IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 October 2024, accepted 13 October 2024, date of publication 18 October 2024, date of current version 14 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3483432

TOPICAL REVIEW

AI in Endoscopic Gastrointestinal Diagnosis: A Systematic Review of Deep Learning and Machine Learning Techniques

JOVITA RELASHA LEWIS [1], SAMEENA PATHAN [1], PREETHAM KUMAR [1], AND CIFHA CRECIL DIAS [2]

[1]Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
[2]Department of Biomedical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India

Corresponding author: Sameena Pathan (sameena.bp@manipal.edu)

ABSTRACT Gastrointestinal (GI) diseases are most common worldwide and the death rate can be reduced by early detection. Endoscopy is widely regarded as the gold standard for diagnosing and managing digestive disorders, affecting both the upper and lower GI tracts. Endoscopy is performed to uncover biopsy tissues used to check the presence of cancerous or benign cells, Helicobacter pylori (H. pylori) infection, or perform colonoscopy in case of the removal of polyps. A systematic review was conducted on databases like PubMed, Scopus, Google Scholar, and IEEE Explore, including research papers published up to May 2023, through the systematic search, 33 papers were identified. This review offers valuable insights to physicians and technological guidance to future researchers by examining GI tract diseases. It provides a detailed analysis of Machine learning (ML) techniques like preprocessing, segmentation, feature extraction, and classification. Additionally, Deep Learning (DL) approaches like transfer learning (TL), Convolution Neural Networks (CNN), optimization, transformer, and reinforcement learning have been analyzed for GI diagnosis. The DL approach has increased its use in GI diseases and CNN was the most commonly used architecture. Lastly, the review highlights the research published in the specialized GI fields and provides technological suggestions and insights for future research prospects. Overall, this study broadens the body of knowledge regarding the existing Artificial Intelligence (AI) techniques in gastroenterology as a manual for creating and assessing AI models.

INDEX TERMS Gastrointestinal disease, artificial intelligence, endoscopy, deep learning, machine learning, colon cancer, gastric cancer, early diagnosis, preprocessing, segmentation, feature extraction, classification, CNN, transfer learning, optimization.

I. INTRODUCTION

GI diseases are increasing globally. According to a survey conducted by the Rome Foundation Global Study International, it was found that about 40% of people globally have Functional GI disorders, which impact healthcare utilization and quality of life [1]. As per the Surveillance, Epidemiology, and End Results (SEER) website, the 5-year relative survival from 2018 to 2022 conveys the death date of colorectal cancer (CRC) as 8.6%, stomach cancer at 1.8%, small intestine cancer at 0.3%, pancreatic cancer at 8.2% and esophageal cancer 2.6% [2], [3], [4], [5], [6].

At present, GI cancer has a poor prognosis, as depicted by the 5-year relative survival rates summarized in Figure 1. Fortunately, advancements in technology and lifestyle changes result in an increasing number of cases being detected yearly. Overweight and obesity were found to be related to GI problems [7]. The survival rate of GI reduces when the cancer is predicated at an advanced stage [8]. Endoscopy is the gold standard for analyzing GI diseases for nodularity, erythema, erosions, polyps, and lesions [9]. A biopsy of the tissue is taken which indicates the presence

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.

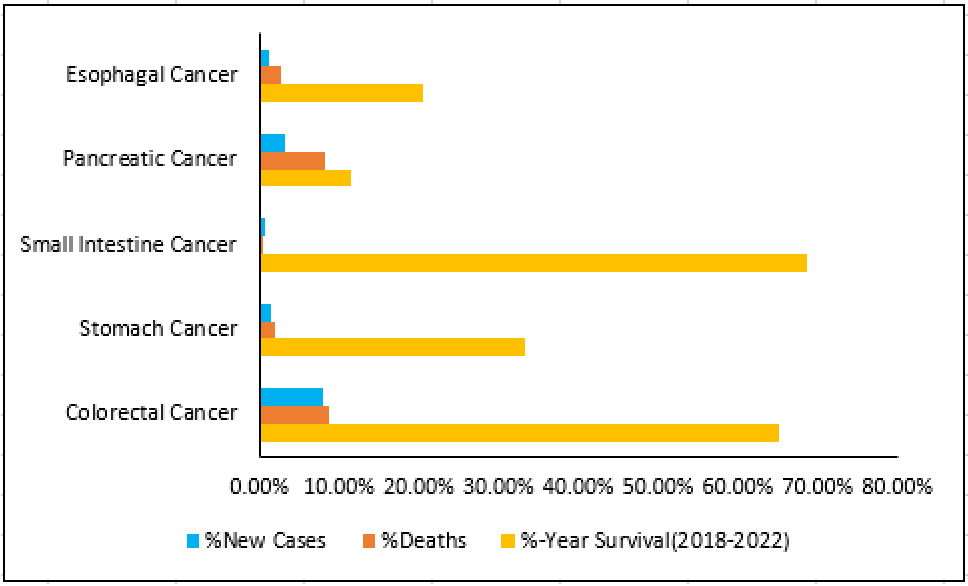


FIGURE 1. Summary of the 5-year relative survival according to SEER website.

of H.pylori infection which is a carcinogen for stomach cancer [10]. Therefore, early prediction using a Computer Aided Diagnosis (CAD) diagnosis system would help the expert predict lesions or abnormalities without performing a biopsy [11]. The ability of AI to fully analyze images, learn and adapt over time, provide decision assistance, and minimize human error contributes to an earlier diagnosis and increased accuracy in the diagnosis of gastric cancer (GC) by reducing false positive rates, which in turn improves patient outcomes [12], [13]. The rest of the article is arranged as follows:- Section I highlights the Introduction to the necessity of AI in detecting GI diseases. Section II considers GI Tract: An overview of the GI tract with the significance of early detection and a background of CAD methods used. It focuses on the contribution of the work together with a research focus. Section III reflects the related reviews and the literature Search Strategy used in the choice of the papers along with the inclusion and exclusion criteria. Section IV analyses the ML and DL approaches concerning pre-processing, segmentation, feature extraction, classification, TL, CNN, optimization, transformer, and reinforcement learning. Section V provides answers to research questions along with the available datasets. Section VI provides limitations of the previous study along with solution and Section VII provides conclusions.

A. RESEARCH AIM

This systematic literature review aims to provide comprehensive knowledge addressing the GI tract and the datasets used to identify GI diseases. An in-depth assessment of ML techniques like pre-processing, segmentation, feature extraction, and classification, in addition to DL approaches such as TL, CNN, optimization, transformer, and reinforcement learning was accomplished.

B. OBJECTIVE

The key objectives regarding comprehensive analysis of AI in the diagnosis of GI diseases through endoscopic procedures are: -

- To evaluate the performance of various ML and DL models as well as newer models like transformers and reinforcement learning to improve the diagnosis accuracy of GI diseases.
- To analyze the techniques like preprocessing, segmentation, feature extraction, classification, TL, and optimization in GI disease diagnosis.
- To provide a comprehensive overview of existing public datasets addressing the limitations and methods to improve the generalization of the datasets.
- To investigate the potential of AI in the early detection of GI diseases.
- To explore the practical challenges in the deployment of real-time GI diagnosis systems.

C. NOVELTY

- The effect of AI on the diagnosis of GI diseases through the evaluation of 33 publications is demonstrated with an overview of ML and DL techniques.
- The focus was on ML techniques like preprocessing, segmentation, and classification along with DL techniques like CNN, TL, and optimization which previous researchers have not addressed totally.
- This review extends the scope of GI diagnosis from CNN to transformers and reinforcement learning.
- The focus of real-time deployment challenges in GI diagnosis is discussed.
- Future researchers will benefit from valuable insights regarding datasets and clinical implications.

D. OVERVIEW OF THE GI DISEASES

- **CRC** starts in the colon or rectum. Colonoscopy is used for screening and diagnosis allowing direct visualization of cancerous and pre-cancerous polyps. Early detection through endoscopy will improve the outcomes.
- **GC** originates from the stomach lining. Gastroscopy is used to detect and diagnose GC. The gastric lesions can be detected using a gastroscopy.
- **Peptic Ulcers** are sores that develop inside the inner lining of the stomach or small intestine. Peptic ulcers can be detected by the physician during a gastroscopy.
- **H. pylori** is a bacterium that infects the stomach lining leading to chronic gastritis, and peptic ulcer, and is a leading carcinogen for GC.
- **Irritable Bowel Syndrome (IBS)** is a GI disorder characterized by chronic abdomen pain, bloating, and diarrhea.
- **Ulcerative Colitis (UC)** causes long-lasting inflammation and sores in the lining of the colon and rectum.
- **Celiac Disease (CD)** is an autoimmune disorder caused by ingestion of protein leading to damage of the small intestine.

  Over is performed by the physicians to detect any lesions, colon, ulcer, or inflammation and accordingly suggests a biopsy test.

II. BACKGROUND STUDY

A. AN OVERVIEW OF THE GASTROINTESTINAL TRACT

The GI Tract begins with the mouth which is the starting point of the 9-meter-long digestive tract that terminates at the anus. Figure 2 shows the upper and lower GI tracts that make up this division [14].

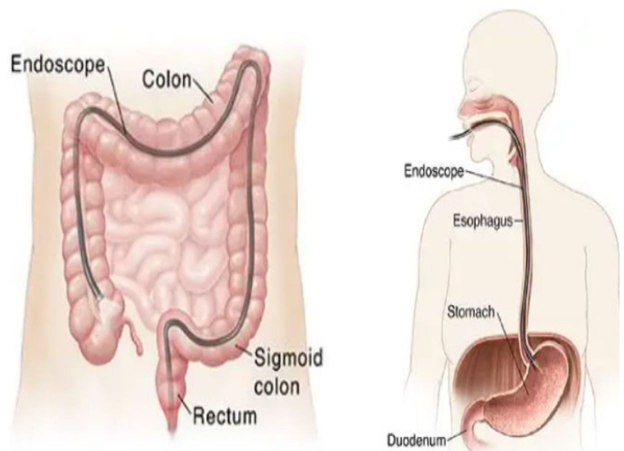


FIGURE 2. Gastrointestinal (GI) tract [14].

The esophagus, stomach, and duodenum comprise the upper GI system, while the small, large intestine and rectum comprise the lower GI system. The endoscopic procedure for GI analysis is performed by placing long, thin, and flexible tubes through the mouth to which a small camera is attached. In upper gastroscopy, the physician inserts the endoscope into the mouth through the stomach; in colonoscopy, the physician inserts the endoscope into the rectum through the large intestine. A primary and widely used method for detecting early stomach cancer is white-light endoscopy (WLE). At present medical professionals use esophagogastroduodenoscopy for monitoring GC [15]. Moreover, it is observed that the detection rate for Early Gastric Cancer (EGC) is inadequate [16]. Previous research indicates that 20–40% of anomalies were missed in EGC patients [17]. Even though, the initial stages are frequently missed, later stages of stomach cancer are simpler to diagnose. The qualitative identification of lesions has improved because of image-enhanced endoscopy (IEE) [18]. The primary and accepted approach for detecting EGC is WLE.

Irritable Bowel Disease (IBD) can be analyzed by medical decision support systems, cross-section image analysis, natural language processing for automated medical text learning, endoscopic image interpretation, and scoring [19]. Endoscopic examination for UC varies across observers and within observers, and biopsies are frequently taken for histologic analysis [20]. CD may be increasing in frequency, and a sizable portion of cases go untreated [21].

B. SIGNIFICANCE OF EARLY DETECTION

The impact of AI on the diagnosis of the GI system is addressed in this work. Early detection and diagnosis play a key role in achieving positive outcomes for GI diseases. Esophagus cancer is usually detected at an advanced stage leading to various difficulties in the treatment. Moreover, White Light Imaging (WLI), is most commonly used and often ineffective [22]. It has been observed that stomach cancer is likely to be present during GI mucus atrophy [23], [24], which gradually leads to stomach cancer [25]. According to the Japanese Stomach Cancer Treatment Guidelines, early identification and treatment are essential for a good prognosis [26]. Patients suffering from chronic atrophic gastritis (CAG) have a lower chance of developing stomach cancer while receiving treatment focused on eliminating H.pylori infection. To prevent the growth and advancement of stomach cancer, it is important to quickly diagnose cases of atrophic gastritis [17]. Early identification of polyps smaller than 5mm can significantly lower the risk of developing CRC [22], [27]. Peptic ulcers, which are digestive system injuries caused by stomach acid, can also develop in the esophagus, proximal duodenum, or stomach [28].

AI can play a crucial role in significantly predicting small polyps that usually go unseen and by predicting H.pylori infection at an early stage. However, an early prediction can reduce further symptoms of colorectal and stomach cancer respectively.

C. BACKGROUND OF CAD METHODS

Several CAD methods were proposed in the 1990s for GI analysis. Early attempts employed region-growing algorithms and a pixel-based technique for picture segmentation, to extract the outlines of the large intestine lumen and identify lower GI tract disease. By the end of the 1990s, most research efforts focused on combining intelligent pattern classification with texture, color, or mixed analytic methods to help find lesions in fixed endoscopic images [21]. More recently, automated CAD models in the identification of GI anomalies have been developed by several AI (computer vision) researchers [29], [30], [31], [32]. The CAD models blend hand-crafted features and deep convolution neural network (CNN) characteristics to recognize features in WLI. Color features [33], [34], textures [35], point attributes [36], as well as a histogram of orientation gradient (HOG) features [37] have all been used in various quantitative investigations. Some researchers have used deep features to recognize GI disorders, thanks to recent advancements in the deep-learning field. The researchers in [35] have included several processes for classifying stomach diseases. With fine-tuned deep CNN models, namely Inception v3 and DenseNet-201, deep features are created by TL.

D. MOTIVATION

The traditional method for diagnosing GI diseases allows the physician to use an endoscopy, a long flexible tube with a lens at one end and a video camera at the other. There is a provision to inspect the inside of the GI tract without serious surgery. An AI-assisted gastroenterology would perform better since they are based on the algorithms in the ML and DL models and not operator-dependent. The performance of an endoscopist varies based on factors like experience, fatigue, and stress. Going through the endoscopic video and identifying the lesions takes time and effort. However, when these tasks are performed by the AI they become machine-dependent and can provide the physician with a second option. These decision models are trained to recognize polyps in CRC [38]. They are even capable of extracting features like pit patterns and micro-vascular patterns in narrow-band imaging (NBI) and Magnifying Endoscopy (ME) [26]. In GI diseases like CRC, Barrett's Esophagus, and UC, AI has already shown a promising role. AI can potentially reduce the burden on health care systems by automating tasks like preprocessing, segmentation, feature extraction, and classification also providing a helping hand for the physician to determine the lesions to be removed in polyps, GC, and ulcers into different stages. The physician couldn't predict H.pylori infection without a follow-up biopsy report. An AI-assisted system would help to predict if the lesions are malignant or benign without performing a biopsy. Therefore, a collaboration between the physician and technical expert would result in a better diagnosis of GI. The gaps like difficulty in labeling a dataset as it requires medical experts and the unavailability of public datasets, for instance, H.pylori infection, would hinder the area of exploration of an interested researcher. The literature on CAD methods for GI disease diagnosis based on both ML and DL models was carefully examined to address this gap for future researchers.

E. CONTRIBUTION TO THE WORK

This systematic literature review (SLR) focuses on the different ML and DL techniques used in GI diseases like GC, H.pylori, CD, and Colorectal polyps. A brief description of the GI tract, the necessity for feature extraction, segmentation, pre-processing, and classification algorithms on GI images, and the variety of datasets available are all briefly covered in the review. By reviewing the existing literature, this study aims to obtain the possible challenges in implementing these techniques and the recent shift towards DL models like CNN and TL. Our study includes papers from Jan 2017 to May 2023 to impart advance in GI diagnosis. The key aspects of this SLR are as follows:-

- Emphasizes vital areas of interest from the existing literature by following data extraction techniques.
- Summarize different ML and DL techniques that were used in the GI diagnosis to provide personalized treatment.
- Recognizing datasets that were used in training the model.
- Examining the possible challenges associated with the clinical implication of the ML and DL approaches.
- Determining the possible gaps in the research and exploring directions for future research.

F. RESEARCH FOCUS

The literature review focuses on the vital areas related to ML and DL techniques in endoscopic image analysis: ML techniques like preprocessing, segmentation, feature extraction, and classification as well as DL techniques like CNN, TL, and optimization, the datasets available, the gaps in the research and the future directions. A review of the available literature will define this scope, offer insights into important areas in research, and highlight opportunities and gaps for further studies to provide a thorough overview of the current state of AI-assisted GI-related models. The following are the research questions addressed in this SLR:-

- RQ1. What recent advancements have been made in ML and DL architectures? Is DL better than ML?
- RQ2. How can combining DL and ML models improve the accuracy in feature extraction and classification, providing a personalized treatment?
- RQ3. Which datasets are being used in the model training process?
- RQ4. What are the Key Findings for Future Researchers?
- RQ5. What are the clinical Implications and open Challenges?

The development of an efficient detection technique can be aided by the literature evaluation's emphasis on these areas, which can highlight the unique potential and problems related to GI detection.

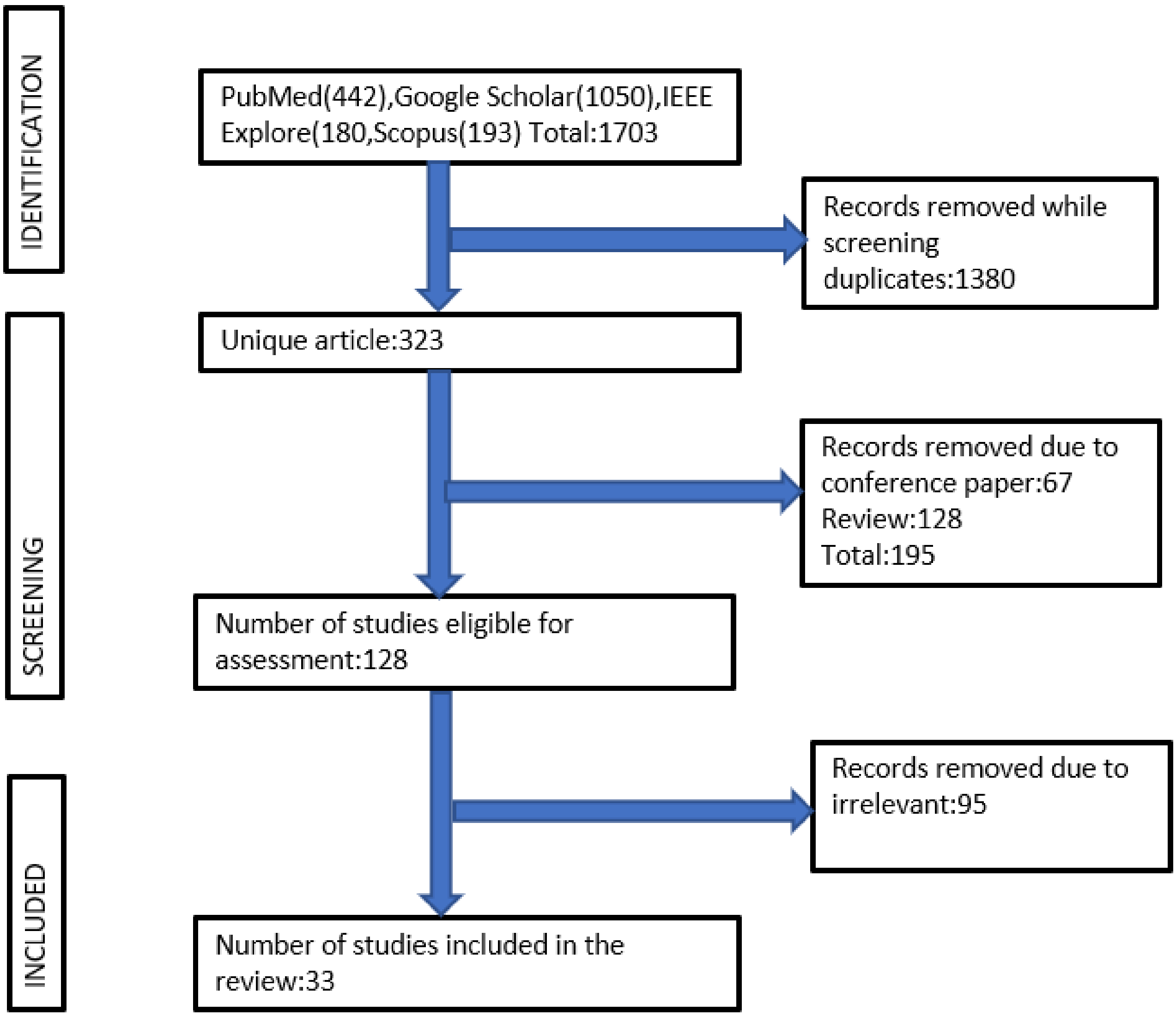


FIGURE 3. Summary of studies included using PRISMA guidelines.

III. STATE OF ART AND REVIEW OF METHODS

A. LITERATURE REVIEW

Recently, AI has been extensively used in the identification and diagnosis of GI disease analysis. The goal of this review paper is to outline the various ML as well as DL techniques that can be effectively utilized in CAD of the GI tract. Our literature search considered peer-reviewed papers that implemented AI technology for the detection and diagnostics of GI diseases, namely esophagus cancer, stomach cancer, H.pylori, CRC, peptic ulceration, IBD, UC, and CD.

B. LITERATURE SEARCH STRATEGY

A systematic literature search on the automated detection and diagnosis of GI diseases including endoscopic images, following the PRISMA guidelines [46]. In total, we gathered 1703 articles from the following journal databases: PubMed, Google Scholar, IEEE (Institute of Electrical and Electronics Engineers) Explore, and Scopus. We used the Boolean search that included terms such as "endoscopic colorectal cancer", "endoscopic gastric cancer", "peptic ulcer", "helicobacter pylori", "irritable bowel disease", "ulcerative colitis", "celiac disease", combined with the keyword "Artificial Intelligence". Our analysis of the peer-reviewed published literature spanned from January 2017 to May 2023 in topics relevant to the GI arena. After screening for duplicates, we obtained 323 unique articles. We then excluded conference papers, review papers, and those that were irrelevant, leaving a total of 60 eligible papers for further review. We narrowed down this list further by excluding additional conference and irrelevant papers, resulting in a final set of 33 papers for our review process, as depicted in Figure 3. The comparison of our proposed work with related review papers is done. Our review covers preprocessing, segmentation, classification,

TABLE 1. Comparisons of the proposed review with related reviews.

Ref	Images included	Pre-processing	Segmentation	Classification	CNN	Transfer learning	Optimisation	Study Period	Disease Type
ZiyiJin et al.,2022 [39]	Endoscopic	-	yes	yes	yes	yes	-	40 related papers	GC, H.pylori. Precancerous conditions
John Gubatan et al. ,2021 [40]	endoscopy and endomicroscopy	-	yes	yes	yes	yes	-	2015-2021	IBD
Patel et al.,2021 [41]	-	-	-	-	yes	-	-	-	Paediatric gastroenterology,CD, IBD, polyps
Ioannis Tziortziotis et al.,2021 [42]	-	-	-	-	yes	-	-	-	Capsule endoscopy
ChrysanthosD Christou et al.,2021 [43]	-	-	-	yes	yes	-	-	-	GC, Crohn's disease, peptic ulcer
Ken Namikawa et al.,2020 [44]	yes	-	-	-	yes	-	-	2016-2020	Ulcer, colorectal polys, GC
Young Joo Yang et al.,2020 [45]	yes	-	-	yes	yes	-	-	-	Ulcer, polyps, GC
Proposed Review	yes	yes	yes	yes	yes	yes	yes	2017-2023	GC H.pylori, Crohn's Disease, Colorectal polyps

IBD: Inflammatory Bowel Disease, GC: Gastric Cancer, CD: Celiac Disease.

CNN, TL, Optimisation, transformers, and reinforcement learning which have not been covered altogether in a single review as shown in Table 1.

IV. DIFFERENT APPROACHES FOR AI IN GASTROENTEROLOGY

A. ML APPROACHES

ML is a fundamental, multidisciplinary area of AI that has drawn inspiration from many different academic fields, such as computer science, statistics, and biology. Finding patterns in data that lead to useful insights is the fundamental purpose of ML. It encompasses a wide range of algorithms that all can learn from the past to perform better in the future [47]. The four basic families of ML algorithms are supervised, unsupervised, semi-supervised, and reinforcement learning.

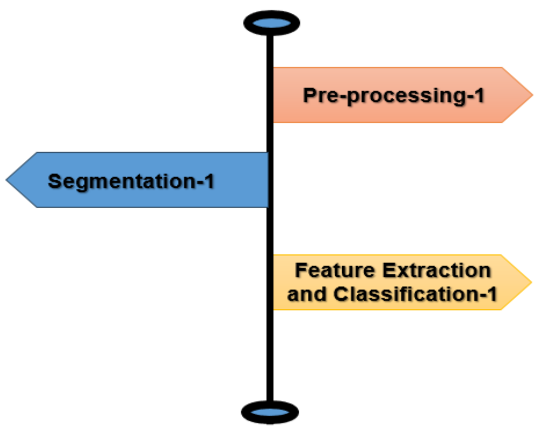


FIGURE 4. Mapping of different ML approaches and the number of studies.

- **Supervised ML:**For this aspect, the data that is used to train the model must be tagged with the correct responses, which necessitates knowledge of the algorithm's potential outputs. For instance, the label to train a system to recognize a particular type of liver cancer might be a cancer's pathologic findings or genetic data. Two categories of learning problems solved using supervised learning are classification and regression. Support vector machines, decision trees, K-nearest neighbors, and logistic regression are some of the commonly used supervised algorithms for this purpose.
- **Unsupervised ML:** The algorithm provides access to unlabeled training data, to produce labels that will effectively categorize it. The unsupervised counterpart examines the underlying similarities in the imagery, and classifies them into appropriate groups, giving each group a sublabel. This type of technique is

mostly utilized for association and clustering. K-means, hierarchical clustering, and self-organizing maps are a few examples of clustering algorithms. The principal component analysis (PCA), non-negative matrix factorization, and singular value decomposition are the basic techniques for association.

- **Semi-supervised ML:**This technique combines the strength of unsupervised and supervised learning into one method. For effective classification, it often makes use of the vast volume of unlabeled data. There is a plentiful amount of unlabeled data but little labeled data in most datasets, which may be costly to produce. Since data can include both labeled and unlabeled components for training, semi-supervised learning algorithms are utilized for the same purposes as supervised learning.
- **Reinforcement learning:** It relies on part upon behavioral psychology and is a different type of ML [48]. Such algorithms are trained to match actions to circumstances, to maximize the reward or the feedback signal. It should be noted that, unlike other kinds of ML, techniques from this family do not obtain instructions on the course of action to follow. Instead, they must experiment to determine which course of action results in the greatest reward. However, in this kind of ML, driven by experience, sequential decision-making takes precedence over pattern recognition.

In this section, we provide a summary of the ML approaches commonly utilized for GI diseases. Specifically, we focus on three application areas: pre-processing, segmentation, and feature extraction with classification as shown in Figure 4.

1) PRE-PROCESSING

"Image pre-processing" describes operations on imagery at a low level of abstraction. When an endoscopic procedure is performed, the operator may find that images may contain extraneous substances such as air bubbles, they may be over- or under-exposed at selected regions, and there can be the presence of extraneous opaque fluids, all of which may cause GI diseases to be misdiagnosed at an early stage. The experience of the physician also matters in the identification of GI tract diseases.

A summary of the various types of imagery pre-processing methods that are accessible include [22]: brightness conversions and adjustments for pixels, geometric modifications, filtering, and image restoration. A system called MAPGI (Modular Adaptive Pre-processing of Gastrointestinal tract Images) was created by the authors in [49]. It may be used to conduct several image processing tasks, such as sharpening or removing edges, improving contrast, mapping using color, scaling, and filtering. Figure 5A displays an original image taken from the Kvasir collection that has been cropped or altered using MAPGI in Figure 5B. It shows the Pylorus original picture from the Kvasir collection with a contrast-enhanced image and exceptionally low mean pixel value Figure 6A. After enhancing contrast, MAPGI filtering can be used to lower out-of-band noise using low, high, and band-pass filtering. Only the Y component of a YUV color image is subjected to filtering, as shown in Figure 6B.

The authors in [50] developed a soft-tissue frame of [110,190] Hounsfield Units (HU) to rescale image intensities within the range [0-255]. This approach enhanced the contrast of soft tissues and provided additional information about the status of abdominal organs. Tumors in the dimension of 28×28 voxels were surgically removed. Table 2 shows the difference in accuracy and F1 scores resulting.

2) SEGMENTATION

Segmentation is a widely used technique in digital image processing to separate an image into multiple portions or areas based on certain characteristics of the image pixels [52]. The foreground and background in an image can be distinguished by segmentation, and pixels can be grouped according to their similarity in color or morphology [53]. During endoscopic procedures, the diagnosis rate is often dependent on the experience of the physician. Segmentation is necessary to divide an image based on features like color, texture, and brightness. In a colonoscopy, the polyps/colons are identified and removed manually by the physician. Segmentation can be assistive in identifying the region of interest (ROI), thereby providing the physician with information regarding the area to be segregated. This would be particularly useful when polyps of size less than 5mm are to be identified, to improve early diagnosis and to avoid unnecessary complications from a missed polyp converting to possible CRC. Segmentation is also helpful in the detection of GC, having the capability of differentiating EGC from non-EGC using a procedure termed magnetic endoscopy-NBI [51]. Researchers in [54] modified Mask-RCNN to use in the segmentation of ulcerous regions. It has been suggested that a deep feature selection method can be helpful to classify the regions. Experiments were conducted using a private dataset as well as the Constant Voltage (CV) Clinic DB, which includes images showing ulceration, blood loss, presence of polyps, or healthy patients [55], as depicted in Figure 7. The segmentation performed by different classifiers is shown in Table 3.

3) FEATURE EXTRACTION AND CLASSIFICATION

Feature extraction is the process of converting raw data into numerical parameters that can be manipulated while retaining the original structure of the data. In other words, it involves identifying and extracting relevant features from complex datasets to make them more manageable for analysis. The authors in [56] have performed classification based on polyps using the UCI and PICCOLO datasets. To investigate the impact of several mucosal features in diagnosing H. pylori infection, a gradient-booster decision-making technique was adopted. The EADHI distinguishes H. pylori infection with significant precision as well as good explaining ability,

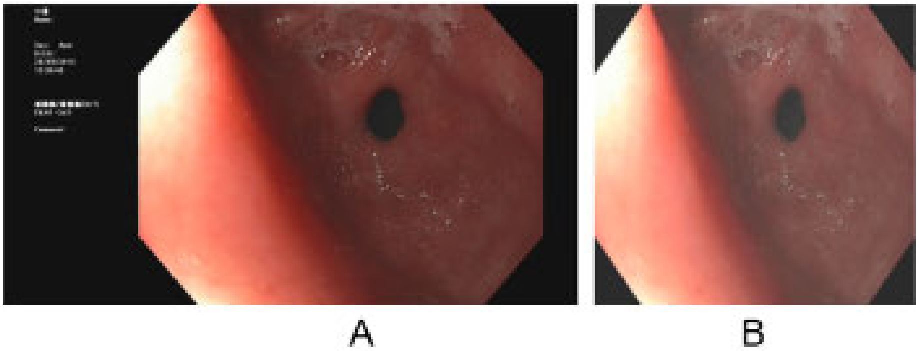


FIGURE 5. A) The original, annotated picture from the Kvasir dataset. B) Cropped/adjusted picture following MAPGI application [49].

FIGURE 6. A)The original, extremely low mean value of pixels, from the Kvasir database. B)Image with improved contrast following MAPGI application [49].

TABLE 2. The pre-processing technique on gastrointestinal diseases.

Author	Pre-processing Technique	Findings
[49]	Edge removal, contrast enhancement, filtering, color mapping, and scaling to each image	Gamma adjustment values are automatically computed for individual photos, normalizing each image's mean pixel value within the 0–255 pixel value range to 90±1.
[52]	Cropped/resized to 299x299	We used several scales to crop the photographs into patches, such as 2 × 2, 3 × 3, 4 × 4, & 5 x 5. Using the z-score technique, the original images have been normalized with a means of.485,.456, and.406 and standard deviations of.229,.224, and.225 for the red, green, and blue channels, accordingly. They were also scaled to 224 x 224.

which can enhance the trust and acceptance of CAD by endoscopists. Atrophy, gastric metaplasia, nodularity, scattered redness, spotty redness, mucosal edema, enlarged folds, and sticky mucus are examples of positive endoscopic results for H. pylori. A red streak, hematin, the fundic glands polyp (FGP), and the regular arrangement for collecting venules (RAC) are additional indicators when H. pylori is not present [57], [58], [59], [60]. Table 4 presents the classification results in different GI images.

B. DEEP LEARNING APPROACHES

DL has emerged as a powerful tool for medical image analysis, with the potential to improve accuracy and efficiency in a variety of clinical applications. In medical image processing, CNNs have been successfully applied to tasks like segmentation, feature extraction, and classification. This section contains works based on common approaches like TL, CNNs, optimization, transformer, and reinforcement learning. Figure 8 shows the mapping of the different Dl approaches and the number of studies.

1) TRANSFER LEARNING

TL is the process of employing a model that has already been trained on one dataset, to solve a new problem. It is now widely regarded and utilized in DL architectures since it can train deep neural networks with a minimal dataset. This is particularly useful in the field of data science because most real-world scenarios frequently lack the millions of

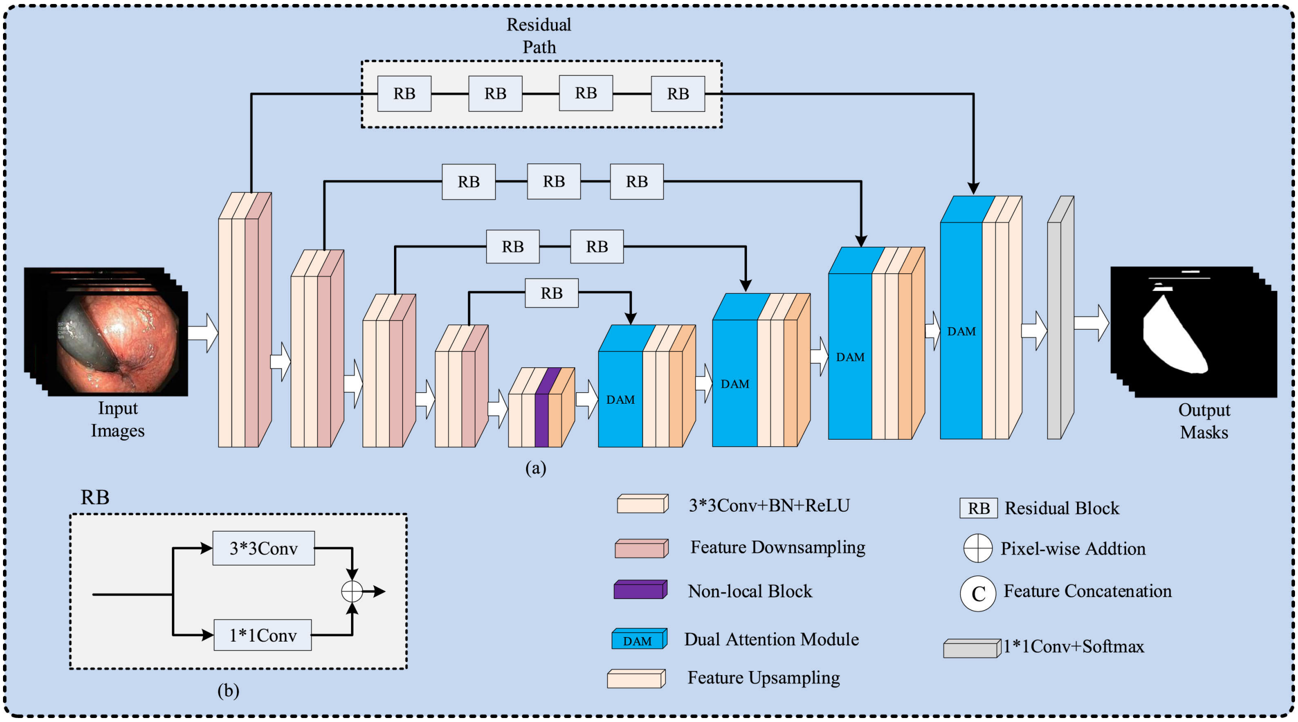


FIGURE 7. Network for segmentation of surgical instruments [55].

TABLE 3. Segmentation performed on GI images.

Ref	Diagnostic Morbidity	Classifier	Disease	Outcome	Inference
[36]	Capsule endoscopy	Suppot Vector Machine(SVM)	Bleeding in ulcer	Grid SVM value was Min:0. 66 Max:0. 18, Avg:0. 84+-0. 10	SVM model performance was better than previous models.

TABLE 4. Classification results in GI images.

Ref	Classifier	Number of Images	Precision	Specificity	F1	Accuracy
[61]	Logistic Regression	4000	0. 7665	0. 966	0. 767	0. 767
[61]	Kernal discriminant Analysis	4000	0. 745	0. 963	0. 7450	936
[61]	Nearest neighbour	4000	0. 564	0. 937	0. 509	0. 891
[61]	K mean clustering	732	0. 688	0. 955	0. 689	0. 922

annotated data points that are normally necessary to train such complex models. TL is the process of transferring as much information as feasible from the task the model has been trained to perform, the current task. Depending on the issue and the information, this knowledge may take any one of many different forms. Although TL offers other benefits as well, its main benefits are shorter training times, better neural network efficiency (in most cases), and less data usage [21]. The two widely used techniques in TL are training a model and reusing it, and the pre-trained model approach. In the first technique, the steps involve choosing a source task, building a proficient model for the initial assignment, and creating a model for the task of interest using the entire model or just certain portions of it. In the pre-trained model approach, the first step is to choose an existing trained model, which can then be used as a foundation for the task of interest, using the entire model or a portion of it. Another strategy is to utilize DL to identify the most crucial elements of a problem's best representation. This strategy, often referred to as representation learning, frequently yields performance that is significantly better than that possible with manually designed representations. Typically, researchers and subject-matter experts manually handcraft the characteristics used in ML. DL can automatically extract features. One still must select which features to include in the network. Yet, neural networks can be used to determine which properties are essential and which can be neglected [58]. Many real-world applications have successfully utilized TL techniques. TL works especially well in fields where the process includes the acquiring of images and ground truth. GastroNet is used as a substitute for ImageNet, which is commonly employed as initialization for training in TL

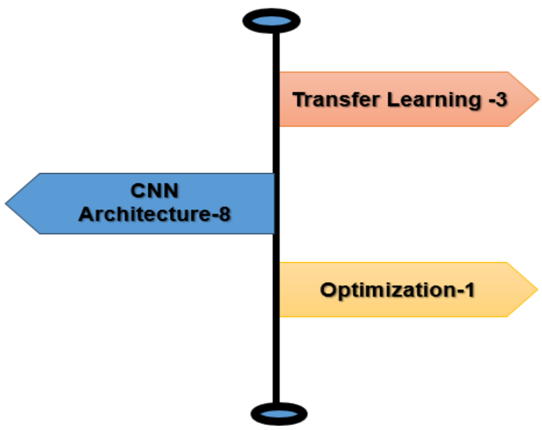


FIGURE 8. Mapping of different DL approaches and the number of studies.

methods leading to image categorization. We expected that the learned features would be more representative of both domains because the images inside the GastroNet dataset closely reflect a target domain (endoscopic images of Barrett's esophagus). Ablation research was conducted to evaluate the critical importance of both GastroNet and the multi-stage TL technique. Ablation studies are often carried out to assess performance when CAD system components are removed (i.e., "ablated"), or changed [62]. When it came to detecting individuals with T1 CRCs who already had lymph node metastases, the ANN (Artificial Neural Networks) model outperformed. Following endoscopic excision of T1 CRCs, this model may be used to assess whether individuals require additional surgery [63]. An innovative system for classifying colorectal polyps photographed by NBI (Narrow Band Imaging) lightning uses a virtual biopsy based on TL. To address the dataset's lack of imagery, the strategy employs TL and image augmentation [64].

In several applications, the deep CNN model developed by Mohammad et al. [35] has been utilized for TL and has been proven successful in object-oriented recognition and classification as shown in Figure 9. This was done on pre-trained, fine-tuned deep CNN models designed for deep feature extraction and implanted TL. Utilizing TL, deep features were collected with the newly acquired data for the categorization of stomach diseases. Based on the results, a suitable strategy for identifying the existence of esophagitis in GI imagery is to add a deep CNN model with values that are pre-trained and fine-tuned. Additionally, using the classification method and subsequently plotting heat maps linked to the classificatory choice makes the model more explainable. Table 5 depicts the TL methods in gastroenterology.

2) CNN ARCHITECTURE

CNNs are a popular DL model for image categorization and identification applications. These have several network layers, including pooling, convolutional, and fully linked layers. CNN uses filters to extract attributes from the input image. The image is down-sampled during the pooling operation to reduce computational expense, and the final prediction is produced by the entirely connected layer. With the help of gradient descent and backpropagation, the network can then learn the best filtering.

During the year 2019, Shichijo et al. investigated the problem of the time required to diagnose an infection with the H.pylori bacteria from endoscopic features, finding that it required 261 seconds to analyze 23,699 images. The trained model outputs a 0 or 1 for H.pylori negative or positive, respectively [67]. In the same year, Stidham et al. carried out studies on UC using a 159-layer CNN model, which had excellent discernment ability from types moderate to severe, with AUROC $= 0.966$. Furthermore, the result was found to rival that of an experienced endoscopist [19]. In the following year, Zhang et al. conducted research on chronic atrophic gastritis and found that CNN would improve the diagnostic accuracy to 0.942. Atrophic gastroenteritis was classified as mild, moderate, and severe with accuracies of 93%, 95%, and 99%, accordingly. Hence, good results were achieved in the classification of gastric lesion severity. In a single unit retrospectively analysis, the LCI-CAD diagnosis was found to be superior to White Light Imaging(WLI) by 9.2% among those who were not infected, 5.0% in cases that were, and 5.0% in cases that were eliminated by the authors [25], who used two models based on CNN for WLI-CAD and LCI-CAD. Thus, the CNN model was determined to be more successful in CAD diagnosis using LCI (Linked Color Imaging) as compared with WLI. In 2021, Xia et al. evaluated the diagnostic capability of CAD systems in Magnetic controlled Capsule Endoscopy, and they developed a novel system to automatically detect gastric lesioning. A region-based convolution Neural Network (RCNN) was formed. The MCE images were divided into 7 categories: erosion, polyp, ulcer, submucosal tumor, xanthoma, normal mucosa, and invalid images. As a result, CNN showed a reliable performance value above 90% for all the categories, except for normal values, which were at 77% [68]. Also, in that year, capsule endoscopy was proven effective in identifying CD and classifying the images as mild, moderate, and severe, as noted by Barash et al. [69]. Their work obtained a classification accuracy of 0.91% for grades 1 and 3, 0.78% for grade 2 versus grade 3, and 62.4% for grade 1 versus grade 2. In contrast, this model was unable to find intermediate categories, and the results were not compared to the endoscopic severity score, because a single image was insufficient, as it required images of the entire bowel. In another study during the same year, Ikenoyama et al. made a comparison between the diagnostic capability of CNN versus the residing endoscopist. When compared to human endoscopists, it was discovered that CNN had found more early cases of stomach cancer in a shorter amount of time [70]. Early detection of GC will be helpful to both the endoscopist and the physician. Early detection would mean

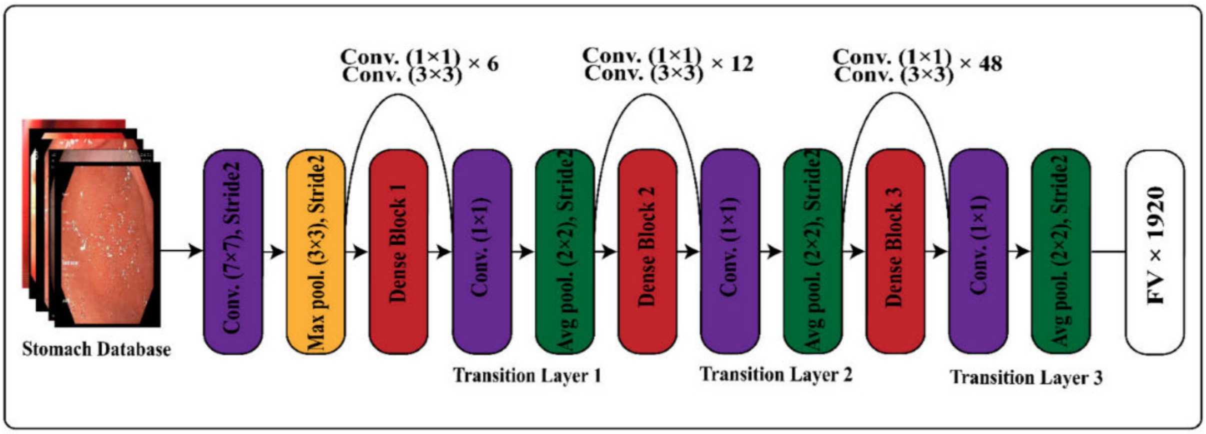


FIGURE 9. DenseNet-201's TL architecture for feature extraction [35].

TABLE 5. Classification results using TL in gastroenterology.

Ref	Diagnostic Morbidity	Classifier	Size of Training/Validating sets/test	Outcome	Inference
[65]	Endoscopic images	Transfer learning techniques with CNN	BE=94 Esophagitis=1663	The accuracy of the CNN ResNet50 using the SVM model is 93. 5%.	In both feature extractors & classification models using transfer learning, DCNN (Deep Convolutional Neural Network) is efficient for diagnosing EAC.
[66]	Endoscopic images	Transfer learning techniques with CNN	1932 esophagitis patients and 1663 normal images.	Adam optimizer trained with 0. 96 AUC and 0. 0001 learning rate. training (80%) and test (20%)	N/A.

TABLE 6. Performance and computation requirement of DL models.

Model	Strength	Performance	Computational Requirement(cost)	Use Cases
ResNet	Deep network with residual connection	Good for classification tasks.	Moderate to High	Object detection and classification.
Inception	Multi-scale feature extraction	Capturing features	High	Multi-scale detection.
DenseNet	Feature reuse due to dense connection	Good for segmentation tasks.	High due to dense connections	Image segmentation.
VGG	Simplicity, feature extraction capabilities.	High accuracy on classification tasks.	High	TL models, feature extraction, and classification tasks.

early diagnosis, which would be time-saving. Zhou et al. [48] created a CNN-based method to obtain 88.3% accuracy for early stomach cancer diagnosis using white light endoscopy.

Table 6 shows the performance and computational requirements of the DL models. The DL models we came across in this study are discussed here. The strength of the ResNet model is the deep network with residual connection. It performs well for classification tasks and the computational requirement is moderate to high. It is mainly used in object detection and classification. Inception is used for multi-scale feature extraction which performs well in capturing the features and has a high computational cost. DenseNet is reuses the feature due to the dense connection and used in segmentation tasks with high computational requirements. VGG models' strength is its simplicity and feature-extracting capabilities with high computational requirements mainly used in TL models, feature extraction, and classification.

Rectified Linear Unit (ReLU) is the activation function commonly used in DL since it avoids the vanishing gradient problem. Sigmoid and tanh are also used. ReLU introduces non-linearity by setting all negative values to zero. It is simple compared to sigmoid or tanh networks. Using ReLU converges faster than sigmoid or tanh.

Endoscopic images can be either of WLI or LCI type and may be in the form of magnetically controlled capsule endoscopy or capsule endoscopy images. CNN was implemented with good accuracy. CNN was found useful in the classification of magnetic-controlled capsule endoscopy images with 90% accuracy. Also used in finding the time for detection of the lesion in endoscopic images. Compared to

TABLE 7. CNN methods used in gastroenterology.

Authors	Image Types	Disease diagnosis	CNN accuracy	Best performance of CNN
[62]	Colonoscopic Image	Colorectal cancer	98. 8% recall rate to predict polyp from non-polyp	Able to assist endoscopist in the procedure.
[71]	WLI-CAD and LCI-CAD (endoscopic)	h pylori detection	LCI outperformed WLI through 9.2% in cases that were not infected, 5.0% in both infected and post-eradicated cases.	LCI-CAD was superior in the detection of all 3 categories
[68]	Magnetic-controlled Capsule Endoscopy	detect gastric lesions	RCNN, 7 categories include xanthoma, submucosal tumor, ulcer, polyp, erosion, and normal mucosa. In addition, they showed a superior performance value above 90%.	RCNN has good detecting performance in all 7 categories of the stomach.
[25]	Endoscopic images	chronic atrophic gastritis	The values used to categorize atrophic gastroenteritis as mild, moderate, or severe were 93%, 95%, and 99%, respectively. The CNN diagnostic accuracy was 0.942.	Excellent accuracy in severity classification of gastric lesioning.
[19]	Endoscopic images	ulcerative colitis	moderate to severe classification AUROC=0. 966	The result was at the same level as an experienced endoscopist.
[70]	Endoscopic images	early gastric cancers	CNN Vs endoscopist	CNN had more detection in less time.
[67]	Endoscopic images	h pylori detection	261 sec/ 23,699 images	Less time in the detection of disease.
[48]	WLE	Early gastric cancer	Accuracy=88.3%	N/A
[72]	Colonoscopy images	Colorectal polyps	The cross-validation's accuracy for 10 holds is 0.751.	Quick identification was helpful.
[73]	Colonoscopy images	Colorectal cancer	Sensitivity is 90% for photographs taken in white light and 97% for images taken in narrow bands.	Promising results in colorectal polys classification.

WLI: White Light Imaging, LCI: Linked color imaging, RCNN: Region-based Convolutional Neural Network,AUROC:Area under the ROC Curve.WLE: White Light Endoscopy, CAD: Computer-aided diagnosis ,N/A: Not Applicable

an endoscopist, they performed more detection in less time. The CNN algorithm was effective in identifying a variety of disorders, including EGC, Crohn's disease, ulcerative colitis, gastric lesions, and chronic atrophic gastritis. Table 7 shows diseases for which CNN is used in gastroenterology and Figure 10 depicts them in the form of a diagram.

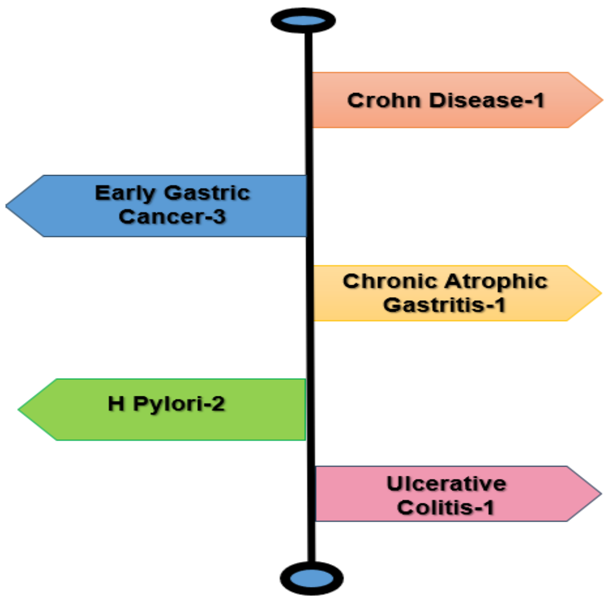


FIGURE 10. Number of papers using CNN for different GI diseases.

3) OPTIMIZATION

Optimization is a process of selecting or transforming the input features into AI models to improve their performance. The goal of feature optimization is to reduce the number of irrelevant or redundant features and select the most informative ones to improve the accuracy and efficiency of the model. By optimizing the features, the ML model can better capture the underlying patterns and relationships within the data, leading to improved performance and better predictive capabilities. As the primary goal of feature optimization is to improve accuracy, researchers in [35] aimed to decrease the number of unimportant characteristics that influence the accuracy of classification and computation time. Figure 11 shows the implanted features optimization approach. This model makes use of a binary dragonfly metaheuristic optimization method and use of the KNN fitness function for reliable feature selection. The dragonfly algorithm (DA), which was recently introduced by Mirjalili et al. [74], was inspired by nature. This population-based metaheuristic method was crafted after dragonfly movement and hunting patterns. The process of hunting involves the movement of a few groups of dragonflies in search of food sources. The bigger dragonfly flocks fly in a single direction during the migration phase. In the study by [28], TL and six current CNNs were tested, and the best-performing architecture was chosen just for the ensemble models. The suggested method's accuracy in performance was assessed using the UCI and PICCOLO datasets (96.3%, 81.2%). The results of trials conducted by other studies on the same data showed that SVM achieved an 82.5% accuracy, and other DL architectures achieved an 85.9% accuracy with 87.6% recall. The suggested method demonstrates how the efficiency of DL-driven CAD has improved by combining a weighted ensemble of optimal networks and data augmentation.

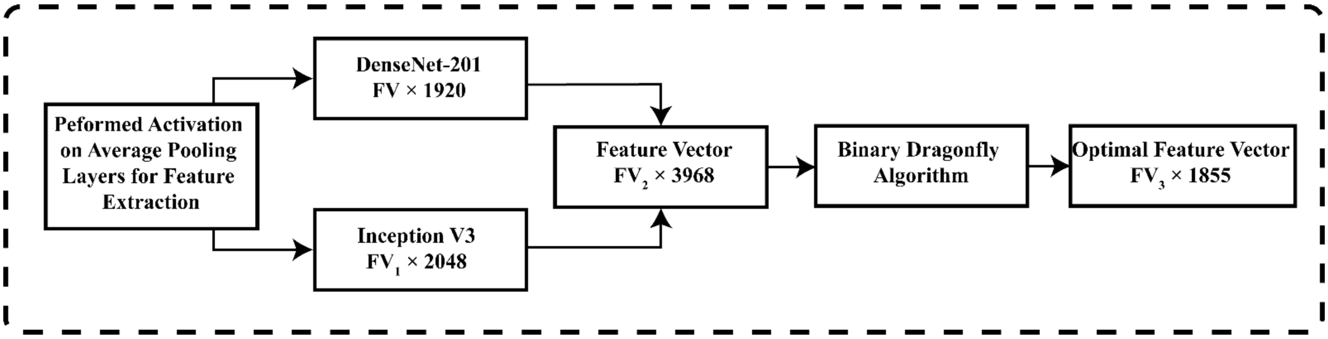


FIGURE 11. Feature optimisation architecture [35].

4) TRANSFORMERS AND REINFORCEMENT LEARNING

- Transformers like Vision Transformers (ViTs) can be used for computer vision tasks.
- Transformers capture dependencies through self-attention mechanisms.
- Image segmentation task can be performed.
- Reinforcement learning can optimize GI disease diagnosis by using image-guided intervention.

Challenges

- The application of transformers to medical imaging is still emerging.
- The transformers need a large amount of data to train the models which may not be suitable when the medical dataset is scarce.
- Reinforcement learning is underexplored in medical imaging but it can be used in robot-assisted surgery.

V. DISCUSSION

RQ1. What recent advancements have been made in ML and DL architectures? Is DL better than ML?

Solution: This review addresses the ML and DL techniques used in GI from 2017 to May 2023. This study found that most studies were based on DL models.

Figure 12 summarize the year-wise ML and DL papers included in the studies. Surely, the trend has moved towards DL. By implementing AI in detecting and diagnosing GI diseases, early detection can be achieved, which benefits both physicians and patients. This review aims to identify the most effective AI tools for optimized results in data analysis.

Discussion of the findings

- In preprocessing techniques, it was observed that when the models were combined, they achieved the highest accuracy level [75].
- In the segmentation process, the classifiers used were ResNet50, ResNet101, DeepLab v3, and R-CNN. It was inferred that when a duo-deep architecture was applied, segmentation was quite effective in characterizing ulceration [76].
- In classification, it was evident that the most used classifiers were CNN, VGG, ResNet- 50, Long Term-Short Memory (LSTM), SVM, random forest, inception, and Xception ensemble model. Ensemble models showed the best performance in classification [27].
- DL models can be applied in real-time during endoscopic procedures, allowing clinicians to receive immediate feedback. This can aid in real-time decision-making.
- CNN-The most common approach in our studies used in the DL models due to reasons like better accuracy, and detection of abnormality in polyps.
- TL- These approaches have been successful in improving accuracy, with CNN, and optimization techniques filling the gaps. TL requires pre-trained models. These methods are particularly useful in real-world scenarios where annotated medical data is limited.
- Endoscopic data is well-suited for DL models, as it enables the detection and analysis of abnormalities, lesions, polyps, tumors, and other pathological conditions.
- Endoscopy is a minimally invasive procedure, which means it can be performed with relatively low risk to the patient.
- In endoscopic imagery, there is a wide variety of methods which include NBI and WLI. The accuracy of a model is related to the quality of the images obtained.

Implication The implications of the findings on various ML techniques and DL Techniques are as follows.

Preprocessing

- Hybrid models can be built by modifying the intermediate layer or by merging the models. Thereby, significantly increasing the accuracy Level.
- Performance Accuracy- The preprocessing models are enhanced by using a combination of models that increase the performance accuracy.

Segmentation

- Duo-deep architecture improves the performance by combining two DL models providing an added advantage of both models. For example- ResNet50 and ResNet101 can be combined and used for feature

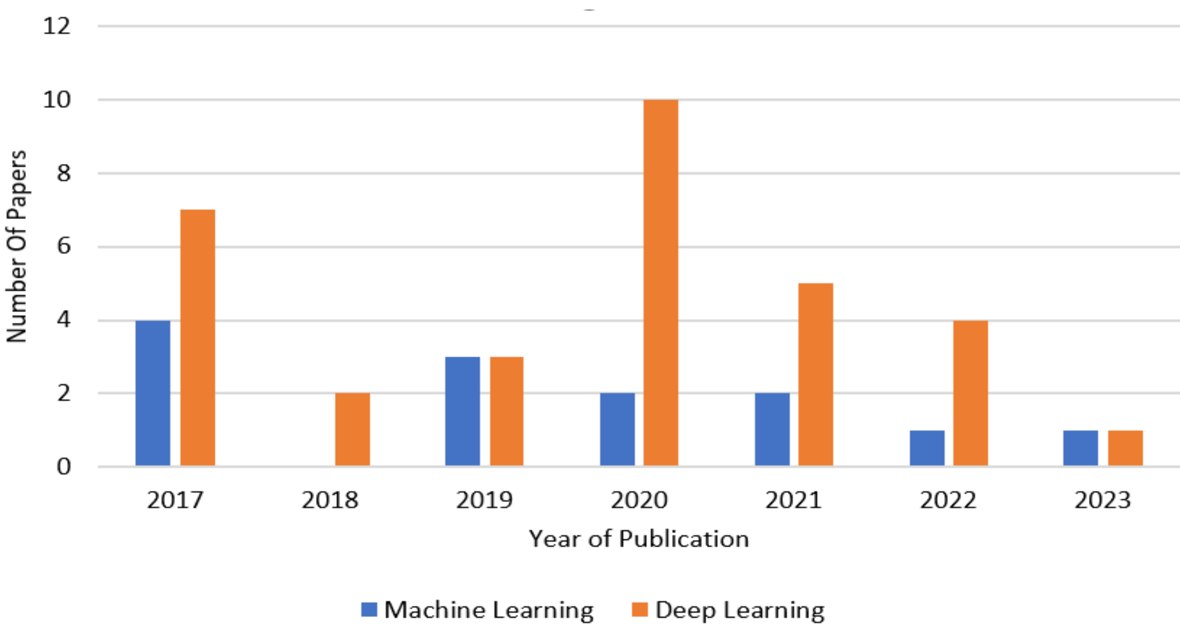


FIGURE 12. Number of studies conducted on GI with AI from 2017 to 2023.

extraction from the image, and DeepLab v3 and R-CNN together can improve the segmentation.

- Increased Accuracy-Segmentation process can be improved by the usage of due-deep architectures.
- Diagnosis Accuracy-Better segmentation can help the physicians in accurate detection of lesions. Eg:-CD, gastric ulcers, and polyps.
- Automation of the segmentation process will provide a great help for the physician by ensuring consistent results.

Classification

- CNN-based architectures -They are helpful in image classification tasks. CNN models can automatically analyze visual data which is essential in GI object detection.
- VGG is useful for easy interpretation in feature extraction.
- ResNet-Its skip connection is useful in avoiding vanishing gradients.
- Inception/Xception's- useful in capturing multi-scale data.
- LSTM is useful in the video-based classification of GI videos.
- SVM and random forest-Traditional ML models are also useful when combined with DL models like an ensemble model to increase robustness.
- Interpretability- The traditional ML models have a greater interpretability compared to DL making it useful in GI diagnosis.

Ensemble models

They are better in generalizability when we have many datasets and diverse real-time data.

- The choice of the model to be used depends on the size of the dataset to be used and the task to be performed.

Using Endoscopic Data

The DL model can automate and standardize the time taken by clinicians by providing early-stage detection.

- The DL model can provide real-time analysis of data.
- Endoscopy is a minimally invasive procedure that, when combined with DL enables safe screening of the patients making it feasible for large populations.
- High quality of the endoscopic image will lead to better diagnosis by providing better feature extraction.
- Through NBI blood vessels and mucosal patterns are visible that help in detecting early gastric cancer. WLI is helpful in providing a broader view of GI diseases

Additional reasons for choosing DL models

- DL models can automatically detect hierarchical features compared to manual feature extraction, performed using traditional ML methods.
- DL models are effective in handling high-dimensional data for tasks like segmentation and classification.
- DL methods can be implemented in real-time diagnosis giving immediate feedback to the physicians.
- DL methods are more generally robust to different imaging techniques WLI or NBI making them suitable for clinical settings.
- TL advantages- When there is limited annotated data, we can use TL models that reduce the training time and model efficiency, a practical choice for real-time GI diagnosis.
- DL models when integrated with ensemble and optimization techniques can be more efficient when compared to the ML models.

RQ2. How can combining of DL and ML models improve the accuracy in feature extraction and classification, providing a personalized treatment?

Solution: This review summarizes all the studies that used a combination of DL and ML approaches are shown in Table 8. The discussion of the findings and their implications are mentioned here.

Discussion of the findings

- DL models can be used to capture comprehensive data for GI diseases in contrast to traditional handcrafted features which can be easily missed.
- DL models can extract comprehensive data and provide a solution based on a patient's GI image.
- ML models have proved successful results in providing an ensemble approach. Here combining multiple base models will give better results. For Example, authors in [77] have used an ensemble approach with classifiers like base feat, LBP, Haralick, and logistic regression and obtained an accuracy of 0. 942.
- TL models are suited when a small dataset is present and trained models can be used. For Example, Cogan et al. [49] used NasNet and Inception as well as Inception-ResNetv2 obtained an accuracy of 0.97, and observed that their performances were similar since the same Inception model was used in the 2nd and 3rd.
- The studies [61], [63], [78], [79] observed that TL is not always advantageous when there is inter and intra-class similarity in data.
- Ensemble models significantly improve accuracy and generalization ability when there is inter and intra-class similarity [79].
- By using Natural Language Processing (NLP) we can use the prescription given by physicians and train the models to provide diagnostic plans. For Example, studies in [80] have used innovation in CAD by promoting electronic reporting as an advancement in the field of GI in the case of bowel preparation by using a combination of ResNet50 and LMT; i.e. a combination of DL and ML methods.

Implication

- **DL** can automatically learn hierarchical features from the endoscopic images compared to the handcrafted traditional methods for edge detection or analysis of texture.
- **DL** can capture even complex patterns that traditional methods cannot capture.
- **Scalability-**DL doesn't require manual intervention to analyze the data even on large datasets making it scalable.
- **Consistency-**DL models provide consistent results by reducing the variability in diagnosis.
- **Adaptability-**DL models can easily learn new data from a dataset and adapt to it.
- **TL** also increases adaptability by fine-tuning to adapt to the specific GI disease which increases the model's performance without using manual feature extraction.
- **DL** can be used in the absence of a GI specialist since manual intervention is not there.
- **DL** models can lead to the discovery of new biomarkers in the early detection of GI conditions.
- When combined in an ensemble approach, ML models can lead to robust and accurate prediction.
- In the **ensemble approach**, diverse feature extraction methods like LBP or Haralick are used, where each method highlights different features.

RQ3. Which datasets are being used in the model training process?

Solution: The GI tract datasets commonly used are shown in Table 9.

Discussion of the findings

- Kvasir is the largest known public CAD dataset for GI diseases.
- The Kvasir dataset is subdivided into Kvasir -Seg used for segmentation purposes and Kvasir Capsule is used for capsule Endoscopy.
- Kvasir and Hyper kvasir are known for various image processing tasks like preprocessing, segmentation, feature extraction, and classification.
- Kvasir-Seg datasets consist of ground truth masks which are helpful in the segmentation of polyps, ulcers, and cancer lesions.
- Kvasir dataset can be used in classification based on anatomy, UC, and bowel preparation scale.
- EAD 2019 dataset was used in Endoscopic Challenge that helps in more research work coming up.
- Most of the studies are single-centered or have combined multiple datasets for training.

Implication

- Kvasir being the largest GI dataset provides extensive resources for researchers working in GI research.
- The size and the variety provided by the Kvasir dataset help to serve as a benchmark for the development of CAD algorithms in the detection, segmentation, and classification of GI diseases.
- Multi-task Analysis dataset includes different modalities like endoscopic videos, ground truth masks, and capsule endoscopy, it can support a wide range of medical image processing tasks.
- Kvasir is an annotated dataset that can enhance GI diagnosis by improving the accuracy of the CAD systems.
- The Kvasir dataset being publicly available encourages open research and collaboration leading to advancement in GI diagnosis.
- Kvasir dataset can help in generalization and it improves diagnosis accuracy by training the CAD systems in various anatomical features.

TABLE 8. Gastrointestinal feature extraction, and classification of gastrointestinal images.

Ref	Disease/Findings	Image Morbidity	Classifier	Size of training/validating set/tests	Outcome	Inference
[77]	Bowel Preparation	Endoscopic	ResNet50 fear. +LMT	21 videos have a total of 5, 525 frames.	Accuracy=0. 957	Innovations in CAD and electronic reporting might advance the field of GI.
[61]	Abnormality in the GI tract	Endoscopic images	Base Feat. + LBP+ Haralick +LR (Logistic Regression)	4000/4000 images	Accuracy=0. 942	The highest accuracy was obtained with logistic regression utilizing the ensemble approach.
[49]	Peptic ulcer	Endoscopic images	NASNet, Inception-v4, and Inception-ResNet v2	6800/1200	Accuracy is 0.9845,0.9848 and 0.9735, correspondingly	Inception-v4 and inception –ResNet showed similar performance.
[78]	Colorectal cancer.	Endoscopic	VGG (Visual Geometry Group)	600 images	83% accuracy	Superior results after fine-tuning.
[63]	Colorectal polyps	Endoscopic Images	Inception-ResNet-v2	6223 images	Among small polyps in the rectum, NPV (Negative Predictive Value) was 97%.	Model surpasses PIVI thresholds for "diagnose and leave" & "resect and discard" techniques.
[79]	Colorectal diseases	Endoscopic Images	ResNet50	3,515 +4000 images	F1-scores of 0. 93 for the Colorectal dataset and 0. 88 for the KVASIR dataset were obtained.	Transfer learning is not always advantageous. Not useful for high intraclass variation and interclass similarity.
[80]	Colorectal polyps	Endoscopic Images	ResNet-101, VGG 19, Xception Ensemble	26512 images	AUC score of 0. 9912.	Use of real-time colonoscopy will help doctors.

- The large dataset will help in avoiding overfitting and generalizing to the real-time scenario.
- Multi-institute data collaboration will help the generalization of the data by validating the dataset.

RQ4. What are the Key Findings for Future Researchers?

Solution: The main issues can be addressed by looking into a variety of variables and strategies, as well as potential future paths for strengthening rapid and accurate assessment, which will raise diagnostic accuracy in GI patients.

Discussion of the findings

- Limited public dataset- Kvasir is the largest known public multi-dataset available for GI diseases. This dataset covers anatomical landmarks like z-line, pylorus, cecum and pathological findings like esophagus, polyps, and UC.
- A close look at the public datasets makes us aware that these datasets have missed having H.pylori infection, a carcinogen leading to GC.
- DL models require more data for training the models to provide beneficial results.
- Future researchers should work on creating a larger and in-depth annotation of the GI features leading to more work in this area and better results.
- pre-trained models are trained on ImageNet and not on GI features. So, these models require fine-tuning.
- Endoscopic images are subjective and the physicians do miss some abnormalities when they are small in size. The CAD systems would help the physicians in diagnosis by reducing these errors.
- The various assessment metrics estimated for segmentation are dice score, (IoU), specificity, sensitivity, accuracy, and precision.
- All assessment parameters are based on four fundamental aspects, which are false positive (FP), false negative (TN), true positive (TP), and false negative (FN).
- Future research should work on creating a larger, annotated dataset, and removal of artifacts like bubbles or specular reflection that hinder the diagnosis accuracy.
- **Focus on the quality and quantity of the dataset-** High-quality and annotated datasets will increase the performance of the DL models.

TABLE 9. GI tract dataset with many classes since 2017.

Year	Dataset	Image Size/video frames	Source	Classes
2017	Kvasir [81]	4,000 images	Norway's VestreViken Health Trust (VV)	Z-line, Ulcerative colitis, pylorus, the cecum, esophagitis, polyps, colored and lifted polyps, and colored resection margins
2017	Nerthus [77]	21 vedios	Baerum Hospital Trust (VV) in Norway	BPS-0, BBPS-1, BBPS2, BBPS-3
2019	EAD2019 [82]	2500 images	There are six distinct data centers, including 1) Oxford University Medical Centre in the UK; 2) Nancy, France's ICL Cancer Institute; 3) The Boulogne-Billancourt AmbroiseParé Hospital in Paris, France; 4) Padova, Italy's Istituto Oncologico Veneto; 5)Switzerland's University Hospital Vaudois, Lausanne; 6) Moscow's Botkin Clinical City Hospital.	It represents multi-tissue (gastroscopy, cystoscopy, gastro-oesophageal, colonoscopy), multi-modal (white light, fluorescence, and narrow-band imaging).
2020	HyperKvasir [83]	110,079 images and 374 videos	Baerumhospital, VestreViken Hospital Trust (VV)in Norway	Z-line, retroflex abdomen, Pylorus, Barrett's, short-segment Barrett's, Esophagitis A, B through D, ileum terminal, Cecum, rectum retroflex, 0-1, 2-3, and BBPS stricken stool, Ulcerative colitis 0-1; Ulcerative colitis 1; Ulcerative colitis 1- 2; Ulcerative colitis 2; Ulcerative colitis 3; Haemorrhoids, Polyps, Dyed resection margins and dyed lifted polyps.
2021	Kvasir-Capsule [84]	4,741,504 images and 117 videos	Norway's Baerum Hospital	VestreViken Hospital Trust (VV)& Ileocecal valve, pylorus, Papal Ampulla, clean and normal mucosa, diminished mucosa view blood fresh and haematin, foreign body, and Erythema, Erosions, ulcers, polyps, lymphangiectasia, and Angi ectasia.

- Researchers should focus on creating a comprehensive dataset that covers various GI conditions.
- **Explainable AI(XAI) has to be included** - Researchers need to integrate XAI to bridge the gap between physicians and AI models [85]. This will help in building trust in the physicians.
- **Collaboration between Discipline**- Collaboration between domain experts in AI and medical professionals will lead to effective AI solutions by focusing on real-time models.
- It is beneficial as more multi-disciplinary researchers can take up topics on GI diseases.
- Data Sharing agreement between the hospitals and research institute will help diagnose GI disease.
- Data Standardization can be performed by image acquisition, labeling, and high-quality datasets in ethical considerations.
- **TL models**- will mitigate the risk of overfitting a small dataset and improve performance.
- **TL models**- are beneficial in a model adaptation that helps in transferring knowledge from labeled, publicly available datasets to unavailable public datasets like H.pylori.
- **Generative Adversarial Network (GAN) models**-help in generating synthetic data with data augmentation techniques. Techniques like rotation and cropping can help in the generation of a diverse and robust dataset.
- Future researchers should focus on translating the AI model from a lab to a clinical setting.
- **AI models should be designed to learn and adapt to new data**-This will make the models relevant and effective to the clinicians.
- **Metrics-** The various assessment metrics estimated for segmentation are dice score, (IoU), specificity, sensitivity, accuracy, and precision. All assessment parameters are based on four fundamental aspects, which are false positive (FP), false negative (TN), true positive (TP), and false negative (FN). Accuracy is calculated as the ratio of correctly classified pixels to all pixels in the image. The total number of actual tumor pixels divided by the frequency of accurately recognizing tumor pixels yields the sensitivity. Specificity is determined by taking the proportion of background or normal tissue that can be correctly recognized. Precision is computed as the rate of correctly recognized tumor pixels regarding backdrop vs. normal tissue. A common metric for comparing the pixel-by-pixel outcomes of anticipated segmentation and ground truth is the dice score. IoU is a metric used to compare the number of pixels for a class that is correctly classified to the sum of the actual and expected numbers of pixels for that class. Equations 1 to 8 display these parameters.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

$$IntersectionoverUnion(IoU) = \frac{TP}{TP + FP + FN} \quad (6)$$

$$DiceScore = \frac{2xTP}{2xTP + FP + FN} \quad (7)$$

$$IntersectionoverUnion(IoU) = \frac{TP}{TP + FP + FN} \quad (8)$$

Implication

- Incomplete coverage of a critical GI condition like H.pylori infection, a carcinogen for GC.
- Limited diversity of the kvasir dataset will not help in generalizing GI conditions.
- A more comprehensive dataset is needed that includes a broader range of landmarks that would enhance the robustness of the CAD systems.
- Opportunities for future researchers to build a large dataset that would support the GI diagnosis process.
- CAD systems can help in the early diagnosis of GI diseases by treating the GI conditions at an early stage.
- CAD systems can serve as a supporting tool for physicians.
- Cost of health care will be reduced due to early intervention of CAD systems.
- Larger, annotated datasets will increase the accuracy of the CAD systems and the removal of artifacts will improve the visibility of GI lesions thereby optimizing the CAD performance.

RQ5 What are the clinical Implications and open Challenges?

Solution: AI is capable of handling large amounts of data with ease and accuracy by providing reliable early detection. Any minor mistakes in medical diagnosis can lead to further complications in saving lives and effective therapy management. The following are the unsolved issues with AI:-

Discussion of the findings

- **Few large public datasets available**- A greater number of public datasets would mean more studies and more results. Most of the studies conducted in the DL platform require large datasets for better prediction.
- **The need for the availability of multicenter data**-This would make a system more robust in generalizing the findings of the study. It would be beneficial not only to the doctors but also to those researchers in the GI domain.
- **The quality of image modalities heavily relies on the operator's experience.** Endoscopy is a procedure wherein the physician should be experienced not only with the device but also in identifying abnormalities.
- **DL models are exact but are considered black boxes.** Since it is difficult to understand the reasons for their predictions. The healthcare industry requires XAI [82], [86], [87]. With the use of more and more XAI in studies the physicians would trust the AI models bridging the gap between the medical and the technical professionals.
- **Real-world challenges in endoscopic image**-It is highly dependent on the operator and affected by the poor quality of the images.
- **Optimization Techniques**- An AI Interface performed parallel with the ongoing procedure is better rather than waiting for the end of the procedure.
- **Hardware Limitation**- Real-time AI deployment relies on factors like computation power, memory, and energy consumption.
- **Processing Power**- High-performing computing resources are needed in real-time AI systems to handle large amounts of visual data.
- **Latency Issue**- In real-time AI in endoscopy, even a slight delay of 100ms per frame might be noticeable.
- Complex models like transformers and CNN have long inference times. So, it is better to minimize any unnecessary layers in DL models as far as possible.
- Quantization reduces the precision and the computational burden.

Implication

- The Presence of a few large datasets will slow the research in GI diseases. There is a risk of overfitting.
- There is a need for standardization of datasets that will help in standardized evaluation.
- A multi-center data will reduce the bias and will be more reliable and accurate which would help in collaboration and validation.
- Multi-center data will strengthen real-time diagnosis and can support Health care systems globally.
- The quality of the endoscopic image can increase diagnostic errors. DL model's accuracy depends on the quality of the image.
- Lack of transparency will be a barrier for the physicians to adapt to the CAD systems. XAI gives better interpretability helping clinicians to adapt to CAD systems.
- High-end GPUs are needed to handle large amounts of data.
- Operator Dependency-Using fully automated AI assistance will reduce human error and increase efficiency.
- For poor-quality image enhancement techniques like noise reduction, contrast enhancement, and artifact removal can be used.
- For optimization processing data locally instead of sending it to the cloud, helps in providing immediate feedback without delay.
- Techniques like pruning and quantization help in reducing the size and computational complexity.
- An alternative for latency problems is cloud-based processing.

VI. LIMITATION

Our comprehensive analysis has identified unsolved issues in AI-based GI diagnosis and we have addressed the possible solutions to them.

Limitations

1) **In Pre-processing and segmentation, while combining the models will increase the computational speed, it results in complex models requiring more computational cost.**
Solution:
   - Use of lightweight models can reduce the computational cost like Mobile Net, Efficient Net, or DeepLabV3.
   - Model Pruning is done by removing the less important parameters improving the performance of the model.
   - A small model (student model) is trained to replicate the output of a large complicated model (teacher model).
   - TL models are already trained on useful features and they can be fine-tuned by reducing the computational cost.
   - Preprocessing techniques like image resizing and cropping can reduce the computational speed.
2) **The effectiveness of DL models is highly dependent on the availability of a large, annotated dataset.**
Solution
   - Data Augmentation techniques like rotation flips, and scaling can be used to increase the dataset size.
   - TL models can be used by taking advantage of the pre-trained model feature and using it to retain the information on a smaller dataset.
   - Semi-supervised learning technique can be used when we have small annotated data and large unannotated data.
   - Using domain-specific models and algorithms can increase the accuracy of even small datasets.
   - The effectiveness of these models is highly dependent on the availability of a large, annotated dataset.
3) **Duo-deep architectures like CNN and LSTM models can lead to overfitting when the dataset is small.**
Solution
   - The complexity of the duo-deep architecture can be reduced by the number of layers and parameters.
   - Early stopping - The validating loss during training is stopped when it increases even after the training loss decreases.
   - Cross-validation-K-fold cross-validation provides a generalized performance by reducing overfitting.
   - Applying regularization techniques like dropout to improve the robustness of the model.
4) **DL models require large datasets that are well-annotated for hierarchical features.**
Solution:
   - Data Augmentation methods can increase the size of the dataset.
   - TL models don't require large datasets since they use pre-trained models.
   - Self-supervised learning can be used when an unlabeled dataset is present.
5) **DL models function as black-boxed making interpretability difficult.**
Solution:
   - The XAI methods used in medical imaging are Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive explanations(SHAP) can be used.
   - Heatmap can be generated using techniques like Gradient-weighted Class Activation Mapping (Grad-CAM).
6) **DL detects complex patterns, and some irrelevant and unimportant patterns are also picked up.**
Solution:
   - The use of techniques like PCA, t-SNE, or autoencoders will help in selecting the most relevant feature.
   - Cross-validation will ensure that the features are extracted across multiple data splits avoiding overfitting.
   - Regularization techniques like L1, L2 regularization, and dropouts can be used to reduce overfitting.
7) **Consistency of the DL model depends on the quality of the dataset(bias).**
Solution:
   - Data cleaning and data validation should be performed to remove bias in the data.
   - Regularization techniques like L1, L2 regularization and dropout should be applied.
8) **In an ensemble model, hyperparameter tuning becomes complex compared to a single model.**
Solution:
   - The complexity of the ensemble model can be reduced by using a smaller set of diverse models.
   - Cross-validation can be performed to increase the robustness of the model.
9) **TL models don't perform well in inter and intra-class dependability. A combination of different feature extraction methods can lead to overfitting.**
Solution:
   - Integration of other imaging types will help the model's ability to discover biomarkers.
   - Overfitting can be avoided by techniques like L1, and L2 regularization, early stopping, cross-validation, and dimension reduction using PCA.
10) **When multiple feature extraction methods are used interpretability becomes difficult.**
Solution:

- Aggregating similar features into groups will help in the interpretability of the results. Example: All methods used for color features can be grouped.
- Visualization techniques like heatmaps can be used to increase the predictability of the models.

VII. CONCLUSION

Implementing AI in GI can transform the present diagnosis of gastro lesions by providing early detection. DL models will become widely accepted in the medical field in the upcoming days.

Promising Techniques for Future Studies

- **Transformers**-ViTs have been emerging as a strong alternative to CNNs due to their ability to capture global contextual information. Further researchers should focus on task like segmentation and classification using refined transformer architecture. Researchers should combine CNNs and transformers to build hybrid models for both global and local features.
- **Reinforcement Learning** has been suited for real-time optimization and robotic endoscopy. Future studies could investigate the integration of reinforcement learning in the medical workflow so that optimization of real-time medical systems can be obtained reducing human intervention and improving diagnostic accuracy.
- **GAN models**- GANs can significantly address the lack of annotated datasets with the creation of synthetic data. Additionally, GANs can use super-resolution imaging to enhance the image quality.

Current Research Gaps

- **Limited high-quality labeled dataset**-Solving the scarcity of large, annotated datasets will help the DL models perform better. While GANs offer synthetic data generation, they must still be clinically validated and generealized.TL models could be further explored as an alternative solution for the absence of a large dataset.
- **Computational cost in real time**- Research should be carried out to optimize the latency, and hardware requirements required for the real-time diagnosis.
- **Interpretablity** -DL models like the transformers and GANs are considered ''black boxes''. So further researchers have to focus on building better XAI methods to provide transparency and Interpretability.
- **Integration with the clinical workflow**- Most models face challenges in adapting to the clinical setting. Further studies can focus on user-friendly systems that work in real-time by providing valuable insights to physicians.

Roadmap to guide future research

- **Collaboration between research institute and hospital**- Collaborative efforts will strengthen the gaps like the creation of large, annotated datasets that can serve as a benchmark for developing and validating DL models. It would help in access to high computational resources and cutting-edge AI.
- **Developing synthetic data using GANs**- GANs can be used in conditions like the unavailability of the dataset. They have to be rigorously validated to ensure their use in training DL models in real-world scenarios.
- **Emphasis on model optimization**-Future research should emphasize optimizing the model for efficiency and accuracy, including pruning, and quantization techniques to reduce computational burden in real-time GI diagnosis. Focus on lightweight architecture would also help.
- **Focus on Explainable AI**- Explain ability is important to gain the trust of the clinician. Researchers should bridge the gap between AI predictions and clinical reasoning by developing better XAI frameworks.
- **Validating clinical trials**-Clinical trials should be designed to evaluate the reliability of AI in real-world scenarios.

In conclusion, transformers, reinforcement learning, and GANs provide exciting opportunities in DL but their real-time deployment faces several challenges. The gaps like data availability, integration with hospitals, interpretability, and computation efficiency will help in the use of AI in GI diagnosis. Future research should focus on various optimization techniques and performance in clinical settings for interdisciplinary collaboration for proven AI-driven. The roadmap gives promising direction by addressing gaps and ensuring AI becomes a promising tool in GI diagnosis.

REFERENCES

[1] A. D. Sperber et al., ''Worldwide prevalence and burden of functional gastrointestinal disorders, results of Rome foundation global study,'' *Gastroenterology*, vol. 160, no. 1, pp. 99–114, 2021.

[2] National Cancer Institute1. *Cancer Stat Facts: Colorectal Cancer*. [Online]. Available: https://seer.cancer.gov/statfacts/html/colorect.html

[3] National Cancer Institute2. *Cancer Stat Facts: Stomach Cancer, National Cancer Institute*. [Online]. Available: https://seer.cancer.gov/statfacts/html/stomach.html

[4] National Cancer Institute3. *Cancer Stat Facts: Small Intestine Cancer, National Cancer Institute*. [Online]. Available: https://seer.cancer.gov/statfacts/html/smint.html

[5] National Cancer Institute4. *Cancer Stat Facts: Pancreatic Cancer, National Cancer Institute*. [Online]. Available: https://seer.cancer.gov/statfacts/html/pancreas.html

[6] National Cancer Institute5. *Cancer Stat Facts: Esophageal Cancer, National Cancer Institute*. [Online]. Available: https://seer.cancer.gov/statfacts/html/esoph.html

[7] S. Emerenziani, M. P. L. Guarino, L. T. Asensio, A. Altomare, M. Ribolsi, P. Balestrieri, and M. Cicala, ''Role of overweight and obesity in gastrointestinal disease,'' *Nutrients*, vol. 12, no. 1, p. 111, Dec. 2019.

[8] X. He, W. Wu, Z. Lin, Y. Ding, J. Si, and L.-M. Sun, ''Validation of the American joint committee on cancer (AJCC) 8th edition stage system for gastric cancer patients: A population-based analysis,'' *Gastric Cancer*, vol. 21, no. 3, pp. 391–400, May 2018.

[9] J.-H. Cho, S. R. Jeon, and S.-Y. Jin, ''Clinical applicability of gastroscopy with narrow-band imaging for the diagnosis of *Helicobacter pylori* gastritis, precancerous gastric lesion, and neoplasia,'' *World J. Clin. Cases*, vol. 8, no. 14, pp. 2902–2916, Jul. 2020.

[10] J. Padda, K. Khalid, A. C. Cooper, and G. Jean-Charles, ''Association between *Helicobacter pylori* and gastric carcinoma,'' *Cureus*, vol. 13, no. 5, May 2021, Art. no. e15165.

[11] R. Pannala, K. Krishnan, J. Melson, M. A. Parsi, A. R. Schulman, S. Sullivan, G. Trikudanathan, A. J. Trindade, R. R. Watson, J. T. Maple, and D. R. Lichtenstein, ''Artificial intelligence in gastrointestinal endoscopy,'' *VideoGIE*, vol. 5, no. 12, pp. 598–613, 2020.

[12] T. Harada, T. Watari, S. Watanuki, J. Hiroshige, S. Kushiro, T. Miyagami, S. Syusa, S. Suzuki, T. Hiyoshi, S. Hasegawa, S. Nabeshima, H. Aihara, S. Yamashita, M. Tago, F. Yoshimura, K. Kunitomo, T. Tsuji, M. Hirose, T. Tsuchida, and T. Shimizu, ''Diagnostic error rates and associated factors for lower gastrointestinal perforation,'' *Sci. Rep.*, vol. 12, no. 1, p. 1028, Jan. 2022.

[13] C. Hassan et al., ''Computer-aided detection-assisted colonoscopy: Classification and relevance of false positives,'' *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 900–904, Oct. 2020.

[14] D. Patel. *Gastrointestinal Endoscopy/Upper GI Endoscopy/Lower GI Endoscopy*. [Online]. Available: https://drruchitpatel.com/upper-lower-gi-endoscopy.php

[15] S. Unno, K. Igarashi, H. Saito, D. Hirasawa, T. Okuzono, Y. Tanaka, M. Nakahori, and T. Matsuda, ''Assigning a different endoscopist for each annual follow-up may contribute to improved gastric cancer detection rates,'' *Endoscopy Int. Open*, vol. 10, no. 10, pp. E1333–E1342, Oct. 2022.

[16] K. Sugano, J. Tack, E. J. Kuipers, D. Y. Graham, E. M. El-Omar, S. Miura, K. Haruma, M. Asaka, N. Uemura, and P. Malfertheiner, ''Kyoto global consensus report on *Helicobacter pylori* gastritis,'' *Gut*, vol. 64, no. 9, pp. 1353–1367, 2015.

[17] Y.-J. Hwang, N. Kim, H. Lee, J. Lee, Y. Choi, H. Yoon, C. Shin, Y. Park, and D. Lee, ''Reversibility of atrophic gastritis and intestinal metaplasia after *Helicobacter pylori* eradication—A prospective study for up to 10 years,'' *Alimentary Pharmacol. Therapeutics*, vol. 47, no. 3, pp. 380–390, 2018.

[18] Y. Mori, S.-E. Kudo, J. E. East, A. Rastogi, M. Bretthauer, M. Misawa, M. Sekiguchi, T. Matsuda, Y. Saito, H. Ikematsu, K. Hotta, K. Ohtsuka, T. Kudo, and K. Mori, ''Cost savings in colonoscopy with artificial intelligence-aided polyp diagnosis: An add-on analysis of a clinical trial (with video),'' *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 905–911, Oct. 2020.

[19] R. W. Stidham and K. Takenaka, ''Artificial intelligence for disease assessment in inflammatory bowel disease: How will it change our practice?'' *Gastroenterology*, vol. 162, no. 5, pp. 1493–1506, Apr. 2022.

[20] K. Takenaka, K. Ohtsuka, T. Fujii, M. Negi, K. Suzuki, H. Shimizu, S. Oshima, S. Akiyama, M. Motobayashi, M. Nagahori, E. Saito, K. Matsuoka, and M. Watanabe, ''Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis,'' *Gastroenterology*, vol. 158, no. 8, pp. 2150–2157, Jun. 2020.

[21] M. K. Sana, Z. M. Hussain, P. A. Shah, and M. H. Maqsood, ''Artificial intelligence in celiac disease,'' *Comput. Biol. Med.*, vol. 125, Oct. 2020, Art. no. 103996.

[22] Y. Horie, T. Yoshio, K. Aoyama, S. Yoshimizu, Y. Horiuchi, A. Ishiyama, T. Hirasawa, T. Tsuchida, T. Ozawa, S. Ishihara, Y. Kumagai, M. Fujishiro, I. Maetani, J. Fujisaki, and T. Tada, ''Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks,'' *Gastrointestinal Endoscopy*, vol. 89, no. 1, pp. 25–32, Jan. 2019.

[23] J. H. Song, S. G. Kim, E. H. Jin, J. H. Lim, and S. Y. Yang, ''Risk factors for gastric tumorigenesis in underlying gastric mucosal atrophy,'' *Gut Liver*, vol. 11, no. 5, pp. 612–619, Sep. 2017.

[24] D. Y. Cheung, ''Atrophic gastritis increases the risk of gastric cancer in asymptomatic population in Korea,'' *Gut Liver*, vol. 11, no. 5, pp. 575–576, Sep. 2017.

[25] Y. Zhang, F. Li, F. Yuan, K. Zhang, L. Huo, Z. Dong, Y. Lang, Y. Zhang, M. Wang, Z. Gao, Z. Qin, and L. Shen, ''Diagnosing chronic atrophic gastritis by gastroscopy using artificial intelligence,'' *Digestive Liver Disease*, vol. 52, no. 5, pp. 566–572, May 2020.

[26] Y. Horiuchi, T. Hirasawa, N. Ishizuka, Y. Tokai, K. Namikawa, S. Yoshimizu, A. Ishiyama, T. Yoshio, T. Tsuchida, J. Fujisaki, and T. Tada, ''Performance of a computer-aided diagnosis system in diagnosing early gastric cancer using magnifying endoscopy videos with narrow-band imaging (with videos),'' *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 856–865.e1, Oct. 2020.

[27] C. Hassan, G. Balsamo, R. Lorenzetti, A. Zullo, and G. Antonelli, ''Artificial intelligence allows leaving-in-situ colorectal polyps,'' *Clin. Gastroenterol. Hepatol.*, vol. 20, no. 11, pp. 2505–2513, Nov. 2022.

[28] A. Lanas and F. K. Chan, ''Peptic ulcer disease,'' *Lancet*, vol. 390, no. 10094, pp. 613–624, 2017.

[29] M. A. Khan, M. Rashid, M. Sharif, K. Javed, and T. Akram, ''Classification of gastrointestinal diseases of stomach from WCE using improved saliency-based method and discriminant features selection,'' *Multimedia Tools Appl.*, vol. 78, no. 19, pp. 27743–27770, Oct. 2019.

[30] S. Charfi and M. E. Ansari, ''Computer-aided diagnosis system for colon abnormalities detection in wireless capsule endoscopy images,'' *Multimedia Tools Appl.*, vol. 77, no. 3, pp. 4047–4064, Feb. 2018.

[31] H. Saito, T. Aoki, K. Aoyama, Y. Kato, A. Tsuboi, A. Yamada, M. Fujishiro, S. Oka, S. Ishihara, T. Matsuda, M. Nakahori, S. Tanaka, K. Koike, and T. Tada, ''Automatic detection and classification of protruding lesions in wireless capsule endoscopy images based on a deep convolutional neural network,'' *Gastrointestinal Endoscopy*, vol. 92, no. 1, pp. 144–151.e1, Jul. 2020.

[32] J. Naz, M. Sharif, M. Yasmin, M. Raza, and M. A. Khan, ''Detection and classification of gastrointestinal diseases using machine learning,'' *Current Med. Imag. Formerly Current Med. Imag. Rev.*, vol. 17, no. 4, pp. 479–490, May 2021.

[33] S. Suman, F. A. B. Hussin, A. S. Malik, K. Pogorelov, M. Riegler, S. H. Ho, I. Hilmi, and K. L. Goh, ''Detection and classification of bleeding region in WCE images using color feature,'' in *Proc. 15th Int. Workshop Content Multimedia Indexing*, Jun. 2017, pp. 1–6.

[34] S. Suman, F. Hussin, A. Malik, S. Ho, I. Hilmi, A. Leow, and K.-L. Goh, ''Feature selection and classification of ulcerated lesions using statistical analysis for WCE images,'' *Appl. Sci.*, vol. 7, no. 10, p. 1097, Oct. 2017.

[35] F. Mohammad and M. Al-Razgan, ''Deep feature fusion and optimization-based approach for stomach disease classification,'' *Sensors*, vol. 22, no. 7, p. 2801, Apr. 2022.

[36] E. Tuba, M. Tuba, and R. Jovanovic, ''An algorithm for automated segmentation for bleeding detection in endoscopic images,'' in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 4579–4586.

[37] S. Charfi and M. El Ansari, ''Computer-aided diagnosis system for ulcer detection in wireless capsule endoscopy videos,'' in *Proc. Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, May 2017, pp. 1–5.

[38] A. Sharma, R. Kumar, G. Yadav, and P. Garg, ''Artificial intelligence in intestinal polyp and colorectal cancer prediction,'' *Cancer Lett.*, vol. 565, Jul. 2023, Art. no. 216238.

[39] Z. Jin, T. Gan, P. Wang, Z. Fu, C. Zhang, Q. Yan, X. Zheng, X. Liang, and X. Ye, ''Deep learning for gastroscopic images: Computer-aided techniques for clinicians,'' *Biomed. Eng.*, vol. 21, no. 1, p. 12, Feb. 2022.

[40] J. Gubatan, S. Levitte, A. Patel, T. Balabanis, M. T. Wei, and S. R. Sinha, ''Artificial intelligence applications in inflammatory bowel disease: Emerging technologies and future directions,'' *World J. Gastroenterol.*, vol. 27, no. 17, pp. 1920–1935, May 2021.

[41] V. Patel, M. N. Khan, A. Shrivastava, K. Sadiq, S. A. Ali, S. R. Moore, D. E. Brown, and S. Syed, ''Artificial intelligence applied to gastrointestinal diagnostics: A review,'' *J. Pediatric Gastroenterol. Nutrition*, vol. 70, no. 1, pp. 4–11, 2020.

[42] I. Tziortziotis, F.-M. Laskaratos, and S. Coda, ''Role of artificial intelligence in video capsule endoscopy,'' *Diagnostics*, vol. 11, no. 7, p. 1192, Jun. 2021.

[43] C. D. Christou and G. Tsoulfas, ''Challenges and opportunities in the application of artificial intelligence in gastroenterology and hepatology,'' *World J. Gastroenterol.*, vol. 27, no. 37, pp. 6191–6223, Oct. 2021.

[44] K. Namikawa, T. Hirasawa, T. Yoshio, J. Fujisaki, T. Ozawa, S. Ishihara, T. Aoki, A. Yamada, K. Koike, H. Suzuki, and T. Tada, ''Utilizing artificial intelligence in endoscopy: A clinician's guide,'' *Expert Rev. Gastroenterol. Hepatol.*, vol. 14, no. 8, pp. 689–706, 2020.

[45] Y. J. Yang, ''The future of capsule endoscopy: The role of artificial intelligence and other technical advancements,'' *Clin. Endoscopy*, vol. 53, no. 4, pp. 387–394, Jul. 2020.

[46] C. Sohrabi, T. Franchi, G. Mathew, A. Kerwan, M. Nicola, M. Griffin, M. Agha, and R. Agha, ''PRISMA 2020 statement: What's new and the importance of reporting guidelines,'' *Int. J. Surgery*, vol. 88, Apr. 2021, Art. no. 105918.

[47] S. Ahmed, M. M. Alshater, A. E. Ammari, and H. Hammami, ''Artificial intelligence and machine learning in finance: A bibliometric review,'' *Res. Int. Bus. Finance*, vol. 61, Oct. 2022, Art. no. 101646.

[48] B. Zhou, X. Rao, H. Xing, Y. Ma, F. Wang, and L. Rong, ''A convolutional neural network-based system for detecting early gastric cancer in white-light endoscopy,'' *Scandin. J. Gastroenterol.*, vol. 58, no. 2, pp. 157–162, Feb. 2023.

[49] T. Cogan, M. Cogan, and L. Tamil, ''MAPGI: Accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning,'' *Comput. Biol. Med.*, vol. 111, Aug. 2019, Art. no. 103351.

[50] T. Chen, S. Liu, Y. Li, X. Feng, W. Xiong, X. Zhao, Y. Yang, C. Zhang, Y. Hu, H. Chen, T. Lin, M. Zhao, H. Liu, J. Yu, Y. Xu, Y. Zhang, and G. Li, "Developed and validated a prognostic nomogram for recurrence-free survival after complete surgical resection of local primary gastrointestinal stromal tumors based on deep learning," *EBioMedicine*, vol. 39, pp. 272–279, Jan. 2019.

[51] L. Gong, M. Wang, L. Shu, J. He, B. Qin, J. Xu, W. Su, D. Dong, H. Hu, J. Tian, and P. Zhou, "Automatic captioning of early gastric cancer using magnification endoscopy with narrow-band imaging," *Gastrointestinal Endoscopy*, vol. 96, no. 6, pp. 929–942.e6, Dec. 2022.

[52] P. M. Vieira, N. R. Freitas, V. B. Lima, D. Costa, C. Rolanda, and C. S. Lima, "Multi-pathology detection and lesion localization in WCE videos by using the instance segmentation approach," *Artif. Intell. Med.*, vol. 119, Sep. 2021, Art. no. 102141.

[53] V. Sadeghi, A. Vard, M. Sharifi, H. Mir, and A. Mehridehnavi, "Segmentation and region quantification of bubbles in small bowel capsule endoscopy images using wavelet transform," *Informat. Med. Unlocked*, vol. 42, Jan. 2023, Art. no. 101364.

[54] M. A. Khan, M. A. Khan, F. Ahmed, M. Mittal, L. M. Goyal, D. J. Hemanth, and S. C. Satapathy, "Gastrointestinal diseases segmentation and classification based on duo-deep architectures," *Pattern Recognit. Lett.*, vol. 131, pp. 193–204, Mar. 2020.

[55] L. Yang, Y. Gu, G. Bian, and Y. Liu, "An attention-guided network for surgical instrument segmentation from endoscopic images," *Comput. Biol. Med.*, vol. 151, Dec. 2022, Art. no. 106216.

[56] F. Younas, M. Usman, and W. Q. Yan, "A deep ensemble learning method for colorectal polyp classification with optimized network parameters," *Int. J. Speech Technol.*, vol. 53, no. 2, pp. 2410–2433, Jan. 2023.

[57] M. Kato, "Endoscopic findings of *H. pylori* infection," *Helicobacter pylori*, pp. 157–167, Apr. 2016.

[58] M. Alhajlah, "Robust ulcer classification: Contrast and illumination invariant approach," *Diagnostics*, vol. 12, no. 12, p. 2898, Nov. 2022.

[59] B. Glover, J. Teare, H. Ashrafian, and N. Patel, "The endoscopic predictors of *Helicobacter pylori* status: A meta-analysis of diagnostic performance," *Therapeutic Adv. Gastrointestinal Endoscopy*, vol. 13, Jan. 2020, Art. no. 263177452095084.

[60] B. Glover, J. Teare, and N. Patel, "Assessment of *Helicobacter pylori* status by examination of gastric mucosal patterns: Diagnostic accuracy of white-light endoscopy and narrow-band imaging," *BMJ Open Gastroenterol.*, vol. 8, no. 1, Aug. 2021, Art. no. e000608.

[61] S. S. A. Naqvi, S. Nadeem, M. Zaid, and M. A. Tahir, "Ensemble of texture features for finding abnormalities in the gastro-intestinal tract," *MediaEval*, vol. 17, pp. 13–15, Sep. 2017.

[62] R. Zhang, Y. Zheng, T. W. C. Mak, R. Yu, S. H. Wong, J. Y. W. Lau, and C. C. Y. Poon, "Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 41–47, Jan. 2017.

[63] R. Zachariah, J. Samarasena, D. Luba, E. Duh, T. Dao, J. Requa, A. Ninh, and W. Karnes, "Prediction of polyp pathology using convolutional neural networks achieves 'resect and discard' thresholds," *Off. J. Amer. College Gastroenterol.*, vol. 115, no. 1, pp. 138–144, 2020.

[64] B. Korbar, A. M. Olofson, A. P. Miraflor, C. M. Nicka, M. A. Suriawinata, L. Torresani, A. A. Suriawinata, and S. Hassanpour, "Deep learning for classification of colorectal polyps on whole-slide images," *J. Pathol. Informat.*, vol. 8, no. 1, p. 30, Jan. 2017.

[65] M. A. Khan, K. Muhammad, S.-H. Wang, S. Alsubai, A. Binbusayyis, A. Alqahtani, A. Majumdar, and O. Thinnukool, *Gastrointestinal Diseases Recognition: A Framework of Deep Neural Network and Improved Moth-Crow Optimization With DCCA Fusion*. Cham, Switzerland: Springer, 2022.

[66] E. C. Silveira, C. F. S. Corrêa, L. M. Silva, B. A. Santos, S. M. Pretti, and F. F. de Melo, "Recognition of esophagitis in endoscopic images using transfer learning," *World J. Gastrointestinal Endoscopy*, vol. 14, no. 5, pp. 311–319, May 2022.

[67] S. Shichijo, Y. Endo, K. Aoyama, Y. Takeuchi, T. Ozawa, H. Takiyama, K. Matsuo, M. Fujishiro, S. Ishihara, R. Ishihara, and T. Tada, "Application of convolutional neural networks for evaluating *Helicobacter pylori* infection status on the basis of endoscopic images," *Scandin. J. Gastroenterol.*, vol. 54, no. 2, pp. 158–163, Feb. 2019.

[68] J. Xia, T. Xia, J. Pan, F. Gao, S. Wang, Y.-Y. Qian, H. Wang, J. Zhao, X. Jiang, W.-B. Zou, Y.-C. Wang, W. Zhou, Z.-S. Li, and Z. Liao, "Use of artificial intelligence for detection of gastric lesions by magnetically controlled capsule endoscopy," *Gastrointestinal Endoscopy*, vol. 93, no. 1, pp. 133–139, Jan. 2021.

[69] Y. Barash, L. Azaria, S. Soffer, R. Margalit Yehuda, O. Shlomi, S. Ben-Horin, R. Eliakim, E. Klang, and U. Kopylov, "Ulcer severity grading in video capsule images of patients with Crohn's disease: An ordinal neural network solution," *Gastrointestinal Endoscopy*, vol. 93, no. 1, pp. 187–192, Jan. 2021.

[70] Y. Ikenoyama, T. Hirasawa, M. Ishioka, K. Namikawa, S. Yoshimizu, Y. Horiuchi, A. Ishiyama, T. Yoshio, T. Tsuchida, Y. Takeuchi, S. Shichijo, N. Katayama, J. Fujisaki, and T. Tada, "Detecting early gastric cancer: Comparison between the diagnostic ability of convolutional neural networks and endoscopists," *Digestive Endoscopy*, vol. 33, no. 1, pp. 141–150, Jan. 2021.

[71] T. Hirasawa, Y. Ikenoyama, M. Ishioka, K. Namikawa, Y. Horiuchi, H. Nakashima, and J. Fujisaki, "Current status and future perspective of artificial intelligence applications in endoscopic diagnosis and management of gastric cancer," *Digestive Endoscopy*, vol. 33, no. 2, pp. 263–272, Jan. 2021.

[72] Y. Komeda, H. Handa, T. Watanabe, T. Nomura, M. Kitahashi, T. Sakurai, A. Okamoto, T. Minami, M. Kono, T. Arizumi, M. Takenaka, S. Hagiwara, S. Matsui, N. Nishida, H. Kashida, and M. Kudo, "Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: Preliminary experience," *Oncology*, vol. 93, no. 1, pp. 30–34, 2017.

[73] T. Ozawa, S. Ishihara, M. Fujishiro, Y. Kumagai, S. Shichijo, and T. Tada, "Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks," *Therapeutic Adv. Gastroenterol.*, vol. 13, Jan. 2020, Art. no. 1756284820910659.

[74] S. Mirjalili, "Dragonfly algorithm: A new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems," *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1053–1073, May 2016.

[75] C. Zhang and Y. Lu, "Study on artificial intelligence: The state of the art and future prospects," *J. Ind. Inf. Integr.*, vol. 23, Sep. 2021, Art. no. 100224.

[76] W. Hu, W. Wang, C. Ai, J. Wang, W. Wang, X. Meng, J. Liu, H. Tao, and S. Qiu, "Machine vision-based surface crack analysis for transportation infrastructure," *Autom. Construct.*, vol. 132, Dec. 2021, Art. no. 103973.

[77] K. Pogorelov, K. R. Randel, T. de Lange, S. L. Eskeland, C. Griwodz, D. Johansen, C. Spampinato, M. Taschwer, M. Lux, P. T. Schmidt, and M. Riegler, "Nerthus: A bowel preparation quality video dataset," in *Proc. 8th ACM Multimedia Syst. Conf.*, 2017, pp. 170–174.

[78] S. Patino-Barrientos, D. Sierra-Sosa, B. Garcia-Zapirain, C. Castillo-Olea, and A. Elmaghraby, "Kudo's classification for colon polyps assessment using a deep learning approach," *Appl. Sci.*, vol. 10, no. 2, p. 501, Jan. 2020.

[79] S. Poudel, Y. J. Kim, D. M. Vo, and S.-W. Lee, "Colorectal disease classification using efficiently scaled dilation in convolutional neural network," *IEEE Access*, vol. 8, pp. 99227–99238, 2020.

[80] M. M. Rahman, M. A. H. Wadud, and M. M. Hasan, "Computerized classification of gastrointestinal polyps using stacking ensemble of convolutional neural network," *Informat. Med. Unlocked*, vol. 24, 2021, Art. no. 100603.

[81] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D. T. Dang-Nguyen, M. Lux, P. T. Schmidt, and M. Riegler, "KVASIR: A multi-class image dataset for computer aided gastrointestinal disease detection," in *Proc. 8th ACM Multimedia Syst. Conf.*, Jun. 2017, pp. 164–169.

[82] S. Ali, F. Zhou, C. Daul, B. Braden, A. Bailey, S. Realdon, J. East, G. Wagnieres, V. Loschenov, E. Grisan, and W. Blondel, "Endoscopy artifact detection (EAD 2019) challenge dataset," 2019, *arXiv:1905.03209*.

[83] H. Borgli, V. Thambawita, P. H. Smedsrud, S. Hicks, D. Jha, S. L. Eskeland, K. R. Randel, K. Pogorelov, M. Lux, D. T. D. Nguyen, D. Johansen, C. Griwodz, H. K. Stensland, E. Garcia-Ceja, P. T. Schmidt, H. L. Hammer, M. A. Riegler, P. Halvorsen, and T. de Lange, "HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Sci. Data*, vol. 7, no. 1, p. 283, Aug. 2020.

[84] P. H. Smedsrud, V. Thambawita, S. A. Hicks, H. Gjestang, O. O. Nedrejord, E. Næss, H. Borgli, D. Jha, T. J. D. Berstad, S. L. Eskeland, and M. Lux, "*Kvasir-Capsule*, a video capsule endoscopy dataset," *Sci. Data*, vol. 8, no. 1, p. 142, 2021.

[85] M. Zhang, J. Pan, J. Lin, M. Xu, L. Zhang, R. Shang, L. Yao, Y. Li, W. Zhou, Y. Deng, Z. Dong, Y. Zhu, X. Tao, L. Wu, and H. Yu, ''An explainable artificial intelligence system for diagnosing *Helicobacter pylori* infection under endoscopy: A case–control study,'' *Therapeutic Adv. Gastroenterol.*, vol. 16, Jan. 2023, Art. no. 17562848231155023.

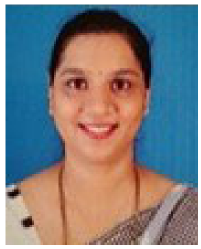
[86] F. Ahmed, S. Abbas, A. Athar, T. Shahzad, W. A. Khan, M. Alharbi, M. A. Khan, and A. Ahmed, ''Identification of kidney stones in KUB X-ray images using VGG16 empowered with explainable artificial intelligence,'' *Sci. Rep.*, vol. 14, no. 1, p. 6173, Mar. 2024.

[87] F. Ahmed, A. Fatima, M. Mamoon, and S. Khan, ''Identification of the diabetic retinopathy using ResNet-18,'' in *Proc. 2nd Int. Conf. Cyber Resilience (ICCR)*, Feb. 2024, pp. 1–6.

JOVITA RELASHA LEWIS received the B.E. degree from Visvesvaraya Technological University, Belgaum, and the M.Sc. degree in computer science. She is currently pursuing the Ph.D. degree with the Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India. Her current research interests include deep learning, medical image analysis, machine learning, and their applications.

SAMEENA PATHAN is currently an Assistant Professor with the Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal. Her research interests include pattern recognition, medical image analysis, artificial intelligence, and machine learning.

PREETHAM KUMAR received the Ph.D. degree in data mining from the National Institute of Technology, Karnataka. He is currently a Deputy Registrar-Academics (Technical) with Manipal Academy of Higher Education, Manipal, and a Professor with the Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education. His research interests include data mining, image processing, bioinformatics, advanced database management systems, operating systems, software architecture, and software engineering.



CIFHA CRECIL DIAS received the B.E. degree in electrical and electronics engineering, the M.Tech. degree in control systems, and the Ph.D. degree from the Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal. She is currently an Assistant Professor Senior Scale with the Department of Biomedical Engineering, Manipal Institute of Technology. She has ten years of teaching and four years of research experience. She has published articles in reputed journals and has been an active reviewer in various journals. Her research interests include precision control in drug delivery, AI and ML, biomedical control and signal, sensor monitoring and medical devices, and disease diagnosis prediction.