 (online)

MINIREVIEWS

Artificial intelligence and human liver allocation: Potential benefits and ethical implications

Marcos Mucenic, Ajacio Bandeira de Mello Brandão, Claudio Augusto Marroni

Specialty type: Transplantation

Provenance and peer review: Invited article; externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): 0
Grade C (Good): C, C, C
Grade D (Fair): D
Grade E (Poor): 0

P-Reviewer: Alkhayyat M, Ferrarese A, Lee KS

Received: December 24, 2021
Peer-review started: December 24, 2021
First decision: January 26, 2022
Revised: February 13, 2022
Accepted: February 23, 2022
Article in press: February 23, 2022
Published online: February 28, 2022

Marcos Mucenic, Liver Transplant Adult Group, Irmandade da Santa Casa de Misericórdia de Porto Alegre, Porto Alegre 90020-090, RS, Brazil

Ajacio Bandeira de Mello Brandão, Claudio Augusto Marroni, Hepatology, Universidade Federal de Ciencias da Saude de Porto Alegre, Porto Alegre 90050-170, RS, Brazil

Corresponding author: Marcos Mucenic, MD, PhD, Doctor, Medical Assistant, Liver Transplant Adult Group, Irmandade da Santa Casa de Misericórdia de Porto Alegre, Av. Independencia, 75, Porto Alegre 90020-090, RS, Brazil. mmucenic@gmail.com

Abstract

Since its implementation almost two decades ago, the urgency allocation policy has improved the survival of patients on the waiting list for liver transplantation worldwide. The Model for End-Stage Liver Disease score is widely used to predict waiting list mortality. Due to some limitations related to its use, there is an active investigation to develop other prognostic scores. Liver allocation (LA) entails complex decision-making, and grafts are occasionally not directed to the recipients who are more likely to survive. Prognostic scores have, thus far, failed to predict post-operatory survival. Furthermore, the increasing use of marginal donors is associated with worse outcomes. Adequate donor-recipient pairing could help avoid retransplantation or futile procedures and reduce postoperative complications, mortality, hospitalization time, and costs. Artificial intelligence has applications in several medical fields. Machine learning algorithms (MLAs) use large amounts of data to detect unforeseen patterns and complex interactions between variables. Artificial neural networks and decision trees were the most common forms of MLA tested on LA. Some researchers have shown them to be superior for predicting waiting list mortality and graft failure than conventional statistical methods. These promising techniques are increasingly being considered for implementation.

Key Words: Liver transplantation; Liver cirrhosis; Artificial intelligence; Prognosis; Survival; Machine learning

Core Tip: This review discusses the ethical aspects and current advancements in liver allocation (LA). It summarizes the concept of artificial intelligence and focuses on the latest developments of machine learning algorithms as applied to predicting waiting list mortality and LA. To date, only a few research groups have published works on this field; they also wrote reviews on the subject. Our minireview offers a thorough and impartial view of the topic, and we hope this will alert other potential researchers to this promising field.

Citation: Mucenic M, de Mello Brandão AB, Marroni CA. Artificial intelligence and human liver allocation: Potential benefits and ethical implications. *Artif Intell Gastroenterol* 2022; 3(1): 21-27
URL: https://www.wjgnet.com/2644-3236/full/v3/i1/21.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i1.21

INTRODUCTION

Liver transplantation (LT) is the treatment of choice for patients with terminal liver disease[1]. LT is increasingly performed worldwide; however, organ scarcity remains a significant challenge for transplant teams[2], placing greater weight on the need for efficient liver allocation (LA). Therefore, correct organ allocation is of paramount importance.

An optimal allocation system for LT should balance considerations of equity (equal opportunity to receive the graft), need (to reduce waiting list mortality), utility (maximizing the overall life-years gained), and benefit (optimizing outcomes from each organ transplanted)[3].

Urgency criteria are based on need, prioritizing grafts to the most critically ill patients. Survival without LT is estimated through prognostic models, such as the model for end-stage liver disease (MELD). MELD is a validated score derived solely from laboratory test results (total bilirubin, serum creatinine, and prothrombin time). MELD is simple and accurate, and it predicts the 3-mo mortality of candidates with an area under the receiver operating characteristic curve (AUROC) of 0.83[4]. Since 2002, the MELD score was adopted by the United Network for Organ Sharing (UNOS) to rank waitlisted patients in order of urgency in the United States (US). Several countries followed this organ allocation system (sickest first)[5]. While post-transplant survival for the sickest is lesser than that of patients with better physiological reserves, they are the ones who benefit the most from LT. Patients with a MELD score of 31-34 had a relative life expectancy 43 times higher than those who remained on the list, and patients with a MELD score of 35-40 were 128 times more likely to survive[6]. MELD was further refined after studies showed adding serum sodium concentration to the formula (MELD-Na) improves risk stratification. This system replaced MELD for LA in some countries, such as the USA, Canada, and Brazil[5,7-9]. Godfrey *et al*[10] advised of a possible loss of predictive accuracy of the MELD score over time, reaching an AUROC of only 0.70 in 2015. This may be due to changes in the epidemiology and treatment of liver diseases and increasing age and comorbidities. Despite several valid concerns about the model, it remains the most widely used.

Urgency allocation models have no value in predicting survival after LT[11,12]. Additionally, the donor pool has been expanded in the last two decades. Although the use of marginal livers (*e.g.*, older donors, steatotic livers, and donation after cardiac death) has been necessary in this regard, it increases the risk of graft failure and postoperative complications, adding further complexity to the matter of allocation[13,14]. Living donor liver transplantation is another strategy to expand the donor pool; however, it poses an inherent risk to healthy donors. Its proportion to the total number of LT is small[3].

The MELD score does not reflect mortality risk in compensated patients with hepatocellular carcinoma (HCC). Exception points are granted to candidates with HCC, one of the leading LT indications worldwide. Currently, the prioritization of HCC candidates varies from one country to another, and there is no international consensus on the matter[15]. Due to the excessive advantage conferred by these exception points, there have been some changes in global allocation policies[16-18]. Notwithstanding these revisions, HCC candidates still have increased transplant rates, decreased risk of delisting, and worse post-transplant prognosis[15].

Outcomes after LT depend on both the preoperative condition of the recipient and donor "quality". Utility criteria have been sought to offer grafts to recipients with greater chances of survival, estimating the outcome based on donor and recipient characteristics[3,16]. While this would decrease the odds for older and sicker patients to receive an organ, the overall post-transplant survival could improve. Better selection avoids retransplantation or futile procedures and reduces postoperative morbidity, hospitalization time, and costs. The survival benefit is quantifiable by estimating waiting list survival and post-transplant outcomes. A benefit-based system could balance urgency and utility in allocation decisions. For a benefit-based allocation to be successful, transplant teams would need an accurate model to predict post-transplant survival.

Although the concept of applying donor-related variables to an algorithm had been used before, Feng *et al*[19] devised the term "donor risk index" (DRI), wherein they identified seven donor characteristics that predicted graft failure. Other researchers further investigated this interesting concept, adding cold ischemia time and organ origin (national or regional). However, DRI has not been widely adopted owing to the following main reasons cited by surgical teams: the inaccuracy to predict survival, exclusion of other relevant risk factors, and difficulty of explaining its concept to the recipients[20].

Since neither candidate nor donor factors are predictive of survival following transplantation, scores that include both variables have been described over the previous years. Halldorson *et al*[21] proposed adding donor age to the MELD score, creating the D-MELD, demonstrating a survival disadvantage when combining higher donor age with higher MELD recipients. Schaubel *et al*[22] proposed the balance of risk (BAR) score, which included donor, recipient, and procurement surgery variables. For an estimated 5-year survival, the c-statistic reached an AUROC of 0.63. Rana *et al*[23] developed a complex score named survival outcomes following liver transplantation (SOFT), containing > 100 variables. It reached an AUROC of 0.70 to predict 3-mo survival after LT. In 2018, the United Kingdom introduced allocation rules based on benefit. Each graft is offered nationally to the recipient predicted to have the greatest survival benefit from that specific graft. LA is based on the transplant benefit score (TBS), calculated by 21 and 7 receptor and donor criteria, respectively. TBS reflects the difference in days between expected 5-year post-operatory survival and expected 5-year waiting list survival. The model reduced deaths on the waiting list and maximized post-operatory life-years[24].

The external validity of these scores is limited by several factors, such as ethnicity, regional differences in the allocation and transplantation procedures, and changes in practice over time. Since the scores are based on logistic regression (LR) models, they depend upon the assumption of independence of each variable and are limited when facing nonlinear variable interactions. Complex donor-related models are considered difficult to implement, and their accuracy can be limited by the large number of variables that impact survival and possible undetected confounding factors. They have not been validated by other researchers or found wide acceptance to date. Moreover, they were not designed for an ideal donor-recipient matching[23,25].

Therefore, transplant teams are faced with a complex decision-making process when having to choose recipients for LA. Objective criteria would exempt the medical staff from difficult decisions and assess whether patients are excessively sick to be transplanted[3,7,26]. "Artificial intelligence" (AI) or machine learning algorithms (MLAs) are under increasingly active investigation for this use[27].

AI

AI is a general term used to describe any application wherein computer systems perform tasks normally associated with human intelligence. It can be a substitute for human subjectivity and limitations[28]. AI encompasses simple automated tasks and increasingly complex fields, such as machine learning, deep learning, and artificial neural networks (ANNs).

In classical programming, the computer is supplied with an algorithm and a dataset to provide an output. Machine learning, in contrast, supplies the computer with data and associated outputs, which it uses to create an algorithm that describes the relationship between the two. These MLAs can detect patterns and improve their analysis over time with further data[29]. MLAs can analyze any number of variables and are not driven (or limited) by hypothesis. This method detects nonlinear patterns within large datasets wherein multiple interactions between variables can occur. MLA can accommodate numerous interdependent variables and improve as more cases are increasingly analyzed[30,31].

Typically, MLA applied to healthcare fall into the category of supervised learning techniques. These algorithms learn the associations between input and labeled outcome data. The following are the basic steps of supervised ML: (1) Acquire a dataset and split it into separate training, validation, and test datasets; (2) use training and validation datasets to create a model that analyzes the association between data and outcomes; and (3) evaluate the model *via* the test dataset to determine how well it predicts outcomes. There are other techniques used, such as unsupervised learning, wherein data are not labeled to find out previously unknown patterns. Semi-supervised learning is particularly useful for datasets that contain both labeled and unlabeled data. Reinforcement learning uses the consequences of their actions to learn to determine the optimal behavior for a given context[29,32].

The decision tree (DT) is a supervised learning technique primarily used for classification tasks (categorical variables). It consists of a hierarchically organized structure of nodes that makes predictions by splitting (branching) the data. Each split can connect to a new root node or attach to a terminal or "leaf" node. A random forest (RF) is an ensemble method that produces multiple DTs[29].

One of the advantages of DTs for healthcare applications is their interpretability. However, each node is determined in isolation without considering the possible impact of future splits. This can fail to capture the dataset's underlying characteristics. This disadvantage stimulated the development of optimal classification trees (OCTs). This type of DT is formed entirely in a single step, allowing each split to be determined with full knowledge of all other splits[33].

ANN is an MLA inspired by biological neural networks. Each ANN contains nodes (analogous to cell bodies) that communicate with other nodes *via* connections (analogous to axons and dendrites), with multiple layers (an input layer, an output layer, and a hidden layer between them) of connected mathematical functions. ANNs can capture complex nonlinear relationships in data, allowing for sophisticated supervised and unsupervised learning tasks[28,29,32].

Support vector machine (SVM) is another type of MLA. This method organizes data by variable classes in a nonlinear modality, subsequently separating by a hyperplane and forming multidimensional planes in space using these data. It can be used for classification or regression problems[34].

AI APPLIED FOR THE PREDICTION OF MORTALITY IN THE WAITING LIST

MLA has shown promising results for predicting 3-mo mortality on the waiting list. The simulation model was based on OCTs. OCTs were fed with > 1.6 million observations and trained, validated, and tested. The result showed a slightly superior AUROC than MELD-Na for predicting death or unsuitability for LT (0.859 *vs* 0.841). The authors argue that this system would save at least 418 more lives annually in the US[35]. An interesting point in this simulation model compared with MELD was the increased allocation of livers to non-HCC patients and a decreased number of waitlist deaths and removals for both HCC and non-HCC patients. Their results await further validation.

Cucchetti *et al*[36] applied an ANN model to predict the 3-mo mortality of patients awaiting LT in the pre-MELD era. The analysis included only laboratory values (liver biochemical and function tests, creatinine, and hemogram). The participants were randomly divided into training and testing groups in a proportion of 75%-25%. After each of the 10 training sessions, ANN was tested on the remaining individuals who were not selected for training. The most accurate ANN system was tested in a retrospective cohort in another LT center. The performance of ANN in predicting the 3-mo mortality was superior to that of MELD (AUROC, 0.98 *vs* 0.86). Results were similar for the external validation cohort (0.96 and 0.86, respectively).

AI APPLIED FOR LA

Hundreds of variables contribute to multiple decisions made in an organ transplant. For each record, the UNOS database collects > 400 parameters. AI can theoretically improve the outcomes of allocation strategies[30,31]. An optimal outcome would be a decreased number of retransplant procedures, excellent graft and overall survival, and decreasing rate of waiting list mortality.

In a large multicenter Spanish study, Briceño *et al*[37] applied 57 variables (26, 19, 6, and 6 from the recipient, donor, retrieval procedure, and transplant procedure, respectively) for each donor-recipient pair to predict 3-mo graft survival. A total of 1003 liver transplants were analyzed. This sample had been previously described in a pilot study by the same group of researchers[38]. The following were the two models of ANN used: a positive-survival (PS) model to predict the 3-mo graft survival rate after LT and a negative-survival (NS) model to predict the 3-mo graft failure rate. These ANN models are MLAs that simulate a biological neural system. In this study, 90% of the data was used for training and 10% for testing, which was repeated ten times to allow all patterns to participate in both phases. Subsequently, the model that correctly classified the most D-R pairs was chosen. PS methodology was slightly superior to common statistical methods (multiple regression, MR) to predict graft survival (90.8% *vs* 87.7%). NS methodology performed worse for predicting graft loss; however, it was far superior to MR (71.4% *vs* 3.4%). Finally, the AUROC curves were compared with previously reported scores (MELD, D-MELD, DRI, P-SOFT, SOFT, and BAR). In the PS model, ANN had an AUROC of 0.81, significantly higher than that of other conventional statistical methods (which varied from 0.42 to 0.68). In the NS model, NN had an AUROC of 0.82, which was also significantly higher than that of the other scores (which varied from 0.42 to 0.61). Of the previously reported scores, BAR showed the best AUROC results.

Ayllón *et al*[30] applied D-R pairing with ANN on 858 cases in a large-volume LT center (King's College Hospital). They used the same PS and NS models described by Briceño *et al*[37], with some differences in the included variables. AUROCs for PS (0.94) and NS (0.94) 3 mo after LT were significantly more accurate than that of BAR, which is the second-best score (AUROC, 0.84). Furthermore, the researchers performed a 12-mo analysis, and when ANN was used to predict graft survival and loss (0.78 and 0.82, respectively), their results were better than that of the best prediction achieved by other scores (BAR, 0.71).

Lau *et al*[39] applied ML techniques in an Australian single-center study with 180 LTs. A bootstrap sample containing approximately 63% of the cases was used for the training set, and the remaining data were used for testing. This process was repeated 1000 times. RF classifiers and ANN were used on the overall top 15 ranked characteristics to determine the performance as measured by AUROC values. Graft failure (NS) within 30 d was the primary outcome, and NS at 3 mo postoperatively was the secondary outcome. The results were subsequently compared with those of DRI and SOFT scores. The AUROC for the 30-d NS was 0.818 with RF and 0.835 with NN compared with 0.64 and 0.68 with SOFT

Table 1 Overview of original works on artificial intelligence applied to liver allocation

Ref.	Sample size and location	AI model(s)	Outcomes analyzed	Results	Comments
Briceño *et al* [37], 2014	1003 LT recipients (multicenter in Spain)	ANN with PS and NS model with D-R pairing	3-mo Graft survival (PS); 3-mo Graft failure (NS)	AUROC 0.81 (PS); AUROC 0.82 (NS)	Superior to BAR score (0.68 for PS, 0.61 for NS). Other conventional statistics fared worse
Ayllón *et al* [30], 2018	858 LT recipients (single-center in England)	ANN (PS and NS) with D-R pairing	3-mo Graft survival (PS); 3-mo Graft failure (NS)	AUROC 0.90 (PS); AUROC 0.90 (NS)	Superior to BAR score (AUROC 0.71). Same model above on a different population (external validation)
Lau *et al*[39], 2017	180 LT recipients (single-center in Australia)	ANN and RF	30-d and 3-mo Graft failure (NS)	30-d prediction: AUROC 0.82 (RF) AUROC 0.835 (ANN)	Superior to SOFT and DRI scores
Guijo-Rubio *et al*[40], 2021	20456 LT recipients (5-yr survival) to 37646 LT recipients (3-mo survival) UNOS database	ANN, RF, DT, SVM, MLP	3-mo, 1 yr, 2 yr, 5 yr survival	AUROC up to 0.618 (3-mo), 0.614 (1-yr), 0.611 (2-yr), 0.644 (5-yr)	No superiority compared to conventional statistics (LR was slightly superior)

AI: Artificial intelligence; ANN: Artificial neural network; AUROC: Area under the receiver operating characteristic curve; BAR: Balance of risk; DRI: Donor risk index; DT: Decision tree; D-R: Donor-receptor; MLP: Multilayer perceptron; NS: Negative-survival; LT: Liver transplantation; LR: Logistic regression; PS: Positive-survival; SOFT: Survival outcomes following liver transplantation; SVM: Support vector machines; RF: Random forest; UNOS: United network of organ sharing.

and DRI scores, respectively. The AUROC decreased to 0.715 with RF and 0.56 with NN to predict the 3-mo NS (including 90 cases in the analysis).

Contrastingly, MLA did not prove to be superior to LR for predicting survival after adult LT using donor-recipient matching in a large database[40]. Four different survival endpoints were analyzed using the UNOS database, including 3-mo and 1-, 2-, and 5-year survivals, varying from 37646 transplants in the 3-mo analysis to 20456 transplants in the 5-year analysis. A total of 28 variables were considered, including recipient, donor, and matching variables. Several types of MLA were used, including ANN, RF, DT, and SVM. The researchers suggested that this lack of accuracy of MLA could be ascribed to database limitations. The highest AUROCs were obtained with LR, followed by RF.

Table 1 summarizes the original works on AI applied to LA that were discussed above.

A systematic review of AI for predicting post-transplant survival was performed by Wingfield *et al* [34] Nine publications were included, and articles were considered of good quality overall. ANN and LR were the most common types of MLA and conventional statistical methods, respectively. MLAs were similar or superior to conventional statistics.

CONCLUSION

Although prioritization criteria have successfully reduced mortality in the waiting list, there is room for refinement in mortality prediction and a growing need for improving LA guidelines. Conventional statistical methods have, thus far, failed to provide a useful and widely applicable allocation score. AI can bring meaningful insights to this field. Paradoxically, MLA could help improve the ethics of LA, increasing waitlist and post-transplant survival, preferably with quality-adjusted life-years gained. The results obtained, thus far, are promising; however, we must consider the limitations of AI in medicine. First, its accuracy depends upon the availability of accurate, organized, and thorough datasets. In this regard, the algorithms also depend upon the data used to feed them, and regional particularities can limit their validation. Further, the clinical relevance of the results must be properly evaluated by experts in the field. Moreover, it can be challenging for the lay population to understand and accept LA decisions based on AI analysis. Finally, the health providers must make the final decision, at least while the concepts of ethics and justice rest upon the human mind.

FOOTNOTES

Author contributions: de Mello Brandão AB wrote about liver allocation; Marroni CA wrote about artificial intelligence; Mucenic M wrote about the applications of artificial intelligence on liver allocation and on the prediction of waiting list mortality, wrote abstract and conclusions, and revised the writing style; all three writers revised the paper.

Conflict-of-interest statement: There is no conflict of interest associated with any of the authors of this manuscript

Country/Territory of origin: Brazil

ORCID number: Marcos Mucenic 0000-0001-9389-2236; Ajacio Bandeira de Mello Brandão 0000-0001-8411-5654; Claudio Augusto Marroni 0000-0002-1718-6548.

S-Editor: Liu JH
L-Editor: A
P-Editor: Liu JH

REFERENCES

1 **Halliday N**, Westbrook RH. Liver transplantation: need, indications, patient selection and pre-transplant care. *Br J Hosp Med (Lond)* 2017; **78**: 252-259 [PMID: 28489446 DOI: 10.12968/hmed.2017.78.5.252]
2 **Trotter JF**. Liver transplantation around the world. *Curr Opin Organ Transplant* 2017; **22**: 123-127 [PMID: 28151809 DOI: 10.1097/MOT.0000000000000392]
3 **Keller EJ**, Kwo PY, Helft PR. Ethical considerations surrounding survival benefit-based liver allocation. *Liver Transpl* 2014; **20**: 140-146 [PMID: 24166860 DOI: 10.1002/lt.23780]
4 **Wiesner R**, Edwards E, Freeman R, Harper A, Kim R, Kamath P, Kremers W, Lake J, Howard T, Merion RM, Wolfe RA, Krom R; United Network for Organ Sharing Liver Disease Severity Score Committee. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology* 2003; **124**: 91-96 [PMID: 12512033 DOI: 10.1053/gast.2003.50016]
5 **Tschuor C**, Ferrarese A, Kuemmerli C, Dutkowski P, Burra P, Clavien PA; Liver Allocation Study Group. Allocation of liver grafts worldwide - Is there a best system? *J Hepatol* 2019; **71**: 707-718 [PMID: 31199941 DOI: 10.1016/j.jhep.2019.05.025]
6 **Locke JE**, Shelton BA, Olthoff KM, Pomfret EA, Forde KA, Sawinski D, Gray M, Ascher NL. Quantifying Sex-Based Disparities in Liver Allocation. *JAMA Surg* 2020; **155**: e201129 [PMID: 32432699 DOI: 10.1001/jamasurg.2020.1129]
7 **Kim WR**, Kremers WK. Benefits of "the benefit model" in liver transplantation. *Hepatology* 2008; **48**: 697-698 [PMID: 18752329 DOI: 10.1002/hep.22497]
8 **Nagai S**, Chau LC, Schilke RE, Safwan M, Rizzari M, Collins K, Yoshida A, Abouljoud MS, Moonka D. Effects of Allocating Livers for Transplantation Based on Model for End-Stage Liver Disease-Sodium Scores on Patient Outcomes. *Gastroenterology* 2018; **155**: 1451-1462.e3 [PMID: 30056096 DOI: 10.1053/j.gastro.2018.07.025]
9 **da Silva Machado AG**, de Medeiros Fleck A Jr, Marroni C, Zanotelli ML, Cantisani G, de Mello Brandão AB. Impact of MELD score implementation on liver allocation: experience at a Brazilian center. *Ann Hepatol* 2013; **12**: 440-447 [PMID: 23619261]
10 **Godfrey EL**, Malik TH, Lai JC, Mindikoglu AL, Galván NTN, Cotton RT, O'Mahony CA, Goss JA, Rana A. The decreasing predictive power of MELD in an era of changing etiology of liver disease. *Am J Transplant* 2019; **19**: 3299-3307 [PMID: 31394020 DOI: 10.1111/ajt.15559]
11 **Freeman RB**, Harper A, Edwards EB. Excellent liver transplant survival rates under the MELD/PELD system. *Transplant Proc* 2005; **37**: 585-588 [PMID: 15848465 DOI: 10.1016/j.transproceed.2004.12.099]
12 **Benckert C**, Quante M, Thelen A, Bartels M, Laudi S, Berg T, Kaisers U, Jonas S. Impact of the MELD allocation after its implementation in liver transplantation. *Scand J Gastroenterol* 2011; **46**: 941-948 [PMID: 21443420 DOI: 10.3109/00365521.2011.568521]
13 **Busuttil RW**, Tanaka K. The utility of marginal donors in liver transplantation. *Liver Transpl* 2003; **9**: 651-663 [PMID: 12827549 DOI: 10.1053/jlts.2003.50105]
14 **Tector AJ**, Mangus RS, Chestovich P, Vianna R, Fridell JA, Milgrom ML, Sanders C, Kwo PY. Use of extended criteria livers decreases wait time for liver transplantation without adversely impacting posttransplant survival. *Ann Surg* 2006; **244**: 439-450 [PMID: 16926570 DOI: 10.1097/01.sla.0000234896.18207.fa]
15 **Sacleux SC**, Samuel D. A Critical Review of MELD as a Reliable Tool for Transplant Prioritization. *Semin Liver Dis* 2019; **39**: 403-413 [PMID: 31242526 DOI: 10.1055/s-0039-1688750]
16 **Cholongitas E**, Burroughs AK. The evolution in the prioritization for liver transplantation. *Ann Gastroenterol* 2012; **25**: 6-13 [PMID: 24713804]
17 **Heimbach JK**. United States liver allocation. *Curr Opin Organ Transplant* 2020; **25**: 104-109 [PMID: 32142481 DOI: 10.1097/MOT.0000000000000740]
18 **Rodríguez S**, Fleck AM Jr, Mucenic M, Marroni C, Brandão A. Hepatocellular carcinoma patients are advantaged in the current brazilian liver transplant allocation system. A competing risk analysis. *Arq Gastroenterol* 2020; **57**: 19-23 [PMID: 32294731 DOI: 10.1590/S0004-2803.202000000-05]
19 **Feng S**, Goodrich NP, Bragg-Gresham JL, Dykstra DM, Punch JD, DebRoy MA, Greenstein SM, Merion RM. Characteristics associated with liver graft failure: the concept of a donor risk index. *Am J Transplant* 2006; **6**: 783-790 [PMID: 16539636 DOI: 10.1111/j.1600-6143.2006.01242.x]

20 **Mataya L**, Aronsohn A, Thistlethwaite JR Jr, Friedman Ross L. Decision making in liver transplantation--limited application of the liver donor risk index. *Liver Transpl* 2014; **20**: 831-837 [PMID: 24692309 DOI: 10.1002/lt.23879]

21 **Halldorson JB**, Bakthavatsalam R, Fix O, Reyes JD, Perkins JD. D-MELD, a simple predictor of post liver transplant mortality for optimization of donor/recipient matching. *Am J Transplant* 2009; **9**: 318-326 [PMID: 19120079 DOI: 10.1111/j.1600-6143.2008.02491.x]

22 **Schaubel DE**, Guidinger MK, Biggins SW, Kalbfleisch JD, Pomfret EA, Sharma P, Merion RM. Survival benefit-based deceased-donor liver allocation. *Am J Transplant* 2009; **9**: 970-981 [PMID: 19341419 DOI: 10.1111/j.1600-6143.2009.02571.x]

23 **Rana A**, Hardy MA, Halazun KJ, Woodland DC, Ratner LE, Samstein B, Guarrera JV, Brown RS Jr, Emond JC. Survival outcomes following liver transplantation (SOFT) score: a novel method to predict patient survival following liver transplantation. *Am J Transplant* 2008; **8**: 2537-2546 [PMID: 18945283 DOI: 10.1111/j.1600-6143.2008.02400.x]

24 **Gimson A**. Development of a UK liver transplantation selection and allocation scheme. *Curr Opin Organ Transplant* 2020; **25**: 126-131 [PMID: 32073485 DOI: 10.1097/MOT.0000000000000743]

25 **Neuberger J**, Heimbach JK. Allocation of deceased-donor livers - Is there a most appropriate method? *J Hepatol* 2019; **71**: 654-656 [PMID: 31451285 DOI: 10.1016/j.jhep.2019.07.013]

26 **Persad G**, Wertheimer A, Emanuel EJ. Principles for allocation of scarce medical interventions. *Lancet* 2009; **373**: 423-431 [PMID: 19186274 DOI: 10.1016/S0140-6736(09)60137-9]

27 **Ferrarese A**, Sartori G, Orrù G, Frigo AC, Pelizzaro F, Burra P, Senzolo M. Machine learning in liver transplantation: a tool for some unsolved questions? *Transpl Int* 2021; **34**: 398-411 [PMID: 33428298 DOI: 10.1111/tri.13818]

28 **Christou CD**, Tsoulfas G. Challenges and opportunities in the application of artificial intelligence in gastroenterology and hepatology. *World J Gastroenterol* 2021; **27**: 6191-6223 [PMID: 34712027 DOI: 10.3748/wjg.v27.i37.6191]

29 **Choi RY**, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl Vis Sci Technol* 2020; **9**: 14 [PMID: 32704420 DOI: 10.1167/tvst.9.2.14]

30 **Ayllón MD**, Ciria R, Cruz-Ramírez M, Pérez-Ortiz M, Gómez I, Valente R, O'Grady J, de la Mata M, Hervás-Martínez C, Heaton ND, Briceño J. Validation of artificial neural networks as a methodology for donor-recipient matching for liver transplantation. *Liver Transpl* 2018; **24**: 192-203 [PMID: 28921876 DOI: 10.1002/lt.24870]

31 **Briceño J**, Ayllón MD, Ciria R. Machine-learning algorithms for predicting results in liver transplantation: the problem of donor-recipient matching. *Curr Opin Organ Transplant* 2020; **25**: 406-411 [PMID: 32487891 DOI: 10.1097/MOT.0000000000000781]

32 **Ahn JC**, Connell A, Simonetto DA, Hughes C, Shah VH. Application of Artificial Intelligence for the Diagnosis and Treatment of Liver Diseases. *Hepatology* 2021; **73**: 2546-2563 [PMID: 33098140 DOI: 10.1002/hep.31603]

33 **Bertsimas D**, Dunn J. Optimal Classification Trees. *Mach Learn* 2017; **106**: 1039-1082 [DOI: 10.1007/s10994-017-5633-9]

34 **Wingfield LR**, Ceresa C, Thorogood S, Fleuriot J, Knight S. Using Artificial Intelligence for Predicting Survival of Individual Grafts in Liver Transplantation: A Systematic Review. *Liver Transpl* 2020; **26**: 922-934 [PMID: 32274856 DOI: 10.1002/lt.25772]

35 **Bertsimas D**, Kung J, Trichakis N, Wang Y, Hirose R, Vagefi PA. Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation. *Am J Transplant* 2019; **19**: 1109-1118 [PMID: 30411495 DOI: 10.1111/ajt.15172]

36 **Cucchetti A**, Vivarelli M, Heaton ND, Phillips S, Piscaglia F, Bolondi L, La Barba G, Foxton MR, Rela M, O'Grady J, Pinna AD. Artificial neural network is superior to MELD in predicting mortality of patients with end-stage liver disease. *Gut* 2007; **56**: 253-258 [PMID: 16809421 DOI: 10.1136/gut.2005.084434]

37 **Briceño J**, Cruz-Ramírez M, Prieto M, Navasa M, Ortiz de Urbina J, Orti R, Gómez-Bravo MÁ, Otero A, Varo E, Tomé S, Clemente G, Bañares R, Bárcena R, Cuervas-Mons V, Solórzano G, Vinaixa C, Rubín A, Colmenero J, Valdivieso A, Ciria R, Hervás-Martínez C, de la Mata M. Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: results from a multicenter Spanish study. *J Hepatol* 2014; **61**: 1020-1028 [PMID: 24905493 DOI: 10.1016/j.jhep.2014.05.039]

38 **Cruz-Ramírez M**, Hervás-Martínez C, Fernández JC, Briceño J, de la Mata M. Predicting patient survival after liver transplantation using evolutionary multi-objective artificial neural networks. *Artif Intell Med* 2013; **58**: 37-49 [PMID: 23489761 DOI: 10.1016/j.artmed.2013.02.004]

39 **Lau L**, Kankanige Y, Rubinstein B, Jones R, Christophi C, Muralidharan V, Bailey J. Machine-Learning Algorithms Predict Graft Failure After Liver Transplantation. *Transplantation* 2017; **101**: e125-e132 [PMID: 27941428 DOI: 10.1097/TP.0000000000001600]

40 **Guijo-Rubio D**, Briceño J, Gutiérrez PA, Ayllón MD, Ciria R, Hervás-Martínez C. Statistical methods versus machine learning techniques for donor-recipient matching in liver transplantation. *PLoS One* 2021; **16**: e0252068 [PMID: 34019601 DOI: 10.1371/journal.pone.0252068]

Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: bpgoffice@wjgnet.com
Help Desk: https://www.f6publishing.com/helpdesk
https://www.wjgnet.com

ISSN 2644-3236 (online)

Artificial Intelligence in *Gastroenterology*

Artif Intell Gastroenterol 2022 April 28; 3(2): 28-79

Published by Baishideng Publishing Group Inc

A I G
Artificial Intelligence in Gastroenterology

Contents **Bimonthly Volume 3 Number 2 April 28, 2022**

MINIREVIEWS

Artificial Intelligence in Gastroenterology

Bimonthly Volume 3 Number 2 April 28, 2022

Contents

ABOUT COVER

Editorial Board Member of *Artificial Intelligence in Gastroenterology*, Michele Finotti, MD, Adjunct Professor, Department of Hepatobiliary Surgery, General Surgeon at the Treviso Hospital, Padova 35123, Italy.
mi6le@libero.it

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

The *AIG* is now abstracted and indexed in Reference Citation Analysis, China Science and Technology Journal Database.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: *Wen-Wen Qi*; **Production Department Director:** *Xiang Li*; **Editorial Office Director:** *Jin-Lei Wang*.

NAME OF JOURNAL
Artificial Intelligence in Gastroenterology

ISSN
ISSN 2644-3236 (online)

LAUNCH DATE
July 28, 2020

FREQUENCY
Bimonthly

EDITORS-IN-CHIEF
Rajvinder Singh, Ferruccio Bonino

EDITORIAL BOARD MEMBERS
https://www.wjgnet.com/2644-3236/editorialboard.htm

PUBLICATION DATE
April 28, 2022

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

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 April 28; 3(2): 28-35

DOI: 10.35712/aig.v3.i2.28

ISSN 2644-3236 (online)

MINIREVIEWS

Liver surgery for colorectal metastasis: New paths and new goals with the help of artificial intelligence

Valeria Tonini, Gabriele Vigutto, Riccardo Donati

Specialty type: Oncology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): 0
Grade D (Fair): D
Grade E (Poor): 0

P-Reviewer: Da Costa AC, United Kingdom; Karamarkovic AR, Serbia

Received: December 20, 2021
Peer-review started: December 21, 2021
First decision: March 12, 2022
Revised: March 28, 2022
Accepted: April 19, 2022
Article in press: April 19, 2022
Published online: April 28, 2022

Valeria Tonini, Department of Medical and Surgical Sciences, Sant' Orsola Hospital University of Bologna, Bologna 40138, Italy

Gabriele Vigutto, Department of Medical and Surgical Sciences, St Orsola Hospital, University of Bologna, Bologna 40138, Italy

Riccardo Donati, Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Bologna 40138, Italy

Corresponding author: Valeria Tonini, MD, PhD, Professor, Surgeon, Department of Medical and Surgical Sciences, University of Bologna, Sant' Orsola Hospital, Via Massarenti 9, Bologna 40138, Italy. valeria.tonini@unibo.it

Abstract

Colorectal cancer is one of the most common neoplasia with an high risk to metastatic spread. Improving medical and surgical treatment is moving along with improving the precision of diagnosis and patient's assessment, the latter two aided more and more with the use of artificial intelligence (AI). The management of colorectal liver metastasis is multidisciplinary, and surgery is the main option. After the diagnosis, a surgical assessment of the patient is fundamental. Reaching a R0 resection with a proper remnant liver volume can be done using new techniques involving also artificial intelligence. Considering the recent application of artificial intelligence as a valid substitute for liver biopsy in chronic liver diseases, several authors tried to apply similar techniques to pre-operative imaging of liver metastasis. Radiomics showed good results in identifying structural changes in a unhealthy liver and in evaluating the prognosis after a liver resection. Recently deep learning has been successfully applied in estimating the remnant liver volume before surgery. Moreover AI techniques can help surgeons to perform an early diagnosis of neoplastic relapse or a better differentiation between a colorectal metastasis and a benign lesion. AI could be applied also in the histopathological diagnostic tool. Although AI implementation is still partially automatized, it appears faster and more precise than the usual diagnostic tools and, in the short future, could become the new gold standard in liver surgery.

Key Words: Colo-rectal cancer; Liver metastasis; Artificial intelligence; Radiomics; Deep learning

Core Tip: Colon cancer is one of the most frequent cancers that unfortunately has a high risk of metastatic spread especially to the liver. The treatment of liver metastases is multidisciplinary, but surgery remains undoubtedly the main act. The results in the treatment of liver metastases have improved significantly over the years, but we continue to seek further paths of improvement. A new path, to which we currently entrust many hopes, is that of artificial intelligence, which could bring revolutionary solutions both in the diagnosis of liver metastases, and as a useful guide for surgical techniques. The purpose of this article is to summarize the latest news reported in the literature and possible research developments on this topic.

Citation: Tonini V, Vigutto G, Donati R. Liver surgery for colorectal metastasis: New paths and new goals with the help of artificial intelligence. *Artif Intell Gastroenterol* 2022; 3(2): 28-35
URL: https://www.wjgnet.com/2644-3236/full/v3/i2/28.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i2.28

INTRODUCTION

Nowadays, colorectal cancer is one of the most common neoplasia in Western countries and among the main causes of death for oncologic diseases[1,2]. Between 30% and 50% of patients with colorectal cancer will develop liver metastasis during their life and surgical resection remains a fundamental treatment[1,2]. The improvement of surgical techniques, along with the use of newer and better schemes of chemotherapy, will increase the chances of a longer disease free survival for these patients[3]. Meanwhile, artificial intelligence (AI) is infiltrating healthcare exponentially and it has already been applied to several fields related to gastroenterology and hepatology[4,5].

HEPATOBILIARY SURGERY FOR COLORECTAL METASTASIS

The treatment of colorectal metastasis is generally multidisciplinary, involving many professional figures and multiples pathways[1,2]. Discussing other therapies, such as chemotherapy or radiotherapy, is beyond the scope of this article.

Surgical treatment always goes with hepatic resection[1]. All metastatic patients need to undergo several pre-operative exams for a better definition of the disease and its extent: a thoraco-abdominal contrast-enhanced CT scan and/or a contrast-enhanced MRI[1,6]. The use of routine PET/CT scan remains controversial[1,7]. The main goals during the assessment are evaluating the extent of the hepatic disease and searching for any extra hepatic localization of disease, the latter one is an exclusion criteria for any kind of hepatic resection[1,8].

Once surgery is considered, the assessment becomes more operative: new main goals are estimating how complex is performing a R0 resection and evaluating the liver remnant volume[1]. Clearly, a R0 resection should be achieved to increase the disease free survival and the overall survival, but the well-known 1cm border of healthy tissue is now reconsidered due to the increasing effectiveness of chemotherapy and the complexity of the resection[1,9,10]. At the same time, the size of the remnant liver must be evaluated with a three dimensional CT volumetry and it should be more than 20% in a healthy liver, more than 30% in post- systemic chemotherapy liver and more than 40% in a cirrhotic liver[1,11]. In case of an insufficient liver remnant volume, a portal vein embolization can be considered to increase the size to the residual liver[1,12], while, in case of bilateral lesions with a majority of them in one lobe, a two-stage hepatectomy with or without contralateral limited resections can be done[1,13]. Finally, a mini invasive approach should be considered if the surgeon is experienced in these techniques, considering the well-known advantages of mini invasive approaches[14].

RADIOMICS AND ARTIFICIAL INTELLIGENCE APPLIED TO MEDICAL IMAGING

The recent advent of artificial intelligence has changed the paradigm in the field of medical imaging interpretation together with radiomics. Artificial intelligence is a discipline that aims at mimicking the function of human brain in solving complex problems using computers. Machine learning and deep learning are branches of AI in which machines are thought how to learn from data using analytical models and algorithms. While machine learning methods usually require less computation on the computer side and more human intervention, deep learning may involve a huge amount of information

(from which stems the adjective "deep") and thus requires high performance computers, but less or no human intervention.

Radiomics is a tool for extensive extraction of quantitative features from medical imaging[4] and can be applied to ultrasound (US), magnetic resonance imaging (MRI), positron emission tomography (PET) and computed tomography (CT). The science of radiomics has taken advantage of machine learning with great benefit for medicine in general. The large amount of information provided by radiomics together with the improvements in AI have given raise to new methods of reading and interpreting medical images. Experts in different domains have now the opportunity to make less challenging the hard task of interpreting images thanks to this machine-aided approach. As shown in Figure 1 the workflow of conventional radiomics and AI applied to medical imaging is split in image acquisition, preprocessing, segmentation, features extraction and selection, model construction and training, model testing and evaluation. In conventional radiomics, one of the main prerequisites during the phase of image acquisition and preprocessing is a certain degree of standardization of the processes, in order to obtain a database with images that have comparable characteristics. Images segmentation consists in locating lesions manually or with the aid of a computer, in order to identify the region of interest o volumes of interests. Feature extraction and selection is a crucial step in machine learning paradigms in order to obtain a subset of quantitative parameters that are given as inputs to train the analytical model. In case of radiomics, these can be shape-based features (*e.g.* size, shape, location), histogram features (or others first-order features like standard deviation and variance), textual features (*e.g.* tumor heterogeneity) and other higher order features extracted with wavelet transforms or Laplacian filters. In the phase of model construction, it is important to choose the analytical engine that gives the best results in term of performance in relation to the selected features. To do so, several models can be chosen and then tested such as linear regression, support vector machines, decision tree, random forest, K-Means. The evaluation of the models and the assessment of their performance is inferred from indicators and methods such as the receiver operating characteristic, nomograms and the decision curve analysis.

Whereas conventional radiomics is still a widely used approach in medical image analysis, in recent years, deep learning has been introduced in the clinical practice thanks to its promising results[7]. This technique can reach high levels of performance while not requiring manual human intervention in the phases of image segmentation and features extraction (Figure 2). In this paradigm, features are in fact automatically selected by a neural network to maximize the performance of the algorithm (called "backpropagation algorithm"). However, a larger amount of data (*e.g.* of number of medical images) is commonly needed to train the neural network models using backpropagation. Among the most popular techniques are multilayer perceptron networks, convolutional neural network, long short-term memory recurrent neural networks. Such as in conventional radiomics, different deep learning techniques can be applied to the input data in order to obtain the best performance.

ARTIFICIAL INTELLIGENCE APPLIED TO LIVER SURGERY

Recently, artificial intelligence was applied to various fields in medicine, including general surgery and hepatology[4,5], as seen in Table 1. Decharatanachart *et al*[4] published a meta-analysis on AI supported imaging and standard liver biopsy. They showed a similar prediction rate for liver cirrhosis without the risk of complications of a biopsy and without the usual interpretation bias of ultrasonography. Meanwhile, Christou *et al*[5] focused more on the possibility of integrating diagnosis and management in several gastroenterological diseases, such as inflammatory bowel disease (IBD), *Helicobacter pylori* infection and gastric cancer, and several hepatic diseases, such as HCV infection and cirrhosis[5]. On one hand, they described how the use of machine learning and CAD can increase sensibility and specificity of a standard endoscopic or radiologic exam; on the other hand they describe the limitations of AI[5].

One the of the main application of AI in liver surgery is in the pre-operative imaging. Park *et al*[15] described the use of radiomics and deep learning in liver diseases. Radiomics appears to be an effective way to analyse the structural changes of an unhealthy liver, comparable to the standard techniques like biopsies[15,16]. Furthermore, radiomics is already in use for determining the prognosis after surgical resection or radiofrequency[17] for hepatocellular carcinoma, especially related to micro vascular invasion[15,18]. Deep learning finds its best application in liver segmentation, where it is fundamental in estimating the liver remnant volume and the fat ratio in post chemotherapy liver[15,19,20]. Fang *et al*[21] focused on the implementation of deep learning in CT-guided biopsy to obtain a better localization of the lesion. In addition they presented a basic algorithm that could offer good results. At the same time, Winkel *et al*[22] compared manual segmentation and automatic segmentation with the use of deep learning showing a similar efficacy of the automatic segmentation with a faster elaboration of the images.

Focusing on focal liver lesions, Zhou *et al*[23] illustrated a 5 categories classification based on dynamic contrast-enhanced CT scan with a deep learning software: applying this classification, the radiologist would be able to make a diagnosis between a carcinoma and a benign lesion without biopsy[23,24]. They reported the application of deep learning to a contrast-enhanced ultrasonography (CEUS) to better

Table 1 Main implementation of artificial intelligence in hepatology and liver surgery

Ref.	Type of paper	Main topic	AI implementation
Decharatanachart *et al*[4], 2021	Meta-analysis	Chronic liver diseases	Diagnosis and staging of liver fibrosis without biopsy
Christou *et al*[5], 2021	Review	IBD, GI bleeding and chronic liver diseases	Increasing accuracy of gold standard diagnostic exams
Park *et al*[15], 2020	Review	Liver diseases	Staging of liver disease and prognosis after liver resection or chemotherapy
Wang *et al*[16], 2012	Survey	Liver imaging	Diagnosis of structural changes in healthy liver
Shan *et al*[19], 2019	Research article	Liver imaging (CT)	Prediction of early recurrence after HCC resection/RF
Hu *et al*[18], 2019	Research article	Liver imaging (US)	Evaluating microvascular invasion in HCC
Iranmanesh *et al*[19], 2014	Research article	Liver imaging (CT)	Evaluating portal pressure without invasive methods
Wang *et al*[23], 2019	Research article	Liver imaging (CT/MRI)	Using liver segmentation to an automatized liver biometry
Fang *et al*[21], 2020	Research article	Liver imaging	Using liver segmentation to more accurate localization of a hepatic lesion
Winkel *et al*[22], 2020	Comparative study	Liver imaging	Comparing a fully automated liver segmentation to a manual one
Zhou *et al*[23], 2019	Review	Liver imaging	Detecting hepatic lesions, characterized them and evaluate a response after treatment
Yasaka *et al*[24], 2018	Retrospective study	Liver imaging (CT)	Differentiation between benign and malignant hepatic lesions
Guo *et al*[25], 2018	Research article	Liver imaging (US)	Differentiation between benign and malignant hepatic lesions
Schmauch *et al*[26], 2019	Research article	Liver imaging (US)	Differentiation between benign and malignant hepatic lesions
Tiyarattanachai *et al*[27], 2021	Retrospective study	Liver imaging (US)	Detect and diagnose hepatic lesions
Perez *et al*[28], 2020	Review	HCC	Improving diagnosis and evaluation after ancillary treatments
Vivanti *et al*[29], 2017	Research article	Liver neoplasia	Evaluating post chemotherapy response
Li *et al*[30], 2015	Research article	Liver imaging (CT)	Differentiation between benign and malignant hepatic lesions
Hamm *et al*[31], 2019	Research article	Liver imaging (MRI)	Differentiation between benign and malignant hepatic lesions
Zhang *et al*[32], 2018	Research article	HCC	Differentiation between healthy and tumoral tissue in patient's liver
Preis *et al*[33], 2011	Research article	Liver imaging (PET)	Differentiation between benign and malignant hepatic lesions
Chen *et al*[34], 2020	Review	Liver surgery	Implementation in pre and post operative care
Nakayama *et al*[35], 2017	Retrospective study	Liver surgery	Use of 3D modeling to improve hepatice resection
Zhang *et al*[36], 2018	Prospective study	Liver surgery	Diagnosis and treatment of perihilar CCC
Vorontsov *et al*[37], 2019	Retrospective study	Liver surgery	Improving CRM identification and segmentation
Chartrand *et al*[39], 2017	Comparative study	Liver imaging	Improving liver segmentation and volumetry
Cancian *et al*[40], 2021	Research article.	Liver pathology	Better assessment pf tumor microenvironment

AI: Artificial intelligence; CCC: Cholangiocarcinoma; CRM: Colo-rectal metastases; CT: Computed tomography; GI: Gastrointestinal; HCC: Hepatocellular carcinoma; IBD: inflammatory bowel disease; MRI: Magnetic resonance imaging; PET: Positron emission tomography; US: Ultrasound.

distinguish between a benign and malignant lesion of the liver, showing again a better performance using AI techniques compared to the conventional technique[23,25]. Schmauch *et al*[26] presented a glimpse of future implementations of the standard ultrasonography where the use of a deep learning technique could drastically improve the diagnostic value of a widespread imaging such as US. Similarly, Tiyarattanachai *et al*[27] implemented a deep learning software for the US reporting a better outcome both in prevention and diagnosis of a focal liver lesion. Closely related to our main topic, Perez *et al*[28]

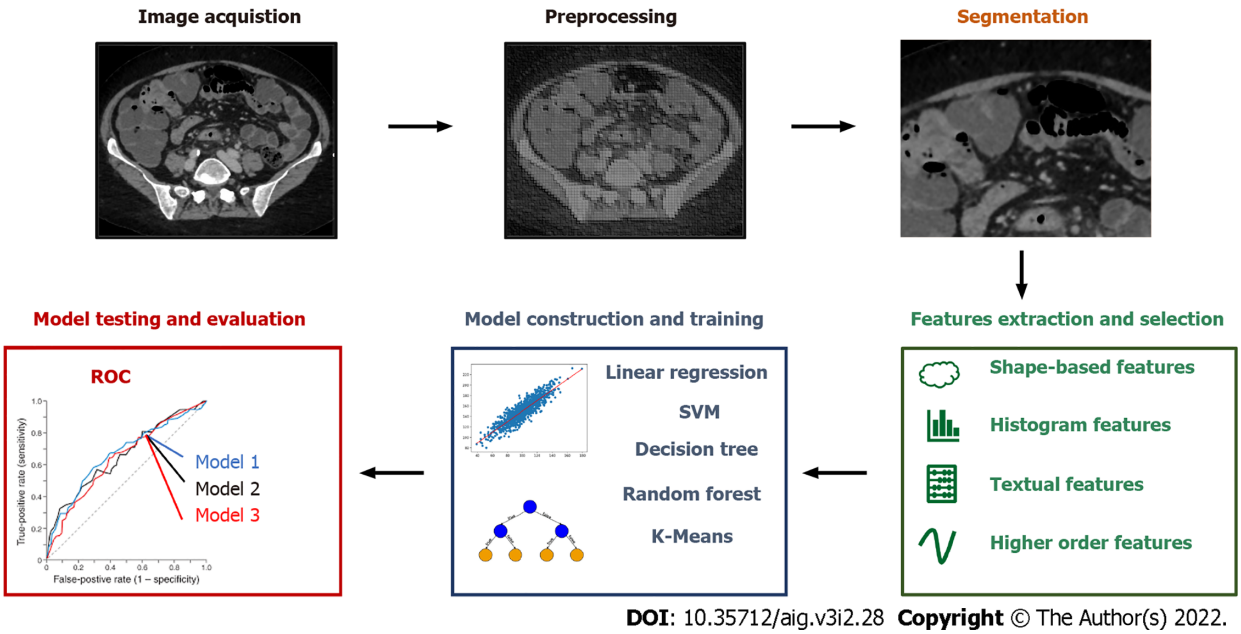


Figure 1 Workflow of conventional radiomics with machine learning.

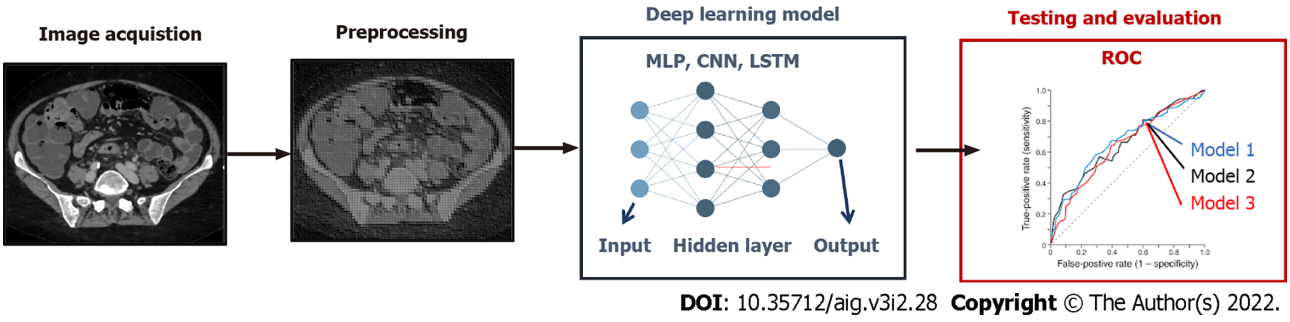


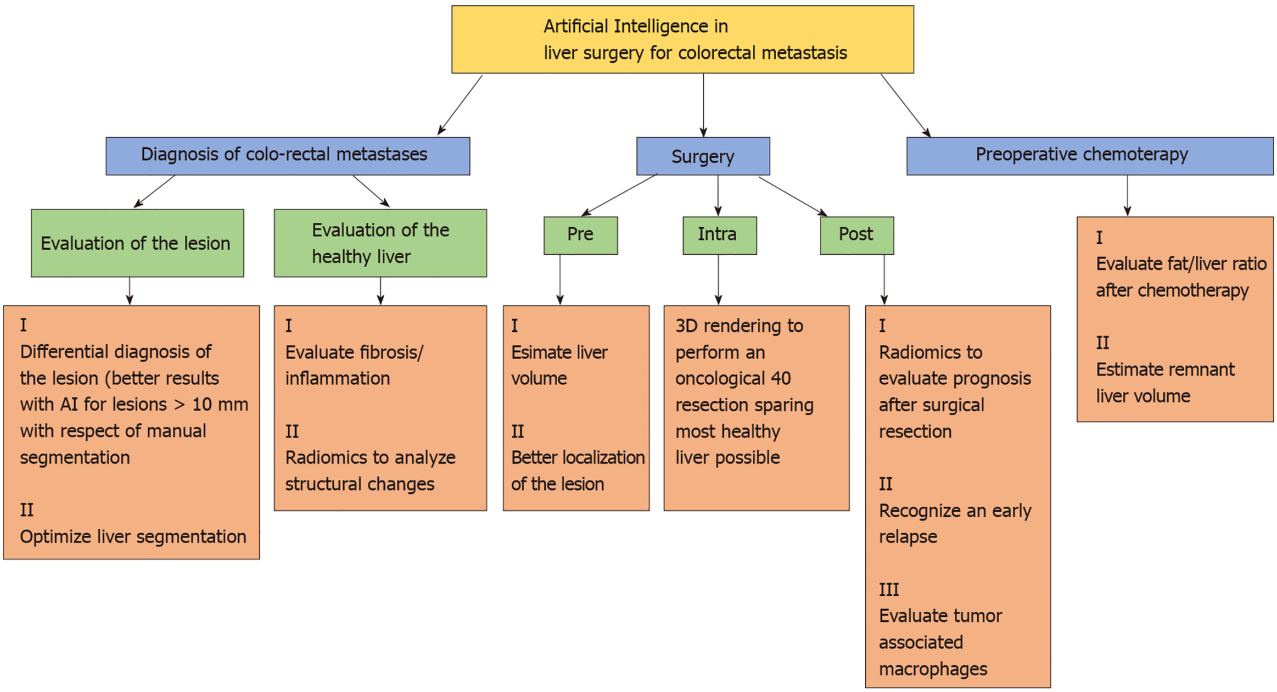
Figure 2 Deep learning techniques applied to radiomics.

proposed a review on the management of hepatocellular carcinoma using AI for diagnosis, treatment and prognosis. Combining the US deep learning software[26] and the contrast-enhanced CT scan deep learning software[24,29,30], the clinician can reach a diagnosis on a focal liver lesion without the use of liver biopsy; in case of more doubts, a deep learning MRI software[31,32] and a deep learning PET software[33] are under external verification, but they appears promising.

Another main application of AI in liver surgery is the pre-operative patient assessment. The second part of the paper of Perez *et al*[28] described how the combined effort of US, CT, MRI scan and deep learning software increase the precision of the hepatic resection and the early recognition of a relapse. Beside the use of AI in the diagnosis, Chen *et al*[34] described the intra-operative advantages of using 3D rendering of the patient's liver to study and apply the best approach for a liver resection and, at the same time, to keep the same 3D model during the operation for a more intuitive way to reach the aforementioned R0 resection[34-36].

About colorectal liver metastasis, Voronstov *et al*[37] proposed a CT-based deep learning software to automatize and improve the recognition of metastasis rather than benign focal liver lesions. Detection performance of the software was still lower for lesion smaller than 10 mm, but it became more precise for lesions between 10 and 20 mm[37]. Manual liver segmentation was still more accurate for lesions smaller than 10mm, but it reached the same value for lesions greater than 10 mm and it was more efficient in lesions greater than 20 mm; the same results appeared considering lesion-volume estimation [37]. The authors also stated that all software calculations for an automatized or semi-automatized recognition and evaluation of metastasis is a significantly faster procedure than the usual manual one, as expected[37-39].

Within the same sphere, Cancian *et al*[40] focused on the analysis of the tumor microenvironment using a deep learning technique to evaluate the morphology of tumor associated macrophages. The same group recently described how different macrophages' morphologies are associated with different outcomes and therapeutic responses in colorectal liver metastasis[41], so they developed a pipeline



DOI: 10.35712/aig.v3i2.28

Figure 3 Main implementation of artificial intelligence in diagnosis and treatment of colo-rectal liver metastases.

using a CAD tool to process faster the histopathological slides. Although the pipeline is still under verification for a fully automatic application, a combined use of a manual and automatic approach showed a better and faster identification of macrophages' morphologies[40,41]. In Figure 3 are shown in a schematic manner the main tools of AI in diagnosis and treatment of colo-rectal liver metastases.

CONCLUSION

Artificial intelligence and deep learning offer new hopes in diagnosis and therapy of the liver metastasis. Therefore new promising research directions open up in this field, that must be confirmed with larger studies in the future.

FOOTNOTES

Author contributions: Tonini V, Vigutto G, and Donati R wrote the paper together and equally contributed to the final manuscript; all authors have read and approved the final manuscript.

Conflict-of-interest statement: The authors declare that they have no conflict of interest.

Country/Territory of origin: Italy

ORCID number: Valeria Tonini 0000-0003-3130-2928; Gabriele Vigutto 0000-0001-5933-0838; Riccardo Donati 0000-0003-0272-817X.

S-Editor: Liu JH
L-Editor: A
P-Editor: Liu JH

REFERENCES

1 **Dhir M**, Sasson AR. Surgical Management of Liver Metastases From Colorectal Cancer. *J Oncol Pract* 2016; **12**: 33-39 [PMID: 26759464 DOI: 10.1200/JOP.2015.009407]

2 **Manfredi S**, Lepage C, Hatem C, Coatmeur O, Faivre J, Bouvier AM. Epidemiology and management of liver metastases from colorectal cancer. *Ann Surg* 2006; **244**: 254-259 [PMID: 16858188 DOI: 10.1097/01.sla.0000217629.94941.cf]

3 **Folprecht G**. Liver Metastases in Colorectal Cancer. *Am Soc Clin Oncol Educ Book* 2016; **35**: e186-e192 [PMID: 27249722 DOI: 10.1200/EDBK_159185]

4 **Decharatanachart P**, Chaiteerakij R, Tiyarattanachai T, Treeprasertsuk S. Application of artificial intelligence in chronic liver diseases: a systematic review and meta-analysis. *BMC Gastroenterol* 2021; **21**: 10 [PMID: 33407169 DOI: 10.1186/s12876-020-01585-5]

5 **Christou CD**, Tsoulfas G. Challenges and opportunities in the application of artificial intelligence in gastroenterology and hepatology. *World J Gastroenterol* 2021; **27**: 6191-6223 [PMID: 34712027 DOI: 10.3748/wjg.v27.i37.6191]

6 **van Kessel CS**, Buckens CF, van den Bosch MA, van Leeuwen MS, van Hillegersberg R, Verkooijen HM. Preoperative imaging of colorectal liver metastases after neoadjuvant chemotherapy: a meta-analysis. *Ann Surg Oncol* 2012; **19**: 2805-2813 [PMID: 22396005 DOI: 10.1245/s10434-012-2300-z]

7 **Moulton CA**, Gu CS, Law CH, Tandan VR, Hart R, Quan D, Fairfull Smith RJ, Jalink DW, Husien M, Serrano PE, Hendler AL, Haider MA, Ruo L, Gulenchyn KY, Finch T, Julian JA, Levine MN, Gallinger S. Effect of PET before liver resection on surgical management for colorectal adenocarcinoma metastases: a randomized clinical trial. *JAMA* 2014; **311**: 1863-1869 [PMID: 24825641 DOI: 10.1001/jama.2014.3740]

8 **Ekberg H**, Tranberg KG, Andersson R, Lundstedt C, Hägerstrand I, Ranstam J, Bengmark S. Determinants of survival in liver resection for colorectal secondaries. *Br J Surg* 1986; **73**: 727-731 [PMID: 3756436 DOI: 10.1002/bjs.1800730917]

9 **Dhir M**, Lyden ER, Wang A, Smith LM, Ullrich F, Are C. Influence of margins on overall survival after hepatic resection for colorectal metastasis: a meta-analysis. *Ann Surg* 2011; **254**: 234-242 [PMID: 21694583 DOI: 10.1097/SLA.0b013e318223c609]

10 **Pawlik TM**, Scoggins CR, Zorzi D, Abdalla EK, Andres A, Eng C, Curley SA, Loyer EM, Muratore A, Mentha G, Capussotti L, Vauthey JN. Effect of surgical margin status on survival and site of recurrence after hepatic resection for colorectal metastases. *Ann Surg* 2005; **241**: 715-722, discussion 722 [PMID: 15849507 DOI: 10.1097/01.sla.0000160703.75808.7d]

11 **Ribero D**, Chun YS, Vauthey JN. Standardized liver volumetry for portal vein embolization. *Semin Intervent Radiol* 2008; **25**: 104-109 [PMID: 21326551 DOI: 10.1055/s-2008-1076681]

12 **van Lienden KP**, van den Esschert JW, de Graaf W, Bipat S, Lameris JS, van Gulik TM, van Delden OM. Portal vein embolization before liver resection: a systematic review. *Cardiovasc Intervent Radiol* 2013; **36**: 25-34 [PMID: 22806245 DOI: 10.1007/s00270-012-0440-y]

13 **Lam VW**, Laurence JM, Johnston E, Hollands MJ, Pleass HC, Richardson AJ. A systematic review of two-stage hepatectomy in patients with initially unresectable colorectal liver metastases. *HPB (Oxford)* 2013; **15**: 483-491 [PMID: 23750490 DOI: 10.1111/j.1477-2574.2012.00607.x]

14 **Ye SP**, Qiu H, Liao SJ, Ai JH, Shi J. Mini-invasive *vs* open resection of colorectal cancer and liver metastases: A meta-analysis. *World J Gastroenterol* 2019; **25**: 2819-2832 [PMID: 31236004 DOI: 10.3748/wjg.v25.i22.2819]

15 **Park HJ**, Park B, Lee SS. Radiomics and Deep Learning: Hepatic Applications. *Korean J Radiol* 2020; **21**: 387-401 [PMID: 32193887 DOI: 10.3348/kjr.2019.0752]

16 **Wang S**, Summers RM. Machine learning and radiology. *Med Image Anal* 2012; **16**: 933-951 [PMID: 22465077 DOI: 10.1016/j.media.2012.02.005]

17 **Shan QY**, Hu HT, Feng ST, Peng ZP, Chen SL, Zhou Q, Li X, Xie XY, Lu MD, Wang W, Kuang M. CT-based peritumoral radiomics signatures to predict early recurrence in hepatocellular carcinoma after curative tumor resection or ablation. *Cancer Imaging* 2019; **19**: 11 [PMID: 30813956 DOI: 10.1186/s40644-019-0197-5]

18 **Hu HT**, Wang Z, Huang XW, Chen SL, Zheng X, Ruan SM, Xie XY, Lu MD, Yu J, Tian J, Liang P, Wang W, Kuang M. Ultrasound-based radiomics score: a potential biomarker for the prediction of microvascular invasion in hepatocellular carcinoma. *Eur Radiol* 2019; **29**: 2890-2901 [PMID: 30421015 DOI: 10.1007/s00330-018-5797-0]

19 **Iranmanesh P**, Vazquez O, Terraz S, Majno P, Spahr L, Poncet A, Morel P, Mentha G, Toso C. Accurate computed tomography-based portal pressure assessment in patients with hepatocellular carcinoma. *J Hepatol* 2014; **60**: 969-974 [PMID: 24362073 DOI: 10.1016/j.jhep.2013.12.015]

20 **Wang K**, Mamidipalli A, Retson T, Bahrami N, Hasenstab K, Blansit K, Bass E, Delgado T, Cunha G, Middleton MS, Loomba R, Neuschwander-Tetri BA, Sirlin CB, Hsiao A; members of the NASH Clinical Research Network. Automated CT and MRI Liver Segmentation and Biometry Using a Generalized Convolutional Neural Network. *Radiol Artif Intell* 2019; **1** [PMID: 32582883 DOI: 10.1148/ryai.2019180022]

21 **Fang X**, Xu S, Wood BJ, Yan P. Deep learning-based liver segmentation for fusion-guided intervention. *Int J Comput Assist Radiol Surg* 2020; **15**: 963-972 [PMID: 32314228 DOI: 10.1007/s11548-020-02147-6]

22 **Winkel DJ**, Weikert TJ, Breit HC, Chabin G, Gibson E, Heye TJ, Comaniciu D, Boll DT. Validation of a fully automated liver segmentation algorithm using multi-scale deep reinforcement learning and comparison versus manual segmentation. *Eur J Radiol* 2020; **126**: 108918 [PMID: 32171914 DOI: 10.1016/j.ejrad.2020.108918]

23 **Zhou LQ**, Wang JY, Yu SY, Wu GG, Wei Q, Deng YB, Wu XL, Cui XW, Dietrich CF. Artificial intelligence in medical imaging of the liver. World J Gastroenterol 2019; 25: 672-682 [PMID: 30783371 DOI: 10.3748/wjg.v25.i6.672]

24 **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 [PMID: 29059036 DOI: 10.1148/radiol.2017170706]

25 **Guo LH**, Wang D, Qian YY, Zheng X, Zhao CK, Li XL, Bo XW, Yue WW, Zhang Q, Shi J, Xu HX. A two-stage multi-view learning framework based computer-aided diagnosis of liver tumors with contrast enhanced ultrasound images. *Clin Hemorheol Microcirc* 2018; **69**: 343-354 [PMID: 29630528 DOI: 10.3233/CH-170275]

26 **Schmauch B**, Herent P, Jehanno P, Dehaene O, Saillard C, Aubé C, Luciani A, Lassau N, Jégou S. Diagnosis of focal liver lesions from ultrasound using deep learning. *Diagn Interv Imaging* 2019; **100**: 227-233 [PMID: 30926443 DOI: 10.1016/j.diii.2019.02.009]

27 **Tiyarattanachai T**, Apiparakoon T, Marukatat S, Sukcharoen S, Geratikornsupuk N, Anukulkarnkusol N, Mekaroonkamol P, Tanpowpong N, Sarakul P, Rerknimitr R, Chaiteerakij R. Development and validation of artificial intelligence to detect and diagnose liver lesions from ultrasound images. *PLoS One* 2021; **16**: e0252882 [PMID: 34101764 DOI: 10.1371/journal.pone.0252882]

28 **Jiménez Pérez M**, Grande RG. Application of artificial intelligence in the diagnosis and treatment of hepatocellular carcinoma: A review. *World J Gastroenterol* 2020; **26**: 5617-5628 [PMID: 33088156 DOI: 10.3748/wjg.v26.i37.5617]

29 **Vivanti R**, Szeskin A, Lev-Cohain N, Sosna J, Joskowicz L. Automatic detection of new tumors and tumor burden evaluation in longitudinal liver CT scan studies. *Int J Comput Assist Radiol Surg* 2017; **12**: 1945-1957 [PMID: 28856515 DOI: 10.1007/s11548-017-1660-z]

30 **Li W**, Jia F, Hu Q. Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks. *J Comput Commun* 2015; **3**: 146-151

31 **Hamm CA**, Wang CJ, Savic LJ, Ferrante M, Schobert I, Schlachter T, Lin M, Duncan JS, Weinreb JC, Chapiro J, Letzen B. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. *Eur Radiol* 2019; **29**: 3338-3347 [PMID: 31016442 DOI: 10.1007/s00330-019-06205-9]

32 **Zhang F**, Yang J, Nezami N, Laage-Gaupp F, Chapiro J, De Lin M, Duncan J. Liver Tissue Classification Using an Auto-context-based Deep Neural Network with a Multi-phase Training Framework. *Patch Based Tech Med Imaging (2018)* 2018; **11075**: 59-66 [PMID: 32432233 DOI: 10.1007/978-3-030-00500-9_7]

33 **Preis O**, Blake MA, Scott JA. Neural network evaluation of PET scans of the liver: a potentially useful adjunct in clinical interpretation. *Radiology* 2011; **258**: 714-721 [PMID: 21339347 DOI: 10.1148/radiol.10100547]

34 **Chen H**, He Y, Jia W. Precise hepatectomy in the intelligent digital era. *Int J Biol Sci* 2020; **16**: 365-373 [PMID: 32015674 DOI: 10.7150/ijbs.39387]

35 **Nakayama K**, Oshiro Y, Miyamoto R, Kohno K, Fukunaga K, Ohkohchi N. The Effect of Three-Dimensional Preoperative Simulation on Liver Surgery. *World J Surg* 2017; **41**: 1840-1847 [PMID: 28271263 DOI: 10.1007/s00268-017-3933-7]

36 **Zhang J**, Qiao QL, Guo XC, Zhao JX. Application of three-dimensional visualization technique in preoperative planning of progressive hilar cholangiocarcinoma. *Am J Transl Res* 2018; **10**: 1730-1735 [PMID: 30018714]

37 **Vorontsov E**, Cerny M, Régnier P, Di Jorio L, Pal CJ, Lapointe R, Vandenbroucke-Menu F, Turcotte S, Kadoury S, Tang A. Deep Learning for Automated Segmentation of Liver Lesions at CT in Patients with Colorectal Cancer Liver Metastases. *Radiol Artif Intell* 2019; **1**: 180014 [PMID: 33937787 DOI: 10.1148/ryai.2019180014]

38 **Boykov YY**, Jolly MP. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. International Conference on Computer Vision. 2001; 105–112

39 **Chartrand G**, Cresson T, Chav R, Gotra A, Tang A, De Guise JA. Liver Segmentation on CT and MR Using Laplacian Mesh Optimization. *IEEE Trans Biomed Eng* 2017; **64**: 2110-2121 [PMID: 27893375 DOI: 10.1109/TBME.2016.2631139]

40 **Cancian P**, Cortese N, Donadon M, Di Maio M, Soldani C, Marchesi F, Savevski V, Santambrogio MD, Cerina L, Laino ME, Torzilli G, Mantovani A, Terracciano L, Roncalli M, Di Tommaso L. Development of a Deep-Learning Pipeline to Recognize and Characterize Macrophages in Colo-Rectal Liver Metastasis. *Cancers (Basel)* 2021; **13** [PMID: 34282750 DOI: 10.3390/cancers13133313]

41 **Donadon M**, Torzilli G, Cortese N, Soldani C, Di Tommaso L, Franceschini B, Carriero R, Barbagallo M, Rigamonti A, Anselmo A, Colombo FS, Maggi G, Lleo A, Cibella J, Peano C, Kunderfranco P, Roncalli M, Mantovani A, Marchesi F. Macrophage morphology correlates with single-cell diversity and prognosis in colorectal liver metastasis. *J Exp Med* 2020; **217** [PMID: 32785653 DOI: 10.1084/jem.20191847]

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 April 28; 3(2): 36-45

DOI: 10.35712/aig.v3.i2.36

ISSN 2644-3236 (online)

MINIREVIEWS

Colorectal cancer: Artificial intelligence and its role in surgical decision making

Nalini Kanta Ghosh, Ashok Kumar

Specialty type: Surgery

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B, B
Grade C (Good): C, C
Grade D (Fair): D
Grade E (Poor): 0

P-Reviewer: Goli A, Iran; Hanada E, Japan; Jheng YC, Taiwan; Tarnawski AS, United States

Received: December 30, 2021
Peer-review started: December 30, 2021
First decision: January 26, 2022
Revised: February 2, 2022
Accepted: April 26, 2022
Article in press: April 26, 2022
Published online: April 28, 2022

Nalini Kanta Ghosh, Ashok Kumar, Department of Surgical Gastroenterology, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow 226014, UP, India

Corresponding author: Ashok Kumar, BSc, FASCRS, FRCS (Ire), FRCS (Ed), FRCS (Eng), MBBS, M. Ch, MS, Professor, Department of Surgical Gastroenterology, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Raebareli Road, Lucknow 226014, UP, India. doc.ashokgupta@gmail.com

Abstract

Despite several advances in the oncological management of colorectal cancer (CRC), there still remains a lacuna in the treatment strategy, which differs from center to center and on the philosophy of the treating clinician that is not without bias. Personalized treatment is essential for the treatment of CRC to achieve better long-term outcomes and to reduce morbidity. Surgery has an important role to play in the treatment. Surgical treatment of CRC is decided based on clinical parameters and investigations and hence likely to have judgmental errors. Artificial intelligence has been reported to be useful in the surveillance, diagnosis, treatment, and follow-up with accuracy in several malignancies. However, it is still evolving and yet to be established in surgical decision making in CRC. It is not only useful preoperatively but also intraoperatively. Artificial intelligence helps to rectify the human surgical decision when clinical data and radiological and laboratory parameters are fed into the computer and may guide correct surgical treatment.

Key Words: Artificial Intelligence; Colorectal cancer; Clinical implications; Treatment strategy; Surgical treatment

Core Tip: Treatment decision making in colorectal cancer significantly affects the outcome, which is a multidisciplinary team approach and is not without bias. Surgery plays a significant role in the treatment. Whether artificial intelligence may improve the outcome of surgery in colorectal cancer is not known. The present review focuses on its current role in surgical decision making and future impact.

Citation: Ghosh NK, Kumar A. Colorectal cancer: Artificial intelligence and its role in surgical decision making. *Artif Intell Gastroenterol* 2022; 3(2): 36-45
URL: https://www.wjgnet.com/2644-3236/full/v3/i2/36.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i2.36

INTRODUCTION

Mr. Alan Turing in 1950 hypothesized that a machine can also think like a human being in his book entitled "Computing Machinery and Intelligence"[1]. The term "artificial intelligence (AI)" was later coined by John McCarthy in a summer workshop[1,2]. AI has evolved from simple tasks to more complex tasks similar to a human brain[1].

AI has proven its worth in various day-to-day life and human requirements, including health care (health tracking devices)[3], automobiles (autopilot)[4], banking and finances (chatbots, robotraders)[5], surveillance (CCTV cameras), social media, entertainment, education, space exploration, industries (aluminum, dairy)[6-8], and disaster management[9,10]. One recent example is the efficient production of facemasks during the coronavirus disease 2019 pandemic[11] (Table 1). Its potential has been exploited in various fields of medicine, including online appointment scheduling, online check-in at hospitals, digitization of medical records, follow-up and immunization reminder, drug dosage algorithm, and adverse effect warnings during the prescription of multidrug combinations. Besides this, its application in the field of oncology is immense. AI is assisting in generating new approaches for cancer detection, screening of healthy subjects, diagnosis, classification of cancers using genomics, tumor microenvironment analysis, prognostication, follow-up, and new drug discovery[12-15].

Colorectal cancer (CRC) is one of the most common types of gastrointestinal (GI) tract malignancy and is the fourth most leading cause of cancer death globally[16,17]. AI has been used to facilitate screening, diagnosis (colonoscopy, advanced endoscopic modalities, imaging), genetic testing, and treatment (chemotherapy, radiotherapy, robotic assisted surgery)[18]. New research and developments are required for better patient management to improve the outcome.

In the past decade, several developments have taken place in the management of CRC, *e.g.*, revised anatomy of the rectum and concept of total mesorectal excision by Heald *et al*[19], concept of complete mesocolic excision and central vascular ligation by Hohenberger[20] for colon cancer, imaging and staging techniques, introduction of staplers[21], newer chemotherapeutic agents and biologicals, radiation therapy, and mode of surgery (laparoscopic and robotic surgery)[22,23] have significantly improved the outcome and sphincter preservation. However, there still remain numerous challenging issues like accurate preoperative diagnosis, staging, individualized and personalized treatment planning, and intraoperative challenges to minimize complications and improve the surgical outcome. Newer tools of AI have been used in various fields of medicine, including drug development, health monitoring, managing medical data, disease diagnostics, digital consultations, personalized treatment, analysis of health plans, and medical and surgical treatment[24] and is quickly finding a role in surgery and surgical decision making.

Two common fields of the AI used in medicine are: virtual and physical[25]. Virtual field is commonly used in medical imaging, clinical diagnosis, treatment, and drug research and development. Surgical and nursing robots are the part of physical fields. Because of ongoing innovations in AI, it is being used widely in medicine, both for diagnosis and management of tumors. AI has played a significant role in CRC at various stages and is reported to have improved the 5-year survival. The subsection of AI used in medicine is deep learning, which is responsible for widespread application of AI[26]. This method encompasses all the concepts of AI and is based on artificial neural networks (ANN), which is inspired by the neurons in a biological brain. Deep learning involves application of training a specific task on a larger data set, extracting information from them, and using them for future predictions about these tasks through flexible adaptation to the new data. Recently, deep learning has been used to predict cardiovascular risk based on retinal images[27], classification of skin lesions[28], mammogram-based breast cancer detection[29], and esophageal carcinoma[30]. However, application of AI in surgery is challenging, as unlike the use of static images, surgery includes dynamic procedural data like the patient clinical parameters, different devices used, and knowledge of clinical guidelines and from the experiences[31]. The uses and applications of various branches of AI in medicine as well as other fields are shown in Table 1.

In 2007, IBM began development of Deep QA technology (Watson). In 2017, Artery's medical imaging platform was the first Food and Drug Administration approved cloud-based deep learning application in healthcare for cardiac disorders, which was faster in giving results as compared to the professionals(15 s *vs* 30 s)[32]. The Food and Drug Administration-approved "GI genius" in the year 2019 is the first device based on machine learning to aid clinicians in detecting polyps or tumors during colonoscopy.

Table 1 Subfields of artificial intelligence and its application in day-to-day human life

S. No	Fields of AI	Description
1	Machine learning	Pattern identification and analysis where machine can help to improve based on past experiences provided from the given data set
2	Deep learning	Consists of multilayered neural networks called artificial neural network, which enables the computer to learn and make decisions on its own
3	Natural language processing	Ability of the computer to extract data from human language and make decisions
4	Computer vision	Potential to obtain information from a series of images or videos
5	Mixed-integer linear programming model[11]	It is helpful in finding the locational, supply, production, distribution, collection, quarantine, recycling, reuse, and disposal decisions within a multiperiod multiechelon multiproduct supply chain
6	Covering tour approach[9]	Optimizing the distribution and allocation of resources among individuals. It is useful at the time of crisis
7	Mixed-integer linear mathematical model[6]	This model optimizes economic, social, and environmental objectives simultaneously
8	Neural network with runner root algorithm[8]	Minimizing risk and maximizing return in industrial production
9	Meta-heuristic algorithms[7]	A comprehensive framework to predict the demand for dairy products
10	Hybrid shapley value and multimoora method[10]	An intelligent performance evaluation system for different supply chains in industries

AI: Artificial intelligence; S. No: Serial number.

This paper reviews the current status of AI in CRC surgical decision making and its future implications.

USES OF AI IN GASTROINTESTINAL DISORDERS AND COLORECTAL CANCER

AI is progressively being used in the understanding of GI diseases[33-35]. Imaging such as X-ray, computed tomography scanning, magnetic resonance imaging, or endoscopic imaging is being used for diagnosis[36-39]. The application of AI has led to early detection of intestinal malignancies or premalignant lesions, and inflammatory or other non-malignant diseases or lesions[40].

With IBM Watson for oncology (WFO), AI has found its increasing role in oncology therapy. It has been used in several malignancies like breast carcinoma, lung carcinoma, gastric cancer, colon and rectal cancer, *etc*. Initially, Memorial Sloan Kettering Cancer Center (New York, United States) started the use of WFO machine learning. WFO uses natural language processing and clinical data from multiple resources (treatment guidelines, expert opinions, literature, and medical records) to formulate treatment recommendations[41]. A recent meta-analysis[42] had shown the highest concordance between WFO and Mass Detection Tool in breast carcinoma and the lowest in stomach carcinoma. The Manipal Comprehensive Cancer Centre (Bangalore, India) has implemented WFO for treatment in 250 CRC patients[43]. There was a concordance in 92.7% of rectal and 81.0% of colon cancer patients between WFO and Mass Detection Tool recommendations[43].

AI IN COLORECTAL CANCER

AI is used in the diagnosis and treatment of colorectal polyps and cancer. In colorectal cancer, it helps in diagnosis, staging (lymph node or liver metastasis), preoperative treatment planning, response to treatment assessment, intraoperative assistance, postoperative prognostic information, *etc*[44-46].

AI in preoperative surgical decision making: staging and planning

After diagnosis of CRC is made, the most important consideration is staging to determine a further plan of management, whether upfront surgery, neoadjuvant treatment, or palliative treatment.

In locally advanced rectal cancer, preoperative chemoradiotherapy is known to reduce the local recurrence. However, selection of patients is essential to avoid unnecessary complications due to overtreatment. Therefore, there is a need for a system that can differentiate between T2 and T3 rectal cancers. Kim *et al*[47] used convolutional neural network models to distinguish T2 from T3 lesions from magnetic resonance imaging with an accuracy of 94%. Similarly, Wu *et al*[48] also used convolutional

neural network to stage rectal cancers.

In addition to its role in preoperative imaging, AI provides faster interpretation compared to radiologists (20 s *vs* 600 s) in the detection of lymph node metastasis in rectal cancer[49]. Preoperatively, positron emission tomography/computed tomography is commonly used in the case of indeterminate lesions on contrast-enhanced computed tomography to potentially find curable M1 disease (National Comprehensive Cancer Network guidelines version 3.2021). Recently, application of AI has improved the sensitivity and specificity of detection of pulmonary nodules[50]. AI can also be used to reconstruct the area of interest from two-dimensional data obtained from imaging and endoscopic findings to generate a three-dimensional structure for better delineation of the tumor in relation to the surrounding vital structures, which may be useful in preoperative surgical planning[51]. This is extremely useful in determining which patient will require a pelvic exenteration or which patient will require a lateral pelvic lymph node dissection. This is also useful to safeguard the important surrounding structures during surgery to reduce the postoperative morbidity and mortality related to it.

In colon cancer, clinical evidence of bulky nodal disease or T4b lesion entails neoadjuvant therapy (National Comprehensive Cancer Network guidelines version 3.2021). It is also recommended that the presence of nodal involvement in T1 cancer requires colectomy and lymphadenectomy. Kudo *et al*[52] applied machine learning ANN in 3134 patients with T1 CRC based on the patient's data on age, gender, tumor size, location, morphology, lymphatic and vascular invasion, and histologic grade to predict nodal involvement. ANN model was significantly better in lymph node metastasis detection compared to guidelines (area under the curve: 0.83 *vs* area under the curve: 0.73, P value = 0.005). Therefore, these patients can be subjected to upfront surgery and lymphadenectomy instead of endoscopic treatment. A meta-analysis by Bedrikovetski *et al*[53] using 17 studies (12 used radiomics models and 5 used deep learning models) concluded that AI was more efficient than radiologists in predicting lymph node metastasis. Similarly, AI was found to be better in detecting metastatic nodes as compared to conventional positron emission tomography/computed tomography imaging[54].

AI in intraoperative surgical decision making

Execution of a surgery depends upon the operating skill and ability of decision making. In 1978, Dr. Spencer[55], a cardiovascular surgeon, mentioned that "a skilfully performed operation is about 75% decision making and 25% dexterity." The decision making can be both technical or non-technical, which impacts patient outcome. Studies of surgical errors have shown that over half of the adverse events are due to cognitive errors[56]. But surgical training is more focused on skill training rather than decision making as it is a challenging task to train[57]. Decision-making skills may vary with experience of operating surgeons[58]. Thus, improving the quality of surgical decision making could help to improve the outcome of surgery.

Decision making is a three-step process, *i.e.* assessment of the situation, action-taking, and re-evaluation of the action's consequences. AI has been used as a decision making aid in a variety of fields, both in medicine and in surgery[59,60]. AI can help surgeons to assess a given situation (*e.g.*, retrieving better data about a clinical situation), the types of actions taken (*e.g.*, through decision suggestion), and the process of re-evaluating the impact of the decision taken. Therefore, it can be achieved in three different ways: (1) Retrieving data and experience from similar clinical scenarios and to supplement sensory input during minimal access surgery, which are not available compared to open surgery; (2) Intraoperative pathology assessment, tumor margin mapping, tumor classification, and tissue identification; and (3) Suggestion of steps of surgery.

Identification of surrounding structures: Harangi *et al*[61] used an ANN model to distinguish ureter from uterine artery during laparoscopic hysterectomy with 94.2% accuracy. Similarly, Quellec *et al*[62] applied a system of retrieving related videos of retinal surgery, and subsequent steps were followed during surgery to minimize the risk of injury. AI made it possible to define dissection planes in the robotic gastrectomy and to identify the recurrent laryngeal nerve during thyroidectomy[63,64]. Various studies have shown improved detection of vital structures during laparoscopic cholecystectomy to prevent bile duct injury using AI (Madani *et al*[65], Mascagni *et al*[66], Tokuyasu *et al*[67]). Table 2 highlights the studies where AI was used for identification of vital structures.

In CRC surgery, AI can be used to detect nearby vital structures (nerve plexus, presacral venous plexus, ureter, bladder, urethra, prostate, seminal vesicles), lymph node metastasis (lateral pelvic nodes, nodes near the root of inferior mesenteric artery), determination of the margin of resection, vascularity, and adequacy of anastomosis.

Augmented reality augments surgeons' intraoperative vision by providing a semi-transparent overlay of preoperative imaging on the area of interest[68]. It has been used in several GI surgical procedures like laparoscopic splenectomy[69] and pancreaticoduodenectomy[70]. Augmented reality can be applied to CRC surgeries to identify and preserve the nearby vital structures.

Deciding the level of resection: In CRC surgery, determination of margin status is important to decide the level of resection and consideration for the feasibility of an anastomosis or the creation of a stoma. Margin status can be obtained with "optical biopsy" (*in vivo* diagnostic imaging), which can avoid time-consuming resection and frozen section analysis. Fluorescence-guided surgery is evolving, and it has

Table 2 Studies having found the role of artificial intelligence in identification of vital structures in surgery

S. No	Primary aim	AI method used	Ref.
1	Recognition of ureter and uterine artery	Convolutional neural network	Harangi *et al*[61], 2017
2	Recognition of surgical steps of retinal surgery	Content-based video retrieval system	Quellec *et al*[62], 2011
3	To define safe dissection plane in robot assisted gastrectomy	Deep learning model based on U-net	Kumazu *et al*[63], 2021
4	Recurrent laryngeal nerve detection during thyroidectomy	Deep learning computer vision algorithm	Gong *et al*[64], 2021

AI: Artificial intelligence; S. No: Serial number.

shown promising results in determination of liver or peritoneal metastasis, anastomotic perfusion, detection of sentinel nodes, ureter, and nerves, and intraoperative detection of primary and recurrent lesions during colorectal cancer surgery[71]. Such a concept can be extrapolated on to AI for more efficient performance. Modalities used for intraoperative optical biopsy are confocal laser endomicroscopy, hyperspectral imaging, optical coherence tomography, and contrast-enhanced ultrasonography. There are several studies where these modalities have been used to distinguish abnormal epithelium from normal with the help of AI (Table 3). Using hyperspectral imaging, Jansen-Winkeln *et al*[72] reported 94% accuracy in distinguishing carcinoma from adenoma and healthy mucosa using ANN on post-resection of colonic lesions during surgery. A couple of experimental studies have shown that laparoscopic hyperspectral imaging can be used to distinguish malignant tissue in CRC from normal tissue. These modalities can be used to help in surgical decision making in CRC as revisional surgery can be done intraoperatively rather than waiting for frozen sections or final histology avoiding another surgery[73,74]. AI has been effective in differentiating glioblastoma, parathyroid gland, and malignant lesions of the colon from adjacent normal tissues[75-77].

Deciding the site of anastomosis: Studies have shown the incidence of colocolic and colorectal anastomosis leak to be 3.3% and 8.6%, respectively[78] and has adverse clinical outcomes and economic burden[79]. It can lead to anastomotic site stricture, recurrence of malignancy, and poor evacuatory function. The literature has shown poor predictive value of surgeons' perceptions of possible anastomotic site leaks that led to investigating other methods like the use of indocyanine green[80]. The robotic platform provides an inbuilt near infrared camera for assessment of vascularity at the resection margin and to reduce anastomotic site leakage[81]. A study by Mazaki *et al*[82], where auto-artificial intelligence was used to develop a predictive model for anastomotic leakage, showed that triple-row staplers can decrease the leak rate. There is an ongoing study by Taha *et al*[83] known as the PANIC study (The Prediction of Anastomotic Insufficiency risk after Colorectal surgery), which utilizes machine learning principles to formulate an algorithm for prediction of anastomotic leak following colonic (PANIC-C) or colorectal (PANIC-R) anastomosis. The results of the study are expected to be available by December 2022.

Helping in operative step suggestion: Operative step suggestion in CRC is at a developmental stage. In the literature, AI has been used in cataract surgery and spinal cord surgery with satisfactory results. Tian *et al*[84] developed VeBIRD (Video-Based Intelligent Recognition and Decision system) to track and classify the cataract grade on videos of phacoemulsification surgeries. It helped to decide the amount of ultrasonic energy needed to emulsify a cataract based on the grade. Therefore, a less experienced surgeon can perform the procedure with as much efficiency as that of an experienced surgeon. Somatosensory evoked potential is used during spinal cord surgeries to detect spinal cord injury. A decrease in somatosensory evoked potential value needs to be confirmed with awakening the patient and checking spinal cord function and this decrease in somatosensory evoked potential can be due to the effect of anesthesia. Fan *et al*[85] applied support vector regression and multi-support vector regression to distinguish spinal cord injury from anesthetic effect. Similarly, in CRC surgery such methods can help to find the area of interest to formulate standardized resection and differentiate intraoperative lymphorrhea from ureter or bladder injury using AI.

Colorectal cancer surgery requires accurate and judicious preoperative decisions to optimize the outcome of surgery (personalized treatment). The decision can be augmented by the use of AI, which is expected to be precise and without errors. It can assist in imaging, tissue diagnosis, and staging before surgery. It can be used preoperatively to choose patients for neoadjuvant therapy and those requiring upfront surgeries. Intraoperatively, it helps in the identification of tumor tissue (to determine the margin of resection), metastatic lymph nodes (for the extent of lymphadenectomy), and important surrounding structures. Its assistance is also useful in assessing the adequate vascularity at the anastomotic site that can decrease the postoperative anastomotic leak and thereby reduce the morbidity and mortality.

Table 3 Studies of artificial intelligence differentiating normal epithelium from abnormal or malignant cells

S. No	Modality used	Primary aim of study	AI method used	Ref.
1	CEUS	To differentiate glioblastoma from normal tissue	Support vector machines	Ritschel *et al*[75], 2015
2	OCT	To distinguish parathyroid tissue from thyroid, lymph node, and adipose tissue	Texture feature analysis and back propagation artificial neural network	Hou *et al*[76], 2017
3	CLE	Normal colonic mucosa from malignant lesion	Fractal analysis and neural network modelling	Ștefănescu *et al*[77], 2016
4	Hyperspectral imaging	Differentiation of colonic carcinoma from adenoma and healthy mucosa	Artificial neural network	Jansen-Winkeln *et al* [72], 2021

AI: Artificial intelligence; CEUS: Contrast-enhanced ultrasonography; OCT: Optical coherence tomography; CLE: Confocal laser endomicroscopy; S. No: Serial number.

Like the application of AI in several domains of medicine and health, it may play a significant role in surgical decision making, enhancing the outcome, in addition to diagnosis (imaging, endoscopy, tissue diagnosis).

FUTURE IMPLICATIONS

The future is promising, where AI is likely to play a significant role in reducing the bias of the Mass Detection Tool in deciding the treatment strategy and reducing the diagnosis and planning time with uniformity and with no or minimum error. The day is not far when the surgical world may be able to find a personalized surgical treatment for each and every patient of CRC, with improved intraoperative technical execution and reduced complications. The overall time taken in the management of CRC will be reduced, the treatment will be standardized, and the outcome will be maximized.

CONCLUSION

The role of AI in CRC is currently limited to preoperative staging and assessment of surgical resection margins and anastomotic sites. Its application to surgical decision making is still evolving, and the literature is very limited. However, the future is promising.

FOOTNOTES

Author contributions: Kumar A designed the concept, corrected, and finalized the manuscript; Ghosh NK wrote the manuscript and reviewed the literature; All authors have read and approved the final manuscript.

Conflict-of-interest statement: The authors declare no conflicts of interest for this article.

Country/Territory of origin: India

ORCID number: Nalini Kanta Ghosh 0000-0003-1213-0235; Ashok Kumar 0000-0003-3959-075X.

S-Editor: Liu JH
L-Editor: Filipodia
P-Editor: Liu JH

REFERENCES

1 **Kaul V**, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointest Endosc* 2020; **92**: 807-812 [PMID: 32565184 DOI: 10.1016/j.gie.2020.06.040]
2 **Hamamoto R**, Suvarna K, Yamada M, Kobayashi K, Shinkai N, Miyake M, Takahashi M, Jinnai S, Shimoyama R, Sakai A, Takasawa K, Bolatkan A, Shozu K, Dozen A, Machino H, Takahashi S, Asada K, Komatsu M, Sese J, Kaneko S. Application of Artificial Intelligence Technology in Oncology: Towards the Establishment of Precision Medicine. *Cancers (Basel)* 2020; **12** [PMID: 33256107 DOI: 10.3390/cancers12123532]
3 **Meng Y**, Speier W, Shufelt C, Joung S, E Van Eyk J, Bairey Merz CN, Lopez M, Spiegel B, Arnold CW. A Machine Learning Approach to Classifying Self-Reported Health Status in a Cohort of Patients With Heart Disease Using Activity Tracker Data. *IEEE J Biomed Health Inform* 2020; **24**: 878-884 [PMID: 31199276 DOI: 10.1109/JBHI.2019.2922178]
4 **Kirsch D**. Autopilot and algorithms: accidents, errors, and the current need for human oversight. *J Clin Sleep Med* 2020; **16**: 1651-1652 [PMID: 32844741 DOI: 10.5664/jcsm.8762]
5 **Bredt S**. Artificial Intelligence (AI) in the Financial Sector-Potential and Public Strategies. *Front Artif Intell* 2019; **2**: 16 [PMID: 33733105 DOI: 10.3389/frai.2019.00016]
6 **Pahlevan SM**, Hosseini SMS, Goli A. Sustainable supply chain network design using products' life cycle in the aluminum industry. *Environ Sci Pollut Res Int* 2021 [PMID: 33474670 DOI: 10.1007/s11356-020-12150-8]
7 **Goli A**, Khademi-Zare H, Tavakkoli-Moghaddam R, Sadeghieh A, Sasanian M, Malekalipour Kordestanizadeh R. An integrated approach based on artificial intelligence and novel meta-heuristic algorithms to predict demand for dairy products: a case study. *Network* 2021; **32**: 1-35 [PMID: 33390063 DOI: 10.1080/0954898X.2020.1849841]
8 **Goli A**, Khademi Zare H, Tavakkoli-Moghaddam R, Sadeghieh A. (2019). Hybrid artificial intelligence and robust optimization for a multi-objective product portfolio problem. *Computers & Industrial Engineering* 2019; **137**: 106090 [DOI: 10.1016/j.cie.2019.106090]
9 **Goli A**, Malmir B. A Covering Tour Approach for Disaster Relief Locating and Routing with Fuzzy Demand. *International Journal of ITS Research* 2019; **18**: 140-152 [DOI: 10.1007/s13177-019-00185-2]
10 **Goli A**, Mohammadi H. Developing a sustainable operational management system using hybrid Shapley value and Multimoora method: case study petrochemical supply chain. *Environ Dev Sustain* 2021 [DOI: 10.1007/s10668-021-01844-9]
11 **Tirkolaee EB**, Goli A, Ghasemi P, Goodarzian F. Designing a sustainable closed-loop supply chain network of face masks during the COVID-19 pandemic: Pareto-based algorithms. *J Clean Prod* 2022; **333**: 130056 [PMID: 34924699 DOI: 10.1016/j.jclepro.2021.130056]
12 **Bhinder B**, Gilvary C, Madhukar NS, Elemento O. Artificial Intelligence in Cancer Research and Precision Medicine. *Cancer Discov* 2021; **11**: 900-915 [PMID: 33811123 DOI: 10.1158/2159-8290.CD-21-0090]
13 **Kann BH**, Hosny A, Aerts HJWL. Artificial intelligence for clinical oncology. *Cancer Cell* 2021; **39**: 916-927 [PMID: 33930310 DOI: 10.1016/j.ccell.2021.04.002]
14 **Huynh E**, Hosny A, Guthier C, Bitterman DS, Petit SF, Haas-Kogan DA, Kann B, Aerts HJWL, Mak RH. Artificial intelligence in radiation oncology. *Nat Rev Clin Oncol* 2020; **17**: 771-781 [PMID: 32843739 DOI: 10.1038/s41571-020-0417-8]
15 **Benzekry S**. Artificial Intelligence and Mechanistic Modeling for Clinical Decision Making in Oncology. *Clin Pharmacol Ther* 2020; **108**: 471-486 [PMID: 32557598 DOI: 10.1002/cpt.1951]
16 **Mármol I**, Sánchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi MJ. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *Int J Mol Sci* 2017; **18** [PMID: 28106826 DOI: 10.3390/ijms18010197]
17 **Ou C**, Sun Z, Li X, Ren W, Qin Z, Zhang X, Yuan W, Wang J, Yu W, Zhang S, Peng Q, Yan Q, Xiong W, Li G, Ma J. MiR-590-5p, a density-sensitive microRNA, inhibits tumorigenesis by targeting YAP1 in colorectal cancer. *Cancer Lett* 2017; **399**: 53-63 [PMID: 28433598 DOI: 10.1016/j.canlet.2017.04.011]
18 **Mitsala A**, Tsalikidis C, Pitiakoudis M, Simopoulos C, Tsaroucha AK. Artificial Intelligence in Colorectal Cancer Screening, Diagnosis and Treatment. A New Era. *Curr Oncol* 2021; **28**: 1581-1607 [PMID: 33922402 DOI: 10.3390/curroncol28030149]
19 **Heald RJ**. The 'Holy Plane' of rectal surgery. *J R Soc Med* 1988; **81**: 503-508 [PMID: 3184105]
20 **Hohenberger W**, Weber K, Matzel K, Papadopoulos T, Merkel S. Standardized surgery for colonic cancer: complete mesocolic excision and central ligation--technical notes and outcome. *Colorectal Dis* 2009; **11**: 354-64; discussion 364 [PMID: 19016817 DOI: 10.1111/j.1463-1318.2008.01735.x]
21 **Moran BJ**. Stapling instruments for intestinal anastomosis in colorectal surgery. *Br J Surg* 1996; **83**: 902-909 [PMID: 8813772 DOI: 10.1002/bjs.1800830707]
22 **Jacobs M**, Verdeja JC, Goldstein HS. Minimally invasive colon resection (laparoscopic colectomy). *Surg Laparosc Endosc* 1991; **1**: 144-150 [PMID: 1688289]
23 **Weber PA**, Merola S, Wasielewski A, Ballantyne GH. Telerobotic-assisted laparoscopic right and sigmoid colectomies for benign disease. *Dis Colon Rectum* 2002; **45**: 1689-94; discussion 1695 [PMID: 12473897 DOI: 10.1007/s10350-004-7261-2]
24 **de Grey AD**. Artificial Intelligence and Medical Research: Time to Aim Higher? *Rejuvenation Res* 2016; **19**: 105-106 [PMID: 26993572 DOI: 10.1089/rej.2016.1827]
25 **Hamet P**, Tremblay J. Artificial intelligence in medicine. *Metabolism* 2017; **69S**: S36-S40 [PMID: 28126242 DOI: 10.1016/j.metabol.2017.01.011]
26 **LeCun Y**, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436-444 [PMID: 26017442 DOI: 10.1038/nature14539]
27 **Poplin R**, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, Peng L, Webster DR. Prediction of cardiovascular risk factors from retinal fundus photographs *via* deep learning. *Nat Biomed Eng* 2018; **2**: 158-164 [PMID: 31015713 DOI: 10.1038/s41551-018-0195-0]
28 **Esteva A**, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer

with deep neural networks. *Nature* 2017; **542**: 115-118 [PMID: 28117445 DOI: 10.1038/nature21056]

29 **McKinney SM**, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, Back T, Chesus M, Corrado GS, Darzi A, Etemadi M, Garcia-Vicente F, Gilbert FJ, Halling-Brown M, Hassabis D, Jansen S, Karthikesalingam A, Kelly CJ, King D, Ledsam JR, Melnick D, Mostofi H, Peng L, Reicher JJ, Romera-Paredes B, Sidebottom R, Suleyman M, Tse D, Young KC, De Fauw J, Shetty S. International evaluation of an AI system for breast cancer screening. *Nature* 2020; **577**: 89-94 [PMID: 31894144 DOI: 10.1038/s41586-019-1799-6]

30 **Horie Y**, Yoshio T, Aoyama K, Yoshimizu S, Horiuchi Y, Ishiyama A, Hirasawa T, Tsuchida T, Ozawa T, Ishihara S, Kumagai Y, Fujishiro M, Maetani I, Fujisaki J, Tada T. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest Endosc* 2019; **89**: 25-32 [PMID: 30120958 DOI: 10.1016/j.gie.2018.07.037]

31 **Maier-Hein L**, Vedula SS, Speidel S, Navab N, Kikinis R, Park A, Eisenmann M, Feussner H, Forestier G, Giannarou S, Hashizume M, Katic D, Kenngott H, Kranzfelder M, Malpani A, März K, Neumuth T, Padoy N, Pugh C, Schoch N, Stoyanov D, Taylor R, Wagner M, Hager GD, Jannin P. Surgical data science for next-generation interventions. *Nat Biomed Eng* 2017; **1**: 691-696 [PMID: 31015666 DOI: 10.1038/s41551-017-0132-7]

32 **Marr B**. First FDA approval for clinical Cloud-Based Deep Learning in Healthcare. 2017. (accessed 1 Jun 2017). Available from: https://www.forbes.com/sites/bernardmarr/2017/01/20/first-fda-approval-for-clinical-cloud-based-deep-learning-in-healthcare/#7a0ed8dc161c

33 **Reismann J**, Romualdi A, Kiss N, Minderjahn MI, Kallarackal J, Schad M, Reismann M. Diagnosis and classification of pediatric acute appendicitis by artificial intelligence methods: An investigator-independent approach. *PLoS One* 2019; **14**: e0222030 [PMID: 31553729 DOI: 10.1371/journal.pone.0222030]

34 **Reichling C**, Taieb J, Derangere V, Klopfenstein Q, Le Malicot K, Gornet JM, Becheur H, Fein F, Cojocarasu O, Kaminsky MC, Lagasse JP, Luet D, Nguyen S, Etienne PL, Gasmi M, Vanoli A, Perrier H, Puig PL, Emile JF, Lepage C, Ghiringhelli F. Artificial intelligence-guided tissue analysis combined with immune infiltrate assessment predicts stage III colon cancer outcomes in PETACC08 study. *Gut* 2020; **69**: 681-690 [PMID: 31780575 DOI: 10.1136/gutjnl-2019-319292]

35 **Maeda Y**, Kudo SE, Mori Y, Misawa M, Ogata N, Sasanuma S, Wakamura K, Oda M, Mori K, Ohtsuka K. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest Endosc* 2019; **89**: 408-415 [PMID: 30268542 DOI: 10.1016/j.gie.2018.09.024]

36 **Ho TY**, Lin CW, Chang CC, Chen HT, Chen YJ, Lo YS, Hsiao PH, Chen PC, Lin CS, Tsou HK. Percutaneous endoscopic unilateral laminotomy and bilateral decompression under 3D real-time image-guided navigation for spinal stenosis in degenerative lumbar kyphoscoliosis patients: an innovative preliminary study. *BMC Musculoskelet Disord* 2020; **21**: 734 [PMID: 33172435 DOI: 10.1186/s12891-020-03745-w]

37 **Bhattacharya S**, Reddy Maddikunta PK, Pham QV, Gadekallu TR, Krishnan S SR, Chowdhary CL, Alazab M, Jalil Piran M. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. *Sustain Cities Soc* 2021; **65**: 102589 [PMID: 33169099 DOI: 10.1016/j.scs.2020.102589]

38 **Karako K**, Song P, Chen Y, Tang W. Realizing 5G- and AI-based doctor-to-doctor remote diagnosis: opportunities, challenges, and prospects. *Biosci Trends* 2020; **14**: 314-317 [PMID: 33100291 DOI: 10.5582/bst.2020.03364]

39 **Shiyam Sundar LK**, Muzik O, Buvat I, Bidaut L, Beyer T. Potentials and caveats of AI in hybrid imaging. *Methods* 2021; **188**: 4-19 [PMID: 33068741 DOI: 10.1016/j.ymeth.2020.10.004]

40 **Le Berre C**, Sandborn WJ, Aridhi S, Devignes MD, Fournier L, Smaïl-Tabbone M, Danese S, Peyrin-Biroulet L. Application of Artificial Intelligence to Gastroenterology and Hepatology. *Gastroenterology* 2020; **158**: 76-94.e2 [PMID: 31593701 DOI: 10.1053/j.gastro.2019.08.058]

41 **Somashekhar SP**, Sepúlveda MJ, Norden AD, Rauthan A, Arun K, Patil P, Ethadka RY, Kumar RC. Early experience with IBM Watson for Oncology (WFO) cognitive computing system for lung and colorectal cancer treatment. [DOI: 10.1200/JCO.2017.35.15_suppl.8527]

42 **Jie Z**, Zhiying Z, Li L. A meta-analysis of Watson for Oncology in clinical application. *Sci Rep* 2021; **11**: 5792 [PMID: 33707577 DOI: 10.1038/s41598-021-84973-5]

43 **Yang SY**, Roh KH, Kim YN, Cho M, Lim SH, Son T, Hyung WJ, Kim HI. Surgical Outcomes After Open, Laparoscopic, and Robotic Gastrectomy for Gastric Cancer. *Ann Surg Oncol* 2017; **24**: 1770-1777 [PMID: 28357674 DOI: 10.1245/s10434-017-5851-1]

44 **Zhang R**, Zheng Y, Poon CCY, Shen D, Lau JYW. Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. *Pattern Recognit* 2018; **83**: 209-219 [PMID: 31105338 DOI: 10.1016/j.patcog.2018.05.026]

45 **Su JR**, Li Z, Shao XJ, Ji CR, Ji R, Zhou RC, Li GC, Liu GQ, He YS, Zuo XL, Li YQ. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). *Gastrointest Endosc* 2020; **91**: 415-424.e4 [PMID: 31454493 DOI: 10.1016/j.gie.2019.08.026]

46 **Zhang S**, Han F, Liang Z, Tan J, Cao W, Gao Y, Pomeroy M, Ng K, Hou W. An investigation of CNN models for differentiating malignant from benign lesions using small pathologically proven datasets. *Comput Med Imaging Graph* 2019; **77**: 101645 [PMID: 31454710 DOI: 10.1016/j.compmedimag.2019.101645]

47 **Kim J**, Oh JE, Lee J, Kim MJ, Hur BY, Sohn DK, Lee B. Rectal cancer: Toward fully automatic discrimination of T2 and T3 rectal cancers using deep convolutional neural network. *Int J Imag Syst Tech* 2019; **29**: 247-259 [DOI: 10.1002/ima.22311]

48 **Wu QY**, Liu SL, Sun P, Li Y, Liu GW, Liu SS, Hu JL, Niu TY, Lu Y. Establishment and clinical application value of an automatic diagnosis platform for rectal cancer T-staging based on a deep neural network. *Chin Med J (Engl)* 2021; **134**: 821-828 [PMID: 33797468 DOI: 10.1097/CM9.0000000000001401]

49 **Ding L**, Liu GW, Zhao BC, Zhou YP, Li S, Zhang ZD, Guo YT, Li AQ, Lu Y, Yao HW, Yuan WT, Wang GY, Zhang DL, Wang L. Artificial intelligence system of faster region-based convolutional neural network surpassing senior radiologists in evaluation of metastatic lymph nodes of rectal cancer. *Chin Med J (Engl)* 2019; **132**: 379-387 [PMID: 30707177 DOI: 10.1097/CM9.0000000000000095]

50 **Schwyzer M**, Martini K, Benz DC, Burger IA, Ferraro DA, Kudura K, Treyer V, von Schulthess GK, Kaufmann PA, Huellner MW, Messerli M. Artificial intelligence for detecting small FDG-positive lung nodules in digital PET/CT: impact of image reconstructions on diagnostic performance. *Eur Radiol* 2020; **30**: 2031-2040 [PMID: 31822970 DOI: 10.1007/s00330-019-06498-w]

51 **Kim HJ**, Choi GS, Park JS, Park SY, Cho SH, Seo AN, Yoon GS. S122: impact of fluorescence and 3D images to completeness of lateral pelvic node dissection. *Surg Endosc* 2020; **34**: 469-476 [PMID: 31139999 DOI: 10.1007/s00464-019-06830-x]

52 **Kudo SE**, Ichimasa K, Villard B, Mori Y, Misawa M, Saito S, Hotta K, Saito Y, Matsuda T, Yamada K, Mitani T, Ohtsuka K, Chino A, Ide D, Imai K, Kishida Y, Nakamura K, Saiki Y, Tanaka M, Hoteya S, Yamashita S, Kinugasa Y, Fukuda M, Kudo T, Miyachi H, Ishida F, Itoh H, Oda M, Mori K. Artificial Intelligence System to Determine Risk of T1 Colorectal Cancer Metastasis to Lymph Node. *Gastroenterology* 2021; **160**: 1075-1084.e2 [PMID: 32979355 DOI: 10.1053/j.gastro.2020.09.027]

53 **Bedrikovetski S**, Dudi-Venkata NN, Kroon HM, Seow W, Vather R, Carneiro G, Moore JW, Sammour T. Artificial intelligence for pre-operative lymph node staging in colorectal cancer: a systematic review and meta-analysis. *BMC Cancer* 2021; **21**: 1058 [PMID: 34565338 DOI: 10.1186/s12885-021-08773-w]

54 **He J**, Wang Q, Zhang Y, Wu H, Zhou Y, Zhao S. Preoperative prediction of regional lymph node metastasis of colorectal cancer based on ^{18}F-FDG PET/CT and machine learning. *Ann Nucl Med* 2021; **35**: 617-627 [PMID: 33738763 DOI: 10.1007/s12149-021-01605-8]

55 **Spencer F**. Teaching and measuring surgical techniques: the technical evaluation of competence. *Bull Am Coll Surg* 1978; **63**: 9-12

56 **Suliburk JW**, Buck QM, Pirko CJ, Massarweh NN, Barshes NR, Singh H, Rosengart TK. Analysis of Human Performance Deficiencies Associated With Surgical Adverse Events. *JAMA Netw Open* 2019; **2**: e198067 [PMID: 31365107 DOI: 10.1001/jamanetworkopen.2019.8067]

57 **Pugh CM**, Santacaterina S, DaRosa DA, Clark RE. Intra-operative decision making: more than meets the eye. *J Biomed Inform* 2011; **44**: 486-496 [PMID: 20096376 DOI: 10.1016/j.jbi.2010.01.001]

58 **Hashimoto DA**, Axelsson CG, Jones CB, Phitayakorn R, Petrusa E, McKinley SK, Gee D, Pugh C. Surgical procedural map scoring for decision-making in laparoscopic cholecystectomy. *Am J Surg* 2019; **217**: 356-361 [PMID: 30470551 DOI: 10.1016/j.amjsurg.2018.11.011]

59 **Hashimoto DA**, Rosman G, Witkowski ER, Stafford C, Navarette-Welton AJ, Rattner DW, Lillemoe KD, Rus DL, Meireles OR. Computer Vision Analysis of Intraoperative Video: Automated Recognition of Operative Steps in Laparoscopic Sleeve Gastrectomy. *Ann Surg* 2019; **270**: 414-421 [PMID: 31274652 DOI: 10.1097/SLA.0000000000003460]

60 **Hashimoto DA**, Rosman G, Rus D, Meireles OR. Artificial Intelligence in Surgery: Promises and Perils. *Ann Surg* 2018; **268**: 70-76 [PMID: 29389679 DOI: 10.1097/SLA.0000000000002693]

61 **Harangi B**, Hajdu A, Lampe R, Torok P. Recognizing ureter and uterine artery in endoscopic images using a convolutional neural network. In 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS) 2017 Jun 22 (pp. 726-727). IEEE

62 **Quellec G**, Lamard M, Cazuguel G, Droueche Z, Roux C, Cochener B. Real-time retrieval of similar videos with application to computer-aided retinal surgery. In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2011 Aug 30 (pp. 4465-4468). IEEE

63 **Kumazu Y**, Kobayashi N, Kitamura N, Rayan E, Neculoiu P, Misumi T, Hojo Y, Nakamura T, Kumamoto T, Kurahashi Y, Ishida Y, Masuda M, Shinohara H. Automated segmentation by deep learning of loose connective tissue fibers to define safe dissection planes in robot-assisted gastrectomy. *Sci Rep* 2021; **11**: 21198 [PMID: 34707141 DOI: 10.1038/s41598-021-00557-3]

64 **Gong J**, Holsinger FC, Noel JE, Mitani S, Jopling J, Bedi N, Koh YW, Orloff LA, Cernea CR, Yeung S. Using deep learning to identify the recurrent laryngeal nerve during thyroidectomy. *Sci Rep* 2021; **11**: 14306 [PMID: 34253767 DOI: 10.1038/s41598-021-93202-y]

65 **Madani A**, Namazi B, Altieri MS, Hashimoto DA, Rivera AM, Pucher PH, Navarrete-Welton A, Sankaranarayanan G, Brunt LM, Okrainec A, Alseidi A. Artificial Intelligence for Intraoperative Guidance: Using Semantic Segmentation to Identify Surgical Anatomy During Laparoscopic Cholecystectomy. *Ann Surg* 2020 [PMID: 33196488 DOI: 10.1097/SLA.0000000000004594]

66 **Mascagni P**, Vardazaryan A, Alapatt D, Urade T, Emre T, Fiorillo C, Pessaux P, Mutter D, Marescaux J, Costamagna G, Dallemagne B, Padoy N. Artificial Intelligence for Surgical Safety: Automatic Assessment of the Critical View of Safety in Laparoscopic Cholecystectomy Using Deep Learning. *Ann Surg* 2020 [PMID: 33201104 DOI: 10.1097/SLA.0000000000004351]

67 **Tokuyasu T**, Iwashita Y, Matsunobu Y, Kamiyama T, Ishikake M, Sakaguchi S, Ebe K, Tada K, Endo Y, Etoh T, Nakashima M, Inomata M. Development of an artificial intelligence system using deep learning to indicate anatomical landmarks during laparoscopic cholecystectomy. *Surg Endosc* 2021; **35**: 1651-1658 [PMID: 32306111 DOI: 10.1007/s00464-020-07548-x]

68 **Bernhardt S**, Nicolau SA, Soler L, Doignon C. The status of augmented reality in laparoscopic surgery as of 2016. *Med Image Anal* 2017; **37**: 66-90 [PMID: 28160692 DOI: 10.1016/j.media.2017.01.007]

69 **Ieiri S**, Uemura M, Konishi K, Souzaki R, Nagao Y, Tsutsumi N, Akahoshi T, Ohuchida K, Ohdaira T, Tomikawa M, Tanoue K, Hashizume M, Taguchi T. Augmented reality navigation system for laparoscopic splenectomy in children based on preoperative CT image using optical tracking device. *Pediatr Surg Int* 2012; **28**: 341-346 [PMID: 22130783 DOI: 10.1007/s00383-011-3034-x]

70 **Onda S**, Okamoto T, Kanehira M, Suzuki F, Ito R, Fujioka S, Suzuki N, Hattori A, Yanaga K. Identification of inferior pancreaticoduodenal artery during pancreaticoduodenectomy using augmented reality-based navigation system. *J Hepatobiliary Pancreat Sci* 2014; **21**: 281-287 [PMID: 23970384 DOI: 10.1002/jhbp.25]

71 **Galema HA**, Meijer RPJ, Lauwerends LJ, Verhoef C, Burggraaf J, Vahrmeijer AL, Hutteman M, Keereweer S, Hilling DE. Fluorescence-guided surgery in colorectal cancer; A review on clinical results and future perspectives. *Eur J Surg Oncol*

2021 [PMID: 34657780 DOI: 10.1016/j.ejso.2021.10.005]

72 **Jansen-Winkeln B**, Barberio M, Chalopin C, Schierle K, Diana M, Köhler H, Gockel I, Maktabi M. Feedforward Artificial Neural Network-Based Colorectal Cancer Detection Using Hyperspectral Imaging: A Step towards Automatic Optical Biopsy. *Cancers (Basel)* 2021; **13** [PMID: 33669082 DOI: 10.3390/cancers13050967]

73 **Köhler H**, Kulcke A, Maktabi M, Moulla Y, Jansen-Winkeln B, Barberio M, Diana M, Gockel I, Neumuth T, Chalopin C. Laparoscopic system for simultaneous high-resolution video and rapid hyperspectral imaging in the visible and near-infrared spectral range. *J Biomed Opt* 2020; **25** [PMID: 32860357 DOI: 10.1117/1.JBO.25.8.086004]

74 **Baltussen EJM**, Kok END, Brouwer de Koning SG, Sanders J, Aalbers AGJ, Kok NFM, Beets GL, Flohil CC, Bruin SC, Kuhlmann KFD, Sterenborg HJCM, Ruers TJM. Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery. *J Biomed Opt* 2019; **24**: 1-9 [PMID: 30701726 DOI: 10.1117/1.JBO.24.1.016002]

75 **Ritschel K**, Pechlivanis I, Winter S. Brain tumor classification on intraoperative contrast-enhanced ultrasound. *Int J Comput Assist Radiol Surg* 2015; **10**: 531-540 [PMID: 24956998 DOI: 10.1007/s11548-014-1089-6]

76 **Hou F**, Yu Y, Liang Y. Automatic identification of parathyroid in optical coherence tomography images. *Lasers Surg Med* 2017; **49**: 305-311 [PMID: 28129441 DOI: 10.1002/lsm.22622]

77 **Ştefănescu D**, Streba C, Cârţână ET, Săftoiu A, Gruionu G, Gruionu LG. Computer Aided Diagnosis for Confocal Laser Endomicroscopy in Advanced Colorectal Adenocarcinoma. *PLoS One* 2016; **11**: e0154863 [PMID: 27144985 DOI: 10.1371/journal.pone.0154863]

78 **Krell RW**, Girotti ME, Fritze D, Campbell DA, Hendren S. Hospital readmissions after colectomy: a population-based study. *J Am Coll Surg* 2013; **217**: 1070-1079 [PMID: 24246621 DOI: 10.1016/j.jamcollsurg.2013.07.403]

79 **Lee SW**, Gregory D, Cool CL. Clinical and economic burden of colorectal and bariatric anastomotic leaks. *Surg Endosc* 2020; **34**: 4374-4381 [PMID: 31720809 DOI: 10.1007/s00464-019-07210-1]

80 **Karliczek A**, Harlaar NJ, Zeebregts CJ, Wiggers T, Baas PC, van Dam GM. Surgeons lack predictive accuracy for anastomotic leakage in gastrointestinal surgery. *Int J Colorectal Dis* 2009; **24**: 569-576 [PMID: 19221768 DOI: 10.1007/s00384-009-0658-6]

81 **Jafari MD**, Lee KH, Halabi WJ, Mills SD, Carmichael JC, Stamos MJ, Pigazzi A. The use of indocyanine green fluorescence to assess anastomotic perfusion during robotic assisted laparoscopic rectal surgery. *Surg Endosc* 2013; **27**: 3003-3008 [PMID: 23404152 DOI: 10.1007/s00464-013-2832-8]

82 **Mazaki J**, Katsumata K, Ohno Y, Udo R, Tago T, Kasahara K, Kuwabara H, Enomoto M, Ishizaki T, Nagakawa Y, Tsuchida A. A Novel Predictive Model for Anastomotic Leakage in Colorectal Cancer Using Auto-artificial Intelligence. *Anticancer Res* 2021; **41**: 5821-5825 [PMID: 34732457 DOI: 10.21873/anticanres.15400]

83 **Taha A**, Taha-Mehlitz S, Hendie A, Staudner T, Adamina M. Development and external validation of an international, multicenter machine learning algorithm for prediction of anastomotic insufficiency after colonic or colorectal anastomosis The Prediction of Anastomotic Insufficiency risk after Colorectal surgery (PANIC) study. Available from: https://clinicaltrials.gov/ct2/show/NCT04985981

84 **Tian S**, Yin XC, Wang ZB, Zhou F, Hao HW. A VidEo-Based Intelligent Recognition and Decision System for the Phacoemulsification Cataract Surgery. *Comput Math Methods Med* 2015; **2015**: 202934 [PMID: 26693249 DOI: 10.1155/2015/202934]

85 **Fan B**, Li HX, Hu Y. An Intelligent Decision System for Intraoperative Somatosensory Evoked Potential Monitoring. *IEEE Trans Neural Syst Rehabil Eng* 2016; **24**: 300-307 [PMID: 26415181 DOI: 10.1109/TNSRE.2015.2477557]

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 April 28; 3(2): 46-53

DOI: 10.35712/aig.v3.i2.46

ISSN 2644-3236 (online)

MINIREVIEWS

Application of artificial intelligence in non-alcoholic fatty liver disease and viral hepatitis

Atchayaa Gunasekharan, Joanna Jiang, Ashley Nickerson, Sajid Jalil, Khalid Mumtaz

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): 0
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): E

P-Reviewer: Cannella R, Italy; Mandal P, India

Received: December 31, 2021
Peer-review started: December 31, 2021
First decision: February 7, 2022
Revised: February 18, 2022
Accepted: April 28, 2022
Article in press: April 28, 2022
Published online: April 28, 2022

Atchayaa Gunasekharan, Department of Medicine, Jewish Hospital, Cincinatti, OH 43201, United States

Joanna Jiang, Ashley Nickerson, Sajid Jalil, Khalid Mumtaz, Department of Medicine, Ohio State University, Columbus, OH 43210, United States

Corresponding author: Khalid Mumtaz, MBBS, MSc, Associate Professor, Department of Medicine, Ohio State University, 395 W. 12th Avenue, Columbus, OH 43210, United States. khalid.mumtaz@osumc.edu

Abstract

Non-alcoholic fatty liver disease (NAFLD) and chronic viral hepatitis are among the most significant causes of liver-related mortality worldwide. It is critical to develop reliable methods of predicting progression to fibrosis, cirrhosis, and decompensated liver disease. Current screening methods such as biopsy and transient elastography are limited by invasiveness and observer variation in analysis of data. Artificial intelligence (AI) provides a unique opportunity to more accurately diagnose NAFLD and viral hepatitis, and to identify patients at high risk for disease progression. We conducted a literature review of existing evidence for AI in NAFLD and viral hepatitis. Thirteen articles on AI in NAFLD and 14 on viral hepatitis were included in our analysis. We found that machine learning algorithms were comparable in accuracy to current methods for diagnosis and fibrosis prediction (MELD-Na score, liver biopsy, FIB-4 score, and biomarkers). They also reliably predicted hepatitis C treatment failure and hepatic encephalopathy, for which there are currently no established prediction tools. These studies show that AI could be a helpful adjunct to existing techniques for diagnosing, monitoring, and treating both NAFLD and viral hepatitis.

Key Words: Non-alcoholic fatty liver disease; Non-alcoholic steatohepatitis; Fatty liver; Artificial intelligences; Steatosis; Fibrosis; Machine learning

Core Tip: Non-alcoholic fatty liver disease (NAFLD) exists on a spectrum from simple hepatocyte steatosis to non-alcoholic steatohepatitis (NASH) with ballooning and fibrosis. Given the lack of efficient screening methods and high rate of asymptomatic disease, it is challenging to identify patients with NAFLD in its various stages. Although liver biopsy remains the gold standard for diagnosing NASH, it is an invasive, costly, and painful procedure. Conventional imaging modalities including ultrasound, computed tomography, magnetic resonance imaging and transient elastography are limited by inter- and intra-observer variability depending on the stage of fibrosis. Similarly, despite recent progress in the prevention and treatment of viral hepatitis, predicting sustained virological response and disease progression remains challenging. Artificial intelligence (AI) is an exciting and increasingly pertinent field in medicine as clinicians incorporate augmenting technology into their daily practice. This review summarizes recent literature on the application of AI in NAFLD and viral hepatitis. Specifically, the review will assess the performance of AI as a non-invasive method for the diagnosis and staging of liver fibrosis and steatosis, as well as for the detection and treatment of chronic viral hepatitis. It will also aim to highlight the potential for AI based methods on their ability to develop therapeutic targets.

Citation: Gunasekharan A, Jiang J, Nickerson A, Jalil S, Mumtaz K. Application of artificial intelligence in non-alcoholic fatty liver disease and viral hepatitis. *Artif Intell Gastroenterol* 2022; 3(2): 46-53
URL: https://www.wjgnet.com/2644-3236/full/v3/i2/46.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i2.46

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) exists on a spectrum from simple hepatocyte steatosis to inflammation, ballooning and fibrosis. Given the lack of efficient screening methods and high rate of asymptomatic disease, it is challenging to identify patients with various stages of NAFLD[1,2]. Non-alcoholic steatohepatitis (NASH) patients with significant fibrosis are at increased risk for cirrhosis and progressive liver failure, which has led NASH to become one of the leading causes of liver transplantation in the United States[3]. NASH affects approximately 3% to 6% of the US population, and this number continues to increase. It affects approximately 25% of the population worldwide[4].

Although liver biopsy remains the gold standard for diagnosing NASH, it is an invasive, costly, and painful procedure. Therefore, serial liver biopsies for surveillance are not always feasible. Conventional imaging modalities including ultrasound, computed tomography (CT), magnetic resonance imaging (MRI) and transient elastography are limited by inter- and intra-observer variability depending on the stage of fibrosis[1,2]. Similarly, despite recent progress in the prevention and treatment of viral hepatitis, predicting sustained virological response (SVR) and disease progression remains challenging.

Artificial intelligence (AI) is an exciting and increasingly pertinent field in medicine as clinicians incorporate augmenting technology into their daily practice. AI is the concept of teaching a computer to simulate the cognitive abilities of the human brain. Machine learning (ML) entails allowing the computer to simulate the human brain independently. It can either be supervised (through specific feedback from humans) or unsupervised, in which case there is no guidance provided and the computer is able to independently synthesize and analyze the output[1]. AI is increasingly applied to the diagnosis and prediction of various diseases. Researchers are developing machine learning (ML) algorithms to predict risk and outcomes using multiple demographic, clinical, biochemical, and imaging parameters for diagnosis and prognosis related to liver fibrosis and steatosis, including NAFLD and viral hepatitis[1].

Current methods of assessing liver fibrosis progression and mortality in both NAFLD and viral hepatitis have many limitations. These include the intra- and inter-observer variability in staging fibrosis, the inability to place fibrosis along a continuum, and the lack of identifiable markers for disease progression[1,2]. These limitations and the ability of ML models to overcome them will be discussed further in this review. This review will also highlight how ML models have the potential to present opportunities for drug discovery and prediction of therapeutic and toxic effects of drugs. Machine learning models based on AI provide promising features that could not only enhance screening for NAFLD, but also help with fibrosis staging in patients with NASH and viral hepatitis.

This review summarizes recent literature on the application of AI in NAFLD and viral hepatitis. The main objective is to assess the performance of AI as a non-invasive method for the diagnosis and staging of liver fibrosis and steatosis, as well as the detection and treatment of chronic viral hepatitis.

METHODS

A review of current literature in the areas of AI in NAFLD and viral hepatitis was conducted using two separate searches on PubMed. First, we used the search terms "non-alcoholic fatty liver disease", "NAFLD", and "deep learning" in combination with "artificial intelligence", "histology", "omics" and "radiology." The second search was conducted using the search terms "viral hepatitis" in combination with "hepatitis A", "hepatitis B", "hepatitis C", "hepatitis E", "machine learning", "artificial intelligence", "histology" and "radiology".

Most articles on NASH and NAFLD published between 2018 and 2021 were included in this review. Articles were excluded if they did not offer comparisons between AI modalities and existing methods for screening or prediction (MELD score, elastography, *etc.*). Twenty-seven articles were included in our review, 13 on NAFLD and 14 on chronic viral hepatitis. For studies on viral hepatitis, described machine learning algorithms fell into one of three categories: Predicting prevalence, screening for complications (including fibrosis, hepatocellular carcinoma, decompensated cirrhosis, and death), and predicting response to treatment.

USE OF AI FOR DIAGNOSING VIRAL HEPATITIS AND NAFLD/NASH

It is estimated that half of patients infected with hepatitis C worldwide are unaware of their diagnosis and only 17% have undergone liver fibrosis staging[5]. This rate is even lower for hepatitis B, for which only 10.5% of infected patients are aware of their status. In March 2020, the USPSTF recommended hepatitis C screening for all adults over 18; however, there are currently no population-based screening recommendations for hepatitis A and B. Primary care offices do not routinely test for hepatitis B. Machine learning has been used both to determine regional prevalence of chronic hepatitis and to identify undiagnosed cases.

Zheng *et al*[6] compared two algorithms (Elman neural network and autoregressive integrated moving average, or ARIMA) designed to predict incidence of hepatitis B in Guangxi, China. ARIMA is a type of model that can capture the randomness of data and is often used for infectious disease prediction. Predictions were compared to the reported cases of hepatitis B cases from the Health Commission of Guangxi, China. The neural network was the more predictive model, with a root-mean-square error (RMSE) of 0.89 and mean absolute error (MAE) of 0.70, while the ARIMA had an RSME of 0.94 and an MAE of 0.81.

A 2020 study by Doyle *et al*[7] aimed to predict chronic hepatitis C (HCV) positive status by using patient claims data to develop four algorithms, all with a predictive accuracy of over 95%. Algorithms included logistic regression, gradient boosted trees, a stacked ensemble, and random forests. The stacked ensemble performed the best, with a precision of 97% at recall levels > 50%. Key predictors of HCV infection included nonsteroidal anti-inflammatory drug use, opioids, healthcare utilization, patient age and osteoarthritis or glomerulonephritis treatment. We were unable to find any study to date using AI to screen for NAFLD/NASH.

USE OF AI TO ASSESS FIBROSIS IN VIRAL HEPATITIS AND NAFLD/NASH

Existing histologic models not only rely on scoring of fibrosis by a pathologist but are also unable to place fibrosis along a continuum. Artificial intelligence enables the placement of fibrosis along a continuum, identifies risk factors for progression of fibrosis, allows enhanced scoring of fibrosis stages, leading to better selection of patients for clinical trials This also allows for identification of therapeutic targets[2].

Lu *et al*[8] developed a light gradient-boosting machine model to predict liver fibrosis and cirrhosis in treatment-naive chronic hepatitis B patients at four centers in China. The model, named Fibro Box, outperformed transient elastography, APRI, and FIB-4, with area under the curve (AUC) 0.88 in external validation sets for significant fibrosis and 0.87 for cirrhosis. Input variables included fibroscan results, platelets, alanine aminotransferase (ALT), Prothrombin time (PT), and splenic vein diameter.

A 2013 study by Zheng *et al*[9], used an artificial neural network (ANN) to predict 3-month mortality of individuals with acute-on-chronic liver failure due to hepatitis B (HBV-ACLF). Patient characteristics included in this model were age, PT, serum sodium, total bilirubin, E antigen positivity status and hemoglobin. The ANN predicted mortality more accurately than MELD-based scoring systems, with area under the curve receiver operating characteristic (AUCROC) 0.765 in the validation cohort compared to 0.599 for MELD.

Similarly, Huo *et al*[10] developed ANNs to predict 28- and 90-d mortality in HBV-ACLF. Data were retrospectively reviewed from 684 patients admitted for ALF at 8 hospitals in various Chinese provinces with 423 cases in the training cohort and 261 in the validation cohort. In the training cohorts, the neural network had a significantly higher accuracy than MELD, MELD-Na, CLIF-ACLF, and Child-Pugh score, with AUC 0.948 and 0.913 for 28- and 90-d mortality, respectively. In the validation cohort, the model

performed significantly better than MELD and insignificantly better than other scoring systems, with AUC 0.748 and 0.754 for 28- and 90-d mortality. Significant mortality predictors included age, presence of HE, sodium, PT, gamma-glutamyl transpeptidase (GGT), e antigen, alkaline phosphatase, and bilirubin.

In another study, Wang *et al*[11] used deep learning radiomics of elastography (DLRE) to assess stages of liver fibrosis in patients with chronic hepatitis B. DLRE was compared to 2D shear wave elastography and biomarkers (AST: Platelet ratio, fibrosis index), with liver biopsy as the reference standard. 1990 images from 398 patients were used to develop the models. AUCROCs for DLRE were 0.97 for cirrhosis, 0.98 for advanced fibrosis, and 0.85 for significant fibrosis; this performed better than other methods except for elastography in severe fibrosis.

Like viral hepatitis, there are several studies establishing the role of AI in assessing fibrosis in NAFLD/NASH. In one study by Forlano *et al*[2], liver biopsy specimens were annotated by two expert pathologists using the clinical research network (CRN) score as a measurable scale of degree of steatosis, inflammation, ballooning and fibrosis. The machine learning model was built using 100 patients with NAFLD in the derivation group and 146 patients in the validation group. There was good concordance when the machine learning model was compared to the scoring of the expert histopathologist on the liver biopsy specimens; the interclass correlation coefficients were 0.97 (95%CI, 0.95-0.99; P value < 0.001) for steatosis, 0.96 (95%CI, 0.9-0.98; P value < 0.001) for inflammation, 0.94 (95%CI, 0.87-0.98; P value < 0.001 for ballooning, and 0.92 for fibrosis (95%CI, 0.88-0.96; P value < 0.001). A subgroup analysis showed that quantitative analysis performed better than the CRN score in differentiating between the various stages of NAFLD. Another CNN model developed by Qu *et al*[12], showed that a convolutional neural network (CNN) model had an area under the curve (AUC) of 63% for all four subsets of the NAFLD scoring, while the AUC's were 90.48% for steatosis, 81.06% for ballooning, 70.18% for inflammation and 83.85% for fibrosis. These studies underscore the utility of ML models in illustrating the heterogeneity of liver pathology in NAFLD[9,26].

In another study by Taylor-Weiner *et al*[13], a CNN model was developed that allowed for assessment of fibrosis along a continuum, which is not possible with pathologist scoring alone. The CRN and Ishak scores were applied to each pixel within a given image, allowing for evaluation of heterogeneity in fibrosis as well. In addition, the CNN served as a prediction model allowing for identification of features associated with disease progression. The model's predictions correlated significantly with the pathologist scoring in all three studies, the STELLAR-3, STELLAR-4, and ATLAS, whose participants were used to build and validate the ML model - steatosis, $\rho = 0.60$; P value < 0.001; lobular inflammation, $\rho = 0.35$; P value < 0.001; and HB, $\rho = 0.41$; P value < 0.001. The model's level of agreement with pathologist scoring was within the range of agreement between individual pathologists. The weighted Cohen's kappa was 0.801 for NASH CRN and 0.817 for the Ishak classifications.

Another study by Gawrieh *et al*[14] built a ML model using support vector machines (SVM) to better characterize architectural patterns in fibrosis. This ML model was built to differentiate between six different patterns of fibrosis and had a strong correlation with the pathologist's semi-quantitative scores for fibrosis, with a coefficient of determination of automated CPA ranging between 0.60 to 0.86 when compared with the pathologist score. The model was built using a trichrome-stained liver biopsy specimen which was marked with 987 annotations for different fibrosis types. As noted in the study, the model's AUROCs were 78.6% for detection of periportal fibrosis, 83.3% for pericellular fibrosis, 86.4% for portal fibrosis, and > 90% for detection of normal fibrosis, bridging fibrosis and presence of nodules/cirrhosis.

AI USING METABOLOMICS FOR NAFLD/NASH

There is an increasing number of studies focusing on metabolomics that allow for non-invasive identification of targets associated with development and progression of NAFLD. These biomarkers may differentiate between patients with and without cirrhosis, and between a healthy liver and NAFLD or NASH[3,15,16]. Several direct and indirect blood-based biomarkers currently exist to assess fibrosis. These have been incorporated to form scoring systems such as NAFLD fibrosis score (NFS), Fibrosis-4 (FIB-4), AST to platelet ratio index (APRI), BARD Score, FibroSURE and Enhanced liver fibrosis score [3]. ML allows for analysis of many multi-omics and clinical variables to screen for NASH and NAFLD and to build models for disease progression.

An eXtreme Gradient Boosting Model (XG Boost) was developed using the NIDDK database by Docherty *et al*[16], which contains a large real-world patient population. This model used confirmed NASH and non-NASH patients within this subset. The unique feature of this study is that it used several demographic variables and clinical biomarkers run through recursive feature elimination in combination with confirmed histologic cases to build an efficient model with a high specificity. When a greater number of markers were used in predicting patients with NASH, the AUROC was 0.82, sensitivity 81%, and precision 81%.

In a study of adults of European ancestry by Atabaski-Pasdar *et al*[15], patients with type 2 diabetes and others with high-risk features for the development of NASH were assessed for liver fat content

using MRI. Several multi-omics and clinical data, including laboratory markers, were entered into the least absolute shrinkage and selection operator to select the most relevant features, which then underwent random forest analysis for the development of the algorithm. The model developed using this method produced a cross-validated AUROC of 0.84 (95%CI 0.82, 0.86; *P* value < 0.001) and outperformed existing prediction tools for NAFLD. However, unlike other studies, the model was built in comparison to MRI fat content, which is not reflective of the continuum of NAFLD, and thus cannot be used to monitor disease progression.

Another study based in China by Ma *et al*[17] identified BMI, triglycerides, GGT, the serum ALT and uric acid as the most common features contributing to NAFLD when a Bayesian network model was used. The model had an accuracy of 83%, specificity of 0.878, sensitivity of 0.675, and F-measure score of 0.655. The F-measure score is an indicator of whether there can be a balance between precision and recall of these variables, and it was higher than for logistic regression models in machine learning.

AI IN IMAGE INTERPRETATION FOR NAFLD/NASH

Like markers discussed previously, many studies have combined machine learning with imaging modalities to more effectively assess liver fat content and to better define fibrosis scores. This would allow for more accurate monitoring of patients for disease progression and their selection for clinical trials.

Current modalities for estimation of liver fat content include conventional ultrasound (US), which is limited by variable accuracy, operator dependency, and its qualitative nature. The measurement of proton density fat fraction (PDFF) by MRI is proving to be an effective method for quantification of hepatic steatosis, but it is expensive and there is variability in results due to dependence on calibration. In a study by Han *et al*[18], one-dimensional CNN was applied to ultrasound radiofrequency signals for the diagnosis of NAFLD and quantitation of hepatic fat content with an AUC of 0.98 (95%CI: 0.94, 1.00). In diagnosing NAFLD, the model had an accuracy of 96%, sensitivity of 97%, and specificity of 94%, PPV of 97% and NPV of 94%. The ML model also correlated with MRI-PDFF with a Pearson correlation coefficient of 0.85 (*P* value < 0.001). The same method was applied to animal models in a study by Nguyen *et al*[19] and it showed that CNN outperformed quantitative ultrasound in differentiating between NAFLD and normal liver. Further support for ML comes from a recent study by Das *et al*[20] on pediatric patients which used an ensemble model comprising SVM, Neural Net and XG Boost that had an AUC of 0.92 (95%CI, 0.91-0.94) when tested in an external validation cohort.

Nonenhanced CT also remains superior to histopathologic quantification of liver fat content like MRI-PDFF, but it is also more commonly performed in clinical practice for other reasons when compared to MRI. It currently uses a manual region-of-interest (ROI) for estimation of liver fat content. A study by Graffy *et al*[21] developed a deep-learning based automated liver segmentation tool and applied it to estimate liver fat content using three-dimensional CNN, without having to depend on manual ROI. The pearson correlation coefficient was 0.93. This allows for large population level estimation of liver fat content to determine the prevalence of NAFLD. It would also determine normal liver fat content based on a large sample. Used in combination with other non-invasive modalities such as serum biomarkers, it could help identify patients who will need closer monitoring for NAFLD progression to cirrhosis. In a similar study by Hou *et al*[22], the automated liver attenuation ROI-based measurement model had a pearson coefficient of 0.94 when compared with manual ROI.

In addition to differentiating healthy liver from NAFLD, ML models have also been used to reduce variability in detecting fibrosis, specifically F2 fibrosis, which is a limiting feature of shear wave elastography. A study by Brattain *et al*[23] combined the use of shear wave elastography with CNN to better assess F2 fibrosis. This approach not only assessed image quality, but also selected ROI, unlike the previous studies. This ML model detected F2 fibrosis with AUC of 0.89 compared to AUC of 0.74 when image quality and ROI were not incorporated into a ML model. This demonstrates the importance of ML models once again in selecting patients for clinical trials, and in assessing response to treatment.

AI IN VIRAL HEPATITIS TREATMENT

The rate of SVR for hepatitis C with modern direct acting antiviral (DAA) regimens is estimated to be over 90%; however, variability remains in treatment length and efficacy. Patients with prior DAA exposure, cirrhosis, and other risk factors may require a longer treatment course[18,24]. Machine learning has been applied to predicting treatment response and duration based on patient-specific factors.

Haga *et al*[24] applied nine machine learning algorithms to identify the optimized combination of HCV genotypic variants that predict SVR after DAA therapy. HCV genomes were sequenced from the serum of 173 patients (including 64 without SVR). The support vector machine algorithm was found to be the most predictive, with a validation accuracy of 0.95. Feldman *et al*[25] used data from 60 million beneficiaries of a managed care plan (including 3943 cases of hepatitis C who received sofos-

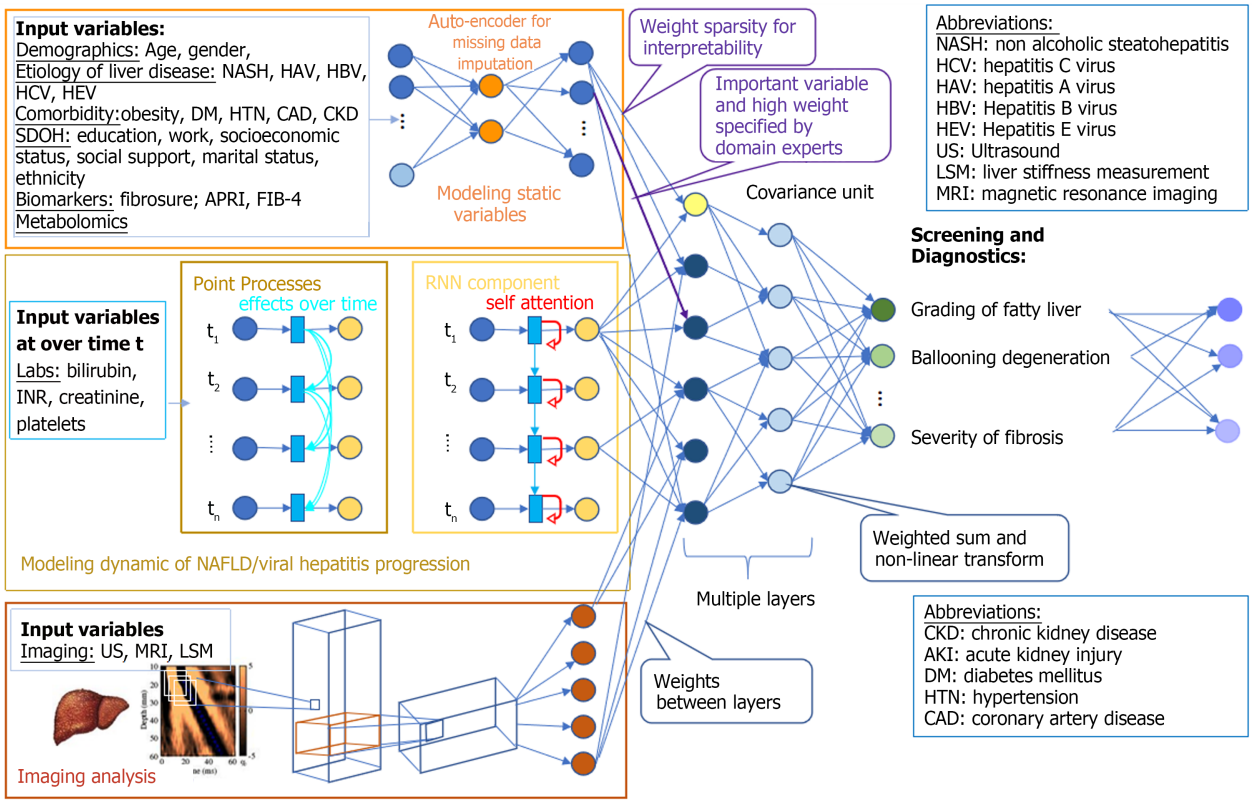


Figure 1 Framework of artificial intelligence based dynamic of non-alcoholic fatty liver disease/viral hepatitis diagnosis, progression and outcomes.

buvir/ledipasvir), to identify demographic and medical factors that may predict a prolonged course of DAA. Machine learning algorithms included extreme gradient boosting (XG Boost), random forest and support vector machine, with XG Boost being the optimal predictive model at an AUC of 0.745. Patient age, comorbidity burden, and type 2 diabetes status were significant predictors. Wei *et al*[26] developed an ANN and logistic regression model to predict fibrosis reversal after 78 wk of hepatitis B treatment. Significant predictors included AST and ALT, platelets, WBC, gender, and Fibroscan results. The ANN outperformed the logistic regression model, with an AUC of 0.81 *vs* 0.75.

The only approved treatment for NAFLD is weight reduction. We were unable to find AI based algorithms and predictive models for NAFLD due to lack of pharmacologic management options.

DISCUSSION

Among the algorithms described, more complex models performed better, with machine learning consistently outperforming more basic logistic regression models. The highest-performing models incorporated both demographic and radiologic/serologic variables. AI models also predicted complications more accurately than biomarkers and scoring systems like MELD and FIB-4. These models could be used to predict the incidence and prevalence of viral hepatitis in regions without robust, widespread screening programs. Additionally, they could be helpful in the initiation of treatment and predicting response to antivirals for individual patients, for which no gold standard currently exists.

Limitations of the current AI models are notably due to the lack of large scale, randomized controlled trials. Further research is necessary to demonstrate the utility of AI. With further advancements, ML models could potentially be incorporated into all aspects of a patient's care, from screening the general population for NAFLD or NASH, to monitoring disease progression and treatment response in clinical trials by enhancing classification of steatosis, ballooning, inflammation, and fibrosis. In this regard, more population-based studies are needed to study the applications of ML models in screening. Additionally, large scale, randomized controlled trials are needed to study serologic and histologic markers for disease progression. Further studies are also warranted to explore the potential of ML algorithms to provide target-specific medications, yielding efficacious pharmacotherapy in a disease such as NASH where good treatment options are lacking at this time. Though AI is promising in terms of its potential to develop therapeutic targets, we were unable to find any studies to date describing the use of AI in drug discovery.

Future directions also include using AI to actively improve outcomes with viral hepatitis by increasing adherence to DAAs or identifying individuals at risk for contracting viral hepatitis. Machine learning models could also help identify barriers to accessing treatment.

CONCLUSION

Machine learning models focus on various aspects of liver disease, including demographics, biochemical labs, histologic assessment and patterns, identification of non-invasive biomarkers, and liver imaging techniques (Figure 1). Overall, the studies outlined above are promising in their reliance on non-invasive methods as opposed to conventional liver biopsy to study the stages of fibrosis, as well as their ability to place fibrosis along a continuum and identify markers for disease progression. This could reduce healthcare costs by allowing better selection of patients in whom a liver biopsy is performed. It would also benefit patients by decreasing the number of them who undergo this invasive procedure. AI can also improve efficiency of pathologist and sonographer scoring of samples when added to existing methods. This will allow for a better understanding of the pathophysiology of diseases like NAFLD, which would not only allow for appropriate screening for disease progression, but also improve the ability to develop therapeutic targets.

FOOTNOTES

Author contributions: Gunasekharan A analyzed articles, wrote and reviewed manuscript; Jiang J analyzed articles relating to viral hepatitis, wrote and reviewed those portions; Nickerson A reviewed manuscript; Jalil S reviewed manuscript; Mumtaz K helped with layout of manuscript, analyzed articles, wrote and revised manuscript.

Conflict-of-interest statement: None of the authors have any conflicts of interest to report.

Country/Territory of origin: United States

ORCID number: Atchayaa Gunasekharan 0000-0001-7868-6513; Joanna Jiang 0000-0002-7995-6499; Ashley Nickerson 0000-0002-4494-9985; Sajid Jalil 0000-0001-6123-153X; Khalid Mumtaz 0000-0001-7868-6514.

S-Editor: Liu JH
L-Editor: A
P-Editor: Cai YX

REFERENCES

1 **Aggarwal P**, Alkhouri N. Artificial Intelligence in Nonalcoholic Fatty Liver Disease: A New Frontier in Diagnosis and Treatment. *Clin Liver Dis (Hoboken)* 2021; **17**: 392-397 [PMID: 34386201 DOI: 10.1002/cld.1071]
2 **Forlano R**, Mullish BH, Giannakeas N, Maurice JB, Angkathunyakul N, Lloyd J, Tzallas AT, Tsipouras M, Yee M, Thursz MR, Goldin RD, Manousou P. High-Throughput, Machine Learning-Based Quantification of Steatosis, Inflammation, Ballooning, and Fibrosis in Biopsies From Patients With Nonalcoholic Fatty Liver Disease. *Clin Gastroenterol Hepatol* 2020; **18**: 2081-2090.e9 [PMID: 31887451 DOI: 10.1016/j.cgh.2019.12.025]
3 **Perakakis N**, Polyzos SA, Yazdani A, Sala-Vila A, Kountouras J, Anastasilakis AD, Mantzoros CS. Non-invasive diagnosis of non-alcoholic steatohepatitis and fibrosis with the use of omics and supervised learning: A proof of concept study. *Metabolism* 2019; **101**: 154005 [PMID: 31711876 DOI: 10.1016/j.metabol.2019.154005]
4 **Leow WQ**, Bedossa P, Liu F, Wei L, Lim KH, Wan WK, Ren Y, Chang JP, Tan CK, Wee A, Goh GB. An Improved qFibrosis Algorithm for Precise Screening and Enrollment into Non-Alcoholic Steatohepatitis (NASH) Clinical Trials. *Diagnostics (Basel)* 2020; **10** [PMID: 32872090 DOI: 10.3390/diagnostics10090643]
5 **Yehia BR**, Schranz AJ, Umscheid CA, Lo Re V 3rd. The treatment cascade for chronic hepatitis C virus infection in the United States: a systematic review and meta-analysis. *PLoS One* 2014; **9**: e101554 [PMID: 24988388 DOI: 10.1371/journal.pone.0101554]
6 **Zheng Y**, Zhang L, Zhu X, Guo G. A comparative study of two methods to predict the incidence of hepatitis B in Guangxi, China. *PLoS One* 2020; **15**: e0234660 [PMID: 32579598 DOI: 10.1371/journal.pone.0234660]
7 **Doyle OM**, Leavitt N, Rigg JA. Finding undiagnosed patients with hepatitis C infection: an application of artificial intelligence to patient claims data. *Sci Rep* 2020; **10**: 10521 [PMID: 32601354 DOI: 10.1038/s41598-020-67013-6]

8 **Lu XJ**, Yang XJ, Sun JY, Zhang X, Yuan ZX, Li XH. FibroBox: a novel noninvasive tool for predicting significant liver fibrosis and cirrhosis in HBV infected patients. *Biomark Res* 2020; **8**: 48 [PMID: 33005419 DOI: 10.1186/s40364-020-00215-2]

9 **Zheng MH**, Shi KQ, Lin XF, Xiao DD, Chen LL, Liu WY, Fan YC, Chen YP. A model to predict 3-month mortality risk of acute-on-chronic hepatitis B liver failure using artificial neural network. *J Viral Hepat* 2013; **20**: 248-255 [PMID: 23490369 DOI: 10.1111/j.1365-2893.2012.01647.x]

10 **Huo Y**, Terry JG, Wang J, Nair S, Lasko TA, Freedman BI, Carr JJ, Landman BA. Fully automatic liver attenuation estimation combing CNN segmentation and morphological operations. *Med Phys* 2019; **46**: 3508-3519 [PMID: 31228267 DOI: 10.1002/mp.13675]

11 **Wang K**, Lu X, Zhou H, Gao Y, Zheng J, Tong M, Wu C, Liu C, Huang L, Jiang T, Meng F, Lu Y, Ai H, Xie XY, Yin LP, Liang P, Tian J, Zheng R. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut* 2019; **68**: 729-741 [PMID: 29730602 DOI: 10.1136/gutjnl-2018-316204]

12 **Qu H**, Minacapelli CD, Tait C, Gupta K, Bhurwal A, Catalano C, Dafalla R, Metaxas D, Rustgi VK. Training of computational algorithms to predict NAFLD activity score and fibrosis stage from liver histopathology slides. *Comput Methods Programs Biomed* 2021; **207**: 106153 [PMID: 34020377 DOI: 10.1016/j.cmpb.2021.106153]

13 **Taylor-Weiner A**, Pokkalla H, Han L, Jia C, Huss R, Chung C, Elliott H, Glass B, Pethia K, Carrasco-Zevallos O, Shukla C, Khettry U, Najarian R, Taliano R, Subramanian GM, Myers RP, Wapinski I, Khosla A, Resnick M, Montalto MC, Anstee QM, Wong VW, Trauner M, Lawitz EJ, Harrison SA, Okanoue T, Romero-Gomez M, Goodman Z, Loomba R, Beck AH, Younossi ZM. A Machine Learning Approach Enables Quantitative Measurement of Liver Histology and Disease Monitoring in NASH. *Hepatology* 2021; **74**: 133-147 [PMID: 33570776 DOI: 10.1002/hep.31750]

14 **Gawrieh S**, Sethunath D, Cummings OW, Kleiner DE, Vuppalanchi R, Chalasani N, Tuceryan M. Automated quantification and architectural pattern detection of hepatic fibrosis in NAFLD. *Ann Diagn Pathol* 2020; **47**: 151518 [PMID: 32531442 DOI: 10.1016/j.anndiagpath.2020.151518]

15 **Atabaki-Pasdar N**, Ohlsson M, Viñuela A, Frau F, Pomares-Millan H, Haid M, Jones AG, Thomas EL, Koivula RW, Kurbasic A, Mutie PM, Fitipaldi H, Fernandez J, Dawed AY, Giordano GN, Forgie IM, McDonald TJ, Rutters F, Cederberg H, Chabanova E, Dale M, Masi F, Thomas CE, Allin KH, Hansen TH, Heggie A, Hong MG, Elders PJM, Kennedy G, Kokkola T, Pedersen HK, Mahajan A, McEvoy D, Pattou F, Raverdy V, Häussler RS, Sharma S, Thomsen HS, Vangipurapu J, Vestergaard H, 't Hart LM, Adamski J, Musholt PB, Brage S, Brunak S, Dermitzakis E, Frost G, Hansen T, Laakso M, Pedersen O, Ridderstråle M, Ruetten H, Hattersley AT, Walker M, Beulens JWJ, Mari A, Schwenk JM, Gupta R, McCarthy MI, Pearson ER, Bell JD, Pavo I, Franks PW. Predicting and elucidating the etiology of fatty liver disease: A machine learning modeling and validation study in the IMI DIRECT cohorts. *PLoS Med* 2020; **17**: e1003149 [PMID: 32559194 DOI: 10.1371/journal.pmed.1003149]

16 **Docherty M**, Regnier SA, Capkun G, Balp MM, Ye Q, Janssens N, Tietz A, Löffler J, Cai J, Pedrosa MC, Schattenberg JM. Development of a novel machine learning model to predict presence of nonalcoholic steatohepatitis. *J Am Med Inform Assoc* 2021; **28**: 1235-1241 [PMID: 33684933 DOI: 10.1093/jamia/ocab003]

17 **Ma H**, Xu CF, Shen Z, Yu CH, Li YM. Application of Machine Learning Techniques for Clinical Predictive Modeling: A Cross-Sectional Study on Nonalcoholic Fatty Liver Disease in China. *Biomed Res Int* 2018; **2018**: 4304376 [PMID: 30402478 DOI: 10.1155/2018/4304376]

18 **Han A**, Byra M, Heba E, Andre MP, Erdman JW Jr, Loomba R, Sirlin CB, O'Brien WD Jr. Noninvasive Diagnosis of Nonalcoholic Fatty Liver Disease and Quantification of Liver Fat with Radiofrequency Ultrasound Data Using One-dimensional Convolutional Neural Networks. *Radiology* 2020; **295**: 342-350 [PMID: 32096706 DOI: 10.1148/radiol.2020191160]

19 **Nguyen TN**, Podkowa AS, Park TH, Miller RJ, Do MN, Oelze ML. Use of a convolutional neural network and quantitative ultrasound for diagnosis of fatty liver. *Ultrasound Med Biol* 2021; **47**: 556-568 [PMID: 33358553 DOI: 10.1016/j.ultrasmedbio.2020.10.025]

20 **Das A**, Connell M, Khetarpal S. Digital image analysis of ultrasound images using machine learning to diagnose pediatric nonalcoholic fatty liver disease. *Clin Imaging* 2021; **77**: 62-68 [PMID: 33647632 DOI: 10.1016/j.clinimag.2021.02.038]

21 **Graffy PM**, Sandfort V, Summers RM, Pickhardt PJ. Automated Liver Fat Quantification at Nonenhanced Abdominal CT for Population-based Steatosis Assessment. *Radiology* 2019; **293**: 334-342 [PMID: 31526254 DOI: 10.1148/radiol.2019190512]

22 **Hou Y**, Zhang Q, Gao F, Mao D, Li J, Gong Z, Luo X, Chen G, Li Y, Yang Z, Sun K, Wang X. Artificial neural network-based models used for predicting 28- and 90-day mortality of patients with hepatitis B-associated acute-on-chronic liver failure. *BMC Gastroenterol* 2020; **20**: 75 [PMID: 32188419 DOI: 10.1186/s12876-020-01191-5]

23 **Brattain LJ**, Telfer BA, Dhyani M, Grajo JR, Samir AE. Objective Liver Fibrosis Estimation from Shear Wave Elastography. *Annu Int Conf IEEE Eng Med Biol Soc* 2018; **2018**: 1-5 [PMID: 30440285 DOI: 10.1109/EMBC.2018.8513011]

24 **Haga H**, Sato H, Koseki A, Saito T, Okumoto K, Hoshikawa K, Katsumi T, Mizuno K, Nishina T, Ueno Y. A machine learning-based treatment prediction model using whole genome variants of hepatitis C virus. *PLoS One* 2020; **15**: e0242028 [PMID: 33152046 DOI: 10.1371/journal.pone.0242028]

25 **Feldman TC**, Dienstag JL, Mandl KD, Tseng YJ. Machine-learning-based predictions of direct-acting antiviral therapy duration for patients with hepatitis C. *Int J Med Inform* 2021; **154**: 104562 [PMID: 34482150 DOI: 10.1016/j.ijmedinf.2021.104562]

26 **Wei W**, Wu X, Zhou J, Sun Y, Kong Y, Yang X. Noninvasive Evaluation of Liver Fibrosis Reverse Using Artificial Neural Network Model for Chronic Hepatitis B Patients. *Comput Math Methods Med* 2019; **2019**: 7239780 [PMID: 31428186 DOI: 10.1155/2019/7239780]

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 April 28; 3(2): 54-65

DOI: 10.35712/aig.v3.i2.54

ISSN 2644-3236 (online)

MINIREVIEWS

Machine learning in endoscopic ultrasonography and the pancreas: The new frontier?

Cem Simsek, Linda S Lee

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): A
Grade B (Very good): B
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Dabbakuti JRKKK, India; Hanada E, Japan; Wang P, China

Received: February 1, 2022
Peer-review started: February 1, 2022
First decision: February 18, 2022
Revised: March 28, 2022
Accepted: April 19, 2022
Article in press: April 19, 2022
Published online: April 28, 2022

Cem Simsek, Department of Gastroenterology, Hepatology and Endoscopy, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02215, United States

Linda S Lee, Division of Gastroenterology, Hepatology and Endoscopy, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, United States

Corresponding author: Linda S Lee, MD, Associate Professor, Division of Gastroenterology, Hepatology and Endoscopy, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA 02115, United States. lslee@partners.org

Abstract

Pancreatic diseases have a substantial burden on society which is predicted to increase further over the next decades. Endoscopic ultrasonography (EUS) remains the best available diagnostic method to assess the pancreas, however, there remains room for improvement. Artificial intelligence (AI) approaches have been adopted to assess pancreatic diseases for over a decade, but this methodology has recently reached a new era with the innovative machine learning algorithms which can process, recognize, and label endosonographic images. Our review provides a targeted summary of AI in EUS for pancreatic diseases. Included studies cover a wide spectrum of pancreatic diseases from pancreatic cystic lesions to pancreatic masses and diagnosis of pancreatic cancer, chronic pancreatitis, and autoimmune pancreatitis. For these, AI models seemed highly successful, although the results should be evaluated carefully as the tasks, datasets and models were greatly heterogenous. In addition to use in diagnostics, AI was also tested as a procedural real-time assistant for EUS-guided biopsy as well as recognition of standard pancreatic stations and labeling anatomical landmarks during routine examination. Studies thus far have suggested that the adoption of AI in pancreatic EUS is highly promising and further opportunities should be explored in the field.

Key Words: Artificial intelligence; Pancreas; Endoscopic ultrasonography; Pancreatic cancer; Autoimmune pancreatitis; Pancreatic cystic lesions

Core Tip: Several reviews in the literature have discussed the use of artificial intelligence in pancreatic disease. However, this is the first review that focuses on the application of artificial intelligence (AI) specifically to endoscopic ultrasonography (EUS) of the pancreas, including pancreatic cystic lesions, pancreatic cancer, chronic pancreatitis, and autoimmune pancreatitis, where it appears to enhance EUS diagnosis. AI may also offer real-time assistance during procedures to direct biopsy towards the highest yield areas as well augment EUS training.

Citation: Simsek C, Lee LS. Machine learning in endoscopic ultrasonography and the pancreas: The new frontier? *Artif Intell Gastroenterol* 2022; 3(2): 54-65
URL: https://www.wjgnet.com/2644-3236/full/v3/i2/54.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i2.54

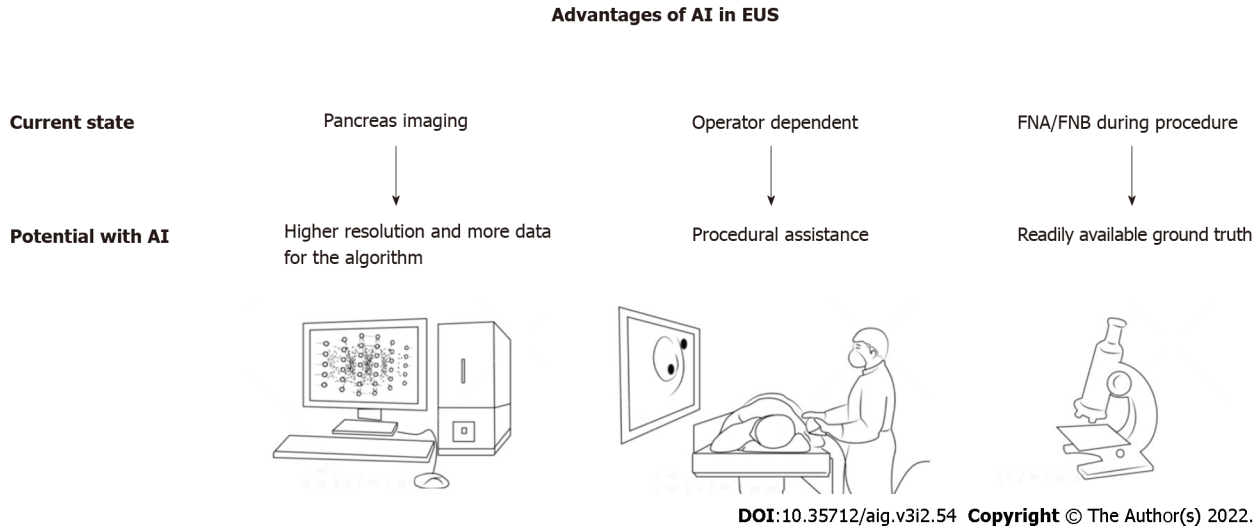
INTRODUCTION

Pancreatic diseases create a substantial burden on society. Pancreatic cancer is the third leading cause of cancer-related death in the United States, and its death count is expected to rise to 460000 by 2040, becoming the second leading cause of cancer related death in 2040[1-3]. Chronic pancreatitis is another cause of the burden with significant morbidity from chronic pain, diabetes mellitus, and even pancreatic cancer[4,5]. Additionally, pancreatic cystic lesions are reported to be detected up to 20% of abdominal imaging studies[6]. Endoscopic ultrasonography (EUS) has surpassed magnetic resonance imaging (MRI), computed tomography (CT) and transabdominal ultrasonography in the diagnosis of pancreatic diseases; however, there remains room for improvement in the diagnostic sensitivity of EUS[7]. In this regard, utilization of artificial intelligence (AI) with EUS has emerged as a promising strategy (Figure 1). Although EUS has better performance than the alternative radiology imaging methods, it is also more operator dependent. The endosonographer's experience and skills can significantly alter the diagnostic or therapeutic outcomes of an EUS procedure. AI may decrease this operator dependency as it can assist the endosonographer in several tasks that include, but are not limited, to identifying anatomical landmarks, detecting lesions, interpreting sonographic findings, and guiding obtaining optimal tissue biopsy with higher diagnostic yield. Because AI algorithms use higher resolution EUS imaging data, they might distinguish patterns and identify details from the images which may not be recognizable with human detection alone currently. Finally, AI research with EUS is more convenient because imaging data used to train the AI models often have readily available definitive histologic diagnoses.

Targeted summary of AI and research

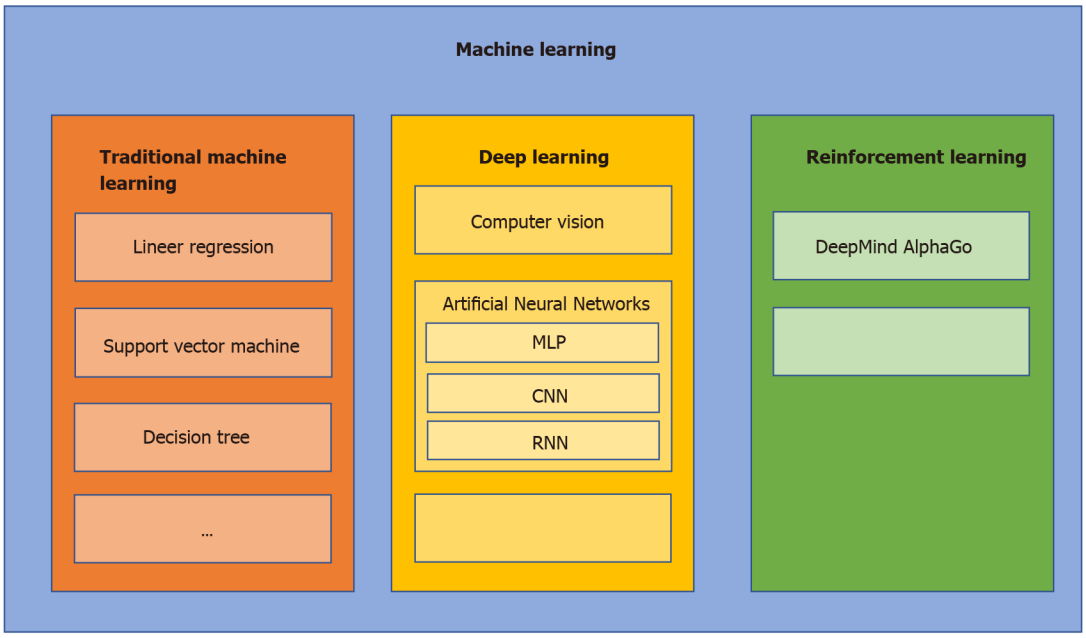
AI is an umbrella term for the computerized performance of complex tasks that normally require human intelligence, such as visual perception, learning, pattern recognition and decision-making[8] (Figure 2). Current medical applications using AI have made significant progress due to advancements in computer technology, data science, and the digitalization of health care. From the development of more complex machine learning algorithms, AI has progressed rapidly to its current front-line role in image-based diagnosis, speech recognition, robotic surgery, drug discovery and patient monitoring[9]. However, the progress of AI in medicine has just begun and has yet to realize its full potential.

Machine learning (ML) is a field of artificial intelligence in which algorithms learn and improve from interactions with the data, obviating the need for explicit programming. Deep learning (DL) is a subfield of ML inspired by the organization and working principle of the human brain and is made up of individual neurons which form multilayered artificial neural networks (ANN). These networks are comprised of input and output layers each of which can execute simple tasks and sequentially interact with one another to produce a conclusion. Among ANNs, Multi-Layered Perceptron are earlier models that are simpler with fewer layers and can only use linear functions[10]. Convolutional neural networks (CNN) include more layers that can also operate in a non-linear fashion allowing more complex tasks such as image classification and have been the most popular DL algorithm. CNNs were inspired by the human visual cortex and designed to process grid pattern data such as images. They have serial neural network layers to recognize and extract features from the input data, learn the patterns of features, and perform hierarchical organization through the layers to search for the intended output (Figure 3)[11]. Most commonly used CNN algorithms are AlexNet, ResNet, U-Net, which all work using the same principle, and the technical details are beyond the scope of this review[12]. Another type of ANN is recurrent neural network (RNN), which also contains a multi-layered structure. In addition, each neuron in this network has its own internal memory, which taken altogether constitutes a collective memory of the network. This neural network can remember previous input data and use it to process subsequent inputs. Therefore, these algorithms are beneficial in processing sequential data such as before and after an intervention or time series data. An example of RNN is the long short-term memory model[13].



DOI:10.35712/aig.v3i2.54

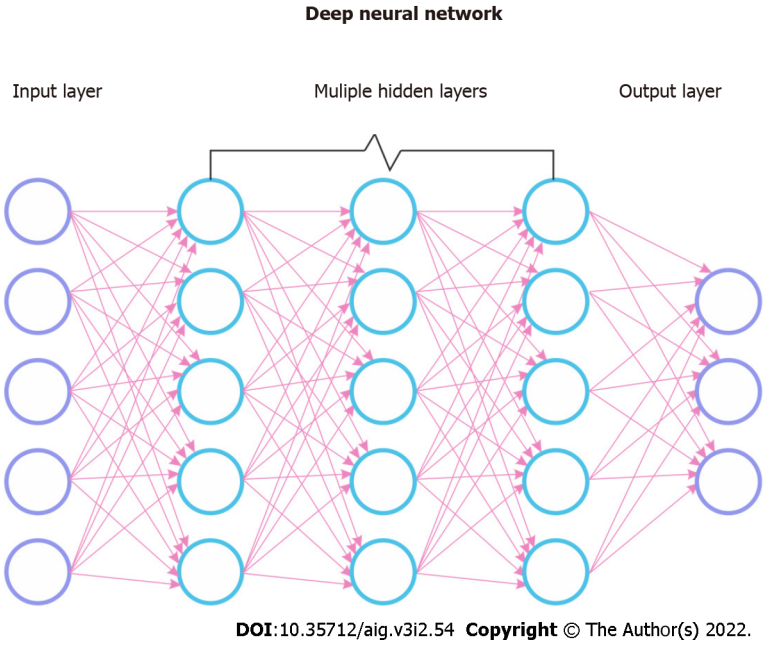
Figure 1 Potential benefit of artificial intelligence in pancreatic endoscopic ultrasonography. Current state of the pancreatic endoscopic ultrasonography (EUS) demonstrates that the procedure yields high resolution pancreatic imaging data, it is operator dependent, and allows acquisition of fine-needle aspiration (FNA) and fine-needle biopsy (FNB). Potentials with artificial intelligence (AI) implementation are utilizing of this higher resolution imaging data for training the algorithms with the readily available histologic ground truth from the FNA and FNB, as well as providing procedural assistance to address operator dependency.



DOI:10.35712/aig.v3i2.54

Figure 2 Overview of machine learning domains. Traditional machine learning algorithms rely on being trained by annotated and processed datasets to perform simpler tasks such as classification and regression. Deep learning algorithms are more autonomous, generally do not require annotation and processing of data for training and can perform more complicated tasks such as image detection and speech recognition. Reinforcement learning algorithms are self-teaching systems that can perform actions and learn by trial and error to achieve the best outcome; they perform most complex tasks such as game playing and learning to walk. CNN: Convolutional neural networks; MLP: Multi-layered perceptron; RNN: Recurrent neural networks.

Machine learning can perform two different types of tasks: Supervised and unsupervised. *Supervised* algorithms aim to reach a previously defined targeted outcome and are used for classification and prediction tasks. Labeled input data is presented to the algorithm and the model is trained with direct feedbacks to predict corresponding outputs. The spectrum of supervised approaches includes statistical methods such as logistic regression, linear regression, decision trees as well as support vector machines and random forest. *Unsupervised algorithms* do not have a predefined target and are used for clustering and dimensionality reduction. Unsupervised models are currently used for disease subtype and biomarker discovery studies[14,15]. Supervised learning has been more commonly used in EUS research; therefore, several important nuances will be summarized to better understand the presented literature. To train supervised learning algorithms, the dataset should be pre-annotated for the targeted



DOI:10.35712/aig.v3i2.54

Figure 3 Design overview of deep neural network model including input, output, and multiple hidden feature-detecting layers.

outcome, which may be a diagnosis, class, or feature. The algorithm aims to optimize its feature detection ability to match the presented inputs to this annotated targeted output, which is defined as "ground truth". This optimization, or training, task requires large datasets, therefore, learning algorithms are data hungry. However, such datasets are not commonly available, which necessitates data augmentation techniques be used to expand the dataset by inserting slightly changed copies of previously collected data or by creating new synthetic data with computerized approaches.

During training of the algorithm, available data is split into three sets: Training, validation, and test. Training and validation sets are used to develop and fine-tune the model, whereas the test set is used to assess the performance of the final model product. Of note, this validation is different from its conventional use in medicine and seeks to optimize parameters of the model during the training phase. Two of the most common validation approaches in medical AI research are cross-validation and hold-out validation. *Cross-validation* occurs when the dataset is randomly resampled and split repetitively – the number of repetitions is designated with *k*- into training and test sets. Each training and test set is then used to develop a new model, and *k* repetitions yield new *k* models. In contrast, *hold-out validation* is a constant single split of a training set and an independent test set to develop one final model which is simpler to perform but brings an increased risk of sampling error. Another important concept in machine learning is *overfitting*, which is defined as a falsely superior performance of the model caused by learning irrelevant features of the dataset or 'noise' as well as the intended signals. Therefore, a separate test set is important to accurately assess the model's performance.

There are several nuances in the performance assessment of a machine learning model. Sensitivity (*recall*), specificity, positive predictive value (*precision*), negative predictive value and area under the rule operator characteristic (AUC) curve are commonly used for assessing the performance of classification. The *area under the precision-recall curve* (*AUPRC*) is used instead of the AUC when observations are not equally distributed for two groups. The *Dice coefficient* (*F1 score*) is the harmonic mean of precision and recall. It is commonly used to assess the labeling performance of an image recognition model. In a model where a ground truth area X is labeled by an image recognition model as area Y, Dice coefficient equals the overlap of X and Y areas divided by the total of X and Y areas, multiplied by two. Another similar metric is the Jaccard index, or intersection over union (IoU), defined as the ratio of overlap and union of two areas: the algorithm labeled area and the ground truth area. Both Jaccard index and Dice coefficient's values range from 0 to 1 signifying 0% to 100% accuracy of labeling with 1 being the highest level of accuracy for both.

While AI has been utilized to investigate numerous gastrointestinal diseases, the study of pancreatic diseases using AI and EUS is limited[5]. In this review, we provide a targeted overview of AI with a summary of the current literature on the use of AI in EUS for the diagnosis of pancreatic diseases.

METHODS

A nonsystematic search of the current literature was performed for 2015 and 2021 in the MEDLINE,

PubMed, Google Scholar, Scopus, Web of Science and Embase databases with the following terms: Machine learning, deep learning, artificial intelligence, EUS, endosonography, endoscopic ultrasound, pancreas, pancreatic disease, pancreatitis, and pancreatic cancer. Review articles were manually screened for any additional studies of interest. Congress abstracts, reviews, correspondences, editorials, and book chapters were excluded. Two authors reviewed all the studies after the initial search and confirmed the appropriateness of each study for inclusion. Our literature search yielded fifteen studies with modern machine learning algorithms (Table 1). Of note, five of the fifteen studies were published in 2021 with only two prospective clinical trials from the same group.

APPLICATIONS OF AI IN PANCREATIC EUS

The application of AI was divided into sonographic image recognition, procedural assistance, and training. Endosonographic images contain cues that may not be recognizable by human visual perception. In this context, deep learning algorithms are promising tools to recognize the patterns from these cues. As such, several important diagnostic challenges in pancreatic diseases with EUS have been addressed, including the classification and risk stratification of pancreatic cysts and the diagnosis of autoimmune pancreatitis (AIP) and pancreatic ductal adenocarcinoma (PDAC).

Pancreatic cystic neoplasms

Pancreatic cysts are increasingly detected in patients undergoing abdominal cross-sectional imaging with up to 20% detection rate on MRI[6,16,17]. Since pancreatic cysts carry a risk of malignancy, this risk should be stratified to guide clinical management. However, in most cases, imaging results are not sufficient for the classification of pancreatic cysts, especially for small lesions[18]. Additionally, assessing the risk of malignant progression remains challenging with current imaging modalities, clinical criteria, cyst fluid analysis or their combinations[18,19]. In this context, ML may help classify pancreatic cysts.

Several studies have investigated the utility of EUS ML models in pancreatic cysts, focusing on malignancy risk assessment and classification. Two studies by Kuwahara *et al*[23] and Nguon *et al*[21] used still images of EUS examinations with data augmentation, while Springer *et al*[20] and Kurita *et al*[22] applied multimodality approaches that included cyst fluid analyses and clinical data[20-23].

The 2019 study by Kuwahara *et al*[23] assessed the accuracy of ML to predict malignant intraductal papillary mucinous neoplasms (IPMN). This single-center study included 50 IPMN patients who underwent surgical resection. Therefore, all diagnoses were made from histopathological examination of surgical specimens. A total of 3970 still images were collected from 50 EUS examinations, and the CNN was fed over 500000 images using data augmentation. Ten-fold cross-validation was performed for training. For each case, the output of the CNN model was given as a predictive continuous value ranging from 0 to 1 for benign and malignant assigned probabilities, respectively. When the final model's predictive values were compared with the surgical diagnoses, predictive values for the benign cases were significantly lower than values for the malignant cases (0.104 *vs* 0.808, respectively). The optimal cutoff for the predictive value was determined using the Youden Index. This cutoff value (0.49) generated an AUC of 98% for the diagnosis of malignancy. The accuracy of the final model (94%) was significantly higher than that of human preoperative diagnosis which incorporated contrast enhanced EUS examination findings of mural nodule size, diameter main pancreatic duct, cyst size, and growth rate (56%). Multivariate analysis showed that the AI predictive value was the only significant factor for diagnosing malignant IPMN. ML outperformed currently used criteria, including serum CA 19-9, presence of mural nodule, and type of IPMN. This study demonstrated the promise of EUS ML algorithms in predicting malignant IPMNs. However, further prospective studies with larger sample sizes that do not rely solely on internal validation are necessary.

Kurita *et al*[22] used a multimodality approach to differentiate benign from malignant cysts. This single center study used 85 patients with pancreatic cystic lesions and final diagnosis from surgical pathology or combination of cyst fluid analysis, radiology imaging, and clinical follow-up. The input data consisted of sex, cyst fluid protein markers, cytologic diagnosis and EUS imaging features of the cyst. A Multi-layered Perceptron was used as the ML model. The final model achieved 95.7% sensitivity, 91.9% specificity, and 0.97 AUC for classifying lesions as benign or malignant, which was the primary endpoint. The model showed 92.9% accuracy which was significantly higher than carcinoembryonic antigen (CEA) (71.8%) and cytology (85.9%) alone[22]. An external data set was not available to test the algorithm. In addition, it is unclear why the algorithm did not mention inclusion of known high-risk features including enhancing nodule, solid mass, and dilated main pancreatic duct.

Another large multicenter study used a ML based approach called CompCyst to guide the management of pancreatic cystic lesions and relied heavily on molecular analysis of cyst fluid in addition to clinical and radiologic imaging features. The study population consisted of 862 patients recruited from 16 centers who underwent surgical resection with final diagnosis based on histologic analysis. DNA from cyst fluid were extracted and evaluated for four types of molecular abnormalities including mutations, loss of heterozygosity, aneuploidy as well as protein markers CEA and vascular

Table 1 Summary of included machine learning studies on endoscopic ultrasonography in pancreatic disease

Field	Ref.	Study population used for training (*n*)	Task	Machine learning method	Performance (in test population if available)
Pancreatic Cysts	Kuwahara *et al*[23], 2019	Benign IPMN (27); Malignant IPMN (23)	Differentiate benign from malignant IPMN	Convolutional neural network	AUC = 0.98
	Springer *et al*[20], 2019	Mucinous cystic neoplasms (153); Serous Cystic Neoplasms (148); IPMN (447); Malignant cysts (114)	Guide clinical management by classify into three risk groups: No risk of malignancyLow risk of progression. High-risk of progression or malignant	Not available	First group: 100% specificity, 46% sensitivity. Second group: 54% specificity, 91% sensitivity. Third group: 30% specificity, 99% sensitivity.
	Kurita *et al* [22], 2019	Mucinous cystic neoplasms (23); Serous Cystic Neoplasms (15); IPMN (30); Other cyst types (17)	Differentiate benign from malignant cyst	Multi-layered perceptron	AUC = 0.96, sensitivity: 95%, specificity: 91.9%
	Nguon *et al* [21], 2021	Mucinous cystic neoplasms (59); Serous Cystic Neoplasms (49)	Differentiate mucinous cystic neoplasm and serous cystadenoma	Convolutional neural network	AUC = 0.88
Pancreatic Cancer	Saftouiu *et al*[27], 2008	PDAC (32); Normal pancreas (22); Chronic pancreatitis (11); Pancreatic neuroendocrine tumor (3)	Differentiate benign from malignant masses	Multi-layered perceptron	AUC = 0.96
	Saftoiu *et al* [28], 2012	PDAC (211); Chronic pancreatitis (47)	Differentiate cancer from benign masses	Multi-layered perceptron	AUC = 0.94
	Ozkan *et al* [30], 2016	PDAC (202); Normal pancreas (130)	Differentiate cancer from normal pancreas	Multi-layered perceptron	Accuracy: 87.5%, sensitivity: 83.3%, and specificity: 93.3%
	Udristou *et al*[31], 2021	PDAC (30); Chronic pancreatitis (20); Pancreatic neuroendocrine tumor (15)	Diagnose focal pancreatic mass	Convolutional neural network and long short-term memory	Mean AUC = 0.98 (Includes PDAC, CP and PNET)
	Tonozuka *et al*[32], 2021	PDAC (76); Chronic pancreatitis (34); Control (29)	Differentiate pancreatic cancer from chronic pancreatitis and normal pancreas	Convolutional neural network and pseudo-colored heatmap	AUC = 0.94
Autoimmune pancreatitis	Zhu *et al* [34], 2015	AIP (81); Chronic pancreatitis (100)	Differentiate AIP from chronic pancreatitis	Support Vector Machine	Accuracy: 89.3%, sensitivity: 84.1%, and specificity: 92.5%
	Marya *et al* [36], 2021	AIP (146); PDAC (292); Chronic pancreatitis (72); Normal pancreas (73)	Differentiate of AIP from PDAC	Convolutional neural network and pseudo-colored heatmap	AUC for AIP from all other = 0.92
Procedural assistance	Iwasa *et al* [38], 2021	Pancreatic mass (100)	Segmentation of pancreatic masses	Convolutional neural network	Intersection over unit = 0.77
	Zhang *et al* [40], 2020	EUS videos (339)	Recognition of stations, and segmentation of anatomical landmarks	Convolutional neural network	Accuracy for classification of stations (average) = 0.824, Dice coefficient for segmentation of pancreas (average) = 0.715

AUC: Area under the rule operator characteristic; AIP: Autoimmune pancreatitis; CP: Chronic pancreatitis; EUS: Endoscopic ultrasonography; IPMN: Intraductal papillary mucinous neoplasms; PDAC: Pancreatic ductal adenocarcinoma; PNET: Primitive neuroectodermal tumors.

endothelial growth factor-A (VEGF-A). Then the CompCyst test was used to classify cysts into one of the three following groups using a combination of molecular and imaging features. The first group was defined as cysts without any malignant potential which would not need surveillance. VHL and GNAS were used in this step and achieved 100% specificity and 46% sensitivity. The second group was cysts with small risk of malignant progression which would require surveillance. Multiple gene mutations and solid component in imaging was used in this step yielding 91% sensitivity and 54% specificity in the test cohort. The third group included cysts with high likelihood of malignant progression or malignancy which should be resected. VEGF-A protein expression was used in this step with 99% sensitivity and 30% specificity. The system was compared to standard of care and demonstrated significantly higher accuracy (69% *vs* 56%, respectively)[20]. This study used a separate validation set and a comprehensive model that incorporated clinical and radiologic findings, however, the wide-ranging molecular analysis is not readily available for routine clinical use.

A recent 2021 study focused on differentiating mucinous cystic neoplasms from serous cystadenomas using a total of 109 cases from two centers[21]. Final diagnoses were determined by endosonographers with over 5 years of experience. Additional cyst fluid or histopathologic examinations were available for only 44% of patients. A total of 221 still images were obtained followed by data augmentation, but the final number of input images was not provided in the study. The ResNet framework was used as the CNN model. Three hold-out validations were performed with 10 cases for testing, and the remaining cases used for training. The result of the study showed 82.75% accuracy and 0.88 AUC to correctly classify mucinous cystic neoplasms and serous cystadenomas from the still EUS images. A pseudo-colored decision map [gradient weighted class activation mapping (GradCAM)] was used to visualize the decision-making process. Presentation of the pseudo-colored decision map is an important asset because it highlights and color codes (red for higher impact and blue for lower impact) the areas in the image which affected the algorithm's final decision; therefore, this allows clinicians to better comprehend the decision-making process by the model. However, this study has several limitations. First, the most commonly encountered cyst, IPMN, was not included in the dataset that decreases the generalizability of the model. Second, ground truth was endosonographers' expert opinion and only 44% of patients had cyst fluid or histologic confirmation of diagnosis. Despite various limitations, the studies presented demonstrate the feasibility of image recognition ML models to perform classification tasks for pancreatic cysts and guide clinical management.

Pancreatic cancer

PDAC is currently the fourth leading cause of cancer-related mortality in Western countries and is predicted to become the second by 2030[24]. Most cases are diagnosed at later stages with 5-year survival rates less than 10%. A promising strategy is earlier diagnosis to combat this disease[25]. For this, EUS with FNA has superseded the cross-sectional imaging modalities such as CT and MRI, especially in the earlier diagnosis of PDAC[26]. However, EUS is operator dependent, and EUS diagnosis of PDAC is more challenging in patients with baseline abnormal pancreatic imaging (*e.g.*, chronic pancreatitis) who also carry a higher risk. Within this context, ML has been used to improve the diagnostic performance of EUS for pancreatic masses. Four studies used histologically confirmed PDAC cases with normal pancreas as control. Additional control groups were used in different studies to reflect clinical scenarios including chronic pancreatitis and neuroendocrine tumors. EUS images served as inputs for the algorithms. Additional EUS diagnostic technology, such as elastography, digital characteristics, contrast-enhancement, and Doppler imaging were also used. Regarding ML methods, Support-Vector-Machines were used in earlier studies to select the best combination of digital imaging features. In later studies the preferred methods were neural networks with different complexity levels depending on the year of the study. Although the models and populations varied, all studies achieved over 80% specificity and 0.94 AUC, demonstrating the feasibility of ML in this area.

In an early 2008 study by Saftoiu *et al*[27], ML for EUS elastography images was evaluated to discriminate pancreatic tumors from 'pseudotumoral' chronic pancreatitis. The prospective study enrolled 68 patients including PDAC, pancreatic neuroendocrine tumor, chronic pancreatitis, and normal pancreas. Final diagnoses were confirmed with additional pathology, imaging findings, and 6-mo follow-up of patients. From each patient, EUS elastography images were converted to vector data. As the sample size was small, 10-fold cross-validation was performed. The vector data was then analyzed with simple three and four layered ANNs. This ML algorithm yielded an AUC of 0.93 to classify malignant tumors from normal and pseudotumoral pancreatitis. This study was followed by a larger prospective blinded study in 2012 with 258 patients enrolled from 13 European centers. The population consisted of 211 PDAC confirmed by pathology diagnosis and 47 chronic pancreatitis patients diagnosed by clinical, imaging and EUS criteria (at least four of the following: hyperechoic foci, hyperechoic strands, lobularity, calcifications, hyperechoic duct wall, dilated main pancreatic duct, irregular main pancreatic duct, dilated side branches, and cysts). EUS elastography images of the regions of interests were converted to vector data and then analyzed with similar ANNs. One hundred training iterations were performed with the model to increase the statistical power of the results. The mean performance of one hundred models to correctly classify PDAC from chronic pancreatitis showed 0.94 (0.91-0.97) AUC with 85.6% sensitivity and 82.9% specificity compared with 0.85 AUC for hue histogram analysis[28]. These two studies present an excellent example for the roadmap of ML research with an initial proof-of-concept study followed by a larger prospective study. Of note, less complex neural networks were used with fewer layers. Multi-layered Perceptron only accepts numeric data as the input unlike newer CNN algorithms that can directly process the image itself. Therefore, the performance of ML in these studies can be improved.

An early study in 2013 used analysis of digital image characteristics as input to the ML model[29]. The study population consisted of 262 PDAC patients diagnosed by cytology with 126 chronic pancreatitis controls diagnosed by standard EUS criteria and over 2-year follow up. Regions of interests were manually selected by blinded endosonographers. Then 105 digital imaging characteristics of these images were extracted with dedicated software. The final combination of 16 characteristics yielded a strong discriminative performance with 94.2% accuracy, 96% sensitivity and 93% specificity.

Another older study evaluated the use of ML to classify PDAC from normal pancreas[30]. This retrospective study in 2016 included 202 PDAC patients and 130 patients with normal pancreas as controls. The regions of interests from EUS images were annotated by endosonographers. Then digital characteristics of the images (wavelet decomposition energy, boundary fractal, gray level cooccurrence matrix, standard statistical) were extracted. Among 112 digital characteristics, 20 were identified as more effective for classification, and therefore served as the input for the ML algorithm. A three-layered Multi-layered Perceptron model was used as the neural network, which is a less-complex approach accepting numerical data such as the digital characteristics of EUS images and does not require extra image processing. As such, because the images themselves are not being used, important information may not be included in the model. The final model yielded 83% sensitivity, 93% specificity and 87% accuracy for differentiating PDAC from normal pancreas. This model also only compared PDAC images to normal pancreatic tissue and not to other commonly encountered differential diagnoses such as chronic pancreatitis, which limits its adoptability to clinical use.

A recent study in 2021 evaluated the performance of ML to classify focal solid lesions. The study population consisted of 30 patients with PDAC, 20 patients with pseudo tumors in chronic pancreatitis, and 15 patients with pancreatic neuroendocrine tumors[31]. The final diagnoses were confirmed with histologic evaluations of fine-needle specimens and clinical follow-ups. From each EUS examination, 5 sets of images were extracted including grayscale images, color Doppler, contrast-enhanced imaging, and elastography. A total of 1300 collected images was increased to 3360 with data augmentation. Regarding the ML method, a CNN algorithm was combined with a Long Short-Term Memory model. Long Short-Term Memory model is a supervised ML model that has additional feedback learning functions and allows the use of sequential pre- and post-contrast appearance from the same EUS images. Cross-validation was performed for each dataset with 80% of images used as training and 20% as test sets. The final combined model's overall specificity was 96.4%, and sensitivity was 98.6% for classifying the pancreatic masses. For PDAC cases, the algorithm yielded 96.7% specificity, 98.1% sensitivity, 97.6% accuracy, and 0.97 AUC. When compared to previous studies, Udristoiu *et al*[31] used a more complex, combined ML approach with CNN and Long Short-Term Memory allowing inclusion of temporal data with contrast-enhanced imaging.

Tonozuka *et al*[32] also evaluated their own ML algorithm for its performance in classifying pancreatic masses. The 139 total patients included 76 with PDAC, 34 with chronic pancreatitis and 29 normal controls. PDAC was diagnosed using histology from EUS-fine needle biopsy or surgery, and chronic pancreatitis was diagnosed using the Rosemont criteria. All patients were followed for over 6 mo. Ten still images of lesions were chosen from each EUS examination, and the input dataset was increased to over 80000 after data augmentation. From 1390 still images, 920 were used for training and cross-validation, while the remaining 470 images were used for testing. A CNN algorithm with seven layers was used. In addition to the CNN model, a pseudo-colored feature mapping was used to highlight the areas in the image with greater impact on the final model, which makes the decision-making process more comprehensible to the endosonographer. In the test dataset, the model yielded 84.1% specificity, 92.4% sensitivity and 0.94 AUC.

Autoimmune pancreatitis

AIP is an increasingly recognized entity that may be challenging to diagnose. Accurate diagnosis is particularly important as the differential often includes PDAC with its different prognostic and management implications. Many diagnostic algorithms have been developed that include clinical, serologic, imaging, and histopathologic criteria, but their performance remains limited. While EUS with biopsy is the most effective diagnostic tool, its diagnostic yield also is suboptimal[33]. Image processing may enhance our ability to diagnose AIP by extracting data and learning from the cues in sonographic images. Two studies have studied the utility of ML in differentiating AIP from other diagnoses, including chronic pancreatitis and PDAC. The studies by Zhu *et al*[34] and Marya *et al*[35] used different ML approaches, but both achieved over 80% sensitivity and specificity for diagnosing AIP only from EUS images[34,35].

The earlier 2015 retrospective study by Zhu *et al*[34] studied a ML algorithm to differentiate AIP from chronic pancreatitis using an EUS image dataset of 81 AIP and 100 chronic pancreatitis cases. AIP diagnoses were based on HISORt criteria. Chronic pancreatitis was diagnosed by standard EUS criteria. Experienced endosonographers selected regions of interest in EUS images, and 115 digital parameters were extracted from each image. Then, a supervised Support Vector Machine algorithm was used to select the best combination of these digital parameters for discriminating AIP from chronic pancreatitis. The final combination of digital parameters yielded 90.6% accuracy, 84.1% sensitivity and 94.0% specificity.

A recent study examined the additive performance of ML with EUS to distinguish AIP from PDAC as well as chronic pancreatitis and normal pancreas. The study included 583 patients (146 AIP, 292 PDAC, 72 chronic pancreatitis, and 74 normal) with all available videos and still images of the pancreatic and peripancreatic regions included in the analysis regardless of whether they included regions of interest [36]. A total of 1174461 still images were extracted from the images and videos. Since all portions of EUS videos were included, there was a risk of oversimplification of diagnosis from certain aspects of the examination, such as presence of metastasis, which were removed from the dataset. The classification

was performed with two datasets: the first one included still images obtained from both EUS videos and captured images, while the second dataset only included EUS videos. The CNN algorithm was trained for both datasets. Pseudo-colored feature mapping was also used to visualize decision making. For comparison, seven independent EUS experts evaluated each case using videos. In the final analysis, ML showed 87% specificity, 90% sensitivity and 0.9 AUC for distinguishing AIP from PDAC in the image-only dataset. In the video-only dataset, the metrics were 90%, 93% and 0.96 for specificity, sensitivity, and AUC, respectively. The ML model was superior to expert endosonographers, who had 82.4% specificity and 53.8% sensitivity in differentiating AIP from PDAC. ML also had high sensitivity (99%) and specificity (98%) for distinguishing AIP from normal pancreas. It had inferior performance in separating AIP from chronic pancreatitis (94% sensitivity, 71% specificity, 0.89 AUC). The heatmap analysis yielded interesting results, which may help guide endosonographers, showing that visualizing a hyperechoic plane between the parenchyma and duct or vessel was highly predictive of AIP while post acoustic enhancement deep to a dilated pancreatic duct or vessel was consistent with PDAC. Regarding AI technology, these two studies differ with respect to their approach of utilizing ML with EUS data. Zhu *et al*[34] used an older ML algorithm, support vector machine, which is a supervised algorithm that classifies two numeric data points. As such, EUS images are converted into numerical data by extracting digital parametric features, and then the ML model is trained with these features. On the other hand, Marya *et al*[35] used a CNN algorithm, ResNet, with 50 layers that can work directly on the EUS images itself.

Procedural assistance and training

EUS is the leading modality for assessing and obtaining tissue from the pancreas with approximately 90% specificity and sensitivity for solid masses[36]. However, interobserver reliability remains an issue in EUS as accuracy relies on the endosonographers' skills and experience and carries the risk of false-negative results. Pancreatic EUS also has a steep learning curve. ML approaches have been developed to potentially augment the diagnostic performance of EUS and biopsy as well as aid in training.

Iwasa *et al*[38] tested ML to augment contrast enhanced EUS by dividing the sonographic image into regions with similar appearance and then differentiating regions of interest, also called automatic segmentation. For this study, videos from 100 contrast enhanced EUS examinations of solid pancreatic masses with histologic diagnosis were used. Each video was transformed into 900 still images as input for a U-Net CNN algorithm. The borders of the lesions were manually annotated by two endosonographers and served as the ground truth. IoU was used as the performance output of the algorithm with median IoU for all cases being 0.77, which is greater than the acceptable 0.5 threshold value[37]. The EUS videos were also classified into different categories to understand the effect of respiratory movements and visibility of boundaries of the lesions by the endosonographers. IoU significantly improved to 0.91 in cases with the most visible boundaries and decreased to 0.13 for cases with the least visible boundaries[39]. On the other hand, respiratory movements did not change the performance of the algorithm. This proof-of-concept study suggests that ML can provide real-time assistance in the detection of pancreatic lesions. The classification of exams with respect to the ease of detecting the border of lesions is an important aspect of this study because it demonstrated that ML can also be affected by the quality of the EUS examination and the sonographic characteristics of the lesion, reflected in this case by how well the border was visible.

A case report suggested that a ML model may help target areas to biopsy within pancreatic masses that have the highest diagnostic yield by avoiding areas of necrosis. A CNN algorithm was used to label and highlight the more cellular region in a 6.5 cm solid pancreatic mass, which was predicted to have the highest probability of yielding a diagnosis by discriminating it from neighboring necrotic or inflammatory regions. EUS-fine needle aspiration was performed and yielded a positive diagnosis for PDAC. The technical details, training dataset and methods, validation and model characteristics were not presented in the report[39]. This is a novel idea that may provide valuable intra-procedural assistance, however, needs further evaluation.

ML may aid EUS training by guiding the steps of routine diagnostic EUS evaluation of the pancreas. A novel AI system aimed to assist recognition of fundamental stations and identification of pancreatic and vascular anatomical landmarks. This was performed in four steps: Identifying images, filtering suitable images, recognizing pancreas stations, and segmenting anatomical landmarks and monitoring for loss of visualization of the pancreas. Two expert endosonographers decided on the criteria for suitable images and annotated video clips that served as ground truth. A ResNet model was used as the CNN algorithm. A separate set of prospective EUS examinations were used as a test set. Three different endosonographers classified each image for comparison with the AI model. The final model was tested using an external test set and demonstrated an accuracy of 82.4% to identify six anatomical stations (abdominal aorta, pancreatic body, pancreatic tail, confluence, pancreatic head from stomach, or pancreatic head from descending duodenum), and a Dice of 0.715 to label pancreas and vessels. Comparison of the AI model with the three expert endosonographers yielded strong interobserver agreement with kappa values of 0.846, 0.853 and 0.826[40]. The results of this study demonstrated that a ML model may aid in recognizing stations and anatomic landmarks in sonographic images. This has the potential to assist procedural navigation during EUS examination and improve cognitive aspects of EUS skills. However, the impact of such real-time procedural assistance on the endosonographer's

performance was not assessed in this study and warrants further evaluation.

CONCLUSION

In this review, we summarize the current literature regarding the use of ML in EUS for diagnosing pancreatic diseases. Our review defined two main areas for AI in the field: visual recognition-classification and procedural assistance and training. AI has been more utilized in transabdominal ultrasonography for detecting liver fibrosis and in CT scans for lesion classification, which have been extensively reviewed elsewhere[41-45]. ML appears to have great potential in assisting EUS examination of the pancreas as sonographic imaging contains vital visual information that the human eye cannot distinguish. The diagnostic accuracy of EUS imaging is highly operator dependent and requires both technical and cognitive skills. Acquisition of these skills currently requires dedicated training with proctorship and procedural experience, which remains limited, apart from dedicated advanced endoscopy fellowship programs. These issues in training limit the widespread adoption of EUS, which is the leading tool for diagnosing pancreatic disorders, including PDAC. AI may assist in the development of cognitive skills and augmentation of procedural efficiency in relatively less experienced endosonographers.

Further opportunities should be explored with AI and pancreatic EUS. However, several limitations exist in the field. First, the number of EUS procedures and the prevalence of pancreatic diseases are lower, which makes it more difficult to train data-hungry machine learning algorithms. Second, annotation of EUS data is more challenging compared to other imaging modalities as the number of experts endosonographers is relatively limited. Third, EUS examinations with histopathologic or cytologic diagnosis is harder to obtain for certain pancreatic diseases and have issues with sensitivity, which further limits the number of studies for AI training. However, these limitations may be overcome with multi-center collaborations and prospective data collection, which will hopefully lead to improved image recognition, procedural assistance, and training for pancreatic EUS.

FOOTNOTES

Author contributions: Simsek C collected data and wrote the paper; Lee L carried out data collection; both authors read, edited, and approved the final manuscript.

Conflict-of-interest statement: Cem Simsek is co-founder of Algomedicus Inc.

Country/Territory of origin: United States

ORCID number: Cem Simsek 0000-0002-7037-5233; Linda S Lee 0000-0001-9921-8920.

S-Editor: Liu JH
L-Editor: A
P-Editor: Liu JH

REFERENCES

1 **Miller KD**, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, Jemal A, Kramer JL, Siegel RL. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin* 2019; **69**: 363-385 [PMID: 31184787 DOI: 10.3322/caac.21565]

2 **Bray F**, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; **68**: 394-424 [PMID: 30207593 DOI: 10.3322/caac.21492]

3 **Rahib L**, Wehner MR, Matrisian LM, Nead KT. Estimated Projection of US Cancer Incidence and Death to 2040. *JAMA Netw Open* 2021; **4**: e214708 [PMID: 33825840 DOI: 10.1001/jamanetworkopen.2021.4708]

4 **Xiao AY**, Tan ML, Wu LM, Asrani VM, Windsor JA, Yadav D, Petrov MS. Global incidence and mortality of pancreatic diseases: a systematic review, meta-analysis, and meta-regression of population-based cohort studies. *Lancet Gastroenterol Hepatol* 2016; **1**: 45-55 [PMID: 28404111 DOI: 10.1016/S2468-1253(16)30004-8]

5 **Kichler A**, Jang S. Chronic Pancreatitis: Epidemiology, Diagnosis, and Management Updates. *Drugs* 2020; **80**: 1155-1168

[PMID: 32647920 DOI: 10.1007/s40265-020-01360-6]
6 **Megibow AJ**, Baker ME, Gore RM, Taylor A. The incidental pancreatic cyst. *Radiol Clin North Am* 2011; **49**: 349-359 [PMID: 21333781 DOI: 10.1016/j.rcl.2010.10.008]
7 **Zhang B**, Zhu F, Li P, Yu S, Zhao Y, Li M. Endoscopic ultrasound elastography in the diagnosis of pancreatic masses: A meta-analysis. *Pancreatology* 2018; **18**: 833-840 [PMID: 30093353 DOI: 10.1016/j.pan.2018.07.008]
8 **Russell S**, Norvig P. Artificial intelligence: a modern approach. 2002
9 **Yu KH**, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng* 2018; **2**: 719-731 [PMID: 31015651 DOI: 10.1038/s41551-018-0305-z]
10 **Penny W**, Frost D. Neural networks in clinical medicine. *Med Decis Making* 1996; **16**: 386-398 [PMID: 8912300 DOI: 10.1177/0272989X9601600409]
11 **Yamashita R**, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018; **9**: 611-629 [PMID: 29934920 DOI: 10.1007/s13244-018-0639-9]
12 **Zhang**, HM. and B. Dong, A Review on Deep Learning in Medical Image Reconstruction. *Journal of the Operations Research Society of China* 2020; **8**: 311-340 [DOI: 10.1007/s40305-019-00287-4]
13 **Van Houdt G**, Mosquera C, Nápoles G. A review on the long short-term memory model. *Artif Intell Rev* 2020; **53**: 5929-5955 [DOI: 10.1007/s10462-020-09838-1]
14 **Rintala TJ**, Federico A, Latonen L, Greco D, Fortino V. A systematic comparison of data- and knowledge-driven approaches to disease subtype discovery. *Brief Bioinform* 2021; **22** [PMID: 34396389 DOI: 10.1093/bib/bbab314]
15 **Sinkala M**, Mulder N, Martin D. Machine Learning and Network Analyses Reveal Disease Subtypes of Pancreatic Cancer and their Molecular Characteristics. *Sci Rep* 2020; **10**: 1212 [PMID: 31988390 DOI: 10.1038/s41598-020-58290-2]
16 **Zerboni G**, Signoretti M, Crippa S, Falconi M, Arcidiacono PG, Capurso G. Systematic review and meta-analysis: Prevalence of incidentally detected pancreatic cystic lesions in asymptomatic individuals. *Pancreatology* 2019; **19**: 2-9 [PMID: 30503370 DOI: 10.1016/j.pan.2018.11.014]
17 **Zaheer A**, Pokharel SS, Wolfgang C, Fishman EK, Horton KM. Incidentally detected cystic lesions of the pancreas on CT: review of literature and management suggestions. *Abdom Imaging* 2013; **38**: 331-341 [PMID: 22534872 DOI: 10.1007/s00261-012-9898-y]
18 **Pinho DF**, Rofsky NM, Pedrosa I. Incidental pancreatic cysts: role of magnetic resonance imaging. *Top Magn Reson Imaging* 2014; **23**: 117-128 [PMID: 24690615 DOI: 10.1097/RMR.0000000000000018]
19 **Wu J**, Wang Y, Li Z, Miao H. Accuracy of Fukuoka and American Gastroenterological Association Guidelines for Predicting Advanced Neoplasia in Pancreatic Cyst Neoplasm: A Meta-Analysis. *Ann Surg Oncol* 2019; **26**: 4522-4536 [PMID: 31617119 DOI: 10.1245/s10434-019-07921-8]
20 **Springer S**, Masica DL, Dal Molin M, Douville C, Thoburn CJ, Afsari B, Li L, Cohen JD, Thompson E, Allen PJ, Klimstra DS, Schattner MA, Schmidt CM, Yip-Schneider M, Simpson RE, Fernandez-Del Castillo C, Mino-Kenudson M, Brugge W, Brand RE, Singhi AD, Scarpa A, Lawlor R, Salvia R, Zamboni G, Hong SM, Hwang DW, Jang JY, Kwon W, Swan N, Geoghegan J, Falconi M, Crippa S, Doglioni C, Paulino J, Schulick RD, Edil BH, Park W, Yachida S, Hijioka S, van Hooft J, He J, Weiss MJ, Burkhart R, Makary M, Canto MI, Goggins MG, Ptak J, Dobbyn L, Schaefer J, Sillman N, Popoli M, Klein AP, Tomasetti C, Karchin R, Papadopoulos N, Kinzler KW, Vogelstein B, Wolfgang CL, Hruban RH, Lennon AM. A multimodality test to guide the management of patients with a pancreatic cyst. *Sci Transl Med* 2019; **11** [PMID: 31316009 DOI: 10.1126/scitranslmed.aav4772]
21 **Nguon LS**, Seo K, Lim JH, Song TJ, Cho SH, Park JS, Park S. Deep Learning-Based Differentiation between Mucinous Cystic Neoplasm and Serous Cystic Neoplasm in the Pancreas Using Endoscopic Ultrasonography. *Diagnostics (Basel)* 2021; **11** [PMID: 34201066 DOI: 10.3390/diagnostics11061052]
22 **Kurita Y**, Kuwahara T, Hara K, Mizuno N, Okuno N, Matsumoto S, Obata M, Koda H, Tajika M, Shimizu Y, Nakajima A, Kubota K, Niwa Y. Diagnostic ability of artificial intelligence using deep learning analysis of cyst fluid in differentiating malignant from benign pancreatic cystic lesions. *Sci Rep* 2019; **9**: 6893 [PMID: 31053726 DOI: 10.1038/s41598-019-43314-3]
23 **Kuwahara T**, Hara K, Mizuno N, Okuno N, Matsumoto S, Obata M, Kurita Y, Koda H, Toriyama K, Onishi S, Ishihara M, Tanaka T, Tajika M, Niwa Y. Usefulness of Deep Learning Analysis for the Diagnosis of Malignancy in Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Clin Transl Gastroenterol* 2019; **10**: 1-8 [PMID: 31117111 DOI: 10.14309/ctg.0000000000000045]
24 **Rahib L**, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res* 2014; **74**: 2913-2921 [PMID: 24840647 DOI: 10.1158/0008-5472.CAN-14-0155]
25 **Howlader N**, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary Cl, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA. SEER cancer statistics review, 1975–2012. National Cancer Institute, 2015. Available from: https://seer.cancer.gov/archive/csr/1975_2012/
26 **Kitano M**, Yoshida T, Itonaga M, Tamura T, Hatamaru K, Yamashita Y. Impact of endoscopic ultrasonography on diagnosis of pancreatic cancer. *J Gastroenterol* 2019; **54**: 19-32 [PMID: 30406288 DOI: 10.1007/s00535-018-1519-2]
27 **Săftoiu A**, Vilmann P, Gorunescu F, Gheonea DI, Gorunescu M, Ciurea T, Popescu GL, Iordache A, Hassan H, Iordache S. Neural network analysis of dynamic sequences of EUS elastography used for the differential diagnosis of chronic pancreatitis and pancreatic cancer. *Gastrointest Endosc* 2008; **68**: 1086-1094 [PMID: 18656186 DOI: 10.1016/j.gie.2008.04.031]
28 **Săftoiu A**, Vilmann P, Gorunescu F, Janssen J, Hocke M, Larsen M, Iglesias-Garcia J, Arcidiacono P, Will U, Giovannini M, Dietrich CF, Havre R, Gheorghe C, McKay C, Gheonea DI, Ciurea T; European EUS Elastography Multicentric Study Group. Efficacy of an artificial neural network-based approach to endoscopic ultrasound elastography in diagnosis of focal pancreatic masses. *Clin Gastroenterol Hepatol* 2012; **10**: 84-90.e1 [PMID: 21963957 DOI: 10.1016/j.cgh.2011.09.014]
29 **Zhu M**, Xu C, Yu J, Wu Y, Li C, Zhang M, Jin Z, Li Z. Differentiation of pancreatic cancer and chronic pancreatitis using computer-aided diagnosis of endoscopic ultrasound (EUS) images: a diagnostic test. *PLoS One* 2013; **8**: e63820 [PMID: 23704940 DOI: 10.1371/journal.pone.0063820]

30 **Ozkan M**, Cakiroglu M, Kocaman O, Kurt M, Yilmaz B, Can G, Korkmaz U, Dandil E, Eksi Z. Age-based computer-aided diagnosis approach for pancreatic cancer on endoscopic ultrasound images. *Endosc Ultrasound* 2016; **5**: 101-107 [PMID: 27080608 DOI: 10.4103/2303-9027.180473]

31 **Udriștoiu AL**, Cazacu IM, Gruionu LG, Gruionu G, Iacob AV, Burtea DE, Ungureanu BS, Costache MI, Constantin A, Popescu CF, Udriștoiu Ș, Săftoiu A. Real-time computer-aided diagnosis of focal pancreatic masses from endoscopic ultrasound imaging based on a hybrid convolutional and long short-term memory neural network model. *PLoS One* 2021; **16**: e0251701 [PMID: 34181680 DOI: 10.1371/journal.pone.0251701]

32 **Tonozuka R**, Itoi T, Nagata N, Kojima H, Sofuni A, Tsuchiya T, Ishii K, Tanaka R, Nagakawa Y, Mukai S. Deep learning analysis for the detection of pancreatic cancer on endosonographic images: a pilot study. *J Hepatobiliary Pancreat Sci* 2021; **28**: 95-104 [PMID: 32910528 DOI: 10.1002/jhbp.825]

33 **Facciorusso A**, Barresi L, Cannizzaro R, Antonini F, Triantafyllou K, Tziatzios G, Muscatiello N, Hart PA, Wani S. Diagnostic yield of endoscopic ultrasound-guided tissue acquisition in autoimmune pancreatitis: a systematic review and meta-analysis. *Endosc Int Open* 2021; **9**: E66-E75 [PMID: 33403238 DOI: 10.1055/a-1293-7279]

34 **Zhu J**, Wang L, Chu Y, Hou X, Xing L, Kong F, Zhou Y, Wang Y, Jin Z, Li Z. A new descriptor for computer-aided diagnosis of EUS imaging to distinguish autoimmune pancreatitis from chronic pancreatitis. *Gastrointest Endosc* 2015; **82**: 831-836.e1 [PMID: 25952089 DOI: 10.1016/j.gie.2015.02.043]

35 **Marya NB**, Powers PD, Chari ST, Gleeson FC, Leggett CL, Abu Dayyeh BK, Chandrasekhara V, Iyer PG, Majumder S, Pearson RK, Petersen BT, Rajan E, Sawas T, Storm AC, Vege SS, Chen S, Long Z, Hough DM, Mara K, Levy MJ. Utilisation of artificial intelligence for the development of an EUS-convolutional neural network model trained to enhance the diagnosis of autoimmune pancreatitis. *Gut* 2021; **70**: 1335-1344 [PMID: 33028668 DOI: 10.1136/gutjnl-2020-322821]

36 **Facciorusso A**, Wani S, Triantafyllou K, Tziatzios G, Cannizzaro R, Muscatiello N, Singh S. Comparative accuracy of needle sizes and designs for EUS tissue sampling of solid pancreatic masses: a network meta-analysis. *Gastrointest Endosc* 2019; **90**: 893-903.e7 [PMID: 31310744 DOI: 10.1016/j.gie.2019.07.009]

37 **Rezatofighi H**, Tsoi N, Gwak JY, Sadeghian A, Reid I, Savaresel S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019

38 **Iwasa Y**, Iwashita T, Takeuchi Y, Ichikawa H, Mita N, Uemura S, Shimizu M, Kuo YT, Wang HP, Hara T. Automatic Segmentation of Pancreatic Tumors Using Deep Learning on a Video Image of Contrast-Enhanced Endoscopic Ultrasound. *J Clin Med* 2021; **10** [PMID: 34441883 DOI: 10.3390/jcm10163589]

39 **Yu X**, Zhang Z, Zhu N, Tang A, Hu S, Wang X, Tian L. Application of a novel artificial intelligence system in guiding the targeted puncture of a pancreatic mass. *Endoscopy* 2021 [PMID: 34624923 DOI: 10.1055/a-1625-3396]

40 **Zhang J**, Zhu L, Yao L, Ding X, Chen D, Wu H, Lu Z, Zhou W, Zhang L, An P, Xu B, Tan W, Hu S, Cheng F, Yu H. Deep learning-based pancreas segmentation and station recognition system in EUS: development and validation of a useful training tool (with video). *Gastrointest Endosc* 2020; **92**: 874-885.e3 [PMID: 32387499 DOI: 10.1016/j.gie.2020.04.071]

41 **Chen Y**, Luo Y, Huang W, Hu D, Zheng RQ, Cong SZ, Meng FK, Yang H, Lin HJ, Sun Y, Wang XY, Wu T, Ren J, Pei SF, Zheng Y, He Y, Hu Y, Yang N, Yan H. Machine-learning-based classification of real-time tissue elastography for hepatic fibrosis in patients with chronic hepatitis B. *Comput Biol Med* 2017; **89**: 18-23 [PMID: 28779596 DOI: 10.1016/j.compbiomed.2017.07.012]

42 **Yang CM**, Shu J. Cholangiocarcinoma Evaluation *via* Imaging and Artificial Intelligence. *Oncology* 2021; **99**: 72-83 [PMID: 33147583 DOI: 10.1159/000507449]

43 **Gatos I**, Tsantis S, Spiliopoulos S, Karnabatidis D, Theotokas I, Zoumpoulis P, Loupas T, Hazle JD, Kagadis GC. A Machine-Learning Algorithm Toward Color Analysis for Chronic Liver Disease Classification, Employing Ultrasound Shear Wave Elastography. *Ultrasound Med Biol* 2017; **43**: 1797-1810 [PMID: 28634041 DOI: 10.1016/j.ultrasmedbio.2017.05.002]

44 **Lee JH**, Joo I, Kang TW, Paik YH, Sinn DH, Ha SY, Kim K, Choi C, Lee G, Yi J, Bang WC. Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network. *Eur Radiol* 2020; **30**: 1264-1273 [PMID: 31478087 DOI: 10.1007/s00330-019-06407-1]

45 **Kuwahara T**, Hara K, Mizuno N, Haba S, Okuno N, Koda H, Miyano A, Fumihara D. Current status of artificial intelligence analysis for endoscopic ultrasonography. *Dig Endosc* 2021; **33**: 298-305 [PMID: 33098123 DOI: 10.1111/den.13880]

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 April 28; 3(2): 66-79

DOI: 10.35712/aig.v3.i2.66

ISSN 2644-3236 (online)

MINIREVIEWS

Artificial intelligence in critically ill diabetic patients: current status and future prospects

Deven Juneja, Anish Gupta, Omender Singh

Specialty type: Critical care medicine

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): C, C
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Li C, China; Villela-Nogueira CA, Brazil

Received: February 16, 2022
Peer-review started: February 16, 2022
First decision: April 17, 2022
Revised: April 21, 2022
Accepted: April 28, 2022
Article in press: April 28, 2022
Published online: April 28, 2022

Deven Juneja, Anish Gupta, Omender Singh, Institute of Critical Care Medicine, Max Super Speciality Hospital, Saket, New Delhi 110092, India

Corresponding author: Deven Juneja, FCCP, MBBS, Director, Institute of Critical Care Medicine, Max Super Speciality Hospital, Saket, 1, Press Enclave Road, Saket, New Delhi 110092, India. devenjuneja@gmail.com

Abstract

Recent years have witnessed increasing numbers of artificial intelligence (AI) based applications and devices being tested and approved for medical care. Diabetes is arguably the most common chronic disorder worldwide and AI is now being used for making an early diagnosis, to predict and diagnose early complications, increase adherence to therapy, and even motivate patients to manage diabetes and maintain glycemic control. However, these AI applications have largely been tested in non-critically ill patients and aid in managing chronic problems. Intensive care units (ICUs) have a dynamic environment generating huge data, which AI can extract and organize simultaneously, thus analysing many variables for diagnostic and/or therapeutic purposes in order to predict outcomes of interest. Even non-diabetic ICU patients are at risk of developing hypo or hyperglycemia, complicating their ICU course and affecting outcomes. In addition, to maintain glycemic control frequent blood sampling and insulin dose adjustments are required, increasing nursing workload and chances of error. AI has the potential to improve glycemic control while reducing the nursing workload and errors. Continuous glucose monitoring (CGM) devices, which are Food and Drug Administration (FDA) approved for use in non-critically ill patients, are now being recommended for use in specific ICU populations with increased accuracy. AI based devices including artificial pancreas and CGM regulated insulin infusion system have shown promise as comprehensive glycemic control solutions in critically ill patients. Even though many of these AI applications have shown potential, these devices need to be tested in larger number of ICU patients, have wider availability, show favorable cost-benefit ratio and be amenable for easy integration into the existing healthcare systems, before they become acceptable to ICU physicians for routine use.

Key Words: Artificial intelligence; Blood glucose; Critical care; Diabetes mellitus; Intensive care unit; Machine learning

Core Tip: Increasing number of applications and devices based on artificial intelligence are being tested and approved for medical care. These devices have the potential to change the way we presently manage chronic diseases like diabetes. Moreover, their application in data rich and dynamic intensive care unit environment may have great implications in detecting hypo or hyperglycemia and reducing glycemic variability, while improving safety and accuracy and reducing nursing workload. Devices like artificial pancreas and continuous glucose monitoring regulated insulin infusion systems have shown promise as comprehensive glucose control solutions and may change the future of care for critically ill diabetic patients.

Citation: Juneja D, Gupta A, Singh O. Artificial intelligence in critically ill diabetic patients: current status and future prospects. *Artif Intell Gastroenterol* 2022; 3(2): 66-79
URL: https://www.wjgnet.com/2644-3236/full/v3/i2/66.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i2.66

INTRODUCTION

As per the International Diabetes Federation 2021 estimates, about 537 million people are living with diabetes signifying a 10% prevalence rate worldwide with an estimated 6.7 million deaths in 2021. This number will rise exponentially in the coming years which will place a heavy burden on the already stressed healthcare system[1]. These patients are at increased risk of developing complications like sepsis, diabetes keto-acidosis and other complications necessitating intensive care unit (ICU) admission. In addition, critically ill diabetic patients are at an increased risk of developing nosocomial infections, having a longer ICU stay and increased ICU mortality[2-4].

All components of diabetes care including prevention and management of hyperglycemia and hypoglycemia, are essential to improve outcomes. In critically ill patients, these complications may be multifactorial and may also occur in non-diabetic patients, complicating their disease course. In addition to hyper- and hypoglycemia, glycemic variability (GV) and time in target range (TITR) are recently recognized components of dysglycemia which may affect patient outcomes[5-7]. However, the exact target for blood glucose (BG) control in ICU is not well established. Moreover, targeting tight glucose control necessitates frequent blood sampling and adjustment of insulin dose, increasing the work-load on ICU staff. In addition, targeting tight glucose control has not shown to have any mortality benefit but is associated with five-fold increased risk of hypoglycemia[8].

It has been difficult to establish a safe blood sugar level but as per American Diabetes Association (ADA) a BG level below 180 mg/dL is acceptable[9]. The surviving sepsis guidelines further recommend a target BG levels between 140-180 mg/dL in patients with sepsis[10].

Artificial intelligence (AI) is a rapidly evolving science which is gradually changing the landscape of many industries including healthcare. As ICUs have a dynamic environment which generates a huge amount of data, AI has a tremendous scope and now is increasingly being used in advanced mechanical ventilation, weaning from ventilation, predicting development of sepsis, antibiotic dosing and radiological assessment and monitoring[11-15]. In this review, we will be discussing the current applications and potential role AI may have in managing critically ill diabetic patients.

ARTIFICIAL INTELLIGENCE

There is no standard definition of AI but as per the Encyclopaedia Britannica, AI refers to "a system endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience"[16]. Basically, AI based systems should be able to perform tasks comparable to human intelligence.

AI has great potential and has been used in the field of medicine for discovery of new drug molecules, diagnostics, radiology and imaging, molecular biology, bioinformatics and therapeutics. AI has the ability to analyze and scrutinize massive amounts of data and help understand disease patterns. The human brain can store a limited amount of information at any one time and may be unable to analyze and visualize patterns embedded in vast quantities of data[17]. In contrast computers have a large storage capacity and can discern even small associations within the data. However, computer programming has limitations as they are able to follow only certain specific patterns, as per the programming instructions. AI in contrast differs from traditional computer programming as it learns from exposure to various experiences and inputs, assimilates the data and can improve on its own

intelligence and modify the output behavior.

AI consists of a wide spectrum of complex algorithms and is broadly divided into machine learning (ML), deep learning, and cognitive computing. In ML, AI systems are trained with large repository of data and algorithms to enable them to follow a format to examine relationships and learn from them. Deep learning based systems develop insights by conducting complex interventions on the available data while cognitive AI systems are the most complex and try and match the human intelligence by understanding, reasoning, interacting, and learning from the data. Such systems are able to process and interpret exponential amounts of data (both structured and unstructured) and thus help in proposing any valid connections or hypothesis[18].

The AI functioning can be broken down in a systematic way and the processes involved can be divided into 3 main functions which occur in succession, which are knowledge discovery followed by learning and finally reasoning.

Knowledge discovery/ retrieval

The discovery of knowledge is the essence of AI. It works by creating algorithms for acquiring relevant and potential information from databases and is referred to as knowledge discovery in databases (KDD). For KDD to be effective it should have an in-depth knowledge of the area of interest as it will evaluate and interpret patterns and models to decide what data constitutes knowledge and what does not. KDD, hence plays a pivotal role in identifying information which is useful and valid.

Learning

Once the KDD process is complete the next step is learning from the knowledge or information acquired. Systems are allowed to automatically learn without human intervention or assistance. It usually consists of an inductive component which could be a simple process or could consist of a convolutional neural network (CNN). The various techniques used are artificial neural networks (ANNs), support vector machines (SVMs), random forest (RF), evolutionary algorithms, deep learning, Naive Bayes (NB), decision trees, and regression algorithms.

Certain types of AI algorithms are more commonly employed in healthcare settings than others. SVMs are used to predict clearly defined outcomes and adherence to medications. ANNs are algorithms which have been inspired by neuronal organization of animal brains, and have been employed to analyze data from computed tomography images, mammograms *etc*., to predict complications and outcomes. Logistic regression, is a ML algorithm which has been used to predict and classify probability of an event using predictor variables. Using data from electronic records or patient's medical history, RF algorithms have been used to predict risk of disease, and NB are the most advanced ML algorithms which have been used recently to predict development of disease in specific patient populations[19].

Reasoning

Reasoning is the final step in the AI process and involves the use of logical techniques to come to a conclusion from the available data. The primary objective of reasoning is to perform tasks at the level of a human intelligence and in a specialized manner with the final objective to generate inferences in the most precise manner.

AI algorithms

AI is a rapidly evolving technology with increasing number of subsets being introduced regularly, each having their own advantages and limitations. For prediction and management of diabetes, commonly used AI algorithms include linear regression (LR), classification/decision trees (DTs), RF, SVMs, ANNs, and NB.

LR is a regression model which analyses the data and predicts a continuous output, finding solution following a linear curve. DTs are predictive models which predict outcome from the given data, but can find solution using both linear and non-linear curves. DTs also fare better than LR models for categorical independent variables. RF is a variation of DT, supporting both linear and non-linear solutions, but is better at handling of missing values and outliers. It is more favorable than DTs as it is more robust, accurate and provides a more generalized solution.

SVMs are supervised learning algorithms which are recently gaining popularity for their applications in healthcare settings. Even though they are mostly used for classification problems in ML, they can also be applied for regression problems. They also support linear and non-linear solutions and are better than LR in handling outliers and analyzing data with large number of features.

ANN is an advanced technology based on the brain and the nerves and programmed to mimic the biological neural system. ANNs can also find non-linear solutions and are sub-classified as convolutional (feedforward networks) and recurrent (feedback loop) neural networks. ANNs have better accuracy but require larger training data as compared to LR.

As compared to LR, DT and RF, which are discriminative models, NB is a generative model which works well even with small data sets. This supervised learning algorithm is based on Bayes theorem and can provide solutions to classification problems. It is easy, fast and performs well in case of categorial data. However, it is a bad estimator and its probability outputs are not reliable.

ROLE OF AI IN MANAGEMENT OF DIABETES MELLITUS

Medical management forms only a small part of the entire spectrum of diabetes care, as diabetes mellitus (DM) is mainly a life-style disorder. Apart from medications, education on self-management (meal schedules, calorie counting, exercising, routine BG monitoring) and continuous medical care is paramount not only to prevent acute complications but also to minimize the risk of long-term complications like nephropathy, retinopathy, diabetic foot, cardiovascular disease, or stroke. As a result, diabetes care is complex and various medical and life-style related factors need to be taken into account to optimize management.

The use of AI in DM is not new and a number of studies have shown the role of AI applications in the care of diabetic patients[20-24]. A number of complex AI systems, and their clinical applications have been described (Table 1). Deep-learning based AI algorithms may help in early diagnosis of diabetic retinopathy using retinal photographs with a reported sensitivity and specificity of more than 90%[25]. IDx-DR is the first such AI-based device approved by US-FDA for screening of diabetic patients for retinopathy[26]. As it does not require a clinician to interpret the results, this automated system can help the non-eye specialists to recognize early signs of retinopathy and send the patients to eye-specialists only if indicated, thereby simplifying the process and achieving higher patient satisfaction[27].

Dreamed Advisor pro assimilates data regarding the glucose levels, insulin dose and carbohydrate intake and using AI-based MD-Logic algorithms it then makes recommendations for insulin dose adjustments. These recommendations have been shown to be similar to those given by experienced physicians in the real-world settings validating the use of such devices in day-to-day clinical practice[23, 28]. Several real-time Continuous Glucose Monitoring (CGM) devices like Medtronic Guardian Connect and Dexcom G6 CGM systems, are commercially available which can act as self-monitoring tools for diabetic patients (Table 1). These devices can provide real-time glucose values which can be displayed on the patient's mobile phones and can raise an alarm if the BG levels go beyond the predefined range. These devices can further be connected to insulin pumps and hence aid in insulin dose adjustments. However, these devices require repeated calibrations with the capillary blood glucose levels, to be measured by finger pricks. Use of these glucose sensors for more than 70% of the time, has shown to improve the HbA1c by 0.4 to 0.6% and reduce the incidence of hypoglycemic episodes[29]. Presently, these devices and applications have not been validated in ICU patients but can be further modified and tested to be applied in the management of critically ill patients.

AI IN DIABETES MANAGEMENT IN ICU

Hyperglycemia is a common phenomenon in the ICU irrespective of the reason for admission and may occur even in the absence of pre-existing DM. The pathophysiology of hyperglycemia in ICU is multifactorial and can occur secondary to release of stress hormones (corticosteroids and catecholamines), proinflammatory mediators, administration of exogenous drugs (corticosteroids, vasopressors, ascorbic acid), parenteral solutions containing dextrose, stress hyperglycemia and use of commercial dietary feeds or supplements[30]. Irrespective of cause, hyperglycemia is associated with an increase in ICU stay, hospitalization costs, morbidity, and mortality[4,31].

Apart from hyperglycemia, hypoglycemia and GV have also been shown to be associated with increase in mortality in critically ill patients[5,6]. Use of variable insulin protocols which are not clinically validated and inaccurate blood sugar measurements are responsible for this GV seen in the ICUs. In addition, insulin sensitivity in critically ill patients follows a very erratic course and is plagued with frequent changes which could be secondary to the underlying illness, dietary changes or medications.

TITR has been recognized as another domain of dysglycemia in critically ill patients[7]. It may be defined as the total time spent in the target range and is expressed as the percentage of time. Data suggests that critically ill patients having more than 70% TITR, have significantly higher survival rates [32]. However, the exact cut-offs for TITR remain unclear with different studies suggesting TITR ranging from 50-80% for improving outcomes[33,34].

In spite of several widely accepted applications for out-patient and long-term management of DM, AI applications in management of critically ill patients are limited. The possible applications of AI in critically ill diabetes patients are given in Table 2[35].

Blood glucose monitoring and prediction

Blood glucose management requires frequent sampling and insulin dose adjustments. Capillary BG monitoring still remains the most commonly employed method, even in critically ill patients. However, its accuracy may be affected in patients with subcutaneous oedema, shock, and hypoxemia, which commonly affect ICU patients. Hence, using arterial blood is preferred but it requires repeated arterial punctures or presence of an invasive arterial line. The characteristics of an ideal method to monitor BG is given in the Table 3.

Table 1 Clinical uses of artificial intelligence in management of diabetes

AI applications	Examples of AI devices	Clinical uses
Retinal screening	IDx-DR device	Screening and diagnosis of diabetic retinopathy
Clinical diagnosis	Advisor Pro	Detection and monitoring of diabetes and its associated complications. Fine-tuning insulin dose
Patient self-management tools	Medtronic Guardian Connect System, Dexcom G6 CGM systems; Mobile applications	Improve blood glucose control, activity and dietary tracking
Risk stratification	AI using random forest and; gradient boosting techniques	Prediction of new-onset diabetes; Prediction of subpopulations at risk for complications, non-compliance to therapy and hospitalization

AI: Artificial intelligence.

Table 2 Possible critical care applications of artificial intelligence in diabetes management

Blood glucose monitoring and prediction
Detection of adverse glycemic events
Blood glucose control strategies
Insulin bolus calculators and advisory systems
Risk and patient stratification

Table 3 Characteristics of an ideal tool to monitor blood glucose in intensive care unit

Ease to use
Minimal burden on staff
Automated data entry
High rate of adherence
Allow for minimal sampling
Comfortable to use for the patient
Use of a proven algorithm to calculate insulin dosage
Quickly correct hyperglycemia
Consistently maintain glucose within the predetermined optimal range
Ensure minimal glycemic variability
Prevent episodes of hypoglycemia
Provide easy interface with other patient measurements and data
Easy to integrate into existing hospital systems
Avoid the need for repeated data entry
Maintain results in a comprehensive, standardized database to facilitate multi-center comparison

Continuous glucose monitoring

Continuous Glucose Monitoring has been employed in the management of DM for more than a decade. Several CGM devices have been developed and are presently commercially available and approved for in-hospital use (Table 4). They can be broadly classified as transdermal (non-invasive), subcutaneous (minimally invasive) and intra-vascular (invasive) devices. Subcutaneous and transdermal devices are not considered ideal in critically ill patients because the presence of subcutaneous oedema, hypoxemia, and shock may affect their accuracy. Hence, intravascular devices may be preferable in these patients. However, the continuous subcutaneous flash glucose monitoring (FGM) system (FreeStyle Libre) has been recently tried in critically ill patients and has shown to have high test-retest reliability and acceptable accuracy[36-38].

Table 4 Continuous glucose monitoring devices

Type of device	Name of device	Comments
Intravenous	GlucoClear by Edwards Lifesciences; (Irvine, CA)	Approved in Europe
Intravenous	Glysure System by Glysure (Abingdon, UK)	Approved in Europe
Intravenous	Eirus by Maquet Getinge Group (Rastatt, Germany)	Approved in Europe
Intravenous	OptiScanner 5000 by OptiScan; (Hayward, CA)	Approved in EuropeFDA-approved for use in US hospitals
Intravenous	GlucoScout (International Biomedical, Austin, TX)	FDA-approved for use in US hospitals
Intravenous	Dexcom G	FDA-approved and CEA approved
Intravenous	Guardian™ Connect system by Medtronic (San Diego, CA)	FDA-approved for use in US hospitals
Subcutaneous	Freestyle Libre by Abbott Diabetes Care	US FDA approved

FDA: Food and Drug Administration; CEA: Carcinoembryonic antigen.

A recently published meta-analysis reported that the use of CGM was associated with significantly reduced HbA1c values and reduced risk of severe hypoglycaemia[39]. In addition, use of FGM was associated with significant reduction in episodes of mild hypoglycemia and was associated with increased treatment satisfaction in patients with type-I diabetes. Hence, it is suggested that real time monitoring with CGM or FGM has the potential to achieve better control in short-time fluctuations in BG levels, improve glycemic control and may also reduce healthcare costs[40]. Although several studies have been conducted testing these devices in critically ill patients, their impact on reducing length of stay in ICU or overall patient outcomes remains unknown[41].

While these devices may not benefit all ICU patients, they may be particularly useful in specific patient populations like those on intravenous insulin or corticosteroids, patients with end stage renal or liver disease, neurosurgery or traumatic brain injury patients and post-transplant patients[42-44]. However, these devices need to be further tested in larger patient cohorts before they find mainstream application.

Detection of adverse glycemic events

Detection of adverse events in the form of both hypoglycemia and hyperglycemia using AI technologies have been studied by various research groups mainly in type 1 and type 2 diabetes patients[35]. The studies used either CGM devices or self-monitoring of blood glucose monitors to detect the individual events. The results were based on the sensitivity and specificity of the modalities used. For example the DCBPN algorithm used by Zhang *et al*[45] provided an accuracy of 88.5% in predicting the BG levels. In the study by Otto *et al*[46], identification of episodes of hypoglycemia, hyperglycemia, severe hypoglycemia, and severe hyperglycemia were 120%, 46%, 123%, and 76% more likely after pattern identification as compared to periods when no pattern was identified. Another study by Nguyen *et al*[47] used electrocardiographic (ECG) parameters to detect episodes of hyperglycemia with a reported sensitivity and specificity of 70.59% and 65.38%, respectively. The results suggested that ECG signal and ANN patterns could be used to detect adverse hyperglycemic events in diabetic patients. Overall, AI has a potential role to predict adverse events and thus help modify treatment protocols so as to rectify them.

Blood glucose control strategies

There are various AI methodologies, fuzzy logic (FL), ANN, RF, which have been used for sugar control. Out of these FL is the most commonly used methodology as it mimics the management strategies by actual diabetes caregivers. Various studies have been performed using the FL methodology for BG control, mainly in type 1 diabetic patients[48,49]. The results have shown better control of nocturnal glucose levels with a low risk of hypoglycaemia as compared to standard insulin pump treatment.

Now, more complex methodologies are being proposed for BG control such as complimentary AI algorithms to support traditional AI controllers. The latest technology is the development of neural networks for regulation of BG[50,51].

From the above data it is evident that AI may potentially help to control BG but similar research in critically ill patients is limited. The LOGIC-1 trial was a single centre randomized control trial (RCT) which compared LOGIC-Insulin computerized algorithm to expert nurses in BG control for critically ill patients[52]. LOGIC-Insulin improved the efficacy of tight glucose control without increasing the risk of hypoglycemia. Encouraged by the results, a larger multi-center RCT, the LOGIC-2 trial, was conducted comparing software guided glucose control to nurse directed orders. This trial also showed better control of BG without an increase in hypoglycemia[53].

Hence, research shows that algorithmic based approach may be beneficial to control BG levels. Even the ability to anticipate excursions in sugar levels could provide early warnings regarding ineffective treatments. Newer CGM could lead to prediction of future glucose levels but reliability may be affected due various physiological and technical factors. Pappada *et al*[54] studied a neural network model for predicting glucose levels in a surgical critical care setting and found CGM to be useful in this patient population. However, further research and studies may be required in real time to test their validity in other critically ill patients.

Artificial pancreas

For BG control one of the most extensively researched modality is the artificial pancreas (AP) which consists of a glucose sensor, a closed-loop control algorithm, and an insulin infusion device. The glucose sensor estimates the BG level which in turn is fed to the control unit with the closed loop algorithm. This is turn directs the infusion device to inject the programmed amount of insulin. Thus, it has been developed to mimic the Islet cells of the pancreas which secrete insulin based on the BG levels. The majority of algorithms used by AP have been derived from control engineering theory and include proportional-integral-derivative (PID), model-predictive control, adaptive control, and FL control[55, 56]. However, the major limiting factor is a reliable glucose sensor and hence, now AI is being used to develop better models of AP.

At present, AP are of two types viz a viz single hormone (insulin only) and dual hormone (insulin and glucagon) systems. Overall, AP has been shown to be safe and effective in controlling BG, reducing episodes of hypoglycemia and hyperglycemia, and increase the proportion of TITR. Weisman *et al*[57] conducted a meta-analysis which showed that AP improves the TITR by 12.59% (equivalent to 172 minutes per day) compared to conventional treatment. Furthermore, this analysis showed that dual-hormone AP systems were associated with greater improvements, especially with respect to hypoglycemic events as compared to single hormone systems. The average time spent in hypoglycemia was reduced by 35 minutes/day. These benefits were more pronounced at night time.

In critically ill patients, use of AP to control BG has shown to reduce the frequency for sampling, reduce the nursing workload, achieve stable glycemic control with reduced episodes of hypo or hyperglycemia, and cause less GV[58-62]. In addition, its use has been associated with significant reduction in postoperative infectious complications in patients undergoing major surgeries[62]. However, use of AP was unable to achieve any significant improvement in mean glucose concentration, improve clinical outcome or show a favorable cost-benefit ratio.

Insulin bolus calculators and advisory systems

Insulin dependent patients routinely require calculation of insulin dosages based on their consumption of carbohydrates. The bolus doses are based on multiple factors like previous insulin dose, BG measurements, approximate calorie count *etc*. This may be a challenging task and could lead to errors in judgement and calculation, eventually leading to adverse glycemic events. Various applications are being developed to simplify this daunting task. Various research groups have used the case-based reasoning methodology for these calculations which has proved to be a safe decision tool. Some studies have also shown that complimenting this system to an AP leads to an improvement in glycemic control [62,63]. Since the cause of hyperglycemia in ICU is multifactorial, probably a combination of an AP with case-based methodology may be of help as glucose excursions could be treated in a more standardized way with better control.

MD-Logic controller, developed on the FL systems, have shown to provide superior glycemic control with fewer nocturnal hypoglycemic episodes as compared to insulin pump treatment[49]. However, it still needs to be validated in ICU patients.

Software based algorithms for insulin dosing

Software based algorithms have been developed to determine insulin dosage depending on the BG levels. These programs, although more complicated than the paper-based protocols, can reduce errors and improve adherence. The simplest of these are based on PID models. Devices based on this model titrate insulin administration based on the previous BG values and predicting the changes in glucose value for a given insulin dose using a dynamic multiplier response to insulin sensitivity. The advantages of this model include the need for minimal patient related information for initiation and its ability to provide real-time dose adjustments. However, this model necessitates multiple blood sampling, which may be up to 18 times per day for BG measurements[64,65].

A more complex modification of software is Glucose Regulation for Intensive Care Patients which not only takes into account the BG values and insulin infusion rates but also includes the change in these values over time. This may increase its effectiveness and may potentially reduce overtreatment and hence, hypoglycemic episodes[66,67].

The most recent algorithms are classified as model predictive controls, which not only include insulin sensitivity and dextrose administration but also include several patient-specific parameters like their age and diabetes status. Based on these factors, these algorithms try to predict the patient's response to hyperglycemia and insulin therapy and adjust the insulin dose accordingly. As the number of

parameters required to be entered at the time of initiation are more, the devices based on these algorithms are more complicated and time consuming but they have advantages of increased accuracy, significantly reduced need for repeated blood sampling and may offer a more individualized insulin therapy[68-70].

CGM regulated insulin infusion system

Newer technologies like CGM which have been validated in non-critically ill patients are now increasingly been used with increased accuracy in ICU patients. Integration of these CGM devices with automated insulin suspension with AI algorithms (Basal-IQ™ technology) have been approved by US-FDA. Use of these predictive low-glucose suspend (PLGS) algorithms offer clinical advantage over the more conventional threshold suspend systems which stop insulin only when the predefined threshold of glucose is breached. Glucose values are obtained by the integrated CGM device (Dexcom G6™) and the Basal-IQ™ has the ability to predict when the glucose value is going to drop below the predefined level and it stops the insulin infusion[71]. Control-IQ is a more advanced hybrid closed-loop system which also uses activity and sleep settings to adjust the insulin requirements. Basal-IQ™ and Control-IQ™ algorithms can predict hypoglycemic events up to 30 minutes in advance and hence, can titrate the insulin dose accordingly.

Integration of CGM with an automated insulin suspension has shown to reduce the frequency and duration of hypoglycaemia with a reported relative risk reduction of 45%[72]. This effect has been shown to exist across different age groups, and is persistent over multiple weeks with real-world use. A large randomized crossover trial comparing the PLGS with sensor-augmented insulin pump showed 31% reduction in time spent in hypoglycemia (< 70 mg/dL) with no increase in incidence of rebound hyperglycemia[73]. It may be suggested that, use of this technology may be feasible and effective for patients with difficult to control DM and those at higher risk for developing hypoglycemia[72].

Risk and patient stratification

Diabetes is a chronic disease associated with many complications. Even though most of the complications develop over a period of time, diabetic patients are also prone to develop acute life-threatening complications like nosocomial infections, acute kidney injury and even cardiovascular complications. AI using deep-learning techniques have been able to produce algorithms which are able to predict long-term micro-angiopathic complications like diabetic retinopathy, diabetic foot, diabetic neuropathy and diabetic nephropathy, with reasonable accuracy[74-77]. Role of AI in predicting the development of macro-angiopathic complications like acute myocardial infarction has also been assessed but there is a dearth of data regarding its role in predicting other acute complications, especially in critically ill patients[78].

AI has been used effectively to determine patients at risk for developing sepsis and life-threatening nosocomial infections like catheter related blood stream infections and *Clostridium difficile* infections and also to predict which ward patients may deteriorate and require ICU admission. However, such models currently do not exist specifically for diabetes patients[13,79-81].

A few studies have also used AI in predicting mortality in critically ill diabetes patients. In their study, Ye *et al*[82] using the MIMIC-III database, reported that AI using CNN was highly accurate in predicting mortality in critically ill diabetes patients with an area under the curve (AUC) of 0.97. Using the same MIMIC-III database, Anand *et al*[83] developed simple predictive tools with AI, to predict mortality in critically ill diabetics. Their models could achieve AUCs of 0.787 and 0.785 to predict mortality. However, these models need to be compared to more widely used and validated models for mortality prediction in ICU patients like acute physiology and chronic health evaluation and sequential organ failure and assessment scores.

Coronavirus disease critical care

The recent pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has put an unprecedented strain on the healthcare with enhanced need for infection control and patient isolation. Separate coronavirus disease 2019 (COVID-19) ICUs had to be developed with negative pressure chambers with treating staff wearing personal protection equipment at all times. Diabetes is one of the most common comorbidities among COVID-19 patients. Diabetic patients developing COVID-19 are at higher risk for requiring ICU admission and have poorer outcomes. The need for personal protection and risk of transmission of infection has put immense pressure on already limited clinical workforce. In such a scenario, labour intensive work like frequent BG monitoring and insulin dose adjustments may get seriously hampered. AI may be especially helpful by reducing the burden on the healthcare workers (HCWs) and reducing their risk of exposure.

Computerized algorithms, automated closed loop systems and remote monitoring may all be used effectively to manage critically ill COVID-19 patients. CGM devices are capable of continuous BG tracking enabling real-time monitoring of BG levels while reducing the need for bedside monitoring, thereby reducing the risk of exposure for the HCWs. The efficacy and safety of CGM in managing critically ill COVID-19 patients has been tested and verified and it has been reported to reduce the need for bedside BG testing by up to 71%. In addition, the efficacy of CGM devices was not significantly

affected by presence of fever, hypoxemia, need for vasopressors, acidosis or with use of corticosteroid or parenteral nutrition[84-86]. Based on this, US-FDA has allowed the use of CGM in COVID-19 ICUs to reduce the exposure of HCWs[87].

AI based devices have the potential to improve patient care and outcomes by providing a better glucose control without increasing the nursing workload and avoiding risk of transmission of infection. Hence, it is recommended to prefer CGM to reduce the need for frequent nurse contact for patients with active COVID-19 infection[88]. Moreover, AI has also been instrumental in achieving glycemic control in COVID-19 patient on extracorporeal membrane oxygenation support by using AP[89].

STRENGTHS OF AI

AI-based devices have the potential to improve glycemic control, reduce GV, increase the TITR, and reduce episodes of hyper and hypoglycemia, thus providing comprehensive diabetes care. AI may allow us to achieve a better and more individualized glycemic control taking into account specific patient requirements as per their calorie intake, exercise and underlying comorbidities. In addition, AI may be better suited to care for patients at risk for adverse effects and those with changing needs, like those in critical care areas. It may enable HCWs to monitor their patients remotely with reduced need for close contact thereby, reducing their workload and exposure to infective patients. By reducing the need for frequent blood sampling and providing close glucose monitoring and insulin dose titration, AI-based algorithms may increase patient safety and satisfaction.

LIMITATIONS OF AI

Healthcare applications of AI are rapidly increasing. However, it still has several limitations affecting its widespread applicability (Table 5). Even though many AI applications have found acceptability in outpatients and ward patients with diabetes, data regarding its safety and accuracy in critically ill patients remains limited. As AI application is largely data-driven, involving collection of sensitive personal data, it may have privacy issues leading to medico-legal problems. Lack of regulations, recommendations and guidelines pertaining to use of AI further limit its applicability. These safety, liability and reliability issues prevent widespread use of AI in critical care practice. In addition, challenges of integrating AI into existing healthcare infrastructure and user acceptance also persist.

FUTURE DIRECTIONS

The future of healthcare development is in AI. Its large-scale applicability requires widespread availability, low cost and ease of use. In addition, AI needs to be adapted gradually in the existing healthcare system and HCWs need to be trained not only to better utilize AI but also to be aware of how to avoid any medico-legal issues arising from its application. Changes in the laws and regulations are also required to safeguard patient's interest and avoid any violation of patient's privacy. With technological improvements in AI, the dosing algorithms for insulin delivery may become individualized for closed-loop control of glycemia. Larger studies, evaluating their efficacy and safety, especially in critically ill patients, along with standardization of AI algorithms and techniques need to be done to improve the acceptability of AI.

CONCLUSION

Many currently available devices and techniques which have proven their role in management of non-critically ill patients, may soon be available for ICU patients, with improved accuracy. CGM is already being recommended for use in critically ill COVID-19 patients and soon may be available for use in all critically ill patients. Its integration with automated insulin suspension holds greater promise. Use of AP may also provide a comprehensive glycemic control option. AI has the potential of reducing the workload of HCWs, provide better glycemic control and prevent related complications, however, larger RCTs may be required before we implement these techniques in our day-to-day critical care. Even though presently AI might not be in its prime for managing critically ill diabetic patients, it is the future of healthcare.

Table 5 Limitations of artificial intelligence

Factors	
Human factors	Inhibition, lack of experience
Technical factors	Cost, availability and implementation
Data limitation	Lack of data in ICU patients, lack of large scale randomized trials
Design limitation	Devices tried in certain patient populations may not be applicable in ICU patients
Ethical	Lack of guidelines

ICU: Intensive care unit.

FOOTNOTES

Author contributions: Juneja D and Gupta A performed the majority of the writing, prepared the tables and performed data accusation; Singh O provided the input in writing the paper and reviewed the manuscript.

Conflict-of-interest statement: The authors declare that they have no conflict of interest.

Country/Territory of origin: India

ORCID number: Deven Juneja 0000-0002-8841-5678; Anish Gupta 0000-0002-0901-4797; Omender Singh 0000-0002-3847-4645.

S-Editor: Liu JH
L-Editor: A
P-Editor: Qi WW

REFERENCES

1 IDF Diabetes Atlas 10th edition Internet. 2217. Cited 26 January 2022. Available from: https://idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html
2 **Kompoti M**, Michalia M, Salma V, Diogou E, Lakoumenta A, Clouva-Molyvdas PM. Glycated hemoglobin at admission in the intensive care unit: clinical implications and prognostic relevance. *J Crit Care* 2015; **30**: 150-155 [PMID: 25239822 DOI: 10.1016/j.jcrc.2014.08.014]
3 **Abu-Ashour W**, Twells L, Valcour J, Randell A, Donnan J, Howse P, Gamble JM. The association between diabetes mellitus and incident infections: a systematic review and meta-analysis of observational studies. *BMJ Open Diabetes Res Care* 2017; **5**: e000336 [PMID: 28761647 DOI: 10.1136/bmjdrc-2016-000336]
4 **Egi M**, Bellomo R, Stachowski E, French CJ, Hart GK, Hegarty C, Bailey M. Blood glucose concentration and outcome of critical illness: the impact of diabetes. *Crit Care Med* 2008; **36**: 2249-2255 [PMID: 18664780 DOI: 10.1097/CCM.0b013e318181039a]
5 **NICE-SUGAR Study Investigators**, Finfer S, Liu B, Chittock DR, Norton R, Myburgh JA, McArthur C, Mitchell I, Foster D, Dhingra V, Henderson WR, Ronco JJ, Bellomo R, Cook D, McDonald E, Dodek P, Hébert PC, Heyland DK, Robinson BG. Hypoglycemia and risk of death in critically ill patients. *N Engl J Med* 2012; **367**: 1108-1118 [PMID: 22992074 DOI: 10.1056/NEJMoa1204942]
6 **Krinsley JS**. Understanding glycemic control in the critically ill: three domains are better than one. *Intensive Care Med* 2011; **37**: 382-384 [PMID: 21210079 DOI: 10.1007/s00134-010-2110-3]
7 **Krinsley JS**. Glycemic control in the critically ill: What have we learned since NICE-SUGAR? *Hosp Pract (1995)* 2015; **43**: 191-197 [PMID: 26224425 DOI: 10.1080/21548331.2015.1066227]
8 **Yamada T**, Shojima N, Noma H, Yamauchi T, Kadowaki T. Glycemic control, mortality, and hypoglycemia in critically ill patients: a systematic review and network meta-analysis of randomized controlled trials. *Intensive Care Med* 2017; **43**: 1-15 [PMID: 27637719 DOI: 10.1007/s00134-016-4523-0]
9 **American Diabetes Association**. 15. Diabetes Care in the Hospital: *Standards of Medical Care in Diabetes-2021*. *Diabetes Care* 2021; **44**: S211-S220 [PMID: 33298426 DOI: 10.2337/dc21-S015]
10 **Evans L**, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, Machado FR, Mcintyre L, Ostermann M, Prescott HC, Schorr C, Simpson S, Joost Wiersinga W, Alshamsi F, Angus DC, Arabi Y, Azevedo L, Beale R, Beilman G, Belley-Cote E, Burry L, Cecconi M, Centofanti J, Yataco AC, De Waele J, Dellinger RP, Doi K, Du B, Estenssoro E,

Ferrer R, Gomersall C, Hodgson C, Møller MH, Iwashyna T, Jacob S, Kleinpell R, Klompas M, Koh Y, Kumar A, Kwizera A, Lobo S, Masur H, McGloughlin S, Mehta S, Mehta Y, Mer M, Nunnally M, Oczkowski S, Osborn T, Papathanassoglou E, Perner A, Puskarich M, Roberts J, Schweickert W, Seckel M, Sevransky J, Sprung CL, Welte T, Zimmerman J, Levy M. Executive Summary: Surviving Sepsis Campaign: International Guidelines for the Management of Sepsis and Septic Shock 2021. *Crit Care Med* 2021; **49**: 1974-1982 [PMID: 34643578 DOI: 10.1097/CCM.0000000000005357]

11 **Sottile PD**, Albers D, Higgins C, Mckeehan J, Moss MM. The Association Between Ventilator Dyssynchrony, Delivered Tidal Volume, and Sedation Using a Novel Automated Ventilator Dyssynchrony Detection Algorithm. *Crit Care Med* 2018; **46**: e151-e157 [PMID: 29337804 DOI: 10.1097/CCM.0000000000002849]

12 **Kuo HJ**, Chiu HW, Lee CN, Chen TT, Chang CC, Bien MY. Improvement in the Prediction of Ventilator Weaning Outcomes by an Artificial Neural Network in a Medical ICU. *Respir Care* 2015; **60**: 1560-1569 [PMID: 26329358 DOI: 10.4187/respcare.03648]

13 **Nemati S**, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Crit Care Med* 2018; **46**: 547-553 [PMID: 29286945 DOI: 10.1097/CCM.0000000000002936]

14 **Turnidge J**. Pharmacodynamics and dosing of aminoglycosides. *Infect Dis Clin North Am* 2003; **17**: 503-528, v [PMID: 14711074 DOI: 10.1016/s0891-5520(03)00057-6]

15 **Rachh P**, Levey AO, Lemmon A, Marinescu A, Auffermann WF, Haycook D, Berkowitz EA. Reducing STAT Portable Chest Radiograph Turnaround Times: A Pilot Study. *Curr Probl Diagn Radiol* 2018; **47**: 156-160 [PMID: 28705527 DOI: 10.1067/j.cpradiol.2017.05.012]

16 **Copeland BJ**. "Artificial Intelligence". Encyclopedia Britannica. Cited 22 January 2022. Available from: https://www.britannica.com/technology/artificial-intelligence

17 **Gobet F**, Clarkson G. Chunks in expert memory: evidence for the magical number four ... or is it two? *Memory* 2004; **12**: 732-747 [PMID: 15724362 DOI: 10.1080/09658210344000530]

18 **Raghupathi V**, Raghupathi W. Healthcare Expenditure and Economic Performance: Insights From the United States Data. *Front Public Health* 2020; **8**: 156 [PMID: 32478027 DOI: 10.3389/fpubh.2020.00156]

19 **Shillan D**, Sterne JAC, Champneys A, Gibbison B. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. *Crit Care* 2019; **23**: 284 [PMID: 31439010 DOI: 10.1186/s13054-019-2564-9]

20 **Nomura A**, Noguchi M, Kometani M, Furukawa K, Yoneda T. Artificial Intelligence in Current Diabetes Management and Prediction. *Curr Diab Rep* 2021; **21**: 61 [PMID: 34902070 DOI: 10.1007/s11892-021-01423-2]

21 **Ellahham S**. Artificial Intelligence: The Future for Diabetes Care. *Am J Med* 2020; **133**: 895-900 [PMID: 32325045 DOI: 10.1016/j.amjmed.2020.03.033]

22 **Abràmoff MD**, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 2018; **1**: 39 [PMID: 31304320 DOI: 10.1038/s41746-018-0040-6]

23 **Nimri R**, Battelino T, Laffel LM, Slover RH, Schatz D, Weinzimer SA, Dovc K, Danne T, Phillip M; NextDREAM Consortium. Insulin dose optimization using an automated artificial intelligence-based decision support system in youths with type 1 diabetes. *Nat Med* 2020; **26**: 1380-1384 [PMID: 32908282 DOI: 10.1038/s41591-020-1045-7]

24 **Dankwa-Mullan I**, Rivo M, Sepulveda M, Park Y, Snowdon J, Rhee K. Transforming Diabetes Care Through Artificial Intelligence: The Future Is Here. *Popul Health Manag* 2019; **22**: 229-242 [PMID: 30256722 DOI: 10.1089/pop.2018.0129]

25 **Lam C**, Yu C, Huang L, Rubin D. Retinal Lesion Detection With Deep Learning Using Image Patches. *Invest Ophthalmol Vis Sci* 2018; **59**: 590-596 [PMID: 29372258 DOI: 10.1167/iovs.17-22721]

26 **US Food and Drug Administration**. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Accessed April 18, 2022. Available from: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm

27 **Keel S**, Lee PY, Scheetz J, Li Z, Kotowicz MA, MacIsaac RJ, He M. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. *Sci Rep* 2018; **8**: 4330 [PMID: 29531299 DOI: 10.1038/s41598-018-22612-2]

28 **Nimri R**, Dassau E, Segall T, Muller I, Bratina N, Kordonouri O, Bello R, Biester T, Dovc K, Tenenbaum A, Brener A, Šimunović M, Sakka SD, Nevo Shenker M, Passone CG, Rutigliano I, Tinti D, Bonura C, Caiulo S, Ruszala A, Piccini B, Giri D, Stein R, Rabbone I, Bruzzi P, Omladič JŠ, Steele C, Beccuti G, Yackobovitch-Gavan M, Battelino T, Danne T, Atlas E, Phillip M. Adjusting insulin doses in patients with type 1 diabetes who use insulin pump and continuous glucose monitoring: Variations among countries and physicians. *Diabetes Obes Metab* 2018; **20**: 2458-2466 [PMID: 29885025 DOI: 10.1111/dom.13408]

29 **Gómez AM**, Henao Carrillo DC, Muñoz Velandia OM. Devices for continuous monitoring of glucose: update in technology. *Med Devices (Auckl)* 2017; **10**: 215-224 [PMID: 28979168 DOI: 10.2147/MDER.S110121]

30 **Dungan KM**, Braithwaite SS, Preiser JC. Stress hyperglycaemia. *Lancet* 2009; **373**: 1798-1807 [PMID: 19465235 DOI: 10.1016/S0140-6736(09)60553-5]

31 **Siegelaar SE**, Hermanides J, Oudemans-van Straaten HM, van der Voort PH, Bosman RJ, Zandstra DF, DeVries JH. Mean glucose during ICU admission is related to mortality by a U-shaped curve in surgical and medical patients: a retrospective cohort study. *Crit Care* 2010; **14**: R224 [PMID: 21143980 DOI: 10.1186/cc9369]

32 **Signal M**, Le Compte A, Shaw GM, Chase JG. Glycemic levels in critically ill patients: are normoglycemia and low variability associated with improved outcomes? *J Diabetes Sci Technol* 2012; **6**: 1030-1037 [PMID: 23063028 DOI: 10.1177/193229681200600506]

33 **Penning S**, Chase JG, Preiser JC, Pretty CG, Signal M, Mélot C, Desaive T. Does the achievement of an intermediate glycemic target reduce organ failure and mortality? *J Crit Care* 2014; **29**: 374-379 [PMID: 24679489 DOI: 10.1016/j.jcrc.2014.01.013]

34 **Omar AS**, Salama A, Allam M, Elgohary Y, Mohammed S, Tuli AK, Singh R. Association of time in blood glucose range with outcomes following cardiac surgery. *BMC Anesthesiol* 2015; **15**: 14 [PMID: 25670921 DOI: 10.1186/1471-2253-15-14]

35 **Contreras I**, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. *J Med Internet Res* 2018; **20**: e10775 [PMID: 29848472 DOI: 10.2196/10775]

36 **Zhang Y**, Liu X, Zhang J, Fu J, Li S, Chen S, Chen Y, Lu B. Evaluation for the feasibility and accuracy of Freestyle Libre Flash Glucose Monitoring System used by COVID-19 patients in intensive care unit. *J Diabetes* 2021; **13**: 603-605 [PMID: 33787006 DOI: 10.1111/1753-0407.13181]

37 **Kotzapanagiotou E**, Tsotridou E, Volakli E, Dimitriadou M, Chochliourou E, Kalamitsou S, Kotzapanagiotou F, Sdougka M, Christoforidis A. Evaluation of continuous flash glucose monitoring in a pediatric ICU setting. *J Clin Monit Comput* 2020; **34**: 843-852 [PMID: 31482363 DOI: 10.1007/s10877-019-00384-y]

38 **Ancona P**, Eastwood GM, Lucchetta L, Ekinci EI, Bellomo R, Mårtensson J. The performance of flash glucose monitoring in critically ill patients with diabetes. *Crit Care Resusc* 2017; **19**: 167-174 [PMID: 28651513]

39 **Dicembrini I**, Cosentino C, Monami M, Mannucci E, Pala L. Effects of real-time continuous glucose monitoring in type 1 diabetes: a meta-analysis of randomized controlled trials. *Acta Diabetol* 2021; **58**: 401-410 [PMID: 32789691 DOI: 10.1007/s00592-020-01589-3]

40 **Wood A**, O'Neal D, Furler J, Ekinci EI. Continuous glucose monitoring: a review of the evidence, opportunities for future use and ongoing challenges. *Intern Med J* 2018; **48**: 499-508 [PMID: 29464891 DOI: 10.1111/imj.13770]

41 **Umpierrez GE**, Klonoff DC. Diabetes Technology Update: Use of Insulin Pumps and Continuous Glucose Monitoring in the Hospital. *Diabetes Care* 2018; **41**: 1579-1589 [PMID: 29936424 DOI: 10.2337/dci18-0002]

42 **Wallia A**, Umpierrez GE, Nasraway SA, Klonoff DC; PRIDE Investigators. Round Table Discussion on Inpatient Use of Continuous Glucose Monitoring at the International Hospital Diabetes Meeting. *J Diabetes Sci Technol* 2016; **10**: 1174-1181 [PMID: 27286715 DOI: 10.1177/1932296816656380]

43 **Wallia A**, Umpierrez GE, Rushakoff RJ, Klonoff DC, Rubin DJ, Hill Golden S, Cook CB, Thompson B; DTS Continuous Glucose Monitoring in the Hospital Panel. Consensus Statement on Inpatient Use of Continuous Glucose Monitoring. *J Diabetes Sci Technol* 2017; **11**: 1036-1044 [PMID: 28429611 DOI: 10.1177/1932296817706151]

44 **Krinsley JS**, Chase JG, Gunst J, Martensson J, Schultz MJ, Taccone FS, Wernerman J, Bohe J, De Block C, Desaive T, Kalfon P, Preiser JC. Continuous glucose monitoring in the ICU: clinical considerations and consensus. *Crit Care* 2017; **21**: 197 [PMID: 28756769 DOI: 10.1186/s13054-017-1784-0]

45 **Zhang Y**, Zhu JM, Liang YB, Chen HB, Yin SM, Chen ZC. Non-invasive blood glucose detection system based on conservation of energy method. *Physiol Meas* 2017; **38**: 325-342 [PMID: 28107204 DOI: 10.1088/1361-6579/aa50cf]

46 **Otto EA**, Tannan V. Evaluation of the utility of a glycemic pattern identification system. *J Diabetes Sci Technol* 2014; **8**: 830-838 [PMID: 24876425 DOI: 10.1177/1932296814532210]

47 **Nguyen LL**, Su S, Nguyen HT. Neural network approach for non-invasive detection of hyperglycemia using electrocardiographic signals. *Annu Int Conf IEEE Eng Med Biol Soc* 2014; **2014**: 4475-4478 [PMID: 25570985 DOI: 10.1109/EMBC.2014.6944617]

48 **Nimri R**, Atlas E, Ajzensztejn M, Miller S, Oron T, Phillip M. Feasibility study of automated overnight closed-loop glucose control under MD-logic artificial pancreas in patients with type 1 diabetes: the DREAM Project. *Diabetes Technol Ther* 2012; **14**: 728-735 [PMID: 22853723 DOI: 10.1089/dia.2012.0004]

49 **Phillip M**, Battelino T, Atlas E, Kordonouri O, Bratina N, Miller S, Biester T, Stefanija MA, Muller I, Nimri R, Danne T. Nocturnal glucose control with an artificial pancreas at a diabetes camp. *N Engl J Med* 2013; **368**: 824-833 [PMID: 23445093 DOI: 10.1056/NEJMoa1206881]

50 **Fernandez de Canete J**, Gonzalez-Perez S, Ramos-Diaz JC. Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes. *Comput Methods Programs Biomed* 2012; **106**: 55-66 [PMID: 22178070 DOI: 10.1016/j.cmpb.2011.11.006]

51 **Catalogna M**, Cohen E, Fishman S, Halpern Z, Nevo U, Ben-Jacob E. Artificial neural networks based controller for glucose monitoring during clamp test. *PLoS One* 2012; **7**: e44587 [PMID: 22952998 DOI: 10.1371/journal.pone.0044587]

52 **Van Herpe T**, Mesotten D, Wouters PJ, Herbots J, Voets E, Buyens J, De Moor B, Van den Berghe G. LOGIC-insulin algorithm-guided versus nurse-directed blood glucose control during critical illness: the LOGIC-1 single-center, randomized, controlled clinical trial. *Diabetes Care* 2013; **36**: 188-194 [PMID: 22961576 DOI: 10.2337/dc12-0584]

53 **Dubois J**, Van Herpe T, van Hooijdonk RT, Wouters R, Coart D, Wouters P, Van Assche A, Veraghtert G, De Moor B, Wauters J, Wilmer A, Schultz MJ, Van den Berghe G, Mesotten D. Software-guided versus nurse-directed blood glucose control in critically ill patients: the LOGIC-2 multicenter randomized controlled clinical trial. *Crit Care* 2017; **21**: 212 [PMID: 28806982 DOI: 10.1186/s13054-017-1799-6]

54 **Pappada SM**, Cameron BD, Rosman PM, Bourey RE, Papadimos TJ, Olorunto W, Borst MJ. Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes. *Diabetes Technol Ther* 2011; **13**: 135-141 [PMID: 21284480 DOI: 10.1089/dia.2010.0104]

55 **Bothe MK**, Dickens L, Reichel K, Tellmann A, Ellger B, Westphal M, Faisal AA. The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas. *Expert Rev Med Devices* 2013; **10**: 661-673 [PMID: 23972072 DOI: 10.1586/17434440.2013.827515]

56 **Bertachi A**, Ramkissoon CM, Bondia J, Vehí J. Automated blood glucose control in type 1 diabetes: A review of progress and challenges. *Endocrinol Diabetes Nutr (Engl Ed)* 2018; **65**: 172-181 [PMID: 29279252 DOI: 10.1016/j.endinu.2017.10.011]

57 **Weisman A**, Bai JW, Cardinez M, Kramer CK, Perkins BA. Effect of artificial pancreas systems on glycaemic control in patients with type 1 diabetes: a systematic review and meta-analysis of outpatient randomised controlled trials. *Lancet Diabetes Endocrinol* 2017; **5**: 501-512 [PMID: 28533136 DOI: 10.1016/S2213-8587(17)30167-5]

58 **Mibu K**, Yatabe T, Hanazaki K. Blood glucose control using an artificial pancreas reduces the workload of ICU nurses. *J Artif Organs* 2012; **15**: 71-76 [PMID: 21947674 DOI: 10.1007/s10047-011-0611-7]

59 **DeJournett L**, DeJournett J. In Silico Testing of an Artificial-Intelligence-Based Artificial Pancreas Designed for Use in the Intensive Care Unit Setting. *J Diabetes Sci Technol* 2016; **10**: 1360-1371 [PMID: 27301982 DOI: 10.1177/1932296816653967]

60 **Hanazaki K**, Tanioka N, Munekage M, Uemura S, Maeda H. Closed-loop artificial endocrine pancreas from Japan. *Artif*

Organs 2021; **45**: 958-967 [PMID: 34105784 DOI: 10.1111/aor.14008]
61 **Yang Z**, Tao G, Guo M, Sun B, Gong L, Ding Y, Ye S, Liu W, Yang X. [Efficacy and safety of simulated artificial pancreas in modulating stress hyperglycemia in critically ill patients: a prospective randomized controlled study]. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue* 2018; **30**: 165-169 [PMID: 29402368 DOI: 10.3760/cma.j.issn.2095-4352.2018.02.014]
62 **Herrero P**, Pesl P, Reddy M, Oliver N, Georgiou P, Toumazou C. Advanced Insulin Bolus Advisor Based on Run-To-Run Control and Case-Based Reasoning. *IEEE J Biomed Health Inform* 2015; **19**: 1087-1096 [PMID: 24956470 DOI: 10.1109/JBHI.2014.2331896]
63 **Herrero P**, Bondia J, Adewuyi O, Pesl P, El-Sharkawy M, Reddy M, Toumazou C, Oliver N, Georgiou P. Enhancing automatic closed-loop glucose control in type 1 diabetes with an adaptive meal bolus calculator - in silico evaluation under intra-day variability. *Comput Methods Programs Biomed* 2017; **146**: 125-131 [PMID: 28688482 DOI: 10.1016/j.cmpb.2017.05.010]
64 **Juneja R**, Roudebush C, Kumar N, Macy A, Golas A, Wall D, Wolverton C, Nelson D, Carroll J, Flanders SJ. Utilization of a computerized intravenous insulin infusion program to control blood glucose in the intensive care unit. *Diabetes Technol Ther* 2007; **9**: 232-240 [PMID: 17561793 DOI: 10.1089/dia.2006.0015]
65 **Newton CA**, Smiley D, Bode BW, Kitabchi AE, Davidson PC, Jacobs S, Steed RD, Stentz F, Peng L, Mulligan P, Freire AX, Temponi A, Umpierrez GE. A comparison study of continuous insulin infusion protocols in the medical intensive care unit: computer-guided vs. standard column-based algorithms. *J Hosp Med* 2010; **5**: 432-437 [PMID: 20945468 DOI: 10.1002/jhm.816]
66 **Vogelzang M**, Loef BG, Regtien JG, van der Horst IC, van Assen H, Zijlstra F, Nijsten MW. Computer-assisted glucose control in critically ill patients. *Intensive Care Med* 2008; **34**: 1421-1427 [PMID: 18389221 DOI: 10.1007/s00134-008-1091-y]
67 **Hoekstra M**, Schoorl MA, van der Horst IC, Vogelzang M, Wietasch JK, Zijlstra F, Nijsten MW. Computer-assisted glucose regulation during rapid step-wise increases of parenteral nutrition in critically ill patients: a proof of concept study. *JPEN J Parenter Enteral Nutr* 2010; **34**: 549-553 [PMID: 20852185 DOI: 10.1177/0148607110372390]
68 **Pachler C**, Plank J, Weinhandl H, Chassin LJ, Wilinska ME, Kulnik R, Kaufmann P, Smolle KH, Pilger E, Pieber TR, Ellmerer M, Hovorka R. Tight glycaemic control by an automated algorithm with time-variant sampling in medical ICU patients. *Intensive Care Med* 2008; **34**: 1224-1230 [PMID: 18297268 DOI: 10.1007/s00134-008-1033-8]
69 **Hoekstra M**, Vogelzang M, Verbitskiy E, Nijsten MW. Health technology assessment review: Computerized glucose regulation in the intensive care unit--how to create artificial control. *Crit Care* 2009; **13**: 223 [PMID: 19849827 DOI: 10.1186/cc8023]
70 **Amrein K**, Ellmerer M, Hovorka R, Kachel N, Fries H, von Lewinski D, Smolle K, Pieber TR, Plank J. Efficacy and safety of glucose control with Space GlucoseControl in the medical intensive care unit--an open clinical investigation. *Diabetes Technol Ther* 2012; **14**: 690-695 [PMID: 22694176 DOI: 10.1089/dia.2012.0021]
71 Tandem Diabetes Care Announces Health Canada Approval of t:slim X2 Insulin Pump with Basal-IQ Predictive Low-Glucose Suspend Technology. Cited 6 January 2022. Available from: http://investor.tandemdiabetes.com/news-releases/news-release-details/tandem-diabetes-care-announces-health-canadaapproval-tslim-x2/
72 **Müller L**, Habif S, Leas S, Aronoff-Spencer E. Reducing Hypoglycemia in the Real World: A Retrospective Analysis of Predictive Low-Glucose Suspend Technology in an Ambulatory Insulin-Dependent Cohort. *Diabetes Technol Ther* 2019; **21**: 478-484 [PMID: 31329468 DOI: 10.1089/dia.2019.0190]
73 **Forlenza GP**, Li Z, Buckingham BA, Pinsker JE, Cengiz E, Wadwa RP, Ekhlaspour L, Church MM, Weinzimer SA, Jost E, Marcal T, Andre C, Carria L, Swanson V, Lum JW, Kollman C, Woodall W, Beck RW. Predictive Low-Glucose Suspend Reduces Hypoglycemia in Adults, Adolescents, and Children With Type 1 Diabetes in an At-Home Randomized Crossover Study: Results of the PROLOG Trial. *Diabetes Care* 2018; **41**: 2155-2161 [PMID: 30089663 DOI: 10.2337/dc18-0771]
74 **Ting DSW**, Cheung CY, Nguyen Q, Sabanayagam C, Lim G, Lim ZW, Tan GSW, Soh YQ, Schmetterer L, Wang YX, Jonas JB, Varma R, Lee ML, Hsu W, Lamoureux E, Cheng CY, Wong TY. Deep learning in estimating prevalence and systemic risk factors for diabetic retinopathy: a multi-ethnic study. *NPJ Digit Med* 2019; **2**: 24 [PMID: 31304371 DOI: 10.1038/s41746-019-0097-x]
75 **Cruz-Vega I**, Hernandez-Contreras D, Peregrina-Barreto H, Rangel-Magdaleno JJ, Ramirez-Cortes JM. Deep Learning Classification for Diabetic Foot Thermograms. *Sensors (Basel)* 2020; **20** [PMID: 32235780 DOI: 10.3390/s20061762]
76 **Williams BM**, Borroni D, Liu R, Zhao Y, Zhang J, Lim J, Ma B, Romano V, Qi H, Ferdousi M, Petropoulos IN, Ponirakis G, Kaye S, Malik RA, Alam U, Zheng Y. An artificial intelligence-based deep learning algorithm for the diagnosis of diabetic neuropathy using corneal confocal microscopy: a development and validation study. *Diabetologia* 2020; **63**: 419-430 [PMID: 31720728 DOI: 10.1007/s00125-019-05023-4]
77 **Norouzi J**, Yadollahpour A, Mirbagheri SA, Mazdeh MM, Hosseini SA. Predicting Renal Failure Progression in Chronic Kidney Disease Using Integrated Intelligent Fuzzy Expert System. *Comput Math Methods Med* 2016; **2016**: 6080814 [PMID: 27022406 DOI: 10.1155/2016/6080814]
78 **Yamada T**, Iwasaki K, Maedera S, Ito K, Takeshima T, Noma H, Shojima N. Myocardial infarction in type 2 diabetes using sodium-glucose co-transporter-2 inhibitors, dipeptidyl peptidase-4 inhibitors or glucagon-like peptide-1 receptor agonists: proportional hazards analysis by deep neural network based machine learning. *Curr Med Res Opin* 2020; **36**: 403-409 [PMID: 31855074 DOI: 10.1080/03007995.2019.1706043]
79 **Churpek MM**, Yuen TC, Winslow C, Hall J, Edelson DP. Differences in vital signs between elderly and nonelderly patients prior to ward cardiac arrest. *Crit Care Med* 2015; **43**: 816-822 [PMID: 25559439 DOI: 10.1097/CCM.0000000000000818]
80 **Beeler C**, Dbeibo L, Kelley K, Thatcher L, Webb D, Bah A, Monahan P, Fowler NR, Nicol S, Judy-Malcolm A, Azar J. Assessing patient risk of central line-associated bacteremia *via* machine learning. *Am J Infect Control* 2018; **46**: 986-991 [PMID: 29661634 DOI: 10.1016/j.ajic.2018.02.021]
81 **Li BY**, Oh J, Young VB, Rao K, Wiens J. Using Machine Learning and the Electronic Health Record to Predict Complicated *Clostridium difficile* Infection. *Open Forum Infect Dis* 2019; **6**: ofz186 [PMID: 31139672 DOI:

10.1093/ofid/ofz186]

82 **Ye J**, Yao L, Shen J, Janarthanam R, Luo Y. Predicting mortality in critically ill patients with diabetes using machine learning and clinical notes. *BMC Med Inform Decis Mak* 2020; **20**: 295 [PMID: 33380338 DOI: 10.1186/s12911-020-01318-4]

83 **Anand RS**, Stey P, Jain S, Biron DR, Bhatt H, Monteiro K, Feller E, Ranney ML, Sarkar IN, Chen ES. Predicting Mortality in Diabetic ICU Patients Using Machine Learning and Severity Indices. *AMIA Jt Summits Transl Sci Proc* 2018; **2017**: 310-319 [PMID: 29888089]

84 **Sadhu AR**, Serrano IA, Xu J, Nisar T, Lucier J, Pandya AR, Patham B. Continuous Glucose Monitoring in Critically Ill Patients With COVID-19: Results of an Emergent Pilot Study. *J Diabetes Sci Technol* 2020; **14**: 1065-1073 [PMID: 33063556 DOI: 10.1177/1932296820964264]

85 **Faulds ER**, Boutsicaris A, Sumner L, Jones L, McNett M, Smetana KS, May CC, Buschur E, Exline MC, Ringel MD, Dungan K. Use of Continuous Glucose Monitor in Critically Ill COVID-19 Patients Requiring Insulin Infusion: An Observational Study. *J Clin Endocrinol Metab* 2021; **106**: e4007-e4016 [PMID: 34100545 DOI: 10.1210/clinem/dgab409]

86 **Agarwal S**, Mathew J, Davis GM, Shephardson A, Levine A, Louard R, Urrutia A, Perez-Guzman C, Umpierrez GE, Peng L, Pasquel FJ. Continuous Glucose Monitoring in the Intensive Care Unit During the COVID-19 Pandemic. *Diabetes Care* 2021; **44**: 847-849 [PMID: 33361145 DOI: 10.2337/dc20-2219]

87 **Dexcom**. Fact sheet for healthcare providers: use of Dexcom continuous glucose monitoring systems during the COVID-19 pandemic. Cited 1 February 2022. Available from: https://www.dexcom.com/hospitalfacts

88 **Galindo RJ**, Umpierrez GE, Rushakoff RJ, Basu A, Lohnes S, Nichols JH, Spanakis EK, Espinoza J, Palermo NE, Awadjie DG, Bak L, Buckingham B, Cook CB, Freckmann G, Heinemann L, Hovorka R, Mathioudakis N, Newman T, O'Neal DN, Rickert M, Sacks DB, Seley JJ, Wallia A, Shang T, Zhang JY, Han J, Klonoff DC. Continuous Glucose Monitors and Automated Insulin Dosing Systems in the Hospital Consensus Guideline. *J Diabetes Sci Technol* 2020; **14**: 1035-1064 [PMID: 32985262 DOI: 10.1177/1932296820954163]

89 **Hinoue T**, Yatabe T, Fujiwara H, Nishida O. Glucose control using an artificial pancreas in a severe COVID-19 patient on extracorporeal membrane oxygenation: a case report. *J Anesth* 2021; **35**: 586-590 [PMID: 34169361 DOI: 10.1007/s00540-021-02965-1]

Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: bpgoffice@wjgnet.com
Help Desk: https://www.f6publishing.com/helpdesk
https://www.wjgnet.com

ISSN 2644-3236 (online)

Artificial Intelligence in *Gastroenterology*

Artif Intell Gastroenterol 2022 June 28; 3(3): 80-95

Published by Baishideng Publishing Group Inc

A I G

Artificial Intelligence in Gastroenterology

Contents

Bimonthly Volume 3 Number 3 June 28, 2022

MINIREVIEWS

Contents

ABOUT COVER

Editorial Board Member of *Artificial Intelligence in Gastroenterology*, Janaina Luz Narciso-Schiavon, PhD, Associate Professor, Department of Internal Medicine, Federal University of Santa Catarina, Florianopolis 88040-900, Santa Catarina, Brazil. janaina.narciso@uol.com.br

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

The *AIG* is now abstracted and indexed in Reference Citation Analysis, China Science and Technology Journal Database.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: *Wen-Wen Qi*; **Production Department Director:** *Xiang Li*; **Editorial Office Director:** *Jin-Lei Wang*.

NAME OF JOURNAL
Artificial Intelligence in Gastroenterology

ISSN
ISSN 2644-3236 (online)

LAUNCH DATE
July 28, 2020

FREQUENCY
Bimonthly

EDITORS-IN-CHIEF
Rajvinder Singh, Ferruccio Bonino

EDITORIAL BOARD MEMBERS
https://www.wjgnet.com/2644-3236/editorialboard.htm

PUBLICATION DATE
June 28, 2022

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

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 June 28; 3(3): 80-87

DOI: 10.35712/aig.v3.i3.80

ISSN 2644-3236 (online)

MINIREVIEWS

Machine learning approaches using blood biomarkers in non-alcoholic fatty liver diseases

Randhall B Carteri, Mateus Grellert, Daniela Luisa Borba, Claudio Augusto Marroni, Sabrina Alves Fernandes

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B, B
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): E

P-Reviewer: Cengiz M, Turkey; Gadekallu TR, India; Masarone M, Italy; Portius D, Germany

Received: December 31, 2021
Peer-review started: December 31, 2021
First decision: March 28, 2022
Revised: April 15, 2022
Accepted: May 8, 2022
Article in press: May 8, 2022
Published online: June 28, 2022

Randhall B Carteri, Department of Nutrition, Methodist University Center - IPA, Porto Alegre 90420-060, Rio Grande do Sul, Brazil

Randhall B Carteri, Department of Health and Behaviour, Catholic University of Pelotas, Pelotas 96015-560, Rio Grande do Sul, Brazil

Mateus Grellert, Department of Informatics and Statistics, Federal University of Santa Catarina, Florianópolis 88040-900, Santa Catarina, Brazil

Daniela Luisa Borba, Sabrina Alves Fernandes, Postgraduate Program in Hepatology, Federal University of Health Sciences of Porto Alegre, Porto Alegre 90050-170, Rio Grande do Sul, Brazil

Claudio Augusto Marroni, Department of Gastroenterology and Hepatology, Federal University of Health Sciences of Porto Alegre, Porto Alegre 90050-170, Rio Grande do Sul, Brazil

Corresponding author: Sabrina Alves Fernandes, PhD, Research Scientist, Teacher, Post-graduate Program in Hepatology, Federal University of Health Sciences of Porto Alegre, Sarmento Leite Street, 245 - Centro Histórico, Porto Alegre 90050-170, Rio Grande do Sul, Brazil. sabrinaafernandes@gmail.com

Abstract

The prevalence of nonalcoholic fatty liver disease (NAFLD) is an important public health concern. Early diagnosis of NAFLD and potential progression to nonalcoholic steatohepatitis (NASH), could reduce the further advance of the disease, and improve patient outcomes. Aiming to support patient diagnostic and predict specific outcomes, the interest in artificial intelligence (AI) methods in hepatology has dramatically increased, especially with the application of less-invasive biomarkers. In this review, our objective was twofold: Firstly, we presented the most frequent blood biomarkers in NAFLD and NASH and secondly, we reviewed recent literature regarding the use of machine learning (ML) methods to predict NAFLD and NASH in large cohorts. Strikingly, these studies provide insights into ML application in NAFLD patients' prognostics and ranked blood biomarkers are able to provide a recognizable signature allowing cost-effective NAFLD prediction and also differentiating NASH patients. Future studies should consider the limitations in the current literature and expand the application of these algorithms in different populations, fortifying an already promising tool in medical science.

Key Words: Artificial intelligence; Liver diseases; Healthcare; Hepatology; Prognosis; Diagnostics

Core Tip: The ability of machine learning approaches to process multiple variables, map linear and nonlinear interactions, ranking the most important features, in addition to the capability of building accurate prediction models, sets a future direction to its application in complex diseases such as nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. Future studies should consider the limitations in the current literature and expand the application of these algorithms in different populations, fortifying an already promising tool in medical science.

Citation: Carteri RB, Grellert M, Borba DL, Marroni CA, Fernandes SA. Machine learning approaches using blood biomarkers in non-alcoholic fatty liver diseases. *Artif Intell Gastroenterol* 2022; 3(3): 80-87
URL: https://www.wjgnet.com/2644-3236/full/v3/i3/80.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i3.80

INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD) affects an expressive part of the population worldwide and is a major cause of liver-disease related morbidity[1]. The most common cause of death in NAFLD patients is related to cardiovascular diseases, which is partially explained by the presence of metabolic comorbidities, such as obesity, type 2 diabetes, dyslipidemia, and hypertension[2]. Recently, there was concordance that the term NAFLD cannot represent the multisystemic metabolic disruption associated with the disease, resulting in the novel term MAFLD - metabolic associated fatty liver disease. Moreover, MAFLD considers the hepatic manifestation of a multimodal disease that is heterogeneous in its causes, symptoms, progression, and outcomes[3]. Nevertheless, the progression of liver fibrosis could lead to Nonalcoholic steatohepatitis (NASH), a condition characterized by histological lobular inflammation and hepatocyte ballooning[2]. Hence, detecting possible elements related to a worse prognosis in these conditions in the early stages of the disease could improve the treatment and its efficiency. Considering the significance of advanced fibrosis in NAFLD patients, differentiating NASH from steatosis is vital, reinforcing the need for cost-effective methods for risk stratification in this population [4]. Although liver biopsy is widely considered the gold standard in liver diseases investigation, it is also invasive, expensive, and prone to sampling error. In this context, the use of non-invasive biomarkers gains considerable importance[5].

The interest in artificial intelligence (AI) methods in different medical specialties, including hepatology, has dramatically increased during the last decade[6]. Advances in technology and data acquisition have simplified the collection and storage of large data sets with long time series, leading to increasingly varied fields of application, including biomedical areas. In this context, large-volume data mining evaluations had been showing promising results in recent clinical studies using machine learning methods[7-9]. More specifically, supervised machine learning (SML), can automatically detect patterns in existing training data and then use the detected patterns to predict future data[6]. Rather than considering differences between groups (as traditional statistical comparisons do), SML methods address individual differences, classifying individuals in ways that contribute to the clinical decision-making process.

The commonly late diagnosis of liver disorders contributes to suboptimal treatment and poor results. More specifically, as the prevalence of NAFLD is an important public health concern, early diagnosis of NAFLD and potential progression to NASH, could reduce the further advance of the disease, and improve patient outcomes. Using SML methods allows for collecting patient data and identifying their profile regarding the risk of developing comorbidities associated with liver damage, such as the development of metabolic syndrome or even predicting the patient's prognosis. Several recent reviews highlighted the application of artificial intelligence in hepatology, while broadly discussing how different approaches present potential applications in several areas of hepatology[10-12]. However, specific discussion of machine learning approaches using cost-effective biomarkers could help to guide future studies towards the improvement of NAFLD diagnosis. Therefore, the objective of this mini-review is to discuss the application of SML approaches using biomarkers for the diagnosis of NAFLD and the prediction of NASH presence.

BLOOD BIOMARKERS IN NAFLD

Biomarkers are a "defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes or responses to an exposure or intervention". This includes a plethora of possible assessments commonly investigated in NAFLD, such as blood profile, imaging (histological/radiographic) exams, specific anthropometric characteristics (body composition), and also phase angle derived from bioimpedance[13]. Noteworthy, blood biomarkers are a less invasive approach from a biological point of view and could complement imaging techniques to improve disease monitoring. In clinical settings, liver biopsy is the diagnostic gold standard for NAFLD, allowing the assessment of lipid content, inflammation, hepatocellular ballooning, and fibrotic alterations, which can also determine NASH diagnostics[14]. However, non-invasive techniques provide limited inflammation and hepatocellular ballooning determination, making objective biomarker panels for the assessment and monitoring of NAFLD or NASH a current challenge[14,15].

Nevertheless, abnormal liver function is often initially identified by nonspecific hepatocellular damage through elevations in alanine aminotransferase (ALT) and aspartate aminotransferase (AST) in addition to alkaline phosphatase and gamma-glutamyl transferase (GGT)[16]. However, ALT and AST can present normal levels while GGT can present a 1.5 - fold elevation, and this response does not reflect hepatic inflammation, fibrosis, or patient metabolic risks[17,18]. Recently, cytokeratin (CK)-18 gained attention as a more specific approach for hepatocyte apoptosis since CK-18 is a major intermediate filament protein cleaved by caspases creating fragments during the apoptotic processes[19]. Assays of CK-18 fragments provide moderate accuracy due to high variability between cut-offs and respective diagnostic accuracy among studies[19]. More specifically, M30 measures caspase-cleaved CK18 produced during apoptosis, and M65 measures the total levels of (both cleaved and intact) CK18[20]. The CK-18 fragments could independently predict NAFLD severity and detect the presence of NASH with a specificity close to 90%[21,22]. In a large and heterogeneous cohort, the blood concentration of CK-18 fragments of patients with NAFLD was higher when compared with healthy volunteers and correlated to several biomarkers of liver damage and steatosis[22]. Moreover, several "biomarker panels" to grade NAFLD patients' steatosis and fibrosis through specific scores comprise different biomarker combinations, summarized in Table 1. Notably, the *FibroTest*, *Fibrometer*, *Hepascore*, and *Enhanced Liver Fibrosis* scores are patented and commercially available panels. Nevertheless, most of the biomarker panels for the diagnosis of NAFLD and NASH, lack validation in specific cohorts, such as bariatric patients and patients with varying ethnicities[23,24]. Further, recent evidence reinforces that a combination of different commonly assessed blood-based biomarkers in addition to direct fibrogenesis markers can provide higher diagnostic accuracy in detecting advanced fibrosis when compared to current protocols. The study of Vilar-Gomez *et al*[25], reviewed the diagnostic accuracy of several blood-based biomarkers, suggesting an algorithm to diagnose NAFLD patients at risk of fibrosis development. Additionally, the European guidelines recommend the combination of different tests to assess NAFLD, stating that the *Fibrometer* is a non-invasive alternative to liver biopsy, albeit the guidelines are not clear regarding which specific version of the *FibroMeter* is preferred[26]. Also, the commercially available biomarker panels and other complementary methods are not accessible for most health services, justifying the search for alternative approaches[25].

The validation study by Wu *et al*[27] compared different panels of biomarkers in 417 NAFLD patients (156 with advanced fibrosis), showing that when predicting liver fibrosis scores Fibrosis-4 (FIB-4), NAFLD Fibrosis Score (NFS), AST to Platelet Ratio Index (APRI) and BARD score (BARD), it is possible to obtain a prediction of moderate fibrosis based on the receptor operator area under the curve (AUROC; 0.724, 0.671 and 0.609, respectively). The authors argued that FIB-4 and NFS performed better compared to both APRI and BARD scores, which resulted in high false-positive rates. Importantly, this study evaluated NAFLD patients based on the new definition of MAFLD, highlighting that the investigated biomarker panels provided poor performance in this setting[27]. In conclusion, the fact that the aforementioned biomarkers come from different types of procedures makes it hard for human experts to jointly analyze all this information, which motivates the use of machine learning techniques. These models can work with different types of data and discovering the relationship between them to obtain a better prediction.

ARTIFICIAL INTELLIGENCE APPLICATION IN NAFLD

Briefly, AI is an umbrella term, referring to a structured utilization of software and algorithms that analyze a wide range of data, ultimately simulating human cognition and intelligence[6]. Machine learning (ML) is one of the subdisciplines of AI, focusing on learning from data and associating specific patterns with different outcomes. An important advantage of ML techniques is that they allow the modeling of complex problems that depend on multiple input variables, justifying the application of ML methods to potentially fill several gaps in the study of complex diseases, such as NAFLD[6]. This is especially important in the case of NAFLD, which is closely related to metabolic disturbances associated with obesity and metabolic syndrome[28]. Given its complexity, NAFLD presents in different forms,

Table 1 Blood biomarker panels for nonalcoholic fatty liver disease and nonalcoholic steatohepatitis

Blood biomarker panels for steatosis			
Panel	**Patient**	**Anthropometry**	**Blood biomarkers**
FLI	-	BMI, Waist circumference	GGT and TG
HSI	Presence of DM	BMI	AST:ASL
Steatotest	Sex	BMI	ALT, GGT, TG, A2M, ApoA1, haptoglobin, bilirubin,cholesterol, and glucose
LAP	Sex	Waist circumference	TG
ION	Sex	Waist to hip ratio	ALT, TG
NAFLD LFS	Presence of DM and MS	-	AST:ALT, Insulin
Blood biomarker panels for fibrosis			
Panel	**Patient**	**Anthropometry**	**Blood biomarkers**
APRI	-	-	Platelet count, AST
FIB-4	Age	-	Platelet count, AST, ALT
FibroTest	Age, sex	BMI	GGT, A2M, ApoA1, haptoglobin, and total bilirubin
Fibrometer	Age	Body weight	Platelet count, AST, ALT, glucose, ferritin
ELF	-	-	Hyaluronic acid, PIIINP and TIMP-1
Hepascore	Age, sex	-	GGT, Hyaluronic acid, PIIINP and TIMP-1
BARD	Presence of DM	BMI	AST:ALT
NFS	Age, sex, Presence of DM	-	Platelet count, AST:ALT, Albumin

A2M: Alpha-2-macroglobulin; ALT: alanine aminotransferase; ApoA1: Apolipoprotein A1; AST: Aspartate aminotransferase; BMI: body mass index; DM: Diabetes mellitus; GGT: gamma-glutamyl transpeptidase; MS: Metabolic syndrome; NAFLD: Nonalcoholic fatty liver disease; PIIINP: Amino-terminal propeptide of type III procollagen; TG: Triglycerides; TIMP1: tissue inhibitor of matrix metalloproteinases-1.

from simple asymptomatic lipid accumulation to symptomatic non-alcoholic steatohepatitis (NASH) characterized by several factors, including steatosis, hepatocellular ballooning, lobular inflammation, and often fibrosis[28]. Machine learning methods are becoming increasingly popular, which has also motivated an increase in the complexity of these models. Particularly, deep learning (DL) models, like convolutional neural networks (CNN), showed promising results in hepatology, especially with high-resolution data such as images and spectrograms[29]. Likewise, CNN models encompass several layers that involve operations like convolution, pooling, and nonlinear activations, making their decisions difficult to understand. Therefore, they represent black-box models, as opposed to interpretable (white-box) techniques, such as regression/decision trees and Bayesian networks[30,31]. Hence, ML could identify patients at risk and guide clinical treatments, whilst considering that the clinical manifestations of NAFLD appear in advanced disease status and the availability and cost of screening methods for the clinicians. Also, ML can help to rank and categorize specific biomarkers and help to elaborate specific "disease signatures", contributing not only to clinical diagnostics, but also provide mechanistic insights for the study of the disease and the development of specific treatments.

MACHINE LEARNING APPROACHES USING BLOOD BIOMARKERS IN HEPATOLOGY

As stated above, the interest in using AI approaches to support clinical decision-making processes in hepatology has increased, albeit current literature is still scarce. Table 2 summarizes the specific studies addressing NAFLD and NASH classification. Initially, the study of Sowa *et al*[32] showed no differences in the investigated biomarkers (ALT, AST, and apoptotic signaling) between patients with a fibrosis score of 1 or 2. However, combining these parameters using random forests (RF) reached 79% accuracy in fibrosis prediction with a sensitivity of more than 60% and specificity of 77%. Moreover, RF identified the cell death markers M30 and M65 as more important for the decision than the classic liver parameters. Similarly, Yip *et al*[33] built a model to predict steatosis in a study including 922 individuals with assessment for NAFLD. The four models developed presented good diagnostic precision for steatosis (AUROC was 0.87-0.9), albeit the authors claimed that the "NAFLD ridge score" offered the best balance between efficacy and simplicity. This model included six parameters (serum triglycerides, alanine aminotransferase, high-density lipoprotein cholesterol, hemoglobin A1c, white cell count, and

Table 2 Machine learning studies in nonalcoholic fatty liver disease and nonalcoholic steatohepatitis patients

Ref.	Patients	Investigated biomarker	Model with best performance	Results
Sowa *et al* [32], 2013	126 patients	Alanine aminotransferase; Aspartate aminotransferase; M30; M60; Hyaluronic acid	Randon forest	79% Accuracy in fibrosis prediction; 60% sensitivity; 77% specificity
Yip *et al* [33], 2017	922 patients	Alanine aminotransferase; High-density lipoprotein cholesterol; Triglycerides; HbA1c; White blood cells; Hypertension	Ridge score	88% Accuracy in steatosis prediction; 92% sensitivity; 90% specificity
Ma *et al* [34], 2018	10.508 patients; 2522 NAFLD patients	Age; Sex; Body mass index; Alanine aminotransferase; Aspartate aminotransferase; Alkaline phosphatase; Gamma-glutamyl transpeptidase; Triglycerides; Blood urea nitrogen; Bilirubin; Cholesterol; Creatinine; Fasting glucose; Uric acid	Bayesian network model	83% Accuracy in NAFLD prediction; 68% sensitivity; 94% specificity
Canbay *et al*[35], 2019	164 patients; 122 (validation)	Age; HbA1c; Gamma-glutamyl transpeptidase; M30; Adiponectin	Logistic regression	70% Accuracy in separate NAFLD and NASH
Liu *et al* [36], 2021	15.315 patients5878 with NAFLD	Body mass index; Waist circumference; Waist-to-height ratio; Alanine aminotransferase; Fasting blood glucose; Gamma-glutamyl transpeptidase; Very-low-density lipoprotein cholesterol; Low-density lipoprotein cholesterol; High-density lipoprotein cholesterol; Systolic blood pressure; Alkaline phosphatase; Diastolic blood pressure	XGBoost model	79% Accuracy in NAFLD prediction; 61% sensitivity; 90% specificity
Pei *et al* [37], 2021	3.419 patients; 845 with fat liver diseases	Age; Height; Hemoglobin; Aspartate aminotransferase; Glucose; Uric acid; Low-density lipoprotein; Alpha-fetoprotein; Triglycerides; High-density lipoprotein; Carcinoembryonic antigen	XGBoost model	94% accuracy of prediction; 90% sensitivity; 95% specificity

NAFLD: Nonalcoholic fatty liver disease; XGBoost: Extreme gradient boosting.

the presence of hypertension) that are routinely available for individuals undergoing medical checkups, and it does not require anthropometric measures, which are not always available. Although there is evident feasibility of the NAFLD ridge score to screen individuals, it still needs additional validation in other ethnicities. The study of Ma *et al*[34], investigated the predictive power for NAFLD of eleven machine learning techniques, demonstrating that the Bayesian network model had the best performance, revealing that the five most discriminating features (based on information gain scores) to be weight, TG, ALT, GGT, and serum uric acid levels. Thus, in practice, users could focus on these features. Furthermore, Canbay *et al*[35] compared different scores for the non-invasive detection of NASH. Briefly, using an ensemble feature selection approach for biomarker selection, the authors built a logistic regression model and validated in an independent study cohort of 122 patients. The logistic regression model generated from age, GGT, hemoglobin A1c, M30, and adiponectin had a strong correlation with the non-alcoholic steatohepatitis activity score and demonstrated reasonable performance to discriminate between NAFL and NASH. Likewise, Liu *et al*[36] performed a retrospective cross-sectional study on 15315 Chinese subjects, where 5878 patients presented NAFLD. The biomarker ranking indicated the body mass index as the most valuable indicator to predict NAFLD, followed by waist circumference, triglycerides, waist-to-height ratio, and alanine aminotransferase. Notably, among seven machine learning models, the extreme gradient boosting (XGBoost) model demonstrated the best prediction ability. Similarly, the XGBoost also presented the highest AUC (0.93), accuracy (0.94), and sensitivity value (0.90) in the study of Pei *et al*[37], comparing different models for predicting fatty liver Disease risk in 3419 participants, of which 845 had diagnostic confirmation. Importantly, regarding the biomarkers, uric acid, body mass index, and triglycerides were the most decisive risk factors for the ML models, whilst high-density lipoprotein and hemoglobin also counted as important risk factors for prediction. Strikingly, these studies provide insights into ML application in a complex context such as NAFLD patients' prognostics. Notably, while there are investigations using AI techniques and common biomarkers to predict NAFLD and NASH, approaches using AI and novel proposed biomarkers are scarce. For instance, a recent meta-analysis showed that CK-18 is the only marker for NASH presenting external validation, with an AUROC of 0.82[38]. Conversely, a large study conducted by the multicenter NASH Clinical Research Network demonstrated that the addition of routinely available clinical-laboratory parameters to CK-18 measurement did not significantly improve its diagnostic performance[22]. However, it remains unknown whether the use of AI techniques combining different biomarkers in a large and diverse cohort could provide different results. Taken together, the data suggests that ranked blood biomarkers can provide a recognizable signature allowing cost-effective NAFLD prediction and also differentiating NASH patients.

CURRENT CHALLENGES IN SML APPROACHES IN HEPATOLOGY

The term "AI-Chasm" describes the gap between developing and testing an algorithm and the definitive application of the algorithm in clinical practice[39]. Unequivocally, the AI application in medical sciences is auspicious, and current literature is shading light on a plethora of potential applications; however, many challenges for SML approaches using biomarkers in hepatology still await scrutiny.

Firstly, the collection, curation, and preprocessing of patient data is a major concern, since SML methods are data-driven[10]. Notably, the cited studies in this mini-review provide relatively small data from specific populations which could lead to sampling bias whilst limiting the generalization of the obtained results. Further, data collection should be standardized and precise, but should also be monitored for privacy and data security breaches. Secondly, as recently discussed by Quinn *et al*[40], one of the main aspects of concern in future studies is the understanding that transdisciplinary approaches require cooperation to build a conceptually appropriate framework while also focusing on evaluating the performance of SML algorithms in terms of clinical endpoints and not just predictive accuracy. In addition to these technical challenges, there is also an increasing demand for transparency concerning the predictions of these models, especially in areas that have no computing background. For instance, healthcare professionals and other stakeholders that can benefit from these solutions are still reluctant to the idea of employing these methods, evidencing the necessity of educational programs aimed to explicit information about the involved decision processes. Nevertheless, the field of explainable AI has emerged to address these issues, with the purpose of creating ML techniques that produce explainable models while maintaining a high level of learning performance, enabling humans to understand and trust the predictions to support their decisions[41].

CONCLUSION

Recent advances in the field of biosciences applying machine learning algorithms resulted in promising results for the diagnosis of disease and biomarker study. The main idea is that SML could overcome the limitations of common statistical techniques. For instance, SML identifies data patterns for classification, considering multiple features at once, allowing the ranking and selection of the available blood biomarkers related to disease pathogenesis for the prediction of NAFLD or NASH, minimizing potential errors between the predicted values and the real data. Although the cited studies provide promising results, there are specific limitations that future studies should reduce. For example, most of the studies involved the Chinese population, and these algorithms still need additional validation in heterogeneous populations. The strong association between NAFLD and metabolic syndrome, obesity, and alcohol consumption may be a confounding factor in previous studies, and the application of these methods in diabetic patients with and without NAFLD could shed light on the influence of specific treatments on the performance of these ML methods. Nevertheless, the ability of ML approaches to process multiple variables, map linear and nonlinear interactions, and rank the most important features, in addition to the capability of building accurate prediction models, sets a future direction to its application in complex diseases, including NAFLD and NASH. Future studies should consider the limitations in the current literature and expand the application of these algorithms in different populations, fortifying an already promising tool in medical science.

FOOTNOTES

Author contributions: Carteri RB organized the structuring of the article in the order of the subheadings covered; Grellert M and Borba DL contributed to manuscript writing; Maroni CA and Fernandes SA supervised and revised the manuscript; All authors contributed to the writing of the article and review of the scientific article.

Conflict-of-interest statement: All authors declare that there are no conflicts of interest related to this article.

Country/Territory of origin: Brazil

ORCID number: Randhall B Carteri 0000-0003-4124-9470; Mateus Grellert 0000-0003-0600-7054; Daniela Luisa Borba (0000-0002- 2141-3993; Claudio Augusto Marroni 0000-0002-1718-6548; Sabrina Alves Fernandes 0000-0001-8504-603X.

S-Editor: Liu JH

L-Editor: Filipodia
P-Editor: Liu JH

REFERENCES

1 **Nd AM**. Non-Alcoholic Fatty Liver Disease, an Overview. *Integr Med (Encinitas)* 2019; **18**: 42-49 [PMID: 31341444]
2 **Buzzetti E**, Pinzani M, Tsochatzis EA. The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD). *Metabolism* 2016; **65**: 1038-1048 [PMID: 26823198 DOI: 10.1016/j.metabol.2015.12.012]
3 **Geier A**, Tiniakos D, Denk H, Trauner M. From the origin of NASH to the future of metabolic fatty liver disease. *Gut* 2021 [PMID: 33632710 DOI: 10.1136/gutjnl-2020-323202]
4 **Juanola O**, Martínez-López S, Francés R, Gómez-Hurtado I. Non-Alcoholic Fatty Liver Disease: Metabolic, Genetic, Epigenetic and Environmental Risk Factors. *Int J Environ Res Public Health* 2021; **18** [PMID: 34069012 DOI: 10.3390/ijerph18105227]
5 **Chalasani N**, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, Harrison SA, Brunt EM, Sanyal AJ. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* 2018; **67**: 328-357 [PMID: 28714183 DOI: 10.1002/hep.29367]
6 **Aggarwal P**, Alkhouri N. Artificial Intelligence in Nonalcoholic Fatty Liver Disease: A New Frontier in Diagnosis and Treatment. *Clin Liver Dis (Hoboken)* 2021; **17**: 392-397 [PMID: 34386201 DOI: 10.1002/cld.1071]
7 **Bertsimas D**, Kung J, Trichakis N, Wang Y, Hirose R, Vagefi PA. Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation. *Am J Transplant* 2019; **19**: 1109-1118 [PMID: 30411495 DOI: 10.1111/ajt.15172]
8 **Shousha HI**, Awad AH, Omran DA, Elnegouly MM, Mabrouk M. Data Mining and Machine Learning Algorithms Using IL28B Genotype and Biochemical Markers Best Predicted Advanced Liver Fibrosis in Chronic Hepatitis C. *Jpn J Infect Dis* 2018; **71**: 51-57 [PMID: 29279441 DOI: 10.7883/yoken.JJID.2017.089]
9 **Konerman MA**, Lu D, Zhang Y, Thomson M, Zhu J, Verma A, Liu B, Talaat N, Balis U, Higgins PDR, Lok ASF, Waljee AK. Assessing risk of fibrosis progression and liver-related clinical outcomes among patients with both early stage and advanced chronic hepatitis C. *PLoS One* 2017; **12**: e0187344 [PMID: 29108017 DOI: 10.1371/journal.pone.0187344]
10 **Spann A**, Yasodhara A, Kang J, Watt K, Wang B, Goldenberg A, Bhat M. Applying Machine Learning in Liver Disease and Transplantation: A Comprehensive Review. *Hepatology* 2020; **71**: 1093-1105 [PMID: 31907954 DOI: 10.1002/hep.31103]
11 **Balsano C**, Alisi A, Brunetto MR, Invernizzi P, Burra P, Piscaglia F; Special Interest Group (SIG) Artificial Intelligence and Liver Diseases; Italian Association for the Study of the Liver (AISF). The application of artificial intelligence in hepatology: A systematic review. *Dig Liver Dis* 2022; **54**: 299-308 [PMID: 34266794 DOI: 10.1016/j.dld.2021.06.011]
12 **Su TH**, Wu CH, Kao JH. Artificial intelligence in precision medicine in hepatology. *J Gastroenterol Hepatol* 2021; **36**: 569-580 [PMID: 33709606 DOI: 10.1111/jgh.15415]
13 **Califf RM**. Biomarker definitions and their applications. *Exp Biol Med (Maywood)* 2018; **243**: 213-221 [PMID: 29405771 DOI: 10.1177/1535370217750088]
14 **Younossi ZM**, Loomba R, Anstee QM, Rinella ME, Bugianesi E, Marchesini G, Neuschwander-Tetri BA, Serfaty L, Negro F, Caldwell SH, Ratziu V, Corey KE, Friedman SL, Abdelmalek MF, Harrison SA, Sanyal AJ, Lavine JE, Mathurin P, Charlton MR, Goodman ZD, Chalasani NP, Kowdley KV, George J, Lindor K. Diagnostic modalities for nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, and associated fibrosis. *Hepatology* 2018; **68**: 349-360 [PMID: 29222917 DOI: 10.1002/hep.29721]
15 **Castera L**, Friedrich-Rust M, Loomba R. Noninvasive Assessment of Liver Disease in Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2019; **156**: 1264-1281.e4 [PMID: 30660725 DOI: 10.1053/j.gastro.2018.12.036]
16 **Neuman MG**, Cohen LB, Nanau RM. Biomarkers in nonalcoholic fatty liver disease. *Can J Gastroenterol Hepatol* 2014; **28**: 607-618 [PMID: 25575111 DOI: 10.1155/2014/757929]
17 **Fracanzani AL**, Valenti L, Bugianesi E, Andreoletti M, Colli A, Vanni E, Bertelli C, Fatta E, Bignamini D, Marchesini G, Fargion S. Risk of severe liver disease in nonalcoholic fatty liver disease with normal aminotransferase levels: a role for insulin resistance and diabetes. *Hepatology* 2008; **48**: 792-798 [PMID: 18752331 DOI: 10.1002/hep.22429]
18 **Mofrad P**, Contos MJ, Haque M, Sargeant C, Fisher RA, Luketic VA, Sterling RK, Shiffman ML, Stravitz RT, Sanyal AJ. Clinical and histologic spectrum of nonalcoholic fatty liver disease associated with normal ALT values. *Hepatology* 2003; **37**: 1286-1292 [PMID: 12774006 DOI: 10.1053/jhep.2003.50229]
19 **He L**, Deng L, Zhang Q, Guo J, Zhou J, Song W, Yuan F. Diagnostic Value of CK-18, FGF-21, and Related Biomarker Panel in Nonalcoholic Fatty Liver Disease: A Systematic Review and Meta-Analysis. *Biomed Res Int* 2017; **2017**: 9729107 [PMID: 28326329 DOI: 10.1155/2017/9729107]
20 **de Haas EC**, di Pietro A, Simpson KL, Meijer C, Suurmeijer AJ, Lancashire LJ, Cummings J, de Jong S, de Vries EG, Dive C, Gietema JA. Clinical evaluation of M30 and M65 ELISA cell death assays as circulating biomarkers in a drug-sensitive tumor, testicular cancer. *Neoplasia* 2008; **10**: 1041-1048 [PMID: 18813353 DOI: 10.1593/neo.08620]
21 **Tamimi TI**, Elgouhari HM, Alkhouri N, Yerian LM, Berk MP, Lopez R, Schauer PR, Zein NN, Feldstein AE. An apoptosis panel for nonalcoholic steatohepatitis diagnosis. *J Hepatol* 2011; **54**: 1224-1229 [PMID: 21145805 DOI: 10.1016/j.jhep.2010.08.023]
22 **Feldstein AE**, Wieckowska A, Lopez AR, Liu YC, Zein NN, McCullough AJ. Cytokeratin-18 fragment levels as noninvasive biomarkers for nonalcoholic steatohepatitis: a multicenter validation study. *Hepatology* 2009; **50**: 1072-1078 [PMID: 19585618 DOI: 10.1002/hep.23050]
23 Fallatah HI. Noninvasive Biomarkers of Liver Fibrosis: An Overview. *Adv Hepatol* 2014; 357287 [DOI: 10.1155/2014/357287]

24 **Soresi M**, Giannitrapani L, Cervello M, Licata A, Montalto G. Non invasive tools for the diagnosis of liver cirrhosis. *World J Gastroenterol* 2014; **20**: 18131-18150 [PMID: 25561782 DOI: 10.3748/wjg.v20.i48.18131]

25 **Vilar-Gomez E**, Chalasani N. Non-invasive assessment of non-alcoholic fatty liver disease: Clinical prediction rules and blood-based biomarkers. *J Hepatol* 2018; **68**: 305-315 [PMID: 29154965 DOI: 10.1016/j.jhep.2017.11.013]

26 **European Association for the Study of the Liver (EASL)**; European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol* 2016; **64**: 1388-1402 [PMID: 27062661 DOI: 10.1016/j.jhep.2015.11.004]

27 **Wu YL**, Kumar R, Wang MF, Singh M, Huang JF, Zhu YY, Lin S. Validation of conventional non-invasive fibrosis scoring systems in patients with metabolic associated fatty liver disease. *World J Gastroenterol* 2021; **27**: 5753-5763 [PMID: 34629799 DOI: 10.3748/wjg.v27.i34.5753]

28 **Byrne CD**, Targher G. NAFLD: a multisystem disease. *J Hepatol* 2015; **62**: S47-S64 [PMID: 25920090 DOI: 10.1016/j.jhep.2014.12.012]

29 **LeCun Y**, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436-444 [PMID: 26017442 DOI: 10.1038/nature14539]

30 **Ribeiro E**, Uhl A, Wimmer G, Häfner M. Exploring Deep Learning and Transfer Learning for Colonic Polyp Classification. *Comput Math Methods Med* 2016; **2016**: 6584725 [PMID: 27847543 DOI: 10.1155/2016/6584725]

31 **Bernal J**, Tajkbaksh N, Sanchez FJ, Matuszewski BJ, Hao Chen, Lequan Yu, Angermann Q, Romain O, Rustad B, Balasingham I, Pogorelov K, Sungbin Choi, Debard Q, Maier-Hein L, Speidel S, Stoyanov D, Brandao P, Cordova H, Sanchez-Montes C, Gurudu SR, Fernandez-Esparrach G, Dray X, Jianming Liang, Histace A. Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results From the MICCAI 2015 Endoscopic Vision Challenge. *IEEE Trans Med Imaging* 2017; **36**: 1231-1249 [PMID: 28182555 DOI: 10.1109/TMI.2017.2664042]

32 **Sowa JP**, Heider D, Bechmann LP, Gerken G, Hoffmann D, Canbay A. Novel algorithm for non-invasive assessment of fibrosis in NAFLD. *PLoS One* 2013; **8**: e62439 [PMID: 23638085 DOI: 10.1371/journal.pone.0062439]

33 **Yip TC**, Ma AJ, Wong VW, Tse YK, Chan HL, Yuen PC, Wong GL. Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease (NAFLD) in the general population. *Aliment Pharmacol Ther* 2017; **46**: 447-456 [PMID: 28585725 DOI: 10.1111/apt.14172]

34 **Ma H**, Xu CF, Shen Z, Yu CH, Li YM. Application of Machine Learning Techniques for Clinical Predictive Modeling: A Cross-Sectional Study on Nonalcoholic Fatty Liver Disease in China. *Biomed Res Int* 2018; **2018**: 4304376 [PMID: 30402478 DOI: 10.1155/2018/4304376]

35 **Canbay A**, Kälsch J, Neumann U, Rau M, Hohenester S, Baba HA, Rust C, Geier A, Heider D, Sowa JP. Non-invasive assessment of NAFLD as systemic disease-A machine learning perspective. *PLoS One* 2019; **14**: e0214436 [PMID: 30913263 DOI: 10.1371/journal.pone.0214436]

36 **Liu YX**, Liu X, Cen C, Li X, Liu JM, Ming ZY, Yu SF, Tang XF, Zhou L, Yu J, Huang KJ, Zheng SS. Comparison and development of advanced machine learning tools to predict nonalcoholic fatty liver disease: An extended study. *Hepatobiliary Pancreat Dis Int* 2021; **20**: 409-415 [PMID: 34420885 DOI: 10.1016/j.hbpd.2021.08.004]

37 **Pei X**, Deng Q, Liu Z, Yan X, Sun W. Machine Learning Algorithms for Predicting Fatty Liver Disease. *Ann Nutr Metab* 2021; **77**: 38-45 [PMID: 33849025 DOI: 10.1159/000513654]

38 **Musso G**, Gambino R, Cassader M, Pagano G. Meta-analysis: natural history of non-alcoholic fatty liver disease (NAFLD) and diagnostic accuracy of non-invasive tests for liver disease severity. *Ann Med* 2011; **43**: 617-649 [PMID: 21039302 DOI: 10.3109/07853890.2010.518623]

39 **Keane PA**, Topol EJ. With an eye to AI and autonomous diagnosis. *NPJ Digit Med* 2018; **1**: 40 [PMID: 31304321 DOI: 10.1038/s41746-018-0048-y]

40 **Quinn TP**, Senadeera M, Jacobs S, Coghlan S, Le V. Trust and medical AI: the challenges we face and the expertise needed to overcome them. *J Am Med Inform Assoc* 2021; **28**: 890-894 [PMID: 33340404 DOI: 10.1093/jamia/ocaa268]

41 **Barredo Arrieta A**, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 2020; 82 [DOI: 10.1016/j.inffus.2019.12.012]

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 June 28; 3(3): 88-95

DOI: 10.35712/aig.v3.i3.88

ISSN 2644-3236 (online)

MINIREVIEWS

Artificial intelligence using advanced imaging techniques and cholangiocarcinoma: Recent advances and future direction

Aaron R Brenner, Passisd Laoveeravat, Patrick J Carey, Danielle Joiner, Samuel H Mardini, Manol Jovani

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Morilla I, France; Naik N, India

Received: March 7, 2022
Peer-review started: March 7, 2022
First decision: April 10, 2022
Revised: April 16, 2022
Accepted: May 5, 2022
Article in press: May 5, 2022
Published online: June 28, 2022

Aaron R Brenner, Patrick J Carey, Danielle Joiner, Department of Internal Medicine, University of Kentucky College of Medicine, Lexington, KY 40536, United States

Passisd Laoveeravat, Division of Digestive Diseases and Nutrition, University of Kentucky College of Medicine, Lexington, KY 40536, United States

Samuel H Mardini, Division of Digestive Diseases and Nutrition, University of Kentucky College of Medicine, Lexington, KENTUCKY 40536, United States

Manol Jovani, Digestive Diseases and Nutrition, University of Kentucky Albert B. Chandler Hospital, Lexington, KY 40536, United States

Corresponding author: Manol Jovani, MD, MSc, Assistant Professor, Attending Doctor, Doctor, Digestive Diseases and Nutrition, University of Kentucky Albert B. Chandler Hospital, 770 Rose St Room MN662, Lexington, KY 40536, United States. manol.jovani@mail.harvard.edu

Abstract

While cholangiocarcinoma represents only about 3% of all gastrointestinal tumors, it has a dismal survival rate, usually because it is diagnosed at a late stage. The utilization of Artificial Intelligence (AI) in medicine in general, and in gastroenterology has made gigantic steps. However, the application of AI for biliary disease, in particular for cholangiocarcinoma, has been sub-optimal. The use of AI in combination with clinical data, cross-sectional imaging (computed tomography, magnetic resonance imaging) and endoscopy (endoscopic ultrasound and cholangioscopy) has the potential to significantly improve early diagnosis and the choice of optimal therapeutic options, leading to a transformation in the prognosis of this feared disease. In this review we summarize the current knowledge on the use of AI for the diagnosis and management of cholangiocarcinoma and point to future directions in the field.

Key Words: Cholangiocarcinoma; Artificial intelligence; Cholangioscopy; Artificial neural network; Machine learning; Therapeutic endoscopy; Endoscopic ultrasound

Core Tip: Artificial intelligence (AI) aided by multiple imaging modalities is accurate and effective for diagnosis and characterization of biliary masses. The advancement and incorporation of imaging into artificial intelligence will help to decrease delay in diagnosis of cholangiocarcinoma and potentially decrease mortality. This review examines studies showing that AI can assist in real-time diagnosis of cholangiocarcinoma and predict outcomes of treatment. Current data suggests that AI will soon become an indispensable part of the armamentarium for the management of cholangiocarcinoma and other biliary diseases.

Citation: Brenner AR, Laoveeravat P, Carey PJ, Joiner D, Mardini SH, Jovani M. Artificial intelligence using advanced imaging techniques and cholangiocarcinoma: Recent advances and future direction. *Artif Intell Gastroenterol* 2022; 3(3): 88-95
URL: https://www.wjgnet.com/2644-3236/full/v3/i3/88.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i3.88

INTRODUCTION

The concept of AI is best explained as a computer program that possesses the ability to perform functions such as data analysis, learning, and problem solving. Medical artificial intelligence involves the development of AI programs to assist in diagnosis and prognosis, therapeutic decision making, drug development, as well as development and data mining from the electronic medical records (EMR)[1-3]. In fact, artificial intelligence is utilized in almost every field of medicine[4-8], including radiology[9], gastroenterology[10], ophthalmology[11], cardiology[12], and surgery[13].

There are many different types of AI. The foundation of the most used form, Artificial Neural Networks (ANN), takes inspiration from the human nervous system[1,3]. The neurons of ANNs are individual computer processors that interconnect and possess the capability of processing and analyzing large amounts of data[1]. ANNs are composed of links of multiple layers of these 'neurons', an input layer linked to multiple hidden layers, which are in turn linked to an output layer[3]. All the layers in an ANN communicate in a feed forward manner with the ability to 'learn' by repeatedly adjusting their links[2]. Thus, one of the attractive qualities of ANNs is in their analytical and pattern recognition ability. One of the first applications of ANNs in medicine was to aid in the diagnosis of myocardial infarction[14]. Since that time, ANNs have been widely used[1]. Support Vector Machines (SVM) is another type of machine learning which uses data analysis algorithms for classification and regression analysis[15]. SVMs are widely used in drug development and cancer detection[16,17].

Convolutional neural networks (CNN) are a type of deep learning network, a network that incorporates three or more layers, that is commonly employed in medicine, in particular because of its easy applicability to imaging[18]. Convolutional neural networks are multi-layer analyses which work by taking an image (*e.g.* from CT, MRI, US) and extracting layers or features at each step of the process. These features are then characterized further by complex mathematical equations to break them down and compare them to similar images, leading to pattern recognition[18]. CNNs also can place weight on the value of a specific feature, thus allowing for the presence or absence of a given variable to haven a greater influence on the overall outcome.

The application of AI has grown at a rapid pace in all fields of medicine, and gastroenterology is no exception[19]. AI has been utilized in gastroenterology to identify esophageal neoplasms[20,21], diagnosis of Helicobacter pylori[22], predict gastric bleeding in patients on anti-thrombotics[23], predict the length of hospitalization for acute pancreatitis[24], differentiate between chronic pancreatitis and pancreatic cancer[25], stratify the need for ERCP[26], and characterization of colonic polyps[27]. These and many other ongoing developments will significantly impact the future of both diagnostic and therapeutic gastroenterology. One area of that has been somewhat neglected in the application of AI in gastroenterology is that of biliary disease, in particular cholangiocarcinoma. In this paper, we will review the current knowledge of the application of artificial intelligence in cholangiocarcinoma and point to the future directions in the field.

CHOLANGIOCARCINOMA

Cholangiocarcinoma (CCA) is a malignant neoplasm that can arise from anywhere along the biliary tree, including within the liver parenchyma, and is classified as distal, perihilar or intrahepatic[28]. Risk factors for CCA usually include long-term inflammatory states, like those associated with primary sclerosing cholangitis (PSC) and helminthic infection, or the continued presence of choledocholithiasis, but the majority for cases are idiopathic[29]. Cholangiocarcinoma accounts for about 3% of all gastrointestinal tumors and 10%-15% of hepatobiliary tumors[30]. Although rare, CCA has a very poor

prognosis, with 5-year survival rates following surgery rarely exceeding 35%[21]. Additionally, CCA incidence and mortality rates are increasing worldwide[31,32]. CCA is usually detected late in the disease stage and found incidentally due to poor screening methods for early detection[32]. Early diagnosis is relatively rare, limiting the possibility of curative surgery to < 30% of patients[33]. Furthermore, even among these, about 20%-50% of patients deemed candidates for resection *via* preoperative evaluation are found to have unresectable disease burden during surgery[34,35].

Given the importance of assessing disease burden, staging and location in determining a patient's treatment plan, it is imperative to have proper preoperative imaging in CCA[36]. While pathological examination remains the gold standard of diagnosis, grading and staging for CCA, advancements in imaging and detection of biomarkers have paved the way for further preoperative predictability of malignancy type and responsiveness to therapies. These advancements have allowed for the incorporation of AI into the sphere of cholangiocarcinoma for a more accurate and personalized management of the disease[37,38].

ARTICLE IDENTIFICATION PROCESS

The article search process was conducted in Medline and Embase[JM1]. Initial search was using different combinations of keywords such as "cholangiocarcinoma", "biliary disease", "cholangioscopy", "artificial intelligence", "artificial neural networks" and "convolutional neural networks". Abstracts of major conferences, such as Digestive Disease Week and United European Gastroenterology Week were also reviewed. Finally, a comprehensive search on clinicaltrial.gov was also conducted using the same keywords to search for active clinical trials involving cholangiocarcinoma and artificial intelligence.

ARTIFICIAL INTELLIGENCE IN BILIARY DISEASES AND CHOLANGIOCARCINOMA

Artificial intelligence has been employed to advance the classification and detection of cholangiocarcinoma by aiding in creating a histopathologic database[39] and characterizing bile acid assays to better predict malignancy[40]. The use of AI to optimize the predictive value of multivariable models, and in improving the diagnostic yield of cross-sectional imaging and endoscopy has been rapidly expanding. Table 1 summarizes currently available studies.

Use of artificial intelligence in aiding the predictive abilities of multivariable models

Artificial Intelligence models have been successfully used to improve the predictive abilities of multivariable models both in the pre-interventional diagnostic phase, as well as in post-operative or post-procedural outcomes in CCA patients. Many of these studies has utilized the area under the curve (AUC), the ability of a test to diagnose a differentiate a disease state from non-disease state, to assess the added benefit of the incorporation of AI in improving the effectiveness of multivariable models.

In the preoperative phase, multiple studies have used AI/radiographic model to predict lymph node metastasis (LNM) in CCA. One study developed and validated a radiographic model for LNM detection in intrahepatic cholangiocarcinoma (ICC) based on computed tomography (CT) imaging features combined with CA19-9 values[41]. In this study, an acceptable calibration and discrimination was observed in the primary study cohort (AUC 0.8462) and in a validation cohort (AUC 0.8921)[41]. Another study developed support vector machine model utilizing magnetic resonance imaging (MRI) imaging to preoperatively evaluate for LNM in ICC. This study found that an SVM model combining CA19-9 levels and select MRI features resulted in better predictive capabilities compared to a model based on imaging features alone (AUC of 0.842 *vs* 0.788, $P = 0.0219$)[42].

One retrospective study was able to use pre-operative MRI combined with post-operative immunohistochemical results to predict early recurrence of ICC after partial hepatectomy[43]. The model that combined AI with pathology and imaging features had a higher AUC (0.949 *vs* 0.889, $P = 0.247$) compared to the model that included only the pathology and imaging features, as well as better sensitivity (0.938 *vs* 0.875), and specificity (0.839 *vs* 0.774)[43]. In another study, inclusion of AI improved the ability of a multivariable model to predict early occlusion of bilateral plastic stents placed in patients with inoperable ICC[44]. In this study, the ANN built with the multivariable model was compared to a multivariable logistic regression model alone that included age, sex, stent diameter, cancer stage, and presence of liver metastasis[44]. Overall, 288 patients were analyzed, and the ANN model outperformed the logistic regression model (AUC 0.9647 *vs* 0.8763, $P = 0.021$)[44]. Artificial intelligence has also been used to identify which serum biomarkers can have higher diagnostic power for CCA[45]. An ANN model analyzed eight biochemical markers of CCA in 85 subjects with CCA and in 82 controls[45]. Alkaline phosphatase and CCA-associated carbohydrate antigen had a higher predictive value for the distinguishing CCA patients from controls[45]. Finally, in a recent study, Müller *et al*[46] developed an ANN utilizing known risk factors for ICC to predict survival in ICC patients. Using 293 patients, the ANN trained model achieved a higher AUC in predicting the 1 year survival

Table 1 Summary of studies assessing computed tomography, magnetic resonance, and endoscopic ultrasound using artificial intelligence-based approach for pancreatic cancer

Ref.	Year	Type of AI	Imaging modality	Training (#)	Testing (#)	AUC	Sensitivity (%)	Specificity (%)
Matake *et al*[47], 2006	2006	ANN	CT	120 - patients	120 - patients	0.934	81.9	94.4
Ji *et al*[48], 2019	2019	ANN	CT	177 - patients	70 - patients	0.961	72	76.2
Logeswaran[49], 2009	2009	MLP	MRI	120 - images	593 - images	N/A	N/A	
Yang *et al*[37], 2020	2020	ANN	MRI	80 - patients	20 - patients	0.9 (LMN)	85.8 (LMN)	81.8 (LMN)
						0.8 (differentiation)	73.2 (differentiation)	68.8 (differentiation)
Ghandour *et al*[51], 2021	2021	CNN	Cholangioscopy	254 - patients	95 - patients	0.86	0.81	0.91
Robles-Medrana *et al*[38], 2021	2021	ML	Cholangioscopy	1714 – images	198 - images	N/A	92	N/A
Pereira *et al*[50], 2022	2022	CNN	Cholangioscopy	5180 - images	1295 - images	1	99.3	99.4
Pattanpairoj *et al*[45], 2015	2015	ANN	Multivariate	85 - patients	22 - patients	N/A	98.71	96.94
Shao *et al*[44], 2018	2018	ANN	Multivariate	231 - patients	57 - patients	0.9544	N/A	N/A
Ji *et al*[41], 2019	2019	N/A	Multivariate	103 - patients	52 - patients	0.8462	86.8	76.3
Xu *et al*[42], 2019	2019	SVM	Multivariate	106 - patients	42 - patients	0.842	89.36	57.63
Zhao *et al*[43], 2019	2019	N/A	Multivariate	92 - patients	33 - patients	0.949	0.938	0.839
Müller *et al*[46], 2021	2021	ANN	Multivariate	233 - patients	60 - patients	0.89	N/A	N/A

ANN: Artificial neural network; MLP: Multi-layer perceptron; CNN: Convolutional neural network; ML: Machine learning; SVM: Support vector machine; CT: Computed tomography; MRI: Magnetic resonance imaging; N/A: Not applicable.

rates compared to one of the most commonly used scoring system, the Fudan score (0.89 *vs* 0.77, P = 0.24). In all of these studies, the addition of AI to commonly used multivariable models significantly improved their predictive abilities, improving therefore the diagnostic and post-procedural management of patient with suspected or diagnosed cholangiocarcinoma.

Use of AI in aiding cross-sectional imaging performance

Artificial Intelligence has been used to aid in the interpretation of cross-sectional imaging for nearly two decades. In a 2006 study, an artificial neural network applied to contrast-enhanced computed tomography (CE-CT) images helped differentiate four types of hepatic masses (intrahepatic peripheral cholangiocarcinoma, hepatocellular carcinoma, hemangioma, and metastatic lesions) from one-another [47]. The study then employed radiologists to evaluate CT scans with and without the assistance of ANN. There was marked improvement in diagnosis the hepatic masses with assistance from ANN compared to traditional radiologic evaluation (AUC 0.934 *vs* 0.888, P = 0.02, respectively)[47]. Another CT-based study was designed to predict survival outcomes and LNM in biliary tract cancers, and CT images were taken from 177 subjects who had previously undergone surgery[48]. An ANN based on CT characteristics was then built to classify the subjects into high risk or low risk for lymph node metastasis [48]. Patients who were classified as high risk based on the ANN model had a significantly lower survival rate compared to those classified as low risk [hazard ratio (HR) 3.37, 95%CI: 1.92, 5.91], underlying the importance of AI in improving prediction of disease course after treatment[48].

Artificial intelligence has also been used with MRI to improve its diagnostic/predictive power in several studies. One such study investigated the ability for an MRI based AI model to predict LNM in extrahepatic cholangiocarcinoma[37]. This was a proof-of-concept study to display the viability of a pre-operative prediction of both LMN and degree of differentiation, which could influence treatment approach. Images from 100 subjects with CCA were analyzed for the degree of CCA differentiation and

lymph node metastasis. The AI model had an AUC of 0.9 (95%CI: 0.66, 1.0) for predicting LMN while the AUC for degree of differentiation was 0.80 (95%CI: 0.58, 0.97)[37]. In another study, an ANN model based on MRCP images was able to distinguish between patients with CCA from those without CCA [49]. A total of 309 images were processed, 248 of which were normal and 61 were taken from patient with CCA. The ANN model achieved an accuracy of 94% for distinguishing between them. Furthermore, ANN achieved an accuracy of 88% in distinguishing between images of CCA and images of other common biliary diseases, such as cholecystitis, choledocholithiasis, PSC, and cholangitis[49].

Use of AI in aiding endoscopic evaluation of biliary diseases/cholangiocarcinoma

Artificial intelligence has also more recently been used to aid in endoscopic diagnosis of cholangiocarcinoma or other biliary diseases, even though most studies are currently in abstract form only. A study by Pereira *et al*[50] developed a CNN that differentiates biliary strictures as benign or malignant based on images from digital single operator cholangioscopy. After an evaluation of 6475 images from 85 patients with indeterminate biliary strictures, the authors found a sensitivity of 99.3%, specificity of 99.4%, and AUC of 1.00 for a correct diagnosis. In another study, currently available only as an abstract, the authors developed a CNN to detect abnormal biliary features *via* cholangioscopy images[51]. They defined abnormal features as presence of papillary mass, tortuous vessels, or ulcerations. Over 1000000 images were from 528 patients were evaluated for the study. The CNN showed an AUC of 0.86 (95%CI: 0.80, 0.92), sensitivity of 0.81 (95%CI: 0.0.72, 0.91), and specificity of 0.91 (95%CI: 0.86, 0.97)[51]. In another recent study, the utility of AI to perform real-time diagnosis of biliary strictures during cholangioscopy was assessed. This model was built using 23 cholangioscopy videos and was then tested on known cases (20 live cholangioscopy and 20 videos of cholangioscopy) of malignant biliary strictures. It accurately predicted malignancy in every case[38]. These initial results suggests that introduction of AI into standard clinical practice could potentially decrease time to diagnosis of indeterminate biliary strictures and allow for better diagnostic accuracy.

Endoscopic ultrasound (EUS) in combination with AI has been used in the assessment of pancreatic disease and may be beneficial in assisting in real-time differentiation between pancreatic masses and other solid masses during endoscopy[52]. However, there has been limited use of AI during EUS evaluations for cholangiocarcinoma. One recent study developed an AI system to recognize standard stations of EUS for biliary duct evaluation. In this study, AI had comparable accuracy to that of expert endosonographers, and significantly improved the learning curve of trainees[53].

CHOLEDOCHOLITHIASIS

Artificial intelligence has also been useful for the study of possible risk factors for CCA, such as choledocholithiasis. Several studies have demonstrated that AI can be used to risk-stratify patients with possible choledocolithiasis and therefore aid in the decision-making of the need for ERCP[54,55]. One study showed that a machine learning model using pre-ERCP imaging, including US and CT, in addition to select demographic features and laboratory findings can achieve a sensitivity of 97.7% and specificity of 100% in identifying choledocholithiasis[55]. Another study found that an AI model outperformed ASGE guidelines for proper indication for an ERCP (AUC 0.79 *vs* 0.59, respectively)[54]. In addition, the use of AI would avoid the need for ERCP in 36% of cases who would have undergone the procedure according to the ASGE guidelines[54]. Once more, the addition of AI can help providers achieve an individualized management program for patients in daily clinical practice.

CONCLUSION

The diagnosis and staging of cholangiocarcinoma is challenging, leading to potential major non-curative surgeries and/or dismal survival rate because of late diagnosis and inadequate prediction of metastases or recurrence using standard diagnostic methods. The introduction of AI technologies to traditional cross-sectional imaging and endoscopy, can create a major shift in the diagnosis and management of CCA. As mentioned above, many studies have already incorporated AI with significant improvement over traditional clinical data. While most of these studies are retrospective in nature, and therefore provide relatively poor quality data, they are very encouraging.

In addition, new studies are currently ongoing in which AI technologies are used to diagnose and risk-stratify patients with cholangiocarcinoma. The Synergy-AI clinical trial for example, is a non-interventional prospective observational study currently enrolling participants with cholangiocarcinoma, along with other malignancies. This trial is employing an Application Programming Interface to help match participants with personalized treatment protocols based on CT imaging, biomarkers, and laboratory results. In this setting, AI is expected to identify both the most cost effective, appropriate, and personalized treatment approach to each individual's malignancy[56]. Considering that most hospitals have incorporated electronic medical records (EMR) for their patients, it is easy to see how AI can be

used to select different patient variables (biochemical, histological or cross-sectional imaging) and use them to help develop personalized management strategies which optimize outcomes. Combining biomarkers, genetic sequencing, and imaging through AI models could lead to new approaches to the diagnosis and treatment of cholangiocarcinoma, including decreasing the need for unnecessary invasive endoscopic procedures for procurement of biopsies, as well as help develop a more targeted approach for therapy[57]. While more research and fine tuning of current AI systems is needed before reaching this stage, the future of AI in the management of cholangiocarcinoma seems clearly within reach.

FOOTNOTES

Author contributions: All authors contributed to the paper with regard to conception and design of the study, literature review and analysis, drafting the manuscript and all authors approved the final version of the manuscript.

Conflict-of-interest statement: The authors declare that they have no conflict of interest.

Country/Territory of origin: United States

ORCID number: Aaron R Brenner 0000-0001-8816-9182; Passisd Laoveeravat 0000-0001-6855-0437; Patrick J Carey 0009-0001-6521-4825; Manol Jovani 0000-0001-6803-8521.

Corresponding Author's Membership in Professional Societies: American Gastroenterological Association, No. 1612153.

S-Editor: Liu JH
L-Editor: A
P-Editor: Liu JH

REFERENCES

1 **Ramesh AN**, Kambhampati C, Monson JR, Drew PJ. Artificial intelligence in medicine. *Ann R Coll Surg Engl* 2004; **86**: 334-338 [PMID: 15333167 DOI: 10.1308/147870804290]
2 **Kaul V**, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointest Endosc* 2020; **92**: 807-812 [PMID: 32565184 DOI: 10.1016/j.gie.2020.06.040]
3 **Nawab K**, Athwani R, Naeem A, Hamayun M, Wazir M. A Review of Applications of Artificial Intelligence in Gastroenterology. *Cureus* 2021; **13**: e19235 [PMID: 34877212 DOI: 10.7759/cureus.19235]
4 **Patil V**, Vineetha R, Vatsa S, Shetty DK, Raju A, Naik N, Malarout N. Artificial neural network for gender determination using mandibular morphometric parameters: A comparative retrospective study. *Cogent Engineering* 2020; **7**: 1723783 [DOI: 10.1080/23311916.2020.1723783]
5 **Musunuri B**, Shetty S, Shetty DK, Vanahalli MK, Pradhan A, Naik N, Paul R. Acute-on-Chronic Liver Failure Mortality Prediction using an Artificial Neural Network. *Engineered Science* 2021; **15**: 187-196 [DOI: 10.1136/gutjnl-2021-iddf.94]
6 **Yang CM**, Shu J. Cholangiocarcinoma Evaluation *via* Imaging and Artificial Intelligence. *Oncology* 2021; **99**: 72-83 [PMID: 33147583 DOI: 10.1159/000507449]
7 **Haghbin H**, Aziz M. Artificial intelligence and cholangiocarcinoma: Updates and prospects. *World J Clin Oncol* 2022; **13**: 125-134 [PMID: 35316928 DOI: 10.5306/wjco.v13.i2.125]
8 **Shah M**, Naik N, Somani BK, Hameed BMZ. Artificial intelligence (AI) in urology-Current use and future directions: An iTRUE study. *Turk J Urol* 2020; **46**: S27-S39 [PMID: 32479253 DOI: 10.5152/tud.2020.20117]
9 **Khan G**, Howells J. A review of the current and future use of artificial intelligence (AI) in diagnostic radiology. *Clinical Radiology* 2020; **75**: e21 [DOI: 10.1016/j.crad.2020.11.086]
10 **Berbís MA**, Aneiros-Fernández J, Mendoza Olivares FJ, Nava E, Luna A. Role of artificial intelligence in multidisciplinary imaging diagnosis of gastrointestinal diseases. *World J Gastroenterol* 2021; **27**: 4395-4412 [PMID: 34366612 DOI: 10.3748/wjg.v27.i27.4395]
11 **Du XL**, Li WB, Hu BJ. Application of artificial intelligence in ophthalmology. *Int J Ophthalmol* 2018; **11**: 1555-1561 [PMID: 30225234 DOI: 10.18240/ijo.2018.09.21]
12 **Lopez-Jimenez F**, Attia Z, Arruda-Olson AM, Carter R, Chareonthaitawee P, Jouni H, Kapa S, Lerman A, Luong C, Medina-Inojosa JR, Noseworthy PA, Pellikka PA, Redfield MM, Roger VL, Sandhu GS, Senecal C, Friedman PA. Artificial Intelligence in Cardiology: Present and Future. *Mayo Clin Proc* 2020; **95**: 1015-1039 [PMID: 32370835 DOI: 10.1016/j.mayocp.2020.01.038]
13 **Moglia A**, Georgiou K, Georgiou E, Satava RM, Cuschieri A. A systematic review on artificial intelligence in robot-assisted surgery. *Int J Surg* 2021; **95**: 106151 [PMID: 34695601 DOI: 10.1016/j.ijsu.2021.106151]

14 **Baxt WG**. Use of an Artificial Neural Network for Data Analysis in Clinical Decision-Making: The Diagnosis of Acute Coronary Occlusion. *Neural Computation* 1990; **2**: 480-489 [DOI: 10.1162/neco.1990.2.4.480]

15 **Cortes C**, Vapnik V. Support-vector networks. *Machine Learning* 1995; **20**: 273-297 [DOI: 10.1007/bf00994018]

16 **Wang H**, Huang G. Application of support vector machine in cancer diagnosis. *Med Oncol* 2011; **28** Suppl 1: S613-S618 [PMID: 20842538 DOI: 10.1007/s12032-010-9663-4]

17 **Zernov VV**, Balakin KV, Ivaschenko AA, Savchuk NP, Pletnev IV. Drug discovery using support vector machines. The case studies of drug-likeness, agrochemical-likeness, and enzyme inhibition predictions. *J Chem Inf Comput Sci* 2003; **43**: 2048-2056 [PMID: 14632457 DOI: 10.1021/ci0340916]

18 **Liu Y**, Chen PC, Krause J, Peng L. How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. *JAMA* 2019; **322**: 1806-1816 [PMID: 31714992 DOI: 10.1001/jama.2019.16489]

19 **Kröner PT**, Engels MM, Glicksberg BS, Johnson KW, Mzaik O, van Hooft JE, Wallace MB, El-Serag HB, Krittanawong C. Artificial intelligence in gastroenterology: A state-of-the-art review. *World J Gastroenterol* 2021; **27**: 6794-6824 [PMID: 34790008 DOI: 10.3748/wjg.v27.i40.6794]

20 **Zhang SM**, Wang YJ, Zhang ST. Accuracy of artificial intelligence-assisted detection of esophageal cancer and neoplasms on endoscopic images: A systematic review and meta-analysis. *J Dig Dis* 2021; **22**: 318-328 [PMID: 33871932 DOI: 10.1111/1751-2980.12992]

21 **Horie Y**, Yoshio T, Aoyama K, Yoshimizu S, Horiuchi Y, Ishiyama A, Hirasawa T, Tsuchida T, Ozawa T, Ishihara S, Kumagai Y, Fujishiro M, Maetani I, Fujisaki J, Tada T. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest Endosc* 2019; **89**: 25-32 [PMID: 30120958 DOI: 10.1016/j.gie.2018.07.037]

22 **Bang CS**, Lee JJ, Baik GH. Artificial Intelligence for the Prediction of Helicobacter Pylori Infection in Endoscopic Images: Systematic Review and Meta-Analysis Of Diagnostic Test Accuracy. *J Med Internet Res* 2020; **22**: e21983 [PMID: 32936088 DOI: 10.2196/21983]

23 **Herrin J**, Abraham NS, Yao X, Noseworthy PA, Inselman J, Shah ND, Ngufor C. Comparative Effectiveness of Machine Learning Approaches for Predicting Gastrointestinal Bleeds in Patients Receiving Antithrombotic Treatment. *JAMA Netw Open* 2021; **4**: e2110703 [PMID: 34019087 DOI: 10.1001/jamanetworkopen.2021.10703]

24 **Pofahl WE**, Walczak SM, Rhone E, Izenberg SD. Use of an artificial neural network to predict length of stay in acute pancreatitis. *Am Surg* 1998; **64**: 868-872 [PMID: 9731816]

25 **Mendoza Ladd A**, Diehl DL. Artificial intelligence for early detection of pancreatic adenocarcinoma: The future is promising. *World J Gastroenterol* 2021; **27**: 1283-1295 [PMID: 33833482 DOI: 10.3748/wjg.v27.i13.1283]

26 **Jovanovic P**, Salkic NN, Zerem E. Artificial neural network predicts the need for therapeutic ERCP in patients with suspected choledocholithiasis. *Gastrointest Endosc* 2014; **80**: 260-268 [PMID: 24593947 DOI: 10.1016/j.gie.2014.01.023]

27 **Hassan C**, Spadaccini M, Iannone A, Maselli R, Jovani M, Chandrasekar VT, Antonelli G, Yu H, Areia M, Dinis-Ribeiro M, Bhandari P, Sharma P, Rex DK, Rösch T, Wallace M, Repici A. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointest Endosc* 2021; **93**: 77-85.e6 [PMID: 32598963 DOI: 10.1016/j.gie.2020.06.059]

28 **Brindley PJ**, Bachini M, Ilyas SI, Khan SA, Loukas A, Sirica AE, Teh BT, Wongkham S, Gores GJ. Cholangiocarcinoma. *Nat Rev Dis Primers* 2021; **7**: 65 [PMID: 34504109 DOI: 10.1038/s41572-021-00300-2]

29 **Cardinale V**, Bragazzi MC, Carpino G, Matteo SD, Overi D, Nevi L, Gaudio E, Alvaro D. Intrahepatic cholangiocarcinoma: review and update. *Hepatoma Research* 2018; **4**: 20 [DOI: 10.20517/2394-5079.2018.46]

30 **Bergquist A**, von Seth E. Epidemiology of cholangiocarcinoma. *Best Pract Res Clin Gastroenterol* 2015; **29**: 221-232 [PMID: 25966423 DOI: 10.1016/j.bpg.2015.02.003]

31 **Everhart JE**, Ruhl CE. Burden of digestive diseases in the United States Part III: Liver, biliary tract, and pancreas. *Gastroenterology* 2009; **136**: 1134-1144 [PMID: 19245868 DOI: 10.1053/j.gastro.2009.02.038]

32 **Razumilava N**, Gores GJ. Surveillance for Cholangiocarcinoma in Patients with Primary Sclerosing Cholangitis: Effective and Justified? *Clin Liver Dis (Hoboken)* 2016; **8**: 43-47 [PMID: 27822365 DOI: 10.1002/cld.567]

33 **DeOliveira ML**, Cunningham SC, Cameron JL, Kamangar F, Winter JM, Lillemoe KD, Choti MA, Yeo CJ, Schulick RD. Cholangiocarcinoma: thirty-one-year experience with 564 patients at a single institution. *Ann Surg* 2007; **245**: 755-762 [PMID: 17457168 DOI: 10.1097/01.sla.0000251366.62632.d3]

34 **Rocha FG**, Matsuo K, Blumgart LH, Jarnagin WR. Hilar cholangiocarcinoma: the Memorial Sloan-Kettering Cancer Center experience. *J Hepatobiliary Pancreat Sci* 2010; **17**: 490-496 [PMID: 19806295 DOI: 10.1007/s00534-009-0205-4]

35 **Nagino M**, Ebata T, Yokoyama Y, Igami T, Sugawara G, Takahashi Y, Nimura Y. Evolution of surgical treatment for perihilar cholangiocarcinoma: a single-center 34-year review of 574 consecutive resections. *Ann Surg* 2013; **258**: 129-140 [PMID: 23059502 DOI: 10.1097/SLA.0b013e3182708b57]

36 **Kim SY**. Preoperative Radiologic Evaluation of Cholangiocarcinoma. *Korean J Gastroenterol* 2017; **69**: 159-163 [PMID: 28329917 DOI: 10.4166/kjg.2017.69.3.159]

37 **Yang C**, Huang M, Li S, Chen J, Yang Y, Qin N, Huang D, Shu J. Radiomics model of magnetic resonance imaging for predicting pathological grading and lymph node metastases of extrahepatic cholangiocarcinoma. *Cancer Lett* 2020; **470**: 1-7 [PMID: 31809800 DOI: 10.1016/j.canlet.2019.11.036]

38 **Robles-Medranda C**, Oleas R, Alcivar-Vasquez J, Mendez JC, Puga-Tejada M, Del Valle R, Lukashok, HP. Artificial Intelligence Model for the Characterization of Biliary Strictures During Real-Time Digital Cholangioscopy: A Pilot Study. *Endoscopy* 2021; **53**: OP53 [DOI: 10.1055/s-0041-1724313]

39 **Zhang Q**, Li Q, Yu G, Sun L, Zhou M, Chu J. A Multidimensional Choledoch Database and Benchmarks for Cholangiocarcinoma Diagnosis. *IEEE access* 2019; **7**: 149414-149421 [DOI: 10.1109/access.2019.2947470]

40 **Negrini D**, Zecchin P, Ruzzenente A, Bagante F, De Nitto S, Gelati M, Salvagno GL, Danese E, Lippi G. Machine Learning Model Comparison in the Screening of Cholangiocarcinoma Using Plasma Bile Acids Profiles. *Diagnostics (Basel)* 2020; **10** [PMID: 32748848 DOI: 10.3390/diagnostics10080551]

41 **Ji GW**, Zhu FP, Zhang YD, Liu XS, Wu FY, Wang K, Xia YX, Jiang WJ, Li XC, Wang XH. A radiomics approach to

predict lymph node metastasis and clinical outcome of intrahepatic cholangiocarcinoma. *Eur Radiol* 2019; **29**: 3725-3735 [PMID: 30915561 DOI: 10.1007/s00330-019-06142-7]

42 **Xu L**, Yang P, Liang W, Liu W, Wang W, Luo C, Wang J, Peng Z, Xing L, Huang M, Zheng S, Niu T. A radiomics approach based on support vector machine using MR images for preoperative lymph node status evaluation in intrahepatic cholangiocarcinoma. *Theranostics* 2019; **9**: 5374-5385 [PMID: 31410221 DOI: 10.7150/thno.34149]

43 **Zhao L**, Ma X, Liang M, Li D, Ma P, Wang S, Wu Z, Zhao X. Prediction for early recurrence of intrahepatic mass-forming cholangiocarcinoma: quantitative magnetic resonance imaging combined with prognostic immunohistochemical markers. *Cancer Imaging* 2019; **19**: 49 [PMID: 31307551 DOI: 10.1186/s40644-019-0234-4]

44 **Shao F**, Huang Q, Wang C, Qiu L, Hu YG, Zha SY. Artificial Neural Networking Model for the Prediction of Early Occlusion of Bilateral Plastic Stent Placement for Inoperable Hilar Cholangiocarcinoma. *Surg Laparosc Endosc Percutan Tech* 2018; **28**: e54-e58 [PMID: 29252936 DOI: 10.1097/SLE.0000000000000502]

45 **Pattanapairoj S**, Silsirivanit A, Muisuk K, Seubwai W, Cha'on U, Vaeteewoottacharn K, Sawanyawisuth K, Chetchotsak D, Wongkham S. Improve discrimination power of serum markers for diagnosis of cholangiocarcinoma using data mining-based approach. *Clin Biochem* 2015; **48**: 668-673 [PMID: 25863112 DOI: 10.1016/j.clinbiochem.2015.03.022]

46 **Müller L**, Mähringer-Kunz A, Gairing SJ, Foerster F, Weinmann A, Bartsch F, Heuft LK, Baumgart J, Düber C, Hahn F, Kloeckner R. Survival Prediction in Intrahepatic Cholangiocarcinoma: A Proof of Concept Study Using Artificial Intelligence for Risk Assessment. *J Clin Med* 2021; **10** [PMID: 34066001 DOI: 10.3390/jcm10102071]

47 **Matake K**, Yoshimitsu K, Kumazawa S, Higashida Y, Irie H, Asayama Y, Nakayama T, Kakihara D, Katsuragawa S, Doi K, Honda H. Usefulness of artificial neural network for differential diagnosis of hepatic masses on CT images. *Acad Radiol* 2006; **13**: 951-962 [PMID: 16843847 DOI: 10.1016/j.acra.2006.04.009]

48 **Ji GW**, Zhang YD, Zhang H, Zhu FP, Wang K, Xia YX, Jiang WJ, Li XC, Wang XH. Biliary Tract Cancer at CT: A Radiomics-based Model to Predict Lymph Node Metastasis and Survival Outcomes. *Radiology* 2019; **290**: 90-98 [PMID: 30325283 DOI: 10.1148/radiol.2018181408]

49 **Logeswaran R**. Cholangiocarcinoma--an automated preliminary detection system using MLP. *J Med Syst* 2009; **33**: 413-421 [PMID: 20052894 DOI: 10.1007/s10916-008-9203-3]

50 **Pereira P**, Mascarenhas M, Ribeiro T, Afonso J, Ferreira JPS, Vilas-Boas F, Parente MPL, Jorge RN, Macedo G. Automatic detection of tumor vessels in indeterminate biliary strictures in digital single-operator cholangioscopy. *Endosc Int Open* 2022; **10**: E262-E268 [PMID: 35295246 DOI: 10.1055/a-1723-3369]

51 **Ghandour B**, Hsieh H-W, Akshintala V, Bejjani M, Szvarca D, Tejaswi S, DiMaio CJ, Trindade AJ, Manvar A, Itoi T, Slivka A, Goenka MK, Moon JH, Kushnir V, Sethi A, Bun Teoh AY, Yan Tang RS, Lau JY, Rerknimitr R, Ang TL, Maydeo AP, Jonica ER, Kaffes AJ, Saxena P, Nguyen NQ, Khokhar AS, D'Souza L, Buscaglia JM, Sherman S, Muniraj T, Jamidar P, Vedula S, Hager GD, Kashab MA. S1 Machine Learning for Classification of Indeterminate Biliary Strictures During Cholangioscopy. 2021; 116: S1 [DOI: 10.14309/01.ajg.0000771984.13549.fa]

52 **Prasoppokakorn T**, Tiyarattanachai T, Chaiteerakij R, Decharatanachart P, Mekaroonkamol P, Ridtitid W, Kongkam P, Rerknimitr R. Application of artificial intelligence for diagnosis of pancreatic ductal adenocarcinoma by EUS: A systematic review and meta-analysis. *Endosc Ultrasound* 2022; **11**: 17-26 [PMID: 34937308 DOI: 10.4103/EUS-D-20-00219]

53 **Yao L**, Zhang J, Liu J, Zhu L, Ding X, Chen D, Wu H, Lu Z, Zhou W, Zhang L, Xu B, Hu S, Zheng B, Yang Y, Yu H. A deep learning-based system for bile duct annotation and station recognition in linear endoscopic ultrasound. *EBioMedicine* 2021; **65**: 103238 [PMID: 33639404 DOI: 10.1016/j.ebiom.2021.103238]

54 **Akshintala VS**, Tang B, Kamal A, Buxbaum JL, Elmunzer BJ, Wani SB, Yu CY, Tieu AH, Kalloo AN, Singh V, Caffo B, Kashab MA. Sa1470 Risk estimation, machine learning based ercp decision-making tool for suspected choledocholithiasis. *Gastrointestinal Endoscopy* 2019; **89**: AB246-AB7 [DOI: 10.1016/j.gie.2019.03.264]

55 **Dalai C**, Azizian J, Trieu H, Rajan A, Chen F, Dong T, Beaven S, Tabibian JH. Machine learning models compared to existing criteria for noninvasive prediction of endoscopic retrograde cholangiopancreatography-confirmed choledocholithiasis. *Liver Res* 2021; **5**: 224-231 [PMID: 35186364 DOI: 10.1016/j.livres.2021.10.001]

56 **Kurnaz S**, Loaiza-Bonilla A, Freedman JL, Arango BA, Johnston K, Bonilla-Reyes PA, Culcuoglu CM, SYNERGY-AI Precision Oncology Group. SYNERGY-AI: Artificial intelligence based precision oncology clinical trial matching and registry. *Journal of Clinical Oncology* 2019; **37** suppl 4: TPS717-TPS [DOI: 10.1200/jco.2019.37.4_suppl.tps717]

57 **Akshintala VS**, Khashab MA. Artificial intelligence in pancreaticobiliary endoscopy. *J Gastroenterol Hepatol* 2021; **36**: 25-30 [PMID: 33448514 DOI: 10.1111/jgh.15343]

Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: bpgoffice@wjgnet.com
Help Desk: https://www.f6publishing.com/helpdesk
https://www.wjgnet.com

ISSN 2644-3236 (online)

Artificial Intelligence in *Gastroenterology*

Artif Intell Gastroenterol 2022 October 28; 3(4): 96-116

Published by Baishideng Publishing Group Inc

A I G

Artificial Intelligence in Gastroenterology

Contents

Bimonthly Volume 3 Number 4 October 28, 2022

MINIREVIEWS

ORIGINAL ARTICLE

Prospective Study

Artificial Intelligence in Gastroenterology

Bimonthly Volume 3 Number 4 October 28, 2022

Contents

ABOUT COVER

Editorial Board Member of *Artificial Intelligence in Gastroenterology*, Wenyu Lin, PhD, Assistant Professor, Department of Gastrointestinal, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, United States. wlin1@mgh.harvard.edu

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

The *AIG* is now abstracted and indexed in Reference Citation Analysis, China Science and Technology Journal Database.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: *Si Zhao*; **Production Department Director:** *Xu Guo*; **Editorial Office Director:** *Jin-Lei Wang*.

NAME OF JOURNAL
Artificial Intelligence in Gastroenterology

ISSN
ISSN 2644-3236 (online)

LAUNCH DATE
July 28, 2020

FREQUENCY
Bimonthly

EDITORS-IN-CHIEF
Rajvinder Singh, Ferruccio Bonino

EDITORIAL BOARD MEMBERS
https://www.wjgnet.com/2644-3236/editorialboard.htm

PUBLICATION DATE
October 28, 2022

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

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 October 28; 3(4): 96-104

DOI: 10.35712/aig.v3.i4.96

ISSN 2644-3236 (online)

MINIREVIEWS

Role of artificial intelligence in the diagnosis and treatment of hepatocellular carcinoma

Rajesh Kumar Mokhria, Jasbir Singh

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): 0
Grade C (Good): C
Grade D (Fair): D
Grade E (Poor): 0

P-Reviewer: Lu G, China; Wan Y, China

Received: June 7, 2022
Peer-review started: June 7, 2022
First decision: July 14, 2022
Revised: July 30, 2022
Accepted: September 13, 2022
Article in press: September 13, 2022
Published online: October 28, 2022

Rajesh Kumar Mokhria, Government Model Sanskriti Senior Secondary School, Chulkana, 132101, Panipat, Haryana, India

Jasbir Singh, Department of Biochemistry, Kurukshetra University, Kurukshetra, 136119, Haryana, India

Corresponding author: Rajesh Kumar Mokhria, PhD, Biology Lecturer, Government Model Sanskriti Senior Secondary School, Chulkana, Panipat 132101, Haryana, India.
mokhria79@gmail.com

Abstract

Artificial intelligence (AI) evolved many years ago, but it gained much advancement in recent years for its use in the medical domain. AI with its different subsidiaries, *i.e.* deep learning and machine learning, examine a large amount of data and performs an essential part in decision-making in addition to conquering the limitations related to human evaluation. Deep learning tries to imitate the functioning of the human brain. It utilizes much more data and intricate algorithms. Machine learning is AI based on automated learning. It utilizes earlier given data and uses algorithms to arrange and identify models. Globally, hepatocellular carcinoma is a major cause of illness and fatality. Although with substantial progress in the whole treatment strategy for hepatocellular carcinoma, managing it is still a major issue. AI in the area of gastroenterology, especially in hepatology, is particularly useful for various investigations of hepatocellular carcinoma because it is a commonly found tumor, and has specific radiological features that enable diagnostic procedures without the requirement of the histological study. However, interpreting and analyzing the resulting images is not always easy due to change of images throughout the disease process. Further, the prognostic process and response to the treatment process could be influenced by numerous components. Currently, AI is utilized in order to diagnose, curative and prediction goals. Future investigations are essential to prevent likely bias, which might subsequently influence the analysis of images and therefore restrict the consent and utilization of such models in medical practices. Moreover, experts are required to realize the real utility of such approaches, along with their associated potencies and constraints.

Key Words: Hepatocellular carcinoma; Artificial intelligence; Deep learning; Machine learning; Support vector machines; Artificial neural networks

Core Tip: Globally, hepatocellular carcinoma is a major cause of illness and fatality. Although substantial progress has been made in the treatment strategy for hepatocellular carcinoma, managing it is still a major issue. Artificial intelligence in the area of gastroenterology, especially in hepatology, is particularly useful for various investigations of hepatocellular carcinoma because it is a commonly found tumor and has specific radiological features that enable diagnostic procedures without the requirement of histological study. Artificial intelligence is utilized to diagnose, curative and prediction goals.

Citation: Mokhria RK, Singh J. Role of artificial intelligence in the diagnosis and treatment of hepatocellular carcinoma. *Artif Intell Gastroenterol* 2022; 3(4): 96-104
URL: https://www.wjgnet.com/2644-3236/full/v3/i4/96.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i4.96

INTRODUCTION

Hepatocellular carcinoma (HCC) is a malignancy of the liver that is very lethal. It is the most commonly found primary adult liver malignancy. Worldwide it is the third most common cause of cancer-related death[1]. According to the American Cancer Society, 42810 new liver and intrahepatic cholangiocarcinoma cases were detected in 2020, of which 30160 died[2]. Surgery (liver transplantation and resection) is the backbone of HCC treatment and is the only possible treatment option. Delamination or removal is an alternative treatment for small tumors. In addition, intra-arterial treatment and chemotherapy can control the disease to some extent[1]. In addition, HCC has certain radiological features that do not require histological examination for diagnosis. Therefore, the analysis and interpretation of diagnostic imaging procedures are not always easy as it changes during the disease course. The same applies to diagnosis/prognosis and treatment response, as they are influenced by numerous factors.

Artificial intelligence (AI) is the computer simulation of the human intelligence process. The concept of AI emerged in the 1950s[3], but only a few years ago it made real progress. It has been used in a variety of industries, *i.e.* image and natural language processing. In the field of medicine, AI is becoming increasingly significant. The utilization of AI is rapidly expanding and is increasingly useful in understanding gastrointestinal diseases[4-6]. The phrase "artificial intelligence" refers to a group of computer programs that attempt to mimic human brain capabilities, *i.e.* learning and problem-solving.

AI has evolved into a separate discipline called machine learning (ML). ML examines data to develop algorithms that can recognize distinct behavior forms and confirm predictive models. ML focuses on developing mathematical models that assist machines in making predictions or judgments without being explicitly programmed. Various ML techniques, for instance, support vector machines (SVM), artificial neural networks (ANNs), classification, and regression trees, seem to be employed in various investigations in the medical discipline[7]. Deep learning (DL) has emerged as an emerging paradigm of ML for developing multilayered neural network algorithms, and approaches like convolutional neural network (CNN), an ANN multilayer, have been widely accepted and used in radiological image analysis[8,9].

In a nutshell, ML is a core branch of AI, and DL is used to implement it. The use of ML and DL to forecast the risk of gastric cancer has been successful[10]. Figure 1 shows the correlation between AI, ML, and DL.

There are limitations in using AI in various areas of medicine. Looking back on many studies and applications of irrelevant databases having biases can influence the truthfulness of AI. Therefore, it is essential to design a bias-free, proposed, well-designed multicenter collaborative study, and various important aspects, such as economics, medical professional regulation, and ethical reviews, should not be ignored. Various terms associated with AI in this minireview are given in Table 1.

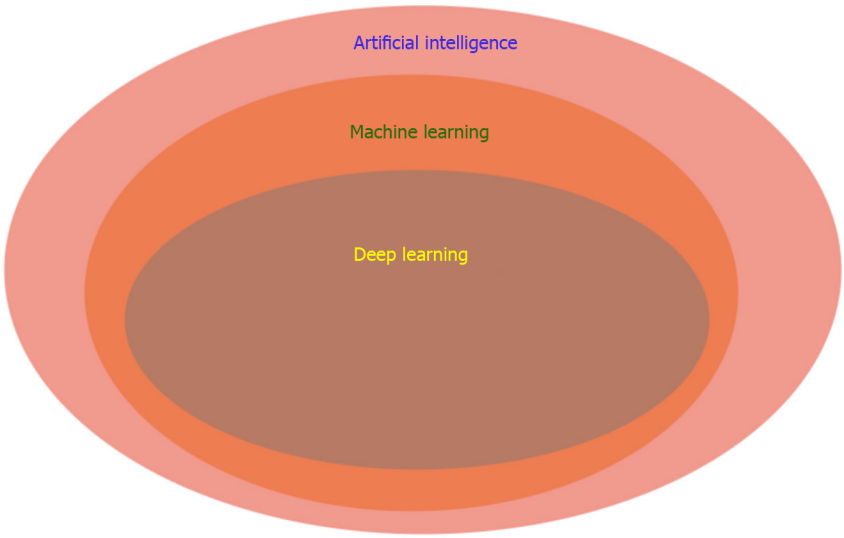
USE OF AI IN HCC DIAGNOSIS

The utility of AI can enhance diagnostic procedures in the area of liver cancer. CNN in the form of multilayered ANN is interlinked, and whole input data passes through every layer before being transformed to give output data. It is a more advanced version of DL that has its own learning capacity. Ultrasound (US) tests, abdominal computed tomography (CT), magnetic resonance imaging (MRI) of the abdomen, positron emission tomography (PET), and histology can benefit from CNN.

Table 1 Various terminology associated with artificial intelligence

Term	Definition
AI	The utilization of computers and associated techniques to mimic the sharp attitude and critical approach of humans
ML	It is a branch of AI and computer science that concerns the usage of data and algorithms to mimic the means that human beings ascertain and step by step upgrading its precision
ANN	It is a computational model in accordance with the structure and functions of biological neural networks. ANNs employ a nonlinear function to a loaded sum of inputs and model relations among them
CNN	It is a deep learning neural network intended to process structured arrays of data, *i.e.* radiological images
Deep learning	A branch of ML that tries to mimic the human brain and has the ability to gather data and do predictions with remarkable precision
AUC	AUC is an approach applied in ML to assess many used models to find out which have the higher performance
Accuracy	AI and ML technology employ algorithms to analyze data and perform predictions on the basis of such data. Although studies report that AI programs may regularly achieve accuracy levels of at least 95% and AI programs cannot verify the veracity of the data being examined, the overall accuracy is typically lower yet still higher than 80%
C-index (c-statistic)	It is an algorithm performance metric that takes values between 0 and 1 and explains how well the model fits the data

AI: Artificial intelligence; ANN: Artificial neural networks; AUC: Area under the curve; CNN: Convolutional neural network; ML: Machine learning.



DOI: 10.35712/aig.v3.i4.96

Figure 1 Relationship between artificial intelligence, machine learning, and deep learning.

Ultrasound of the abdomen

HCC develops in cirrhotic livers most of the time but not always. Clinical practice recommendations advocate routine abdominal US in hepatic cirrhosis patients. This approach is used for detecting lesions that occupy space. US is the primary machine for detecting hepatic disease and fresh lesions. Though, analysis of images is not straightforward and can be subject to interobserver variations.

To review the fundamental disorder, Bharti *et al*[11] established an ANN model that discriminated various phases of hepatic infection by analyzing US images: normal liver, chronic liver disease, cirrhosis, and HCC. Further, this model's accuracy was found to be 96.6%[11]. An algorithm to analyze US images was developed by Liu *et al*[12]. Liu *et al*[12] preferred the liver capsule to detect the existence of cirrhosis, even at an early stage when radiological findings are not clearly visible. By investigating the morphology of the liver capsule, Liu *et al*[12] predicted the presence or absence of cirrhosis with an area under the curve (AUC) of 0.968.

The human output is defined when it comes to identifying liver lesions from US images. Schmauch *et al*[13] developed a DL approach that could reveal and label benign and malignant space-occupying liver lesions. This system requires acceptance. It has the potential to improve the diagnostic yield of US and inform clinicians about potentially malignant lesions[13].

To improve the ability of contrast-enhanced US (C-US) for the detection of cancer-related characteristics, the use of AI has been utilized. Guo *et al*[14] confirmed how applying DL to the behavior of liver lesions observed on C-US in three phases (arterial, portal, and late) improved the accuracy, sensitivity, and specificity of the investigation undertaken.

Abdominal CT scan with intravenous contrast

When an US reveals a fresh liver lesion, further imaging procedures, primarily dynamic contrast-enhanced CT or MRI, are used to get an accurate diagnosis. In dynamic CT or MRI scans, the radiological behavior of liver lesions can be used to characterize the lesion. If CT scans of liver nodules reveal unclear behavior, then lesion biopsy is prescribed as per the recommendation of the European Association for the Study of the Liver guidelines[15]. As suggested by the American Association for the Study of Liver Diseases guidelines[16], there is the possibility of non-detection of a malignant lesion involved during the procedure or during close follow-up. A study was performed on 178 patients with cirrhosis and liver nodules by Mokrane *et al*[17], and they were unable to differentiate between neoplastic and non-neoplastic lesions in these patients, hence requiring a biopsy. On doing a biopsy, 77% of the lesions were malignant. By applying DL techniques, the AUC for classifying nodules as HCC or non-HCC was 0.70. By analyzing the output of three-layered ANN, Yasaka *et al*[18] with the help of contrast-enhanced CT classified liver masses into five groups: A (cholangiocarcinoma, hepatocholangiocarcinoma, or metastasis); B (other malignant tumors, *i.e.* cholangiocarcinoma, hepatocholangiocarcinoma, or metastasis); C (ambiguous masses, dysplastic nodules, or early HCC, and benign masses other than cysts or haemangiomas); D (haemangiomas); and E (cysts).

Assessing tumor load could be beneficial for detecting tumor relapse in follow-up CT scans. Vivanti *et al*[19] proposed an automated detecting procedure for recurrence on the basis of early manifestation of the tumor, its CT behavior, baseline tumor load/mass quantification, and follow-up. With an accuracy of 86%, this approach demonstrated a higher proportion of true positives in detecting tumor relapse.

The usefulness of liver segmentation in assessing lesions in the liver and managing good treatment is critical. Li *et al*[20] developed a CNN that could cause the segmentation of liver tumors on the basis of CT images having an accuracy of 82.67% ± 1.43%, which is better than existing approaches, allowing for more appropriate treatment planning.

Abdominal MRI

The use of CNN in MRI has also been investigated. Hamm *et al*[21] prepared and verified a CNN-based DL approach that identified MRI liver lesions with 92% accuracy, 92% sensitivity, and 98% specificity with a mean computation time of 5.6 milliseconds.

Further research has used more MRI sequences, risk components, and clinical information of the patient to create an automated classification method that classifies hepatic lesions as adenoma, cyst, haemangioma, HCC, and metastasis, having sensitivity/specificity of 0.80/0.78, 0.93/0.93, 0.84/0.82, 0.73/0.56, and 0.62/0.77 respectively[22].

PET

Preis *et al*[23] used a neural network to study hepatic intake of fluorodeosyglucose 18F along with data from the patient and clinical details to assess the results of 18F-FDG PET/CT (Fluorine 18 fluorodeosyglucose positron emission tomography/computed tomography). Preis *et al*[23] obtained higher sensitivity and specificity to find malignancy of the liver, which remained unrevealed visibly. This method can help the radiologist in the analysis of PET.

Histology

Even for experienced pathologists, determining the histopathological categorization of a liver lesion and distinction of tumor strain is critical to planning the treatment and prognosis assessment of the disease. Kiani *et al*[24] were concerned with the histopathological distinction between HCC and cholangiocarcinoma and employed AI to assist pathologists.

Others reported how a deep CNN can perform an automatic identification of HCC and discriminate normal tissue from malignant tissue as well as identify key biological predictors, utilizing previous histopathological images of HCC[25].

USE OF AI FOR TREATING HCC

The specific biological variance among HCC patients hampers evidence-based clinical assessment among all patients. Hence, for optimizing treatment techniques and measuring the results, powerful standardized risk classification tools are required. AI has the potential to play a significant role in the treatment of HCC in this area. The majority of studies about the applicability of AI in HCC treatment are focused on analyzing specific tumor attributes, *i.e.* radiological, histological, or genetic traits, or combining clinical data to estimate treatment response. Therefore, patients will be able to be better

selected for certain treatment alternatives.

Use of radiomics

The examination and remedy measure of HCC is generally performed with imaging facilities i.e. C-US, CT, and MRI following investigation of assured tumor characteristics, *i.e.* vascularization or behavior after the addition of a contrasting substance[26]. These attributes are amenable to biases after analysis by radiologists, along with the absence of high-resolution dimensional images. Recently an advanced technology has emerged in the area of radiology and cancer which is known as radiomics[27]. This technology extracts a large amount of significant data from the radiological images and links this data with the related biological system. The study of complete data with AI software can give effective and accurate reports for proper diagnosis and prognosis[27,28]. Figure 2 shows various stages of radiomics where AI can play a role.

Assessment of surgical resection

The early reappearance of the tumor following operative removal is due to an unsatisfactory prognostic process. The recognition of clinical cases before surgical operation with more risk of relapse is essential to escape irrelevant treatment. Various computer models help to analyze specific tumor markers/features and assist in the prognosis of the risk of relapse before operative procedures. These models also help in the assessment of survival after surgical removal.

Vascular microinvasion (VMI) is a self-sufficient prognostic component of relapse. VMI is linked with poor outcomes following tumor excision[29]. The accessibility of data regarding VMI preoperatively can be of high use. The radiological approach presently used in medical practice does not give a fair diagnosis.

Several studies explain radiomic signatures that presume the status of VMI preoperatively on the basis of contrast-enhanced CT[30,31] or MRI[32]. These techniques include exposure to radiation, are hard to execute, and are expensive. In a recent study, Dong *et al*[33] used grayscale US images based on radiomic algorithms to proceed with radiomic signatures in the prediction of VMI. By using radiomic techniques, Ji *et al*[34] developed prognostic models for relapse after excision surgery for assessing contrast-enhanced CT images and had a C-index value of 0.633-0.699. These models could be utilized for providing an individualized risk stratification for managing HCC individually.

ML techniques help in assessing survival after surgical resection as observed in many studies[35-37]. Recently, more advanced DL models helped in assessing survival after surgical resection on the basis of digitalized histological images of tumors.

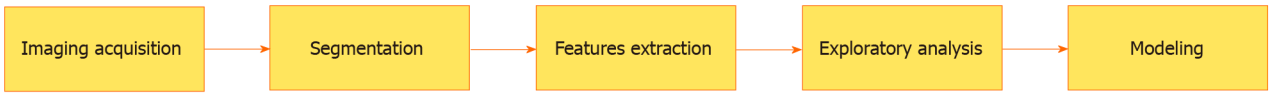
Assessment of transcatheter arterial chemoembolization

According to Barcelona Clinical Liver Cancer (BCLC) classification, transcatheter arterial chemoembolization (TACE) exists as the preferred option for the treatment of intermediary B stage HCC[38]. The right choice of patients who can get benefit from this treatment is critical in order to minimize superfluous investigations that can lead to unfavorable side effects and waste healthcare resources. Studies based on AI approaches have been created as a trial to infer the feedback of TACE treatment and facilitate the proper selection of patients. The majority of the studies rely on image analysis, but some studies have also utilized genomic signatures. Morshid *et al*[39] developed an automatic ML algorithm that predicted TACE response using a mixture of quantitative CT image attributes and pretreatment patient clinical data. They obtained a prediction accuracy rate of 74.2% while working on combining the Barcelona Clinic Liver Cancer stage and quantitative image characteristics instead of applying the Barcelona Clinic Liver Cancer stage alone. Peng *et al*[40] used CT scans from 789 patients from three separate hospitals to verify a DL model for predicting TACE response. They were able to predict complete responses with an accuracy of 84% and an AUC of 0.97. Liu *et al*[41] developed and verified a DL radiomics-based C-US approach as a result of a quantitative assessment of C-US cine recordings. They demonstrated a high level of reproducibility and an AUC of 0.93 (95% confidence interval: 0.80-0.98) for predicting TACE reaction.

Further research has combined MRI and clinical data with ML approaches to predict TACE response. Abajian *et al*[42] worked on 36 patients who had an MRI prior to TACE. They built a response prediction model with 78.0% accuracy, 62.5% sensitivity, and 82% specificity.

The efficacy of TACE has also been tested by a post-treatment survival analysis of patients. Mähringer-Kunz *et al*[43] designed an ANN with every variable of main traditional prediction scores to produce a survival prediction model following TACE (ART[44], ABCR[45], and SNACOR[46]). With an AUC of 0.77, 78% sensitivity, and 81% specificity, they expected a 1-year survival rate that was better than the conventional scores.

Although radiomics have been used in the majority of investigations estimating the usage of AI to examine TACE. Some have also looked at genetic analysis to predict TACE response. Ziv *et al*[47] analyzed genetic mutations by applying SVM algorithms to look for tumor responses following TACE. However, this study involved a small number of cases.



DOI: 10.35712/aig.v3.i4.96

Figure 2 Stages of radiomics wherever artificial intelligence can play a role.

Radiofrequency ablation evaluation

Radiofrequency ablation has also been studied as a treatment for HCC in its early stages[38]. Liang *et al*[48] used SVM to create a prognostic model of HCC relapse. They investigated 83 HCC cases that had undergone radiofrequency ablation and secured an AUC of 0.69, 67% sensitivity, and 86% specificity. From this data, they could recognize patients with a greater chance of relapse.

HCC OVERALL SURVIVAL PREDICTION

Apart from the use of any therapy, AI approaches have been used to predict the overall survival of HCC patients. The observations by Dong *et al*[49] were based on current information on the relationship between anomalies in DNA methylation and HCC[50-52]. They employed ML techniques (SVM) for the evaluation of DNA methylation data from 377 HCC samples and created three risk groups to expect complete survival and achieved a mean 10-fold cross-validation score of 0.95.

FUTURE PERSPECTIVES

To illustrate the effectiveness of AI for medical assistance, further research is required that compares the output of medical staff with AI assistance *vs* experts lacking AI assistance. These studies should target elements linked to curing and prognosis (for instance, identifying ambiguous hepatic wounds, the existence of vascular invasion, and the reaction to percutaneous treatments) to analyze liver masses and explore HCC. Additional significant points are the utilization of AI for interpretation of HCC behavior in cirrhotic and non-cirrhotic patients, in the differential diagnosis of primary and metastatic liver lesions[53], and particularly in the clinical detection of cholangiocarcinoma, which is difficult to differentiate from HCC with existing approaches and has distinct treatment methods from HCC. Simultaneously, healthcare providers must be trained for the integration of AI into everyday practice in the area of liver cancer.

SIGNIFICANCE OF THE STUDY

AI has guided the detection of HCC (on the basis of premalignant variations, imaging, and biomarkers) as a result of its capability to examine huge datasets and combine data effectively. The perspective of AI techniques is immense in every stage in the handling of HCC, *e.g.*, from early diagnosis to treatment options and prognostic and therapeutic response prognosis. These methods could promote accurate and personalized medicine to assist clinical practice and better utilize healthcare resources. Numerous datasets (radiological images or pathologic data) could be utilized individually or in conjunction for accuracy better than that of conventional statistical means. Moreover, AI-based approaches can also assist in lowering interobserver variance while studying images and leads to standardization.

INNOVATIVE CONTRIBUTIONS OF THE STUDY

The outcomes from many studies endorse the consolidation of the ML models with clinical/pathologic data and created clinical scores or biomarkers. Biomarkers detected by the incorporation of several '-omics' datasets lead to the recognition of a biochemical tumor signature, which revolutionizes HCC detection in the near future.

CONCLUSION

One of the most significant advancements in recent years has been the utilization of AI technologies in medicine. It will almost certainly grow in popularity as a result of its utility in processing and analyzing

massive amounts of available data. However, we should be attentive that there are some limitations that may reduce its acceptability and application in the medical field. Medical professionals need to understand the genuine value of AI and recognize the necessity for it to coexist with the essential requirement for human assessment. Regardless of the significant advancements, it is critical to ensure that medical protocols remain completely transparent.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Jitender Bhardwaj, Assistant Professor, Zoology Department, Kurukshetra University, Kurukshetra, Haryana, India, and Sh. Ram Narayan, Lecturer (English), Government High School, Karsa Dod, Karnal, Haryana, India for timely support.

FOOTNOTES

Author contributions: Mokhria RK designed the outline, performed data acquisition, contributed to the majority of the writing, and proofread the paper; Singh J coordinated the writing of the paper and proofread the paper.

Conflict-of-interest statement: There is no conflict of interest associated with any of the senior authors or other coauthors who contributed their efforts to this manuscript.

Country/Territory of origin: India

ORCID number: Rajesh Kumar Mokhria 0000-0003-2602-7714.

S-Editor: Liu JH
L-Editor: Filipodia
P-Editor: Liu JH

REFERENCES

1 **Vitale A**, Trevisani F, Farinati F, Cillo U. Treatment of Hepatocellular Carcinoma in the Precision Medicine Era: From Treatment Stage Migration to Therapeutic Hierarchy. *Hepatology* 2020; **72**: 2206-2218 [PMID: 32064645 DOI: 10.1002/hep.31187]
2 **Siegel RL**, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 2020; **70**: 7-30 [PMID: 31912902 DOI: 10.3322/caac.21590]
3 **Turing AM**. I.-Computing Machinery and Intelligence. *Mind* 1950; LIX: 433-460 [DOI: 10.1093/mind/Lix.236.433]
4 **Reismann J**, Romualdi A, Kiss N, Minderjahn MI, Kallarackal J, Schad M, Reismann M. Diagnosis and classification of pediatric acute appendicitis by artificial intelligence methods: An investigator-independent approach. *PLoS One* 2019; **14**: e0222030 [PMID: 31553729 DOI: 10.1371/journal.pone.0222030]
5 **Reichling C**, Taieb J, Derangere V, Klopfenstein Q, Le Malicot K, Gornet JM, Becheur H, Fein F, Cojocarasu O, Kaminsky MC, Lagasse JP, Luet D, Nguyen S, Etienne PL, Gasmi M, Vanoli A, Perrier H, Puig PL, Emile JF, Lepage C, Ghiringhelli F. Artificial intelligence-guided tissue analysis combined with immune infiltrate assessment predicts stage III colon cancer outcomes in PETACC08 study. *Gut* 2020; **69**: 681-690 [PMID: 31780575 DOI: 10.1136/gutjnl-2019-319292]
6 **Maeda Y**, Kudo SE, Mori Y, Misawa M, Ogata N, Sasanuma S, Wakamura K, Oda M, Mori K, Ohtsuka K. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest Endosc* 2019; **89**: 408-415 [PMID: 30268542 DOI: 10.1016/j.gie.2018.09.024]
7 **Kaul V**, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointest Endosc* 2020; **92**: 807-812 [PMID: 32565184 DOI: 10.1016/j.gie.2020.06.040]
8 **Briceño J**. Artificial intelligence and organ transplantation: challenges and expectations. *Curr Opin Organ Transplant* 2020; **25**: 393-398 [PMID: 32487888 DOI: 10.1097/MOT.0000000000000775]
9 **Yang YJ**, Bang CS. Application of artificial intelligence in gastroenterology. *World J Gastroenterol* 2019; **25**: 1666-1683 [PMID: 31011253 DOI: 10.3748/wjg.v25.i14.1666]
10 **Taninaga J**, Nishiyama Y, Fujibayashi K, Gunji T, Sasabe N, Iijima K, Naito T. Prediction of future gastric cancer risk using a machine learning algorithm and comprehensive medical check-up data: A case-control study. *Sci Rep* 2019; **9**: 12384 [PMID: 31455831 DOI: 10.1038/s41598-019-48769-y]

11 **Bharti P**, Mittal D, Ananthasivan R. Preliminary Study of Chronic Liver Classification on Ultrasound Images Using an Ensemble Model. *Ultrason Imaging* 2018; **40**: 357-379 [PMID: 30015593 DOI: 10.1177/0161734618787447]

12 **Liu X**, Song JL, Wang SH, Zhao JW, Chen YQ. Learning to Diagnose Cirrhosis with Liver Capsule Guided Ultrasound Image Classification. *Sensors (Basel)* 2017; **17** [PMID: 28098774 DOI: 10.3390/s17010149]

13 **Schmauch B**, Herent P, Jehanno P, Dehaene O, Saillard C, Aubé C, Luciani A, Lassau N, Jégou S. Diagnosis of focal liver lesions from ultrasound using deep learning. *Diagn Interv Imaging* 2019; **100**: 227-233 [PMID: 30926443 DOI: 10.1016/j.diii.2019.02.009]

14 **Guo LH**, Wang D, Qian YY, Zheng X, Zhao CK, Li XL, Bo XW, Yue WW, Zhang Q, Shi J, Xu HX. A two-stage multi-view learning framework based computer-aided diagnosis of liver tumors with contrast enhanced ultrasound images. *Clin Hemorheol Microcirc* 2018; **69**: 343-354 [PMID: 29630528 DOI: 10.3233/CH-170275]

15 **European Association for the Study of the Liver**. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol* 2018; **69**: 182-236 [PMID: 29628281 DOI: 10.1016/j.jhep.2018.03.019]

16 **Heimbach JK**, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, Zhu AX, Murad MH, Marrero JA. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology* 2018; **67**: 358-380 [PMID: 28130846 DOI: 10.1002/hep.29086]

17 **Mokrane FZ**, Lu L, Vavasseur A, Otal P, Peron JM, Luk L, Yang H, Ammari S, Saenger Y, Rousseau H, Zhao B, Schwartz LH, Dercle L. Radiomics machine-learning signature for diagnosis of hepatocellular carcinoma in cirrhotic patients with indeterminate liver nodules. *Eur Radiol* 2020; **30**: 558-570 [PMID: 31444598 DOI: 10.1007/s00330-019-06347-w]

18 **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 [PMID: 29059036 DOI: 10.1148/radiol.2017170706]

19 **Vivanti R**, Szeskin A, Lev-Cohain N, Sosna J, Joskowicz L. Automatic detection of new tumors and tumor burden evaluation in longitudinal liver CT scan studies. *Int J Comput Assist Radiol Surg* 2017; **12**: 1945-1957 [PMID: 28856515 DOI: 10.1007/s11548-017-1660-z]

20 **Li W**, Jia F, Hu Q. Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks. *J Comput Commun* 2015; **3**: 146-151 [DOI: 10.4236/jcc.2015.311023]

21 **Hamm CA**, Wang CJ, Savic LJ, Ferrante M, Schobert I, Schlachter T, Lin M, Duncan JS, Weinreb JC, Chapiro J, Letzen B. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. *Eur Radiol* 2019; **29**: 3338-3347 [PMID: 31016442 DOI: 10.1007/s00330-019-06205-9]

22 **Jansen MJA**, Kuijf HJ, Veldhuis WB, Wessels FJ, Viergever MA, Pluim JPW. Automatic classification of focal liver lesions based on MRI and risk factors. *PLoS One* 2019; **14**: e0217053 [PMID: 31095624 DOI: 10.1371/journal.pone.0217053]

23 **Preis O**, Blake MA, Scott JA. Neural network evaluation of PET scans of the liver: a potentially useful adjunct in clinical interpretation. *Radiology* 2011; **258**: 714-721 [PMID: 21339347 DOI: 10.1148/radiol.10100547]

24 **Kiani A**, Uyumazturk B, Rajpurkar P, Wang A, Gao R, Jones E, Yu Y, Langlotz CP, Ball RL, Montine TJ, Martin BA, Berry GJ, Ozawa MG, Hazard FK, Brown RA, Chen SB, Wood M, Allard LS, Ylagan L, Ng AY, Shen J. Impact of a deep learning assistant on the histopathologic classification of liver cancer. *NPJ Digit Med* 2020; **3**: 23 [PMID: 32140566 DOI: 10.1038/s41746-020-0232-8]

25 **Liao H**, Xiong T, Peng J, Xu L, Liao M, Zhang Z, Wu Z, Yuan K, Zeng Y. Classification and Prognosis Prediction from Histopathological Images of Hepatocellular Carcinoma by a Fully Automated Pipeline Based on Machine Learning. *Ann Surg Oncol* 2020; **27**: 2359-2369 [PMID: 31916093 DOI: 10.1245/s10434-019-08190-1]

26 **Villanueva A**. Hepatocellular Carcinoma. *N Engl J Med* 2019; **380**: 1450-1462 [PMID: 30970190 DOI: 10.1056/NEJMra1713263]

27 **Gillies RJ**, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016; **278**: 563-577 [PMID: 26579733 DOI: 10.1148/radiol.2015151169]

28 **Lambin P**, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; **14**: 749-762 [PMID: 28975929 DOI: 10.1038/nrclinonc.2017.141]

29 **Erstad DJ**, Tanabe KK. Prognostic and Therapeutic Implications of Microvascular Invasion in Hepatocellular Carcinoma. *Ann Surg Oncol* 2019; **26**: 1474-1493 [PMID: 30788629 DOI: 10.1245/s10434-019-07227-9]

30 **Ma X**, Wei J, Gu D, Zhu Y, Feng B, Liang M, Wang S, Zhao X, Tian J. Preoperative radiomics nomogram for microvascular invasion prediction in hepatocellular carcinoma using contrast-enhanced CT. *Eur Radiol* 2019; **29**: 3595-3605 [PMID: 30770969 DOI: 10.1007/s00330-018-5985-y]

31 **Xu X**, Zhang HL, Liu QP, Sun SW, Zhang J, Zhu FP, Yang G, Yan X, Zhang YD, Liu XS. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. *J Hepatol* 2019; **70**: 1133-1144 [PMID: 30876945 DOI: 10.1016/j.jhep.2019.02.023]

32 **Zhou W**, Zhang L, Wang K, Chen S, Wang G, Liu Z, Liang C. Malignancy characterization of hepatocellular carcinomas based on texture analysis of contrast-enhanced MR images. *J Magn Reson Imaging* 2017; **45**: 1476-1484 [PMID: 27626270 DOI: 10.1002/jmri.25454]

33 **Dong Y**, Zhou L, Xia W, Zhao XY, Zhang Q, Jian JM, Gao X, Wang WP. Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma: Initial Application of a Radiomic Algorithm Based on Grayscale Ultrasound Images. *Front Oncol* 2020; **10**: 353 [PMID: 32266138 DOI: 10.3389/fonc.2020.00353]

34 **Ji GW**, Zhu FP, Xu Q, Wang K, Wu MY, Tang WW, Li XC, Wang XH. Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: A multi-institutional study. *EBioMedicine* 2019; **50**: 156-165 [PMID: 31735556 DOI: 10.1016/j.ebiom.2019.10.057]

35 **Ho WH**, Lee KT, Chen HY, Ho TW, Chiu HC. Disease-free survival after hepatic resection in hepatocellular carcinoma patients: a prediction approach using artificial neural network. *PLoS One* 2012; **7**: e29179 [PMID: 22235270 DOI:

10.1371/journal.pone.0029179]
36 **Chiu HC**, Ho TW, Lee KT, Chen HY, Ho WH. Mortality predicted accuracy for hepatocellular carcinoma patients with hepatic resection using artificial neural network. *ScientificWorldJournal* 2013; **2013**: 201976 [PMID: 23737707 DOI: 10.1155/2013/201976]
37 **Shi HY**, Lee KT, Lee HH, Ho WH, Sun DP, Wang JJ, Chiu CC. Comparison of artificial neural network and logistic regression models for predicting in-hospital mortality after primary liver cancer surgery. *PLoS One* 2012; **7**: e35781 [PMID: 22563399 DOI: 10.1371/journal.pone.0035781]
38 **Forner A**, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet* 2018; **391**: 1301-1314 [PMID: 29307467 DOI: 10.1016/S0140-6736(18)30010-2]
39 **Morshid A**, Elsayes KM, Khalaf AM, Elmohr MM, Yu J, Kaseb AO, Hassan M, Mahvash A, Wang Z, Hazle JD, Fuentes D. A machine learning model to predict hepatocellular carcinoma response to transcatheter arterial chemoembolization. *Radiol Artif Intell* 2019; **1** [PMID: 31858078 DOI: 10.1148/ryai.2019180021]
40 **Peng J**, Kang S, Ning Z, Deng H, Shen J, Xu Y, Zhang J, Zhao W, Li X, Gong W, Huang J, Liu L. Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. *Eur Radiol* 2020; **30**: 413-424 [PMID: 31332558 DOI: 10.1007/s00330-019-06318-1]
41 **Liu D**, Liu F, Xie X, Su L, Liu M, Kuang M, Huang G, Wang Y, Zhou H, Wang K, Lin M, Tian J. Accurate prediction of responses to transarterial chemoembolization for patients with hepatocellular carcinoma by using artificial intelligence in contrast-enhanced ultrasound. *Eur Radiol* 2020; **30**: 2365-2376 [PMID: 31900703 DOI: 10.1007/s00330-019-06553-6]
42 **Abajian A**, Murali N, Savic LJ, Laage-Gaupp FM, Nezami N, Duncan JS, Schlachter T, Lin M, Geschwind JF, Chapiro J. Predicting Treatment Response to Image-Guided Therapies Using Machine Learning: An Example for Trans-Arterial Treatment of Hepatocellular Carcinoma. *J Vis Exp* 2018 [PMID: 30371657 DOI: 10.3791/58382]
43 **Mähringer-Kunz A**, Wagner F, Hahn F, Weinmann A, Brodehl S, Schotten S, Hinrichs JB, Düber C, Galle PR, Pinto Dos Santos D, Kloeckner R. Predicting survival after transarterial chemoembolization for hepatocellular carcinoma using a neural network: A Pilot Study. *Liver Int* 2020; **40**: 694-703 [PMID: 31943703 DOI: 10.1111/liv.14380]
44 **Sieghart W**, Hucke F, Pinter M, Graziadei I, Vogel W, Müller C, Heinzl H, Trauner M, Peck-Radosavljevic M. The ART of decision making: retreatment with transarterial chemoembolization in patients with hepatocellular carcinoma. *Hepatology* 2013; **57**: 2261-2273 [PMID: 23316013 DOI: 10.1002/hep.26256]
45 **Adhoute X**, Penaranda G, Naude S, Raoul JL, Perrier H, Bayle O, Monnet O, Beaurain P, Bazin C, Pol B, Folgoc GL, Castellani P, Bronowicki JP, Bourlière M. Retreatment with TACE: the ABCR SCORE, an aid to the decision-making process. *J Hepatol* 2015; **62**: 855-862 [PMID: 25463541 DOI: 10.1016/j.jhep.2014.11.014]
46 **Kim BK**, Shim JH, Kim SU, Park JY, Kim DY, Ahn SH, Kim KM, Lim YS, Han KH, Lee HC. Risk prediction for patients with hepatocellular carcinoma undergoing chemoembolization: development of a prediction model. *Liver Int* 2016; **36**: 92-99 [PMID: 25950442 DOI: 10.1111/liv.12865]
47 **Ziv E**, Yarmohammadi H, Boas FE, Petre EN, Brown KT, Solomon SB, Solit D, Reidy D, Erinjeri JP. Gene Signature Associated with Upregulation of the Wnt/β-Catenin Signaling Pathway Predicts Tumor Response to Transarterial Embolization. *J Vasc Interv Radiol* 2017; **28**: 349-355.e1 [PMID: 28126478 DOI: 10.1016/j.jvir.2016.11.004]
48 **Liang JD**, Ping XO, Tseng YJ, Huang GT, Lai F, Yang PM. Recurrence predictive models for patients with hepatocellular carcinoma after radiofrequency ablation using support vector machines with feature selection methods. *Comput Methods Programs Biomed* 2014; **117**: 425-434 [PMID: 25278224 DOI: 10.1016/j.cmpb.2014.09.001]
49 **Dong RZ**, Yang X, Zhang XY, Gao PT, Ke AW, Sun HC, Zhou J, Fan J, Cai JB, Shi GM. Predicting overall survival of patients with hepatocellular carcinoma using a three-category method based on DNA methylation and machine learning. *J Cell Mol Med* 2019; **23**: 3369-3374 [PMID: 30784182 DOI: 10.1111/jcmm.14231]
50 **Xu RH**, Wei W, Krawczyk M, Wang W, Luo H, Flagg K, Yi S, Shi W, Quan Q, Li K, Zheng L, Zhang H, Caughey BA, Zhao Q, Hou J, Zhang R, Xu Y, Cai H, Li G, Hou R, Zhong Z, Lin D, Fu X, Zhu J, Duan Y, Yu M, Ying B, Zhang W, Wang J, Zhang E, Zhang C, Li O, Guo R, Carter H, Zhu JK, Hao X, Zhang K. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat Mater* 2017; **16**: 1155-1161 [PMID: 29035356 DOI: 10.1038/nmat4997]
51 **Villanueva A**, Portela A, Sayols S, Battiston C, Hoshida Y, Méndez-González J, Imbeaud S, Letouzé E, Hernandez-Gea V, Cornella H, Pinyol R, Solé M, Fuster J, Zucman-Rossi J, Mazzaferro V, Esteller M, Llovet JM; HEPTROMIC Consortium. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. *Hepatology* 2015; **61**: 1945-1956 [PMID: 25645722 DOI: 10.1002/hep.27732]
52 **Yeh CC**, Goyal A, Shen J, Wu HC, Strauss JA, Wang Q, Gurvich I, Safyan RA, Manji GA, Gamble MV, Siegel AB, Santella RM. Global Level of Plasma DNA Methylation is Associated with Overall Survival in Patients with Hepatocellular Carcinoma. *Ann Surg Oncol* 2017; **24**: 3788-3795 [PMID: 28593503 DOI: 10.1245/s10434-017-5913-4]
53 **Azer SA**. Deep learning with convolutional neural networks for identification of liver masses and hepatocellular carcinoma: A systematic review. *World J Gastrointest Oncol* 2019; **11**: 1218-1230 [PMID: 31908726 DOI: 10.4251/wjgo.v11.i12.1218]

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 October 28; 3(4): 105-116

DOI: 10.35712/aig.v3.i4.105

ISSN 2644-3236 (online)

ORIGINAL ARTICLE

Prospective Study

Dietary counseling based on artificial intelligence for patients with nonalcoholic fatty liver disease

Yumi Kusano, Kei Funada, Mayumi Yamaguchi, Miwa Sugawara, Masaya Tamano

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Unsolicited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): 0
Grade C (Good): C, C
Grade D (Fair): D
Grade E (Poor): E, E, E

P-Reviewer: Liu Y, China; MORYA AK, India; Yang JS, China; Zhang LL, China

Received: June 22, 2022
Peer-review started: June 22, 2022
First decision: July 11, 2022
Revised: July 13, 2022
Accepted: October 26, 2022
Article in press: October 26, 2022
Published online: October 28, 2022

Yumi Kusano, Kei Funada, Mayumi Yamaguchi, Masaya Tamano, Department of Gastroenterology, Dokkyo Medical University Saitama Medical Center, Koshigaya 343-8555, Saitama, Japan

Miwa Sugawara, Nutrition Unit, Dokkyo Medical University Saitama Medical Center, Koshigaya 343-8555, Saitama, Japan

Corresponding author: Masaya Tamano, PhD, Professor, Department of Gastroenterology, Dokkyo Medical University Saitama Medical Center, 2-1-50 Minami-Koshigaya, Koshigaya 343-8555, Saitama, Japan. mstamano@dokkyomed.ac.jp

Abstract

BACKGROUND

About 25% of the general population in Japan are reported to have nonalcoholic fatty liver disease (NAFLD). NAFLD and nonalcoholic steatohepatitis carry a risk of progressing further to hepatocellular carcinoma. The primary treatment for NAFLD is dietary therapy. Dietary counseling plays an essential role in dietary therapy. Although artificial intelligence (AI)-based nutrition management software applications have been developed and put into practical use in recent years, the majority focus on weight loss or muscle strengthening, and no software has been developed for patient use in clinical practice.

AIM

To examine whether effective dietary counseling is possible using AI-based nutrition management software.

METHODS

NAFLD patients who had been assessed using an AI-based nutrition management software application (Calomeal) that automatically analyzed images of meals photographed by patients and agreed to receive dietary counseling were given dietary counseling. Blood biochemistry tests were performed before (baseline) and 6 mo after (6M follow-up) dietary counseling. After the dietary counseling, the patients were asked to complete a questionnaire survey.

RESULTS

A total of 29 patients diagnosed with NAFLD between August 2020 and March 2022 were included. There were significant decreases in liver enzyme and triglyceride levels at the 6M follow-up compared to baseline. The food analysis

capability of the AI used by Calomeal in this study was 75.1%. Patient satisfaction with the AI-based dietary counselling was high.

CONCLUSION
AI-based nutrition management appeared to raise awareness of dietary habits among NAFLD patients. However, it did not directly alleviate the burden of registered dietitians, and improvements are much anticipated.

Key Words: Artificial intelligence; Dietary counselling; Nonalcoholic fatty liver disease; Nonalcoholic steatohepatitis; Nutrition management software applications

Core Tip: Use of artificial intelligence (AI)-based nutrition management software (Calomeal) appeared to raise awareness of nonalcoholic fatty liver disease patients' dietary habits, and they showed significant decreases in liver enzyme and triglyceride levels at the 6-mo follow-up compared to baseline. The food analysis capability of the AI package used in this study was 75.1%, and patient satisfaction with the AI-based dietary counselling was high. However, due to the limitations of the food analysis capabilities of AI, it did not directly alleviate the burden of registered dietitians, and improvements in the analytical capabilities of AI are much anticipated.

Citation: Kusano Y, Funada K, Yamaguchi M, Sugawara M, Tamano M. Dietary counseling based on artificial intelligence for patients with nonalcoholic fatty liver disease. *Artif Intell Gastroenterol* 2022; 3(4): 105-116
URL: https://www.wjgnet.com/2644-3236/full/v3/i4/105.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i4.105

INTRODUCTION

About 25% of the general population in Japan are reported to have nonalcoholic fatty liver disease (NAFLD), and this figure is expected to increase to 39.3% by 2030[1]. NAFLD includes nonalcoholic fatty liver (NAFL) and nonalcoholic steatohepatitis (NASH). NAFL and NASH carry a risk of progressing further to hepatocellular carcinoma. Although the risk of developing liver cancer from NAFL is low (0.44 per 1000 persons per year), the risk increases with the progression of liver pathology (5.29 per 1000 persons per year in patients with NASH and 0.45 to 22.6 per 1000 persons per year in patients with liver cirrhosis)[2,3].

The primary treatment for NAFLD is dietary therapy. Dietary counseling plays an essential role in dietary therapy. In dietary counseling, patients self-report all the food items that they consumed over the previous 1 wk, and during the interview, registered dietitians calculate the caloric and nutritive values of the food items. Limitations exist with patients' memory certainty and the listening and calculation skills of the dietitians in the field. In recent years, pictures taken with digital cameras have been used in combination with interviews. Neither of these is very efficient, because dietary counseling begins when patients come in for their consultation, and the caloric and nutrient intakes are calculated based on patients' recalled food items consumed in the past.

Artificial intelligence (AI)-based image analysis has penetrated deeply in our daily lives, and the field of medicine is no exception. In gastroenterology, it is used to diagnose colon polyps, *Helicobacter pylori* infection, and stomach cancer[4-8]. Although AI-based nutrition management software applications have been developed and put into practical use in recent years, the majority focus on weight loss or muscle strengthening, and no software has been developed for patient use in clinical practice.

Thus, the present study examined whether effective dietary counseling is possible using food intake data of NAFLD patients that had been automatically analyzed with an AI-based nutrition management software application.

MATERIALS AND METHODS

Patients

This prospective study was conducted in compliance with the ethics guidelines of the 2008 Declaration of Helsinki. Approval was obtained from the Biomedical Ethics Committee of the authors' affiliated hospital (No. 2014). Written informed consent was obtained from all patients.

Patients clinically diagnosed with NAFLD between August 2020 and March 2022 were included as subjects. NAFLD was diagnosed when fatty liver was observed in patients with an alcohol consumption equivalent to ≤ 30 g of ethanol per day in men and ≤ 20 g of ethanol per day in women[9]. Fatty liver was diagnosed when the following three criteria were confirmed on abdominal ultrasound examination: (1) Increased hepatic echogenicity; (2) positive liver-kidney contrast; and (3) deep ultrasound attenuation in the liver. Patients with chronic hepatitis B, chronic hepatitis C, autoimmune hepatitis, and primary biliary cholangitis were excluded. Patients with non-compensated liver cirrhosis and those with concomitant hepatocellular carcinoma were also excluded.

Dietary counseling using AI

"Calomeal" is a software application developed by Life Log Technology, Inc. (Tokyo, Japan) that can easily record and manage one's daily meals and physical activity. It has been available in Japan since December 2015. The "Calomeal" software is commercially available, and anyone can purchase it. When a user takes a picture of a meal using devices such as smartphones, the cloud-based AI calculates the nutritive values of the food in the photograph, and the results are sent to the user's device (Figure 1). The AI of Calomeal was compiled using machine learning of food image data collected from major restaurant chains, food manufacturers, and everyday home-cooked meals. It can identify approximately 18000 food images. In addition to total caloric intake, Calomeal simultaneously calculates the nutritive values of proteins, lipids, carbohydrates, glucides, dietary fiber, and salt content.

This study used a customized version of Calomeal that was developed in cooperation with Life Log Technology. The following three areas were customized: (1) Changing the device from a smartphone to an iPad Mini; (2) simplifying food selection and offering more detailed options on consumed quantities; and (3) redirecting the AI analysis results to the study's dedicated personal computer and not to the patients' devices.

More specifically, since some NAFLD patients were older, iPad Minis were used as the device instead of smartphones because iPad Minis have larger display screens (Figure 2). When patients took pictures of a food item, the AI suggested the name of three possible food choices. The patients then selected from the three choices the food item that they consumed. When none of the offered three choices matched, "no suitable choice" was selected. Next, the patients chose the quantity of food that they consumed from the list of options. Eight options were available. Setting 100% as the normal serving size, the options were as follows: 200%, 150%, 100%, 66% (about two-thirds), 50%, 33% (about one-third), 10% (about one bite), and 0% (although a picture was taken, no food was consumed). AI calculated the nutritive values once these tasks were completed. The analysis results were forwarded to the study's dedicated personal computer managed by the authors (Figure 3). Breakfast, lunch, dinner, or snacks were automatically determined based on the time when the foods were photographed.

After photographing one week's worth of pictures of meals, the patients came in for their dietary counseling session. By the time of the patients' visits, the AI had prepared a list of a patient's one week's worth of photographed foods (Figure 4) and bar charts showing the amounts of caloric, protein, lipid, carbohydrate, glucide, dietary fiber, and salt intakes (Figure 5). Registered dietitians participating in this study conducted dietary counseling while presenting the abovementioned data.

The photographed food images for which the patients could not find a suitable choice at the time of taking the picture were examined by the registered dietitians before the patients' visits, and after checking the photographed images, the registered dietitians entered the correct item name.

Principle of Calomeal

Around dozens of pieces of photograph data for one food (or one product) are prepared for machine learning. These photographs are learned by deep neural network, and foods (or a product) are analyzed by the pattern of the color and form.

The nutrient of common foods was calculated based on "standard table of food composition" announced by Japanese Ministry of Education, Culture, Sports, Science and Technology by dietitian of Life Log Technology company. The nutrient of foods of major restaurant chains and food manufacturers was calculated using the data published in the home page of each company. When the nutrient was not announced in home page of the company, the dietitian calculated the nutrient from announced raw materials.

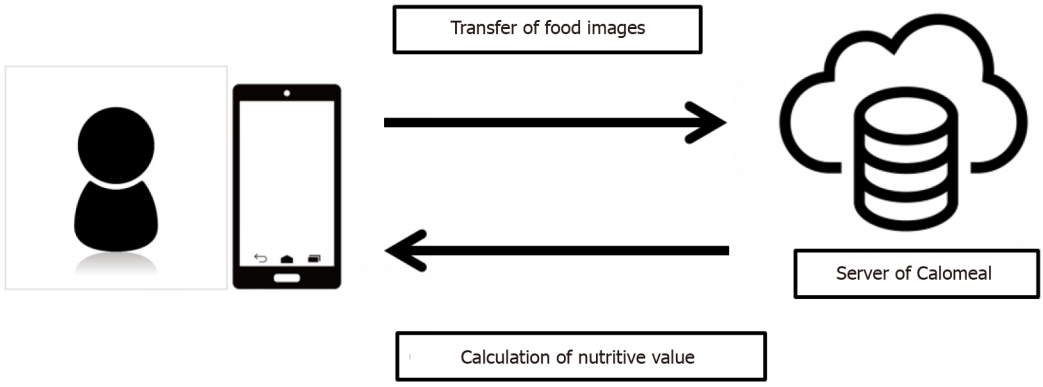
Effects of dietary counseling using Calomeal

Of all the patients enrolled in the study, patients who agreed to receive dietary counseling were given dietary counseling using Calomeal. Blood biochemistry tests were performed before (baseline) and 6 mo after (6M follow-up) dietary counseling. Aspartate aminotransferase (AST), alanine aminotransferase (ALT), γ-glutamyltransferase (GGT), total cholesterol (T-cho), and triglyceride (TG) levels were compared between the baseline and 6M follow-up.

Body weight of all patients was compared between the baseline and 6M follow-up.

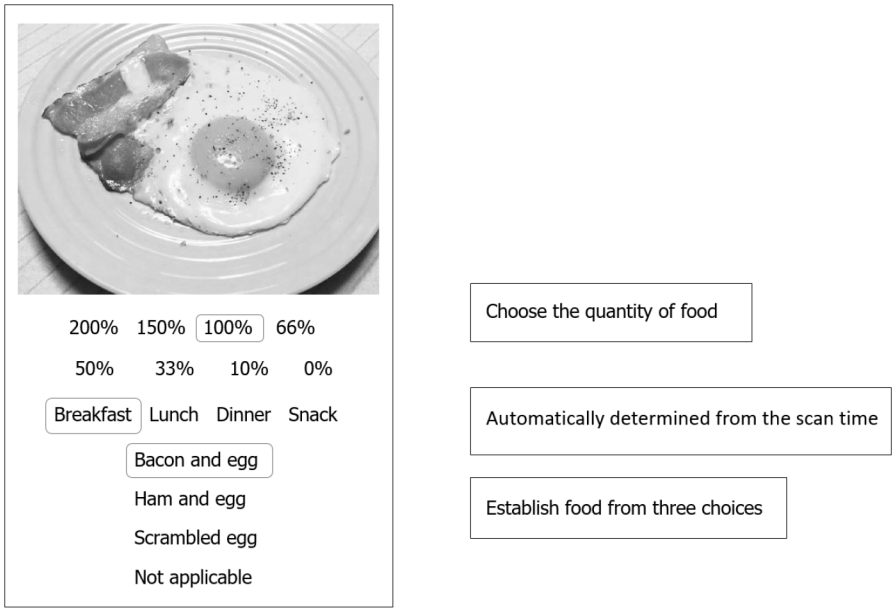
Analysis capability of AI

As noted earlier, the AI used in the present study was capable of analyzing approximately 18000 meal



DOI: 10.35712/aig.v3.i4.105 **Copyright** ©The Author(s) 2022.

Figure 1 The concept of Calomeal. When users take a photograph of their meals with their smartphones, the artificial intelligence on the Calomeal server calculates the nutritive value of the food included in the photograph, and the results are sent to the user's device.



DOI: 10.35712/aig.v3.i4.105 **Copyright** ©The Author(s) 2022.

Figure 2 Display on the iPad Mini. Instead of using a smartphone as the device, an iPad Mini, which has a larger screen, is used. When a patient photographs a food item, the artificial intelligence suggests the names of three possible food choices. The patient selects the food that he or she ate from one of the three choices. If none of the three choices matches, the patient chooses "no suitable choice." Next, the patient selects the quantity of food consumed from the list of options. Eight options are available. Setting 100% as the normal serving size, the choices are as follows: 200%, 150%, 100%, 66% (about two-thirds), 50%, 33% (about one-third), 10% (about one bite), and 0% (picture was taken but no food was consumed). Breakfast, lunch, dinner, or snacks are automatically determined based on the time that the foods were photographed.

items. However, since most of these items were collected from menus of major restaurant chains and food manufacturers, its ability to identify everyday home-cooked meals was unknown. The AI's ability to analyze food images was evaluated by calculating the percentage (%) of unidentified foods (*i.e.*, foods for which the results of the AI's automated analysis did not match the foods actually consumed by the patients, thereby categorized under "no suitable choice") among all food items.

In addition, the time spent by registered dietitians before the patient visits in entering the correct food names of the unidentified "no suitable choice" food items was also measured.

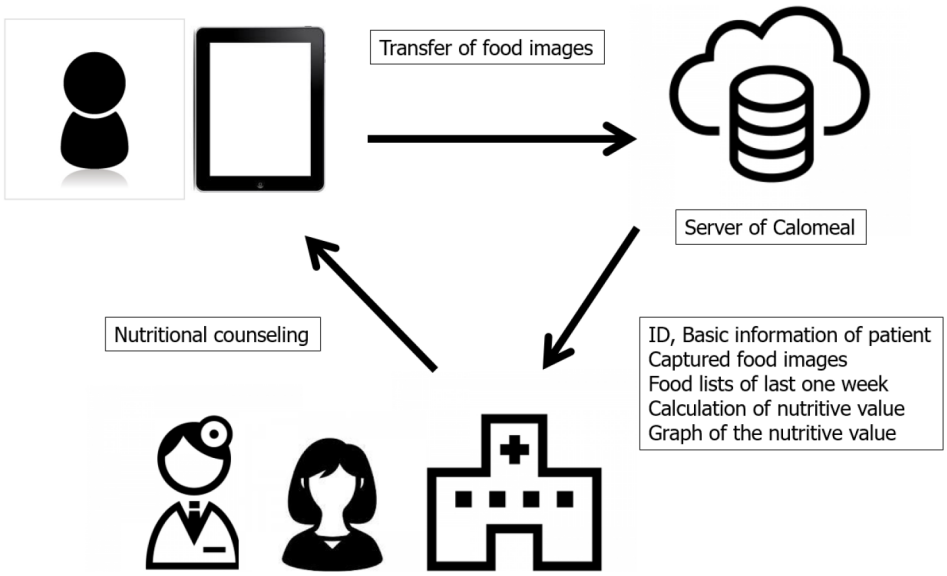
Acceptance of Calomeal

After the dietary counseling, subjects in the Calomeal Group were asked to complete a questionnaire survey consisting of the following questions:

Question 1: Were you glad that you were given an AI-based dietary counseling session? (Yes or No).

Question 2: Did you find the dietary counseling rewarding? (Yes or No).

Question 3: Have you become more conscious of improving your dietary lifestyle? (Yes or No).



DOI: 10.35712/aig.v3.i4.105

Figure 3 Calomeal concepts that were customized in this study. The food images photographed with an iPad Mini are analyzed by the artificial intelligence on the Calomeal server. The following information is sent to the dedicated personal computer at the authors' hospital: ID, basic information of patients, captured food images, food lists of last one week, calculation of nutritive value, and bar charts of the nutritive value. Registered dietitians conduct dietary counseling based on these data.

Question 4: Would you like to receive another AI-based dietary counseling session? (Yes or No).

Statistical analysis

Continuous data, such as those from the blood biochemistry tests, are shown as the mean ± standard deviation. The paired Wilcoxon test was used to test the difference in each parameter between the start of observation (baseline) and after 6 mo (6M follow-up). $P < 0.05$ was considered significant.

RESULTS

There were 29 patients who agreed and received dietary counseling using Calomeal. The operation of Calomeal using iPads was accepted by all 29 patients. Table 1 shows the patients' characteristics. The patients had been taking all of these medications before starting this study, and no new drugs were initiated after the start of the observation in this study.

Effects of dietary counseling

Table 2 shows the AST, ALT, GGT, hemoglobin A1c (HbA1c), T-cho, and TG levels and body weight at baseline and 6M follow-up. AST, ALT, GGT, and TG levels were significantly lower ($P = 0.0088$, 0.0133, 0.0494, and 0.0246, respectively) at the 6M follow-up compared to the levels at baseline. Body weight was significantly lower ($P = 0.0472$) at the 6M follow-up compared to that at baseline.

Analysis capability of AI

Table 3 shows the total number of food items photographed in 1 wk of all patients in the Calomeal Group, the number of food items categorized under "no suitable choice", and the food analysis capability of the AI. The mean number of total photographed food items was 62.6 (20 to 104), the mean number of food items categorized under "no suitable choice" was 15.0 (1 to 32), and the mean analysis capability was 75.1% (51.5 to 98.6%). Before dietary counseling sessions, registered dietitians spent on average 25.9 min (4.5 to 67.0 min) identifying the food items categorized under "no suitable choice."

Acceptance of Calomeal

Table 4 shows the findings of the responses to the questionnaire survey. When the patients were asked "Were you glad that you were given an AI-based dietary counseling session?" in Question 1, all 29 patients said "Yes." When the patients were asked "Did you find the dietary counseling rewarding?" in Question 2, 15 of the 29 patients responded "Yes." When the patients were asked "Have you become more conscious of improving your dietary lifestyle after the dietary counseling session?" in Question 3, all 29 patients responded "Yes." When the patients were asked "Would you like to receive another AI-based dietary counseling session?" in Question 4, four of the 29 patients responded "No."

Table 1 Patients' characteristics (*n* = 29)

Male/Female	16/13
Age (yr)	56.4 ± 14.3
Body weight (kg)	74.1 ± 13.1
BMI (kg/m^2)	28.4 ± 5.2
Metabolic diseases	
Diabetes mellitus (Yes/No)	10/19
Dyslipidemia (Yes/No)	14/15
Concomitant drugs	
SGLT2 inhibitor	2
DPP-4 inhibitor	3
Thiazolidinedione	
GLP-1 agonist	1
Statin	7
Bezafibrate	
Pemafibrate	
EPA and DHA preparation	1
AST (U/L)	50.2 ± 33.4
ALT (U/L)	53.8 ± 39.0
GGT (U/L)	80.3 ± 84.3
T-B (mg/dL)	1.3 ± 0.9
Alb (mg/dL)	4.2 ± 0.6
eGFR (mL/min)	71.1 ± 11.5
HbA1c (%)	6.4 ± 0.8
T-cho (mg/dL)	191.3 ± 35.0
TG (mg/dL)	126.5 ± 57.8
WBC (10^3/μL)	5.6 ± 1.5
Hb (g/dL)	14.5 ± 1.7
Plts (10^4/μL)	17.7 ± 7.4

AST: Aspartate aminotransferase; ALT: Alanine aminotransferase; GGT: γ-Glutamyltransferase; T-Bil: Total bilirubin; Alb: Albumin; eGFR: Estimated glomerular filtration rate; HbA1c: Glycated hemoglobin A1c; T-chol: Total cholesterol; TG: Triglycerides; WBC: White blood cells; Hb: Hemoglobin; Plts: Platelets.

DISCUSSION

AI helps diagnose NAFLD through its use in diagnostic imaging procedures such as ultrasound, computed tomography, and magnetic resonance imaging and pathological diagnostic procedures[10-12]. However, there have been no reports of using AI in nutritional therapies and dietary counseling for NAFLD. Thus, the authors focused on this aspect and planned the present study.

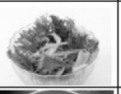
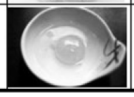
There is a long history of using AI in food analysis. In 1983, Chen *et al*[13] reported that the caloric intake of foods could be calculated using a small wearable computer that can be attached to clothing like a tin badge. This small computer called the eButton is being used for the healthy transformation of homemade foods[14,15]. When the eButton is used solely for identifying foods, the sensitivity is 85.0% and the specificity is 85.8%. Since the small eButton can be worn by attaching it to clothing, privacy is maintained[13]; however, its widespread use by the general public may be difficult because it is a purpose-built computer.

In Japan, there are five types of nutrition management software applications that can be downloaded on smartphones. Although there may be some differences in their analysis capability, all of them are very easy to use[16]. Of these software applications, Oka *et al*[17] used Asken for dietary counseling of

Table 2 Changes of blood biochemistry parameters and body weight

	Baseline	6M Follow-up	*P* value
AST (U/L)	50.2 ± 33.4	34.7 ± 14.7	0.0088
ALT (U/L)	53.8 ± 39.0	35.3 ± 16.8	0.0113
GGT (U/L)	80.3 ± 84.3	66.3 ± 82.9	0.0494
HbA1c (%)	6.4 ± 0.8	6.2 ± 0.6	0.27832
T-cho (mg/dL)	191.3 ± 35.0	189 ± 34.3	0.2109
TG (mg/dL)	126.5 ± 57.8	104.1 ± 60.6	0.0426
Body weight (kg)	74.0 ± 13.1	71.2 ± 12.3	0.0472

AST: Aspartate aminotransferase; ALT: Alanine aminotransferase; GGT: γ-Glutamyltransferase; HbA1c: Glycated hemoglobin A1c; T-chol: Total cholesterol; TG: Triglycerides.

Food name	Image	Intake (%)	Energy (kcal)	Protein (g)	Lipids (g)	Carbohyd rate (g)	Salt (g)	Glucids (g)	Dietary fiber (g)
			1978	64.7	105.7	182.8	11.7	172.5	10.4
Breakfast			450	17.1	25.8	35.9	2.1	34.5	1.4
Butter toast		100	231	5.5	10.6	28.0	0.9	26.8	1.4
Ham and egg		100	190	11.4	15.2	0.6	1.1	0.6	0.0
Ice tea with lemon		100	29	0.2	0.0	7.3	0.1	7.3	0.0
Lunch			515	15.6	23.3	56.4	5.4	54.5	1.9
Soy sauce ramen		66	515	15.6	23.3	56.4	5.4	54.5	1.9
Dinner			1013	32.0	56.6	90.5	4.3	83.5	7.1
Sukiyaki		100	597	21.0	37.6	43.2	3.8	37.9	5.3
Rice		100	202	3.0	0.4	44.5	0.0	44.2	0.4
Salad		100	131	1.2	12.9	2.6	0.2	1.2	1.4
Egg		100	8	6.8	5.7	0.2	0.2	0.2	0.0

DOI: 10.35712/aig.v3.i4.105

Figure 4 Analysis results of the photographed food images. A list is created based on the results of the artificial intelligence calculation using food names and intakes (%) selected by patients. Energy, protein, lipid, carbohydrate, salt, glucide, and dietary fiber content of each meal over 1 wk are shown.

type 2 diabetes mellitus patients and reported its benefit. Of these five software applications, the present study focused on Calomeal. The main reason for choosing Calomeal was its ease of operation.

In the present study, significant decreases in AST, ALT, GGT, and TG levels and body weight were observed at the 6M follow-up compared to the levels at baseline. However, there is an important limitation of this study. The present study did not compare dietary counseling using conventional methods with dietary counseling using Calomeal. The reason was that, in contrary to the authors' expectations, fewer patients gave consent for dietary counseling after being diagnosed with NAFLD during the study period. Thus, it cannot be asserted that this study's dietary counseling using Calomeal is superior to conventional methods. This issue needs to be addressed in future research.

The food analysis capability of the AI used by Calomeal in this study was 75.1%. The food analysis capability using Life Log Technology's own data was reported to be 85.7%, which showed a discrepancy with the present study's findings. One possible reason may be the small percentage of data for everyday

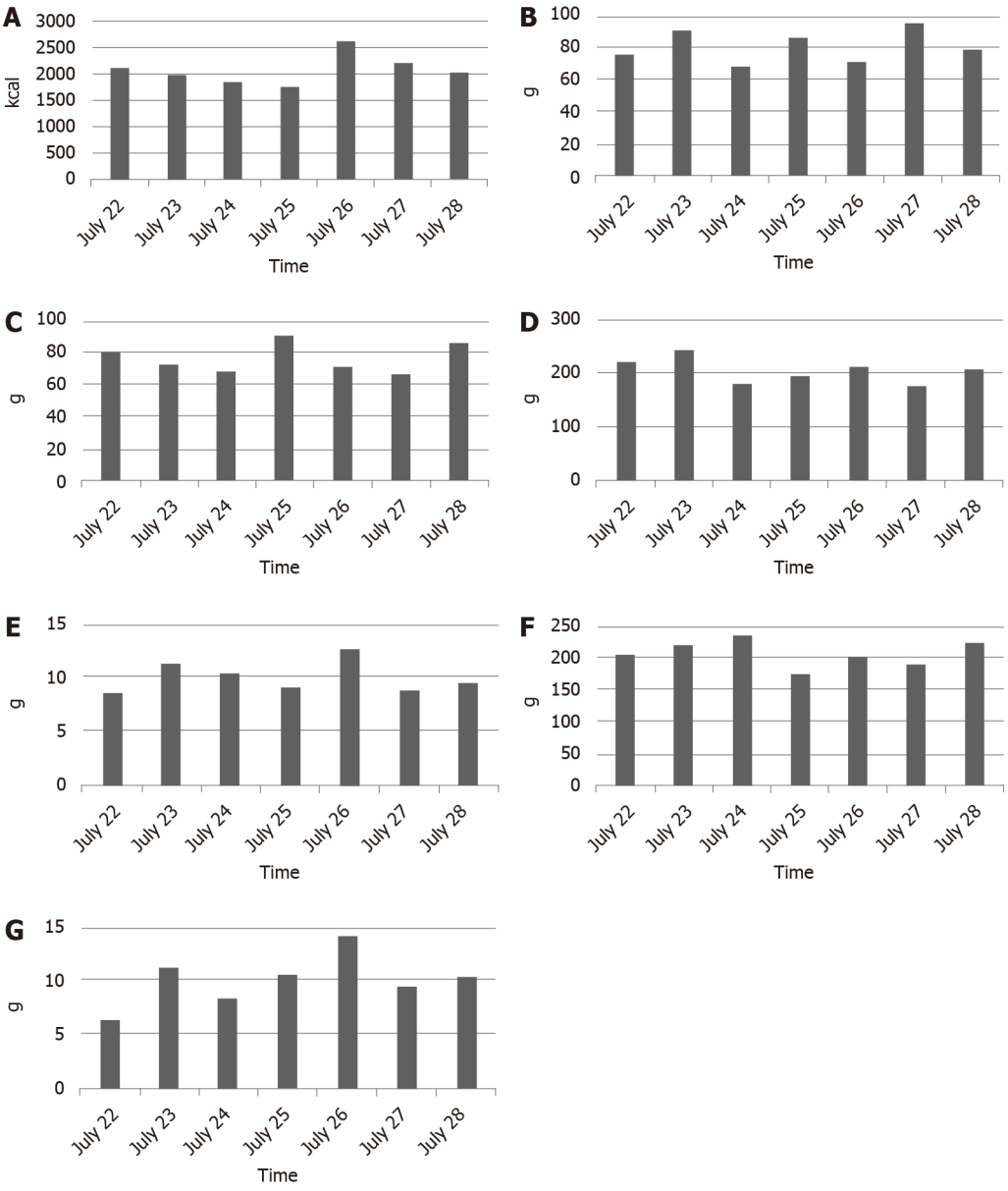
Table 3 Number of photographed food items, analysis rate, and time for confirmation

Case	All foods	No of candidates	Analysis rate (%)	Time for confirmation (min)
1	57	13	77.2	25.0
2	31	2	93.5	4.5
3	90	31	65.6	67.0
4	40	7	82.5	20.5
5	74	20	73.0	31.0
6	69	1	98.6	6.0
7	24	3	87.5	9.5
8	20	9	55.0	12.0
9	47	9	80.9	28.5
10	87	29	66.7	35.5
11	26	12	53.8	14.0
12	35	8	77.1	10.0
13	85	25	70.6	36.5
14	95	8	91.6	8.5
15	53	15	71.7	20.0
16	84	32	61.9	39.5
17	77	13	83.1	14.0
18	51	7	86.3	13.0
19	33	16	51.5	20.0
20	103	9	91.3	19.5
21	32	6	81.3	5.5
22	68	25	63.2	32.5
23	88	30	65.9	50.0
24	41	18	56.1	66.0
25	74	22	70.3	36.0
26	104	16	84.6	34.0
27	84	22	73.8	47.5
28	83	13	84.3	25.5
29	61	13	78.7	20.5
Average	62.6	15.0	75.1	25.9

Table 4 Questionnaire results

Question	Yes	No
1	29	0
2	15	14
3	29	0
4	25	29

home-cooked meals, since the food items registered in Calomeal's AI consisted primarily of major restaurant chain menus, food manufacturers' items, and convenience store products. In the present study, NAFLD patients, especially elderly patients, often consumed unique homemade dishes whose food names were difficult to identify just by looking at the pictures.



DOI: 10.35712/aig.v3.i4.105

Figure 5 Bar charts of nutritive values for 1 wk. A: Daily energy intake; B: Daily protein intake; C: Daily lipid intake; D: Daily carbohydrate intake; E: Daily salt intake; F: Daily glucide intake; G: Daily dietary fiber intake. The bar charts show the daily energy, protein, lipid, carbohydrate, salt, glucide, and dietary fiber intakes for 1 wk.

In other words, there were various daily diets, especially the cooking methods of family diets cannot be unified. These made food illegible and energy intake difficult to calculate. This can lead to significant error in the study.

Assuming that a patient eats four dish items per meal, three meals a day, this means that 4 × 3 × 7 = 84 items should have been photographed in 1 wk. However, as shown in Table 2, there was a large difference among patients, some with very few photographed food items (20 items) and some who had photographed many food items (104 items). The following reasons were given: (1) Patients could not take pictures while they were dining out because they were conscious of their surroundings; (2) patients could not bring the iPads to work; and (3) Japanese "teishoku" set lunches often plate various dishes on one plate, and some patients took pictures of the whole plate as one photograph. In the case of (3), the AI could not analyze multiple food items at once. This issue was reported to Life Log Technology as feedback, and the Calomeal software at the time of writing this paper had already taken this into account.

In the present study, the names of all food images that patients selected as "no suitable choice" on the screen were examined and entered individually by the registered dietitians. This procedure, which on average took 25.9 min (4.5 to 67.0 min), was conducted before the dietary counseling sessions. Since the duration of one dietary counseling session was 30 min, this means that preparation required about the same amount of time. However, when meeting the patient face-to-face, a list of food items pho-

tographed by the patient over the previous 1 wk and bar charts of various nutrients were already prepared. Thus, it is possible that more productive dietary counseling was offered within the allocated 30-min session.

The questionnaire results showed that all of the respondents said that they were glad to have undergone dietary counseling using Calomeal. Similarly, they all responded that they had become aware of their dietary habits. On the other hand, 15 of 29 patients found the content of the dietary counseling rewarding. The majority of patients in the Calomeal Group in this study had a good command of their smartphones, and these patients had free access to the internet. Perhaps there was a lack of originality in the hospital-based dietary counseling, since nutritional information, to some extent, is available when one searches the internet. When the patients were asked whether they would like to receive dietary counseling again using Calomeal, four patients responded "No." During outpatient follow-up interviews, all four patients commented that they "did not want to bring the devices out of their homes or to work." This issue could be resolved by changing the device to a smartphone.

CONCLUSION

When an AI-based nutrition management software application automatically analyzed images of meals photographed by NAFLD patients, there were significant decreases in AST, ALT, GGT, and TG levels after 6 mo (6M follow-up). Thus, this method appeared to raise awareness of dietary habits of NAFLD patients. On the other hand, due to the limitations of the food analysis capabilities of AI, it did not directly alleviate the burden of registered dietitians, and improvements in the analytical capabilities of AI are much anticipated.

ARTICLE HIGHLIGHTS

Research background

Approximately 27000 people a year die from liver cancer in Japan. Liver cancer from non-viral liver disease increases while cancerogenesis from viral liver decreases. In the non-viral liver disease, nonalcoholic fatty liver disease (NAFLD) increases in particular. Therefore, carcinogenesis restraint from NAFLD is urgent business to reduce liver cancer death. Diet therapy is the first choice for the treatment of NAFLD and nutrition education for this purpose becomes extremely important.

Research motivation

The authors paid attention to the nutrition education using the artificial intelligence and led to the idea of this study using the application software called the "Calomeal". The authors have the patients understand the importance of the diet by performing the nutrition education using the artificial intelligence for the NAFLD patients and want to help inhibit the cancerogenesis from NAFLD. A study on optimization of the nutrition education using the artificial intelligence (AI) for NAFLD is the attempt that leads the world and thinks with pioneer positioning of the future health promotion medical care.

Research objectives

Patients clinically diagnosed with NAFLD between August 2020 and March 2022 were included as subjects. "Calomeal" as a software application developed by Life Log Technology, Inc. (Tokyo, Japan) was used for the nutrition education. Blood biochemistry tests were performed before (baseline) and 6 mo after (6M follow-up) dietary counseling. After the dietary counseling, the patients were asked to complete a questionnaire survey.

Research methods

There were significant decreases in liver enzyme and triglyceride levels at the 6M follow-up compared to baseline. The food analysis capability of the AI used by Calomeal in this study was 75.1%. Patient satisfaction with the AI-based dietary counselling was high.

Research results

The authors have the patients understand the importance of the diet because the NAFLD patients receive a nutrition education using the artificial intelligence, and the purpose of this study is to carry a help of the cancerogenesis restraint.

Research conclusions

When an AI-based nutrition management software application automatically analyzed images of meals photographed by NAFLD patients, liver function was improved significantly. On the other hand, due to the limitations of the food analysis capabilities of AI, improvements in the analytical capabilities of AI

are much anticipated.

Research perspectives

The direction of future research is nutrition education using more advanced artificial intelligence to inhibit the carcinogenesis from NAFLD.

FOOTNOTES

Author contributions: Kusano Y reviewed the literature and contributed to manuscript drafting; Funada K analyzed and interpreted the imaging findings; Yamaguchi M drafted the tables and figures; Sugawara M conducted dietary counseling for patients; Tamano M revised the manuscript for important intellectual content; all authors issued final approval for the version to be submitted.

Institutional review board statement: Approval was obtained from the Biomedical Ethics Committee of the authors' affiliated hospital (No. 2014).

Clinical trial registration statement: The clinical trial is registered with clinical research support office of the authors' affiliated hospital. Details can be found at https://dept.dokkyomed.ac.jp/dep-k/gast/.

Informed consent statement: Written informed consent was obtained from all patients for publication of this report and any accompanying images.

Conflict-of-interest statement: None of the authors have any commercial or financial involvements in connection with this study that represent or appear to represent any conflicts of interest.

Data sharing statement: Participants gave informed consent for data sharing.

CONSORT 2010 statement: The authors have read the CONSORT Statement—checklist of items, and the manuscript was prepared and revised according to the CONSORT Statement—checklist of items.

Country/Territory of origin: Japan

ORCID number: Yumi Kusano 0000-0001-8000-0884; Kei Funada 0000-0003-1190-5588; Masaya Tamano 0000-0001-5595-2330.

S-Editor: Liu JH
L-Editor: Wang TQ
P-Editor: Liu JH

REFERENCES

1 **Ito T**, Ishigami M, Zou B, Tanaka T, Takahashi H, Kurosaki M, Maeda M, Thin KN, Tanaka K, Takahashi Y, Itoh Y, Oniki K, Seko Y, Saruwatari J, Kawanaka M, Atsukawa M, Hyogo H, Ono M, Ogawa E, Barnett SD, Stave CD, Cheung RC, Fujishiro M, Eguchi Y, Toyoda H, Nguyen MH. The epidemiology of NAFLD and lean NAFLD in Japan: a meta-analysis with individual and forecasting analysis, 1995-2040. *Hepatol Int* 2021; **15**: 366-379 [PMID: 33580453 DOI: 10.1007/s12072-021-10143-4]

2 **Younossi ZM**, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016; **64**: 73-84 [PMID: 26707365 DOI: 10.1002/hep.28431]

3 **Amarapurkar DN**, Dharod M, Gautam S, Patel N. Risk of development of hepatocellular carcinoma in patients with NASH-related cirrhosis. *Trop Gastroenterol* 2013; **34**: 159-163 [PMID: 24851525 DOI: 10.7869/tg.120]

4 **Misawa M**, Kudo SE, Mori Y, Cho T, Kataoka S, Yamauchi A, Ogawa Y, Maeda Y, Takeda K, Ichimasa K, Nakamura H, Yagawa Y, Toyoshima N, Ogata N, Kudo T, Hisayuki T, Hayashi T, Wakamura K, Baba T, Ishida F, Itoh H, Roth H, Oda M, Mori K. Artificial Intelligence-Assisted Polyp Detection for Colonoscopy: Initial Experience. *Gastroenterology* 2018; **154**: 2027-2029.e3 [PMID: 29653147 DOI: 10.1053/j.gastro.2018.04.003]

5 **Nakashima H**, Kawahira H, Kawachi H, Sakaki N. Artificial intelligence diagnosis of *Helicobacter pylori* infection using blue laser imaging-bright and linked color imaging: a single-center prospective study. *Ann Gastroenterol* 2018; **31**: 462-468 [PMID: 29991891 DOI: 10.20524/aog.2018.0269]

6 **Bang CS**, Lee JJ, Baik GH. Artificial Intelligence for the Prediction of Helicobacter Pylori Infection in Endoscopic Images: Systematic Review and Meta-Analysis Of Diagnostic Test Accuracy. *J Med Internet Res* 2020; **22**: e21983 [PMID: 32936088 DOI: 10.2196/21983]

7 **Jin P**, Ji X, Kang W, Li Y, Liu H, Ma F, Ma S, Hu H, Li W, Tian Y. Artificial intelligence in gastric cancer: a systematic review. *J Cancer Res Clin Oncol* 2020; **146**: 2339-2350 [PMID: 32613386 DOI: 10.1007/s00432-020-03304-9]

8 **Joseph J**, LePage EM, Cheney CP, Pawa R. Artificial intelligence in colonoscopy. *World J Gastroenterol* 2021; **27**: 4802-4817 [PMID: 34447227 DOI: 10.3748/wjg.v27.i29.4802]

9 **Watanabe S**, Hashimoto E, Ikejima K, Uto H, Ono M, Sumida Y, Seike M, Takei Y, Takehara T, Tokushige K, Nakajima A, Yoneda M, Saibara T, Shiota G, Sakaida I, Nakamuta M, Mizuta T, Tsubouchi H, Sugano K, Shimosegawa T; Japanese Society of Gastroenterology; Japan Society of Hepatology. Evidence-based clinical practice guidelines for nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *J Gastroenterol* 2015; **50**: 364-377 [PMID: 25708290 DOI: 10.1007/s00535-015-1050-7]

10 **Decharatanachart P**, Chaiteerakij R, Tiyarattanachai T, Treeprasertsuk S. Application of artificial intelligence in chronic liver diseases: a systematic review and meta-analysis. *BMC Gastroenterol* 2021; **21**: 10 [PMID: 33407169 DOI: 10.1186/s12876-020-01585-5]

11 **Su TH**, Wu CH, Kao JH. Artificial intelligence in precision medicine in hepatology. *J Gastroenterol Hepatol* 2021; **36**: 569-580 [PMID: 33709606 DOI: 10.1111/jgh.15415]

12 **Wong GL**, Yuen PC, Ma AJ, Chan AW, Leung HH, Wong VW. Artificial intelligence in prediction of non-alcoholic fatty liver disease and fibrosis. *J Gastroenterol Hepatol* 2021; **36**: 543-550 [PMID: 33709607 DOI: 10.1111/jgh.15385]

13 **Chen HC**, Jia W, Yue Y, Li Z, Sun YN, Fernstrom JD, Sun M. Model-based measurement of food portion size for image-based dietary assessment using 3D/2D registration. *Meas Sci Technol* 2013; **24** [PMID: 24223474 DOI: 10.1088/0957-0233/24/10/105701]

14 **Jia W**, Li Y, Qu R, Baranowski T, Burke LE, Zhang H, Bai Y, Mancino JM, Xu G, Mao ZH, Sun M. Automatic food detection in egocentric images using artificial intelligence technology. *Public Health Nutr* 2019; **22**: 1168-1179 [PMID: 29576027 DOI: 10.1017/S1368980018000538]

15 **Raber M**, Baranowski T, Crawford K, Sharma SV, Schick V, Markham C, Jia W, Sun M, Steinman E, Chandra J. The Healthy Cooking Index: Nutrition Optimizing Home Food Preparation Practices across Multiple Data Collection Methods. *J Acad Nutr Diet* 2020; **120**: 1119-1132 [PMID: 32280056 DOI: 10.1016/j.jand.2020.01.008]

16 **Shinozaki N**, Murakami K. Evaluation of the Ability of Diet-Tracking Mobile Applications to Estimate Energy and Nutrient Intake in Japan. *Nutrients* 2020; **12** [PMID: 33138088 DOI: 10.3390/nu12113327]

17 **Oka R**, Nomura A, Yasugi A, Kometani M, Gondoh Y, Yoshimura K, Yoneda T. Study Protocol for the Effects of Artificial Intelligence (AI)-Supported Automated Nutritional Intervention on Glycemic Control in Patients with Type 2 Diabetes Mellitus. *Diabetes Ther* 2019; **10**: 1151-1161 [PMID: 30877556 DOI: 10.1007/s13300-019-0595-5]

Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: bpgoffice@wjgnet.com
Help Desk: https://www.f6publishing.com/helpdesk
https://www.wjgnet.com

ISSN 2644-3236 (online)

Artificial Intelligence in *Gastroenterology*

Artif Intell Gastroenterol 2022 December 28; 3(5): 117-162

Published by Baishideng Publishing Group Inc

AIG

Artificial Intelligence in Gastroenterology

Contents

Bimonthly Volume 3 Number 5 December 28, 2022

REVIEW

Artificial Intelligence in Gastroenterology

Bimonthly Volume 3 Number 5 December 28, 2022

Contents

ABOUT COVER

Editorial Board Member of *Artificial Intelligence in Gastroenterology*, Alina Ioana Tantau, FACE, PhD, Academic Research, Chief Doctor, Professor, Senior Researcher, Department of Internal Medicine and Gastroenterology, "Iuliu Hatieganu" University of Medicine and Pharmacy, 4th Medical Clinic, Cluj-Napoca 400012, Romania. alitantau@gmail.com

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

The *AIG* is now abstracted and indexed in Reference Citation Analysis, China Science and Technology Journal Database.

RESPONSIBLE EDITORS FOR THIS ISSUE

NAME OF JOURNAL
Artificial Intelligence in Gastroenterology

ISSN
ISSN 2644-3236 (online)

LAUNCH DATE
July 28, 2020

FREQUENCY
Bimonthly

EDITORS-IN-CHIEF
Rajvinder Singh, Ferruccio Bonino

EDITORIAL BOARD MEMBERS
https://www.wjgnet.com/2644-3236/editorialboard.htm

PUBLICATION DATE
December 28, 2022

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

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 December 28; 3(5): 117-141

DOI: 10.35712/aig.v3.i5.117

ISSN 2644-3236 (online)

REVIEW

Artificial intelligence in gastroenterology: A narrative review

Jonathan S Galati, Robert J Duve, Matthew O'Mara, Seth A Gross

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): C
Grade D (Fair): D
Grade E (Poor): 0

P-Reviewer: Liu XQ, China; Qi XS, China

Received: October 9, 2022
Peer-review started: October 9, 2022
First decision: October 29, 2022
Revised: November 21, 2022
Accepted: December 21, 2022
Article in press: December 21, 2022
Published online: December 28, 2022

Jonathan S Galati, Department of Medicine, NYU Langone Health, New York, NY 10016, United States

Robert J Duve, Department of Internal Medicine, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, NY 14203, United States

Matthew O'Mara, Seth A Gross, Division of Gastroenterology, NYU Langone Health, New York, NY 10016, United States

Corresponding author: Jonathan S Galati, MD, Department of Medicine, NYU Langone Health, 550 First Avenue, New York, NY 10016, United States. jonathan.galati@nyulangone.org

Abstract

Artificial intelligence (AI) is a complex concept, broadly defined in medicine as the development of computer systems to perform tasks that require human intelligence. It has the capacity to revolutionize medicine by increasing efficiency, expediting data and image analysis and identifying patterns, trends and associations in large datasets. Within gastroenterology, recent research efforts have focused on using AI in esophagogastroduodenoscopy, wireless capsule endoscopy (WCE) and colonoscopy to assist in diagnosis, disease monitoring, lesion detection and therapeutic intervention. The main objective of this narrative review is to provide a comprehensive overview of the research being performed within gastroenterology on AI in esophagogastroduodenoscopy, WCE and colonoscopy.

Key Words: Artificial intelligence; Colonoscopy; Computer-aided detection; Deep learning; Endoscopy; Machine learning

Core Tip: Artificial intelligence (AI) is a complex concept that has the capacity to revolutionize medicine. Within gastroenterology, recent research efforts have focused on using AI in esophagogastroduodenoscopy, wireless capsule endoscopy (WCE) and colonoscopy to assist in diagnosis, disease monitoring, lesion detection and therapeutic intervention. This narrative review provides a comprehensive overview of the research being performed within gastroenterology on AI in esophagogastroduodenoscopy, WCE and colonoscopy.

Citation: Galati JS, Duve RJ, O'Mara M, Gross SA. Artificial intelligence in gastroenterology: A narrative review. *Artif Intell Gastroenterol* 2022; 3(5): 117-141
URL: https://www.wjgnet.com/2644-3236/full/v3/i5/117.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i5.117

INTRODUCTION

Artificial intelligence (AI) is a complex concept, broadly defined in medicine as the development of computer systems to perform tasks that require human intelligence[1]. Since its inception in the 1950s, the field of AI has grown considerably (Figure 1)[2]. Often AI is accompanied by the terms machine learning (ML) and deep learning (DL), techniques used within the field of AI to develop systems that can learn and adapt without explicit instructions. Machine learning uses self-learning algorithms that derive knowledge from data to predict outcomes[1]. There are two main categories within ML: Supervised and unsupervised learning. In supervised learning, the AI is trained on a dataset in which human intervention has previously assigned a hierarchy of features which allows the algorithm to understand differences between data inputs and classify or predict outcomes[3]. In unsupervised learning, the system is provided a dataset that has not been categorized by human intervention. The algorithm then analyzes the data with the goal of identifying labels or patterns[3].

Deep learning is a subfield of ML that utilizes artificial neural networks (ANN) to analyze data. In DL, the system is able to analyze raw data and determine features that distinguish between data inputs. ANN systems are composed of interconnected nodes in a layered structure similar to how neurons are organized in the human brain. The weight of the connections between each node influences how the system can recognize, classify, and describe objects within data[3,4]. ANNs with multiple layers of nodes are classified as deep neural networks which form the backbone of deep learning.

Artificial intelligence has the capacity to revolutionize medicine. It can be used to increase efficiency by aiding in appointment scheduling, reviewing insurance eligibility, or tracking patient history. AI can also expedite data and image analysis and detect patterns, trends and associations[5]. Within gastroenterology, AI's prominence stems from its utility in image analysis[5,6]. Many gastrointestinal diseases rely on endoscopic evaluation for diagnosis, disease monitoring, lesion detection and therapeutic intervention. However, endoscopic evaluation is heavily operator dependent and thus subject to operator bias and human error. As such, recent efforts have focused on using AI in esophagogastroduodenoscopy, wireless capsule endoscopy (WCE) and colonoscopy to mitigate these issues, serving as an additional objective observer of the intestinal tract. The main objective of this narrative review is to provide a comprehensive overview of the research being performed within gastroenterology on artificial intelligence in esophagogastroduodenoscopy, WCE and colonoscopy. While other narrative reviews have been published regarding the use of artificial intelligence in esophagogastroduodenoscopy, WCE and colonoscopy, this narrative review goes a step further by providing a granular and more technical assessment of the literature. As such, this narrative review is intended for medical providers and researchers who are familiar with the use of artificial intelligence in esophagogastroduodenoscopy, WCE and colonoscopy and are interested in obtaining an in-depth review in a specific area.

LITERATURE REVIEW

Electronic databases Embase, Ovid Medicine, and PubMed were searched from inception to September 2022 using multiple search queries. Combinations of the terms "artificial intelligence", "AI", "computer aided", "computer aided detection", "CADe", "convolutional neural network", "deep learning", "DCNN", "machine learning", "colonoscopy", "endoscopy", "wireless capsule endoscopy", "capsule endoscopy", "WCE", "esophageal cancer", esophageal adenocarcinoma", "esophageal squamous cell carcinoma", "gastric cancer", "gastric neoplasia", "gastric lesions", "Barrett's esophagus", "celiac disease", "Helicobacter pylori", "Helicobacter pylori infection", "H pylori", "H pylori infection", "gastric ulcers", "duodenal ulcers", "inflammatory bowel disease", "IBD", "ulcerative colitis", "Crohn's disease", "parasitic infections", "hookworms", "bleeding", "gastrointestinal bleeding", "vascular lesions", "angioectasias", "polyp", "polyp detection", "tumor", "gastrointestinal tumor", "small bowel tumor", "bowel preparation", "Boston bowel preparation scale", "BBPS", "adenoma", "adenoma detection", "adenoma detection rate", "sessile serrated lesion", and "sessile serrated lesion rate" were used. We subsequently narrowed the results to clinical trials in human published within the last 10 years.

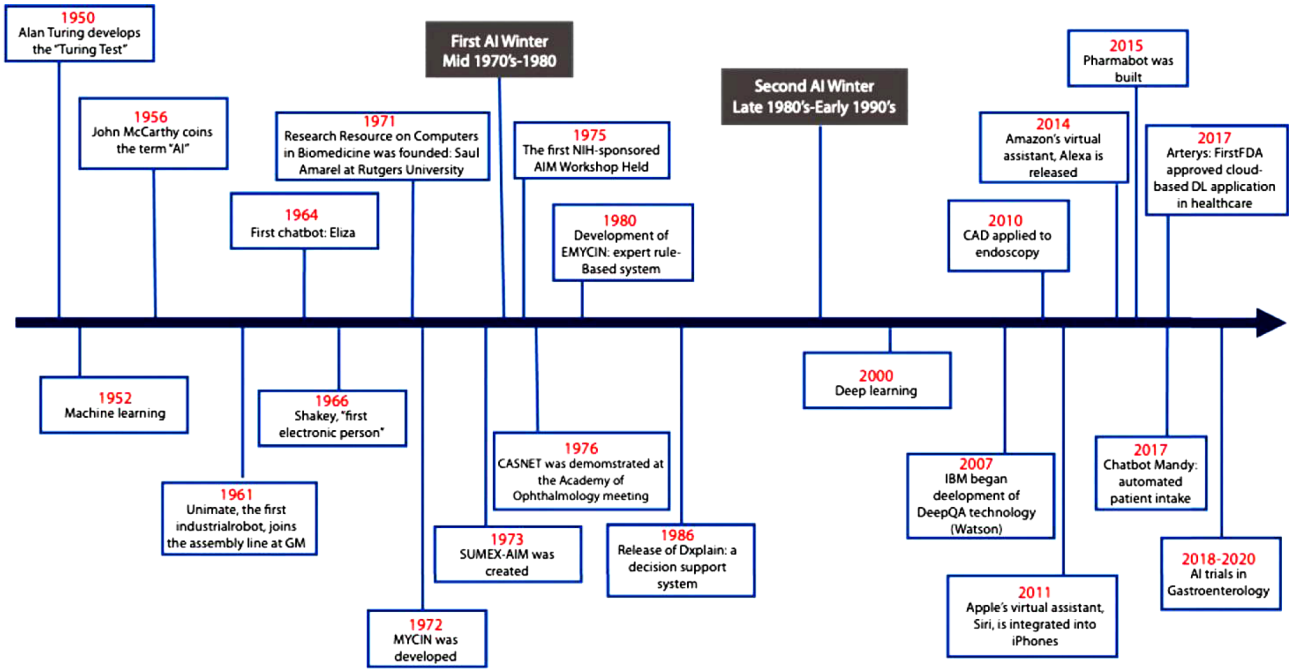


Figure 1 Timeline of the development and use of artificial intelligence in medicine. AI: Artificial intelligence; DL: Deep learning; FDA: U.S. Food and Drug Administration; CAD: Computer-aided diagnosis. Reprinted with permission from Elsevier Science & Technology Journals[2].

ESOPHAGOGASTRODUODENOSCOPY

Barrett's esophagus and esophageal adenocarcinoma

Barrett's esophagus (BE) is a premalignant condition associated with esophageal adenocarcinoma (EAC)[7-9]. It is caused by chronic inflammation and tissue injury of the lower esophagus as a result of gastric reflux[7-9]. Early detection and diagnosis can prevent the progression of BE to EAC[7-9]. Patients with BE should undergo routine surveillance endoscopies to monitor for progression. However, even with surveillance, dysplastic changes can be easily missed[7]. To improve the detection of dysplastic changes in BE, researchers have focused on developing AI systems to assist with the identification of dysplasia and early neoplasia during endoscopic evaluation.

Since 2016, a group of researchers from the Netherlands have developed numerous AI systems to identify neoplastic lesions in BE[10-16]. Their first publication detailed their experience using a support vector machine (SVM), a ML method, to identify early neoplastic lesions from white light endoscopy (WLE) images[10]. Their SVM achieved a sensitivity and specificity of 83% with respect to per-image detection and sensitivity of 86% and specificity of 87% with respect to per-patient detection[10]. In their next study, the group trialed several different feature extraction and ML methods using volumetric laser endomicroscopy (VLE) images[11]. They received the best results with the feature extraction module "layering and signal decay statistics", achieving high sensitivity (90%) and specificity (93%) with area under the curve (AUC) 0.95 for neoplastic lesion detection[11]. Following this, they conducted a second studying again using ML in VLE to identify neoplastic lesions in BE, however, they used a multiframe analysis approach, including frames neighboring the region of interest in the analysis[12]. With this approach, they found that multiframe analysis resulted in a significantly higher median AUC when compared to single frame analysis (0.91 *vs* 0.83; $P < 0.001$)[12]. Continuing to use ML methods, the group published their finding from the ARGOS project – a consortium of three international tertiary referral centers for Barrett's neoplasia[13]. In this study, de Groof *et al*[13] created a computer-aided detection (CADe) system that used SVM to classify images. The group tested the CADe with 60 images – 40 images from patients with a neoplastic lesion, 20 images from patients with non-dysplastic Barrett's esophagus. The CADe achieved an AUC of 0.92 and a sensitivity, specificity and accuracy of 95%, 85% and 92% respectively for detecting neoplastic lesions[13].

Following their successes creating ML systems for neoplastic lesion detection, the group of researchers from the Netherlands shifted their focus to DL methods. In their first foray into DL, they developed a hybrid CADe system using architecture from ResNet and U-Net models. The CADe was trained with 494364 labeled endoscopic images and subsequently refined with a data set comprised of 1247 WLE images. It was finally tested on a set of 297 images (129 images with early neoplasia, 168 with non-dysplastic BE) where the hybrid CADe system attained a sensitivity of 87.6%, specificity of 88.6% and accuracy of 88.2% for identifying early neoplasia[14]. The system was also tested in two external validation sets where it achieved similar results. A secondary outcome of the study was to see if within the images classified as having neoplasia if the CADe could delineate the neoplasia and recommend a

site for biopsy. The ground truth was determined by expert endoscopists. In two external data sets (external validation data set 4 and 5), the CADe identified the optimal biopsy site in 97.2% of cases and 91.9% of cases respectively[14]. Using a similar hybrid CADe, the group performed a pilot study testing the CADe during live endoscopic procedures[15]. Overall, the CADe achieved a sensitivity of 75.8%, specificity of 86.5% and accuracy of 84% in per-image analyses[15]. Their most recent study again used their hybrid ResNet and U-Net CADe to identify neoplastic lesions in narrow-band imaging (NBI)[16]. With respect to NBI images, the CADe was found to have sensitivity of 88% (95%CI 86%-94%), specificity of 78% (95%CI 72%-84%), and accuracy of 84% (95%CI 81%-88%) for identifying BE neoplasia [16]. In per frame and per video analyses, the CADe achieved sensitivities of 75% and 85%, specificities of 90% and 83% and accuracies of 85% and 83% respectively[16].

Outside of this group from the Netherlands, several other researchers have created DL systems for the detection of BE neoplasia[17-21]. Hong *et al*[17] created a CNN that could distinguish between intestinal metaplasia, gastric metaplasia and neoplasia from images obtained by endomicroscopy in patients with Barrett's esophagus with accuracy of 80.8%. Ebigbo *et al*[18] created a DL-CADe capable of detecting BE neoplasia with sensitivity 83.7%, specificity of 100.0% and accuracy of 89.9%. Two other groups achieved similar results to Ebigbo *et al*[18]: Hashimoto *et al*'s CNN detected early neoplasia with sensitivity of 96.4%, specificity of 94.2%, and accuracy of 95.4% and Hussein *et al*'s CNN detected early neoplasia with sensitivity 91%, specificity 79%, area under the receiver operating characteristic (AUROC) of 93%[19,20]. An overview of these studies is provided in Table 1.

In addition to neoplasia detection, some groups started to use AI to grade BE and predict submucosal invasion of lesions. Ali *et al*[22] recently published the results from a pilot study using a DL system to quantitatively assess BE area (BEA), circumference and maximal length (C&M). They tested their DL system on 3D printed phantom esophagus models with different BE patterns and 194 videos from 131 patients with BE. In the phantom esophagus models, the DL system achieved an accuracy of 98.4% for BEA and 97.2% for C&M[22]. In the patient videos, the DL system differed from expert endoscopists by 8% and 7% for C&M respectively[22]. Ebigbo *et al*[23], building upon their earlier success using a DL CADe to detect neoplasia, performed a pilot study using a 101-layer CNN to differentiate T1a (mucosal) and T1b (submucosal) BE related cancers. Using 230 WLE images obtained from three tertiary care centers in Germany, their CNN was capable of discerning T1a lesions from T1b lesions with sensitivity, specificity and accuracy of 77%, 64% and 71% respectively, comparable to the expert endoscopists enrolled in the study[23].

Despite BE's potential progression to EAC if left unmanaged, few studies have explicitly looked at using AI to detect EAC. Ghatwary *et al*[24] tested several DL models on 100 WLE images (50 featuring EAC, 50 featuring normal mucosa) to determine which was best at identifying EAC. They found that the Single-Shot Multibox Detector (SSD) method achieved the best results, attaining a sensitivity of 96% and specificity of 92%[24]. In 2021, Iwagami *et al*[25] focused on developing an AI system to identify esophagogastric junctional adenocarcinomas. They used SSD for their CNN, achieving a sensitivity, specificity and accuracy of 94%, 42% and 66% for detecting esophagogastric junctional adenocarcinomas. Their CNN performed similarly to endoscopists enrolled in the study (sensitivity 88%, specificity 43%, accuracy 66%)[25].

Esophageal squamous cell carcinoma

Esophageal squamous cell carcinoma (ESCC) is the most common histologic type of esophageal cancer in the world[26]. While certain imaging modalities such as Lugol's chromoendoscopy and confocal microendoscopy are effective at improving the accuracy, sensitivity and specificity of targeted biopsies, they are expensive and not universally available[27]. In recent years, efforts have focused on developing AI systems to support lower cost imaging modalities in order to improve their ability to detect ESCC.

Shin *et al*[27] and Quang *et al*[28] created ML algorithms which they tested on high-resolution microendoscope images, obtaining comparable sensitivities for the detection of ESCC (98% and 95% respectively). Following these studies, several groups created DL systems to detect ESCC[29-38]. In Cai *et al*'s study, their deep neural network-CADe was tested on 187 images obtained from WLE. The system obtained good sensitivity (97.8%), specificity (85.4%) and accuracy (91.4%) for identifying ESCC [29]. Similar findings occurred in three separate studies that used deep convolutional neural networks (DCNNs) to detect ESCC in WLE[30-32]. Using NBI, Guo *et al*[33] created a CADe that achieved high sensitivity (98.0%), specificity (95.0%) and an AUC of 0.99 for detecting ESCC in still images. Similar results were obtained in Li *et al*'s study[35]. For detecting ESCC in NBI video clips, Fukuda *et al*[34] obtained different results, finding similar sensitivity (91%) to Guo *et al*[33] however substantially lower specificity (51%). Three studies compared a DL-CADe with WLE to DL-CADe with NBI for the detection of ESCC[32,35,36]. The results from these three studies were quite discordant and as such a statement regarding whether a DL-CADe with WLE or DL-CADe with NBI is better for the detection of ESCC cannot be made at this time.

Interestingly, several studies used DL algorithms to assess ESCC invasion depth[39,40]. Everson *et al* [39] and Zhao *et al*[40] created CNNs to detect intrapapillary capillary loops, a feature of ESCC that correlates with invasion depth, in images obtained from magnification endoscopy with NBI. They achieved similar findings with Everson *et al*'s CNN achieving an accuracy of 93.7% and Zhao *et al*'s achieving an accuracy of 89.2%[39,40]. Using DL, two groups created DCNNs to directly detect ESCC

Table 1 Overview of findings from studies evaluating the detection accuracy of computer-aided detection for Barrett's esophagus-related neoplasia

Ref.	Country	Study design	AI Classifier	Lesions	Training dataset	Test dataset	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUROC
Swager *et al* [11], 2017	Netherlands	Retrospective	ML[2] methods	NPL	-	60 VLE images	90	93	-	0.95
van der Sommen *et al* [10], 2016	Netherlands	Retrospective	SVM	NPL	-	100 WLE images	83	83	-	-
Hong *et al* [17], 2017	South Korea	Retrospective	CNN	NPL, IM, GM	236 endomicroscopy images	26 endomicroscopy images	-	-	80.77	-
de Groof *et al* [13], 2019	Netherlands, Germany, Belgium	Prospective	SVM	NPL	-	60 WLE images	95	85	91.7	0.92
Ebigbo *et al* [21], 2019	Germany, Brazil	Retrospective	CNN	EAC	Augsburg dataset: 148 WLE images and NBI; MICCAI dataset: 100 WLE images		97; 94[a]; 92	88; 80[a]; 100	-	-
Ghatwary *et al*[24], 2019	England, Egypt	Retrospective	Multiple CNNs	EAC	Images from 21 patients	Images from 9 patients	96	92	-	-
de Groof *et al* [14], 2020	Netherlands, France, Sweden, Germany, Belgium, Australia	Ambispective	CNN	NPL	Dataset 1: 494364 images; Dataset 2:1; 247 images; Dataset 3: 297 images	Dataset 3: 297 images; Dataset 4: 80 images; Dataset 5: 80 images	90[b]	87.5[b]	88.8[b]	-
de Groof *et al* [15], 2020	Netherlands, Belgium	Prospective	CNN	NPL	495611 images	20 patients; 144 WLE images	75.8	86.5	84	-
Ebigbo *et al* [18], 2020	Germany, Brazil	Prospective	CNN	EAC	129 images	62 images	83.7	100	89.9	-
Hashimoto *et al*[19], 2020	United States	Retrospective	CNN	NPL	1374 images	458 images	96.4	94.2	95.4	-
Struyvenberg *et al*[12], 2020	Netherlands	Prospective	ML[2] methods	NPL	-	3060 VLE frames	-	-	-	0.91
Iwagami *et al* [25], 2021	Japan	Retrospective	CNN	EJC	3443 images	232 images	94	42	66	-
Struyvenberg *et al*[16], 2021	Netherlands, Sweden, Belgium	Retrospective	CNN	NPL	495611 images	157 NBI zoom videos; 30021 frames	85[1]; 75	83[1]; 90	83[1]; 85	-
Hussein *et al* [20], 2022	England, Spain, Belgium, Austria	Prospective	CNN	DPL	148936 frames	264 *iscan-1* images	91	79	-	0.93

[a]Sensitivity and specificity reported by white light endoscopy images from the Augsburg dataset, narrow band images from the Augsburg dataset, and from the MICCAI dataset respectively.

[b]Results found from convolutional neural network analyzing dataset 4.

[1]Sensitivity, specificity and accuracy obtained from per-video analysis and from per-frame analysis respectively.

[2]Multiple machine learning (ML) methods tested. Results from best performing ML method reported.

AI: Artificial intelligence; AUROC: Area under the receiver operating characteristic; CNN: Convolutional neural network; DPL: Dysplasia; EAC: Esophageal adenocarcinoma; EJC: Esophagogastric junctional adenocarcinoma; GM: Gastric metaplasia; IM: Intestinal metaplasia; ML: Machine learning; NBI: Narrow band images; NPL: Neoplasia; SVM: Support vector machine; VLE: Volumetric laser endomicroscopy; WLE: White light endoscopy.

invasion depth[41-43]. One group from Osaka International Cancer Institute conducted two studies using SSD to create their DCNNs[41,42]. The DCNNs were made to classify images as EP-SM1 or EP-SM2-3 as this distinction in ESCC bares clinical significance. The studies (Nakagawa *et al*[41] and Shimamoto *et al*[42]) attained similar accuracies and specificities, however had substantially different sensitivities (90.1% *vs* 50% and 71%)[41,42]. The third study, Tokai *et al*[43], used SSD as well for their DCNN and also programed the DCNN to classify images as EP-SM1 or EP-SM2-3. Their observed sensitivity, specificity and accuracy were lower than those found by Nakagawa *et al*[41] (84.1%, 73.3% and 80.9% respectively).

Gastric cancer

Gastric cancer is the third leading cause of cancer-related mortality in the world[44,45]. Early detection of precancerous lesions or early gastric cancer with endoscopy can prevent progression to advanced disease[46]. However, a substantial number of upper gastrointestinal cancers are missed placing patients at risk for interval development[45]. To mitigate this risk, AI systems are being develop to assist with lesion detection.

In 2013, Miyaki *et al*[47] used a bag-of-features framework with densely sampled scale-invariant feature transform descriptors to classify still images obtained from magnifying endoscopy with flexible spectral imaging color enhancement as having or not having gastric cancer. Their system, a rudimentary version of ML, obtained good sensitivity (84.8%), specificity (87.0%) and accuracy (85.9%) for identifying gastric cancer[47]. Using SVM, Kanesaka *et al*[48] found higher sensitivity (96.7%), specificity (95%) and accuracy (96.3%).

Following these successes, several groups began using CNNs for the identification of gastric cancer [44-46,49-59]. In 2018, Hirasawa *et al*[44] published one of the first papers to use a CNN (SSD) to detect gastric cancer. In a test set of 2296 images, the CNN had a sensitivity of 92.2% for identifying gastric cancer lesions[44]. In a larger study, Tang *et al*[49] created a DCNN to detect gastric cancer in a test set of 9417 images and 26 endoscopy videos. With respect to their test set, the DCNN performed well, achieving a sensitivity of 95.5% (95%CI 94.8%–96.1%), specificity of 81.7% (95%CI 80.7%–82.8%), accuracy of 87.8% (95%CI 87.1%–88.5%) and AUC 0.94[49]. The DCNN continued to perform well in external validation sets, achieving sensitivity of 85.9%-92.1%, specificity of 84.4%-90.3%, accuracy of 85.1%-91.2% and AUC 0.89-0.93[49]. Compared to expert endoscopists, the DCNN attained higher sensitivity, specificity and accuracy. In the video set, the DCNN achieved a sensitivity of 88.5% (95%CI 71.0%-96.0%)[49]. Several studies using DCNN to detect gastric cancer in endoscopy images obtained similar sensitivities, specificities and accuracies to Tang *et al*[49]. While one study reported a sensitivity of 58.4% for detecting gastric cancer, the sensitivity for the study's 67 endoscopists was 31.9%[55].

Recently, several groups from China and Japan have published studies using CNNs with magnified endoscopy with NBI (ME-NBI) in an effort to improve early gastric cancer detection[56-59]. Using a 22-layer CNN, Horiuchi *et al*[56] achieved a sensitivity, specificity and accuracy of 95.4%, 71.0% and 85.3% respectively for identifying early gastric cancer from a set of 258 ME-NBI images (151 gastric cancer, 107 gastritis). The same group published a similar study the following year however using ME-NBI videos instead of still images[57]. They obtained similar results: sensitivity of 87.4% (95%CI 78.8%-92.8%), specificity of 82.8% (95%CI 73.5%-89.3%) and accuracy of 85.1% (955 CI 79.0%-89.6%)[57]. Hu *et al*[58] and Ueyama *et al*[59] in their studies using CNN to identify gastric cancer in ME-NBI achieved similar sensitivities, specificities and accuracies as Horiuchi *et al*[56]. An overview of these studies is provided in Table 2.

Of increasing interest to researchers within this field is predicting invasion depth of gastric cancer using AI. Few studies have used CNNs to predict invasion depth[60-63]. Yoon *et al*[60] created a CNN to predict gastric cancer lesion depth from standard endoscopy images. The CNN achieved good sensitivity (79.2%) and specificity (77.8%) for differentiating T1a (mucosal) from T1b (submucosal) gastric cancers (AUC 0.851)[60]. Also using standard endoscopy images, Zhu *et al*[61] attained similar results. They trained their CNN to identify P0 (restricted to the mucosa or < 0.5 mm within the muscularis mucosae) *vs* P1 (≥ 0.5 mm deep into the muscularis mucosae) lesions. The CNN achieved a sensitivity of 76.6%, specificity of 95.6%, accuracy of 89.2% and AUROC 0.94 (95%CI 0.90-0.97). Cho *et al*[62] using DenseNet-161 as their CNN and Nagao *et al*[62] using ResNet50 as their CNN obtained comparable results to Zhu *et al*[61] for predicting gastric cancer invasion depth from endoscopy images.

Gastric ulcers

Within recent years, numerous studies have been published regarding the use of AI to assist with the detection and classification of gastric lesions. Few of these studies explicitly used AI systems to detect duodenal and gastric ulcers, however they report data pertaining to ulcer detection.

Using YOLOv5, a deep learning object detection model, Ku *et al*[64] created a CADe system capable of detecting multiple gastric lesions with good precision (98%) and sensitivity (89%). Also using YOLO for their DCNN, Yuan *et al*[53] achieved an overall system accuracy of 85.7% for gastric lesion identification. With respect to peptic ulcer detection, their system achieved an accuracy of 95.4% (93.5%-97.2%), sensitivity of 86.2% (77.5%–94.8%) and specificity of 96.8% (95.1%–98.4%)[53]. Guo *et al*[54] used ResNet50 to construct their CADe designed to detect gastric lesions. Their CADe achieved lower sensitivity 71.4% (95%CI 69.5–73.2%) and specificity 70.9% (95%CI 70.3–71.4%) than Yuan *et al*'s DCNN [53], however Guo *et al*[54] combined erosions and ulcers into one category for analysis. With their primary outcome being classifying gastric cancers and ulcers, Namikawa *et al*[52] developed a CNN capable of identifying gastric ulcers with high sensitivity (93.3%; 95%CI 87.3%−97.1%) and specificity (99.0%; 95%CI 94.6%-100%).

Helicobacter pylori infection

As a risk factor for future development of gastric cancer, early detection and eradication of *Helicobacter pylori* (*H. pylori*) in infected individuals is important. Endoscopic evaluation for *H. pylori* is highly

Table 2 Overview of findings from studies evaluating the detection accuracy of computer-aided detection for gastric cancer

Ref.	Country	Study design	AI classifier	Lesions	Training dataset	Test dataset	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUROC
Miyaki *et al* [47], 2013	Japan	Prospective[a]	SVM	Gastric cancer	493 FICE-derived magnifying endoscopic images	92 FICE-derived magnifying endoscopic images	84.8	97	85.9	-
Kanesaka *et al*[48], 2018	Japan, Taiwan	Retrospective	SVM	EGC	126 M-NBI images	81 M-NBI images	96.7	95	96.3	-
Wu *et al* [50], 2019	China	Retrospective	CNN	EGC	9151 images	200 images	94	91	92.5	-
Cho *et al* [51], 2019	South Korea	Ambispective	CNN	Advanced gastric cancer, EGC, high grade dysplasia, low grade dysplasia, non-neoplasm	4205 WLE images	812 WLE images; 200 WLE images	-	-	86.6[b]; 76.4	0.877[b]
Tang *et al* [49], 2020	China	Retrospective	CNN	EGC	35823 WLE images	Internal: 9417 WLE images; External: 1514 WLE images[1]	95.5[1]; 85.9-92.1	81.7[1]; 84.4-90.3	87.8[1]; 85.1-91.2	0.94[1]; 0.887-0.925
Namikawa *et al*[52], 2020	Japan	Retrospective	CNN	Gastric cancer	18410 images	1459 images	99	93.3	99	-
Horiuchi *et al*[56], 2020	Japan	Retrospective	CNN	EGC	2570 M-NBI images	258 M-NBI images	95.4	71	85.3	0.852
Horiuchi *et al*[57], 2020	Japan	Retrospective	CNN	EGC	2570 M-NBI images	174 videos	87.4	82.8	85.1	0.8684
Guo *et al* [54], 2021	China	Retrospective	CNN	Gastric cancer, erosions/ulcers, polyps, varices	293162 WLE images	33959 WLE images	67.5[2]; 85.1	70.9[2]; 90.3	-	-
Ikenoyama *et al*[55], 2021	Japan	Retrospective	CNN	EGC	13584 WLE and NBI images	2940 WLE and NBI images	58.4	87.3	-	-
Hu *et al* [58], 2021	China	Retrospective	CNN	EGC	M-NBI images from 170 patients	Internal: M-NBI from 73 patients External: M-NBI images from 52 patients	79.2[3]; 78.2	74.5[3]; 74.1	77[3]; 76.3	0.808[3]; 0.813
Ueyama *et al*[59], 2021	Japan	Retrospective	CNN	EGC	5574 M-NBI images	2300 M-NBI	98	100	98.7	-
Yuan *et al* [53], 2022	China	Retrospective	CNN	EGC, advanced gastric cancer, submucosal tumor, polyp, peptic ulcer, erosion, and lesion-free gastric mucosa	29809 WLE images	1579 WLE images	59.2[4]; 100	99.3[4]; 98.1	93.5[4]; 98.4	-

[a]Presumed prospective based on manuscript.

[b]Accuracy of convolutional neural network (CNN) for detecting the five different lesions and detecting gastric cancer respectively. Area under the receiver operating characteristic (AUROC) pertains to detecting gastric cancer.

[1]The external dataset was comprised of images from 3 external sites. Sensitivity, specificity, accuracy and AUROC for the internal dataset and external dataset respectively.

[2]Sensitivity and specificity of CNN for detecting gastric cancers in a dataset comprised of images without annotations and for detecting gastric cancers in a dataset comprised of annotated images respectively.

[3]Sensitivity, specificity, accuracy and AUROC for the internal dataset and external dataset respectively.

[4]Sensitivity, specificity and accuracy for detecting early gastric cancer and for detecting advanced gastric cancer respectively.

"Internal" and "External" refer to internal and external datasets respectively.

CNN: Convolutional neural network; AUROC: Area under the receiver operating characteristic; EGC: Early gastric cancer; FICE: Flexible spectral imaging

color enhancement; M-NBI: Magnifying endoscopy with narrow band imaging; NBI: Narrow band imaging; SVM: Support vector machine; WLE: White light endoscopy.

operator dependent[65]. Pairing artificial intelligence with endoscopy for the detection of *H. pylori* could possibly reduce false results.

Shichijo *et al*[66] used GoogLeNet, a DCNN consisting of 22 layers, to evaluate 11481 images obtained from 397 patients (72 *H. pylori* positive, 325 negative) for the presence or absence of *H. pylori* infection. GoogLeNet attained a sensitivity of 81.9% (95%CI 71.1%-90.0%), specificity of 83.4% (95%CI 78.9%-87.3%) and accuracy of 83.1% (95%CI 79.1%-86.7%) with AUROC 0.89 for detecting *H. Pylori* infection [66]. When compared to endoscopists enrolled in the study, the sensitivity, specificity and accuracy attained by GoogLeNet was comparable to those attained by the endoscopists[66]. This same group published a second study in 2019 again using GoogLeNet for their DCNN[67]. However, a different optimization technique was used to prepare GoogLeNet. The DCNN was tasked with classifying images as *H. pylori* positive, negative or eradicated. In a set of 23699 images, the DCNN attained an accuracy of 80% for *H. pylori* negative, 84% for *H. pylori* eradicated, and 48% for *H. pylori* positive[67]. Also using GoogLeNet, Itoh *et al*[68] obtained similar results to Shichijo *et al*'s 2017 study with respect to sensitivity (86.7%) and specificity (86.7%)[66]. Using ResNet-50 as their architectural unit for their DCNN, Zheng *et al*[69] were successful in classifying images as *H. pylori* positive or negative, achieving a sensitivity, specificity, accuracy and AUC of 81.4% (95%CI 79.8%–82.9%), 90.1% (95%CI 88.4%–91.7%), 84.5% (95%CI 83.3%–85.7%) and 0.93 (95%CI 0.92–0.94) respectively.

Taking a different approach, Yasuda *et al*[70] used linked color imaging (LCI) with SVM to identify *H. pylori* infection. The LCI images were classified into high-hue and low-hue images based on redness and classified by SVM as *H. pylori* positive or negative. This method attained a sensitivity, specificity and accuracy of 90.4%, 85.7% and 87.6% respectively[70]. Combining LCI with a deep learning CADe system, Nakashima *et al*[71] achieved a sensitivity, specificity and accuracy of 92.5%, 80.0%, 84.2% for identifying *H. pylori* negative images, 62.5%, 92.5%, 82.5% for *H pylori* positive images, 65%, 86.2%, 79.2% for *H. pylori* post-eradication images respectively.

Celiac disease

While immunological tests can support the diagnosis of celiac disease, definitive diagnosis requires histological assessment of duodenal biopsies[72]. As such being able to identify changes in the duodenal mucosa consistent with celiac disease is important. However, these changes can be subtle and difficult to appreciate. Few studies have been published using a CADe system to detect or diagnose celiac disease.

In 2016, Gadermayr *et al*[73] created a system that combined expert knowledge acquisition with feature extraction to classify duodenal images obtained from 290 children as Marsh-0 (normal mucosa) or Marsh-3 (villous atrophy). Expert knowledge acquisition was achieved by having one of three study endoscopists assign a Marsh grade of 0 or 3 to an image. Feature extraction was accomplished using one of three methods: (1) multi-resolution local binary patterns; (2) multi-fractal spectrum; and (3) improved Fisher vectors. From expert knowledge acquisition and feature extraction, their classification algorithm identified images as Marsh-0 or Marsh-3. With optimal settings, the classification algorithm achieved an accuracy of 95.6%-99.6%[73]. In 2016, Wimmer *et al*[74] used CNN to detect celiac disease in a set of 1661 images (986 images of normal mucosa, 675 images of celiac disease) with varying convolutional blocks. Their CNN achieved the best overall classification rate (90.3%) with 4 convolutional blocks[74]. Taking their CNN a step further, they combined the CNN with 4 convolutional blocks with SVM which increased overall classification rate by 6.7%[74]. While interesting, Gadermayr *et al*'s method requires human intervention and the paper's methodology is quite complicated[73], largely in part to the extensive number of systems tested. Wimmer *et al*[74] provided a simpler method that attained a good overall classification rate.

WIRELESS CAPSULE ENDOSCOPY

Celiac disease

Few studies have assessed the utility of AI in the detection of celiac disease using WCE. In 2017, Zhou *et al*[75] trained GoogLeNet, a DCNN, to identify celiac disease using clips obtained during WCE. Their DCCN achieved a sensitivity and specificity of 100% for identifying patients with celiac disease from 10 WCE videos (5 from patients with celiac disease, 5 from healthy controls)[75]. Similarly, Wang *et al*[76] used DL to diagnose celiac disease from WCE videos, however their CNN utilized a block-wise channel squeeze and excitations attenuation module, a newer architectural unit thought to better mimic human visual perception[76]. Their system attained an accuracy of 95.9%, sensitivity of 97.2% and specificity of 95.6% for diagnosing celiac disease.

Inflammatory bowel disease

WCE is often used in patients with inflammatory bowel disease (IBD) to detect small bowel ulcers and erosions. While computed tomography enterography and MRI have been used to detect areas of disease activity and inflammation along the gastrointestinal tract in patients with IBD, these imaging modalities can miss early or small lesions. While WCE can directly visualize lesions, endoscopists reviewing the video may miss lesions or mistakenly identify imaging artifacts as lesions. AI systems could help reduce these errors. Several studies have been published using AI in WCE to detect intestinal changes consistent with Crohn's disease[77-83].

To discriminate ulcers from normal mucosa in Crohn's disease, Charisis *et al*[78] proposed combining bidimensional ensemble empirical mode decomposition and differential lacunarity to pre-process images followed by classification using several ML algorithms and a multilayer neural network. Using a dataset consisting of 87 ulcer and 87 normal mucosa images, their CADe achieved accuracy 89.0%-95.4%, sensitivity 88.2%-98.8%, and specificity 84.2%-96.6%[78]. Subsequently, Charisis and Hadjileontiadis published a paper in 2016 combining hybrid adaptive filtering and differential lacunarity (HAF-DLac) to process images followed by SVM to detect Crohn's disease related lesions in WCE[79]. In a set of 800 WCE images, the HAF-DLac system achieved a sensitivity, specificity and accuracy of 95.2%, 92.4% and 93.8% respectively for detecting lesions[79]. Using a similar approach to Charisis *et al*[78], Kumar *et al*[80] used MPEG-7 edge, color and texture features to pre-process images followed by image classification using SVM to detect and classify lesions in patients with Crohn's disease. Their system, tested against 533 images (212 normal mucosa, 321 images with lesions), obtained an accuracy of 93.0%-93.8% for detecting lesions and an accuracy of 78.5% for classifying them based on severity.

With respect to deep learning, few groups have used deep learning algorithms in WCE to identify Crohn's disease related lesions. Recently, Ferreira *et al*[82] used a DCNN to identify erosions and ulcers in patients with Crohn's disease. Their DCNN achieved a sensitivity of 98.0%, specificity of 99.0%, accuracy of 98.8% and AUROC of 1.00. Interestingly, Klang *et al*[83] developed a DCNN to detect intestinal strictures. Overall, their DCNN achieved an accuracy of 93.5% ± 6.7% and AUC of 0.989 for detecting strictures.

Hookworm infections

Three studies have used artificial intelligence to detect hookworms using WCE. The first to publish on this topic was Wu *et al*[84] in 2016. Using SVM, they were able to create a system that achieved a specificity of 99.0% and accuracy of 98.4% for detecting hookworms in WCE[84]. However, the system's sensitivity was 11.1%. He *et al*[85] created a DCNN using a novel deep hookworm detection framework that modeled the tubular appearance of hookworms. Their DCNN had an accuracy of 88.5% for identifying hookworm[85]. Gan *et al*[86] performed a similar study, finding an AUC of 0.97 (95%CI 0.967-0.978), sensitivity of 92.2%, specificity of 91.1% and accuracy of 91.2% The concordant findings of these three studies suggest a possible utility of using AI to diagnose hookworm infections.

Intestinal bleeding

One of the most common reasons to perform WCE is to evaluate for gastrointestinal bleeding after prior endoscopic attempts have failed to localize a source. Since the implementation of WCE in clinical practice, many methods, notably AI, have been employed to improve the detection of gastrointestinal sources of bleeding.

Several studies have looked at using supervised learning to identify bleeding in WCE. In 2014, Sainju *et al*[87] used an ML algorithm to interpret color quantization images and determine if bleeding was present. One of their models achieved a sensitivity, specificity and accuracy of 96%, 90% and 93%, respectively[87]. Using SVM, Usman *et al*[88] achieved similar results - sensitivity, specificity and accuracy of 94%, 91% and 92% respectively.

More recently, several groups have created DCNNs to identify bleeding and sources of bleeding in WCE. In 2021, Ghosh *et al*[89] used a system comprised of two CNN systems (CNN-1, CNN-2) to classify WCE images as bleeding or non-bleeding and subsequently to identify sources of bleeding within the bleeding images. For classifying images as bleeding or non-bleeding, CNN-1 had a sensitivity, specificity, accuracy and AUC of 97.5%, 99.9%, 99.4% and 0.99[89]. For identifying sources of bleeding within the bleeding images, CNN-2 had an accuracy of 94.4% and intersection over union (IoU) of 90.7%[89].

In 2020, Tsuboi *et al*[90] published the first study to use DCNN to detect small bowel angioectasias from WCE images. In their test set which included 488 images of small bowel angioectasias and 10000 images of normal small bowel mucosa, their DCNN achieved an AUC of 0.99 with sensitivity and specificity of 98.8% and 98.4%[90]. Similarly, in 2021 Ribeiro *at al*[91] developed a DCNN to identify vascular lesions, categorizing them by bleeding risk according to Saurin's classification: P0 – no hemorrhagic potential, P1 – uncertain/intermediate hemorrhagic potential and red spots, and P2 – high hemorrhagic potential (angioectasias, varices). In their validation set, the DCNN had a sensitivity, specificity, accuracy and AUROC of 91.7%, 95.3%, 94.1% and 0.97 respectively for identifying P1 lesions [91]. Regarding P2 lesions, the network had a sensitivity, specificity, accuracy and AUROC of 94.1%, 95.1%, 94.8% and 0.98 respectively[91]. This group published a similar study in 2022 however now using

their DCNN to detect and differentiate mucosal erosions and ulcers based on bleeding potential[92]. Saurin's classification was again used to classify lesions, additionally labeling P1 lesions as mucosal erosions or small ulcers and P2 lesions as large ulcers (> 2 cm)[92]. The DCNN achieved an overall sensitivity of 90.8% ± 4.7%, specificity of 97.1% ± 1.7%, and accuracy of 93.4% ± 3.3% in their test set of 1226 images[92]. For the detection of mucosal erosions (P1), their DCNN achieved a sensitivity of 87.2%, specificity of 95.0% and accuracy of 93.3% with AUROC of 0.98 (95%CI 0.97-0.99)[92]. With respect to small ulcers (P1), their DCNN achieved a sensitivity of 86.4%, specificity of 96.9% and accuracy of 94.5% with AUROC of 0.99 (95%CI 0.97-1.00)[92]. Finally, with respect to large ulcers (P2), their DCNN achieved a sensitivity of 95.3%, specificity of 99.2% and AUROC of 1.00 (95%CI 0.98-1.00)[92]. A third study published by this group aimed to develop a DCNN to identify colonic lesions and luminal blood/hematic vestiges had similar findings. In their training set of 1801 images, the DCNN achieved an overall sensitivity, specificity and accuracy of 96.3%, 98.2%, and 97.6% respectively[93]. For detecting mucosal lesions, the DCNN achieved a sensitivity of 92.0%, specificity of 98.5% and AUROC of 0.99 (95%CI 0.98-1.00)[93]. For luminal blood/hematic vestiges, the DCNN achieved a sensitivity of 99.5%, specificity of 99.8% and AUROC of 1.00 (95%CI 0.99-1.00)[93].

Polyp and tumor detection

Gastrointestinal tumors can be difficult to discern from normal mucosa and thus pose a higher degree of diagnostic difficulty compared to other lesions on traditional WCE[94]. As such, developing an AI system to aid with the detection of these easy to miss lesions could be beneficial.

Several groups have developed ML systems to aid with detection. Using SVM, Li *et al*[95] were able to develop a system capable of detecting small bowel tumors with sensitivity, specificity and accuracy of 88.6%, 96.2% and 92.4%. Similarly, Liu *et al*[96] and Faghih Dinevari *et al*[97] used SVM to identify tumors in WCE, however they used different image pre-processing algorithms. Liu *et al*[96] used discrete curvelet transform to pre-process images prior to being classified by SVM. Their ML system achieved a sensitivity of 97.8% ± 0.5, specificity of 96.7% ± 0.4 and accuracy of 97.3% ± 0.5 for identifying small bowel tumors[96]. Faghih Dinevari *et al*[97] relied on discrete wavelet transform and singular value decomposition for image pre-processing prior to classification by SVM. Their system achieved a sensitivity of 94.0%, specificity of 93.0% and accuracy of 93.5% for identifying small bowel tumors[97]. Sundaram and Santhiyakumari built upon these methodologies, using a region of interest-based color histogram to enhance WCE images prior to being classified by two SVM algorithms: SVM1 and SVM2 [98]. SVM1 classified the WCE image as normal or abnormal. If SVM1 classified the image as abnormal, it was further classified by SVM2 as benign, malignant or normal[98]. The system attained an overall sensitivity of 96.0%, specificity of 95.4% and accuracy of 95.7% for small bowel tumor detection and classification[98].

With respect to DL methods, Blanes-Vidal *et al*[99] created a DCNN to autonomously detect and localize colorectal polyps. Their study included 255 patients who underwent WCE and standard colonoscopy for positive fecal immunochemical tests. Of the 255 patients, 131 had at least 1 polyp. The DCNN obtained a sensitivity of 97.1%, specificity of 93.3% and accuracy of 96.4% for detecting polyps in WCE[99]. Saraiva *et al*[100] and Mascarenhas *et al*[101] similarly used DCNNs to detect colonic polyps in WCE and obtained similar results to Blanes-Vidal *et al*[99-101]. Using an ANN, Constantinescu *et al*[102] created a DL system able to detect small bowel polyps with sensitivity of 93.6% and specificity of 91.4%. For gastric polyps and tumors, Xia *et al*[103] created a novel CNN – a region-based convolutional neural network (RCNN) – to evaluate magnetically controlled capsule endoscopy (MCE) images. Tested on 201365 MCE images obtained from 100 patients, the RCNN detected gastric polyps with sensitivity of 96.5%, specificity of 94.8%, accuracy of 94.9% and AUC of 0.898 (95%CI 0.84-0.96)[103]. For submucosal tumors, the RCNN achieved a sensitivity of 87.2%, specificity of 95.3%, accuracy of 95.2% and AUC of 0.88 (95%CI 0.81-0.96)[103]. Taking a different approach, Yuan and Meng used a novel deep learning method – stacked sparse autoencoder image manifold constraint – to identify intestinal polyps on WCE, finding an accuracy of 98.00% for poly detection[104]. However, sensitivity, specificity and AUC analyses were not reported.

COLONOSCOPY

Bowel preparation assessment

Inadequate bowel preparation, present in 15% to 35% of colonoscopies, is associated with lower rates of cecal intubation, lower adenoma detection rate (ADR), and higher rates of procedure-related adverse events[105,106]. For patients with inadequate bowel preparation, the United States Multi-Society Task Force of Colorectal Cancer (MSTF) which represents the American College of Gastroenterology, the American Gastroenterological Association and the American Society for Gastrointestinal Endoscopy (ASGE), and the European Society of Gastrointestinal Endoscopy recommend repeating a colonoscopy within 1 year[105,107-109]. In addition, the MSTF and ASGE recommend that endoscopists document bowel preparation quality at time of colonoscopy[108,109].

Despite these recommendations and variety of bowel preparation rating scales available, documentation of bowel preparation quality remains variable with studies reporting appropriate documentation in 20% to 88% of colonoscopies[110-112]. Few studies have been published regarding the use of DCNN to assist in the objective assessment of bowel preparation. The first group to do so, Zhou *et al*[113] in 2019, found that their DCNN (ENDOANGEL) was more accurate (93.3%) at grading the bowel preparation quality of still images than novice (< 1 year of experience performing colonoscopies; 75.91%), senior (1-3 years of experience performing colonoscopies; 74.36%) and expert (> 3 years of experience performing colonoscopies; 55.11%) endoscopists. When tested on colonoscopy videos, ENDOANGEL remained accurate at grading bowel preparation quality (89.04%)[113].

Building upon their experience with ENDOANGEL, Zhou *et al*[114] created a new system using two DCNNs: DCNN1 filtered unqualified frames while DCNN2 classified images by Boston Bowel Preparation Scale (BBPS) scores. The BBPS is a validated rating scale for assessing bowel preparation quality[115]. Colonic segments are assigned scores on a scale from 0 to 3. Colonic segments unable to be evaluated due to the presence of solid, unremovable stool are assigned a score of 0 whereas colonic segments that are able to be easily evaluated and contain minimal to no stool are assigned a score of 3 [115]. Zhou *et al*'s DCNN2 classified images into two categories: well-prepared (BBPS score 2-3) and poorly prepared (BPPS score 0-1)[114]. There was no difference between the dual DCNN system and study endoscopists when calculating the unqualified image portion (28.35% *vs* 29.58%, $P = 0.285$) and e-BBPS scores (7.81% *vs* 8.74%, $P = 0.088$). In addition, a strong inverse relationship between e-BBPS and ADR ($\rho = -0.976$, $P = 0.022$) was found.

Two other groups developed similar dual DCNN systems as Zhou *et al*[114] to calculate BBPS and obtained concordant findings[116,117]. Lee *et al*[116] tested their system on colonoscopy videos and found the system had an accuracy of 85.3% and AUC of 0.918 for detecting adequate bowel preparation. Using still images, Low *et al*'s system was able to accurately determine bowel preparation adequacy (98%) and subclassify by BBPS (91%)[117].

Using a different approach, Wang *et al*[118] used U-Net to create a DCNN to perform automatic segmentation of fecal matter from still images. Compared to images segmented by endoscopists, U-Net achieved an accuracy of 94.7%.

Inflammatory bowel disease

Colonoscopy is essential for the assessment of IBD as it allows for real-time evaluation of colonic inflammation[119,120]. Despite there being endoscopic scoring systems available to quantify disease activity, assessment is operator-dependent resulting in high interobserver variability[119-121]. Recent efforts have focused on using artificial intelligence to objectively grade colonic inflammation[121,122].

Several studies have investigated using DCNNs to classify images obtained from patients with ulcerative colitis (UC) by endoscopic inflammation scoring systems. The most commonly used endoscopic scoring system in these studies is the Mayo Endoscopic Score (MES). Physicians assign scores on a scale from 0 to 3 based on the absence or presence of erythema, friability, erosions, ulceration and bleeding[123]. A score of 0 indicates normal or inactive mucosa whereas a score of 3 indicates severe disease activity[123]. In 2018, Ozawa *et al*[121] published the first study to use a DCNN to classify still images obtained from patients with UC into MES 0 *vs* MES 1-3 and MES 0-1 *vs* MES 2-3. Their DCNN had an AUROC of 0.86 (95%CI 0.84-0.87) and AUROC 0.98 (95%CI 0.97-0.98) when differentiating MES 0 *vs* MES 1-3 and MES 0-1 *vs* MES 2-3 respectively[121]. Stidham *et al*[122] performed a similar study and found an AUROC of 0.966 (95%CI 0.967-0.972) for differentiating still images into MES 0-1 *vs* MES 2-3. Using a combined deep learning and machine learning system, Huang *et al*[124] were able to achieve an AUC of 0.938 with accuracy of 94.5% for identifying MES 0-1 *vs* MES 2-3 from still images. While the binary classification used in the aforementioned studies can differentiate remission/mucosal healing (MES 0-1) and active inflammation (MES 2-3), knowing exact MESs also has clinical significance[125,126]. Bhambhvani and Zamora created a DCNN to assign individual MESs to still images. The model achieved an AUC of 0.89, 0.86 and 0.96 for classifying images into MES 1, MES 2 and MES 3 respectively and achieved an average specificity of 85.7%, average sensitivity of 72.4% and overall accuracy of 77.2%[127].

In order to simulate how MES is performed in practice, several groups developed systems using DL to predict MES from colonoscopy videos. Yao *et al*'s DCNN had good agreement with MES scoring performed by gastroenterologists in their internal video test set (k = 0. 84; 95%CI 0.75-0.92), however their DCCN did not perform as well in the external video test set (k = 0.59; 95%CI 0.46-0.71)[128]. Gottlieb *et al*[129] reported similar findings to Yao *et al*[128], finding that their DCNN had good agreement with MES scoring performed by gastroenterologists (quadratic weighted kappa of 0.844; 95%CI 0.787–0.901). Gutierrez Becker *et al*[130] created a DL system designed to perform multiple binary tasks: discriminating MES < 1 *vs* MES ≥ 1, MES < 2 *vs* MES ≥ 2, and MES < 3 *vs* MES ≥ 3. For these tasks, their DL system attained an AUROC of 0.84, 0.85, and 0.85 respectively.

A group from Japan published several studies using AI on endoscopic images to predict histologic activity in patients with UC[131-134]. Their first study in 2016 used machine learning to predict persistent histologic inflammation[131]. Their system attained a sensitivity of 74% (95%CI 65%-81%), specificity of 97% (95%CI 95%-99%) and accuracy of 91% (95%CI 83%-95%) for predicting persistent histologic inflammation in still images[131]. Their following studies used a deep neural network labeled

DNUC (deep neural network for evaluation of UC) to identify endoscopic remission and histologic remission[132,134]. In still images, DNUC had a sensitivity of 93.3% (95%CI 92.2%–94.3%), specificity of 87.8% (95%CI 87.0%–88.4%) and diagnostic accuracy of 90.1% (95%CI 89.2%–90.9%) for determining endoscopic remission[132]. With respect to histologic remission, DNUC had a sensitivity of 92.4% (95%CI 91.5%–93.2%), specificity of 93.5% (95%CI 92.6%–94.3%) and diagnostic accuracy of 90.1% (92.9%; 95%CI 92.1%–93.7%)[132]. In colonoscopy videos, DNUC showed a sensitivity of 81.5% (95%CI 78.5%–83.9%) and specificity of 94.7% (95%CI 92.5%–96.4%) for endoscopic remission[134]. For histologic remission, DNUC had a sensitivity of 97.9% (95%CI 97.0%–98.5%) and specificity of 94.6% (95%CI 91.1%–96.9%) in colonoscopy videos[134].

To date, only one study has been published using an AI system to distinguish normal from inflamed colonic mucosa in Crohn's disease[135]. The group paired a DCNN with a long short-term memory (LSTM), a type of neural network that uses previous findings to interpret its current input, and confocal laser endomicroscopy. Their DCNN-LSTM system attained an accuracy of 95.3% and AUC of 0.98 for differentiating normal from inflamed mucosa[135].

Polyp detection

Colorectal cancer is the third most common malignancy and second leading cause of cancer-related mortality in the world[136]. While colonoscopy is the gold standard for detection and treatment of premalignant and malignant lesions, a substantial number of adenomas are missed[137,138]. As such, efforts have focused on using AI to improve ADR and decrease adenoma miss rate (AMR).

At present, numerous pilot, validation and prospective studies[139-161], randomized controlled studies[162-174], and systematic reviews and meta-analyses[175-183] have been published regarding the use of AI for the detection of colonic polyps. Furthermore, there are commercially available AI systems for both polyp detection and interpretation. With respect to the systematic reviews and meta-analyses published on this topic, AI-assisted colonoscopy has consistently been shown to have higher ADR, polyp detection rate (PDR) and adenoma per colonoscopy (APC) compared to standard colonoscopy [175-183]. Recently, several large, randomized controlled trials have been published supporting these findings. Shaukat *et al*[162] published their findings from their multicenter, randomized controlled trial comparing CADe colonoscopy to standard colonoscopy. Their study included 1359 patients: 677 randomized to standard colonoscopy, 682 to CADe colonoscopy. They found an increase in ADR (47.8% *vs* 43.9%; $P = 0.065$) and APC (1.05 *vs* 0.83; $P = 0.002$) in the CADe colonoscopy group. However, they also found a decrease in the overall sessile serrated lesions per colonoscopy rate (0.20 *vs* 0.28; $P = 0.042$) and sessile serrated lesion detection rate (12.6% *vs* 16.0%; $P = 0.092$) in the CADe colonoscopy group [162]. Brown *et al*[163] in their CADeT-CS Trial which was a multicenter, single-blind randomized tandem colonoscopy study comparing CADe colonoscopy to high-definition white light colonoscopy found similar increases in ADR (50.44% *vs* 43.64%; $P = 0.3091$) and APC (1.19 *vs* 0.90; $P = 0.0323$) in their patients who underwent CADe colonoscopy first[163]. Additionally, polyp miss rate (PMR) (20.70% *vs* 33.71%; $P = 0.0007$), AMR (20.12% *vs* 31.25%; $P = 0.0247$), and sessile serrated lesion miss rate (7.14% *vs* 42.11%; $P = 0.0482$) were lower in the CADe colonoscopy first group. In a similarly designed study to Brown *et al*[163], Kamba *et al*'s multicenter, randomized tandem colonoscopy study comparing CADe colonoscopy to standard colonoscopy found lower AMR (13.8% *vs* 26.7%; $P < 0.0001$), PMR (14.2% *vs* 40.6%; $P < 0.0001$), and sessile serrated lesion miss rate (13.0% *vs* 38.5%' $P = 0.03$) and higher ADR (64.5% *vs* 53.6%; $P = 0.036$) and PDR (69.8% *vs* 60.9%; $P = 0.084$) in patients who underwent CADe colonoscopy first[164]. Similar to Shaukat *et al*[162], the sessile serrated lesion detection rate was lower in the CADe colonoscopy first group compared to standard colonoscopy first (7.6% *vs* 8.1%; $P = 0.866$) [164]. Similar increases in ADR, APC and PDR were appreciated in randomize controlled trials by Xu *et al*[172], Liu *et al*[173], Repici *et al*[170], Gong *et al*[166], Wang *et al*[167], and Su *et al*[169] as well[166-172].

The majority of AI-assisted colonoscopy studies focus on adenoma detection. While these studies report sessile serrated lesion rates, it is often a secondary outcome despite sessile serrated lesions being the precursors of 15%-30% of all colorectal cancers[184]. Few studies have created AI systems optimized for dedicating sessile serrated lesions. Recently, Yoon *et al*[184] used a generative adversarial network (GAN) to generate endoscopic images of sessile serrated lesions which were used to train their DCNN with the hope of improving sessile serrated lesion detection. In the validation set which was comprised of 1141 images of polyps and 1000 normal images, their best performing GAN-DCNN model, GAN-aug2, achieved a sensitivity of 95.44% (95%CI 93.71%-97.17%), specificity of 90.10% (95%CI 88.38%-91.77%), accuracy of 92.95% (95%CI 91.86%-94.04%) and AUROC of 0.96 (95%CI 0.9547-0.9709)[184]. In a type-separated polyp validation dataset, the GAN-aug2 achieved a sensitivity of 95.24%, 19.1% higher than the DCNN without augmentation[184]. Given the small number of sessile serrate lesions present in the initial set, Yoon *et al*[184] collected an additional 130 images depicting 133 sessile serrated lesions to create an additional validation set titled SSL temporal validation dataset[184]. The GAN-aug2 continued to outperform the DCNN without augmentation (sensitivity 93.98% *vs* 84.21%). Nemoto *et al*[185] created a DCNN to differentiate (1) tubular adenomas from serrated lesions; and (2) serrated lesions from hyperplastic polyps. In their 215-image training set, the DCNN was able to differentiate tubular adenomas from sessile serrated lesions with sensitivity of 72% (95%CI 62%-81%), specificity 89% (95%CI 82%-94%), accuracy 82% (95%CI 77%-87%) and AUC 0.86 (95%CI 0.80-0.91). For differentiating sessile serrated lesions from hyperplastic polyps, the DCNN achieved a sensitivity of 17% (95%CI 7%-32%),

specificity 85% (95%CI 76%-92%), accuracy 63% (95%CI 54%-72%) and AUC 0.55 (95%CI 0.44-0.66)[185]. An overview of studies investigating the detection accuracy of CADe is provided in Table 3. An overview of studies investigating ADR and PDR using CADe is provided in Table 4.

FUTURE DIRECTIONS

Artificial intelligence is in its early stages for medicine, especially in gastroenterology and endoscopy. AI will help is in the areas of "augmentation" and "automation". Augmentation like what is happening with polyp detection and interpretation. Automation by eliminating electronic paperwork, such as the use of natural language processing for procedure documentation. Artificial intelligence systems have repeatedly been shown to be effective at identifying gastrointestinal lesions with high sensitivity, specificity and accuracy. While lesion detection is important, this is only the beginning of AI's utility in esophagogastroduodenoscopy, WCE and colonoscopy.

After refining their AI systems for lesion detection, several groups discussed in this narrative review were able to add additional functions to their AI systems. In BE, ESCC and gastric cancer, several AI systems were capable of predicting tumor invasion depth. Within IBD, AI systems were able to generate endoscopic disease severity scores. One group was able to train their CADe to recommend neoplasia biopsy sites in BE[14]. Additional efforts should be dedicated to developing these functions, testing them in real-time and having the AI system provide management recommendations when clinically appropriate.

Additional areas in need of future research are using AI systems to make histologic predictions, to assist with positioning of the endoscopic ultrasound (EUS) transducer and interpretation of EUS images, to detect biliary diseases and make therapeutic recommendations in endoscopic retrograde cholangiopancreatography (ERCP), and, in combination with endoscopic mechanical attachments, to improve colorectal cancer screening and surveillance. While endoscopists may perform optical biopsies of gastrointestinal lesions to predict histology and make real-time management decisions, these predictions are highly operator-dependent and often require expensive equipment that is not readily available. Thus, developing an AI system capable of performing objective optical biopsies, especially in WLE, would preserve the quality of histologic predictions, be cost effective, and avoid the risks associated with endoscopic biopsy and resection.

Similarly, EUS is highly operator-dependent, requiring endoscopists to place the transducer in specific positions to obtain adequate views of the hepatopancreatobiliary system. Research should focus on using AI systems to assist with appropriate transducer positioning and perform real-time EUS image analysis[186-194].

Presently, several clinical studies are actively recruiting patients to evaluate the utility of AI systems in ERCP. Of particular interest is the diagnosis and management of biliary diseases. Some groups are planning to use AI to classify bile duct lesions and provide biopsy site recommendations[195]. One group is planning to use an AI system in patients requiring biliary stents to assist with biliary stent choice and stent placement[196]. It will be interesting to see how AI performs in these tasks as successes could pave the way for future studies investigating the utility of AI systems to make real-time management recommendations.

While this narrative review focused on the use of AI in colonoscopy, of growing interest is the use of endoscopic mechanical attachments in colonoscopy to assist with polyp detection in colorectal cancer screening and surveillance. Independently, AI systems and endoscopic mechanical attachments are known to increase ADR and PDR. Few studies have investigated how combining AI with endoscopic mechanical attachments impacts ADR and PDR. Future research should examine the impact that combining these modalities has on ADR and PDR.

LIMITATIONS

While substantial advances have been made in AI, it is important to note that AI is not without limitations. In many of the studies discussed in this narrative review, the authors trained their AI systems using internally obtained images labeled by a single endoscopist. Thus, the AI is subject to the same operator biases and human error as the labeling endoscopist[1,197]. In addition, by using internally obtained data, several of these training sets may have inherent institutional or geographic biases resulting in AI systems that are biased and nongeneralizable[197]. As AI continues to progress, large datasets comprised of high-quality images should be created and used for training AI systems to reduce these biases[1].

With the implementation of AI in clinical practice, medical error accountability must also be addressed. While many of the AI systems discussed in this narrative review boast high detection accuracies, none are perfect. It is undeniable that errors in detection and diagnosis will arise when using these technologies. Regulatory bodies are needed to continually supervise these AI systems and oversee problems as they arise[198].

Table 3 Overview of findings from studies evaluating the detection accuracy of computer-aided detection for colonic polyps

Ref.	Country	Study design	Lesions	Training dataset	Test dataset	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUROC
Komeda *et al*[139], 2017	Japan	Retrospective	Adenomas	1200 images	10 images	80	60	70	-
Misawa *et al*[140], 2018	Japan	Retrospective	Polyps	411 video clips	135 video clips	90	63.3	76.5	0.87
Wang *et al* [149], 2018	China, United States	Retrospective	Polyps	4495 images	Dataset A: 27113 images; Dataset C: 138 video clips; Dataset D: 54 full-length videos	Dataset A: 94.38; Dataset C: 91.64	Dataset A: 95.92; Dataset D: 95.4	-	Dataset A: 0.984
Horiuchi *et al*[154], 2019	Japan	Prospective	Diminutive polyps	-	[a]	80	95.3	91.5	-
Hassan *et al*[141], 2020	Italy, United States	Retrospective	Polyps	-	338 video clips	99.7	-	-	-
Guo *et al* [142], 2021	Japan	Retrospective	Polyps	1991 images	100 video clips; 15 full videos	87[b]	98.3[b]	-	-
Neumann *et al*[143], 2021	Germany	Retrospective[1]	Polyps	> 500 videos	240 polyps within full-length videos	100	0	-	-
Li *et al* [144], 2021	Singapore	Retrospective	Polyps	6038 images	2571 images	74.1	85.1	-	-
Livovsky *et al*[151], 2021	Israel	Ambispective	Polyps	3611 h of videos	1393 h of videos	97.1	0	-	-
Pfeifer *et al* [158], 2021	Germany, Italy, Netherlands	Retrospective	Polyps	10467 images	45 videos	90	80	-	0.92
Ahmad *et al*[145], 2022[2]	England	Prospective	Polyps	Dataset A: 58849 frames; Dataset B: 10993 videos and still images	Dataset C: 110985 frames; Dataset D: 8950 frames; Dataset E: 542484 frames	Dataset C: 100, 84.1; Dataset D&E: 98.9, 85.2	Dataset C: 79.6; Dataset D&E: 79.3%		
Hori *et al* [146], 2022	Japan	Prospective	Polyps	1456 images	600 images	97	97.7	97.3	-
Pacal *et al* [152], 2022	Turkey	Retrospective	Polyps	Used images from 3 publicly available datasets (SUN, PICCOLO, Etis-Larib) to create training and test datasets		91.04	-	-	-
Yoon *et al* [184], 2022	South Korea	Retrospective	SSL	4397 images	Validation Set 2106; SSL Temporal Validation set 133	95.44; 93.89	90.1	92.95	0.96
Nemoto *et al*[185], 2022	Japan	Retrospective	TA, SSL	1849 images	400 images	72	89	82	0.86
Lux *et al* [148], 2022	Germany	Retrospective	Polyps	506338 images	41 full-length videos	-	-	95.3	-

[a]Tested CADe in a cohort of 95 patients.
[b]Per-frame analysis from full-length video dataset.
[1]Presumed retrospective based on manuscript.
[2]Sensitivity is reported as per-polyp and per-frame respectively. Specificity is reported as per-frame.
AUROC: Area under the receiver operating characteristic; SSL: Sessile serrate lesion; TA: Tubular adenoma. All studies used a convolutional neural network to classify images.

CONCLUSION

In this narrative review, we provide an objective overview of the AI-related research being performed within esophagogastroduodenoscopy, WCE and colonoscopy. We attempted to be comprehensive by using several electronic databases including Embase, Ovid Medicine, and PubMed. However, it is possible that some publications pertinent to our narrative review were missed.

Undoubtedly, AI within esophagogastroduodenoscopy, WCE and colonoscopy is rapidly evolving, moving from retrospectively tested supervised learning algorithms to large, multicenter clinical trials using completely autonomous systems within the span of 10 years. The systems developed by these researchers show promise for detecting lesions, diagnosing conditions, and monitoring diseases. In fact, two of the computer aided detection systems discussed in this narrative review designed to aid with colorectal polyp detection were approved by the United States Food and Drug Administration in 2021 [171,199]. Thus, the question is no longer if but when will AI become integrated with clinical practice. Medical providers at all levels of training should prepare to incorporate artificial intelligence systems into routine practice.

Table 4 Overview of findings from studies evaluating computer-aided detection for adenoma detection rate and polyp detection rate

Ref.	Country	Study design	Patients (*n*)		PDR (%)			ADR (%)		
			CADe	SC	CADe	SC	*P* value	CADe	SC	*P* value
Wang *et al* [168], 2019	China, United States	Randomized	522	536	45.02	29.1	< 0.001	29.12	20.34	< 0.001
Becq *et al*[155], 2020	United States, Turkey, Costa Rica	Prospective	50[b]		82	62	Not reported	-	-	-
Gong *et al*[166], 2020	China	Randomized	355	349	47	34	0.0016	16	8	0.001
Liu *et al*[171], 2020	China, United States	Randomized	393	397	47.07	33.25	< 0.001	29.01	20.91	0.009
Liu *et al*[173], 2020	China	Prospective	508	518	43.65	27.81	< 0.001	39.1	23.89	< 0.001
Repici *et al* [170], 2020	Italy, Kuwait, United States, Germany	Randomized	341	344	-	-	-	54.8	40.4	< 0.001
Su *et al*[169], 2020	China	Randomized	308	315	38.3	25.4	0.001	28.9	16.5	< 0.001
Wang *et al* [156], 2020	China, United States	Prospective, Tandem[1]	184	185	65.59	55.14	0.099	42.39	35.68	0.186
Wang *et al* [167], 2020	China, United States	Randomized	484	478	52	37	< 0.0001	34	28	0.03
Kamba *et al* [164], 2021	Japan	Randomized, Tandem[2]	172	174	69.8	60.9	0.084	64.5	53.6	0.036
Luo *et al*[174], 2021	China	Randomized, Tandem[1]	72	78	38.7	34	< 0.001	-	-	-
Pfeifer *et al* [158], 2021	Germany, Italy, Netherlands	Prospective, Tandem[1]	42[b]		50	38	0.023	36	26	0.044
Shaukat *et al* [157], 2021	United States, England	Prospective	83	283	-	-	-	54.2	40.6	0.028
Shen *et al*[150], 2021	China	Ambispective	64	64	78.1	56.3	0.008	53.1	29.7	0.007
Xu *et al*[172], 2021	China	Randomized	1177	1175	38.8	36.2	0.183	-	-	-
Glissen Brown *et al*[163], 2022	China, United States	Randomized, Tandem[2]	113	110	70.8	65.45	0.3923	50.44	43.64	0.3091
Ishiyama *et al* [159], 2022	Japan, Norway	Prospective	918	918	59	52.1	0.003	26.4	19.9	0.001
Lux *et al*[148], 2022	Germany	Retrospective	41	-	-	-	-	-	41.5	-

Quan *et al*[153], 2022	United States	Prospective	300	300	-	-	-	43.7[a]; 66.7	37.8[a]; 59.72	0.37[a]; 0.35
Repici *et al* [165], 2022	Italy, Switzerland, United States, Germany	Randomized	330	330	-	-	-	53.3	44.5	0.017
Shaukat *et al* [162], 2022	United States	Randomized	682	677	64.4	61.2	0.242	47.8	43.9	0.065
Zippelius *et al* [160], 2022	Germany, United States	Prospective	150[b]		-	-	-	50.7	52	0.5

[a]Quan *et al*[153] reported results by indication, screening and surveillance respectively.
[b]The same patients were used to compare CADe versus standard colonoscopy; Becq *et al*[155] recorded 50 colonoscopy videos that were analyzed by CADe and reviewed by endoscopists separately; Pfeifer *et al*[158] performed standard colonoscopy followed by CADe-assisted colonoscopy in all 42 patients; Zippelius *et al*[160] had their CADe analyze their patients while the endoscopists performed their colonoscopies.
[1]Performed analyses using data obtained from whole process.
[2]Performed analyses using data obtained from first pass.
ADR: Adenoma detection rate; CADe: Computer-aided detection; PDR: Polyp detection rate; SC: Standard colonoscopy. All studies used a convolutional neural network to classify images.

FOOTNOTES

Author contributions: Galati JS, Gross SA contributed to manuscript concept and design; Galati JS, Duve RJ, O'Mara M contributed to obtaining and interpreting literary sources, drafting of manuscript; Galati JS, Duve RJ, O'Mara M, Gross SA contributed to revision of manuscript; All authors read and approved the final version of the manuscript.

Conflict-of-interest statement: All the authors declare that they have no conflict of interest.

Country/Territory of origin: United States

ORCID number: Jonathan S Galati 0000-0001-8075-3376; Seth A Gross 0000-0001-8942-865X.

S-Editor: Liu JH
L-Editor: A
P-Editor: Liu JH

REFERENCES

1 **Kröner PT**, Engels MM, Glicksberg BS, Johnson KW, Mzaik O, van Hooft JE, Wallace MB, El-Serag HB, Krittanawong C. Artificial intelligence in gastroenterology: A state-of-the-art review. *World J Gastroenterol* 2021; **27**: 6794-6824 [PMID: 34790008 DOI: 10.3748/wjg.v27.i40.6794]
2 **Kaul V**, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointest Endosc* 2020; **92**: 807-812 [PMID: 32565184 DOI: 10.1016/j.gie.2020.06.040]
3 **Deo RC**. Machine Learning in Medicine. *Circulation* 2015; **132**: 1920-1930 [PMID: 26572668 DOI: 10.1161/CIRCULATIONAHA.115.001593]
4 **Le Berre C**, Sandborn WJ, Aridhi S, Devignes MD, Fournier L, Smaïl-Tabbone M, Danese S, Peyrin-Biroulet L. Application of Artificial Intelligence to Gastroenterology and Hepatology. *Gastroenterology* 2020; **158**: 76-94.e2 [PMID: 31593701 DOI: 10.1053/j.gastro.2019.08.058]
5 **Christou CD**, Tsoulfas G. Challenges and opportunities in the application of artificial intelligence in gastroenterology and hepatology. *World J Gastroenterol* 2021; **27**: 6191-6223 [PMID: 34712027 DOI: 10.3748/wjg.v27.i37.6191]
6 **Pannala R**, Krishnan K, Melson J, Parsi MA, Schulman AR, Sullivan S, Trikudanathan G, Trindade AJ, Watson RR, Maple JT, Lichtenstein DR. Artificial intelligence in gastrointestinal endoscopy. *VideoGIE* 2020; **5**: 598-613 [PMID: 33319126 DOI: 10.1016/j.vgie.2020.08.013]
7 **Hussein M**, González-Bueno Puyal J, Mountney P, Lovat LB, Haidry R. Role of artificial intelligence in the diagnosis of oesophageal neoplasia: 2020 an endoscopic odyssey. *World J Gastroenterol* 2020; **26**: 5784-5796 [PMID: 33132634 DOI: 10.3748/wjg.v26.i38.5784]
8 **Sharma P**. Barrett Esophagus: A Review. *JAMA* 2022; **328**: 663-671 [PMID: 35972481 DOI: 10.1001/jama.2022.13298]
9 **Bujanda DE**, Hachem C. Barrett's Esophagus. *Mo Med* 2018; **115**: 211-213 [PMID: 30228724]
10 **van der Sommen F**, Zinger S, Curvers WL, Bisschops R, Pech O, Weusten BL, Bergman JJ, de With PH, Schoon EJ.

Computer-aided detection of early neoplastic lesions in Barrett's esophagus. *Endoscopy* 2016; **48**: 617-624 [PMID: 27100718 DOI: 10.1055/s-0042-105284]

11 **Swager AF**, van der Sommen F, Klomp SR, Zinger S, Meijer SL, Schoon EJ, Bergman JJGHM, de With PH, Curvers WL. Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy. *Gastrointest Endosc* 2017; **86**: 839-846 [PMID: 28322771 DOI: 10.1016/j.gie.2017.03.011]

12 **Struyvenberg MR**, van der Sommen F, Swager AF, de Groof AJ, Rikos A, Schoon EJ, Bergman JJ, de With PHN, Curvers WL. Improved Barrett's neoplasia detection using computer-assisted multiframe analysis of volumetric laser endomicroscopy. *Dis Esophagus* 2020; **33** [PMID: 31364700 DOI: 10.1093/dote/doz065]

13 **de Groof J**, van der Sommen F, van der Putten J, Struyvenberg MR, Zinger S, Curvers WL, Pech O, Meining A, Neuhaus H, Bisschops R, Schoon EJ, de With PH, Bergman JJ. The Argos project: The development of a computer-aided detection system to improve detection of Barrett's neoplasia on white light endoscopy. *United European Gastroenterol J* 2019; **7**: 538-547 [PMID: 31065371 DOI: 10.1177/2050640619837443]

14 **de Groof AJ**, Struyvenberg MR, van der Putten J, van der Sommen F, Fockens KN, Curvers WL, Zinger S, Pouw RE, Coron E, Baldaque-Silva F, Pech O, Weusten B, Meining A, Neuhaus H, Bisschops R, Dent J, Schoon EJ, de With PH, Bergman JJ. Deep-Learning System Detects Neoplasia in Patients With Barrett's Esophagus With Higher Accuracy Than Endoscopists in a Multistep Training and Validation Study With Benchmarking. *Gastroenterology* 2020; **158**: 915-929.e4 [PMID: 31759929 DOI: 10.1053/j.gastro.2019.11.030]

15 **de Groof AJ**, Struyvenberg MR, Fockens KN, van der Putten J, van der Sommen F, Boers TG, Zinger S, Bisschops R, de With PH, Pouw RE, Curvers WL, Schoon EJ, Bergman JJGHM. Deep learning algorithm detection of Barrett's neoplasia with high accuracy during live endoscopic procedures: a pilot study (with video). *Gastrointest Endosc* 2020; **91**: 1242-1250 [PMID: 31926965 DOI: 10.1016/j.gie.2019.12.048]

16 **Struyvenberg MR**, de Groof AJ, van der Putten J, van der Sommen F, Baldaque-Silva F, Omae M, Pouw R, Bisschops R, Vieth M, Schoon EJ, Curvers WL, de With PH, Bergman JJ. A computer-assisted algorithm for narrow-band imaging-based tissue characterization in Barrett's esophagus. *Gastrointest Endosc* 2021; **93**: 89-98 [PMID: 32504696 DOI: 10.1016/j.gie.2020.05.050]

17 **Jisu Hong**, Bo-Yong Park, Hyunjin Park. Convolutional neural network classifier for distinguishing Barrett's esophagus and neoplasia endomicroscopy images. *Annu Int Conf IEEE Eng Med Biol Soc* 2017; **2017**: 2892-2895 [PMID: 29060502 DOI: 10.1109/EMBC.2017.8037461]

18 **Ebigbo A**, Mendel R, Probst A, Manzeneder J, Prinz F, de Souza LA Jr, Papa J, Palm C, Messmann H. Real-time use of artificial intelligence in the evaluation of cancer in Barrett's oesophagus. *Gut* 2020; **69**: 615-616 [PMID: 31541004 DOI: 10.1136/gutjnl-2019-319460]

19 **Hashimoto R**, Requa J, Dao T, Ninh A, Tran E, Mai D, Lugo M, El-Hage Chehade N, Chang KJ, Karnes WE, Samarasena JB. Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). *Gastrointest Endosc* 2020; **91**: 1264-1271.e1 [PMID: 31930967 DOI: 10.1016/j.gie.2019.12.049]

20 **Hussein M**, González-Bueno Puyal J, Lines D, Sehgal V, Toth D, Ahmad OF, Kader R, Everson M, Lipman G, Fernandez-Sordo JO, Ragunath K, Esteban JM, Bisschops R, Banks M, Haefner M, Mountney P, Stoyanov D, Lovat LB, Haidry R. A new artificial intelligence system successfully detects and localises early neoplasia in Barrett's esophagus by using convolutional neural networks. *United European Gastroenterol J* 2022; **10**: 528-537 [PMID: 35521666 DOI: 10.1002/ueg2.12233]

21 **Ebigbo A**, Mendel R, Probst A, Manzeneder J, Souza LA Jr, Papa JP, Palm C, Messmann H. Computer-aided diagnosis using deep learning in the evaluation of early oesophageal adenocarcinoma. *Gut* 2019; **68**: 1143-1145 [PMID: 30510110 DOI: 10.1136/gutjnl-2018-317573]

22 **Ali S**, Bailey A, Ash S, Haghighat M; TGU Investigators, Leedham SJ, Lu X, East JE, Rittscher J, Braden B. A Pilot Study on Automatic Three-Dimensional Quantification of Barrett's Esophagus for Risk Stratification and Therapy Monitoring. *Gastroenterology* 2021; **161**: 865-878.e8 [PMID: 34116029 DOI: 10.1053/j.gastro.2021.05.059]

23 **Ebigbo A**, Mendel R, Rückert T, Schuster L, Probst A, Manzeneder J, Prinz F, Mende M, Steinbrück I, Faiss S, Rauber D, de Souza LA Jr, Papa JP, Deprez PH, Oyama T, Takahashi A, Seewald S, Sharma P, Byrne MF, Palm C, Messmann H. Endoscopic prediction of submucosal invasion in Barrett's cancer with the use of artificial intelligence: a pilot study. *Endoscopy* 2021; **53**: 878-883 [PMID: 33197942 DOI: 10.1055/a-1311-8570]

24 **Ghatwary N**, Zolgharni M, Ye X. Early esophageal adenocarcinoma detection using deep learning methods. *Int J Comput Assist Radiol Surg* 2019; **14**: 611-621 [PMID: 30666547 DOI: 10.1007/s11548-019-01914-4]

25 **Iwagami H**, Ishihara R, Aoyama K, Fukuda H, Shimamoto Y, Kono M, Nakahira H, Matsuura N, Shichijo S, Kanesaka T, Kanzaki H, Ishii T, Nakatani Y, Tada T. Artificial intelligence for the detection of esophageal and esophagogastric junctional adenocarcinoma. *J Gastroenterol Hepatol* 2021; **36**: 131-136 [PMID: 32511793 DOI: 10.1111/jgh.15136]

26 **Zhang Y**. Epidemiology of esophageal cancer. *World J Gastroenterol* 2013; **19**: 5598-5606 [PMID: 24039351 DOI: 10.3748/wjg.v19.i34.5598]

27 **Shin D**, Protano MA, Polydorides AD, Dawsey SM, Pierce MC, Kim MK, Schwarz RA, Quang T, Parikh N, Bhutani MS, Zhang F, Wang G, Xue L, Wang X, Xu H, Anandasabapathy S, Richards-Kortum RR. Quantitative analysis of high-resolution microendoscopic images for diagnosis of esophageal squamous cell carcinoma. *Clin Gastroenterol Hepatol* 2015; **13**: 272-279.e2 [PMID: 25066838 DOI: 10.1016/j.cgh.2014.07.030]

28 **Quang T**, Schwarz RA, Dawsey SM, Tan MC, Patel K, Yu X, Wang G, Zhang F, Xu H, Anandasabapathy S, Richards-Kortum R. A tablet-interfaced high-resolution microendoscope with automated image interpretation for real-time evaluation of esophageal squamous cell neoplasia. *Gastrointest Endosc* 2016; **84**: 834-841 [PMID: 27036635 DOI: 10.1016/j.gie.2016.03.1472]

29 **Cai SL**, Li B, Tan WM, Niu XJ, Yu HH, Yao LQ, Zhou PH, Yan B, Zhong YS. Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma (with video). *Gastrointest Endosc* 2019; **90**: 745-753.e2 [PMID: 31302091 DOI: 10.1016/j.gie.2019.06.044]

30 **Liu G**, Hua J, Wu Z, Meng T, Sun M, Huang P, He X, Sun W, Li X, Chen Y. Automatic classification of esophageal

lesions in endoscopic images using a convolutional neural network. *Ann Transl Med* 2020; **8**: 486 [PMID: 32395530 DOI: 10.21037/atm.2020.03.24]

31 **Liu W**, Yuan X, Guo L, Pan F, Wu C, Sun Z, Tian F, Yuan C, Zhang W, Bai S, Feng J, Hu Y, Hu B. Artificial Intelligence for Detecting and Delineating Margins of Early ESCC Under WLI Endoscopy. *Clin Transl Gastroenterol* 2022; **13**: e00433 [PMID: 35130184 DOI: 10.14309/ctg.0000000000000433]

32 **Ohmori M**, Ishihara R, Aoyama K, Nakagawa K, Iwagami H, Matsuura N, Shichijo S, Yamamoto K, Nagaike K, Nakahara M, Inoue T, Aoi K, Okada H, Tada T. Endoscopic detection and differentiation of esophageal lesions using a deep neural network. *Gastrointest Endosc* 2020; **91**: 301-309.e1 [PMID: 31585124 DOI: 10.1016/j.gie.2019.09.034]

33 **Guo L**, Xiao X, Wu C, Zeng X, Zhang Y, Du J, Bai S, Xie J, Zhang Z, Li Y, Wang X, Cheung O, Sharma M, Liu J, Hu B. Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos). *Gastrointest Endosc* 2020; **91**: 41-51 [PMID: 31445040 DOI: 10.1016/j.gie.2019.08.018]

34 **Fukuda H**, Ishihara R, Kato Y, Matsunaga T, Nishida T, Yamada T, Ogiyama H, Horie M, Kinoshita K, Tada T. Comparison of performances of artificial intelligence versus expert endoscopists for real-time assisted diagnosis of esophageal squamous cell carcinoma (with video). *Gastrointest Endosc* 2020; **92**: 848-855 [PMID: 32505685 DOI: 10.1016/j.gie.2020.05.043]

35 **Li B**, Cai SL, Tan WM, Li JC, Yalikong A, Feng XS, Yu HH, Lu PX, Feng Z, Yao LQ, Zhou PH, Yan B, Zhong YS. Comparative study on artificial intelligence systems for detecting early esophageal squamous cell carcinoma between narrow-band and white-light imaging. *World J Gastroenterol* 2021; **27**: 281-293 [PMID: 33519142 DOI: 10.3748/wjg.v27.i3.281]

36 **Shiroma S**, Yoshio T, Kato Y, Horie Y, Namikawa K, Tokai Y, Yoshimizu S, Yoshizawa N, Horiuchi Y, Ishiyama A, Hirasawa T, Tsuchida T, Akazawa N, Akiyama J, Tada T, Fujisaki J. Ability of artificial intelligence to detect T1 esophageal squamous cell carcinoma from endoscopic videos and the effects of real-time assistance. *Sci Rep* 2021; **11**: 7759 [PMID: 33833355 DOI: 10.1038/s41598-021-87405-6]

37 **Horie Y**, Yoshio T, Aoyama K, Yoshimizu S, Horiuchi Y, Ishiyama A, Hirasawa T, Tsuchida T, Ozawa T, Ishihara S, Kumagai Y, Fujishiro M, Maetani I, Fujisaki J, Tada T. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest Endosc* 2019; **89**: 25-32 [PMID: 30120958 DOI: 10.1016/j.gie.2018.07.037]

38 **Kumagai Y**, Takubo K, Kawada K, Aoyama K, Endo Y, Ozawa T, Hirasawa T, Yoshio T, Ishihara S, Fujishiro M, Tamaru JI, Mochiki E, Ishida H, Tada T. Diagnosis using deep-learning artificial intelligence based on the endocytoscopic observation of the esophagus. *Esophagus* 2019; **16**: 180-187 [PMID: 30547352 DOI: 10.1007/s10388-018-0651-7]

39 **Everson M**, Herrera L, Li W, Luengo IM, Ahmad O, Banks M, Magee C, Alzoubaidi D, Hsu HM, Graham D, Vercauteren T, Lovat L, Ourselin S, Kashin S, Wang HP, Wang WL, Haidry RJ. Artificial intelligence for the real-time classification of intrapapillary capillary loop patterns in the endoscopic diagnosis of early oesophageal squamous cell carcinoma: A proof-of-concept study. *United European Gastroenterol J* 2019; **7**: 297-306 [PMID: 31080614 DOI: 10.1177/2050640618821800]

40 **Zhao YY**, Xue DX, Wang YL, Zhang R, Sun B, Cai YP, Feng H, Cai Y, Xu JM. Computer-assisted diagnosis of early esophageal squamous cell carcinoma using narrow-band imaging magnifying endoscopy. *Endoscopy* 2019; **51**: 333-341 [PMID: 30469155 DOI: 10.1055/a-0756-8754]

41 **Nakagawa K**, Ishihara R, Aoyama K, Ohmori M, Nakahira H, Matsuura N, Shichijo S, Nishida T, Yamada T, Yamaguchi S, Ogiyama H, Egawa S, Kishida O, Tada T. Classification for invasion depth of esophageal squamous cell carcinoma using a deep neural network compared with experienced endoscopists. *Gastrointest Endosc* 2019; **90**: 407-414 [PMID: 31077698 DOI: 10.1016/j.gie.2019.04.245]

42 **Shimamoto Y**, Ishihara R, Kato Y, Shoji A, Inoue T, Matsueda K, Miyake M, Waki K, Kono M, Fukuda H, Matsuura N, Nagaike K, Aoi K, Yamamoto K, Nakahara M, Nishihara A, Tada T. Real-time assessment of video images for esophageal squamous cell carcinoma invasion depth using artificial intelligence. *J Gastroenterol* 2020; **55**: 1037-1045 [PMID: 32778959 DOI: 10.1007/s00535-020-01716-5]

43 **Tokai Y**, Yoshio T, Aoyama K, Horie Y, Yoshimizu S, Horiuchi Y, Ishiyama A, Tsuchida T, Hirasawa T, Sakakibara Y, Yamada T, Yamaguchi S, Fujisaki J, Tada T. Application of artificial intelligence using convolutional neural networks in determining the invasion depth of esophageal squamous cell carcinoma. *Esophagus* 2020; **17**: 250-256 [PMID: 31980977 DOI: 10.1007/s10388-020-00716-x]

44 **Hirasawa T**, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, Ohnishi T, Fujishiro M, Matsuo K, Fujisaki J, Tada T. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer* 2018; **21**: 653-660 [PMID: 29335825 DOI: 10.1007/s10120-018-0793-2]

45 **Lee JH**, Kim YJ, Kim YW, Park S, Choi YI, Park DK, Kim KG, Chung JW. Spotting malignancies from gastric endoscopic images using deep learning. *Surg Endosc* 2019; **33**: 3790-3797 [PMID: 30719560 DOI: 10.1007/s00464-019-06677-2]

46 **Li L**, Chen Y, Shen Z, Zhang X, Sang J, Ding Y, Yang X, Li J, Chen M, Jin C, Chen C, Yu C. Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. *Gastric Cancer* 2020; **23**: 126-132 [PMID: 31332619 DOI: 10.1007/s10120-019-00992-2]

47 **Miyaki R**, Yoshida S, Tanaka S, Kominami Y, Sanomura Y, Matsuo T, Oka S, Raytchev B, Tamaki T, Koide T, Kaneda K, Yoshihara M, Chayama K. Quantitative identification of mucosal gastric cancer under magnifying endoscopy with flexible spectral imaging color enhancement. *J Gastroenterol Hepatol* 2013; **28**: 841-847 [PMID: 23424994 DOI: 10.1111/jgh.12149]

48 **Kanesaka T**, Lee TC, Uedo N, Lin KP, Chen HZ, Lee JY, Wang HP, Chang HT. Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging. *Gastrointest Endosc* 2018; **87**: 1339-1344 [PMID: 29225083 DOI: 10.1016/j.gie.2017.11.029]

49 **Tang D**, Wang L, Ling T, Lv Y, Ni M, Zhan Q, Fu Y, Zhuang D, Guo H, Dou X, Zhang W, Xu G, Zou X. Development and validation of a real-time artificial intelligence-assisted system for detecting early gastric cancer: A multicentre retrospective diagnostic study. *EBioMedicine* 2020; **62**: 103146 [PMID: 33254026 DOI: 10.1016/j.ebiom.2020.103146]

50 **Wu L**, Zhou W, Wan X, Zhang J, Shen L, Hu S, Ding Q, Mu G, Yin A, Huang X, Liu J, Jiang X, Wang Z, Deng Y, Liu M, Lin R, Ling T, Li P, Wu Q, Jin P, Chen J, Yu H. A deep neural network improves endoscopic detection of early gastric cancer without blind spots. *Endoscopy* 2019; **51**: 522-531 [PMID: 30861533 DOI: 10.1055/a-0855-3532]

51 **Cho BJ**, Bang CS, Park SW, Yang YJ, Seo SI, Lim H, Shin WG, Hong JT, Yoo YT, Hong SH, Choi JH, Lee JJ, Baik GH. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. *Endoscopy* 2019; **51**: 1121-1129 [PMID: 31443108 DOI: 10.1055/a-0981-6133]

52 **Namikawa K**, Hirasawa T, Nakano K, Ikenoyama Y, Ishioka M, Shiroma S, Tokai Y, Yoshimizu S, Horiuchi Y, Ishiyama A, Yoshio T, Tsuchida T, Fujisaki J, Tada T. Artificial intelligence-based diagnostic system classifying gastric cancers and ulcers: comparison between the original and newly developed systems. *Endoscopy* 2020; **52**: 1077-1083 [PMID: 32503056 DOI: 10.1055/a-1194-8771]

53 **Yuan XL**, Zhou Y, Liu W, Luo Q, Zeng XH, Yi Z, Hu B. Artificial intelligence for diagnosing gastric lesions under white-light endoscopy. *Surg Endosc* 2022; **36**: 9444-9453 [PMID: 35879572 DOI: 10.1007/s00464-022-09420-6]

54 **Guo L**, Gong H, Wang Q, Zhang Q, Tong H, Li J, Lei X, Xiao X, Li C, Jiang J, Hu B, Song J, Tang C, Huang Z. Detection of multiple lesions of gastrointestinal tract for endoscopy using artificial intelligence model: a pilot study. *Surg Endosc* 2021; **35**: 6532-6538 [PMID: 33185766 DOI: 10.1007/s00464-020-08150-x]

55 **Ikenoyama Y**, Hirasawa T, Ishioka M, Namikawa K, Yoshimizu S, Horiuchi Y, Ishiyama A, Yoshio T, Tsuchida T, Takeuchi Y, Shichijo S, Katayama N, Fujisaki J, Tada T. Detecting early gastric cancer: Comparison between the diagnostic ability of convolutional neural networks and endoscopists. *Dig Endosc* 2021; **33**: 141-150 [PMID: 32282110 DOI: 10.1111/den.13688]

56 **Horiuchi Y**, Aoyama K, Tokai Y, Hirasawa T, Yoshimizu S, Ishiyama A, Yoshio T, Tsuchida T, Fujisaki J, Tada T. Convolutional Neural Network for Differentiating Gastric Cancer from Gastritis Using Magnified Endoscopy with Narrow Band Imaging. *Dig Dis Sci* 2020; **65**: 1355-1363 [PMID: 31584138 DOI: 10.1007/s10620-019-05862-6]

57 **Horiuchi Y**, Hirasawa T, Ishizuka N, Tokai Y, Namikawa K, Yoshimizu S, Ishiyama A, Yoshio T, Tsuchida T, Fujisaki J, Tada T. Performance of a computer-aided diagnosis system in diagnosing early gastric cancer using magnifying endoscopy videos with narrow-band imaging (with videos). *Gastrointest Endosc* 2020; **92**: 856-865.e1 [PMID: 32422155 DOI: 10.1016/j.gie.2020.04.079]

58 **Hu H**, Gong L, Dong D, Zhu L, Wang M, He J, Shu L, Cai Y, Cai S, Su W, Zhong Y, Li C, Zhu Y, Fang M, Zhong L, Yang X, Zhou P, Tian J. Identifying early gastric cancer under magnifying narrow-band images with deep learning: a multicenter study. *Gastrointest Endosc* 2021; **93**: 1333-1341.e3 [PMID: 33248070 DOI: 10.1016/j.gie.2020.11.014]

59 **Ueyama H**, Kato Y, Akazawa Y, Yatagai N, Komori H, Takeda T, Matsumoto K, Ueda K, Hojo M, Yao T, Nagahara A, Tada T. Application of artificial intelligence using a convolutional neural network for diagnosis of early gastric cancer based on magnifying endoscopy with narrow-band imaging. *J Gastroenterol Hepatol* 2021; **36**: 482-489 [PMID: 32681536 DOI: 10.1111/jgh.15190]

60 **Yoon HJ**, Kim S, Kim JH, Keum JS, Oh SI, Jo J, Chun J, Youn YH, Park H, Kwon IG, Choi SH, Noh SH. A Lesion-Based Convolutional Neural Network Improves Endoscopic Detection and Depth Prediction of Early Gastric Cancer. *J Clin Med* 2019; **8** [PMID: 31454949 DOI: 10.3390/jcm8091310]

61 **Zhu Y**, Wang QC, Xu MD, Zhang Z, Cheng J, Zhong YS, Zhang YQ, Chen WF, Yao LQ, Zhou PH, Li QL. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. *Gastrointest Endosc* 2019; **89**: 806-815.e1 [PMID: 30452913 DOI: 10.1016/j.gie.2018.11.011]

62 **Cho BJ**, Bang CS, Lee JJ, Seo CW, Kim JH. Prediction of Submucosal Invasion for Gastric Neoplasms in Endoscopic Images Using Deep-Learning. *J Clin Med* 2020; **9** [PMID: 32549190 DOI: 10.3390/jcm9061858]

63 **Nagao S**, Tsuji Y, Sakaguchi Y, Takahashi Y, Minatsuki C, Niimi K, Yamashita H, Yamamichi N, Seto Y, Tada T, Koike K. Highly accurate artificial intelligence systems to predict the invasion depth of gastric cancer: efficacy of conventional white-light imaging, nonmagnifying narrow-band imaging, and indigo-carmine dye contrast imaging. *Gastrointest Endosc* 2020; **92**: 866-873.e1 [PMID: 32592776 DOI: 10.1016/j.gie.2020.06.047]

64 **Ku Y**, Ding H, Wang G. Efficient Synchronous Real-Time CADe for Multicategory Lesions in Gastroscopy by Using Multiclass Detection Model. *Biomed Res Int* 2022; **2022**: 8504149 [PMID: 36093395 DOI: 10.1155/2022/8504149]

65 **Watanabe K**, Nagata N, Shimbo T, Nakashima R, Furuhata E, Sakurai T, Akazawa N, Yokoi C, Kobayakawa M, Akiyama J, Mizokami M, Uemura N. Accuracy of endoscopic diagnosis of Helicobacter pylori infection according to level of endoscopic experience and the effect of training. *BMC Gastroenterol* 2013; **13**: 128 [PMID: 23947684 DOI: 10.1186/1471-230X-13-128]

66 **Shichijo S**, Nomura S, Aoyama K, Nishikawa Y, Miura M, Shinagawa T, Takiyama H, Tanimoto T, Ishihara S, Matsuo K, Tada T. Application of Convolutional Neural Networks in the Diagnosis of Helicobacter pylori Infection Based on Endoscopic Images. *EBioMedicine* 2017; **25**: 106-111 [PMID: 29056541 DOI: 10.1016/j.ebiom.2017.10.014]

67 **Shichijo S**, Endo Y, Aoyama K, Takeuchi Y, Ozawa T, Takiyama H, Matsuo K, Fujishiro M, Ishihara S, Ishihara R, Tada T. Application of convolutional neural networks for evaluating Helicobacter pylori infection status on the basis of endoscopic images. *Scand J Gastroenterol* 2019; **54**: 158-163 [PMID: 30879352 DOI: 10.1080/00365521.2019.1577486]

68 **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 [PMID: 29399610 DOI: 10.1055/s-0043-120830]

69 **Zheng W**, Zhang X, Kim JJ, Zhu X, Ye G, Ye B, Wang J, Luo S, Li J, Yu T, Liu J, Hu W, Si J. High Accuracy of Convolutional Neural Network for Evaluation of Helicobacter pylori Infection Based on Endoscopic Images: Preliminary Experience. *Clin Transl Gastroenterol* 2019; **10**: e00109 [PMID: 31833862 DOI: 10.14309/ctg.0000000000000109]

70 **Yasuda T**, Hiroyasu T, Hiwa S, Okada Y, Hayashi S, Nakahata Y, Yasuda Y, Omatsu T, Obora A, Kojima T, Ichikawa H, Yagi N. Potential of automatic diagnosis system with linked color imaging for diagnosis of Helicobacter pylori infection. *Dig Endosc* 2020; **32**: 373-381 [PMID: 31398276 DOI: 10.1111/den.13509]

71 **Nakashima H**, Kawahira H, Kawachi H, Sakaki N. Endoscopic three-categorical diagnosis of Helicobacter pylori infection using linked color imaging and deep learning: a single-center prospective study (with video). *Gastric Cancer* 2020; **23**: 1033-1040 [PMID: 32382973 DOI: 10.1007/s10120-020-01077-1]

72 **Stoleru CA**, Dulf EH, Ciobanu L. Automated detection of celiac disease using Machine Learning Algorithms. *Sci Rep* 2022; **12**: 4071 [PMID: 35260574 DOI: 10.1038/s41598-022-07199-z]

73 **Gadermayr M**, Kogler H, Karla M, Merhof D, Uhl A, Vécsei A. Computer-aided texture analysis combined with experts' knowledge: Improving endoscopic celiac disease diagnosis. *World J Gastroenterol* 2016; **22**: 7124-7134 [PMID: 27610022 DOI: 10.3748/wjg.v22.i31.7124]

74 **Wimmer G**, Hegenbart S, Vecsei A, Uhl A. Convolutional Neural Network Architectures for the Automated Diagnosis of Celiac Disease. Presented at: International Workshop on Computer-Assisted and Robotic Endoscopy. CARE 2016; 104–113 [DOI: 10.1007/978-3-319-54057-3_10]

75 **Zhou T**, Han G, Li BN, Lin Z, Ciaccio EJ, Green PH, Qin J. Quantitative analysis of patients with celiac disease by video capsule endoscopy: A deep learning method. *Comput Biol Med* 2017; **85**: 1-6 [PMID: 28412572 DOI: 10.1016/j.compbiomed.2017.03.031]

76 **Wang X**, Qian H, Ciaccio EJ, Lewis SK, Bhagat G, Green PH, Xu S, Huang L, Gao R, Liu Y. Celiac disease diagnosis from videocapsule endoscopy images with residual learning and deep feature extraction. *Comput Methods Programs Biomed* 2020; **187**: 105236 [PMID: 31786452 DOI: 10.1016/j.cmpb.2019.105236]

77 **Yogapriya J**, Chandran V, Sumithra MG, Anitha P, Jenopaul P, Suresh Gnana Dhas C. Gastrointestinal Tract Disease Classification from Wireless Endoscopy Images Using Pretrained Deep Learning Model. *Comput Math Methods Med* 2021; **2021**: 5940433 [PMID: 34545292 DOI: 10.1155/2021/5940433]

78 **Charisis VS**, Hadjileontiadis LJ, Liatsos CN, Mavrogiannis CC, Sergiadis GD. Capsule endoscopy image analysis using texture information from various colour models. *Comput Methods Programs Biomed* 2012; **107**: 61-74 [PMID: 22056811 DOI: 10.1016/j.cmpb.2011.10.004]

79 **Charisis VS**, Hadjileontiadis LJ. Potential of hybrid adaptive filtering in inflammatory lesion detection from capsule endoscopy images. *World J Gastroenterol* 2016; **22**: 8641-8657 [PMID: 27818583 DOI: 10.3748/wjg.v22.i39.8641]

80 **Kumar R**, Zhao Q, Seshamani S, Mullin G, Hager G, Dassopoulos T. Assessment of Crohn's disease lesions in wireless capsule endoscopy images. *IEEE Trans Biomed Eng* 2012; **59**: 355-362 [PMID: 22020661 DOI: 10.1109/TBME.2011.2172438]

81 **Wei Z**, Wang W, Bradfield J, Li J, Cardinale C, Frackelton E, Kim C, Mentch F, Van Steen K, Visscher PM, Baldassano RN, Hakonarson H; International IBD Genetics Consortium. Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease. *Am J Hum Genet* 2013; **92**: 1008-1012 [PMID: 23731541 DOI: 10.1016/j.ajhg.2013.05.002]

82 **Ferreira JPS**, de Mascarenhas Saraiva MJDQEC, Afonso JPL, Ribeiro TFC, Cardoso HMC, Ribeiro Andrade AP, de Mascarenhas Saraiva MNG, Parente MPL, Natal Jorge R, Lopes SIO, de Macedo GMG. Identification of Ulcers and Erosions by the Novel Pillcam™ Crohn's Capsule Using a Convolutional Neural Network: A Multicentre Pilot Study. *J Crohns Colitis* 2022; **16**: 169-172 [PMID: 34228113 DOI: 10.1093/ecco-jcc/jjab117]

83 **Klang E**, Grinman A, Soffer S, Margalit Yehuda R, Barzilay O, Amitai MM, Konen E, Ben-Horin S, Eliakim R, Barash Y, Kopylov U. Automated Detection of Crohn's Disease Intestinal Strictures on Capsule Endoscopy Images Using Deep Neural Networks. *J Crohns Colitis* 2021; **15**: 749-756 [PMID: 33216853 DOI: 10.1093/ecco-jcc/jjaa234]

84 **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 [PMID: 26886971 DOI: 10.1109/TMI.2016.2527736]

85 **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 [PMID: 29470172 DOI: 10.1109/TIP.2018.2801119]

86 **Gan T**, Yang Y, Liu S, Zeng B, Yang J, Deng K, Wu J, Yang L. Automatic Detection of Small Intestinal Hookworms in Capsule Endoscopy Images Based on a Convolutional Neural Network. *Gastroenterol Res Pract* 2021; **2021**: 5682288 [PMID: 34868306 DOI: 10.1155/2021/5682288]

87 **Sainju S**, Bui FM, Wahid KA. Automated bleeding detection in capsule endoscopy videos using statistical features and region growing. *J Med Syst* 2014; **38**: 25 [PMID: 24696394 DOI: 10.1007/s10916-014-0025-1]

88 **Usman MA**, Satrya GB, Usman MR, Shin SY. Detection of small colon bleeding in wireless capsule endoscopy videos. *Comput Med Imaging Graph* 2016; **54**: 16-26 [PMID: 27793502 DOI: 10.1016/j.compmedimag.2016.09.005]

89 **Ghosh T**, Chakareski J. Deep Transfer Learning for Automated Intestinal Bleeding Detection in Capsule Endoscopy Imaging. *J Digit Imaging* 2021; **34**: 404-417 [PMID: 33728563 DOI: 10.1007/s10278-021-00428-3]

90 **Tsuboi A**, Oka S, Aoyama K, Saito H, Aoki T, Yamada A, Matsuda T, Fujishiro M, Ishihara S, Nakahori M, Koike K, Tanaka S, Tada T. Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images. *Dig Endosc* 2020; **32**: 382-390 [PMID: 31392767 DOI: 10.1111/den.13507]

91 **Ribeiro T**, Saraiva MM, Ferreira JPS, Cardoso H, Afonso J, Andrade P, Parente M, Jorge RN, Macedo G. Artificial intelligence and capsule endoscopy: automatic detection of vascular lesions using a convolutional neural network. *Ann Gastroenterol* 2021; **34**: 820-828 [PMID: 34815648 DOI: 10.20524/aog.2021.0653]

92 **Afonso J**, Saraiva MM, Ferreira JPS, Cardoso H, Ribeiro T, Andrade P, Parente M, Jorge RN, Macedo G. Automated detection of ulcers and erosions in capsule endoscopy images using a convolutional neural network. *Med Biol Eng Comput* 2022; **60**: 719-725 [PMID: 35038118 DOI: 10.1007/s11517-021-02486-9]

93 **Mascarenhas M**, Ribeiro T, Afonso J, Ferreira JPS, Cardoso H, Andrade P, Parente MPL, Jorge RN, Mascarenhas Saraiva M, Macedo G. Deep learning and colon capsule endoscopy: automatic detection of blood and colonic mucosal lesions using a convolutional neural network. *Endosc Int Open* 2022; **10**: E171-E177 [PMID: 35186665 DOI: 10.1055/a-1675-1941]

94 **Lewis BS**, Eisen GM, Friedman S. A pooled analysis to evaluate results of capsule endoscopy trials. *Endoscopy* 2005; **37**: 960-965 [PMID: 16189768 DOI: 10.1055/s-2005-870353]

95 **Li B**, Meng MQ. Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection. *IEEE Trans Inf Technol Biomed* 2012; **16**: 323-329 [PMID: 22287246 DOI: 10.1109/TITB.2012.2185807]

96 **Liu G**, Yan G, Kuang S, Wang Y. Detection of small bowel tumor based on multi-scale curvelet analysis and fractal technology in capsule endoscopy. *Comput Biol Med* 2016; **70**: 131-138 [PMID: 26829705 DOI: 10.1016/j.compbiomed.2016.01.021]

97 **Faghih Dinevari V**, Karimian Khosroshahi G, Zolfy Lighvan M. Singular Value Decomposition Based Features for Automatic Tumor Detection in Wireless Capsule Endoscopy Images. *Appl Bionics Biomech* 2016; **2016**: 3678913 [PMID: 27478364 DOI: 10.1155/2016/3678913]

98 **Shanmuga Sundaram P**, Santhiyakumari N. An Enhancement of Computer Aided Approach for Colon Cancer Detection in WCE Images Using ROI Based Color Histogram and SVM2. *J Med Syst* 2019; **43**: 29 [PMID: 30612188 DOI: 10.1007/s10916-018-1153-9]

99 **Blanes-Vidal V**, Baatrup G, Nadimi ES. Addressing priority challenges in the detection and assessment of colorectal polyps from capsule endoscopy and colonoscopy in colorectal cancer screening using machine learning. *Acta Oncol* 2019; **58**: S29-S36 [PMID: 30836800 DOI: 10.1080/0284186X.2019.1584404]

100 **Saraiva MM**, Ferreira JPS, Cardoso H, Afonso J, Ribeiro T, Andrade P, Parente MPL, Jorge RN, Macedo G. Artificial intelligence and colon capsule endoscopy: development of an automated diagnostic system of protruding lesions in colon capsule endoscopy. *Tech Coloproctol* 2021; **25**: 1243-1248 [PMID: 34499277 DOI: 10.1007/s10151-021-02517-5]

101 **Mascarenhas M**, Afonso J, Ribeiro T, Cardoso H, Andrade P, Ferreira JPS, Saraiva MM, Macedo G. Performance of a Deep Learning System for Automatic Diagnosis of Protruding Lesions in Colon Capsule Endoscopy. *Diagnostics (Basel)* 2022; **12** [PMID: 35741255 DOI: 10.3390/diagnostics12061445]

102 **Constantinescu AF**, Ionescu M, Iovănescu VF, Ciurea ME, Ionescu AG, Streba CT, Bunescu MG, Rogoveanu I, Vere CC. A computer-aided diagnostic system for intestinal polyps identified by wireless capsule endoscopy. *Rom J Morphol Embryol* 2016; **57**: 979-984 [PMID: 28002513]

103 **Xia J**, Xia T, Pan J, Gao F, Wang S, Qian YY, Wang H, Zhao J, Jiang X, Zou WB, Wang YC, Zhou W, Li ZS, Liao Z. Use of artificial intelligence for detection of gastric lesions by magnetically controlled capsule endoscopy. *Gastrointest Endosc* 2021; **93**: 133-139.e4 [PMID: 32470426 DOI: 10.1016/j.gie.2020.05.027]

104 **Yuan Y**, Meng MQ. Deep learning for polyp recognition in wireless capsule endoscopy images. *Med Phys* 2017; **44**: 1379-1389 [PMID: 28160514 DOI: 10.1002/mp.12147]

105 **Sharma P**, Burke CA, Johnson DA, Cash BD. The importance of colonoscopy bowel preparation for the detection of colorectal lesions and colorectal cancer prevention. *Endosc Int Open* 2020; **8**: E673-E683 [PMID: 32355887 DOI: 10.1055/a-1127-3144]

106 **Ahmad OF**. Deep learning for automated bowel preparation assessment during colonoscopy: time to embrace a new approach? *Lancet Digit Health* 2021; **3**: e685-e686 [PMID: 34538737 DOI: 10.1016/S2589-7500(21)00143-6]

107 **Hassan C**, East J, Radaelli F, Spada C, Benamouzig R, Bisschops R, Bretthauer M, Dekker E, Dinis-Ribeiro M, Ferlitsch M, Fuccio L, Awadie H, Gralnek I, Jover R, Kaminski MF, Pellisé M, Triantafyllou K, Vanella G, Mangas-Sanjuan C, Frazzoni L, Van Hooft JE, Dumonceau JM. Bowel preparation for colonoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2019. *Endoscopy* 2019; **51**: 775-794 [PMID: 31295746 DOI: 10.1055/a-0959-0505]

108 **Rex DK**, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, Levin TR, Lieberman D, Robertson DJ. Colorectal Cancer Screening: Recommendations for Physicians and Patients from the U.S. Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol* 2017; **112**: 1016-1030 [PMID: 28555630 DOI: 10.1038/ajg.2017.174]

109 **ASGE Standards of Practice Committee**, Saltzman JR, Cash BD, Pasha SF, Early DS, Muthusamy VR, Khashab MA, Chathadi KV, Fanelli RD, Chandrasekhara V, Lightdale JR, Fonkalsrud L, Shergill AK, Hwang JH, Decker GA, Jue TL, Sharaf R, Fisher DA, Evans JA, Foley K, Shaukat A, Eloubeidi MA, Faulx AL, Wang A, Acosta RD. Bowel preparation before colonoscopy. *Gastrointest Endosc* 2015; **81**: 781-794 [PMID: 25595062 DOI: 10.1016/j.gie.2014.09.048]

110 **Hadlock SD**, Liu N, Bernstein M, Gould M, Rabeneck L, Ruco A, Sutradhar R, Tinmouth JM. The Quality of Colonoscopy Reporting in Usual Practice: Are Endoscopists Reporting Key Data Elements? *Can J Gastroenterol Hepatol* 2016; **2016**: 1929361 [PMID: 27579299 DOI: 10.1155/2016/1929361]

111 **Gavin DR**, Valori RM, Anderson JT, Donnelly MT, Williams JG, Swarbrick ET. The national colonoscopy audit: a nationwide assessment of the quality and safety of colonoscopy in the UK. *Gut* 2013; **62**: 242-249 [PMID: 22661458 DOI: 10.1136/gutjnl-2011-301848]

112 **Coe SG**, Panjala C, Heckman MG, Patel M, Qumseya BJ, Wang YR, Dalton B, Tran P, Palmer W, Diehl N, Wallace MB, Raimondo M. Quality in colonoscopy reporting: an assessment of compliance and performance improvement. *Dig Liver Dis* 2012; **44**: 660-664 [PMID: 22579446 DOI: 10.1016/j.dld.2012.03.022]

113 **Zhou J**, Wu L, Wan X, Shen L, Liu J, Zhang J, Jiang X, Wang Z, Yu S, Kang J, Li M, Hu S, Hu X, Gong D, Chen D, Yao L, Zhu Y, Yu H. A novel artificial intelligence system for the assessment of bowel preparation (with video). *Gastrointest Endosc* 2020; **91**: 428-435.e2 [PMID: 31783029 DOI: 10.1016/j.gie.2019.11.026]

114 **Zhou W**, Yao L, Wu H, Zheng B, Hu S, Zhang L, Li X, He C, Wang Z, Li Y, Huang C, Guo M, Zhang X, Zhu Q, Wu L, Deng Y, Zhang J, Tan W, Li C, Zhang C, Gong R, Du H, Zhou J, Sharma P, Yu H. Multi-step validation of a deep learning-based system for the quantification of bowel preparation: a prospective, observational study. *Lancet Digit Health* 2021; **3**: e697-e706 [PMID: 34538736 DOI: 10.1016/S2589-7500(21)00109-6]

115 **Lai EJ**, Calderwood AH, Doros G, Fix OK, Jacobson BC. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. *Gastrointest Endosc* 2009; **69**: 620-625 [PMID: 19136102 DOI: 10.1016/j.gie.2008.05.057]

116 **Lee JY**, Calderwood AH, Karnes W, Requa J, Jacobson BC, Wallace MB. Artificial intelligence for the assessment of bowel preparation. *Gastrointest Endosc* 2022; **95**: 512-518.e1 [PMID: 34896100 DOI: 10.1016/j.gie.2021.11.041]

117 **Low DJ**, Hong Z, Jugnundan S, Mukherjee A, Grover SC. Automated Detection of Bowel Preparation Scoring and Adequacy With Deep Convolutional Neural Networks. *JCAG* 2022; **XX**: 1-5 [DOI: 10.1093/jcag/gwac013]

118 **Wang YP**, Jheng YC, Sung KY, Lin HE, Hsin IF, Chen PH, Chu YC, Lu D, Wang YJ, Hou MC, Lee FY, Lu CL. Use of U-Net Convolutional Neural Networks for Automated Segmentation of Fecal Material for Objective Evaluation of Bowel Preparation Quality in Colonoscopy. *Diagnostics (Basel)* 2022; **12** [PMID: 35328166 DOI: 10.3390/diagnostics12030613]

119 **Kohli A**, Holzwanger EA, Levy AN. Emerging use of artificial intelligence in inflammatory bowel disease. *World J Gastroenterol* 2020; **26**: 6923-6928 [PMID: 33311940 DOI: 10.3748/wjg.v26.i44.6923]

120 **Cohen-Mekelburg S**, Berry S, Stidham RW, Zhu J, Waljee AK. Clinical applications of artificial intelligence and

machine learning-based methods in inflammatory bowel disease. *J Gastroenterol Hepatol* 2021; **36**: 279-285 [PMID: 33624888 DOI: 10.1111/jgh.15405]

121 **Ozawa T**, Ishihara S, Fujishiro M, Saito H, Kumagai Y, Shichijo S, Aoyama K, Tada T. Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis. *Gastrointest Endosc* 2019; **89**: 416-421.e1 [PMID: 30367878 DOI: 10.1016/j.gie.2018.10.020]

122 **Stidham RW**, Liu W, Bishu S, Rice MD, Higgins PDR, Zhu J, Nallamothu BK, Waljee AK. Performance of a Deep Learning Model vs Human Reviewers in Grading Endoscopic Disease Severity of Patients With Ulcerative Colitis. *JAMA Netw Open* 2019; **2**: e193963 [PMID: 31099869 DOI: 10.1001/jamanetworkopen.2019.3963]

123 **Dave M**, Loftus EV Jr. Mucosal healing in inflammatory bowel disease-a true paradigm of success? *Gastroenterol Hepatol (N Y)* 2012; **8**: 29-38 [PMID: 22347830]

124 **Huang TY**, Zhan SQ, Chen PJ, Yang CW, Lu HH. Accurate diagnosis of endoscopic mucosal healing in ulcerative colitis using deep learning and machine learning. *J Chin Med Assoc* 2021; **84**: 678-681 [PMID: 34050105 DOI: 10.1097/JCMA.0000000000000559]

125 **Barreiro-de Acosta M**, Vallejo N, de la Iglesia D, Uribarri L, Bastón I, Ferreiro-Iglesias R, Lorenzo A, Domínguez-Muñoz JE. Evaluation of the Risk of Relapse in Ulcerative Colitis According to the Degree of Mucosal Healing (Mayo 0 vs 1): A Longitudinal Cohort Study. *J Crohns Colitis* 2016; **10**: 13-19 [PMID: 26351390 DOI: 10.1093/ecco-jcc/jjv158]

126 **Osterman MT**, Scott FI, Fogt FF, Gilroy ED, Parrott S, Galanko J, Cross R, Moss A, Herfarth HH, Higgins PDR. Endoscopic and Histological Assessment, Correlation, and Relapse in Clinically Quiescent Ulcerative Colitis (MARQUEE). *Inflamm Bowel Dis* 2021; **27**: 207-214 [PMID: 32170946 DOI: 10.1093/ibd/izaa048]

127 **Bhambhvani HP**, Zamora A. Deep learning enabled classification of Mayo endoscopic subscore in patients with ulcerative colitis. *Eur J Gastroenterol Hepatol* 2021; **33**: 645-649 [PMID: 33079775 DOI: 10.1097/MEG.0000000000001952]

128 **Yao H**, Najarian K, Gryak J, Bishu S, Rice MD, Waljee AK, Wilkins HJ, Stidham RW. Fully automated endoscopic disease activity assessment in ulcerative colitis. *Gastrointest Endosc* 2021; **93**: 728-736.e1 [PMID: 32810479 DOI: 10.1016/j.gie.2020.08.011]

129 **Gottlieb K**, Requa J, Karnes W, Chandra Gudivada R, Shen J, Rael E, Arora V, Dao T, Ninh A, McGill J. Central Reading of Ulcerative Colitis Clinical Trial Videos Using Neural Networks. *Gastroenterology* 2021; **160**: 710-719.e2 [PMID: 33098883 DOI: 10.1053/j.gastro.2020.10.024]

130 **Gutierrez Becker B**, Arcadu F, Thalhammer A, Gamez Serna C, Feehan O, Drawnel F, Oh YS, Prunotto M. Training and deploying a deep learning model for endoscopic severity grading in ulcerative colitis using multicenter clinical trial data. *Ther Adv Gastrointest Endosc* 2021; **14**: 2631774521990623 [PMID: 33718871 DOI: 10.1177/2631774521990623]

131 **Maeda Y**, Kudo SE, Mori Y, Misawa M, Ogata N, Sasanuma S, Wakamura K, Oda M, Mori K, Ohtsuka K. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest Endosc* 2019; **89**: 408-415 [PMID: 30268542 DOI: 10.1016/j.gie.2018.09.024]

132 **Takenaka K**, Ohtsuka K, Fujii T, Negi M, Suzuki K, Shimizu H, Oshima S, Akiyama S, Motobayashi M, Nagahori M, Saito E, Matsuoka K, Watanabe M. Development and Validation of a Deep Neural Network for Accurate Evaluation of Endoscopic Images From Patients With Ulcerative Colitis. *Gastroenterology* 2020; **158**: 2150-2157 [PMID: 32060000 DOI: 10.1053/j.gastro.2020.02.012]

133 **Takenaka K**, Ohtsuka K, Fujii T, Oshima S, Okamoto R, Watanabe M. Deep Neural Network Accurately Predicts Prognosis of Ulcerative Colitis Using Endoscopic Images. *Gastroenterology* 2021; **160**: 2175-2177.e3 [PMID: 33485853 DOI: 10.1053/j.gastro.2021.01.210]

134 **Takenaka K**, Fujii T, Kawamoto A, Suzuki K, Shimizu H, Maeyashiki C, Yamaji O, Motobayashi M, Igarashi A, Hanazawa R, Hibiya S, Nagahori M, Saito E, Okamoto R, Ohtsuka K, Watanabe M. Deep neural network for video colonoscopy of ulcerative colitis: a cross-sectional study. *Lancet Gastroenterol Hepatol* 2022; **7**: 230-237 [PMID: 34856196 DOI: 10.1016/S2468-1253(21)00372-1]

135 **Udristoiu AL**, Stefanescu D, Gruionu G, Gruionu LG, Iacob AV, Karstensen JG, Vilman P, Saftoiu A. Deep Learning Algorithm for the Confirmation of Mucosal Healing in Crohn's Disease, Based on Confocal Laser Endomicroscopy Images. *J Gastrointestin Liver Dis* 2021; **30**: 59-65 [PMID: 33723558 DOI: 10.15403/jgld-3212]

136 **Xi Y**, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Transl Oncol* 2021; **14**: 101174 [PMID: 34243011 DOI: 10.1016/j.tranon.2021.101174]

137 **Roselló S**, Simón S, Cervantes A. Programmed colorectal cancer screening decreases incidence and mortality. *Transl Gastroenterol Hepatol* 2019; **4**: 84 [PMID: 32039289 DOI: 10.21037/tgh.2019.12.13]

138 **Ahn SB**, Han DS, Bae JH, Byun TJ, Kim JP, Eun CS. The Miss Rate for Colorectal Adenoma Determined by Quality-Adjusted, Back-to-Back Colonoscopies. *Gut Liver* 2012; **6**: 64-70 [PMID: 22375173 DOI: 10.5009/gnl.2012.6.1.64]

139 **Komeda Y**, Handa H, Watanabe T, Nomura T, Kitahashi M, Sakurai T, Okamoto A, Minami T, Kono M, Arizumi T, Takenaka M, Hagiwara S, Matsui S, Nishida N, Kashida H, Kudo M. Computer-Aided Diagnosis Based on Convolutional Neural Network System for Colorectal Polyp Classification: Preliminary Experience. *Oncology* 2017; **93** Suppl 1: 30-34 [PMID: 29258081 DOI: 10.1159/000481227]

140 **Misawa M**, Kudo SE, Mori Y, Cho T, Kataoka S, Yamauchi A, Ogawa Y, Maeda Y, Takeda K, Ichimasa K, Nakamura H, Yagawa Y, Toyoshima N, Ogata N, Kudo T, Hisayuki T, Hayashi T, Wakamura K, Baba T, Ishida F, Itoh H, Roth H, Oda M, Mori K. Artificial Intelligence-Assisted Polyp Detection for Colonoscopy: Initial Experience. *Gastroenterology* 2018; **154**: 2027-2029.e3 [PMID: 29653147 DOI: 10.1053/j.gastro.2018.04.003]

141 **Hassan C**, Wallace MB, Sharma P, Maselli R, Craviotto V, Spadaccini M, Repici A. New artificial intelligence system: first validation study versus experienced endoscopists for colorectal polyp detection. *Gut* 2020; **69**: 799-800 [PMID: 31615835 DOI: 10.1136/gutjnl-2019-319914]

142 **Guo Z**, Nemoto D, Zhu X, Li Q, Aizawa M, Utano K, Isohata N, Endo S, Kawarai Lefor A, Togashi K. Polyp detection algorithm can detect small polyps: Ex vivo reading test compared with endoscopists. *Dig Endosc* 2021; **33**: 162-169 [PMID: 32173917 DOI: 10.1111/den.13670]

143 **Neumann H**, Kreft A, Sivanathan V, Rahman F, Galle PR. Evaluation of novel LCI CAD EYE system for real time detection of colon polyps. *PLoS One* 2021; **16**: e0255955 [PMID: 34437563 DOI: 10.1371/journal.pone.0255955]

144 **Li JW**, Chia T, Fock KM, Chong KW, Wong YJ, Ang TL. Artificial intelligence and polyp detection in colonoscopy: Use of a single neural network to achieve rapid polyp localization for clinical use. *J Gastroenterol Hepatol* 2021; **36**: 3298-3307 [PMID: 34327729 DOI: 10.1111/jgh.15642]

145 **Ahmad OF**, González-Bueno Puyal J, Brandao P, Kader R, Abbasi F, Hussein M, Haidry RJ, Toth D, Mountney P, Seward E, Vega R, Stoyanov D, Lovat LB. Performance of artificial intelligence for detection of subtle and advanced colorectal neoplasia. *Dig Endosc* 2022; **34**: 862-869 [PMID: 34748665 DOI: 10.1111/den.14187]

146 **Hori K**, Ikematsu H, Yamamoto Y, Matsuzaki H, Takeshita N, Shinmura K, Yoda Y, Kiuchi T, Takemoto S, Yokota H, Yano T. Detecting colon polyps in endoscopic images using artificial intelligence constructed with automated collection of annotated images from an endoscopy reporting system. *Dig Endosc* 2022; **34**: 1021-1029 [PMID: 34748658 DOI: 10.1111/den.14185]

147 **Brand M**, Troya J, Krenzer A, Saßmannshausen Z, Zoller WG, Meining A, Lux TJ, Hann A. Development and evaluation of a deep learning model to improve the usability of polyp detection systems during interventions. *United European Gastroenterol J* 2022; **10**: 477-484 [PMID: 35511456 DOI: 10.1002/ueg2.12235]

148 **Lux TJ**, Banck M, Saßmannshausen Z, Troya J, Krenzer A, Fitting D, Sudarevic B, Zoller WG, Puppe F, Meining A, Hann A. Pilot study of a new freely available computer-aided polyp detection system in clinical practice. *Int J Colorectal Dis* 2022; **37**: 1349-1354 [PMID: 35543874 DOI: 10.1007/s00384-022-04178-8]

149 **Wang P**, Xiao X, Glissen Brown JR, Berzin TM, Tu M, Xiong F, Hu X, Liu P, Song Y, Zhang D, Yang X, Li L, He J, Yi X, Liu J, Liu X. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat Biomed Eng* 2018; **2**: 741-748 [PMID: 31015647 DOI: 10.1038/s41551-018-0301-3]

150 **Shen P**, Li WZ, Li JX, Pei ZC, Luo YX, Mu JB, Li W, Wang XM. Real-time use of a computer-aided system for polyp detection during colonoscopy, an ambispective study. *J Dig Dis* 2021; **22**: 256-262 [PMID: 33742774 DOI: 10.1111/1751-2980.12985]

151 **Livovsky DM**, Veikherman D, Golany T, Aides A, Dashinsky V, Rabani N, Ben Shimol D, Blau Y, Katzir L, Shimshoni I, Liu Y, Segol O, Goldin E, Corrado G, Lachter J, Matias Y, Rivlin E, Freedman D. Detection of elusive polyps using a large-scale artificial intelligence system (with videos). *Gastrointest Endosc* 2021; **94**: 1099-1109.e10 [PMID: 34216598 DOI: 10.1016/j.gie.2021.06.021]

152 **Pacal I**, Karaman A, Karaboga D, Akay B, Basturk A, Nalbantoglu U, Coskun S. An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets. *Comput Biol Med* 2022; **141**: 105031 [PMID: 34802713 DOI: 10.1016/j.compbiomed.2021.105031]

153 **Quan SY**, Wei MT, Lee J, Mohi-Ud-Din R, Mostaghim R, Sachdev R, Siegel D, Friedlander Y, Friedland S. Clinical evaluation of a real-time artificial intelligence-based polyp detection system: a US multi-center pilot study. *Sci Rep* 2022; **12**: 6598 [PMID: 35449442 DOI: 10.1038/s41598-022-10597-y]

154 **Horiuchi H**, Tamai N, Kamba S, Inomata H, Ohya TR, Sumiyama K. Real-time computer-aided diagnosis of diminutive rectosigmoid polyps using an auto-fluorescence imaging system and novel color intensity analysis software. *Scand J Gastroenterol* 2019; **54**: 800-805 [PMID: 31195905 DOI: 10.1080/00365521.2019.1627407]

155 **Becq A**, Chandnani M, Bharadwaj S, Baran B, Ernest-Suarez K, Gabr M, Glissen-Brown J, Sawhney M, Pleskow DK, Berzin TM. Effectiveness of a Deep-learning Polyp Detection System in Prospectively Collected Colonoscopy Videos With Variable Bowel Preparation Quality. *J Clin Gastroenterol* 2020; **54**: 554-557 [PMID: 31789758 DOI: 10.1097/MCG.0000000000001272]

156 **Wang P**, Liu P, Glissen Brown JR, Berzin TM, Zhou G, Lei S, Liu X, Li L, Xiao X. Lower Adenoma Miss Rate of Computer-Aided Detection-Assisted Colonoscopy vs Routine White-Light Colonoscopy in a Prospective Tandem Study. *Gastroenterology* 2020; **159**: 1252-1261.e5 [PMID: 32562721 DOI: 10.1053/j.gastro.2020.06.023]

157 **Shaukat A**, Colucci D, Erisson L, Phillips S, Ng J, Iglesias JE, Saltzman JR, Somers S, Brugge W. Improvement in adenoma detection using a novel artificial intelligence-aided polyp detection device. *Endosc Int Open* 2021; **9**: E263-E270 [PMID: 33553591 DOI: 10.1055/a-1321-1317]

158 **Pfeifer L**, Neufert C, Leppkes M, Waldner MJ, Häfner M, Beyer A, Hoffman A, Siersema PD, Neurath MF, Rath T. Computer-aided detection of colorectal polyps using a newly generated deep convolutional neural network: from development to first clinical experience. *Eur J Gastroenterol Hepatol* 2021; **33**: e662-e669 [PMID: 34034272 DOI: 10.1097/MEG.0000000000002209]

159 **Ishiyama M**, Kudo SE, Misawa M, Mori Y, Maeda Y, Ichimasa K, Kudo T, Hayashi T, Wakamura K, Miyachi H, Ishida F, Itoh H, Oda M, Mori K. Impact of the clinical use of artificial intelligence-assisted neoplasia detection for colonoscopy: a large-scale prospective, propensity score-matched study (with video). *Gastrointest Endosc* 2022; **95**: 155-163 [PMID: 34352255 DOI: 10.1016/j.gie.2021.07.022]

160 **Zippelius C**, Alqahtani SA, Schedel J, Brookman-Amissah D, Muehlenberg K, Federle C, Salzberger A, Schorr W, Pech O. Diagnostic accuracy of a novel artificial intelligence system for adenoma detection in daily practice: a prospective nonrandomized comparative study. *Endoscopy* 2022; **54**: 465-472 [PMID: 34293812 DOI: 10.1055/a-1556-5984]

161 **Wu L**, Zhang J, Zhou W, An P, Shen L, Liu J, Jiang X, Huang X, Mu G, Wan X, Lv X, Gao J, Cui N, Hu S, Chen Y, Hu X, Li J, Chen D, Gong D, He X, Ding Q, Zhu X, Li S, Wei X, Li X, Wang X, Zhou J, Zhang M, Yu HG. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. *Gut* 2019; **68**: 2161-2169 [PMID: 30858305 DOI: 10.1136/gutjnl-2018-317366]

162 **Shaukat A**, Lichtenstein DR, Somers SC, Chung DC, Perdue DG, Gopal M, Colucci DR, Phillips SA, Marka NA, Church TR, Brugge WR; SKOUT™ Registration Study Team. Computer-Aided Detection Improves Adenomas per Colonoscopy for Screening and Surveillance Colonoscopy: A Randomized Trial. *Gastroenterology* 2022; **163**: 732-741 [PMID: 35643173 DOI: 10.1053/j.gastro.2022.05.028]

163 **Glissen Brown JR**, Mansour NM, Wang P, Chuchuca MA, Minchenberg SB, Chandnani M, Liu L, Gross SA, Sengupta N, Berzin TM. Deep Learning Computer-aided Polyp Detection Reduces Adenoma Miss Rate: A United States Multi-center Randomized Tandem Colonoscopy Study (CADeT-CS Trial). *Clin Gastroenterol Hepatol* 2022; **20**: 1499-1507.e4

[PMID: 34530161 DOI: 10.1016/j.cgh.2021.09.009]

164 **Kamba S**, Tamai N, Saitoh I, Matsui H, Horiuchi H, Kobayashi M, Sakamoto T, Ego M, Fukuda A, Tonouchi A, Shimahara Y, Nishikawa M, Nishino H, Saito Y, Sumiyama K. Reducing adenoma miss rate of colonoscopy assisted by artificial intelligence: a multicenter randomized controlled trial. *J Gastroenterol* 2021; **56**: 746-757 [PMID: 34218329 DOI: 10.1007/s00535-021-01808-w]

165 **Repici A**, Spadaccini M, Antonelli G, Correale L, Maselli R, Galtieri PA, Pellegatta G, Capogreco A, Milluzzo SM, Lollo G, Di Paolo D, Badalamenti M, Ferrara E, Fugazza A, Carrara S, Anderloni A, Rondonotti E, Amato A, De Gottardi A, Spada C, Radaelli F, Savevski V, Wallace MB, Sharma P, Rösch T, Hassan C. Artificial intelligence and colonoscopy experience: lessons from two randomised trials. *Gut* 2022; **71**: 757-765 [PMID: 34187845 DOI: 10.1136/gutjnl-2021-324471]

166 **Gong D**, Wu L, Zhang J, Mu G, Shen L, Liu J, Wang Z, Zhou W, An P, Huang X, Jiang X, Li Y, Wan X, Hu S, Chen Y, Hu X, Xu Y, Zhu X, Li S, Yao L, He X, Chen D, Huang L, Wei X, Wang X, Yu H. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol Hepatol* 2020; **5**: 352-361 [PMID: 31981518 DOI: 10.1016/S2468-1253(19)30413-3]

167 **Wang P**, Liu X, Berzin TM, Glissen Brown JR, Liu P, Zhou C, Lei L, Li L, Guo Z, Lei S, Xiong F, Wang H, Song Y, Pan Y, Zhou G. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. *Lancet Gastroenterol Hepatol* 2020; **5**: 343-351 [PMID: 31981517 DOI: 10.1016/S2468-1253(19)30411-X]

168 **Wang P**, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, Liu P, Li L, Song Y, Zhang D, Li Y, Xu G, Tu M, Liu X. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2019; **68**: 1813-1819 [PMID: 30814121 DOI: 10.1136/gutjnl-2018-317500]

169 **Su JR**, Li Z, Shao XJ, Ji CR, Ji R, Zhou RC, Li GC, Liu GQ, He YS, Zuo XL, Li YQ. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). *Gastrointest Endosc* 2020; **91**: 415-424.e4 [PMID: 31454493 DOI: 10.1016/j.gie.2019.08.026]

170 **Repici A**, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, Ferrara E, Spadaccini M, Alkandari A, Fugazza A, Anderloni A, Galtieri PA, Pellegatta G, Carrara S, Di Leo M, Craviotto V, Lamonaca L, Lorenzetti R, Andrealli A, Antonelli G, Wallace M, Sharma P, Rosch T, Hassan C. Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology* 2020; **159**: 512-520.e7 [PMID: 32371116 DOI: 10.1053/j.gastro.2020.04.062]

171 **Liu P**, Wang P, Glissen Brown JR, Berzin TM, Zhou G, Liu W, Xiao X, Chen Z, Zhang Z, Zhou C, Lei L, Xiong F, Li L, Liu X. The single-monitor trial: an embedded CADe system increased adenoma detection during colonoscopy: a prospective randomized study. *Therap Adv Gastroenterol* 2020; **13**: 1756284820979165 [PMID: 33403003 DOI: 10.1177/1756284820979165]

172 **Xu L**, He X, Zhou J, Zhang J, Mao X, Ye G, Chen Q, Xu F, Sang J, Wang J, Ding Y, Li Y, Yu C. Artificial intelligence-assisted colonoscopy: A prospective, multicenter, randomized controlled trial of polyp detection. *Cancer Med* 2021; **10**: 7184-7193 [PMID: 34477306 DOI: 10.1002/cam4.4261]

173 **Liu WN**, Zhang YY, Bian XQ, Wang LJ, Yang Q, Zhang XD, Huang J. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. *Saudi J Gastroenterol* 2020; **26**: 13-19 [PMID: 31898644 DOI: 10.4103/sjg.SJG_377_19]

174 **Luo Y**, Zhang Y, Liu M, Lai Y, Liu P, Wang Z, Xing T, Huang Y, Li Y, Li A, Wang Y, Luo X, Liu S, Han Z. Artificial Intelligence-Assisted Colonoscopy for Detection of Colon Polyps: a Prospective, Randomized Cohort Study. *J Gastrointest Surg* 2021; **25**: 2011-2018 [PMID: 32968933 DOI: 10.1007/s11605-020-04802-4]

175 **Barua I**, Vinsard DG, Jodal HC, Løberg M, Kalager M, Holme Ø, Misawa M, Bretthauer M, Mori Y. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. *Endoscopy* 2021; **53**: 277-284 [PMID: 32557490 DOI: 10.1055/a-1201-7165]

176 **Aziz M**, Fatima R, Dong C, Lee-Smith W, Nawras A. The impact of deep convolutional neural network-based artificial intelligence on colonoscopy outcomes: A systematic review with meta-analysis. *J Gastroenterol Hepatol* 2020; **35**: 1676-1683 [PMID: 32267558 DOI: 10.1111/jgh.15070]

177 **Hassan C**, Spadaccini M, Iannone A, Maselli R, Jovani M, Chandrasekar VT, Antonelli G, Yu H, Areia M, Dinis-Ribeiro M, Bhandari P, Sharma P, Rex DK, Rösch T, Wallace M, Repici A. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointest Endosc* 2021; **93**: 77-85.e6 [PMID: 32598963 DOI: 10.1016/j.gie.2020.06.059]

178 **Li J**, Lu J, Yan J, Tan Y, Liu D. Artificial intelligence can increase the detection rate of colorectal polyps and adenomas: a systematic review and meta-analysis. *Eur J Gastroenterol Hepatol* 2021; **33**: 1041-1048 [PMID: 32804846 DOI: 10.1097/MEG.0000000000001906]

179 **Zhang Y**, Zhang X, Wu Q, Gu C, Wang Z. Artificial Intelligence-Aided Colonoscopy for Polyp Detection: A Systematic Review and Meta-Analysis of Randomized Clinical Trials. *J Laparoendosc Adv Surg Tech A* 2021; **31**: 1143-1149 [PMID: 33524298 DOI: 10.1089/lap.2020.0777]

180 **Deliwala SS**, Hamid K, Barbarawi M, Lakshman H, Zayed Y, Kandel P, Malladi S, Singh A, Bachuwa G, Gurvits GE, Chawla S. Artificial intelligence (AI) real-time detection vs. routine colonoscopy for colorectal neoplasia: a meta-analysis and trial sequential analysis. *Int J Colorectal Dis* 2021; **36**: 2291-2303 [PMID: 33934173 DOI: 10.1007/s00384-021-03929-3]

181 **Nazarian S**, Glover B, Ashrafian H, Darzi A, Teare J. Diagnostic Accuracy of Artificial Intelligence and Computer-Aided Diagnosis for the Detection and Characterization of Colorectal Polyps: Systematic Review and Meta-analysis. *J Med Internet Res* 2021; **23**: e27370 [PMID: 34259645 DOI: 10.2196/27370]

182 **Xu Y**, Ding W, Wang Y, Tan Y, Xi C, Ye N, Wu D, Xu X. Comparison of diagnostic performance between convolutional neural networks and human endoscopists for diagnosis of colorectal polyp: A systematic review and meta-analysis. *PLoS One* 2021; **16**: e0246892 [PMID: 33592048 DOI: 10.1371/journal.pone.0246892]

183 **Bang CS**, Lee JJ, Baik GH. Computer-Aided Diagnosis of Diminutive Colorectal Polyps in Endoscopic Images:

Systematic Review and Meta-analysis of Diagnostic Test Accuracy. *J Med Internet Res* 2021; **23**: e29682 [PMID: 34432643 DOI: 10.2196/29682]

184 **Yoon D**, Kong HJ, Kim BS, Cho WS, Lee JC, Cho M, Lim MH, Yang SY, Lim SH, Lee J, Song JH, Chung GE, Choi JM, Kang HY, Bae JH, Kim S. Colonoscopic image synthesis with generative adversarial network for enhanced detection of sessile serrated lesions using convolutional neural network. *Sci Rep* 2022; **12**: 261 [PMID: 34997124 DOI: 10.1038/s41598-021-04247-y]

185 **Nemoto D**, Guo Z, Peng B, Zhang R, Nakajima Y, Hayashi Y, Yamashina T, Aizawa M, Utano K, Lefor AK, Zhu X, Togashi K. Computer-aided diagnosis of serrated colorectal lesions using non-magnified white-light endoscopic images. *Int J Colorectal Dis* 2022; **37**: 1875-1884 [PMID: 35861862 DOI: 10.1007/s00384-022-04210-x]

186 **Yao L**, Zhang J, Liu J, Zhu L, Ding X, Chen D, Wu H, Lu Z, Zhou W, Zhang L, Xu B, Hu S, Zheng B, Yang Y, Yu H. A deep learning-based system for bile duct annotation and station recognition in linear endoscopic ultrasound. *EBioMedicine* 2021; **65**: 103238 [PMID: 33639404 DOI: 10.1016/j.ebiom.2021.103238]

187 **Săftoiu A**, Vilmann P, Gorunescu F, Janssen J, Hocke M, Larsen M, Iglesias-Garcia J, Arcidiacono P, Will U, Giovannini M, Dietrich CF, Havre R, Gheorghe C, McKay C, Gheonea DI, Ciurea T; European EUS Elastography Multicentric Study Group. Efficacy of an artificial neural network-based approach to endoscopic ultrasound elastography in diagnosis of focal pancreatic masses. *Clin Gastroenterol Hepatol* 2012; **10**: 84-90.e1 [PMID: 21963957 DOI: 10.1016/j.cgh.2011.09.014]

188 **Kuwahara T**, Hara K, Mizuno N, Okuno N, Matsumoto S, Obata M, Kurita Y, Koda H, Toriyama K, Onishi S, Ishihara M, Tanaka T, Tajika M, Niwa Y. Usefulness of Deep Learning Analysis for the Diagnosis of Malignancy in Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Clin Transl Gastroenterol* 2019; **10**: 1-8 [PMID: 31117111 DOI: 10.14309/ctg.0000000000000045]

189 **Goyal H**, Sherazi SAA, Gupta S, Perisetti A, Achebe I, Ali A, Tharian B, Thosani N, Sharma NR. Application of artificial intelligence in diagnosis of pancreatic malignancies by endoscopic ultrasound: a systemic review. *Therap Adv Gastroenterol* 2022; **15**: 17562848221093873 [PMID: 35509425 DOI: 10.1177/17562848221093873]

190 **Dumitrescu EA**, Ungureanu BS, Cazacu IM, Florescu LM, Streba L, Croitoru VM, Sur D, Croitoru A, Turcu-Stiolica A, Lungulescu CV. Diagnostic Value of Artificial Intelligence-Assisted Endoscopic Ultrasound for Pancreatic Cancer: A Systematic Review and Meta-Analysis. *Diagnostics (Basel)* 2022; **12** [PMID: 35204400 DOI: 10.3390/diagnostics12020309]

191 **Minoda Y**, Ihara E, Komori K, Ogino H, Otsuka Y, Chinen T, Tsuda Y, Ando K, Yamamoto H, Ogawa Y. Efficacy of endoscopic ultrasound with artificial intelligence for the diagnosis of gastrointestinal stromal tumors. *J Gastroenterol* 2020; **55**: 1119-1126 [PMID: 32918102 DOI: 10.1007/s00535-020-01725-4]

192 **Jang SI**, Kim YJ, Kim EJ, Kang H, Shon SJ, Seol YJ, Lee DK, Kim KG, Cho JH. Diagnostic performance of endoscopic ultrasound-artificial intelligence using deep learning analysis of gallbladder polypoid lesions. *J Gastroenterol Hepatol* 2021; **36**: 3548-3555 [PMID: 34431545 DOI: 10.1111/jgh.15673]

193 **Zhang B**, Zhu F, Li P, Zhu J. Artificial intelligence-assisted endoscopic ultrasound in the diagnosis of gastrointestinal stromal tumors: a meta-analysis. *Surg Endosc* 2022 [PMID: 36100781 DOI: 10.1007/s00464-022-09597-w]

194 **Ye XH**, Zhao LL, Wang L. Diagnostic accuracy of endoscopic ultrasound with artificial intelligence for gastrointestinal stromal tumors: A meta-analysis. *J Dig Dis* 2022; **23**: 253-261 [PMID: 35793389 DOI: 10.1111/1751-2980.13110]

195 **Marya NB**, Powers PD, Petersen BT, Law R, Storm A, Abusaleh RR, Rau P, Stead C, Levy MJ, Martin J, Vargas EJ, Abu Dayyeh BK, Chandrasekhara V. Identification of patients with malignant biliary strictures using a cholangioscopy-based deep learning artificial intelligence (with video). *Gastrointest Endosc* 2022 [PMID: 36007584 DOI: 10.1016/j.gie.2022.08.021]

196 **Huang L**, Lu X, Huang X, Zou X, Wu L, Zhou Z, Wu D, Tang D, Chen D, Wan X, Zhu Z, Deng T, Shen L, Liu J, Zhu Y, Gong D, Zhong Y, Liu F, Yu H. Intelligent difficulty scoring and assistance system for endoscopic extraction of common bile duct stones based on deep learning: multicenter study. *Endoscopy* 2021; **53**: 491-498 [PMID: 32838430 DOI: 10.1055/a-1244-5698]

197 **Uche-Anya E**, Anyane-Yeboa A, Berzin TM, Ghassemi M, May FP. Artificial intelligence in gastroenterology and hepatology: how to advance clinical practice while ensuring health equity. *Gut* 2022; **71**: 1909-1915 [PMID: 35688612 DOI: 10.1136/gutjnl-2021-326271]

198 **Stidham RW**. Artificial Intelligence for Understanding Imaging, Text, and Data in Gastroenterology. *Gastroenterol Hepatol (N Y)* 2020; **16**: 341-349 [PMID: 34035738]

199 **Spadaccini M**, Marco A, Franchellucci G, Sharma P, Hassan C, Repici A. Discovering the first US FDA-approved computer-aided polyp detection system. *Future Oncol* 2022; **18**: 1405-1412 [PMID: 35081745 DOI: 10.2217/fon-2021-1135]

Artificial Intelligence in Gastroenterology

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Gastroenterol 2022 December 28; 3(5): 142-162

DOI: 10.35712/aig.v3.i5.142

ISSN 2644-3236 (online)

REVIEW

Artificial intelligence applications in predicting the behavior of gastrointestinal cancers in pathology

Aysen Yavuz, Anil Alpsoy, Elif Ocak Gedik, Mennan Yigitcan Celik, Cumhur Ibrahim Bassorgun, Betul Unal, Gulsum Ozlem Elpek

Specialty type: Pathology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Tsoulfas G, Greece; Wan XH, China

Received: October 16, 2022
Peer-review started: October 16, 2022
First decision: November 15, 2022
Revised: November 25, 2022
Accepted: December 13, 2022
Article in press: December 13, 2022
Published online: December 28, 2022

Aysen Yavuz, Anil Alpsoy, Elif Ocak Gedik, Mennan Yigitcan Celik, Cumhur Ibrahim Bassorgun, Betul Unal, Gulsum Ozlem Elpek, Department of Pathology, Akdeniz University Medical School, Antalya 07070, Turkey

Corresponding author: Gulsum Ozlem Elpek, MD, Professor, Department of Pathology, Akdeniz University Medical School, Dumlupinar bulvarı, Antalya 07070, Turkey. elpek@akdeniz.edu.tr

Abstract

Recent research has provided a wealth of data supporting the application of artificial intelligence (AI)-based applications in routine pathology practice. Indeed, it is clear that these methods can significantly support an accurate and rapid diagnosis by eliminating errors, increasing reliability, and improving workflow. In addition, the effectiveness of AI in the pathological evaluation of prognostic parameters associated with behavior, course, and treatment in many types of tumors has also been noted. Regarding gastrointestinal system (GIS) cancers, the contribution of AI methods to pathological diagnosis has been investigated in many studies. On the other hand, studies focusing on AI applications in evaluating parameters to determine tumor behavior are relatively few. For this purpose, the potential of AI models has been studied over a broad spectrum, from tumor subtyping to the identification of new digital biomarkers. The capacity of AI to infer genetic alterations of cancer tissues from digital slides has been demonstrated. Although current data suggest the merit of AI-based approaches in assessing tumor behavior in GIS cancers, a wide range of challenges still need to be solved, from laboratory infrastructure to improving the robustness of algorithms, before incorporating AI applications into real-life GIS pathology practice. This review aims to present data from AI applications in evaluating pathological parameters related to the behavior of GIS cancer with an overview of the opportunities and challenges encountered in implementing AI in pathology.

Key Words: Digital pathology; Colorectal cancer; Gastric cancer; Machine learning; Deep learning; Prognosis

Core Tip: This review outlines the potential of artificial intelligence applications for evaluating pathological parameters related to the behavior of gastrointestinal cancers. The role of these methods in determining the behavior of esophageal cancers remains to be investigated. On the other hand, the results are promising, supporting that these models can assist in the determination of conventional pathological parameters and perform molecular subtyping in gastric and colorectal cancers. Furthermore, these applications encourage digital prognostic biomarker discovery by revealing predictions that are impossible when using traditional visual methods. However, further studies are needed to overcome the obstacles to implementing these applications into pathology practice.

Citation: Yavuz A, Alpsoy A, Gedik EO, Celik MY, Bassorgun CI, Unal B, Elpek GO. Artificial intelligence applications in predicting the behavior of gastrointestinal cancers in pathology. *Artif Intell Gastroenterol* 2022; 3(5): 142-162
URL: https://www.wjgnet.com/2644-3236/full/v3/i5/142.htm
DOI: https://dx.doi.org/10.35712/aig.v3.i5.142

INTRODUCTION

Gastrointestinal (GIS) cancers, including tumors of the esophagus, stomach, colon, and rectum, are an important health problem worldwide. Although the incidence of esophageal cancer (EC) is relatively low, gastric cancer (GC) and colorectal cancer (CRC) are among the most common types of cancer (fifth and third, respectively)[1]. They are also responsible for a substantial proportion of cancer mortality, with GC being the third and CRC the second most common cause of cancer-related death[2]. Although various predictive and prognostic parameters are currently available, the mortality rates for patients with GIS cancer are, unfortunately, still very high[2]. It has been shown that rectifying this situation may depend on paving the way for more personalized treatment strategies that lead to a better prognosis and/or fewer treatment side effects[3,4]. Therefore, the meticulous and complete evaluation of patients to determine the appropriate treatment is critical.

In this context, in addition to providing a definitive diagnosis, the role of an accurate evaluation of pathological parameters related to the behavior and proper treatment of GIS tumors cannot be ignored. However, pathology, a morphology-based specialty, is susceptible to subjectivity regarding intraobserver and interobserver variations, particularly in oncology. That is why, in recent years, the search for more objective criteria to eliminate bias, as well as to reduce the growing workload and to contribute time-saving, has allowed the improvement of image analysis-based digital pathology (DP), which has an important place in modern pathological applications[5,6].

In particular, significant advances in slide scanner technology, which can rapidly digitize all pathological slides at high resolution whole slide images (WSIs), has enabled not only the analysis of a wide range of morphological parameters but also the detection of biomarkers/genetic changes in many types of tumors[7-9]. The ability of computer-based analysis to detect prognostic and predictive markers from these images, depending on the fact that they are composed of number matrices containing a large amount of information that is not accessible to the human eye, has led to the adoption of artificial intelligence (AI) for DP[10,11]. Accordingly, the number of studies on AI applications associated with the diagnosis, follow-up, and treatment of many tumors has increased significantly over time. Regarding GIS, data from previous studies evaluating pathological prognostic parameters with various AI models suggest that using these methods may be beneficial. Unfortunately, these encouraging results have not overcome the wide range of challenges to be solved, from laboratory infrastructure to improving the robustness of algorithms, before incorporating AI applications into real-life pathology practice.

This review presented the applications of AI in the evaluation of pathological parameters related to the behavior of GIS cancer, along with a brief overview of the opportunities and challenges encountered in its implementation in pathology.

GENERAL VIEW OF AI IN PATHOLOGY LABORATORIES

In parallel with technological developments, the evolution of whole slide imaging (WSI) has provided remote diagnosis, consultation, and education[12-14]. In the recent past, it was suggested that the use of WSI is comparable to, or even better than, conventional microscopic examination for decision-making in pathology[15-17]. On the other hand, WSIs are also crucial in applying AI methods in pathological practice. They not only provide quick access to the archive without loss of image quality, but they can also render gigabit images, which are very difficult to process, suitable for processing by "tessellation"

[18]. This preprocessing is based on cutting a large image into nonoverlapping smaller patches called "tiles," making them amenable to computational analysis. It should be noted that although some pathological studies use selected images captured manually with a camera, WSI is currently recommended as a standard for AI applications, especially in tumors where heterogeneity is frequent, such as those of the GIS[19].

To achieve reliable results with WSIs, many steps, from preserving the structure of the tissue to the preparation of sections, must be carried out with care in the pathological laboratory. In particular, it is imperative to evaluate and check slides for artifacts (tears, floating contamination, thickness) that have the potential to adversely affect digitization and, thus, AI applications[20,21]. However, it should be noted that even with optimal protocols and slide scanner standardization, the importance of color normalization to ensure consistency in WSI databases should not be overlooked, as it can affect the robustness of deep learning (DL) models. Accordingly, histogram-matching color transfer and spectral matching methods can be applied[22-24]. However, as these methods depend on the expertise of pathologists and are impractical for manual adjustment, various algorithms have been proposed by researchers capable of performing this normalization. Although promising results have been obtained, there is a need for future studies on the performance of AI models using color normalization systems[25, 26].

The gradual evolution of traditional pathology into DP has led to the development of powerful and user-friendly WSI analysis software tools with the ability to manage substantial WSIs and metadata from different hardware manufacturers, as well as interactive drawing annotation capabilities to facilitate decision-making and reporting. Moreover, a significant proportion of them is freely available [27-29]. In addition, the high costs of hardware required for high-performance computation in software development have become more affordable, leading to the implementation of DP in major medical centers[16,30-32]. Increasing the number of centers capable of using DP will allow for the generation of large and high-quality WSI databases, enabling the acquisition of large datasets and the design of algorithms for AI. However, the requirement of a significant investment is still an obstacle to overcome for the widespread application of these technologies[33]. In addition, the problem of proprietary datasets persists, limiting the repeatability of the proposed methodologies and hindering advancement in this field.

As mentioned above, the ability of AI to extract meaningful information from images that the naked human eye cannot discriminate makes it an attractive tool in the field of image processing and analysis in pathology. Therefore, contemporary AI models have evolved from expert systems to different types, such as machine learning (ML) and DL (Table 1). In brief, ML is a subtype of AI that provides a computer system to automatically learn and develop from datasets on its own and solve problems without explicit programming[34-36]. DL is a subfield of ML that employs sophisticated algorithmic structures inspired by the neural network of the human brain (artificial neural network, ANN) in which statistical models are established from input training data[37-39]. Therefore, DL requires large, annotated datasets to develop its algorithms. At present, the annotation of datasets is a complex task in model development[9,40]. In practice, the time-consuming and challenging nature of annotation, especially in systems where heterogeneous lesions are common, such as GIS, may affect the accuracy of the model being trained[41]. Another limitation is that the dataset obtained by a study group does not show the same performance when compared to external validation sets from other institutions. Recently, studies have been conducted to overcome the hindering properties of annotation[42-44]. It has also been suggested that the adoption of DP for diagnosis could indirectly facilitate the generation of valuable datasets for future algorithm development by enabling pathologists to describe areas of interest during evaluation and reporting[45].

It has often been emphasized that the validation of AI-based technologies requires an evidence-based approach[42,46]. This should also be considered in a laboratory-based medical specialty such as pathology. On the other hand, analyzing the performance of AI techniques to that of pathologists is a significant challenge regarding interobserver and interobserver heterogeneity. Currently, the problems related to establishing "ground truth" in AI methods should not be overlooked[40,47]. It should be noted that this requires repeated testing of the effectiveness and consistency of AI applications in many different patient populations. The relative lack of a validation cohort in developing AI-powered DP applications is also related to the possible drawbacks of sharing histopathological slides. Despite interobserver heterogeneity and variability in pathological assessment also demonstrating the uncertainty of "ground truth" in this regard, multi center assessments involving multiple pathologists and datasets may be the best way to overcome this obstacle.

Before the integration of AI into the pathology workflow, the need to validate its benefits and address ethical recommendations increases the importance of AI-based tools being transparent and interpretable, resulting in an increasing demand for more explainable AI models. In this respect, there is a dilemma about the application of AI. Because most algorithms developed use DL, ensemble methods called "black box" models to tackle multidimensional problems are very complex. However, more straightforward methods that are not complex are not powerful enough to achieve the expected results [48]. For this reason, model interpretability, ethical concerns, and potential regulatory barriers should also be considered in newly developed AI tools to meet these expectations.

Table 1 General features of machine learning methods in the development of artificial intelligence models in gastrointestinal pathology

AI models	Strengths	Weaknesses
ML, Traditional, Supervised	Data output can be produced from the previously labeled training set	Labeling big data takes a considerable amount of time and can be challenging
	Allows users to reflect domain knowledge features	Feature extraction quality significantly affects the accuracy
ML, Traditional, Supervised	Users do not supervise the model or label any data	Input data is unknown and not labeled
	Patterns are detected automatically	Precise information related to data sorting is not provided
	Save time	Interpretation is challenging
SVM	Suitable for more efficient regression and classification analysis with high-dimensional data	Not suitable for large data sets. Requires more time for training; Low performance in overlapping classes
CNN	No labeling is required for important information and features	Lack of interpretability due to black boxes
	The performance capacity in image recognition is high	
FCN	Provides computational speed	A large amount of labeled data for training is required
	The background noise is automatically eliminated	The labeling cost is high
RNN	Able to decide which information to remember from past experiences	The model is hard to train
	A suitable deep learning model for sequential data	The computational cost is high
MIL	A detailed annotation is not required	A large amount of training data is required
	Suitable to be performed on large datasets	The computational cost is high
GAN	The potential to produce new realistic data that resembles the original data	The model is hard to train

AI: Artificial intelligence; ML: Machine learning; SVM: Support Vector Machine; CNN: Convolutional neural networks; FCN: Fully convolutional neural networks; RNN: Recurrent neural networks; MIL: Multi-instance learning; GAN: Generative adversarial networks.

AI IN THE PATHOLOGICAL DETERMINATION OF PRENEOPLASTIC LESIONS IN GIS

Barrett's esophagus

The majority of AI studies in EC consist of imaging studies. In pathology, there have been recent studies on the diagnosis of Barrett's esophagus (BE) and the evaluation of dysplasia in these lesions to predict the risk of EC[49,50]. A proposed attention-based deep NN framework for detecting BE and adenocarcinoma (ADC) was found to be reliable with a mean accuracy of 0.83[49]. Unlike existing methods based on the region of interest, this model is based on tissue-level annotations, suggesting that it may provide a new approach for applying DL in pathology. On the other hand, the fact that the study was performed in a single center and on a relatively small data set necessitates the development of the proposed model with further studies. Since trefoil factor 3 expression is the key finding of BE, a DL model (VGG16) using immunohistochemically stained sections showed significant adaptability, with an area under the curve (AUC) of 0.88[50]. Although the proposed approach reduced the pathologist workload by 57%, the underlying ML model still needs further optimization.

Colorectal polyp classification

In CRC, unlike GC, the classification of polyps is an important task to determine the risk of CRC and the future surveillance needs of patients[51]. In routine examinations, high-risk polyps are evaluated based on their histopathological features with considerable interobserver variability among pathologists[52,53]. However, a precise diagnosis of high-risk polyps is required for efficient and early detection of cancer. In addition, the recommendation for endoscopic screening of these lesions for an early diagnosis of CRC, especially in elderly individuals, increases the workload of daily pathology practice[54].

Therefore, AI applications have been developed to classify high-risk colorectal polyps and/or adenomas with high-grade dysplasia. In studies on the classification of these lesions and the identification of CRC, datasets of three to six specific categories and five models were used[55-62] (Table 2). Although most studies showed good performance with generally high AUCs and accuracies, because of the following restrictions, the evidence level of each model needed to be improved. The number of patches and WSIs that make up the datasets are different. Accordingly, in some studies, the number of datasets may affect the reliability of the results. In various studies, the annotation process is not delineated in detail. In addition, the fact that each model has a different focus and characteristics makes

Table 2 AI-based applications in pathology for the determination of tumor behavior in colorectal carcinomas

Ref.	Task	Data sets	Algorithm/Model	Performance	Comments
Xu *et al*[55]	NL/ADC/MC/SC/PC/CCTA	717 patches	AlexNet	Accuracy: 97%	The model provides the classifications of tumor subtypes
Korbar *et al*[56]	NL/HP/SSP/TSA/TA/TVA-VA	Training set: 458 WSIs; Test set: 239 WSIs	ResNET	F1 Score: 88.8%; Accuracy: 93%; Precision: 89.7%; Recall: 88.3%	The model may reduce the workload of pathologists in the assessment of colorectal polyps
Haj-Hassan *et al*[57]	NL/AD/ADC	30 patients, Multispectral image patches	CNN	Accuracy: 99.2%	CNN allows the classification of CRC tissue types using pre-segmented regions of interest
Ponzio *et al*[58]	NL/AD/ADC	27 WSIs	VGG16	Accuracy: 96%	TL considerably outperforms the CNN fully trained on CRC samples on the same test dataset
Sena *et al*[59]	NL/HP/AD/ADC	393 images	CNN	Accuracy: 80%	DL may provide a valuable tool to assist pathologists in the histological classification of CR tumors
Iizuka *et al*[60]	NL/AD/ADC	4036 WSIs + 500WSIs	CNN/RNN	AUCs: 0.96-0.99	Integrating DL models in pathology workflow would be of high benefit for easing the workload of pathologists
Wei *et al*[61]	NL//TA/TVA/VA/HP	1182 WSIs	ResNet	Accuracy: 93.5% (Internal test set); Accuracy: 87% (External test set)	This model may assist pathologists by improving the accuracy of CRC screening
Awan *et al*[62]	NL/Low GR/High GR	139 images	CNN	Accuracy: 97% (two-class), 91% (three-class)	The model provides the classifications of tumor subtypes based on the shape of glands
Sirinukunwattana *et al*[97]	Prediction of MSTs	510 WSIs (FOCUS), 431 WSIs (TCGA), 265 WSIs (GRAMPIAN cohort)	Inception V3	AUCs: 0.9 (FOCUS); 0.94 (TCGA), 0.85 (GRAMPIAN cohort)	RNA expression classifiers can predict from H-E stained images, opening the door to cheap and reliable biological stratification within routine workflows
Echle *et al*[98]	MSI *vs* MSS	6406 WSIs (Training); 771 WSIs (External validation)	ShuffleNet	AUC: 0.92 (Training); AUC: 0.96 (External validation)	The model provides a low-cost evaluation of MSI without molecular testing
Kather *et al*[80]	MSI *vs* MSS	60894 patches (TCGA-CRC-KR); 93408 patches (TCGA-CRC-DX)	ResNet18	AUC: 0.84 (TCGA-CRC-KR); AUC: 0.77 (TCGA-CRC-DX)	This method may lead to improvements in molecular subtype screening workload in pathology
Kather *et al*[77]	Prediction of molecular Als	426 patients (TCGA-CRC); 379 patients (DACHS)	ShuffleNet	AUROC: 0.76	The algorithm predicts a wide range of molecular alterations from routine, H-E stained slides
Kruger *et al*[99]	Prediction of MSTs	919 WSIs	ResNet 34	AUCs: Mean: 0.87; CMS1: 0.85; CMS2: 0.92, CMS3: 0.85; CMS4: 0.86	The MIL framework can identify morphological features indicative of different molecular subtypes
Popovici *et al*[100]	Prediction of MSTs	300 WSIs	VGG-F	Accuracy: 0.84;	The image-based

				Recall: 0.85; Precision: 0.84	classifier shows a significant prognostic value similar to the molecular counterparts
Cao *et al*[101]	MSI *vs* MSS	429 patients (TCGA-COAD); 785 patients (Asian-CRC)	EPLA	AUC: 0.88 (TCGA-COAD); AUC: 0.85 (Asian-CRC)	This pathomics-based model provides MSI estimation directly from images without molecular testing
Bilal *et al*[102]	Prediction of molecular AIs	502 slides (TCGA-CRC-DX); 47 slides (PAIP)	ResNet18, ResNet34, HoVerNet	AUROCS: HM (0.81 *vs* 0.71); MSI (0.86 *vs* 0.74); CIN (0.83 *vs* 0.73), BRAFmut (0.79 *vs* 0.66), TP53mut (0 *vs* 0.64), KRASmut (0.60), CIMP (0.79)	This algorithm is based on non-annotated images and uses only slide-level labels to predict the status of CRC pathways and mutations
Kwak *et al*[110]	LNM prediction	164 patients	CNN, U-Net	AUROC: 67%	PTS score is a potential prognostic parameter for LNM in CRC
Pai *et al*[111]	LNM prediction	230 patients (training), (136 testing)	CNN	AUROC: 79%	The model allows to identify and quantify a broad spectrum of histological features, including LNM in CRC
Kiehl *et al*[112]	LNM prediction	3013 patients	ResNET18	AUROC: 74.1%	DL-based analysis may help predict the LNM of patients with CRC using routine HE-stained slides
Weis *et al*[120]	Tumor Budding (Pan-CK)	381 patients	CNN	Spatial clusters of tumor buds correlates to N status (*P*: 0.003)	The model is a feasible and valid assessment tool for tumor budding on WSIs and can predict prognosis
Kather *et al*[121]	ADI, DEB, LYM, MUC, SM	86 slides (Training), 25 slides (Testing); 862 slide (TCGA-COAD)	VGG19	AUC: 98.7% HR: 2.29 (OS); 1.92 (RFS); Deep stroma score HR: 1.99 (*P*: 0.002), Shorter OS	This model can assess the human TME and predict prognosis directly from histopathological images
Shapcott *et al*[122]	TME (EC/IC/FC/MC)	853 patches, 142 images (TCGA-COAD)	CNN	Accuracy: 76% (detection), 65% (classification)	The model provides the assessment of TME in CRC slides
Sirinukunwattana *et al*[123]	a-4 tissues classes; b- prediction of DM	102 cases	Spatially Constrained CNN	a-AUROC: 90.4-99.9%; b-AUROC: 58.6-63.8%	The algorithm provides a digital marker for estimating the risk of DM
Swiderska-Chadaj *et al*[124]	TME Detection of ICs	28 WSIs	FCN/LSM/U-Net	F1-score of 0.80; Sensitivity: 74%; Precision: 86%	DL approaches are reliable for automatically detecting lymphocytes in IHC-stained CRC tissue sections
Geessink *et al*[115]	TSR	129 slides	CNN	HR: 2.48 (DSS); 2.05 (DFS)	CNN defined TSR as an independent prognosticator
Zhao *et al*[125]	TSR	499 patients (Discovery cohort); 315 patients (Validation cohort:)	CNN	TSR, independent prognostic parameter. HRs: 2.48 (Discovery cohort); 2.08 (Validation cohort)	CNN allows objective evaluation of TSR
Zhao *et al*[126]	Mucus tumor ratio low *vs* mucus tumor ratio high	814 patients	CNN	HRs: 1.88 (Discovery cohort); 2.09 (Validation cohort)	The DL quantified mucus tumor ratio is an independent prognostic factor in CRC
Bychkov *et al*[132]	Prognosis LR *vs* HR	420 TMA	VGG-16	HR: 2.3	The model extracts

					more prognostic information from the tissue morphology than the experienced human observer
Skrede *et al*[133]	Prognosis (CSS)	1122 patients (Validation cohort)	DoMorev1	HRs: 1.89 (uncertain *vs* good); 3.84 (poor *vs* good)	The digital marker has the potential to identify patients at LR and HR and provides the selection of treatment
Jiang *et al*[134]	a-HRR *vs* LRR b-Poor *vs* good prognosis	101 patients (Traning); 67 patients (Validation); 47 (TCGA-COAD)	InceptionResNetV2	a-HRs: 8.98 (training); 10.69 (other 2 test groups); b-HRs: 10.687 (training); 5.03 (other 2 test groups)	The selected model offers an independent prognostic predictor which allows stratification of stage III CRC into risk groups

NL: Normal; ADC: Adenocarcinoma; MC: Mucinous carcinoma; SC: Serrated carcinoma; PC: Papillary carcinoma; CCTA: Cribriform comedo-type adenocarcinoma; CRC: Colorectal cancer; HP: Hyperplastic polyp; SSP: Sessile serrated polyp; TSA: Traditional serrated adenoma; TA: Tubular adenoma; TVA: Tubulovillous adenoma; VA: Villous adenoma; AD: Adenoma; CNN: Convolutional neural networks; WSIs: Whole slide images; RNN: Recurrent neural networks; AUC: Area under the curve; GR: Grade; MSI: Microsatellite instable; MSS: Microsatellite stable; TCGA-CRC: Tumor Cancer Genome Atlas-Colorectal cancer; KR: Frozen tissues; DX: Formalin fixed paraffin embedded tissues; COAD: Colon adenocarcinoma; CRC: Colorectal carcinoma; EPLA: Ensemble patch likelihood aggregation; MST: Molecular subtype; CMS1: Tumor with MSI; CMS2: Tumors exhibiting epithelial gene expression, activated WNT and MYC signaling; CMS3: Tumors with metabolic disregulations; CMS4: Tumors that possess TGF-β; MIL: Multi instance learning; Als: Alterations; AUROC: Area under the receiver operating characteristics; PAIP: Pathology artificial intelligence platform; HM: Hypermutation; CIN: Chromosomally unstable; CIMP: CpG island methylator phenotype; CK: Cytokeratin; ML: Machine learning; ADI: Adipocyte; DEB: Debris; LYM: Lymphocytes; MUC: Mucus; SM: Smooth muscle; HR: Hazard ratio; OS: Overall survival; RFS: Recurrence free-survival; TME: Tumor microenvironment; ICs: Immune cells; FCN: Fully convolutional network; LSM: Liquid state machine; IHC: Immunohistochemistry; EC: Epithelial cell; FC: Fibroblast; MC: Miscellaneous; TSR: Tumor stroma ratio; LR: Low risk; HR: High risk; TMA: Tissue microarray; CSS: Cancer specific survival; HRR: High recurrence risk; LRR: Low recurrence risk; DM: Distant metastasis; LNM: Lymph node metastasis; PTS: The predictive value of the peritumoral stroma score.

their comparison across studies impossible. One of the most striking examples of these studies is Korbar *et al*[56], where a DL model (ResNet-152) trained with over 400 WSIs showed a high overall accuracy in subtyping polyps. In another study, Wei *et al*[61], who ensembled five layers of ResNet, could classify these lesions with WSIs from a single institution, even in external datasets with a performance comparable to that of histopathological evaluation. This data indicates that further manual annotations by various qualified GI pathologists may be required to decrease classification problems in future AI systems for colorectal polyp detection.

AI IN THE PATHOLOGICAL DETERMINATION OF TUMOR BEHAVIOR IN GIS

In this section relevant data on GC and CRC will be discussed. Unfortunately, no AI studies have identified the parameters that are important in determining tumor behavior and survival in EC. Similarly, studies of EC concerning molecular characterization have not been found. Therefore, in EC, a tumor with extremely high mortality, it is clear that additional pathology studies are necessary to reveal the effectiveness of AI applications in predicting tumor behavior.

TUMOR SUBTYPING

Gastric cancer

Although nearly all GC are ADC, the clinicopathological features and behaviors show considerable variation depending on the histopathological diversity of tumor cells[63,64]. In recent years, it has been reported that the survival of patients with GC at the same stage differs significantly among the different subtypes. Therefore, accurate histopathological classification is critical in determining their prognosis, monitoring, and treatment.

GC is often classified based on the ADC differentiation grade, including well-differentiated ADC and poorly differentiated ADC. The grading depends on the presence or absence of glandular structure formation. ADCs are divided into intestinal and diffuse subtypes based on the Lauren classification[65]. While the diffuse form comprises a poorly differentiated type and signet ring cell carcinoma (SRCC), the intestinal type exhibits glands with papillae, tubules, or solid regions. Diffuse-type carcinomas are commonly confused with other nonneoplastic diseases. Because they usually consist of solitary dispersed cells in a desmoplastic stroma and inflammation.

In most of the reported studies, the adenocarcinoma differentiation grade is judged through manual identification by pathologists. Although there have been many studies on AI applications in the pathological diagnosis of GC in the recent past, there are few studies regarding tumor subclassification (Table 3). Yasuda *et al*[66] investigated the features and classification of GC tissues by using supervised ML algorithms. The results showed that this method reliably identifies morphological changes in tumors with different grades. Interestingly, PD-L1 expression levels have been found to serve as a morphological classification in hematoxylin and eosin (HE)-stained slides and correlate with histological grades. Therefore, quantitative analyses of tissue morphology may reveal molecular alterations in malignancies, and molecular analyses may aid in the pathological evaluation of cancer tissues. In another study, four different DL models were used to classify GC into diffuse ADC *vs* other ADC subtypes[67]. From biopsy WSIs, the trained model performed well at identifying both poorly differentiated ADC and SRCC cells. The authors pointed out that while higher magnification can reduce the false positive rate in classification, applying an RNN model with a more comprehensive dataset yields good results even at low magnifications. Hybrid models such as StoHisNet have also distinguished tubular, mucinous, and papillary subtypes of GC. This model showed a higher performance for multiclassification of pathological images of GC than other CNN-based models[68]. Although the model performed well in the four classifications of gastric pathological images, the study group does not include SRCC and other types. Also, the inability of the supervised network in the study to use unlabeled data and the lack of information on which combination maximizes the performance of the model performance warrant further studies. More recently, Su *et al*[69] demonstrated that DL models constructed using a pre-trained ResNet-18 model based on ImageNet27 achieved tumor differentiation recognition or poorly differentiated ADC and well-differentiated ADC classes, respectively. Although these results suggest that AI may be useful in GC classification, the scarcity of data and the differences in classification parameters used in these studies make it difficult to come to any solid conclusions.

Recently, GC has also been classified by the Tumor Cancer Genome Atlas (TCGA) into four molecular subtypes that are also included in the latest World Health Organization classification: Epstein–Barr-virus (EBV)-positive (9%), microsatellite unstable (MSI) (22%), genomically stable (GS) (19%) and chromosomally unstable (CIN) (50%)[70,71]. The clinical significance of this classification comes from the fact that various factors, such as the prognosis and treatment response, differ among these subtypes [72,73]. In particular, among all subclasses of GC, tumors with MSI and positive EBV are associated with a better response to immunotherapy[72]. Consequently, recognizing these subtypes is crucial for categorizing patients who benefit from these treatments. Nevertheless, such classification requires the application of costly techniques, such as immunohistochemistry, and molecular testing, such as polymerase chain reaction, into pathological practice.

On the other hand, these two types have known characteristic histopathological findings. While EBV-positive GCs show prominent infiltration of lymphocytes into the neoplastic epithelium and the stroma, MSI subtype shows significant lymphocytic infiltration, intestinal-type histology, and expanding growth characteristics[63,74,75]. Therefore, these morphological features could be used to make predictions about the molecular subtype. In recent years, it has been suggested that molecular findings can be detected with AI *via* WSIs from HE-stained sections produced for pathological assessment[76-78]. Various models have been applied for molecular subtyping of GIS cancers. However, most of these studies have been conducted on CRCs (see below), whereas relatively few studies are available for GC (Table 3). For the detection of GC subtypes, Muti *et al*[79] demonstrated that DL could detect MSI and EBV positivity independently from each other in GC directly from HE-stained tissues in multi center pooled cohorts. They observed a high classification performance for the detection of MSI and EBV status. The relatively limited number of cases with positive findings and the fact that the ground truth methods for MSI were developed in CRC are presented as potential limitations of this study. On the other hand, their findings align with previous observations[69,80,81]. In addition, large-scale and multicenter validation broadens their work, which has considerable potential for integration into clinical procedures, suggesting that the application of DL could be a substitute for molecular techniques in the classification of GC. Furthermore, because these two subtypes share common morphological features and they are immunotherapy-sensitive tumors, Hinata *et al*[82] combined MSI and EBV in DL models and found they had a higher detection accuracy. This finding has been interpreted based on the possibility that these subtypes have similar distinctive pathological features, such as abundant stromal lymphocytic infiltration and intraepithelial lymphocytosis. On the other hand, the use of tissue microarray and manual labeling of tumor regions for TCGA presented as sources of bias compared to whole tissue slides, given the heterogeneity of tumor tissue. It was also emphasized that manual annotation by a pathologist might be a challenge to overcome by some weakly supervised methods (for example, attention-based deep multi instance learning) in the field of DL for the broad application of the proposed model.

Recently, a DL model called EBVNet that assists pathologists in predicting EBV from HE-stained slides has been introduced in GC[83]. The results suggested that human-machine fusion dramatically enhances the diagnostic ability of both EBVNet and the pathologist. However, this study has some limitations regarding its retrospective evaluation of training and validation. Additionally, the logistic regression model applied in the assessment is still an indirect way to interpret the model. More importantly, as in many DL models, the EBVNet decision-making procedure by the neural network is

Table 3 Artificial intelligence-based applications in pathology for the determination of tumor behavior in gastric cancer

Ref.	Task	Data sets	Algorithm/Model	Performance	Comments
Yasuda *et al*[66]	NC, GR1, GR2, GR3; PDL-1, ATF7IP/MCAF1	66 WSIs	SV, ML, wndchrm	AUCs: 0.98-0.99	The model allows grading emphasizing a correlation between molecular expression and tissue structures
Kanavati *et al*[67]	NC, ADC-D, ADC-O	1-stage training: 1950 WSIs, 2-stage training: 874 WSIs	CNN and RNN	AUCs: 0.95-0.99	The tool can aid pathologists by potentially accelerating their diagnostic workflow
Fu *et al*[68]	NC, TC, MC, PC	Training 2938 WSIs, Testing 980 WSIs	StoHisNet	The accuracy: 94.69%, F1 score: 94.96%, Recall: 94.95%, Precision: 94.97%	The model has high performance in the multi-classification on gastric images and shows strong generalization ability on other pathological datasets
Su *et al*[69]	NC, WD, PD, MSS *vs* MSI	GR: Training 348 WSIs, Testing 88 WSIs MSS: Training 212 WSIs, Testing: 52 WSIs, MSI: Training 136 WSIs, Testing: 36 WSIs	ResNet-18	PD *vs* WD, F1 score: 0.8615, PD *vs* WD *vs* NC, F1 score: 0.8977; MSI *vs* MSS accuracy: 0.7727	The proposed system integrated the tumor GR and MSI status recognition problems into the same workflow and was suitable for exploring the relationships between pathological features and molecular status
Muti *et al*[79]	MSI *vs* MSS; EBV (+) *vs* EBV (-)	2823 patients with known MSI status; 2685 patients with known EBV status	CNN, Shufflenet	MSI *vs* MSS, AUROCs: 0.723-0.863; EBV (+) *vs* EBV (-), AUROCs: 0.672-0.859	DL-based classifiers have the potential to provide faster decisions for pathologists and to offer therapeutic options tailored to the molecular profile of the individual patient
Kather *et al*[80]	MSI *vs* MSS	Training 81 patients +216 patients (TCGA-STAD)	ResNet-18	AUC: 0.84	This system provides significant improvements in molecular alterations screening workflow
Kather *et al*[81]	EBV (+) vs. EBV (-)	Training 317 patients (TCGA-STAD)	CNN, VGG19	AUC: 0.80	This workflow enables a fast and low-cost method to identify EBV and enables pathologists to check the plausibility of computer-based image classification (the black box of DL)
Hinata *et al*[82]	EBV+MSI/dMMR *vs* EBV- non MSI/dMMR	UTokyo training cohort: 326 patients; TCGA training cohort: 48 patients	CNNs,VGG16, VGG19, ResNet50, EfficientNetB0	AUCs: 0.901–0.992 (Utokyo cohort); AUCs: 0.809–0.931 (TCGA cohort)	The model detects immuno-therapy-sensitive GC subtypes from histological images at a lower cost and in a shorter time than the conventional methods
Zheng *et al*[83]	EBV (+) *vs* EBV (-)	EBV (+) 203 WSIs; EBV (-) 803 WSIs	EBVNet	AUROC: 0.969, Internal validation; AUROC: 0.941, External dataset AUROC: 0.895, TCGA dataset	The human-machine fusion significantly improves the diagnostic performance of both the EBVNet and the pathologist, provides an approach for the identification of EBV(+) GC, and may help effectively select patients for immuno-therapy
Flinner *et al*[87]	EBV, MSI, GS, CIN	Training 84 WSIs (TCGA-STAD); Testing: 133 WSIs (TCGA-STAD)	CNN, DenseNet161	AUC: 0.76 for four classes	The simplified molecular TCGA and GC subclasses could be predicted by DL directly based on H-E staining
Jang *et al*[88]	CDH1, ERBB2, KRAS, PIK3CA, TP53 mutations	425 FF slides (TCGA-STAD); 320 FT slides (TCGA-STAD)	CNN, Inception-v3	AUCs (FF-FT): CDH1 (0.667-0.778), ERBB2(0.63-0.833), KRAS (0.657-0.838); PIK3CA (0.688-0.761), TP53 (0.572-0.775)	When trained with appropriate tissue data, DL could predict genetic mutations in H-E-stained tissue slides

Huang *et al*[109]	Metastatic LNs	983 WSIs	ESCNN	AUC: 0.9936	ESCNN improves the accuracy of pathologists in identifying metastatic LNs, micrometastases, and isolated tumor cells, allowing for shortening the review time
Hu *et al*[107]	Metastatic LNs	222 patients	RCNN, Xception and DenseNet-121	Accuracy 97.13%; PPV: 93.53, NPV: 97.99%	The system can be implemented into clinical workflow to assist pathologists in preliminary screening for LN metastases in GC patients
Matsushima *et al* [108]	Metastatic LNs	827 lymph nodes	CNN	AUROC: 0.9994	This DL-based diagnosis-aid system can assist pathologists in detecting LN metastasis in GC and reduce their workload
Wang *et al*[106]	Metastatic LNs, T/LNM	9366 slides (7736 with metastasis)	Resnet-50	LNM (+) *vs* (-): Sensitivity 98.5%, Specificity 96.1%; T/LNM: HR: 2.05 (univariate analysis); 1.39 (multivariate analysis)	This system can assist pathologists in detecting LN metastasis in GC and reduce their workload. Besides, T/LNM is prognostic of OS in GC patients
Hong *et al*[116]	dTSR (HE and CK7)	Training 13 WSIs; Testing 358 WSIs	cGAN	Kappa value: 0.623 (dTSR and vTSR); AUROC: 0.907; OS (*P*: 0.0024)	By diagnosing TSR in GC, this model predicts OS in the advanced stage of GC
Meier *et al*[127]	TME + Ki-67	248 patients	CNN	HRs: Ki67&CD20: 1.364, CD20&CD68: 1.338; Ki67&CD68: 1.473	In combination with a panel of IHC markers, this model predicts the prognosis of patients with GC
Huang *et al*[128]	OS	Training: 2261 pictures; Internal validation: 960 pictures	GastroMIL	HR: 2.414 (univariate analysis), 1.843 (multivariate analysis)	The risk score computed by MIL-GC was proved to be the independent prognostic value of GC
Jiang *et al*[129]	5-YS, 5-YDFS	786 patients	ML, SVM	AUCs: 5-YS: 0.834; 5-YDFS: 0.828	The classifier can accurately distinguishes GC patients with different OS and DFS and identifies a subgroup of patients with stage II and III disease who could benefit from adjuvant chemotherapy
Jiang *et al*[130]	Low SVM *vs* High SVM, 5-YS, 5-YDFS	Training: 223 patients; Internal validation: 218 patientsExternal validation: 227 patients	ML, SVM	AUCs: 5-YS: 0.818; 5-YDFS: 0.827	SVM signature distinguish GC patients with different OS and DFS and identifies a subgroup of patients with stage II and III disease who could benefit from adjuvant chemotherapy
Wang *et al*[131]	TME	172 patients	$CG_{Signature}$ powered by AI	AUROCs: 0.960 ± 0.01 (binary classification), 0.771 ± 0.024 to 0.904 ± 0.012 (ternary classification)	Digital grade cancer staging produced by CGSignature predicts the prognosis of GC and significantly outperforms the AJCC 8th edition Tumor Node Metastasis staging system

NC: Non cancer; GR: Grade; ATF7IP/MCAF1: Activating Transcription Factor 7 Interacting Protein; WSIs: Whole slide images; SV: Supervised; ML: Machine learning; wndchrm: weighted neighbor distances using a compound hierarchy of algorithms representing morphology; AUC: Area under the curve; ADC-D: Diffuse adenocarcinoma; ADC-O: Adenocarcinoma other; DL: Deep learning; CNN: Convolutional neural networks; RNN: Recurrent neural network; TC: Tubular carcinoma; MC: Mucinous carcinoma; PC: Papillary carcinoma; WD: Well differentiated; PD: Poorly differentiated; MSI: Microsatellite instable; MSS: Microsatellite stable EBV: Epstein-Barr virus; TCGA-STAD: Tumor Cancer Genome Atlas, Stomach adenocarcinoma dMMR: Deficient mismatch repair; GC: Gastric cancer; AUROC: Area under the receiver operating characteristics; GS: Genomically stable; CIN: Chromosomally unstable; LN: Lymph node; ESCNN: Enhanced streaming CNN; RCNN: Region based CNN; PPV: Positive predictive value; NPV: Negative predictive value; T/LNM: Tumor area-to-metastatic LN-area ratio; dTSR: Digital tumor-stroma ratio; HE: Hematoxylin and eosin; CK7: Cytokeratin 7; cGAN: Conditional generative adversarial network; vTSR: Visual tumor-stroma ratio; OS: Overall survival; TME: Tumor microenvironment; HR: Hazard ratio; 5-YS: Five year survival; 5-YDFS: Five year disease free survival; SVM: Support vector machine; AI: Artificial intelligence.

nontransparent (black boxes). Since various methods have been proposed to solve black boxes in DL in the recent past, additional studies applying these methods will contribute to the determination of the molecular subtypes of AI models of GC[84-86]. In a more recent study, Flinner *et al*[87], in their study emphasizing the error-proneness of the morphological and staining methods used to determine GC subtypes for subclassification, found that DL could be more effective in this regard. On the other hand, they also pointed out that image tiles labeled with false ground truth associated with GC heterogeneity may reduce the accuracy of DL but this can be overcome by first experimentally defining the test data.

Recently, the feasibility of a DL approach has also been evaluated in the classification of GC for mutations in the CDH1, ERBB2, KRAS, PIK3CA, and TP53 genes[88]. High AUCs observed in both frozen and formalin-fixed tissues highlight that DL-based classifiers could predict the mutational status of these tumors. Although these results are promising for the application of AI to subtyping GC, additional studies are necessary, with further refinement of these methods.

Colorectal cancer

Similar to GC, molecular subtyping of CRC is essential for targeted treatment against critical oncogenic signaling pathways. CRCs are divided by molecular consensus into four types (CMS): 1. CMS1: Tumors with MSI that have a good prognosis in non metastatic stages; CMS2: Tumors with intermediate prognosis exhibiting epithelial gene expression, activated WNT and MYC signaling; CMS3: Tumors with intermediate prognosis demonstrating metabolic dysregulations; CMS4: Tumors with a poor prognosis that possess transforming growth factor beta (TGF-β) activation[89-91]. The identification of CRC with MSI is paramount because this group is susceptible to immunomodulating therapies[92,93]. Although some findings, such as tissue architecture, growth pattern, cellular morphology, and distributions of tumor stroma ratio (TSR) and tumor microenvironment (TME) provide some clues about the subclassification of these tumors, molecular stratification of patients necessitates RNA analyses that are expensive and difficult to standardize[94-96]. Accordingly, some studies have investigated the contribution of AI to tumor subclassification from HE-stained tissue sections by DL models (Table 2). Sirinukunwattana *et al*[97] demonstrated that a CNN-based model could detect CMS subtypes. At the same time, they criticized the potential over fitting of the computational model to the training cohort as a limitation of the study. In a more recent study, Echle *et al*[98] developed a DL model in a large series of 8836 cases of CRC to predict MSI tumors. In the international validation of the study group, the algorithm achieved a high performance [area under the receiver operating curve (AUROC) of 0.96][80]. Other investigators have also reported similar results, pointing out the potential use of DL models for detecting molecular subtypes of CRC[77,99-101]. In a retrospective study, a DL pipeline method was developed based on experimental setups similar to previous studies[102]. Three models were used to predict mutation density (low *vs* high), MSI, CIN, and GpG island methylator phenotype. The mutated and wild-type BRAF, TP53, and KRAS types were also investigated. This method showed higher AUROCs for the prediction of hypermutation, MSI, CIN, BRAF, and TP53 compared to previously reported data, suggesting that AI methods may provide the stratification of patients with CRC for targeted therapies. However, further large-scale validations with multicenter datasets are required before their implementation in pathological practice.

LYMPH NODE METASTASIS

Gastric cancer

Another important parameter that predicts GC behavior and treatment is lymph node metastasis (LNM) [103]. However, identifying LNM is still a challenging and tedious task in pathological practice, making the implementation of AI an attractive tool to reduce the workload[104,105]. Although numerous studies have demonstrated that DL-based algorithms can detect metastatic lymph nodes in GC with a similar level of accuracy to human specialists, these algorithms have not yet been implemented into pathology practice[106-108] (Table 3). The failure to integrate these algorithms is related to the characteristics of WSIs, the excessive effort required to apply the annotation, and the limited associated data. Recently, Huang *et al*[109] developed a weakly supervised end-to-end technique termed enhanced streaming CNN (ESCNN). Their results revealed that the routine pathological evaluation benefitted from the AI-assisted LN assessment workflow regarding review time, sensitivity, and consistency. On the other hand, AI-attributable false alarms that misled the pathologists on negative results led to a decrease in specificity from 94% to 84%, which needs more large-scale or multicenter studies to check the effectiveness of the workflow.

Colorectal cancer

Recent evidence indicates that features extracted by DL models from routine histologic slides can predict LNM in CRC[110-112] (Table 2). For example, Kwak *et al*[110] detected LNM by generating a score based on the ratio of peritumoral stroma to tumor tissue on a test set. In another study, the presence of LNM was detected with a model which segmented WSIs into areas such as tumor budding or poorly differentiated clusters[111]. More recently, Kiehl *et al*[112] performed an approach that uses

DL-based image analysis (slide-based artificial intelligence predictor) in association with patient data to estimate LNM in CRC patients. Their results indicated that LNM could be predicted in patients with CRC through AI applications from histological slides to a similar level to using a classifier containing clinical data.

THE TUMOR STROMA RATIO, TUMOR MICROENVIRONMENT AND TUMOR BUDDING

Gastric cancer

In recent years, it has been shown that the TSR in many organ tumors is an important clue to the course of the disease. In particular, stromal dominance has been observed to be an independent prognostic factor in many tumors, including GIS[113,114]. However, TSRs are not included in pathology report protocols because of the lack of a standard procedure among different methodologies and a low reproducibility related to the high interobserver variation[115]. Recently, a DL pipeline has been introduced to facilitate the automated assessment of TSR in GC[116]. Although this model has been shown to be effective in detecting survival according to the low and high TSR rates in advanced GC, it was emphasized that some limitations, such as the nonautomatic selection of hot spots and the use of a single test, should be eliminated. Therefore, there is a need for many studies on the use of AI applications in TSR determination of GC.

In a recent study, a DL model determined the tumor-to-metastatic lymph node-area ratio in metastatic lymph nodes in patients with GC[106]. Statistical analysis also revealed that this ratio is an independent prognostic factor warranting further investigation.

Colorectal cancer

In CRC, recent studies have demonstrated that lymphocytes and fibroblasts profoundly shape the TME and significantly impact tumor behavior[117-119]. In addition, it has been shown that CRC may have a poor prognosis due to tumor budding (1-5 cells in the invasive area)[120]. In the literature, seven studies of AI methods have been identified to determine these parameters in a more objective and time-saving manner (Table 2). However, many of them used different methods. Three models focused on the classification of the cell types, such as epithelial, inflammatory, fibroblast, lymphocytes, and others (mucus, smooth muscle, normal mucosa, stroma, and cancer epithelium)[121-123]. In an elegant study, a DL algorithm was proposed for estimating the risk of distant metastasis by analyzing the TME[123]. Cell detection and cell classification were evaluated in two CNNs used to build a cell network. In each tumor, a tissue phenotype signature was obtained by proportioning the area of tissue phenotypes to the total tissue area. Statistical analysis revealed that the connection frequency (CF) of the smooth muscle ratio, the CF of the inflammation ratio, and the appearance (AP) based on inflammation could independently estimate the development of distant metastasis. Distant metastasis-free survival analysis indicated that CF smooth muscle and AP inflammation ratios were potential prognosticators. Although the hazard ratios for CF of the smooth muscle ratio and AP inflammation were 2.11 and 0.39, respectively, the AUC values for distant metastasis prediction were 0.59 for the CF of the smooth muscle ratio and 0.64 for AP based on inflammation. As emphasized by the authors, specific immunohistochemical staining can improve the prediction of distant metastases by increasing the informative value of histological slides. Another limitation of this study is the small number of metastatic cases. Another recent study was performed to detect CD3- and CD8-positive immune cells on WSIs of slides stained by immunohistochemistry in a multicenter cohort by four different methods[124]. U-Net obtained the highest performance and highest agreement with manual evaluation (0.72), which was higher than that of pathologists (K = 0.64), supporting that DL models are helpful for automatically detecting lymphocytes in immunohistochemically stained tissue sections.

In CRC, the automatic tumor budding evaluation on immunohistochemical pankeratin-stained slides revealed that the absolute number of buds per image was significantly correlated with manually segmented ground truth (R: 0.86)[120]. Interestingly, the number of spatial clusters of buds in hot spots was significantly correlated with the prognosis. In three studies, the impact of detecting the TSR or deep stroma score in CRC by DL algorithms was found to be an independent parameter to predict tumor behavior[115,121,125] (Table 2).

Recently, Zhao *et al*[126] demonstrated that the ratio of the mucinous component in the tumor area (MTR) quantified by AI is an independent prognostic factor in CRC. On the other hand, the most invasive part of primary tumors was selected for evaluation. As noted by the authors, measuring the exact proportion and prognostic value of mucus in the entire tumor is still worthy of further investigation.

SURVIVAL OUTCOMES

Gastric cancer

Another continuing research topic is evaluating survival outcomes in GC with AI models[127-129] (Table 3). Recently, support vector machine (SVM), one of the popular algorithms in ML, has been applied to predict the survival of GC. Jiang *et al*[129] demonstrated that SVM could be useful in predicting the outcome and identifying patients with GC who might benefit from adjuvant therapy. In this study, the classifier incorporated patient gender, carcinoembryonic antigen levels, LNM, and the protein expression level of eight features, composed of CD3 invasive margin (IM), CD3 center of the tumor (CT), CD8IM, CD45ROCT, CD57IM, CD66bIM, CD68CT, and CD34. There were significant variations between the high- and low-GC-SVM classifiers. Recently, Huang *et al*[128] designed MIL-GC (a DL-based model) to predict overall survival (OS) in patients with GC. They observed C-indices of 0.728 and 0.671 in the training and internal validation sets, respectively. The external validation likewise exhibited strong prognostic prediction performance (C-index = 0.657), confirming the resilience of the two models. Furthermore, univariate and multivariate Cox analyses demonstrated that the risk score derived by MIL-GC has independent prognostic significance, indicating the potential of AI approaches to predict GC behavior. Additionally, tumor progression includes complex interactions between malignant cells and their surrounding microenvironment (TME)[130]. TME targeting and reprogramming is, in fact, can be a potential strategy to achieve antitumor effects in many cancers. Several AI studies involving the TME have recently demonstrated that these methods can determine the prognosis of GIS cancers. Regarding GC, Wang *et al*[131], suggested a graph NN-based solution, CellGraph Signature powered AI, for the digital staging of TME and the exact prediction of patient survival by combining and converting multiplexed immunohistochemistry (mIHC) images as Cell-Graphs. The survival prediction achieved outstanding model performance for both binary and ternary classifications. Furthermore, survival analysis revealed that this method outperforms the AJCC 8th edition Tumor Node Metastasis staging system in discriminating both binary and ternary classes with statistical significance (P value < 0.0001), implying the effectiveness and advantages of such an AI-powered digital staging system in DP and precision oncology.

These data demonstrate that AI-based models allow prognosis prediction in GC. However, developing efficient models requires training on large sets reflecting scanning and staining protocols variability.

Colorectal cancer

Regarding prognostic evaluations from HE-stained slides by AI in CRC, some DL models have been developed for prognostication (Table 2). Bychkov *et al*[132] combined a CNN and a recurrent NN model to estimate the disease-specific five-year survival from tumor tissue microarray samples without tissue classification. The model classified patients into a low- or high-risk group (AUC of 0.69). This result was more significant than the AUC of the visual evaluation of the pathologist (AUC of 0.58) or the histological grade determined at the time of the original diagnosis (AUC of 0.57). However, an external dataset was not included. In another study by Skrede *et al*[133], diverse data from four different cohorts were used to develop an automatic prognostic marker to predict the outcome. The model included a CNN used to separate tumor tissue and two other CNN ensembles that identified individuals as having a favorable or poor survival. Patients were assigned as uncertain when the two CNN ensembles predicted different outcomes. In an external test group, the classifier was a strong predictor of survival. In addition, the output of the two CNN ensembles produced a strong predictive score related to patient outcome (AUC of 0.71). A generalization of this approach has been recommended, as an external test cohort from more than one medical center demonstrated similar hazard ratios.

Jiang *et al*[134], to achieve a shorter computational time, developed a hybrid model by synergizing ML algorithms with DL (InceptionResNetV2 and gradient boosting decision machine classifier) to predict the survival of patients with stage III CRC. While the internal test sets constituted a Chinese cohort, external testing was performed on the TCGA cohort. They revealed that the model stratifies patients with stage III colon cancer into high- and low-risk recurrence and poor and favorable prognostic groups directly from tissue sections. These data suggest that the analysis of H-E-stained tissue samples by AI methods could serve as a digital prognostic biomarker in CRC. However, additional studies are warranted to support the evaluation of the performance of these methods in larger patient series.

OVERALL LIMITATIONS OF AI-BASED APPLICATIONS IN REAL-LIFE PRACTICE

In the literature, there are some frequently discussed topics considering the general challenges of AI such as identification of the clinical need, ethical considerations, funding, optimization of data-sets, annotation of the dataset, regulation, validation, and implementation[46].

Recognizing the actual clinical need and defining a potential solution is the first stage in developing the AI application. However, there can be an imbalance between the benefits in daily pathological practice and the total cost of its implementation. As a result, the market for a particular AI tool may be too tiny and it may not be profitable.

Although patients can provide permission for data to be used for studies, constructing AI models may have issues if commercial use is not approved[135]. In order to develop a framework for global data sharing, patient consent should include the possibility of its commercial use for product development[40].

Training on huge datasets is necessary for developing AI systems with high performance in digital pathology. Changes related to differences in fixation, tissue thickness, and variations in staining and scanning protocols encountered in preanalytical and analytical phases may influence data accuracy[136,137]. For example, it is difficult to convert a glass slide to WSI, and changing the hue of the slide could affect AI accuracy. Many AI algorithms have emerged for this purpose recently, including staining and color features[138,139]. In addition, a number of algorithms are presented to optimize WSI quality. These algorithms identify areas of the highest quality and exclude areas that are out of focus or affected by artifacts[140,141].

Concerning the implementation of AI, to enable users to shift the daily routine practice in the pathology laboratory, from glass slides to WSIs, the first step is to install an institutional IT infrastructure. In addition to these changes in infrastructure, pathology residency training might need to be adjusted in accordance with the availability of this new tool. Preventing residents from relying completely on AI while also allowing them to benefit from it as a helping instrument would require fine balancing and planning prior to its installation[142].

Similar to other clinical tests, quality assurance is crucial, hence it is urgently necessary to develop a plan for external quality assurance for applications. Furthermore, laboratory workers should also be familiar with the quality management system.

Although some algorithms and automated AI models are thought to perform better than pathologists, pathologists will always be required to audit technology and control mechanisms in AI implementation [143].

CONCLUSION

In this review, we outlined the potential of AI applications for evaluating pathological parameters related to the behavior of GIS cancers. Current data suggest the merit of AI-based approaches in assessing tumor grading, subtyping, detection of metastasis, and prognosis in GC and CRC. In addition, these methods encourage biomarker discovery by revealing predictions that are impossible when using traditional visual methods. Regarding EC, there is still much room for improvement in developing AI models to predict the behavior of these tumors in pathology. On the other hand, the enormous potential of AI in improving workflows, eliminating simple errors, and increasing objectivity during pathological evaluations to determine the behavior of GIS cancers should motivate researchers to overcome the many remaining hurdles. In algorithm development, variations in imaging data, interobserver variability during interpretations, model transparency, and interpretability are significant challenges to be solved. A large number of studies with external validation and quality controls implemented on large datasets are essential in meeting the standards of these methods. Thereby, AI applications that are practical, interpretable, manageable, and cost-effective can play a crucial role in the development of pathological evaluations to be performed in the prognosis and treatment of GIS tumors.

FOOTNOTES

Author contributions: Yavuz A, Alpsoy A, Gedik EO, Celik MY performed the data acquisition; Bassorgun CI, Unal B and Elpek GO designed the outline and coordinated the writing of the paper; all authors equally contributed to the writing of the paper and preparation of the tables.

Conflict-of-interest statement: All the authors declare that they have no conflict of interest.

Country/Territory of origin: Turkey

ORCID number: Aysen Yavuz 0000-0001-9991-5515; Anil Alpsoy 0000-0003-4978-7652; Elif Ocak Gedik 0000-0003-2618-498X; Mennan Yigitcan Celik 0000-0001-8769-5156; Cumhur Ibrahim Bassorgun 0000-0003-2440-511X; Betul Unal 0000-

0001-7680-1808; Gulsum Ozlem Elpek 0000-0002-1237-5454.

S-Editor: Liu JH
L-Editor: A
P-Editor: Liu JH

REFERENCES

1 **Arnold M**, Abnet CC, Neale RE, Vignat J, Giovannucci EL, McGlynn KA, Bray F. Global Burden of 5 Major Types of Gastrointestinal Cancer. *Gastroenterology* 2020; **159**: 335-349.e15 [PMID: 32247694 DOI: 10.1053/j.gastro.2020.02.068]
2 **Bray F**, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; **68**: 394-424 [PMID: 30207593 DOI: 10.3322/caac.21492]
3 **Johdi NA**, Sukor NF. Colorectal Cancer Immunotherapy: Options and Strategies. *Front Immunol* 2020; **11**: 1624 [PMID: 33042104 DOI: 10.3389/fimmu.2020.01624]
4 **Li K**, Zhang A, Li X, Zhang H, Zhao L. Advances in clinical immunotherapy for gastric cancer. *Biochim Biophys Acta Rev Cancer* 2021; **1876**: 188615 [PMID: 34403771 DOI: 10.1016/j.bbcan.2021.188615]
5 **Niazi MKK**, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol* 2019; **20**: e253-e261 [PMID: 31044723 DOI: 10.1016/S1470-2045(19)30154-8]
6 **Abels E**, Pantanowitz L, Aeffner F, Zarella MD, van der Laak J, Bui MM, Vemuri VN, Parwani AV, Gibbs J, Agosto-Arroyo E, Beck AH, Kozlowski C. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *J Pathol* 2019; **249**: 286-294 [PMID: 31355445 DOI: 10.1002/path.5331]
7 **Dangott B**, Parwani A. Whole slide imaging for teleconsultation and clinical use. *J Pathol Inform* 2010; **1** [PMID: 20805956 DOI: 10.4103/2153-3539.65342]
8 **Evans AJ**, Depeiza N, Allen SG, Fraser K, Shirley S, Chetty R. Use of whole slide imaging (WSI) for distance teaching. *J Clin Pathol* 2021; **74**: 425-428 [PMID: 32646928 DOI: 10.1136/jclinpath-2020-206763]
9 **Saillard C**, Schmauch B, Laifa O, Moarii M, Toldo S, Zaslavskiy M, Pronier E, Laurent A, Amaddeo G, Regnault H, Sommacale D, Ziol M, Pawlotsky JM, Mulé S, Luciani A, Wainrib G, Clozel T, Courtiol P, Calderaro J. Predicting Survival After Hepatocellular Carcinoma Resection Using Deep Learning on Histological Slides. *Hepatology* 2020; **72**: 2000-2013 [PMID: 32108950 DOI: 10.1002/hep.31207]
10 **Wong ANN**, He Z, Leung KL, To CCK, Wong CY, Wong SCC, Yoo JS, Chan CKR, Chan AZ, Lacambra MD, Yeung MHY. Current Developments of Artificial Intelligence in Digital Pathology and Its Future Clinical Applications in Gastrointestinal Cancers. *Cancers (Basel)* 2022; **14** [PMID: 35954443 DOI: 10.3390/cancers14153780]
11 **Courtiol P**, Maussion C, Moarii M, Pronier E, Pilcer S, Sefta M, Manceron P, Toldo S, Zaslavskiy M, Le Stang N, Girard N, Elemento O, Nicholson AG, Blay JY, Galateau-Sallé F, Wainrib G, Clozel T. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat Med* 2019; **25**: 1519-1525 [PMID: 31591589 DOI: 10.1038/s41591-019-0583-3]
12 **Dun XP**, Parkinson DB. Visualizing peripheral nerve regeneration by whole mount staining. *PLoS One* 2015; **10**: e0119168 [PMID: 25738874 DOI: 10.1371/journal.pone.0119168]
13 **Kim D**, Pantanowitz L, Schüffler P, Yarlagadda DVK, Ardon O, Reuter VE, Hameed M, Klimstra DS, Hanna MG. (Re) Defining the High-Power Field for Digital Pathology. *J Pathol Inform* 2020; **11**: 33 [PMID: 33343994 DOI: 10.4103/jpi.jpi_48_20]
14 **Heffner S**, Colgan O, Doolan C. Digital Pathology. Available from: https://www.leicabiosystems.com/en-br/knowledgepathway/digitalpathology/
15 **Borowsky AD**, Glassy EF, Wallace WD, Kallichanda NS, Behling CA, Miller DV, Oswal HN, Feddersen RM, Bakhtar OR, Mendoza AE, Molden DP, Saffer HL, Wixom CR, Albro JE, Cessna MH, Hall BJ, Lloyd IE, Bishop JW, Darrow MA, Gui D, Jen KY, Walby JAS, Bauer SM, Cortez DA, Gandhi P, Rodgers MM, Rodriguez RA, Martin DR, McConnell TG, Reynolds SJ, Spigel JH, Stepenaskie SA, Viktorova E, Magari R, Wharton KA, Qiu J, Bauer TW. Digital Whole Slide Imaging Compared With Light Microscopy for Primary Diagnosis in Surgical Pathology. *Arch Pathol Lab Med* 2020; **144**: 1245-1253 [PMID: 32057275 DOI: 10.5858/arpa.2019-0569-OA]
16 **Hanna MG**, Reuter VE, Ardon O, Kim D, Sirintrapun SJ, Schüffler PJ, Busam KJ, Sauter JL, Brogi E, Tan LK, Xu B, Bale T, Agaram NP, Tang LH, Ellenson LH, Philip J, Corsale L, Stamelos E, Friedlander MA, Ntiamoah P, Labasin M, England C, Klimstra DS, Hameed M. Validation of a digital pathology system including remote review during the COVID-19 pandemic. *Mod Pathol* 2020; **33**: 2115-2127 [PMID: 32572154 DOI: 10.1038/s41379-020-0601-5]
17 **Aloqaily A**, Polonia A, Campelos S, Alrefae N, Vale J, Caramelo A, Eloy C. Digital Versus Optical Diagnosis of Follicular Patterned Thyroid Lesions. *Head Neck Pathol* 2021; **15**: 537-543 [PMID: 33128731 DOI: 10.1007/s12105-020-01243-y]
18 **Yeh FC**, Ye Q, Hitchens TK, Wu YL, Parwani AV, Ho C. Mapping stain distribution in pathology slides using whole slide imaging. *J Pathol Inform* 2014; **5**: 1 [PMID: 24672736 DOI: 10.4103/2153-3539.126140]
19 **Calderaro J**, Kather JN. Artificial intelligence-based pathology for gastrointestinal and hepatobiliary cancers. *Gut* 2021; **70**: 1183-1193 [PMID: 33214163 DOI: 10.1136/gutjnl-2020-322880]
20 **Taqi SA**, Sami SA, Sami LB, Zaki SA. A review of artifacts in histopathology. *J Oral Maxillofac Pathol* 2018; **22**: 279 [PMID: 30158787 DOI: 10.4103/jomfp.JOMFP_125_15]
21 **Salvi M**, Acharya UR, Molinari F, Meiburger KM. The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. *Comput Biol Med* 2021; **128**: 104129

[PMID: 33254082 DOI: 10.1016/j.compbiomed.2020.104129]
22 **Tabesh A**, Teverovskiy M, Pang HY, Kumar VP, Verbel D, Kotsianti A, Saidi O. Multifeature prostate cancer diagnosis and Gleason grading of histological images. *IEEE Trans Med Imaging* 2007; **26**: 1366-1378 [PMID: 17948727 DOI: 10.1109/TMI.2007.898536]
23 **Macenko M**, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, Schmitt C, Thomas NE. A method for normalizing histology slides for quantitative analysis. Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009 June 28–July 1; Boston, USA, 2009: 1107–1110 [DOI: 10.1109/ISBI.2009.5193250]
24 **Yagi Y**. Color standardization and optimization in whole slide imaging. *Diagn Pathol* 2011; **6** Suppl 1: S15 [PMID: 21489185 DOI: 10.1186/1746-1596-6-S1-S15]
25 **Lee JS**, Ma YX. Stain Style Transfer for Histological Images Using S3CGAN. *Sensors (Basel)* 2022; **22** [PMID: 35161789 DOI: 10.3390/s22031044]
26 **Kausar T**, Kausar A, Ashraf MA, Siddique MF, Wang M, Sajid M, Siddique MZ, Haq AU, Riaz I. SA-GAN: Stain Acclimation Generative Adversarial Network for Histopathology Image Analysis. Appl Sci 2022; 12: 288 [DOI: 10.3390/app12010288]
27 **Schneider CA**, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012; **9**: 671-675 [PMID: 22930834 DOI: 10.1038/nmeth.2089]
28 **Bankhead P**, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, James JA, Salto-Tellez M, Hamilton PW. QuPath: Open source software for digital pathology image analysis. *Sci Rep* 2017; **7**: 16878 [PMID: 29203879 DOI: 10.1038/s41598-017-17204-5]
29 **Aubreville M**, Bertram C, Klopfleisch R, Maier A. SlideRunner. In Bildverarbeitung für die Medizin 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 309–314 [DOI: 10.007/978-3-662-56-537-7_81]
30 **Tizhoosh HR**, Pantanowitz L. Artificial Intelligence and Digital Pathology: Challenges and Opportunities. *J Pathol Inform* 2018; **9**: 38 [PMID: 30607305 DOI: 10.4103/jpi.jpi_53_18]
31 **Williams B**, Hanby A, Millican-Slater R, Verghese E, Nijhawan A, Wilson I, Besusparis J, Clark D, Snead D, Rakha E, Treanor D. Digital pathology for primary diagnosis of screen-detected breast lesions - experimental data, validation and experience from four centres. *Histopathology* 2020; **76**: 968-975 [PMID: 31994224 DOI: 10.1111/his.14079]
32 **Eloy C**, Vale J, Curado M, Polónia A, Campelos S, Caramelo A, Sousa R, Sobrinho-Simões M. Digital Pathology Workflow Implementation at IPATIMUP. *Diagnostics (Basel)* 2021; **11** [PMID: 34829458 DOI: 10.3390/diagnostics11112111]
33 **The Royal College of Pathologists**. Digital Pathology. (accessed on 20 March 2022). Available from: https://www.rcpath.org/profession/digital-pathology.html
34 **Gurcan MN**, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: a review. *IEEE Rev Biomed Eng* 2009; **2**: 147-171 [PMID: 20671804 DOI: 10.1109/RBME.2009.2034865]
35 **Rashidi HH**, Tran NK, Betts EV, Howell LP, Green R. Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods. *Acad Pathol* 2019; **6**: 2374289519873088 [PMID: 31523704 DOI: 10.1177/2374289519873088]
36 **Saxena S**, Gyanchandani M. Machine Learning Methods for Computer-Aided Breast Cancer Diagnosis Using Histopathology: A Narrative Review. *J Med Imaging Radiat Sci* 2020; **51**: 182-193 [PMID: 31884065 DOI: 10.1016/j.jmir.2019.11.001]
37 **LeCun Y**, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436-444 [PMID: 26017442 DOI: 10.1038/nature14539]
38 **Wang X**, Chen H, Gan C, Lin H, Dou Q, Tsougenis E, Huang Q, Cai M, Heng PA. Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis. *IEEE Trans Cybern* 2020; **50**: 3950-3962 [PMID: 31484154 DOI: 10.1109/TCYB.2019.2935141]
39 **Silva-Rodríguez J**, Colomer A, Naranjo V. WeGleNet: A weakly-supervised convolutional neural network for the semantic segmentation of Gleason grades in prostate histology images. *Comput Med Imaging Graph* 2021; **88**: 101846 [PMID: 33485056 DOI: 10.1016/j.compmedimag.2020.101846]
40 **Yoshida H**, Kiyuna T. Requirements for implementation of artificial intelligence in the practice of gastrointestinal pathology. *World J Gastroenterol* 2021; **27**: 2818-2833 [PMID: 34135556 DOI: 10.3748/wjg.v27.i21.2818]
41 **Cosatto E**, Laquerre PF, Malon C, Graf HP, Saito A, Kiyuna T, Marugame A, Kamijo K. Automated gastric cancer diagnosis on H and E-stained sections; training a classifier on a large scale with multiple instance machine learning. Proceedings of SPIE - Progress in Biomedical Optics and Imaging. MI: 2013 [DOI: 10.1117/12.2007047]
42 **Mattocks CJ**, Morris MA, Matthijs G, Swinnen E, Corveleyn A, Dequeker E, Müller CR, Pratt V, Wallace A; EuroGentest Validation Group. A standardized framework for the validation and verification of clinical molecular genetic tests. *Eur J Hum Genet* 2010; **18**: 1276-1288 [PMID: 20664632 DOI: 10.1038/ejhg.2010.101]
43 **Joshi**, R. ; Kruger, A.J.; Sha, L.; Kannan, M.; Khan, A.A.; Stumpe, M.C. Learning relevant H&E slide morphologies for prediction of colorectal cancer tumor mutation burden using weakly supervised deep learning. *J Clin Oncol* 2020; **38**: e15244 [DOI: 10.1200/JCO.2020.38.15_suppl.e15244]
44 **Lu MY**, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* 2021; **5**: 555-570 [PMID: 33649564 DOI: 10.1038/s41551-020-00682-w]
45 **van der Laak J**, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. *Nat Med* 2021; **27**: 775-784 [PMID: 33990804 DOI: 10.1038/s41591-021-01343-4]
46 **Colling R**, Pitman H, Oien K, Rajpoot N, Macklin P; CM-Path AI in Histopathology Working Group, Snead D, Sackville T, Verrill C. Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. *J Pathol* 2019; **249**: 143-150 [PMID: 31144302 DOI: 10.1002/path.5310]
47 **Elmore JG**, Longton GM, Carney PA, Geller BM, Onega T, Tosteson AN, Nelson HD, Pepe MS, Allison KH, Schnitt SJ, O'Malley FP, Weaver DL. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA* 2015; **313**: 1122-1132 [PMID: 25781441 DOI: 10.1001/jama.2015.1405]

48 **Amann J**, Blasimme A, Vayena E, Frey D, Madai VI; Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak* 2020; **20**: 310 [PMID: 33256715 DOI: 10.1186/s12911-020-01332-6]

49 **Tomita N**, Abdollahi B, Wei J, Ren B, Suriawinata A, Hassanpour S. Attention-Based Deep Neural Networks for Detection of Cancerous and Precancerous Esophagus Tissue on Histopathological Slides. *JAMA Netw Open* 2019; **2**: e1914645 [PMID: 31693124 DOI: 10.1001/jamanetworkopen.2019.14645]

50 **Gehrung M**, Crispin-Ortuzar M, Berman AG, O'Donovan M, Fitzgerald RC, Markowetz F. Triage-driven diagnosis of Barrett's esophagus for early detection of esophageal adenocarcinoma using deep learning. *Nat Med* 2021; **27**: 833-841 [PMID: 33859411 DOI: 10.1038/s41591-021-01287-9]

51 **Vu HT**, Lopez R, Bennett A, Burke CA. Individuals with sessile serrated polyps express an aggressive colorectal phenotype. *Dis Colon Rectum* 2011; **54**: 1216-1223 [PMID: 21904135 DOI: 10.1097/DCR.0b013e318228f8a9]

52 **Foss FA**, Milkins S, McGregor AH. Inter-observer variability in the histological assessment of colorectal polyps detected through the NHS Bowel Cancer Screening Programme. *Histopathology* 2012; **61**: 47-52 [PMID: 22486166 DOI: 10.1111/j.1365-2559.2011.04154.x]

53 **Osmond A**, Li-Chang H, Kirsch R, Divaris D, Falck V, Liu DF, Marginean C, Newell K, Parfitt J, Rudrick B, Sapp H, Smith S, Walsh J, Wasty F, Driman DK. Interobserver variability in assessing dysplasia and architecture in colorectal adenomas: a multicentre Canadian study. *J Clin Pathol* 2014; **67**: 781-786 [PMID: 25004943 DOI: 10.1136/jclinpath-2014-202177]

54 **US Preventive Services Task Force**, Davidson KW, Barry MJ, Mangione CM, Cabana M, Caughey AB, Davis EM, Donahue KE, Doubeni CA, Krist AH, Kubik M, Li L, Ogedegbe G, Owens DK, Pbert L, Silverstein M, Stevermer J, Tseng CW, Wong JB. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA* 2021; **325**: 1965-1977 [PMID: 34003218 DOI: 10.1001/jama.2021.6238]

55 **Xu Y**, Jia Z, Wang LB, Ai Y, Zhang F, Lai M, Chang EI. Large scale tissue histopathology image classification, segmentation, and visualization *via* deep convolutional activation features. *BMC Bioinformatics* 2017; **18**: 281 [PMID: 28549410 DOI: 10.1186/s12859-017-1685-x]

56 **Korbar B**, Olofson AM, Miraflor AP, Nicka CM, Suriawinata MA, Torresani L, Suriawinata AA, Hassanpour S. Deep Learning for Classification of Colorectal Polyps on Whole-slide Images. *J Pathol Inform* 2017; **8**: 30 [PMID: 28828201 DOI: 10.4103/jpi.jpi_34_17]

57 **Haj-Hassan H**, Chaddad A, Harkouss Y, Desrosiers C, Toews M, Tanougast C. Classifications of Multispectral Colorectal Cancer Tissues Using Convolution Neural Network. *J Pathol Inform* 2017; **8**: 1 [PMID: 28400990 DOI: 10.4103/jpi.jpi_47_16]

58 **Ponzio F**, Macii E, Ficarra E, Di Cataldo S. Colorectal Cancer Classification using Deep Convolutional Networks-An Experimental Study. Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2. Bioimaging, 2018: 58-66. Available from: https://www.sciencepress.org/papers/2018/66431.pdf

59 **Sena P**, Fioresi R, Faglioni F, Losi L, Faglioni G, Roncucci L. Deep learning techniques for detecting preneoplastic and neoplastic lesions in human colorectal histological images. *Oncol Lett* 2019; **18**: 6101-6107 [PMID: 31788084 DOI: 10.3892/ol.2019.10928]

60 **Iizuka O**, Kanavati F, Kato K, Rambeau M, Arihiro K, Tsuneki M. Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours. *Sci Rep* 2020; **10**: 1504 [PMID: 32001752 DOI: 10.1038/s41598-020-58467-9]

61 **Wei JW**, Suriawinata AA, Vaickus LJ, Ren B, Liu X, Lisovsky M, Tomita N, Abdollahi B, Kim AS, Snover DC, Baron JA, Barry EL, Hassanpour S. Evaluation of a Deep Neural Network for Automated Classification of Colorectal Polyps on Histopathologic Slides. *JAMA Netw Open* 2020; **3**: e203398 [PMID: 32324237 DOI: 10.1001/jamanetworkopen.2020.3398]

62 **Awan R**, Sirinukunwattana K, Epstein D, Jefferyes S, Qidwai U, Aftab Z, Mujeeb I, Snead D, Rajpoot N. Glandular Morphometrics for Objective Grading of Colorectal Adenocarcinoma Histology Images. *Sci Rep* 2017; **7**: 16852 [PMID: 29203775 DOI: 10.1038/s41598-017-16516-w]

63 **Arai T**, Sakurai U, Sawabe M, Honma N, Aida J, Ushio Y, Kanazawa N, Kuroiwa K, Takubo K. Frequent microsatellite instability in papillary and solid-type, poorly differentiated adenocarcinomas of the stomach. *Gastric Cancer* 2013; **16**: 505-512 [PMID: 23274922 DOI: 10.1007/s10120-012-0226-6]

64 **Sugimura H**. Editorial: an obsession with subtyping gastric cancer. *Gastric Cancer* 2013; **16**: 451-453 [PMID: 23483302 DOI: 10.1007/s10120-013-0243-0]

65 **Lauren P**. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An Attempt At A Histo-Clinical Classification. *Acta Pathol Microbiol Scand* 1965; **64**: 31-49 [PMID: 14320675 DOI: 10.1111/apm.1965.64.1.31]

66 **Yasuda Y**, Tokunaga K, Koga T, Sakamoto C, Goldberg IG, Saitoh N, Nakao M. Computational analysis of morphological and molecular features in gastric cancer tissues. *Cancer Med* 2020; **9**: 2223-2234 [PMID: 32012497 DOI: 10.1002/cam4.2885]

67 **Kanavati F**, Tsuneki M. A deep learning model for gastric diffuse-type adenocarcinoma classification in whole slide images. *Sci Rep* 2021; **11**: 20486 [PMID: 34650155 DOI: 10.1038/s41598-021-99940-3]

68 **Fu B**, Zhang M, He J, Cao Y, Guo Y, Wang R. StoHisNet: A hybrid multi-classification model with CNN and Transformer for gastric pathology images. *Comput Methods Programs Biomed* 2022; **221**: 106924 [PMID: 35671603 DOI: 10.1016/j.cmpb.2022.106924]

69 **Su F**, Li J, Zhao X, Wang B, Hu Y, Sun Y, Ji J. Interpretable tumor differentiation grade and microsatellite instability recognition in gastric cancer using deep learning. *Lab Invest* 2022; **102**: 641-649 [PMID: 35177797 DOI: 10.1038/s41374-022-00742-6]

70 **Fukayama M**, Goldblum JR, Miettinen M, et al Digestive System Tumours, WHO Classification of Tumours (5th edn). International Agency for Research on Cancer: Lyon, 2019. Available from: http://iarc.who.int

71 **Cancer Genome Atlas Research Network**. Comprehensive molecular characterization of gastric adenocarcinoma.

Nature 2014; **513**: 202-209 [PMID: 25079317 DOI: 10.1038/nature13480]

72 **Kim HS**, Shin SJ, Beom SH, Jung M, Choi YY, Son T, Kim HI, Cheong JH, Hyung WJ, Noh SH, Chung H, Park JC, Shin SK, Lee SK, Lee YC, Koom WS, Lim JS, Chung HC, Rha SY, Kim H. Comprehensive expression profiles of gastric cancer molecular subtypes by immunohistochemistry: implications for individualized therapy. *Oncotarget* 2016; **7**: 44608-44620 [PMID: 27331626 DOI: 10.18632/oncotarget.10115]

73 **Sohn BH**, Hwang JE, Jang HJ, Lee HS, Oh SC, Shim JJ, Lee KW, Kim EH, Yim SY, Lee SH, Cheong JH, Jeong W, Cho JY, Kim J, Chae J, Lee J, Kang WK, Kim S, Noh SH, Ajani JA, Lee JS. Clinical Significance of Four Molecular Subtypes of Gastric Cancer Identified by The Cancer Genome Atlas Project. *Clin Cancer Res* 2017; **23**: 4441-4449 [PMID: 28747339 DOI: 10.1158/1078-0432.CCR-16-2211]

74 **Grogg KL**, Lohse CM, Pankratz VS, Halling KC, Smyrk TC. Lymphocyte-rich gastric cancer: associations with Epstein-Barr virus, microsatellite instability, histology, and survival. *Mod Pathol* 2003; **16**: 641-651 [PMID: 12861059 DOI: 10.1097/01.MP.0000076980.73826.C0]

75 **Fukayama M**, Abe H, Kunita A, Shinozaki-Ushiku A, Matsusaka K, Ushiku T, Kaneda A. Thirty years of Epstein-Barr virus-associated gastric carcinoma. *Virchows Arch* 2020; **476**: 353-365 [PMID: 31836926 DOI: 10.1007/s00428-019-02724-4]

76 **Fu Y**, Jung AW, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, Yates LR, Jimenez-Linan M, Moore L, Gerstung M. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Cancer* 2020; **1**: 800-810 [PMID: 35122049 DOI: 10.1038/s43018-020-0085-8]

77 **Kather JN**, Heij LR, Grabsch HI, Loeffler C, Echle A, Muti HS, Krause J, Niehues JM, Sommer KAJ, Bankhead P, Kooreman LFS, Schulte JJ, Cipriani NA, Buelow RD, Boor P, Ortiz-Brüchle NN, Hanby AM, Speirs V, Kochanny S, Patnaik A, Srisuwananukorn A, Brenner H, Hoffmeister M, van den Brandt PA, Jäger D, Trautwein C, Pearson AT, Luedde T. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat Cancer* 2020; **1**: 789-799 [PMID: 33763651 DOI: 10.1038/s43018-020-0087-6]

78 **Schmauch B**, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, Kamoun A, Sefta M, Toldo S, Zaslavskiy M, Clozel T, Moarii M, Courtiol P, Wainrib G. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat Commun* 2020; **11**: 3877 [PMID: 32747659 DOI: 10.1038/s41467-020-17678-4]

79 **Muti HS**, Heij LR, Keller G, Kohlruss M, Langer R, Dislich B, Cheong JH, Kim YW, Kim H, Kook MC, Cunningham D, Allum WH, Langley RE, Nankivell MG, Quirke P, Hayden JD, West NP, Irvine AJ, Yoshikawa T, Oshima T, Huss R, Grosser B, Roviello F, d'Ignazio A, Quaas A, Alakus H, Tan X, Pearson AT, Luedde T, Ebert MP, Jäger D, Trautwein C, Gaisa NT, Grabsch HI, Kather JN. Development and validation of deep learning classifiers to detect Epstein-Barr virus and microsatellite instability status in gastric cancer: a retrospective multicentre cohort study. *Lancet Digit Health* 2021; **3**: e654-e664 [PMID: 34417147 DOI: 10.1016/S2589-7500(21)00133-3]

80 **Kather JN**, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, Marx A, Boor P, Tacke F, Neumann UP, Grabsch HI, Yoshikawa T, Brenner H, Chang-Claude J, Hoffmeister M, Trautwein C, Luedde T. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* 2019; **25**: 1054-1056 [PMID: 31160815 DOI: 10.1038/s41591-019-0462-y]

81 **KatherJN**, Schulte J, Grabsch HI, Loeffler C, Muti H, Dolezal J, Srisuwananukorn A, Agrawal N, Kochanny S, von Stillfried S, Boor P, Yoshikawa T, JaegerD, Trautwein C, Bankhead P, Cipriani NA, Luedde T, Pearson AT. Deep learning detects virus presence in cancer histology. bioRxiv [DOI: 10.1101/690206]

82 **Hinata M**, Ushiku T. Detecting immunotherapy-sensitive subtype in gastric cancer using histologic image-based deep learning. *Sci Rep* 2021; **11**: 22636 [PMID: 34811485 DOI: 10.1038/s41598-021-02168-4]

83 **Zheng X**, Wang R, Zhang X, Sun Y, Zhang H, Zhao Z, Zheng Y, Luo J, Zhang J, Wu H, Huang D, Zhu W, Chen J, Cao Q, Zeng H, Luo R, Li P, Lan L, Yun J, Xie D, Zheng WS, Cai M. A deep learning model and human-machine fusion for prediction of EBV-associated gastric cancer from histopathology. *Nat Commun* 2022; **13**: 2790 [PMID: 35589792 DOI: 10.1038/s41467-022-30459-5]

84 **Montavon G**, Samek W, Müller KR. Methods for interpreting and understanding deep neuralnetworks. Digit Signal Process 2018; 73: 1-15 [DOI: 10.1016/j.dsp.2017.10.011]

85 **Yang JH**, Wright SN, Hamblin M, McCloskey D, Alcantar MA, Schrübbers L, Lopatkin AJ, Satish S, Nili A, Palsson BO, Walker GC, Collins JJ. A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action. *Cell* 2019; **177**: 1649-1661.e9 [PMID: 31080069 DOI: 10.1016/j.cell.2019.04.016]

86 **Tosun AB**, Pullara F, Becich MJ, Taylor DL, Fine JL, Chennubhotla SC. Explainable AI (xAI) for Anatomic Pathology. *Adv Anat Pathol* 2020; **27**: 241-250 [PMID: 32541594 DOI: 10.1097/PAP.0000000000000264]

87 **Flinner N**, Gretser S, Quaas A, Bankov K, Stoll A, Heckmann LE, Mayer RS, Doering C, Demes MC, Buettner R, Rueschoff J, Wild PJ. Deep learning based on hematoxylin-eosin staining outperforms immunohistochemistry in predicting molecular subtypes of gastric adenocarcinoma. *J Pathol* 2022; **257**: 218-226 [PMID: 35119111 DOI: 10.1002/path.5879]

88 **Jang HJ**, Lee A, Kang J, Song IH, Lee SH. Prediction of genetic alterations from gastric cancer histopathology images using a fully automated deep learning approach. *World J Gastroenterol* 2021; **27**: 7687-7704 [PMID: 34908807 DOI: 10.3748/wjg.v27.i44.7687]

89 **Guinney J**, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, Fessler E, De Sousa E Melo F, Missiaglia E, Ramay H, Barras D, Homicsko K, Maru D, Manyam GC, Broom B, Boige V, Perez-Villamil B, Laderas T, Salazar R, Gray JW, Hanahan D, Tabernero J, Bernards R, Friend SH, Laurent-Puig P, Medema JP, Sadanandam A, Wessels L, Delorenzi M, Kopetz S, Vermeulen L, Tejpar S. The consensus molecular subtypes of colorectal cancer. *Nat Med* 2015; **21**: 1350-1356 [PMID: 26457759 DOI: 10.1038/nm.3967]

90 **Dienstmann R**, Vermeulen L, Guinney J, Kopetz S, Tejpar S, Tabernero J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat Rev Cancer* 2017; **17**: 268 [PMID: 28332502 DOI: 10.1038/nrc.2017.24]

91 **Ganesh K**, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, Diaz LA Jr. Immunotherapy in colorectal cancer:

rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol* 2019; **16**: 361-375 [PMID: 30886395 DOI: 10.1038/s41575-019-0126-x]

92 **Li K**, Luo H, Huang L, Zhu X. Microsatellite instability: a review of what the oncologist should know. *Cancer Cell Int* 2020; **20**: 16 [PMID: 31956294 DOI: 10.1186/s12935-019-1091-8]

93 **André T**, Shiu KK, Kim TW, Jensen BV, Jensen LH, Punt C, Smith D, Garcia-Carbonero R, Benavides M, Gibbs P, de la Fouchardiere C, Rivera F, Elez E, Bendell J, Le DT, Yoshino T, Van Cutsem E, Yang P, Farooqui MZH, Marinello P, Diaz LA Jr; KEYNOTE-177 Investigators. Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. *N Engl J Med* 2020; **383**: 2207-2218 [PMID: 33264544 DOI: 10.1056/NEJMoa2017699]

94 **De Smedt L**, Lemahieu J, Palmans S, Govaere O, Tousseyn T, Van Cutsem E, Prenen H, Tejpar S, Spaepen M, Matthijs G, Decaestecker C, Moles Lopez X, Demetter P, Salmon I, Sagaert X. Microsatellite instable vs stable colon carcinomas: analysis of tumour heterogeneity, inflammation and angiogenesis. *Br J Cancer* 2015; **113**: 500-509 [PMID: 26068398 DOI: 10.1038/bjc.2015.213]

95 **Sepulveda AR**, Hamilton SR, Allegra CJ, Grody W, Cushman-Vokoun AM, Funkhouser WK, Kopetz SE, Lieu C, Lindor NM, Minsky BD, Monzon FA, Sargent DJ, Singh VM, Willis J, Clark J, Colasacco C, Bryan Rumble R, Temple-Smolkin R, B Ventura C, Nowak JA. Molecular Biomarkers for the Evaluation of Colorectal Cancer: Guideline From the American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and American Society of Clinical Oncology. *Arch Pathol Lab Med* 2017; **141**: 625-657 [PMID: 28165284 DOI: 10.5858/arpa.2016-0554-CP]

96 **Punt CJ**, Koopman M, Vermeulen L. From tumour heterogeneity to advances in precision treatment of colorectal cancer. *Nat Rev Clin Oncol* 2017; **14**: 235-246 [PMID: 27922044 DOI: 10.1038/nrclinonc.2016.171]

97 **Sirinukunwattana K**, Domingo E, Richman SD, Redmond KL, Blake A, Verrill C, Leedham SJ, Chatzipli A, Hardy C, Whalley CM, Wu CH, Beggs AD, McDermott U, Dunne PD, Meade A, Walker SM, Murray GI, Samuel L, Seymour M, Tomlinson I, Quirke P, Maughan T, Rittscher J, Koelzer VH; S:CORT consortium. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut* 2021; **70**: 544-554 [PMID: 32690604 DOI: 10.1136/gutjnl-2019-319866]

98 **Echle A**, Grabsch HI, Quirke P, van den Brandt PA, West NP, Hutchins GGA, Heij LR, Tan X, Richman SD, Krause J, Alwers E, Jenniskens J, Offermans K, Gray R, Brenner H, Chang-Claude J, Trautwein C, Pearson AT, Boor P, Luedde T, Gaisa NT, Hoffmeister M, Kather JN. Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning. *Gastroenterology* 2020; **159**: 1406-1416.e11 [PMID: 32562722 DOI: 10.1053/j.gastro.2020.06.021]

99 **Kruger AJ**, Sha L, Kannan M, Joshi RP, Leibowitz BD, Zhang R, Khan AA, Stumpe M. H&E Image-based Consensus Molecular Subtype Classification of Colorectal Cancer Using Weak Labeling. *J Clin Oncol* 2020; **38**: e16097 [DOI: 10.1200/JCO.2020.38.15_suppl.e16097]

100 **Popovici V**, Budinská E, Dušek L, Kozubek M, Bosman F. Image-based surrogate biomarkers for molecular subtypes of colorectal cancer. *Bioinformatics* 2017; **33**: 2002-2009 [PMID: 28158480 DOI: 10.1093/bioinformatics/btx027]

101 **Cao R**, Yang F, Ma SC, Liu L, Zhao Y, Li Y, Wu DH, Wang T, Lu WJ, Cai WJ, Zhu HB, Guo XJ, Lu YW, Kuang JJ, Huan WJ, Tang WM, Huang K, Huang J, Yao J, Dong ZY. Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer. *Theranostics* 2020; **10**: 11080-11091 [PMID: 33042271 DOI: 10.7150/thno.49864]

102 **Bilal M**, Raza SEA, Azam A, Graham S, Ilyas M, Cree IA, Snead D, Minhas F, Rajpoot NM. Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. *Lancet Digit Health* 2021; **3**: e763-e772 [PMID: 34686474 DOI: 10.1016/S2589-7500(21)00180-1]

103 **Sun F**, Liu S, Song P, Zhang C, Liu Z, Guan W, Wang M. Impact of retrieved lymph node count on short-term complications in patients with gastric cancer. *World J Surg Oncol* 2020; **18**: 224 [PMID: 32838799 DOI: 10.1186/s12957-020-02000-9]

104 **Veta M**, Pluim JP, van Diest PJ, Viergever MA. Breast cancer histopathology image analysis: a review. *IEEE Trans Biomed Eng* 2014; **61**: 1400-1411 [PMID: 24759275 DOI: 10.1109/TBME.2014.2303852]

105 **Bhargava R**, Madabhushi A. Emerging Themes in Image Informatics and Molecular Analysis for Digital Pathology. *Annu Rev Biomed Eng* 2016; **18**: 387-412 [PMID: 27420575 DOI: 10.1146/annurev-bioeng-112415-114722]

106 **Wang X**, Chen Y, Gao Y, Zhang H, Guan Z, Dong Z, Zheng Y, Jiang J, Yang H, Wang L, Huang X, Ai L, Yu W, Li H, Dong C, Zhou Z, Liu X, Yu G. Predicting gastric cancer outcome from resected lymph node histopathology images using deep learning. *Nat Commun* 2021; **12**: 1637 [PMID: 33712598 DOI: 10.1038/s41467-021-21674-7]

107 **Hu Y**, Su F, Dong K, Wang X, Zhao X, Jiang Y, Li J, Ji J, Sun Y. Deep learning system for lymph node quantification and metastatic cancer identification from whole-slide pathology images. *Gastric Cancer* 2021; **24**: 868-877 [PMID: 33484355 DOI: 10.1007/s10120-021-01158-9]

108 **Matsushima J**, Sato T, Ohnishi T, Yoshimura Y, Mizutani H, Koto S, Ikeda JI, Kano M, Matsubara H, Hayashi H. The Use of Deep Learning-Based Computer Diagnostic Algorithm for Detection of Lymph Node Metastases of Gastric Adenocarcinoma. *Int J Surg Pathol* 2022; 10668969221113475 [PMID: 35898183 DOI: 10.1177/10668969221113475]

109 **Huang SC**, Chen CC, Lan J, Hsieh TY, Chuang HC, Chien MY, Ou TS, Chen KH, Wu RC, Liu YJ, Cheng CT, Huang YJ, Tao LW, Hwu AF, Lin IC, Hung SH, Yeh CY, Chen TC. Deep neural network trained on gigapixel images improves lymph node metastasis detection in clinical settings. *Nat Commun* 2022; **13**: 3347 [PMID: 35688834 DOI: 10.1038/s41467-022-30746-1]

110 **Kwak MS**, Lee HH, Yang JM, Cha JM, Jeon JW, Yoon JY, Kim HI. Deep Convolutional Neural Network-Based Lymph Node Metastasis Prediction for Colon Cancer Using Histopathological Images. *Front Oncol* 2020; **10**: 619803 [PMID: 33520727 DOI: 10.3389/fonc.2020.619803]

111 **Pai RK**, Hartman D, Schaeffer DF, Rosty C, Shivji S, Kirsch R, Pai RK. Development and initial validation of a deep learning algorithm to quantify histological features in colorectal carcinoma including tumour budding/poorly differentiated clusters. *Histopathology* 2021; **79**: 391-405 [PMID: 33590485 DOI: 10.1111/his.14353]

112 **Kiehl L**, Kuntz S, Höhn J, Jutzi T, Krieghoff-Henning E, Kather JN, Holland-Letz T, Kopp-Schneider A, Chang-Claude J,

Brobeil A, von Kalle C, Fröhling S, Alwers E, Brenner H, Hoffmeister M, Brinker TJ. Deep learning can predict lymph node status directly from histology in colorectal cancer. *Eur J Cancer* 2021; **157**: 464-473 [PMID: 34649117 DOI: 10.1016/j.ejca.2021.08.039]

113 **van Pelt GW**, Kjær-Frifeldt S, van Krieken JHJM, Al Dieri R, Morreau H, Tollenaar RAEM, Sørensen FB, Mesker WE. Scoring the tumor-stroma ratio in colon cancer: procedure and recommendations. *Virchows Arch* 2018; **473**: 405-412 [PMID: 30030621 DOI: 10.1007/s00428-018-2408-z]

114 **Fu M**, Chen D, Luo F, Li M, Wang Y, Chen J, Li A, Liu S. Association of the tumour stroma percentage in the preoperative biopsies with lymph node metastasis in colorectal cancer. *Br J Cancer* 2020; **122**: 388-396 [PMID: 31787749 DOI: 10.1038/s41416-019-0671-7]

115 **Geessink OGF**, Baidoshvili A, Klaase JM, Ehteshami Bejnordi B, Litjens GJS, van Pelt GW, Mesker WE, Nagtegaal ID, Ciompi F, van der Laak JAWM. Computer aided quantification of intratumoral stroma yields an independent prognosticator in rectal cancer. *Cell Oncol (Dordr)* 2019; **42**: 331-341 [PMID: 30825182 DOI: 10.1007/s13402-019-00429-z]

116 **Hong Y**, Heo YJ, Kim B, Lee D, Ahn S, Ha SY, Sohn I, Kim KM. Deep learning-based virtual cytokeratin staining of gastric carcinomas to measure tumor-stroma ratio. *Sci Rep* 2021; **11**: 19255 [PMID: 34584193 DOI: 10.1038/s41598-021-98857-1]

117 **Kather JN**, Poleszczuk J, Suarez-Carmona M, Krisam J, Charoentong P, Valous NA, Weis CA, Tavernar L, Leiss F, Herpel E, Klupp F, Ulrich A, Schneider M, Marx A, Jäger D, Halama N. In Silico Modeling of Immunotherapy and Stroma-Targeting Therapies in Human Colorectal Cancer. *Cancer Res* 2017; **77**: 6442-6452 [PMID: 28923860 DOI: 10.1158/0008-5472.CAN-17-2006]

118 **Scheer R**, Baidoshvili A, Zoidze S, Elferink MAG, Berkel AEM, Klaase JM, van Diest PJ. Tumor-stroma ratio as prognostic factor for survival in rectal adenocarcinoma: A retrospective cohort study. *World J Gastrointest Oncol* 2017; **9**: 466-474 [PMID: 29290917 DOI: 10.4251/wjgo.v9.i12.466]

119 **Kather JN**, Halama N, Jaeger D. Genomics and emerging biomarkers for immunotherapy of colorectal cancer. *Semin Cancer Biol* 2018; **52**: 189-197 [PMID: 29501787 DOI: 10.1016/j.semcancer.2018.02.010]

120 **Weis CA**, Kather JN, Melchers S, Al-Ahmdi H, Pollheimer MJ, Langner C, Gaiser T. Automatic evaluation of tumor budding in immunohistochemically stained colorectal carcinomas and correlation to clinical outcome. *Diagn Pathol* 2018; **13**: 64 [PMID: 30153844 DOI: 10.1186/s13000-018-0739-3]

121 **Kather JN**, Krisam J, Charoentong P, Luedde T, Herpel E, Weis CA, Gaiser T, Marx A, Valous NA, Ferber D, Jansen L, Reyes-Aldasoro CC, Zörnig I, Jäger D, Brenner H, Chang-Claude J, Hoffmeister M, Halama N. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Med* 2019; **16**: e1002730 [PMID: 30677016 DOI: 10.1371/journal.pmed.1002730]

122 **Shapcott M**, Hewitt KJ, Rajpoot N. Deep Learning With Sampling in Colon Cancer Histology. *Front Bioeng Biotechnol* 2019; **7**: 52 [PMID: 30972333 DOI: 10.3389/fbioe.2019.00052]

123 **Sirinukunwattana K**, Snead D, Epstein D, Aftab Z, Mujeeb I, Tsang YW, Cree I, Rajpoot N. Novel digital signatures of tissue phenotypes for predicting distant metastasis in colorectal cancer. *Sci Rep* 2018; **8**: 13692 [PMID: 30209315 DOI: 10.1038/s41598-018-31799-3]

124 **Swiderska-Chadaj Z**, Pinckaers H, van Rijthoven M, Balkenhol M, Melnikova M, Geessink O, Manson Q, Sherman M, Polonia A, Parry J, Abubakar M, Litjens G, van der Laak J, Ciompi F. Learning to detect lymphocytes in immunohistochemistry with deep learning. *Med Image Anal* 2019; **58**: 101547 [PMID: 31476576 DOI: 10.1016/j.media.2019.101547]

125 **Zhao K**, Li Z, Yao S, Wang Y, Wu X, Xu Z, Wu L, Huang Y, Liang C, Liu Z. Artificial intelligence quantified tumour-stroma ratio is an independent predictor for overall survival in resectable colorectal cancer. *EBioMedicine* 2020; **61**: 103054 [PMID: 33039706 DOI: 10.1016/j.ebiom.2020.103054]

126 **Zhao K**, Wu L, Huang Y, Yao S, Xu Z, Lin H, Wang H, Liang Y, Xu Y, Chen X, Zhao M, Peng J, Liang C, Li Z, Li Y, Liu Z. Deep learning quantified mucus-tumor ratio predicting survival of patients with colorectal cancer using whole-slide images. *Precis Clin Med* 2021; **4**: 17-24 [PMID: 35693123 DOI: 10.1093/pcmedi/pbab002]

127 **Meier A**, Nekolla K, Hewitt LC, Earle S, Yoshikawa T, Oshima T, Miyagi Y, Huss R, Schmidt G, Grabsch HI. Hypothesis-free deep survival learning applied to the tumour microenvironment in gastric cancer. *J Pathol Clin Res* 2020; **6**: 273-282 [PMID: 32592447 DOI: 10.1002/cjp2.170]

128 **Huang B**, Tian S, Zhan N, Ma J, Huang Z, Zhang C, Zhang H, Ming F, Liao F, Ji M, Zhang J, Liu Y, He P, Deng B, Hu J, Dong W. Accurate diagnosis and prognosis prediction of gastric cancer using deep learning on digital pathological images: A retrospective multicentre study. *EBioMedicine* 2021; **73**: 103631 [PMID: 34678610 DOI: 10.1016/j.ebiom.2021.103631]

129 **Jiang Y**, Xie J, Han Z, Liu W, Xi S, Huang L, Huang W, Lin T, Zhao L, Hu Y, Yu J, Zhang Q, Li T, Cai S, Li G. Immunomarker Support Vector Machine Classifier for Prediction of Gastric Cancer Survival and Adjuvant Chemotherapeutic Benefit. *Clin Cancer Res* 2018; **24**: 5574-5584 [PMID: 30042208 DOI: 10.1158/1078-0432.CCR-18-0848]

130 **Jiang Y**, Xie J, Huang W, Chen H, Xi S, Han Z, Huang L, Lin T, Zhao LY, Hu YF, Yu J, Cai SR, Li T, Li G. Tumor Immune Microenvironment and Chemosensitivity Signature for Predicting Response to Chemotherapy in Gastric Cancer. *Cancer Immunol Res* 2019; **7**: 2065-2073 [PMID: 31615816 DOI: 10.1158/2326-6066.CIR-19-0311]

131 **Wang Y**, Wang YG, Hu C, Li M, Fan Y, Otter N, Sam I, Gou H, Hu Y, Kwok T, Zalcberg J, Boussioutas A, Daly RJ, Montúfar G, Liò P, Xu D, Webb GI, Song J. Cell graph neural networks enable the precise prediction of patient survival in gastric cancer. *NPJ Precis Oncol* 2022; **6**: 45 [PMID: 35739342 DOI: 10.1038/s41698-022-00285-5]

132 **Bychkov D**, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, Walliander M, Lundin M, Haglund C, Lundin J. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep* 2018; **8**: 3395 [PMID: 29467373 DOI: 10.1038/s41598-018-21758-3]

133 **Skrede OJ**, De Raedt S, Kleppe A, Hveem TS, Liestøl K, Maddison J, Askautrud HA, Pradhan M, Nesheim JA, Albregtsen F, Farstad IN, Domingo E, Church DN, Nesbakken A, Shepherd NA, Tomlinson I, Kerr R, Novelli M, Kerr

DJ, Danielsen HE. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet* 2020; **395**: 350-360 [PMID: 32007170 DOI: 10.1016/S0140-6736(19)32998-8]

134 **Jiang D**, Liao J, Duan H, Wu Q, Owen G, Shu C, Chen L, He Y, Wu Z, He D, Zhang W, Wang Z. A machine learning-based prognostic predictor for stage III colon cancer. *Sci Rep* 2020; **10**: 10333 [PMID: 32587295 DOI: 10.1038/s41598-020-67178-0]

135 **Kotsenas AL**, Balthazar P, Andrews D, Geis JR, Cook TS. Rethinking Patient Consent in the Era of Artificial Intelligence and Big Data. *J Am Coll Radiol* 2021; **18**: 180-184 [PMID: 33413897 DOI: 10.1016/j.jacr.2020.09.022]

136 **Inoue T**, Yagi Y. Color standardization and optimization in whole slide imaging. *Clin Diagn Pathol* 2020; **4** [PMID: 33088926 DOI: 10.15761/cdp.1000139]

137 **Yoshida H**, Yokota H, Singh R, Kiyuna T, Yamaguchi M, Kikuchi S, Yagi Y, Ochiai A. Meeting Report: The International Workshop on Harmonization and Standardization of Digital Pathology Image, Held on April 4, 2019 in Tokyo. *Pathobiology* 2019; **86**: 322-324 [PMID: 31707388 DOI: 10.1159/000502718]

138 **Janowczyk A**, Basavanhally A, Madabhushi A. Stain Normalization using Sparse AutoEncoders (StaNoSA): Application to digital pathology. *Comput Med Imaging Graph* 2017; **57**: 50-61 [PMID: 27373749 DOI: 10.1016/j.compmedimag.2016.05.003]

139 **Vahadane A**, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, Steiger K, Schlitter AM, Esposito I, Navab N. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. *IEEE Trans Med Imaging* 2016; **35**: 1962-1971 [PMID: 27164577 DOI: 10.1109/TMI.2016.2529665]

140 **Janowczyk A**, Zuo R, Gilmore H, Feldman M, Madabhushi A. HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides. *JCO Clin Cancer Inform* 2019; **3**: 1-7 [PMID: 30990737 DOI: 10.1200/CCI.18.00157]

141 **Senaras C**, Niazi MKK, Lozanski G, Gurcan MN. DeepFocus: Detection of out-of-focus regions in whole slide digital images using deep learning. *PLoS One* 2018; **13**: e0205387 [PMID: 30359393 DOI: 10.1371/journal.pone.0205387]

142 **Arora A**, Arora A. Pathology training in the age of artificial intelligence. *J Clin Pathol* 2021; **74**: 73-75 [PMID: 33020175 DOI: 10.1136/jclinpath-2020-207110]

143 **Hekler A**, Utikal JS, Enk AH, Solass W, Schmitt M, Klode J, Schadendorf D, Sondermann W, Franklin C, Bestvater F, Flaig MJ, Krahl D, von Kalle C, Fröhling S, Brinker TJ. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *Eur J Cancer* 2019; **118**: 91-96 [PMID: 31325876 DOI: 10.1016/j.ejca.2019.06.012]

Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: bpgoffice@wjgnet.com
Help Desk: https://www.f6publishing.com/helpdesk
https://www.wjgnet.com