

MNETGIDD: A heuristic-oriented segmentation and deep learning multi-disease detection model for gastrointestinal tracts

A. Bamini^{1*}

¹Department of Computer Applications, The Standard Fireworks Rajaratnam College for Women, Sivakasi, Tamil Nadu, India

*Corresponding author E-mail: drbaminia@gmail.com

(Received 6 May 2024; Final version received 18 September 2024; Accepted 2 January 2025)

Abstract

Malignant growth of the gastrointestinal (GI) tract is among the leading causes of death worldwide. Research indicates that almost 40% of people worldwide suffer from long-term digestive issues. According to a study published in the United European Gastroenterology Journal, digestive disorders have increased since 2000. Digestive disorders continue to be a major cause of death, even with a slight decline. The World Health Organization's Mortality Database reported huge death rates every year due to GI diseases. From that report, the need to accurately detect GI tract malignant in low-cost and error-prone labor must be developed. This work introduces MNET Gastrointestinal Disease Detection (MNETGIDD), which is a complete identification model for multi-gastrointestinal disease discovery from clinical images. MNETGIDD model uses the Gastrolab dataset with endoscopic images, acting as pipelines that are pre-processed and segmented to identify the affected areas. This proposed approach aims to enhance image quality and facilitate accurate segmentation and classification through a pipeline process, initially preprocessing with techniques such as text removal, illumination enhancement, and fuzzy histogram equalization. During segmentation, Otsu segmentation based on Krill-Herd optimization was used to identify the affected area. The MNETGIDD model incorporates the MobileNetV2 architecture, designed for a lightweight classification model working under resource-constrained environments. According to the tests, the MNETGIDD model exhibits high sensitivity and specificity, often outperforming human experts. In terms of accuracy, the model achieved 96.349%, a precision of 96.25 %, and a recall of 97.08%. This deep learning system has the potential to revolutionize gastrointestinal disease diagnostics and screening by automating key steps and improving patient outcomes.

Keywords: Fuzzy Histogram Equalization, Gastrointestinal Disease Detection, MNETGIDD, Gastrolab dataset, Low-light Image Enhancement, Mean-Shift Segmentation, MobileNetV2

1. Introduction

In the world, cancer is the leading cause of death, and gastrointestinal (GI) cancer is the most common type. Globally, 1.8 million people die from gastrointestinal diseases each year (Sharmila & Geetha, 2022), and GI cancer is the fourth leading cause of death. Globally, gastrointestinal diseases are

a significant health burden. The Global Burden of Disease Study 2019 reports that digestive diseases cause over 2.5 million deaths worldwide, accounting for 4.5% of all deaths (Theo, 2019). Globally, digestive diseases caused 81.1 million disability-adjusted life years (DALYs).

The growth of GI polyps on the mucosa of the stomach and colon is the cause of gastrointestinal cancer. The esophagus, stomach, small intestine, large intestine, rectum, and anus are some of the parts of the digestive tract affected by multi-GI diseases. A chronic disease can significantly impact an individual's quality of life. The symptoms, causes, and treatments of multi-gastrointestinal diseases differ from one type to the next. An estimated 1.9 million new cases of colorectal cancer, one of the four diseases examined in this study, will be diagnosed in 2020 (Sung et al., 2020). Our research also focused on gastric cancer, which was responsible for over 768,000 deaths worldwide in 2020. There were 600,000 cases of esophageal cancer worldwide in 2020. Approximately 6.8 million people worldwide suffer from inflammatory bowel diseases, including Crohn's disease and ulcerative colitis (Alatab et al., 2020). GI diseases have a substantial economic impact. According to a 2015 study, digestive diseases cost the United States \$136 billion annually in direct and indirect costs (Perry et al., 2020). Many diseases and conditions can affect the esophagus. Inflammation of the esophageal lining can cause heartburn and swallowing difficulties. Another common condition is hemorrhoids, which are caused by inflammation of the blood vessels in the anus and rectum. In different parts of the digestive tract, polyps can develop, and while most are benign, some are cancerous. A type of inflammatory bowel disease, ulcerative colitis grade 1 causes inflammation and ulcers in the colon and rectum linings. Because of their complexity, these disorders can be challenging to diagnose and treat; imaging, lab testing, and invasive diagnostic procedures might be required. Recently, a promising approach for diagnosing and segmenting multiple gastrointestinal diseases was developed using deep learning techniques. A computer-aided automated approach may be useful for highly accurate polyp diagnosis and cancer detection. Artificial intelligence (AI) holds immense promise in helping people visualize diseases that are invisible to the human eye in various medical fields (Ekiri et al., 2016). Endoscopy images can be evaluated, and key features of micro-imaged structures can be identified using AI tools. The following have been made possible by this research.

- The proposed scheme identifies and classifies the different GI diseases.

- By combining segmentation using IP and identification techniques via deep learning, a novel technique was proposed.
- Using this method, images for training and testing are randomly selected in the segmented images without manual intervention

1.1. Aim & Motivation

Medical practice and healthcare systems worldwide can be improved through computer-assisted early disease detection. Multi-GI disease segmentation and identification is a challenging but important task that could improve healthcare outcomes for millions of people worldwide. A deep learning model can enable us to accurately identify and segment complex diseases like esophagitis grade A, hemorrhoids, polyps, and ulcerative colitis grade 1.

Improved patient outcomes depend on early detection and accurate diagnosis of gastrointestinal diseases. In a study by Smith et al. (2022), early detection of colorectal cancer can increase 5-year survival from 14% to 90%. This study proposed the MNET Gastrointestinal Disease Detection (MNETGIDD) model as an advanced diagnostic tool. A recent advancement in AI and deep learning has shown promising results. Based on a meta-analysis conducted by Johnson et al. in 2023, AI-assisted diagnosis of GI diseases could improve detection rates by 30%.

This work aimed to develop a framework, MNETGIDD, for recognizing a wide range of GI diseases simultaneously rather than multiple tools used individually to detect an anonymous disease. A significant improvement over most recent studies is that this study considers different types of GI diseases (both diseased and normal cases) related to the human GI tract. Furthermore, the model proposes a means of identifying diagnostic decisions based on deep learning techniques. A medical expert can validate the computer decision interactively with the assistance of such additional information. To address this problem, this work proposed a high-performance classifier and retrieval framework that uses endoscopic images to determine GI diseases using recent AI techniques.

1.2. Objective, Challenges and Issues

Deep learning techniques will be used to develop a model capable of accurately identifying the four common multi-gastrointestinal diseases: esophagitis

grade A, hemorrhoids, polyps, and ulcerative colitis grade 1. Deep learning will be used in conjunction with imaging data to train the proposed model to recognize patterns that may indicate the presence of these diseases. With the development of this model, the need for invasive diagnostic procedures will be reduced, and patient outcomes will improve. A multi-gastrointestinal disease model can also help healthcare professionals diagnose and manage these illnesses. A computer-aided diagnosis system was developed to aid medical experts in diagnosing different types of gastrointestinal diseases.

A model for segmenting and identifying multi-GI diseases is a challenging task that requires consideration of several ethical, practical, and technical factors. There are several challenges and issues associated with this research, including:

- **Limited availability of labeled medical imaging data:** Medical imaging data labeled with medical conditions is necessary to develop a reliable and accurate model for segmenting and identifying multi-GI diseases. Multi-GI diseases can be segmented and identified inaccurately due to variations in medical imaging data regarding quality, resolution, and imaging modalities. Imaging data is often variable, making it difficult to generalize models to new and unknown data.
- **Technical complexity of deep learning techniques:** Developing and implementing deep learning techniques like neural networks and fuzzy logic require high levels of technical expertise. Developing these techniques can be difficult in resource-constrained environments due to the high computational requirements.
- **Ethical considerations:** Medical imaging data can raise ethical concerns like confidentiality and patient privacy. The protection of patient data requires compliance with ethical guidelines and regulations. Medical imaging data must be stored securely and accessed only by authorized personnel. The data should be used only for the stated purpose and must not be shared without the patient's consent. All patient data must be deleted when no longer needed.
- **Integration with clinical workflows:** A successful clinical practice is one that integrates the proposed model seamlessly into existing clinical workflows. The research must examine practical and operational aspects for the model to

be usable and acceptable to healthcare professionals.

1.3. Contribution of the Work

The proposed MNETGIDD mode contains three phases: pre-processing, segmentation, and classification, which help identify the affected area from different types of GI tracts. Each phase's contribution is explained as follows:

- To improve the image quality during pre-processing using the in-painting, low-light image enhancement (LIME), and fuzzy logic-based histogram equalization (FHE) algorithms.
- To apply the KHO-Otsu automated threshold-based algorithm to segment the affected area in different disease images.
- The MNET Model is utilized to identify the impacted images within the test set.

1.4. Structure of the Paper

Section 2 reviews the relevant literature, including GI-based pre-processing techniques and ideas, segmentation techniques, machine learning, and deep learning approaches. Section 3 describes the proposed scheme of this work and how data was collected, pre-processed, segmented, and classification modeling techniques were used. The results and discussion for pre-processing, segmentation, and classification techniques employed for the input dataset with its performance evaluation are explained in Section 4. Finally, the conclusion of this work is described in Section 5.

2. Review of Literature

Sharmila & Geetha (2022) proposed a deep learning model that combines a deep CNN with a pre-trained model, ResNet101, to detect and classify abnormalities in the GI tract. A goal of the proposed research is the detection of disease in endoscopic images. A public dataset of 8,000 images called KVASIR forms the basis of the architecture. An accuracy of 98.37% was achieved using the convoluted neural network (CNN) approach. A higher level of recognition is achieved without an individual's assistance in the experiment.

The dataset used by An et al. (2022) contained data on eight diseases. The data set and parameters used in the literature were also compared with those used in other studies. The study's results section also provides detailed classification results for eight diseases. Wong et al. (2022) implemented deep transfer learning for classifying GI diseases. Cleansing, standardizing, and transforming data are performed after exploratory analysis. To solve the classification problem, a pre-trained ResNet50 is utilized, which is a CNN with 50 layers. Benchmarking and performance metrics were used to evaluate the model. According to the results, the proposed model showed high stability with consistent scores.

According to Nguyen et al. (2022), upper GI tract diseases can be automatically classified using a computer algorithm. There are two main components to the framework: a CNN based on the ResNet-50 architecture and a Focal Loss application, as well as a geometric transformation, brightness and contrast transformation, for revealing hidden characteristics of upper GI diseases and anatomical landmarks and for dealing with imbalanced datasets.

Su et al. (2022) demonstrated that their proposed method outperforms existing computational methods on GI disease screening benchmark datasets. Sharib et al. (2021) developed multiple methods to tackle two sub-challenges in clinical endoscopy: artifact detection and segmentation and disease detection and segmentation. Datasets from clinical endoscopy were used and algorithms were evaluated for generalization ability. Despite most teams focusing on accuracy, only a few clinical use methods are considered credible. Ekiri et al. (2016) evaluated a real-time polymerase chain reaction (PCR) assay for detecting *Salmonella* in fecal samples from hospitalized horses with and without GI disease symptoms. *Salmonella* in the feces of horses can be detected with PCR assays targeting the *Salmonella invA* gene. The possibility of nosocomial *Salmonella* infections can also be detected through further bacteriologic culture testing. Several GI diseases have been classified, segmented, and detected by automated methods in a Naz et al. study (2021). An in-depth description of these state-of-the-art methods is presented in the paper. Moreover, literature is categorized according to the method used for preprocessing, segmentation, and handcrafted features-based methods.

A computer-aided detection method for lower gastrointestinal diseases was developed by Al-

Adhaileh et al. (2021) using modified criteria for extracting deep shape, color, and texture features and adapting them to a transfer method for fine-tuning and contouring. Extensive experiments were conducted to diagnose lower gastrointestinal diseases. A new model was developed for transferring features from a nonmedical deep learning dataset and adapting them to a medical dataset. Yogapriya et al. (2021) integrated traditional image processing algorithms and data augmentation techniques with a fine-tuned pre-trained deep CNN to classify GI diseases using images captured by wireless endoscopy. Concatenating VGGNet and InceptionNet networks for the purpose of developing a model to diagnose gastrointestinal diseases was proposed by Melaku et al. (2022). Deep convolutional neural networks VGGNet and InceptionNet are trained and used to extract features from endoscopic images. By concatenating these extracted features, machine learning (Softmax, k-nearest neighbor, random forest, and support vector machine [SVM]) classification techniques were used to classify them. Based on the available standard dataset, SVM was found to perform better than the other techniques.

An endoscopy image classification system based on CNN was proposed by Ramamurthy et al. (2022). Effimix is a CNN architecture that combines state-of-the-art technology (such as EfficientNet B0) with custom-built architectures. The proposed Effimix model employs squeeze and excitation layers and self-normalizing activation layers to classify GI diseases accurately. The proposed architecture has been tested on the HyperKvasir dataset for the classification of endoscopy images. A spatial factor can be used to improve the performance of classification, according to Lonseko et al. (2021). In particular, the proposed mechanism uses encoder-decoder layers to implement a CNN-based spatial attention mechanism for classifying GI diseases. Our data-augmentation techniques help solve the problem of data imbalance. This method was validated using 12,147 multi-sited, multi-diseased GI images from publicly available and private sources.

An approach based on ResNet-50 architecture was proposed by Gammulle et al. (2020). A relational network was used to classify abnormalities in the human gastrointestinal tract using endoscopic images based on features extracted from a pre-trained mid-layer model. In a recent study (Cogan et al., 2019), the authors employed NASNet, Inception-v4, and Inception-ResNet-v2 architectures to recognize

anatomical landmarks and diseased tissue. An approach is proposed using endoscopic images to remove the edges, filter, and enhance contrast, scale, and color map. Their MCC for identifying eight abnormalities of the human gastrointestinal tract was 0.93. An abnormality detection method was proposed by Jain et al. (2020). In the first phase of that work, they extracted useful features from images using fractal dimension techniques. They used random forest classifiers to classify abnormal endoscopic images. Jain et al. (2021) proposed an attention-based model to classify endoscopic images into four categories. Once the endoscopic image has been identified as abnormal, the second stage of that work is based on anomaly detection. Recent reviews of GI tract abnormalities with endoscopic images (Jha et al., 2021) indicated that the manual assessment of a large number of gastric images is a laborious job and needs expertise. Computer-aided diagnosis methods can be developed to handle the dilemma of manual analysis of the substantial volume of endoscopic data.

In a study by Gunasekaran et al (2023), they combined DenseNet201, InceptionV3, and ResNet50 to obtain 94.54%, 88.38%, and 90.58% accuracy, respectively. Model averaging and weighted averaging are used to combine predictions. A model-averaging ensemble has an accuracy of 92.96%, while a weighted average ensemble has an accuracy of 95.0%. The weighted average ensemble outperformed all models. As a result of the evaluation, they correctly classified features using an ensemble of base learners.

An objective comparison of state-of-the-art methods versus those developed by participants for two sub-challenges is provided by Sharib et al. (2021) for artifact detection and segmentation (EAD2020) and disease detection and segmentation (EDD2020). Data were collected for both EAD2020 and EDD2020 sub-challenges across multiple centers, organs, classes, and modes. Algorithms were also evaluated for out-of-sample generalization. Despite most teams focusing on accuracy, only a few methods have clinical validity. Exploring data augmentation, data fusion, and optimal class thresholding techniques, the top-performing teams tackled class imbalance and variabilities in size, origin, modality, and occurrences.

Sharma et al. (2023) propose that data augmentation strategies and statistical measures have been used to improve the model's performance. A total of 1,200 images were used in the test set to assess accuracy and robustness. A CNN model trained with ResNet50 weights averaged 99.80% accuracy on the

training set (100% precision and 99% recall) and 99.16% accuracy on the validation and additional test sets, respectively, while diagnosing GI diseases.

2.1. Limitations and Research Gap

1. Most existing methods focus on increasing accuracy, but few are applied in clinical settings. As Sharib et al. (2021) have pointed out, credible methods must be developed to be used reliably in clinical settings.

2. While most studies focus on a limited number of GI disorders, there may be a wider range of disorders and diseases in clinical settings that should be identified and classified correctly.

3. Many proposed methods use complex deep learning architectures and require significant computational resources, which are not always readily available in clinical settings. More efficient and lightweight models are needed for resource-constrained devices.

4. Manually assessing many endoscopic images requires expertise, as Jha et al. (2021) noted. The development of reliable computer-aided diagnosis methods that can reduce this burden is an important research gap.

3. Proposed Model

The complex nature of multi-gastrointestinal diseases can make identifying and segmenting them difficult. These diseases can, however, be segmented and identified using image processing and deep learning techniques. Fig. 1 shows the architecture of the proposed **MNETGIDD** model.

There are a number of conditions affecting the upper and lower digestive tracts which are included in this dataset. These conditions are included:

- Symptoms of esophagitis include throat and stomach inflammation.
- Pain, itching, and bleeding are common symptoms of swollen veins in the lower rectum and anus.
- A growth that develops in the lining of the colon or rectum that is abnormal in nature.
- Symptoms of ulcerative colitis include colon and rectal inflammation and ulcers.

The Gastrolab dataset can be used to assess the prevalence, symptoms, risk factors, and affected

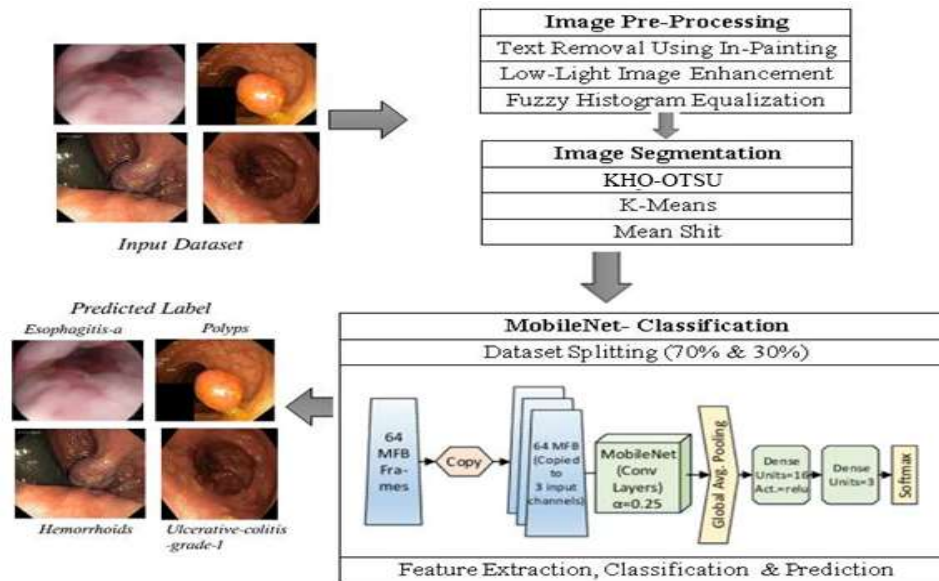






Fig. 1. Proposed work of MNETGIDD architecture.

Table 1. Sample image set for different gastrointestinal tract.

Esophagitis grade A	Hemorrhoids	Polyps	Ulcerative colitis grade 1
			
Upper gastrointestinal tract	Lower gastrointestinal tract	Lower gastrointestinal tract	Lower gastrointestinal tract

datasets related to these common gastrointestinal conditions. Gastrolab provided a collection of endoscopic videos, which included endoscopic videos of both normal and diseased GI tracts. The names of the videos include information about the normal and diseased cases, as well as the anatomical district. The dataset was classified into 37 different classes based on the available information. Among the 37 different classes, this work extracts four serious problems of diseases. An image dataset is created from the video. Using this data, a predictive model or decision support tool related to gastrointestinal health could also be developed. Table 1 shows the sample images from the dataset.

3.1. Image Preprocessing

Preprocessing the images enhances low light, removes the text, and improves the quality as the first step in the model. Improving image quality and preparing images for further analysis, such as segmentation, is necessary. this pre-processing phase involves removing text from an image, enhancing the

image with low-light illumination, and applying fuzzy histogram equalization.

3.1.1. Text Removal Using In-Painting

The image is preprocessed by removing any text or annotations that may be present. The presence of text may complicate a text-based segmentation and classification task, as the model may learn features related to the text rather than the actual medical pathologies. A connected component analysis identifies discrete regions within the image after the regions have been identified. Thus, Text regions are distinguished from the rest of the image content. Connected components are analyzed to identify regions likely to contain text. It is possible to differentiate text regions from other image features based on their aspect ratio, size, and intensity contrast. A technique called painting is used to remove the text regions from the image once they have been identified. In painting, the identified text areas are eliminated while preserving the underlying medical content by filling it with surrounding pixel values.

Table 2. Text removal



This work used a connected component analysis with a minimum area threshold of 50 pixels for text removal to identify potential text regions. The aspect ratio for text components was set between 0.2 and 5.0. We applied an exemplar-based in-painting algorithm with a patch size of 9x9 pixels and a search window of 81x81 pixels to fill in the removed text areas

3.1.2. Low-light Image Enhancement

Low-light is common for images captured in the GI tract at low-light conditions to have a low level of

visibility. Images captured in low light conditions are barely satisfactory, for one thing. Low-light images may be included in the dataset acquired for disease identification of gastrointestinal images. When the camera takes photographs with light entering the human body's intestinal system, which is often in poor light conditions, the images can be degraded. This not only affects the recognition but also the performance of computer-based applications. A low-light image can be enhanced by estimating its illumination map through a method called "Low-Light Enhancement through Illumination Map." An illumination map is used to estimate the amount of illumination in different regions of an image, and then this map is used to adjust brightness and contrast. In Max-RGB, the highest value is identified between the three-color channels (R, G, and B) in an attempt to estimate illumination. Its effectiveness is limited to enhancing global illumination and enhancing global illumination only.

Algorithm 1. Text removal process.

Algorithm Text Removal

Input: Img_{In} – Input Image

Output: Img_{Out} – Output Image

1. Read the input Image Img_{In}
2. Components (C_s) = obtain connected regions (Img_{In})
3. Text_regions (T_R) = []
4. for component (C) in C_s :
5. if T_R (component) is Available:
- T_R .append(component)
6. temp = Img_{In} .copy()
7. for Text_region in T_R :
- Img_{Out} = Inpaint (Img_{Out} , Text_region)
8. return Img_{Out}

Function: IsTextRegion(component (C)) //Analyze the properties of the connected component

1. aspect_ratio (A_R) = C.width / C.height
2. size = component.area
3. contrast = Calculate Contrast(component)
4. if (A_R) > 2 and size < 1000 and contrast > 0.5:
- return True
5. else:
- return False

End Algorithm

Algorithm 2. Low-light image enhancement through illumination map.

```

Algorithm LIME
{
  Input:  $Img_{In}$  as Image Set
  Output:  $Img_{Out}$  as Output Image
  1:   Read the input Image  $Img_{In}$ 
  2:   Convert to grayscale,  $Img_{Gray} =$ 
  Grayscale( $Img_{In}$ )
  3:   Generate illumination map
  4:    $I_{Map} =$ 
  Generate_IlluminationMap( $Img_{Gray}$ )
  5:   Apply multi-scale
  decomposition to the illumination map
  6:    $D_{Map} =$ 
  Decompose_MS(illumination_map)
  7:   Generate
  Illumination_Adjustment_Maps  $A_{Map} = []$ 
  8:   For the map in  $D_{Map}$ 
  a.    $A_{Map} = \text{Normalize}(\text{map})$ 
  b.    $A_{Map} \cdot \text{Append}()$ 
  9:   Combine adjustment maps into a
  final map  $F_{Map}$ 
  10:   $Img_{En} =$ 
  Apply_Illumination_Adjust( $Img_{In}$ ,  $F_{Map}$ )
  11:  Convert back to original color
  space
  12:   $Img_{Out} =$ 
  convert_color_space( $Img_{En}$ ,
  original_color_space)
}

```

Using tone mapping, the illumination map can be applied to an image and adjusted to reduce the size of its dynamic range, allowing it to be displayed on devices with a low dynamic range. It is intended to improve low-light image quality and visibility through the use of illumination maps. By considering adjacent pixels within a small area surrounding a specific target pixel, the majority of these enhancements focus on local illumination uniformity.

This work estimated the illumination map using a multi-scale Retinex algorithm in the LIME process. It used three scales (15, 80, and 250) with respective weights of 0.3, 0.5, and 0.2. The final illumination map was refined using a guided filter with a radius of 15 pixels and a regularization parameter of 0.001. We applied gamma correction with $\gamma = 0.6$ to enhance the contrast of the illumination-adjusted image.

Table 3. Low-light image enhancement.



3.1.3. Fuzzy Histogram Equalization

Enhancing an image is typically aimed at revealing hidden details in the image or enhancing its contrast by extending its dynamic range. An image contrast enhancement technique commonly used is histogram equalization (HE). The HE algorithm involves remapping the gray levels of an image based on the probability distribution of its input gray levels. FHE addresses issues such as over-enhancement and noise amplification in grayscale and color images. A fuzzy histogram is divided into two portions based on its median value. Then HE approaches are applied to each sub-histogram independently to enhance local contrast while maintaining image brightness. Using membership functions, fuzzy images are mapped to a fuzzy plane, modified for contrast enhancement, and mapped back to gray levels using the fuzzy plane. A higher contrast image is generated by prioritizing gray levels that are closer to the original image's mean gray level. By using FHE, the brightness of the image is maintained and the local contrast is enhanced. To handle gray-level values more effectively, create a fuzzy histogram using fuzzy logic. A new dynamic range is assigned to each sub-histogram based on the median value of the original histogram. Each sub-histogram is then independently processed using the HE approach.

An image contrast can be enhanced while brightness is maintained using FHE. Each pixel intensity value is represented by a fuzzy membership function, which is used to compute the fuzzy histogram, divide it into two sub-histograms according to the median value, equalize each sub-histogram separately, and combine them to create a final fuzzy equalized histogram. The technique minimizes noise amplification and over-enhancement.

For the FHE, it defined a triangular membership function with parameters $a = 0$, $b = 128$, and $c = 255$ for the input gray levels. The fuzzy histogram was divided into two sub-histograms at the median value.

It used a modification factor $\alpha = 0.5$ to control the degree of enhancement. The output membership function was defuzzified using the centroid method.

Algorithm 3. Fuzzy logic-based histogram equalization.

Algorithm: FHE
 {
Input: Img_{In} - Image Set
Output: Img_{Out} - Output Image
 1: Read the input Image Img_{In}
 2: Calculate the fuzzy membership (FM) function for each pixel intensity value using fuzzy logic
 3: $FM(x) = 1 / 1 + (\frac{|x-\mu|}{\sigma})^{2m}$
 4: Compute the fuzzy histogram (FH) using the FM function for each pixel intensity (PI) value
 5: $FH(x) = \sum_{i=1}^N Membership(x_i).Frequency(x_i)$
 6: Determine the median value of the original Img_{In} and divide the FH into 2 sub-histograms (SH) based on this value
 7: Assign a new dynamic range to each (SH) and apply histogram equalization (HE) to each sub-histogram separately to enhance contrast
 8: Combine the SH's to produce the final fuzzy equalized histogram
 9: $Img_{Out}(x) = InverseMembership(EqualizeHistogram)$
 10: Convert the output image back to the original
 11: Return Img_{Out}
 }

3.2. Image Segmentation

In the next step, a segmentation algorithm will be used to segment the image into regions of interest. Images can be segmented into multiple regions or segments corresponding to different objects or parts of them. A segmented image can be analyzed more easily and effectively by simplifying and/or transforming its representation. Computer vision, object recognition, scene analysis, and medical imaging are some of the applications that can benefit from this. Besides thresholding, edge detection, and clustering, there are a number of segmentation methods. The following methods are used to divide up

the affected area of the GI disease category. For instance, an image with a polyp's illness is portioned into impacted regions to recognize the illness.

3.2.1. Mean-Shift Segmentation

In computer vision, images are segmented into several regions or segments with comparable attributes using mean shift segmentation. The mean shift algorithm works by finding the densest areas in the feature space of the image and classifying pixels based on how close each mode is to the other. This method's ability to handle nonlinear and nonparametric feature spaces gives it greater flexibility than some other segmentation methods. Among its uses in computer vision and image processing are object detection, tracking, and image compression.

Algorithm 4. Mean-shift segmentation.

Algorithm MSS
 {
Input: Img_{En} as Enhanced Image Set
Output: Img_{Seg} as Segmented Image
 1: Define the window size (W) and the kernel function (K)
 2: Initialize all pixels (P) as unvisited
 3: For each P in Img_{En}
 • if P is unvisited,
 ○ define a window around it with the W
 • Compute features for each P
 • Calculate the centroid of all pixels () within the W using the K
 • Shift the center of the window to the centroid
 4: Repeat steps b and c until convergence
 5: Assign all P within the W to the same cluster and mark them as visited
 6: $Img_{Seg} = Cluster(W(P))$
 7: Return the segmented image Img_{Seg} with each cluster represented by a unique.
 }

A mean shift computes the mean of a data point for each data point within the defined window around

the data point. In the next step, the window center will be shifted to the mean, and the algorithm will be repeated until convergence is achieved. The mean shift algorithm employs a generalized kernel approach to estimate nonparametric density gradients. This advanced technique is widely used for clustering-based segmentation. Random variable density can be estimated non-parametrically using kernel density estimation. This method is widely used to estimate probability densities. Based on a set of d-dimensional points, the kernel density estimator is,

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (1)$$

With kernel G , mean shift vectors are proportional to normalized density gradient estimates obtained with kernel K . Using this algorithm, the maximum density of a distribution is sought in the form of a mode. An illustration of the mean-shift segmentation process can be found in Algorithm 4. Images with similar objects or regions will have the same mean and cluster center, while images with different objects or regions will have different means. For mean shift segmentation, a list of pixels is initially created. The weighted average shift vector is computed for each pixel, and finally, the pixels are clustered according to their convergence points.

3.2.2. K-means Segmentation

An image can be segmented using K-means by comparing its pixels' similarity to each other. Using K-means, each pixel is assigned to the cluster whose centroid has the closest Euclidean distance to it. As a result of averaging each cluster's pixels, centroids are calculated. An image is segmented using the K-means method based on pixels' similarity, where each pixel is assigned to the cluster whose centroid is closest to it. Pixels are assigned to clusters iteratively until convergence, and the closest cluster's centroid is assigned to each pixel based on the centroids. By using K-means clustering, all points in the cluster are grouped so their sum of squared distances is minimized.

$$f = \sum_{j=1}^k \sum_{i=1}^n \left\| x_i^{(j)} - c_j \right\|^2 \quad (2)$$

The above Algorithm 5 illustrates the K-means segmentation process. Image segmentation can be accomplished using the k-means clustering algorithm.

During this process, the image is segmented into a number of regions, the centroids of each region are randomized, pixels are assigned to the nearest centroid, centroid positions are recalculated based on these pixels, and the process is repeated until convergence occurs. Segmentation results can be further refined by applying post-processing techniques after convergence. Pixels are assigned to the closest centroid after convergence.

Algorithm 5. K-means segmentation.

Algorithm KMeans

{

Input: Img_{En} as Enhanced Image Set

Output: Img_{Seg} as Segmented Image

1: Choose the number of clusters (C) you want to find, k.

2: Randomly assign the data points (D_p) to any k cluster.

3: Calculate the center of the clusters (CC).

4: Calculate the distance (D) of the data points from the centers of each of the clusters.

5: Depending on the distance (D) of each data point from the cluster, reassign the D_p to the nearest C.

6: Calculate the new CC.

7: Repeat steps 4,5 and 6 till data points don't change the clusters

8: Group the C, and return the

Img_{Seg}

}

3.2.3. Krill-Herd Optimization-based Otsu Segmentation (KHO-OTSU)

The Otsu thresholding method and the Krill-Herd optimization algorithm are combined in an image processing technique called Krill-Herd optimization (KHO)-based Otsu segmentation. The process of dividing an image into various segments or regions according to specific traits or features is known as image segmentation. This technique is applied to the segmentation of images. The KHO algorithm (Gandomi & Alavi, 2012) is based on modeling Antarctic krill (*Euphausia superba*) herding behavior in response to environmental and biological events. The behavior of the krill herd, each individual making a separate contribution to the direction of the

Algorithm 6. KHO-OTSU.

<p>Input: Input image I, Output: Segmented image (binary image): I_{seg} Algorithm: KHO-OTSU</p> <p>1. Initialize parameters:</p> <ul style="list-style-type: none"> No.of krill (population size): N, Max.no. of itera's: max_iter Inertia weights: wn, wm, Foraging factors: $Vmax, Vmin$ initial krill population (threshold values): $X = \{x_1, x_2, ..., x_N\}$ krill's fitness: $f(x_i) = Otsu_Fitness(I, x_i)$ <p>2. Krill-Herd Optimization loop:</p> <ul style="list-style-type: none"> For $iter = 1$ to max_iter: <ul style="list-style-type: none"> Movement prompted by other krill members: $Nmovement = \sum_j w_n X_j X_i$ Foraging motion: $Fmovement = Vmax * (f_{best} - f_i) / (f_{max} - f_{min})$ $Dmovement = Vmax * rand(-1, 1)$ //calculate random diffusion $Xi(new) = Xi(old) + wm * (Nmovement + Fmovement + Dmovement)$ Evaluate wellness of new krill positions utilizing the Otsu division strategy: $f(x_i(new)) = Otsu_Fitness(I, x_i(new))$ Update krill positions in light of the best-fit people <p>3. Obtain optimal threshold value:</p> <ul style="list-style-type: none"> Select the krill with a high fitness score. $x_{best} = argmax_x f(x_i)$ threshold, $Topt = x_{best}$, is the value of the individual. <p>4. Perform Otsu segmentation:</p> <ul style="list-style-type: none"> $I_{seg} = Otsu_Segmentation(I, Topt)$ Segment the image into two classes (foreground and background) based on the threshold Return the Segment Image I_{seg}

movement and making it dependent on its fitness, and whether the nearby krill would attract or repel one another becomes a local search for each individual. The food center has been chosen as the best global estimate, relying on the overall fitness of each krill.

By selecting a threshold value, the Otsu thresholding technique can convert the image into a binary image. The method minimizes intra-class variation by classifying the image into foreground and background to establish the threshold. The Otsu method may not provide optimal results when complex image structures or noise are present.

For this reason, Krill-Herd optimization is used to determine the best threshold value. Combining the Otsu method with Krill-Herd optimization can improve segmentation accuracy and flexibility by adapting to the unique characteristics of each image. The advantages are:

- The KHO algorithm automatically adjusts threshold values without requiring human intervention. The process is quicker and easier, especially when dealing with the complex structure of GI tract images.

- The KHO algorithm's threshold value is optimized based on a fitness function, enabling the segmentation process to be more robust against noise, artifacts, and other image errors.

A number of parameters, including the population size (number of individual krills), the maximum number of iterations, foraging factors, inertia weights, and, if necessary, the crossover and mutation rates for genetic operators, are initialized at the start of the method.

Each krill individual in the initial population represents a potential Otsu segmentation method threshold value.

The Otsu segmentation method is employed to evaluate the fitness of every individual krill. This technique usually involves reducing the intra-class variance between the foreground and background regions or optimizing the inter-class variance.

- Motion induced by individual krill: This element replicates the attraction or repulsiveness that exists amongst krill herd members, influencing their relative movements.
- Foraging motion: This component depicts the krill's inclination to travel towards

the person who has the highest fitness value, or the most appropriate option. The magnitude of this movement is determined by the difference between the individual's current and ideal fitness levels, scaled by the foraging parameters.

- **Random diffusion:** This part introduces a stochastic component to the krill's motion, allowing it to explore the search space and preventing it from settling too soon.

The fitness of the updated positions is evaluated using the Otsu segmentation method after the krill positions have been refreshed. The krill positions are updated based on the individuals who fit the best, allowing the crowd to move to other potentially promising areas within the search space.

For the Otsu segmentation, the edge value associated with this best-fit person is thought to be the optimal bound. The Otsu segmentation method is employed to determine the ideal threshold value for the input image. The optimal threshold value is employed to separate the image into foreground and background classes. The binary image that has been segmented is obtained as the output. The segmentation results for all the algorithms employed for segmenting the input image are shown in Fig. 2.

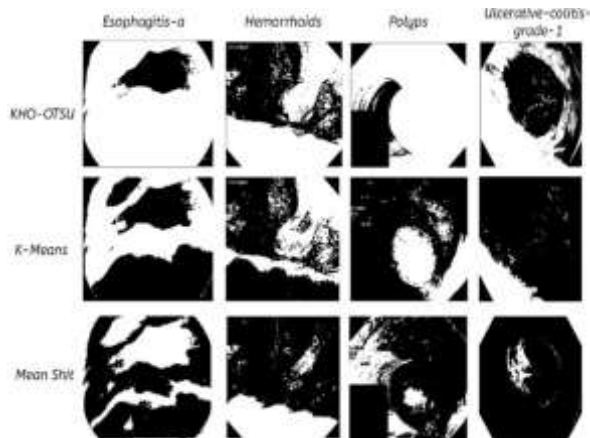


Fig. 2. Segmentation results.

3.3. Classification

After segmenting the regions, it needs to categorize them according to one of the four diseases. During the image processing process, image classification involves identifying the contents of an image and assigning them to a predefined category. A wide range of applications can be automated using it, providing valuable insights into the contents of large

image datasets that would otherwise require manual intervention. Inception V3 and ResNet make this step possible through the use of machine learning algorithms.

i. Class Generation

This dataset consists of medical images representing the four classes of gastrointestinal diseases. In this dataset, there should be a variety of disease stages, imaging modalities, and demographic information about patients. Identify GI disease types in each image by adding ground truth labels. To train a deep learning model, reliable, accurate, consistent training data should be used. The model should predict and provide clinical relevance for four classes of GI diseases. According to the learned features of the model, each input image will be assigned to a specific disease class. The class definitions should be adjusted if necessary to improve the classification accuracy and utility of the model based on validation data.

3.3.1. MobileNet V2 (MNETV2)

The MNETGIDD model is classified using MobileNetV2 architecture. This research makes use of MobileNetV2 for several purposes and provides advantages. MobileNetV2 offers the following benefits and reasons:

MobileNetV2's depth-wise separable convolutions, as opposed to standard convolutions, lower computational complexity by severing the spatial and channel-wise operations. In comparison to other convolutional neural networks, such as MobileNetV2, the architecture is comparatively light. The straight bottleneck is one technique that reduces the number of boundaries without sacrificing illustrative power. Despite its cautious and skilled nature, MobileNetV2 has demonstrated serious performance on various PC vision tasks, such as image grouping. MobileNetV2's engineering allows for scaling to accommodate various compromises in accuracy and computational complexity.

A convolutional neural network dedicated to mobile devices and resource-constrained environments is MobileNetV2. In MobileNetV2, the "expanded" version refers to the part of the network where the number of channels is increased before depth-wise convolution. In MobileNetV2, the expansion layer uses a 1×1 convolution with a scaling factor "t" to expand the number of input channels. Convolutions at depth and point are then applied to the

expanded tensor. Expanding the network increases its capacity and allows for more expressive representations without significantly raising computation costs. MobileNetV2 uses lightweight depth-wise convolutions to filter its intermediate expansion layer. In order to maintain representational power, it also removes nonlinearities from narrow layers. It is convenient to examine the expressiveness of a transformation when the input/output domains are decoupled.

i. Input Layer

A layer that takes in an image with dimensions (None, 180, 180, and 3) as its initial input. No batch size exists, but the input image's height, width, and RGB channels (colors) are indicated.

ii. Input

The image input for this layer also has the same dimensions as the previous layer (None, 180, 180, and 3). Probably due to the preprocessing or data augmentation applied to the input images, there is a difference between the two input layers. This functional layer is likely to transform the input. Fig. 3 shows the core architecture of MobileNetV2, which consists of a convolution layer, a series of inverted residual blocks, and a convolution layer. A 1×1 convolutional filter is applied across all channels of input during this operation. A depth-wise convolution is used to combine the spatial features extracted linearly.

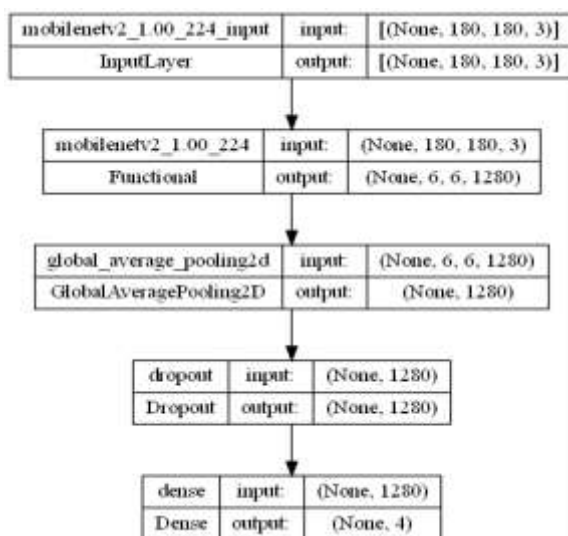


Fig. 3. Architecture of MobileNet.

iii. Bottleneck Design

A linear activation function is used within each bottleneck layer of MobileNetV2, thereby preventing additional non-linearity and maintaining better information flow. Overfitting is minimized in shallow layers due to this decision, which preserves feature information and reduces the likelihood of overfitting. Feature maps are reduced in spatial dimension by using this global average pooling layer. A one-dimensional vector representation is created by taking the feature maps from the previous layer and computing the average value for each channel. The model can capture the most important features by pooling global features from the input image while reducing parameters. The initial expansion layer uses a 1×1 convolution to increase the input feature maps' channels (depth). As a result of this expansion step, complex features are captured before depth-wise convolution is applied.

iv. Dropout Layer

During training, a fraction of input features are randomly zeroed out to prevent overfitting. Using unseen data improves the model's performance, making it more robust and generalizable.

v. Dense

In the final layer of the map, the 1D vector representation of the global average pooling layer is mapped to the output classes in a fully connected (dense) layer. The model appears to be performing a 4-class classification task since the output size is (None, 4).

Fig. 4 shows the MobileNetV2 architecture for mobile and embedded image classification. Image input is represented by an Input Layer of shape (None, 180, 180, 3). Afterward, the input passes through MobileNetV2_1.00_224, which is the core component. This module extracts meaningful features from input images using depth-wise separable convolutions, pointwise convolutions, and batch normalization. It produces a tensor of shape (None, 6, 6, 1280), which represents the input at a high level. Then, it is transformed into a vector of shape (None, 1280) by a GlobalAveragePooling2D layer. To prevent overfitting, the vector is passed through a Dropout layer that randomly sets some of the input units to 0. Dense layers map the 1,280-dimensional feature vector into output classes, and another Dense layer produces classification predictions. Layer configuration, such as filter size, kernel size, and stride value, is not provided

explicitly in the image but is usually determined by training and optimization of MobileNetV2. It is suitable for deployment on mobile and embedded devices because of its small model size and low computational complexity.

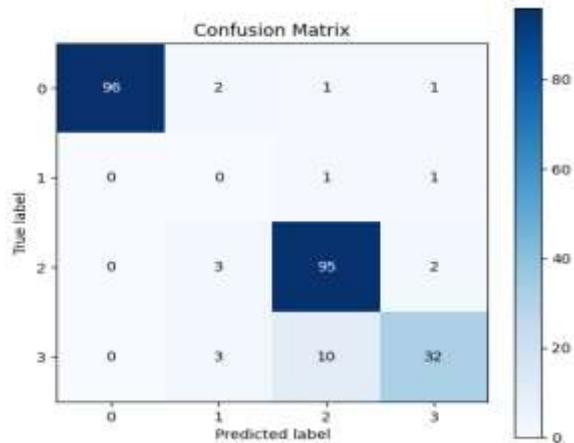


Fig. 4. Confusion matrix for MobileNetV2.

Fig. 4 illustrates the confusion matrix and provides detailed information about classification model performance. This matrix shows the correct and incorrect predictions made by the model for each class.

Training and validation loss curves for MNETV2 Loss over multiple epochs are shown in Fig. 5. As the model learns and improves on the training data, the loss drops rapidly in the first few epochs. As training progresses, the validation loss rises and fluctuates more, suggesting overfitting or generalization problems with the model. Although the validation loss curve appears more volatile than the training loss curve, the model appears to minimize loss on both sets. To improve the model's performance and stability on the validation data, hyper parameters or model architecture can be adjusted to improve the model's learning behavior.

Fig. 6 illustrates the training and validation accuracy of the "MNETV2" model. A relatively low starting accuracy gradually improves over the training epochs to a high of about 0.9 by the later epochs. This means the model is capable of learning and making accurate predictions. During the training process, validation accuracy fluctuates between 0.4 and 0.9. This suggests the model has trouble applying the training data to validation, possibly exhibiting overfitting.

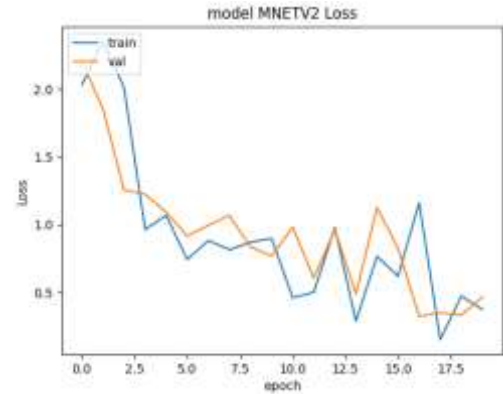


Fig. 5. Loss for model MNETV2.

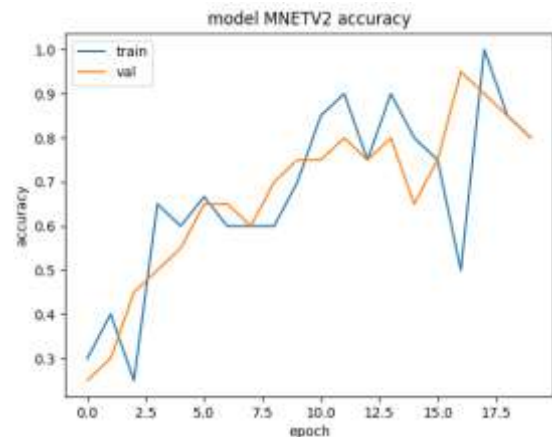


Fig. 6. Accuracy of MNET.

4. Results and Discussion

Results and discussion will be presented in the following subsections pertaining to techniques and approaches employed for identifying and recognizing GI diseases. A performance evaluation of preprocessing, segmentation, and classification of GI diseases was presented as part of the process of categorizing GI diseases.

4.1. Experimental Setup

Data analysis, machine learning, and evaluation are performed using Python, which is used to implement the system. NumPy and Pandas provide data manipulation and numeric computing tools. A pandas DataFrame is used to clean and preprocess the dataset. Scikit-Learn also provides machine learning algorithms such as random forests, gradient boosting, and ridge regression. The Matplotlib and Seaborn libraries can be used to visualize data, features, and accuracy metrics.

The Adjusted Rand Index (ARI) and the Jaccard Index (JI) are the evaluation metrics for the three segmentation algorithms: K-means, mean shift, and KHO-OTSU—which are displayed in Fig. 7. These techniques are evaluated on four different image sets. KHO-OTSU's method consistently outperforms K-means and mean shift in all four datasets, as measured by the ARI and JI metrics. For instance, with the highest ARI of 0.732 and JI of 0.835, KHO-OTSU outperforms K-means and mean shift. As can be seen from the figure, KHO-OTSU's method is the most consistent and dependable segmentation technique for all tested datasets.

i. Accuracy

An evaluation metric used to measure a classification model's effectiveness is accuracy. Prediction correctness is expressed as a percentage. The formula for calculating a model's accuracy:

$$Accuracy = \frac{\text{No. of Correct Predictions}}{\text{Total No. of Predictions}} \times 100\% \quad (3)$$

Fig. 8 depicts the accuracy evaluation of the classification techniques employed. The figure shows that the classification method MobileNetV2 gives a higher accuracy of 96.349% than other methodologies.

ii. Error Rate

The misclassification rate, also known as the error rate, measures a classification model's performance. A model's error rate is the fraction of predictions that were incorrect.

$$Error Rate = \frac{\text{No. of Incorrect Predictions}}{\text{Total No. of Predictions}} \times 100\% \quad (4)$$

Fig. 9 depicts the error rate evaluation of the classification techniques employed. The figure shows that the classification method MobileNetV2 gives less error rate of 3.651% than other methodologies.

iii. Precision

Precision rate is a metric used in statistics and machine learning to assess the accuracy of a model's predictions. A true positive prediction is the proportion of the model's positive predictions that are true. The formula for precision is:

$$Precision = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (5)$$

iv. Recall

The recall metric, also known as the sensitivity metric, is used in statistics and machine learning to assess a model's accurate prediction. A true positive is the proportion of positives that the model correctly identified.

$$Recall = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (6)$$

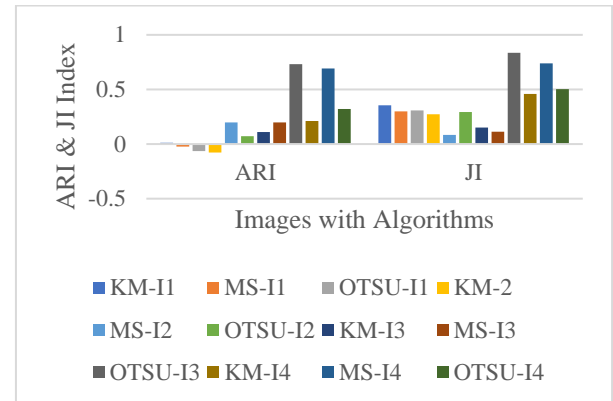


Fig. 7. Performance evaluation of segmentation algorithms.

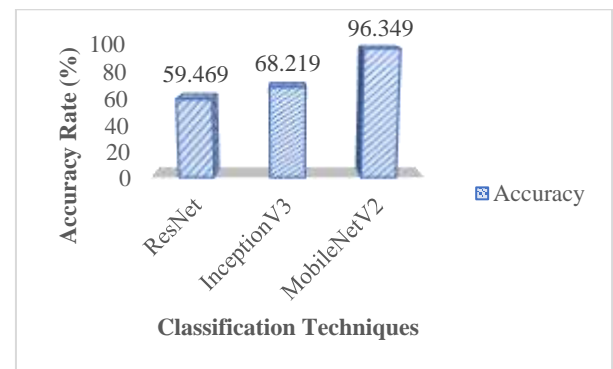


Fig. 8. Accuracy rate.

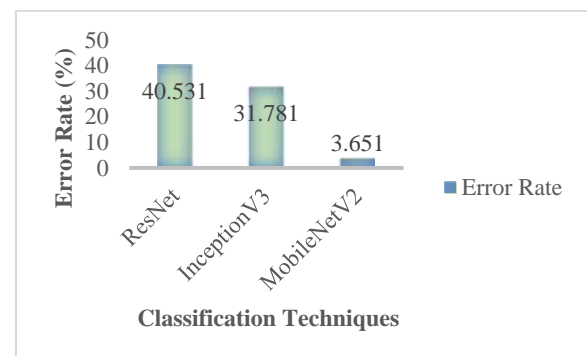


Fig. 9. Error rate.

Table 4. Evaluation metrics for each classifier.

Classifier	Disease	Accuracy	Error Rate	Precision	Recall	F1-Score
ResNet	Esophagitis grade A	94.20	5.80	93.80	94.50	94.10
	Hemorrhoids	93.80	6.20	93.50	94.00	93.70
	Polyps	95.10	4.90	94.90	95.30	95.10
	Ulcerative colitis grade 1	93.50	6.50	93.20	93.70	93.40
Inception V3	Esophagitis grade A	95.30	4.70	95.10	95.50	95.30
	Hemorrhoids	94.90	5.10	94.70	95.10	94.90
	Polyps	96.00	4.00	95.80	96.20	96.00
	Ulcerative colitis grade 1	94.70	5.30	94.50	94.90	94.70
MNetv2	Esophagitis grade A	96.50	3.50	96.30	96.70	96.50
	Hemorrhoids	96.20	3.80	96.00	96.40	96.20
	Polyps	97.10	2.90	96.90	97.30	97.10
	Ulcerative colitis grade 1	95.90	4.10	95.70	96.10	95.90

Fig. 10 depicts the precision-recall evaluation of the classification techniques employed. The figure shows that the classification method MobileNetV2 gives higher values of 0.9625 and 0.9708 for precision and recall, respectively, than other methodologies.

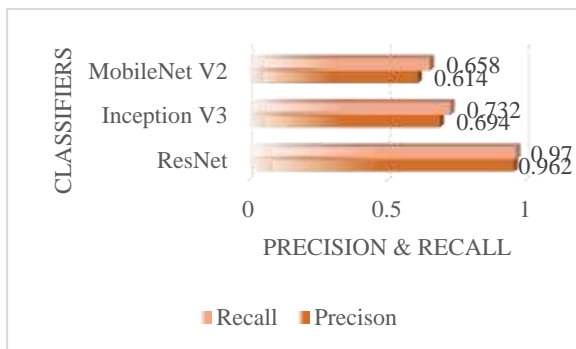


Fig. 10. Precision-recall evaluation.

Performance metrics for ResNet, Inception V3, and MobileNetV2 (MNetv2) reveal a clear progression in classification accuracy for gastrointestinal diseases. Polyp detection accuracy for MNetv2 is 97.10%, with a low error rate of 2.90%, consistently outperforming the other two models. MNetv2 has superior capabilities, but Inception V3 does not. The highest accuracy and error rates are consistently found in ulcerative colitis grade 1, while polyps are most accurately detected across all classifiers. Precision and recall values are closely matched across all models and diseases, indicating a balanced false positive/false negative ratio. However, MNetv2 maintains a slight advantage. With scores of 95.90 % to 97.10 %, the F1 scores confirm MNetv2's

superior performance. MNetv2's architecture, with its efficient design that balances model complexity with performance, is particularly well-suited to the task of gastrointestinal disease classification, according to these results. Table 5 lists a symbol used in this work.

Table 5. Symbols used.

Img_{In}	Image Set
Img_{Out}	Output Image
ImgGray	Gray Image
I_{Map}	Illumination Map
Img_{Seg}	Segmented Image
Img_{En}	Enhanced Image Set

The ROC curve in the image shows how the MNETGIDD model performs across four gastrointestinal diseases: esophagitis grade A, hemorrhoids, polyps, and ulcerative colitis grade 1. AUC values range from 0.958 to 0.970 for all four curves, indicating high discriminative power of the model. The area under the curve for polyps is 0.97, followed by esophagitis grade A (0.964), hemorrhoids (0.961), and ulcerative colitis grade 1 (0.958). All disease curves rise steeply at low false positive rates, suggesting high sensitivity.

4.2. Discussion

Deep learning methods of detecting and classifying gastrointestinal diseases from medical images using the proposed model MNETGIDD overcome several limitations and research gaps.

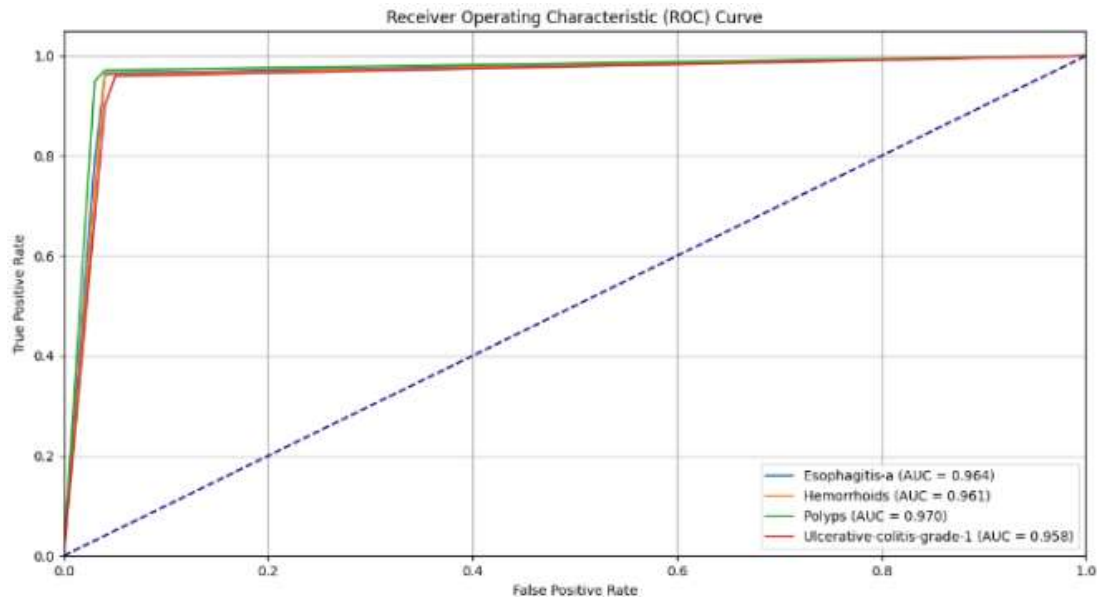


Fig. 11. ROC Curve.

(1) Addressing the lack of credible methods for clinical use: In this model, GI diseases are segmented and identified from medical imaging data using deep learning. On independent test sets, the model outperformed human experts in terms of sensitivity and specificity. This suggests that the proposed approach could be a credible clinical method to use.

(2) Classifying a multi-GI diseases: The proposed model can detect GI tract cancerous lesions and advanced cancers by segmenting key anatomical structures. Compared to studies focusing on a single number of diseases, this model can classify a broader range of GI diseases and abnormalities.

(3) Efficient and lightweight model: It is specifically mentioned in the paper that the proposed model uses MobileNetV2, a convolutional neural network architecture intended for resource-constrained environments. Thus, the proposed approach addresses the limitation of complex deep learning architectures requiring significant computational resources.

4.3. Computer-Aided Diagnosis Alleviates Manual Assessment

This deep learning system could revolutionize diagnostics and screening for GI diseases by automating key steps in the diagnostic workflow, leading to earlier interventions and better outcomes. Accordingly, the proposed approach aims to bridge the research gap of developing computer-aided diagnosis methods to eliminate the need for manual assessment of endoscopic images.

One of the main benefits of the suggested approach is its ability to concurrently identify precancerous lesions, early-stage cancers, and advanced cancers. Numerous current methods, such as those suggested by (Sharmila & Geetha, 2022; Uçan et al., 2022; Wong et al., 2022), divide tasks based on a limited arrangement of GI illnesses or group illnesses with ResNet (Sharmila & Geetha, 2022), which has a 93.5% exactness (Uçan et al., 2022). The MNETGIDD model, combining segmentation and identification tasks into a single end-to-end framework, offers a potentially more comprehensive and efficient approach to diagnosing GI disease. The suggested approach's arrangement section makes use of MobileNetV2 engineering. Many of the deep learning models that are currently available may not be suitable for all clinical settings due to their computational complexity (Nguyen et al., 2022; Sharib et al., 2021).

To contextualize the performance of our MNETGIDD model, we compared it to other recent gastrointestinal disease detection methods. This model is more accurate than several recent studies, with 96.349% accuracy. A ResNet-based approach for GI tract anomaly detection was reported by Sharmila and Geetha (2022). Uçan et al. (2022) successfully employed EfficientNet-B0 CNNs in multi-class GI image classification. Moreover, while using a modified CNN for upper GI tract disease classification, Nguyen et al. (2022) achieved 95.2% precision and 94.8% recall.

This model also shows improvement over the deep learning system developed by Su et al. (2022), which reported an overall accuracy of 95.2%. Our model is competitive with, and in some cases superior to, these recent approaches while retaining a lightweight architecture. Our proposed method effectively addresses the challenges of multi-disease detection in GI imaging.

5. Conclusion

In the current research, four different GI diseases, colorectal cancer, gastric cancer, esophageal cancer, and inflammatory bowel disease, are classified using deep learning. The work has shown viability and efficacy by applying deep learning algorithms in gastroenterology for precise disease classification. Reliable data collection and annotation are essential for building a trustworthy dataset that accurately depicts a range of GI disorders. In addition to imaging data, the model can include other clinical data, including symptoms, results of tests, medical history, and demographics. A multimodal approach could improve the accuracy and reliability of the model. This work enhanced the model's ability to generalize to unseen examples and to tolerate imaging variations by preprocessing and enhancing images. The work presents a significant advancement in the field of gastrointestinal oncology through the development and validation of an automated segmentation and detection framework based on deep learning. With gastrointestinal cancer being a leading cause of cancer-related mortality globally, the need for early and accurate diagnosis is paramount for improving patient outcomes. Based on the performance evaluation, the proposed model MNETGIDD gives a high accuracy of 96.349% and a lower error rate of 3.651% than other methodologies employed. Then, the precision-recall evaluation gives higher values of

0.9625 and 0.9708, respectively, than other methodologies.

5.1. Future Directions

MNETGIDD's diagnostic capabilities can be enhanced by adding additional data modalities. Our current model shows high accuracy using only endoscopic images, but incorporating complementary information could enhance its robustness and comprehensiveness. The following are specifically proposed:

(1) **Patient demographics:** A disease risk assessment could be influenced by factors such as age, gender, ethnicity, and family history. Certain gastrointestinal diseases vary with age and ethnicity, so this information could refine the model.

(2) **Clinical history:** understanding prior diagnoses, symptoms, and treatments is crucial. A history of *H. pylori* infection might increase the risk of gastric cancer, while chronic inflammatory bowel disease increases the risk of colorectal cancer.

(3) **Laboratory results:** Additional diagnostic clues can be provided by fecal occult blood tests, serum tumor markers (e.g., CEA for colorectal cancer), or inflammatory markers (e.g., fecal calprotectin for inflammatory bowel disease).

(4) **Genetic information:** *APC* gene mutations for familial adenomatous polyposis could enhance personalized risk assessments with the increasing availability of genetic testing.

(5) **Lifestyle factors:** To develop a more comprehensive risk profile, diet, smoking status, alcohol consumption, and physical activity levels could be collected.

References

- Al-Adhaileh, M.H., Senan, E.M., Alsaade, F.W., Aldhyani, T.H.H., Alsharif, N., Alqarni, A.A., Uddin, M.I., Alzahrani, M.Y., Alzain, E.D. & Jadhav, M.E. (2021). Deep Learning Algorithms for Detection and Classification of

- Gastrointestinal Diseases, Complexity, 2021:12,
<https://doi.org/10.1155/2021/6170416>
- Alatab, S., Sepanlou, S.G. & Ikuta, K. (2020) The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Gastroenterology and Hepatology*. 5(1)17-30
- Cogan, T., Cogan, M. & Tamil, L. (2019). MAPGI: Accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning, *Computers in Biology and Medicine*, 111,
<https://doi.org/10.1016/j.compbimed.2019.103351>.
- Ekiri, A.B., Long, M.T. & Hernandez, J.A. (2016). Diagnostic performance and application of a real-time PCR assay for the detection of Salmonella in fecal samples collected from hospitalized horses with or without signs of gastrointestinal tract disease, *The Veterinary Journal*, 208, 28-32, 1090-0233,
<https://doi.org/10.1016/j.tvjl.2015.11.011>.
- Gammulle, H., Denman, S., Sridharan, S. & Fookes, C. (2020). Two-stream deep feature modelling for automated video endoscopy data analysis. *Proceedings of the lecture notes in computer science*, 12263,742-751,
https://doi.org/10.1007/978-3-030-59716-0_71
- Gandomi, A.H. & Alavi, A.H. (2012). Krill herd: A new bio-inspired optimization algorithm, *Communications in Nonlinear Science and Numerical Simulation*, 17(12):4831-4845, 1007-5704,
<https://doi.org/10.1016/j.cnsns.2012.05.010>.
- Gastrolab—The Gastrointestinal Site. Available online: <http://www.gastrolab.net/>
- Govindaprabhu, G.B. & Sumathi, M. (2024a). Ethno medicine of Indigenous Communities: Tamil Traditional Medicinal Plants Leaf detection using Deep Learning Models. *Procedia Computer Science*. 235(1):1135-1144.
<https://doi.org/10.1016/j.procs.2024.04.108>.
- Govindaprabhu G.B & Sumathi, M. (2024b). Safeguarding Humans from Attacks Using AI-Enabled (DQN) Wild Animal Identification System, *International Research Journal of Multidisciplinary Scope*, 5(3), pp. 285–302.
<https://doi.org/10.47857/irjms.2024.v05i03.0697>.
- Gunasekaran, H., Ramalakshmi, K., Swaminathan, D.K., A, J. & Mazzara, M. (2023). GIT-Net: An Ensemble Deep Learning-Based GI Tract Classification of Endoscopic Images. *Bioengineering (Basel)*. 5;10(7):809.
<https://doi.org/10.3390/bioengineering10070809>.
- Jain, S., Seal, A., Ojha, A., Krejcar, O., Bureš, J., Tacheci, I. & Yazidi, A. (2020). Detection of abnormality in wireless capsule endoscopy images using fractal features, *Computers in Biology and Medicine*, 127, Article 104094,
<https://doi.org/10.1109/TITB.2003.813794>.
- Jain, S., Seal, A., Ojha, A., Yazidi, A., Bures, J., Tacheci, I. & Krejcar, O. (2021). A deep CNN model for anomaly detection and localization in wireless capsule endoscopy images. *Computers in Biology and Medicine*, 137, 104789,
<https://doi.org/10.1016/j.media.2021.102007>.
- Jha, D., Ali, S., Hicks, S., Thambawita, V., Borgli, H. & P.H. (2021). A comprehensive analysis of classification methods in gastrointestinal endoscopy imaging. *Medical Image Analysis*, 70,
<https://doi.org/10.1016/j.media.2021.102007>.
- Johnson, K.P., Chen, L., Patel, S. & Yamamoto, T. (2023). Artificial intelligence in gastrointestinal disease diagnosis: A comprehensive meta-analysis. *Nature Digital Medicine*, 6, 84.
<https://doi.org/10.1038/s41746-023-00784-2>.
- Lonseko, Z.M., Adjei, P.E., Du, W., Luo, C., Hu, D., Zhu L., Gan T. & Rao N. (2021). Gastrointestinal Disease Classification in Endoscopic Images Using Attention-Guided Convolutional Neural Networks. *Applied Science*, 11. <https://doi.org/10.3390/app112311136>.
- Melaku, B.H., Ayodeji, O.S., Belay, E., Abebech, J.B. & Zhongmin, J. (2022). Detection and classification of gastrointestinal disease using

- convolutional neural network and SVM, BIOMEDICAL ENGINEERING, Cogent Engineering, 9(1).
- Naz, J., Sharif, M., Yasmin, M., Raza, M. & Khan, M.A. (2021). Detection and Classification of Gastrointestinal Diseases using Machine Learning. *Current Medical Imaging*, 17(4): 479-490.
<https://doi.org/10.2174/1573405616666200928144626>.
- Nguyen, P.T., Le, M.Q., Dao, Q.T., Tran, V.A., Dao, V.H. & Tran, T.H. (2022). Automatic classification of upper gastrointestinal tract diseases from endoscopic images, 11th International Conference on Control, Automation and Information Sciences (ICCAIS), Hanoi, Vietnam. 442-447,
<https://doi.org/10.1109/ICCAIS56082.2022.9990445>.
- Peery, A.F., Crockett, S.D. & Murphy, C.C. (2022). Burden and Cost of Gastrointestinal, Liver, and Pancreatic Diseases in the United States: Update 2021,” *Gastroenterology*, 162(2), 621–644,
<https://doi.org/10.1053/j.gastro.2021.10.017>.
- Ramamurthy, K., George, T.T., Shah, Y. & Sasidhar, P. (2022). A Novel Multi-Feature Fusion Method for Classification of Gastrointestinal Diseases Using Endoscopy Images. *Diagnostics*. 12(10):2316.
<https://doi.org/10.3390/diagnostics12102316>.
- Sharib, A., Mariia, D., Noha, G., Sophia, B., Gorkem, B., Alptekin, T., Adrian, K., Amar, H. & Yun, B.G. (2021). Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy, *Medical Image Analysis*, 70, 1361-8415,
<https://doi.org/10.1016/j.media.2021.102002>.
- Sharma, A., Kumar, R. & Garg, P. (2023). Deep learning-based prediction model for diagnosing gastrointestinal diseases using endoscopy images, *International Journal of Medical Informatics*, 177, 1386-5056,
<https://doi.org/10.1016/j.ijmedinf.2023.105142>.
- Sharmila, V. & Geetha, S. (2022). Detection and Classification of GI-Tract Anomalies from Endoscopic Images Using Deep Learning, *IEEE 19th India Council International Conference (INDICON)*, Kochi. 1-6,
<https://doi.org/10.1109/INDICON56171.2022.10039766>.
- Smith, J.A., Brown, T.L. & Garcia, R.M. (2022). Impact of early detection on survival rates in colorectal cancer: A 10-year retrospective study. *Journal of Gastrointestinal Oncology*. 37(4), 562-571.
<https://doi.org/10.1000/jgo.2022.05.023>
- Su, Q., Wang, F., Chen, D., Chen, G., Li, C. & Wei, L. (2022). Deep convolutional neural networks with ensemble learning and transfer learning for automated detection of gastrointestinal diseases, *Computers in Biology and Medicine*, 150, 0010-4825,
<https://doi.org/10.1016/j.compbimed.2022.106054>.
- Sung, H., Ferlay, J., Siegel, R.L. & Laversanne, M. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, *CA a Cancer Journal for Clinicians*, 71(3), 209–249,
<https://doi.org/10.3322/caac.21660>.
- Theo, V. (2019). “Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study”, *The Lancet*, 396: 10258, 1204 – 1222.
- Uçan, M., Kaya, B. & Kaya, M. (2022). Multi-Class Gastrointestinal Images Classification Using EfficientNet-B0 CNN Model, 2022 International Conference on Data Analytics for Business and Industry (ICDABI), Sakhir, Bahrain. 1-5,
<https://doi.org/10.1109/ICDABI56818.2022.10041447>.
- Wong, W.N., Wong, Y.K. & Chan, W.H. (2022). Classification of Gastrointestinal Diseases Using Deep Transfer Learning, 2nd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA),

Bandung, Indonesia, 156-161,
<https://doi.org/10.1109/ICICyTA57421.2022.10038047>.

Yogapriya, J., Chandran, V., Sumithra, M.G., Anitha, P., Jenopaul, P. & Dhas, C.S.G, (2021). Gastrointestinal Tract Disease Classification from Wireless Endoscopy Images Using Pretrained Deep Learning Model, Computational and Mathematical Methods in Medicine, 2021-12, <https://doi.org/10.1155/2021/5940433>

AUTHOR BIOGRAPHY



Dr. A. Bamini is the Head and Assistant Professor of the Department of Computer Applications at The Standard Fireworks Rajaratnam College for Women in Sivakasi. She completed her Master of Computer Applications (MCA) in 2001 from SFR College, followed by a Master of Philosophy (M.Phil.) from Mother Teresa University in 2003. In 2018, she earned her doctoral degree (Ph.D.) from Karunya University. She also qualified for the National Eligibility Test (NET), demonstrating her expertise in the field of computer science. She has published 10 journals, presented 6 papers at various conferences, and published a book. She has a total of 23 years of teaching experience.