A Comprehensive Review on Gastrointestinal Disease Detection and Classification Using a Tailored Convolutional Neural Network Layer

Hemanth T S[1], Hemalatha B M[2], Chinmayi M U[3], Vijayalaxmi[4], Hemanth P[5], Manoj H P[6]
Malnad College of Engineering, Hassan,Karnataka, , India

***Abstract*—Gastrointestinal (GI) diseases represent a significant global health burden, with endoscopy being the primary diagnostic tool, though manual image analysis is time-consuming and prone to human error, potentially delaying treatment. Deep learning, particularly Convolutional Neural Networks (CNNs), offers an automated and accurate solution for disease classification, with this review comprehensively analyzing techniques such as foundational CNN architectures (e.g., ResNet), hybrid and ensemble models, and transformer-based systems, alongside feature extraction strategies, data preprocessing methods, and public datasets like KVASIR Research has proven that, for many applications, hybrid and ensemble models show superior accuracy, often more than 98%, by mixing diverse architectures, while different types of augmentations and strategies to counter class imbalance, BL-SMOTE for example, show a considerable improvement in robustness. Despite this, there were still some problems such as costs of computing, large annotated datasets and lack of interpretability. Hence, future work must focus on light-weight models for real-time use in clinics and explainable-AI (XAI) for greater trust and adoption in practice.**

***Index Terms*—Endoscopy, Deep Learning, Convolutional Neural Networks (CNN), ResNet101V2, Alimentary Tract**

I. INTRODUCTION

GI Diseases are a global health problem. They affect millions of people. They cause a lot of morbidity. They kill too many people, and they also cost too much money. Millions of cases and deaths from colorectal, gastric, and esophageal cancers are recorded every year, signaling the need for accurate and timely diagnoses. According to the National Institute of Health, 60 and 70 million American suffers from GI diseases each year (these diseases are not just restricted to America, but in developed nations [6][7][5][4], Endoscopy is the mainstay of diagnosis in these diseases in which the GI tract is visually examined through a flexible tube with camera placed on it. It is a useful procedure for identifying pathological findings of polyps, ulcers, and tumors. Despite this, the diagnostic process has its own limitations. Gastroenterologists face a daunting task of analyzing thousands of images and videos obtained from an endoscopic procedure, which is tedious and time-consuming. This process strongly relies on the expertise of specialists and is subject to inter-observer variability. Lesions can go unnoticed and misdiagnosis can occur, which are both caused by fatigue and inexperience. Inaccuracies of this nature can adversely affect patient care by causing delays in treatment [5][8][4]. Deep learning (DL) and more specifically convolutional neural networks (CNNs) have ushered in a paradigm shift in medical imaging. These algorithms can learn to identify intricate patterns and features in images on their own, which helps to make accurate predictions when diagnosing diseases. In gastroenterology, the use of computer-assisted techniques for the automated diagnosis of endoscopic images is under development to improve the accuracy of physicians' diagnosis, reduce the rate of errors, as well as to save time, as endoscopists' third eye [3][5].A thorough literature review is required to document the state of the art, compare the effectiveness of different techniques, and identify common challenges and issues, given the rapid growth of research in this. The aim of this review is to give a systematic overview of the DL approaches used to classify GI diseases from endoscopic images. We aim to analyze the different architectures, data handling strategies, and evaluation frameworks to highlight key trends and inform future research directions.

II. ANATOMY AND CLINICAL CATEGORIES OF GI DISEASES

The GI tract is a tube in which diseases manifest as visible alterations in the mucosa (which lines the tube). Endoscopy is used to detect diverse clinical types [8]:

- Polyps: Polyps are mucosa growths that can turn cancerous, causing tissue damage. Early detection and removal are important for colorectal cancer prevention [11].
- Ulcers and inflammation: Esophagitis (inflammation of the esophagus), ulcerative colitis (inflammation of the large bowel) and peptic ulcers result in sores, redness, and swelling on the GI lining [4][11].
- Anatomical Markers: Key landmarks like the Z-line (esophagus-stomach junction) pylorus (stomach-duodenum opening) and cecum are critical references for appropriately orienting the procedure and ensuring completeness [8][11]. The early diagnosis of these findings is crucial for medical intervention that improves mortality outcomes and is timely [8].

III. DATASETS FOR GI DISEASE CLASSIFICATION

Deep learning models for classifying gastrointestinal diseases rely on large public datasets with expert annotations for their development and validation. Different important resources are often used for training and scaling new algorithms are standards.

- KVASIR: This is a fundamental dataset, used for gastrointestinal disease detection. The initial version contained 4,000 images. These classes go over landmarks, diseases, and therapies that could be seen on endoscopy. The images of the clinical devices vary in sizes from 720×576 pixels to 1920×1072 pixels [8].
- HyperKvasir: This is an unusually large and comprehensive data collection of more than 110,000 images and 374 videos from the examinations at Bae rum Hospital in Norway. This big collection has 10,662 labelled and sorted images in 23 classes. HyperKvasir is a dataset that contains images of endoscopic procedures, but has a serious distribution problem because of the high imbalance of samples between the classes. Thus, some classes contain very few images which will hamper training generalizable models [8].
- The Kvasir-Capsule WCE: This dataset has a specific focus on image data produced by Wireless Capsule Endoscopy (WCE) technology. The dataset can be extracted from 117 videos with more than 4.7 million images. Out of the above, a total of 47,238 frames have been annotated by medical experts into 14 classes. This dataset, like HyperKvasir, suffers from considerable class imbalance which has to be accounted for when building models [9].

Table 1.COMPARATIVE ANALYSIS OF DATASETS FOR GI DISEASE CLASSIFICATION

Dataset	Total Images	Labeled Images	Annotation Quality & Modality
KVASIR	4,000	4,000	Tagged and validated by experienced endoscopists; standard endoscopic images.
HyperKvasir	>110,000	10,662	Labeled by experienced endoscopists; includes images and videos.
Kvasir-Capsule (WCE)	>4.7 million frames	47,238	Medically confirmed by specialists; Wireless Capsule Endoscopy (WCE) images.

IV. DEEP LEARNING APPROACHES FOR GI IMAGE CLASSIFICATION

Different deep learning models, ranging from basic to complicated hybrid, have widely improved the classification of gastrointestinal images. There are three main categories of these methods.

A. Convolutional Architectures

Modern GI image classification mostly utilizes deep learning models with convolutional layers that can learn the hierarchy of spatial features. Transfer learning is the most popular strategy, where the models are fine-tuned for endoscopic image analysis. Here, established architectures pretrained on large-scale image datasets are used [11].

Architectures: Popular models like VGGNet, ResNet, Inception, DenseNet, and EfficientNet are often used for this. The models are performing well, with reported accuracies of 98.01% with EfficientNetB0 and 98.30% with a VGG-based model in several research works. Due to their skip connection ability to mitigate the vanishing gradient problem, ResNet architectures are very effective for training very deep networks [1][8][11].

- Customization and Limitations: Some researchers will take it further than transfer learning and build their own custom deep learning models. With this method, it is possible to find architectures for GI imaging characteristics. They are usually less parameterized to reduce computing costs and model complexity. Nonetheless, a typical drawback of these designs is that their last fully-connected layers may suffer from over-fitting, thus impairing the model's performance on non-training data [9][11].

B. Hybrid and Ensemble Models

To enhance behavior and strength of individual models, many researchers chose hybrid and ensemble strategies that combine multiple models to work together.

- Stacking models: Stacking model is a complete advanced technique of ensemble models. It refers to training various models and the predictions of those individual base models are passed to another model called a meta-learner, which can be a CNN + SVM model or a Multi-Layer Perceptron. By using diverse predictive strengths of the base learners, this fusion strategy achieves state of the art results with accuracies of 98.42% and 98.53% on public benchmark datasets [8].
- Voting Classifier: The Voting classifier is a more straightforward ensemble method. It takes predictions from independent models and makes a final prediction based on the majority of the votes. The idea behind this is that one model's mistakes can be corrected through the right results generated from another model. One study, for example, utilized a voting classifier that combined features obtained from ResNet50 and EfficientNetB7[1].
- Other Hybrids: Besides the previously mentioned hybrid systems, researchers also experimented with other hybrid strategies, but no notable improvement was reported concerning the classifier choice. One such strategy involved appending a Long Short-Term Memory (LSTM) based classifier to the feature extraction layers of a deep network in [8].

C. Transformer and Attention based Models

There are new image classification tasks in GI with architectures designed by the successes of transformer architectures in AI domains.

- Architectures: A Vision Transformer (ViT) is an architecture that takes an image as a sequence of patches. They use the relationships between patches to perform tasks like classification or segmentation based on their spatial arrangement. The Swin Transformer is a newer hierarchical variant specifically designed for computer vision that has achieved increased efficiency and performance. Models are often used in a hybrid fashion, for instance, by fusing Swin Transformer with a convolutional architecture like Xception to reap the complementary benefits of the two architectures [6].
- Role of Attention: The attention mechanism is the core strength of these models. This enables the model to attend to the most useful parts of an image while ignoring less useful background areas. Items of interest are classified and integrated with GI surveillance imaging. This capability is particularly impactful for GI diagnosis, as it facilitates the automatic specification of localization and analysis of small polyp, subtle ulcer, or distinct area of inflammation.

V. PREPROCESSING AND DATA AUGMENTATION TECHNIQUES

Prior to feeding endoscopic pictures to a deep learning model, preprocessing and data augmentation techniques are performed to standardize the data, increase the dataset diversity, and enhance the overall performance and robustness of the model. Making reliable diagnostic tools requires these steps.

Image Preprocessing:

To improve image quality and consistent among all the images, preprocessing is done.

- Resizing and Normalization: We resize the image in our image processing project since images are of a very high resolution that uses very high computational power. Moreover, deep learning frameworks also require the input image to be of the same size. Resize the image to have the same shape, preferably a square shape. This step is important because the pixels must be normalized and can help accelerate training. A common normalization step is to rescale pixel values between 0 and 1 [7][8][6].

• Filtering and Cleaning: Noise and artifacts, reflections, etc. may not lead to a problem but can decrease a model's performance. BFs are a common filtering method that smoothens pixels while retaining edge details. Clustering to filter out useless patches having useless mostly image background is also used [3][5][7].

• Stain Color Normalization: Due to different stains and scanners there are heavy color variations in histopathological images. This step reduces this variation. To prevent bias, it is necessary to perform stain color normalization to standardize the color profile across all images prior to analysis [7].

Data Augmentation:

Data augmentation is used to create data artificially to increase the size of training set. This is a very useful technique when the data at hand is medical data and may be small in size. The process improves the network's performance on the training set and prevents overfitting.

• Geometric Transformations: An image augmentation technique alters the original image and its ground truth mask. They consist of rotating, flipping, translating, and zooming [8][3].

• Handling Class Imbalance: Medical datasets often have significantly more examples of healthy tissue than of rare diseases. To solve this problem, oversampling techniques were adopted to generate synthetic samples for the minority classes. An interesting technique called BORDERLINE SMOTE (BL-SMOTE), has been used to create new instances of the different minority classes (rather than the majority) in order to make it more balanced dataset so that the model does not get biased towards the majority class [9]. Using these techniques, the researchers are able to ensure a more diverse and balanced training set which enhances the ability of the model to generalize to new and previously unseen endoscopic images.

VI. FEATURE EXTRACTION AND SELECTION TECHNIQUES

The success of a deep learning model in classifying gastrointestinal (GI) diseases relies significantly on its ability to extract and select the important visual features present in endoscopic images. Researchers employ several key strategies to achieve this.

A. Use of Pretrained Models (Transfer Learning).

Transfer learning is the most powerful and effective tactic for feature extraction. This method uses deep learning algorithms that were initially trained on other large, diverse image datasets, such as ImageNet. The pretrained models like VGGNet, ResNet, EfficientNet have strong fixed features which can be utilized from their first layer. Researchers can create rich hierarchical feature maps by passing endoscopic pictures through these layers, avoiding the need to train a deep network from scratch. This method greatly diminishes the requirement for large quantities of labeled data and saves computational expenses, making it a highly efficient method widely accepted.

B. Custom CNNs for Feature Maps.

Some researchers do not use pretrained models but instead create their own Convolutional Neural Networks (CNNs) from scratch to generate feature maps. This can create models tailored to the specific visual features of endoscopic images where only minor variation in mucosal texture, coloration and morphology takes place. These custom-made models are often designed to be lightweight with fewer parameters. This reduces model complexity and speeds up processing which is vital for possible real-time clinical use [9].

C. Optimization Algorithms.

Various optimization algorithms are employed to refine extracted features and improve their quality. Further, an optimization algorithm is used to select the best features or tune the hyperparameters of the used feature extraction models. Many studies have used nature-inspired and some other optimization techniques. These include.

• The Improved Spotted Hyena Optimizer (ISHO) has been used to improve the performance of a ShuffleNet feature extraction model.

• Bayesian optimization has been used to optimize the hyperparameters of a MobileNet-V2 architecture to find the best deep learning features.

• Moth-crow optimization integrated with Canonical Correlation Analysis fusion to improve GI diseases recognition [8].

These algorithms help speed up the identification of best model ensembles and thereby improve diagnostic accuracy.

d. Comparative Analysis of Extraction Depth and Clinical Impact

The clinical utility of a feature is determined by the depth at which it is extracted from a network. Deep architectures learn a hierarchy of features with shallow and deep layers performing different but complementary diagnostics roles [5][7][9].

Table 2.Comparative Analysis of Extraction Depth and Clinical Impact

Extraction Depth	Feature Type	Clinical Impact & Application
Shallow Layer Extraction	Low-level features such as: Edges and lines and Colors, simple textures, Basic local patterns	Extracts low-level features like edges, colors, and textures. These are useful for capturing basic visual differences but are less specific to complex diseases. They can be used to classify broad categories, such as different organ types
Deeper Layer Extraction	High-level, abstract features corresponding to: Complex pathologies (e.g., polyp structures, inflammatory patterns), Learned combinations of low-level features into meaningful concepts	Extracts high-level, abstract features that correspond directly to complex pathologies like polyp structures or inflammatory patterns. These features are essential for the final, fine-grained classification of specific diseases

VII. PERFORMANCE METRICS AND EVALUATION

The performance metrics almost always arise from the four outcomes of a confusion matrix. These are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) [8].

• Accuracy: Accuracy is the most intuitive metric that gives the ratio of all correctly classified images to the total number of images. The results can be misleading when classes have unequal instances.

• Precision: Tells us how many of the items we predicted are actually true. In other words, out of all the Zoomers we predicted will enjoy the song, how many actually did. This tells us how many of the diseased images were really diseased. “The necessity of high precision in medical applications stems from the consequences of a Type I error, or false alarm”

• Recall or Sensitivity: Tells how well the model identifies relevant occurrences in a dataset. It addresses the query, how many of the actual diseased images were correctly identified by the model?[9].

• F1−Score: The F1−score is the harmonic mean of precision and recall. That is the weighted average of the two. It is especially good in evaluating models on imbalanced classification problems [8][9].

AUC - ROC Curve: ROC is a probability curve and AUC is the measure of separability. It tells how much the model is capable of distinguishing between classes. AUC values range from 0.5 (no discriminative ability) to 1 (perfect classification) for a Single Classifier [8].

• The Matthews correlation coefficient (MCC): MCC is the very reliable metric for unbalanced classification problems. The strength of this metric is that it takes into account all four values of the confusion matrix. It generates a score that ranges between -1 and +1. A score of +1 indicates a perfect prediction, and a score of 0 indicates a random guess. Finally, -1 indicates complete disagreement [8].

We often use k-fold cross-validation in order to obtain a better evaluation that is not reliant on any particular split of the data. The dataset is split into ‘k’ folds. The model will train and test in ‘k’ times. This produces a better estimate of the performance of the model on unseen data [8]

VIII. REVIEW OF EXISTING APPROACHES

Gupta and others used Kvasir v2 dataset and features using ResNet50 and EfficientNetB7. We achieved 98.18% accuracy using CNN, random forest, and Kvasir classifier for classification. It was noted that a manual review of the GI images was a limitation due to its time-consuming nature [1].

According to Iyer et al, a deep learning model was developed using GI endoscopic images was created utilizing KVASIR dataset. Masking methods were used for noise reduction and transfer learning for improved performance. The model achieved 96.89% accuracy. There were various drawbacks, such as manual diagnosis which is subjective, the look of the lesion inconsistent, and the frames' image quality fairly poor [2].

According to Varalaxmi et al. (2023), ResNet50, a type of Residual Networks, is used by them for classification of diseases. The accuracy achieved by CNN Based Model was 88.05%, sensitivity was 87.05%, specificity was 92.33% The authors of the study recognized that endoscopy takes a long time to perform. Furthermore, the process is repetitive. Thus, it may inconvenience patients. Additionally, it may lead to misdiagnosis when relying on manual observations.

The team composed of Sharmila et al. (2022) is a work that uses both deep CNN and the pre-trained ResNet101 classification system for abnormalities in GI tract images from the KVASIR image dataset. The model achieved 98.37% accuracy. There are certain limitations that the authors discussed. For instance, it is hard for CADs to identify morphologically uneven abnormalities. In addition, the analysis of the endoscopic data is very time-consuming as well. The expert who analyses it also requires a lot of concentration and the task may also be prone to errors [4].

Alruban and colleagues improved the ShuffleNet used for feature extraction with another robust ShuffleNet, and proposed the EIAGTD-NIADL approach which made use of bilateral filtering for image preprocessing. They use layered long short-term memory networks for classification. The system performed well on benchmark medical datasets. Nonetheless, the paper did not mention any limitations explicitly [5].

Shahriar Hossain et al. (2024) developed a combined vision transformers-based transfer learning model based on Xception architecture. The proposed hybrid system was evaluated with ViT models like CCT, EANet, and was effective in achieving 97.22% accuracy. Some automated techniques are crucial due to the time-consuming manual classification processes and worldwide impacts of GI diseases [6].

Rasoul Sali and colleagues (2023) suggest a multi-category GI disorder diagnosis model based on an ensemble of 18 VGGNet CNNs. The model is hierarchical and takes in histopathological images. Each model branch focuses on a specific part of the GI tract, with their results merged for a well-rounded diagnosis. This hierarchical model is better than the flat classification models and it can discriminate, at a high resolution, diseases that have only subtle morphological differences. The high similarity of various GI disorders and the challenges associated with scaling multi-category classification beyond binary classification approaches posed significant challenges to the model despite the positive outcomes [7].

Esra Sivari et.al. A hybrid stacking ensemble model is introduced in (2023) that is based on deep learning. This is done for detection and classification of the finding related to the gastrointestinal tract that has been done using endoscopic images. The model employs a distributed search method to optimize the hyperparameters of various CNNs. The model achieved 98.42% accuracy on kvasir2 and 98.99% MCC on hyperkvasir showing high diagnostic precision. It is put forward as a tool to help specialists and decrease the burden of diagnosis. Yet, variability in image quality, inter-observer differences and the time-consuming nature for reviewing massive volume of endoscopic videos, remain hurdles [8].

Saqib Mahmood et.al. A deep CNN model for the classification of a peptic ulcer and digestive disorder from endoscopic images was proposed (2022). To deal with class imbalance, we apply a technique called BORDERLINE-SMOTE, while relying on the class activation maps (CAMs) to interpret the model. It displayed a good performance with 98.9% accuracy and 98.7% F1 score. The technique successfully automates diagnosis, but clinical workflow bottlenecks emerge when large image sets are interpreted manually. Furthermore, traditional endoscopy is invasive, though Wireless Capsule Endoscopy (WCE) is available [9].

The Kvasir-V2 dataset (8000 images across 8 classes) is transferred learning in this transfer learning research by Muhammad Nouman Noor et.al. (2023) through the pre-trained deep CNNs ResNet50, InceptionV3, EfficientNetB0. The EfficientNetB0 had the best accuracy at 97.8% by adding the customized classification head and checking the performance at accuracy, precision, recall and F1 score. The study doesn't talk about limitations directly, while it shows that transfer learning works in this domain; we can still think of challenges like class imbalance or real-world variability [10].

J. V. Thomas Abraham et.al. (2023) proposed a transfer-learning-based model to classify digestive diseases using pre-trained architectures such as EfficientNetB0. The best performing model was EfficientNetB0 with a score of 98.01%, a precision of 98%, and a recall of 98%. The model is promising for clinical applications, but it highlights the wider issues in the classification of digestive diseases. There is variation in the data available and high quality and diverse datasets are needed to generalise [11].

Karthik Ramamurthy's group purposed a model that combines the features from EfficientNetB0 and MobileNetV2 to classify GI diseases using endoscopic images. The model is trained and tested on HyperKvasirs Dataset using concatenation of feature vectors and SVM classifier. The fusion strategy improved classification performance over individual models. Even though no specific limitations are outlined, the method presumes superior feature complementarity and may suffer from feature redundancy or dataset variation [12].

Melaku Bitew Haile et.al. The technique utilizes the SVM classifier and Kvasir image datasets, including a minimalistic feature extraction model. The workflow is to resize and normalize a standard way. Findings indicate that this hybrid CNN-SVM model demonstrates good ability to identify abnormalities of the GI tract as it offers a good learning ability and interpretability of the classifier. Though, the authors haven't outlined the limitations, classical SVMs might get vulnerable in richer and noisy data settings [13].

Yaw Afriyie et.al (2022) proposes a denoising capsule network for improved gastrointestinal disease classification in noisy endoscopic images. The model keeps the spatial features related through capsule architecture and combines it with denoising for robustness. The Kvasir dataset shows that the accuracy of the strategy is high and fairly robust. Even though there are no specified constraints there could be challenges such as increased model complexity and computational cost of capsule networks [14].

In their 2023 study, Mousa Alhajlah and colleagues applied pre-trained CNN models such as VGG16, ResNet50, and InceptionV3 for feature extraction, followed by techniques that optimize features from the dataset, improving classification and accuracy. Kvasir dataset: The model achieved high classification accuracy. Removal of irrelevant or redundant features improved the performance of the model. A key challenge pointed out is the visual similarity of healthy and infected regions, which can cause even the best classifiers to fail. Subsequently, refined selection of features is essential [15]. The AASAN model is proposed by Sheng Li et.al. (2022) which mimics how endoscopists diagnose using global, local and fusion branches. It implements a self-attention module and a relative position encoding (SA-RPE) and an adaptive aggregation feature (AAF) module. The accuracy of the model on Kvasir is 96.37%. Still, it continues to deal with ambiguous boundary lesions, shape and size variability, and bubbles, turbidity artifacts, etc. [16].

Yixin Liu et al (2022) combine the Xception CNN model with a residual attention mechanism for an improvement of ulcer classification. Using the Sobel operator for preprocessing offers a solution to the illumination unevenness in endoscopic frames. The improved model did the job with an accuracy of 81.41%. The model's evaluation on the clinical data remains a limitation [17].

Diego Marin-Santos et al. (2022) paper presents a CNN designed to detect Crohn's disease from CE images. The model attained an AUC above 0.997 and 95–99% sensitivity and 96–99% specificity on a test set, with the use of over 15,000 images from 31 video cases. Finally, the efficient net and ResNet models were surpassed in classifying stumped and unmoved positions using the model. One of the major problems raised is the high manual workload that arises from the excessive length of CE reviews [18].

The study done by Geonhui Son et.al (2022) presents an automated method for small bowel detection in WCE videos using a temporal ResNet50 and hybrid filtering. Gained 96.0% accuracy and helps decrease endoscopy reading time Due to motion artifacts, uneven illumination, and subjective nature of manual assessments, there can be misdetections [19].

Zeyad Ghaleb and others suggested using models like VGG-16 and DenseNet-121 alongside features GLCM, DWT, and LBP for diagnosis. A fused features ANN gave best accuracy of 98.9% and specificity of 99.69%. A small set of datasets, similarity of image parameters between diseases and manual review is too time consuming [20].

IX. RESULTS AND DISCUSSIONS

Researchers have created a variety of powerful deep learning models for detecting gastrointestinal diseases,

with many achieving over 95% accuracy on benchmark datasets like Kvasir. Most successful methods combine techniques, such as fusing features from CNN (e.g. DenseNet-121) with handcrafted features or utilizing advanced architectures like EIAGTD-NIADL, which have achieved 98.9% and 98.96% accuracy, respectively [5][16][20]. Comparative analyses reveal that more complex multi-branch feature fusion models outperform simple CNNs across multiple datasets, thus demonstrating the effectiveness of well-designed model architectures. For instance, the attention mechanism-enhanced AASAN model (96.37 percent accuracy) outperformed the baseline DenseNet version (91.67 percent accuracy) and application of a temporal filter to the ResNet50 model resulted in an accuracy upgrade from 88.0 percent to 99.8 percent for organ detection. A hybrid Swin Transformer-Xception model achieves an accuracy of 87.23% while exploring newer architectures like Vision Transformers. Some models are found to perform at par or even better than expert gastroscopists in classification
[6][16][17][19].

X. SUMMARY

Table 3.A summary of 20 important studies on Gastrointestinal Disease Detection and classification Using a Tailored Convolutional Neural Network Layer

Reference No	Paper Title	Authors	Methods Used	Key Results	Research gap and Limitations
1	Classification of Endoscopic Images and Identification of Gastrointestinal Diseases	Gaurish Anand, Dev Gupta (2022)	Machine learning approach using KVASIR-V2 dataset, Voting Classifier for classification.	Achieved 88.19% accuracy, automated disease detection, reduces human error.	Limited to specific dataset, might not generalize well to all GI diseases.
2	Deep Learning Model for Disease Prediction Using Gastrointestinal Endoscopic Images	Iyer, Slobodan Narmadha, Sundar (2023)	Transfer learning and dataset expansion to achieve better accuracy for GI disease prediction.	96.89% accuracy, robust model with expanded dataset, improved computational efficiency.	High computational cost, dependent on quality of dataset and pre-trained models.
3	Diagnosis of Gastrointestinal Diseases Using Modern CNN Techniques	Guduru Varalaxmi, K. Swaraja, Sahithi Reddy Baddam (2023)	Feature extraction using ResNet50 and EfficientNetB7, Voting Classifier for GI disease detection.	Achieved 88.05% accuracy, faster diagnosis, outperforms traditional methods.	Requires large dataset, traditional methods still widely used in practice.
4	Detection and Classification of GI-tract Anomalies from Endoscopic Images Using Deep Learning	Sharmila, V. Geetha, S (2022)	Deep learning models applied to detect GI anomalies from endoscopic images.	Improved diagnostic accuracy, automated and scalable solution for GI diseases.	May not handle all types of GI anomalies, requires high quality labeled data for training.

5	Endoscopic Image Analysis for GI Disease Diagnosis Using Nature Inspired Algorithm	Abdulrahman Alruban, Eatedal Alabdulkreem (2023)	Bilateral filtering, Enhanced ShuffleNet, Spotted Hyena Optimizer, Stacked LSTM.	96.89% accuracy, robust model with expanded dataset, improved computational efficiency.	98% accuracy, real-time potential, superior performance.
6	Gastrointestinal Insights Redefined: Hybrid Model Using Vision Transformer & Transfer Learning	Shahriar Hossain, Md. Fahim-Ul-Islam (2024)	Swin Transformer + Xception, Transfer Learning.	87.23% accuracy, improved automation, robust evaluation.	Lower accuracy, high computational needs, dataset reliance.
7	Hierarchical Deep Convolutional Neural Networks for Multi-category Diagnosis of Gastrointestinal Disorders on Histopathological Images	Rasoul Sali, Sodiq Adewole, Lubaina Ehsan, et al. (2020)	Hierarchical deep CNN, ensemble model, transfer learning	Mean micro- and macro-level F1-score of 99.4%	Misclassification by flat models, hierarchical model reduces errors
8	A New Approach for Gastrointestinal Tract Findings Detection and Classification: Deep Learning-Based Hybrid Stacking Ensemble Models	Esra Sivari, Erkan Bostanci, Mehmet Serdar Guzel, et al. (2023)	Hybrid stacking ensemble, pre-trained models, Python (Keras, Scikit-learn)	98.42% accuracy, 98.42% precision, 98.39% recall, 98.39% F1-score	Variability in endoscopic observations, costly repetitive endoscopies
9	A Robust Deep Model for Classification of Peptic Ulcer and Other Digestive Tract Disorders Using Endoscopic Images	Saqib Mahmood, Mian Muhammad Sadiq Fareed, et al. (2022)	Custom CNN, BORDERLINE-SMOTE for imbalanced datasets	98.9% accuracy, 98.8% precision, 98.8% F1-score, 98.7% recall	Imbalance in medical datasets, handled by BORDERLINE-SMOTE
10	Efficient Gastrointestinal Disease Classification Using Pretrained Deep Convolutional Neural Network	Muhammad Nouman Noor, Muhammad Nazir, et al. (2023)	Pretrained deep CNNs, brightness-contrast enhancement, SGA, VIF similarity index	15.26% improvement in accuracy, 13.3% precision, 16.77% recall	Wireless Capsule Endoscopy challenges, large variations in images
11	A Deep-Learning Approach for Identifying and	J. V. T. Abraham, A. Muralidhar, K. Sathyarajasekaran, N. Ilakiyaselvan	Transfer learning (ResNet50, InceptionV3, DenseNet121,	EfficientNetB0: 98.01% accuracy, 98% precision/recall;	Manual review time-consuming; expert inconsistency; high resource needs; WCE

	Classifying Digestive Diseases	(2023)	EfficientNetB0); Custom CNN; Kvasir dataset (5 classes); TensorFlow/Keras	Custom CNN: 89% accuracy (overfitting)	image quality issues; overfitting in custom CNN
12	A Novel Multi-Feature Fusion Method for Classification of Gastrointestinal Diseases Using Endoscopy Images	Karthik Ramamurthy et al. (2022)	Handcrafted (LBP, HOG) + Deep features (EfficientNetB0); PCA for selection; SVM, KNN classifiers; Kvasir-V2	98.24% accuracy (fusion) vs. 97.4% (EffNet only); PCA improved performance	No explicit section; implies single-feature/model methods underperform; potential high-dimensionality
13	Detection and classification of gastrointestinal disease using CNN and SVM	Melaku Bitew Haile et al. (2022)	VGG16 for feature extraction; SVM classifier; Kvasir dataset	96.8% accuracy	Generalizability concerns; computational cost; interpretability; possible class imbalance
14	Gastrointestinal tract disease recognition based on denoising capsule network	Yaw Afriyie et al. (2022)	Denoising capsule networks; activation maps	94.16% accuracy; 83.1% precision; 86.7% sensitivity; 96.1% specificity	CapsuleNets struggle with complex images; manual diagnosis slow; CNNs [need augmentation
15	Gastrointestinal Diseases Classification Using Deep Transfer Learning and Features Optimization	Mousa Alhajlah et al. (2022)	ResNet-50/152 via transfer learning; Mask R-CNN; Ant Colony Optimization	96.43% accuracy (optimized); ResNet-152 alone: 90.62%	Similarity in infected/healthy regions causes misclassification; WCE manual review is slow
16	Adaptive aggregation with self-attention network for gastrointestinal image classification	Sheng Li et al. (2022)	Self-attention and global/local/fusion branches	96.37% accuracy on Kvasir dataset and Outperformed SOTA DL models	Ambiguous lesions, varied sizes, artifacts
17	An Xception model based on residual attention mechanism for the classification of benign and malignant gastric ulcers	Yixin Liu et al. (2022)	Xception and attention modules, Sobel preprocessing	81.41% accuracy, 83.75% sensitivity	Needs real-world validation
18	Automatic detection of Crohn's disease in wireless capsule endoscopic images using a deep	Diego Marin-Santos et al. (2022)	CNN on CE images (15,972)	AUC > 0.997, 95–99% sensitivity	Long manual review time

	convolutional neural network				
19	Small Bowel Detection for Wireless Capsule Endoscopy Using CNNs with Temporal Filtering	Geonhui Son et al. (2022)	ResNet50 and temporal filtering (WCE videos)	96% accuracy for small bowel	Misalignment, lighting, frame noise
20	Hybrid Techniques for Diagnosing Endoscopy Images for Early Detection of GI Disease Based on Fusion Features	Zeyad Ghaleb Al-Mekhlafi et al. (2023)	VGG/DenseNet ,handcrafted features and PCA	98.9% accuracy, 99.69% specificity	Image overload, visual similarity issues

XI. KEY CHALLENGES IDENTIFIED

1. Data and Image Quality Issues.

A. Poor Image Quality: The problem of poor image quality has been highlighted in various studies. Issues such as inconsistent lighting, motion artifacts, bubbles, turbidity and poor image quality of video frames hinder the performance of some diagnostic models [2][16][19].

B. Limited and Imbalanced Datasets: Having datasets that are large enough, diverse enough and well-annotated enough is an all-too-common problem. Some models do train on limited data which impacts their ability to generalize to real-world clinical scenarios.

C. Data Variability: Endoscopic images can vary significantly, influencing performance in models [2][11].

2. Challenges in Manual Diagnosis (The problems automation aims to solve).

D. Time-Consuming and Repetitive: Going through thousands of images of an endoscopic procedure is very time-consuming, responsible for a heavy workload on the specialists, and prone to fatigue-induced errors [1][3][4][6][8][18][20].

E. Subjectivity and Misdiagnosis: Diagnosis of a patient is subjective in nature and can vary from clinician to clinician (inter-observer variability). This could result in inconsistencies and possible wrong diagnoses, particularly in the presence of subtle abnormalities with the affected individual [2][3][8].

F. High Level of Concentration Required: The researcher must use high-level concentration while doing the analysis. Otherwise, we can miss a lot of things [4].

3. Intrinsic Difficulties in Lesion/Disease Identification.

G. High Visual Similarity Between Diseases: Gastrointestinal illnesses or even healthy tissue and diseased tissue looking alike morphologically makes it difficult for systems to detect diseases accurately [7][15][20].

H. Ambiguous or Uneven Lesion Characteristics: A challenge for detection and classification algorithms is presented by the ambiguous boundaries, varying shape and size of lesions [4][16].

4. Model and Algorithm Limitations

I. Scalability of Models: The complexity of real-life problems makes it difficult to scale the classification models beyond basic binary problem (disease or no disease) or a limited multi-class problems especially for a large number of diseases which are quite similar [7].

J. Real World Validation: There is a significant gap in achieving high accuracy on a carefully curated dataset and validating the performance and reliability of the model in real clinical settings [17].

K. Model Complexity and Cost: The use of advanced model, such as capsule networks, can bring better performance as discussed by [14], however, such advanced models come at a higher cost of computational complexity, which can create a hurdle to implementation.

XII. RESEARCH GAP

• Although models display remarkable accuracy in laboratory conditions, they rarely undergo testing in clinical trials. In other words, they are never clinically validated, rendering them useless in practice.

• Instead of just naming a disease, more clinically useful tasks such as localization of a lesion (where the lesion is), and segmentation (where the lesion ends), are required.

• Some models can assess the severity (e.g. mild vs severe ulceration) of a disease, which is useful for planning treatment. • Poor Generalization: Models trained on one dataset (like Kvasir) may not generalize well to image data coming from another hospital with different equipment, a problem known as domain shift.

• Overuse of imaging from video data during endoscopy has been quite common among the researchers, who are ignoring the time-based information.

• Current models cannot explain any better why they reached a diagnosis, thus the biggest roadblock of earning the trust of clinicians is explainability.

• The absence of multi-modal fusion implies that the studies mainly focus on the images only and ignore the others. These others can be patient symptoms, history, lab results, etc. However, the inclusion of these non-image data can improve the accuracy of the diagnosis.

• Dealing with multiple diseases: Model is designed to detect single disease in an image. However, in clinical practice, it is a common occurrence to find co-existing findings in the same procedure.

XIII. FUTURE SCOPE

Future of Predicting GI Disease with AI will focus on its clinical applicability and robustness. Progress will entail the advancement of convolutional neural network (CNN) architectures that go beyond simple classification towards more complex tasks such as lesion segmentation, automated severity grading and identification of multiple diseases in one go. Developing models that generalize well enough for different clinical settings will be an important area of research. The CNN models will become more robust through manipulations of training data. Training data manipulations will heavily rely on complex data-augmentation strategies. These strategies are those involving GANs to develop synthetic images, etc. We will develop user-friendly, rigorously validated tools to analyze real-time video and ensure integration with patient health records after testing on a larger scale in clinical trials.

XIV. CONCLUSION

The automated diagnosis of gastrointestinal (GI) diseases relies heavily on deep learning, most importantly on convolutional neural networks (CNNs), based on the important 20 studies. Typical methods involve transfer learning using established architectures such as ResNet, VGG, DenseNet, and Xception. More sophisticated methods develop hybrid models that integrate characteristics from various CNNs or combine deep features with handcrafted features (for example, GLCM, DWT). In order to enhance performance, various techniques are used such as ensemble classifiers and attention mechanisms to focus on the relevant region of the lesion. Novel architectures like Vision Transformers (ViT) are used, some of which are also coupled with CNNs. Other specialized techniques process video data by temporally filtering CNN output.

There are key gaps in research that show the need to have computationally efficient models and these models should be able to support real-time diagnostics and deployment on edge devices. This is necessary because many frozen models are resource intensive.

Experts always recommend validating these models on bigger and diverse multi-center and multi-vendor datasets for ensuring clinical generalizability and expanding their utility to a wider spectrum of GI diseases. Applying these technologies in primary care settings with limited data is challenging. Various studies have been conducted on various topics in quantum physics. However, all of them faced the problem of model complexity and data complexity. Datasets are typically small, drawn from a single institution, and often of low resolution with noise and other artifacts. These issues can hinder performance. It's hard for doctors and AI to accurately classify GI lesions since polyps, ulcers and the like can look very similar to each other.

REFERENCES

[1] D. Gupta, G. Anand, P. Kirar, and P. Meel, "Classification of Endoscopic Images and Identification of Gastrointestinal Diseases," in *Proc. 2022 Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput. (COM-IT-CON)*, Faridabad, India, May 2022, pp. 231–235, doi: 10.1109/COM-IT-CON54601.2022.9850571.

[2] S. Iyer, D. Narmadha, G. N. Sundar, S. J. Priya, and K. M. Sagayam, "Deep Learning Model for Disease Prediction Using Gastrointestinal-Endoscopic Images," *2023 4th International Conference on Signal Processing and Communication (ICSPC)*, Coimbatore, India, Mar. 2023,pp.133–136,doi: 10.1109/ICSPC57692.2023.10126043

[3] G. Varalaxmi, K. Swaraja, S. R. Baddam, K. R. Madhavi, E. S. Yalamarthi, and C. N. Sujatha, "Diagnosis of Gastrointestinal Diseases Using Modern CNN Techniques," in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, Pune, India, Apr. 2023, pp. 1–6, doi: 10.1109/I2CT57861.2023.10126259

[4] S. V. and G. S., "Detection and Classification of GI-Tract Anomalies from Endoscopic Images Using Deep Learning," in *2022 IEEE 19th India Council International Conference (INDICON)*, 2022, pp. 1–6, doi: 10.1109/INDICON56171.2022.10039766

[5] Alruban, E. Alabdulkreem, M. M. Eltahir, A. Alharbi, I. Issaoul, and A. Sayed, "Endoscopic Image Analysis for Gastrointestinal Tract Disease Diagnosis Using Nature Inspired Algorithm With Deep Learning Approach," *IEEE Access*, vol. 11, pp. 130024–130030, 2023. doi: 10.1109/COM-IT-CON54601.2022.9850571

[6] S. Hossain, M. Fahim-Ul-Islam, R. Rahman and A. Chakrabarty, "Gastrointestinal Insights Redefined: An Integrated Hybrid Model Fusing Vision Transformer and Transfer Learning," *2024 6th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, Dhaka, Bangladesh, May 2024, pp. 19–23, doi: 10.1109/ICEEICT62016.2024.10534523.

[7] R. Sali, S. Adewole, L. Ehsan, L. A. Denson, P. Kelly, B. C. Amadi, L. Holtz, S. A. Ali, S. R. Moore, S. Syed, and D. E. Brown, "Hierarchical deep convolutional neural networks for multi-category diagnosis of gastrointestinal disorders on histopathological images," *arXiv preprint arXiv:2005.03868v2*, Aug. 2020, doi: 10.48550/arXiv.2005.03868.

[8] E. Sivari, E. Bostanci, M. S. Guzel, K. Acici, T. Asuroglu, and T. E. Ayyildiz, "A New Approach for Gastrointestinal Tract Findings Detection and Classification: Deep Learning-Based Hybrid Stacking Ensemble Models," *Diagnostics*, vol. 13, no. 4, p. 720, 2023, doi: 10.3390/diagnostics13040720.

[9] S. Mahmood, M. M. S. Fareed, G. Ahmed, F. Dawood, S. Zikria, A. Mostafa, S. F. Jilani, M. Asad, and M. Aslam, "A Robust Deep Model for Classification of Peptic Ulcer and Other Digestive Tract Disorders Using Endoscopic Images," *Biomedicines*, vol. 10, no. 9, p. 2195, 2022, doi: 10.3390/biomedicines10092195.

[10] M. N. Noor, M. Nazir, S. A. Khan, O.-Y. Song, and I. Ashraf, "Efficient Gastrointestinal Disease Classification Using Pretrained Deep Convolutional Neural Network," *Electronics*, vol. 12, no. 7, p. 1557, 2023, doi: 10.3390/electronics12071557.

[11] J. V. T. Abraham, A. Muralidhar, K. Sathyarajasekaran, and N. Ilakiyaselvan, "A Deep-Learning Approach for Identifying and Classifying Digestive Diseases," *Symmetry*, vol. 15, no. 3, p. 723, 2023, doi: 10.3390/symmetry15030723.

[12] K. Ramamurthy, T. T. George, Y. Shah, and P. Sasidhar, "A Novel Multi-Feature Fusion Method for Classification of Gastrointestinal Diseases Using Endoscopy Images," *Diagnostics*, vol. 12, no. 10, p. 2316, 2022, doi: 10.3390/diagnostics12102316.

[13] M. B. Haile, A. O. Salau, B. Enyew, and A. J. Belay, "Detection and classification of gastrointestinal disease using convolutional neural network and SVM," *Cogent Engineering*, vol. 9, no. 1, p. 2084878, 2022, doi: 10.1080/23311916.2022.2084878.

[14] Y. Afriyie, B. A. Weyori, and A. A. Opoku, "Gastrointestinal tract disease recognition based on denoising capsule network," *Cogent Engineering*, vol. 9, no. 1, p. 2142072, 2022, doi: 10.1080/23311916.2022.2142072.

[15] M. Alhajlah, M. N. Noor, M. Nazir, A. Mahmood, I. Ashraf, and T. Karamat, "Gastrointestinal Diseases Classification Using Deep Transfer Learning and Features Optimization," *Computers, Materials & Continua*, vol. 75, no. 1, pp. 2235–2247, 2023, doi: 10.32604/cmc.2023.031890.

[16] S. Li, J. Cao, J. Yao, J. Zhu, X. He, and Q. Jiang, "Adaptive aggregation with self-attention network for gastrointestinal image classification," *IET Image Processing*, vol. 16, no. 8, pp. 2384–2397, 2022, doi: 10.1049/ipr2.12495.

[17] Y. Liu, L. Zhang, Z. Hao, Z. Yang, S. Wang, X. Zhou, and Q. Chang, "An xception model based on residual attention mechanism for the classification of benign and malignant gastric ulcers," *Scientific Reports*, vol. 12, no. 1, p. 15365, 2022, doi: 10.1038/s41598-022-19639-x.

[18] D. Marin-Santos, J. A. Contreras-Fernandez, I. Perez-Borrero, H. Pallares-Manrique, and M. E. Gegundez-Arias, "Automatic detection of crohn disease in wireless capsule endoscopic images using a deep convolutional neural network," *Applied Intelligence*, vol. 53, no. 14, pp. 12632–12646, 2023, doi: 10.1007/s10489-022-04146-3.

[19] G. Son *et al.*, "Small Bowel Detection for Wireless Capsule Endoscopy Using Convolutional Neural Networks with Temporal Filtering," *Diagnostics*, vol. 12, no. 8, p. 1858, 2022, doi: 10.3390/diagnostics12081858.

[20] Z. G. Al-Mekhlafi, E. M. Senan, J. S. Alshudukhi, and B. A. Mohammed, "Hybrid Techniques for Diagnosing Endoscopy Images for Early Detection of Gastrointestinal Disease Based on Fusion Features," *International Journal of Intelligent Systems*, vol. 2023, Article ID 8616939, 2023, doi: 10.1155/2023/8616939.