

Original Article

# Mathematical Modelling for Generalized Lower Gastrointestinal Image Based Classification and Detection on Both ML and DL Techniques

S. Vasudevan<sup>1</sup>, Vedyappan Govindan<sup>2</sup>, Haewon Byeon<sup>3</sup>

<sup>1,2</sup>Department of Mathematics, Hindustan Institute of Technology and Science, Tamilnadu, India.

<sup>3</sup>Department of AI Big Data, Inje University, Gimhae, Republic of Korea.

<sup>2</sup>Corresponding Authors : [govindoviya@gmail.com](mailto:govindoviya@gmail.com)

Received: 25 June 2024

Revised: 26 December 2024

Accepted: 30 December 2024

Published: 31 January 2025

**Abstract** - In this paper, the authors investigated an essential diagnostic and therapeutic technique for inspecting the colon, the distal portion of the small intestine, and the rectum in lower Gastrointestinal (GI) endoscopy. The condition of lower GI endoscopy today is thoroughly examined in this study, including its methods, uses, difficulties, and new developments. This research delves into the development of lower gastrointestinal endoscopy, emphasizing technological breakthroughs, improved procedural techniques, and its growing significance in medical practice. This review delves into the diagnostic potential of lower gastrointestinal endoscopy, highlighting its efficacy in identifying pathologies such as polyps, inflammatory bowel disorders, and colorectal malignancies. Discussion is held about the difficulties of lower GI endoscopy, such as patient pain, problems, and visual impairments. We examine ways to overcome these obstacles, including better sedation methods, better endoscope designs, and innovations in healthcare professional training. The study also discusses current technical advancements to improve lesion detection efficiency and accuracy, such as merging computer-aided detection techniques with artificial intelligence. Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) are investigated in this research for classification, lesion identification, and real-time picture processing during lower GI endoscopy. Furthermore, integrating sophisticated computer vision methods, such as feature extraction and picture segmentation, are examined to improve the visualization and comprehension of gastrointestinal diseases.

**Keywords** - Python, Machine Learning models, Deep Learning models, Gastrointestinal, Convolutional Neural Networks.

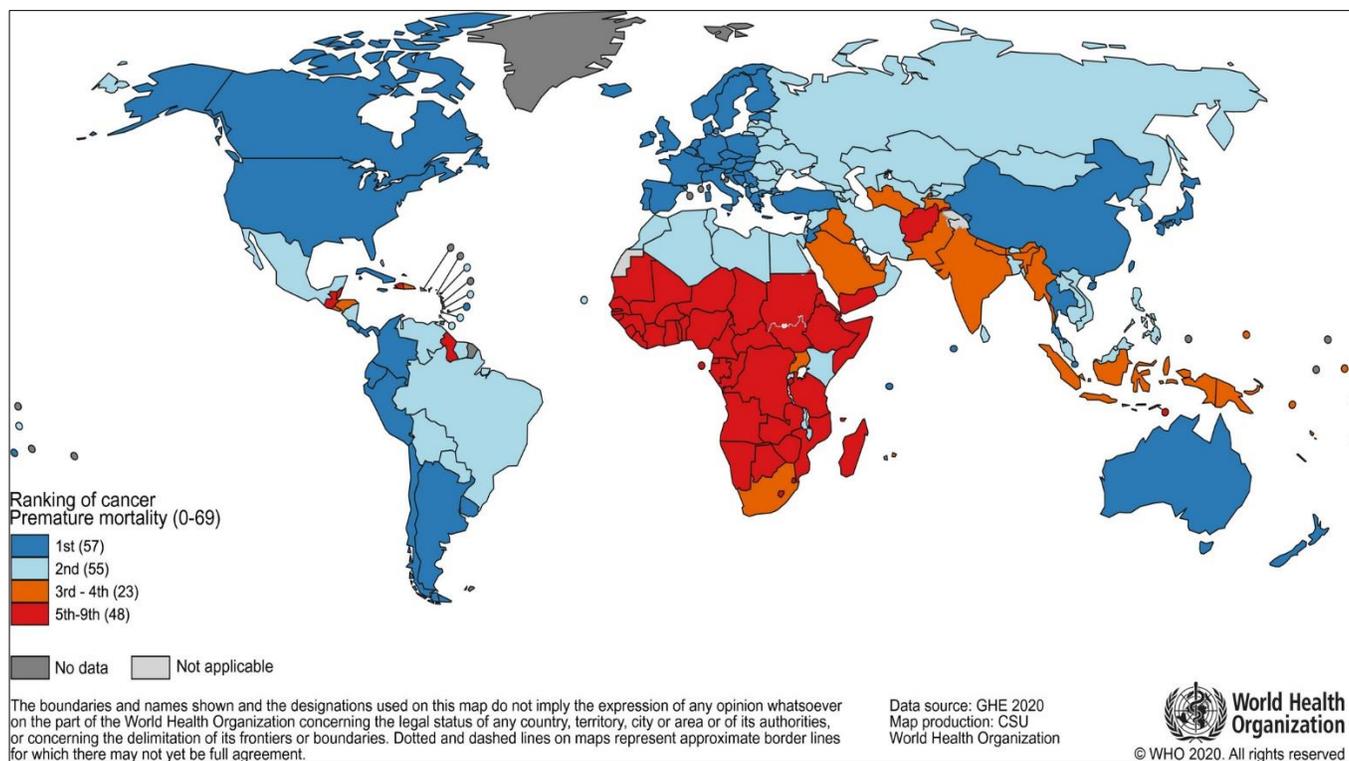
## 1. Introduction

The global cancer burden is estimated using GLOBOCAN 2020 cancer incidence and mortality statistics from the International Agency for Research on Cancer. In 2020, there are predicted to have been around 19.3 million new cases of cancer (excluding nonmelanoma skin cancer) and 9.9 million cancer deaths worldwide. Lung cancer is no longer the most common cancer diagnosed in women, with an estimated 2.3 million new cases (11.7%) of breast cancer emerging. Lung cancer is still the most prevalent cancer at 11.4%, followed by colorectal (10.0%), prostate (7.3%), stomach (5.6%), and lung (11.4%). With an anticipated 1.8 million fatalities (18%), lung cancer continued to be the most common cause of cancer-related mortality.

Colorectal (9.4%), liver (8.3%), stomach (7.7%), and female breast (6.9%) cancers were the next most common causes. The database was used to extract statistics on the number of new cancer cases and cancer deaths for 36 different cancer types and all cancers combined (ICD-10 codes C00–

C97). lip, oral cavity (C00-C06), salivary glands (C07-C08), oropharynx (C09-C10), Kaposi sarcoma (C46), female breast (C50), vulva (C51), vagina (C52), cervix uteri (C53), corpus uteri (C54), ovary (C56), penis (C60), prostate (C61), testis (C62), kidney (C64-C65, including renal pelvis), bladder (C67), brain, central nervous system (C70-C72), thyroid (C73), Hodgkin lymphoma (C81), non-Hodgkin lymphoma (C82-C86, C96), multiple myeloma (C88 and C90), nasopharynx (C11), hypopharynx (C12-C13), esophagus (C15), stomach (C16), colon (C18), rectum (C19-C20), anus (C21), liver (C22, including intrahepatic bile ducts), gallbladder (C23), pancreas (C25), larynx (C32), lung (C33-C34, including trachea and bronchus), melanoma of skin (C43), NMSC (C44, excluding basal cell carcinoma for incidence), mesothelioma (C45), including immunoproliferative diseases), and leukemia (C91-C95). In terms of the leading cause of mortality before the age of 70, cancer ranks first or second in 112 out of 183 countries and third or fourth in 23 additional countries, according to the World Health Organization (WHO) 2019 estimations [1].





**Fig. 1 2019 National cancer death index: a comprehensive analysis of deaths at ages <70. Included are the total number of nations in each ranking category**

The rising prevalence of cancer as a leading cause of death is partly due to the fact that, in many of the nations shown in Figure 1, the mortality rates from stroke and coronary heart disease have declined considerably in contrast to cancer. When it comes to lower Gastrointestinal (GI) cancers, endoscopy is essential, particularly colorectal cancer, because of its extensive diagnostic and therapeutic uses. Endoscopic treatments are particularly useful for the identification and intervention of colorectal cancer, a tumor that is commonly occurring worldwide [3]. During screening exams, endoscopy is useful for identifying precancerous lesions like polyps because it provides a clear image of the colon and rectum.

This skill is essential for preventing colorectal cancer because it eliminates the precursors before they have a chance to develop into cancer. Moreover, because endoscopy allows for the direct sight of the gastrointestinal system, it is an essential tool in diagnosing lower GI malignancies. The method makes tissue biopsies, abnormality identification, and sometimes even real-time treatments easier.

Advanced endoscopic modalities such as virtual chromoendoscopy and enhanced imaging technology boost diagnostic precision and lesion characterization. These technological advancements improve the capacity to identify minute irregularities that traditional techniques could miss. Lower GI tumors are especially important in the field of

endoscopy, where the procedure's diagnostic and therapeutic capacities intersect. Endoscopic screenings, in conjunction with developments in artificial intelligence and imaging technology, may be able to prevent lower gastrointestinal tract tumors. Endoscopy is, therefore, an essential tool in the continuous quest to enhance early detection, diagnostic precision, and treatment options for lower gastrointestinal cancer.

## 2. Global Health Metrics

Lower Gastrointestinal (GI) cancers, particularly colorectal cancer, are a significant and influential subset of cancers when seen through the lens of global health indicators. With high incidence rates seen in a variety of populations, these malignancies considerably increase the burden of illness worldwide. Specifically, colorectal cancer is a major global source of cancer-related morbidity and death.

Lower GI malignancies are common in industrialized and developing countries, highlighting their significance for global health. Because these tumours require efficient screening methods, early detection programs, and easily accessible treatment options, they provide special potential and difficulties to healthcare systems. The economic impact of lower gastrointestinal malignancies stems from lost production in society and the expense of treatment. It goes beyond simple medical concerns.

The effect of lower GI cancers on mortality, their widespread prevalence, and the difficulties in prevention, early identification, and treatment make them a major concern in global health measures. Improving overall health outcomes and addressing the worldwide burden of lower GI malignancies require strategies incorporating advances in screening, diagnosis, and treatment techniques.

Globally, colon and rectal cancer caused 2.17 million (95% UI 2.00–2.34) incident cases, 1.09 million (1.00–1.15) deaths, and 24.3 million (22.6–25.7) DALYs in 2019. In the world in 2019, colon and rectal cancer ranked as the fifteenth most common Level 3 cause of death [2]. The GBD research is being carried out by the University of Washington’s Institute for Health Metrics and Evaluation (IHME). The precise count of references consulted for the colorectal cancer estimate in the Global Burden of Disease (GBD) Table 1. Researchers, organizations, and data sources from across the world are working together on the Global Burden of Disease Study.

The precise proportion of Disability-Adjusted Life Years (DALYs) in Figure 2 in 2019 attributed to the highest risk factors combined for both sexes, broken down by GDP. Such

information is usually included in extensive reports released by the Institute for Health Metrics and Evaluation (IHME) as part of its Global Burden of Disease (GBD) research in Figure 3.

The detailed breakdown of Disability-Adjusted Life Years (DALYs) for the year 2019 in Figure 4, broken down by Gross Domestic Product (GDP), age group, and sex, as well as the separation between Years of Life Lost (YLLs) and Years Lived with Disability (YLDs). The age-standardized Disability-Adjusted Life Years (DALY) rates, broken down by Gross Domestic Product (GDP), for each area in 2019 using the Socio-Demographic Index (SDI). Usually, the Institute for Health Metrics and Evaluation (IHME) releases specialized Global Burden of Disease (GBD) in Figure 5 studies that provide this level of comprehensive data.

Table 1. Overall resources utilized for estimating GBD 2019

	Total Sources
Occurrence	3357
Frequentness	3
Remission	0
Cause of Death	5354
Others	0

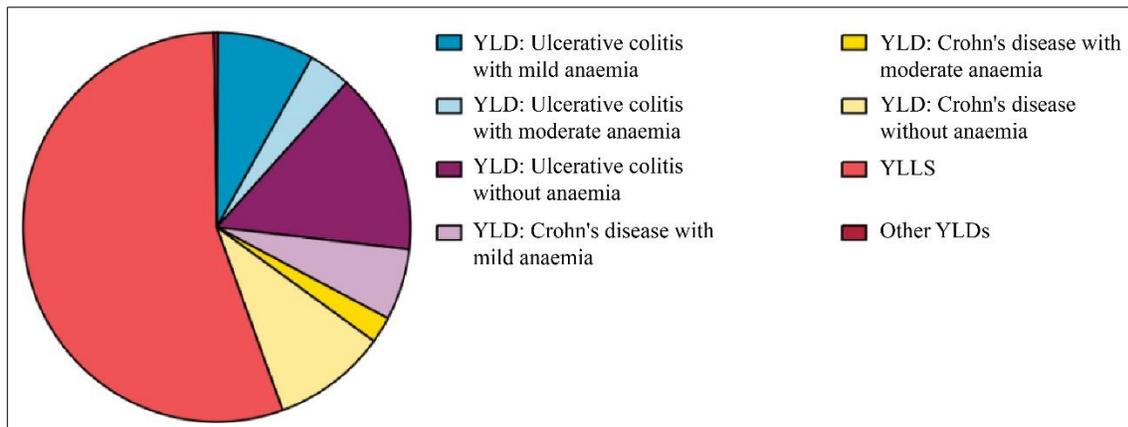


Fig. 2 Distribution of DALYs by individual sequelae YLDs and YLLs aggregated for both sexes, 2019

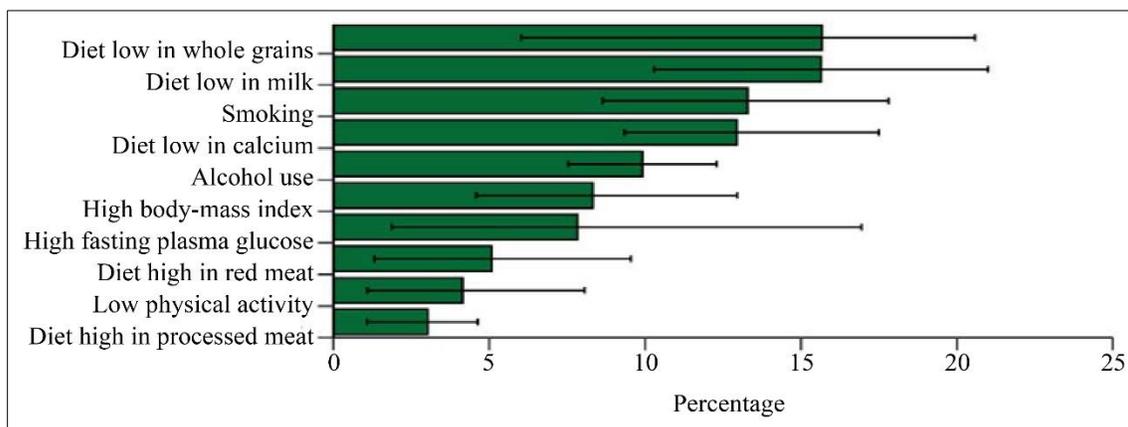


Fig. 3 2019 DALYs due to the major risk factors for both sexes aggregated as a percentage

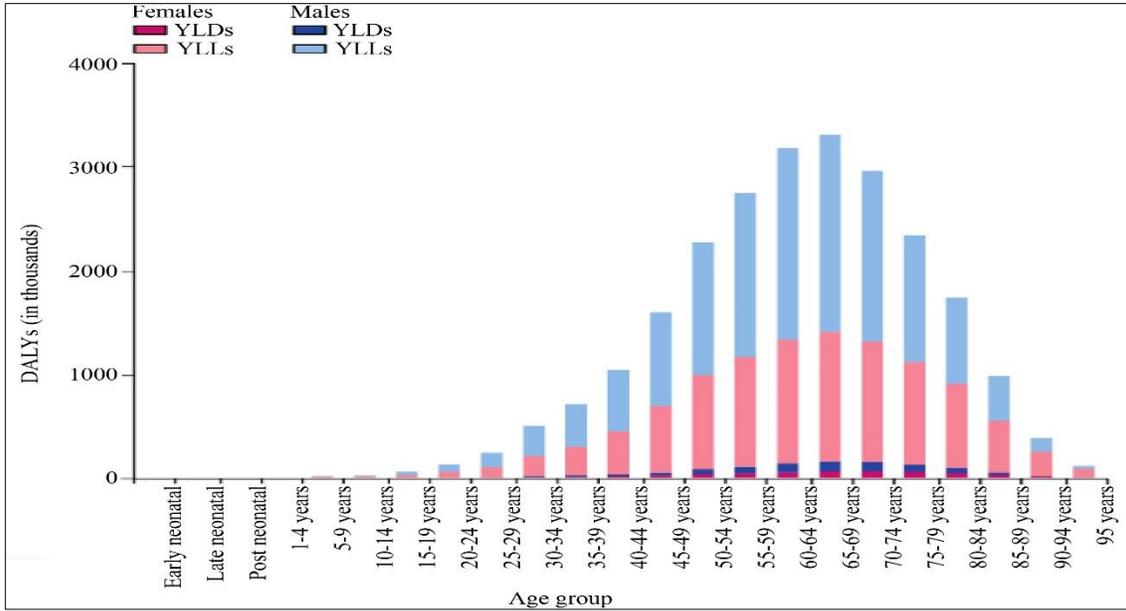


Fig. 4 Age group, sex, and the breakdown of DALYs by YLLs and YLDs in 2019

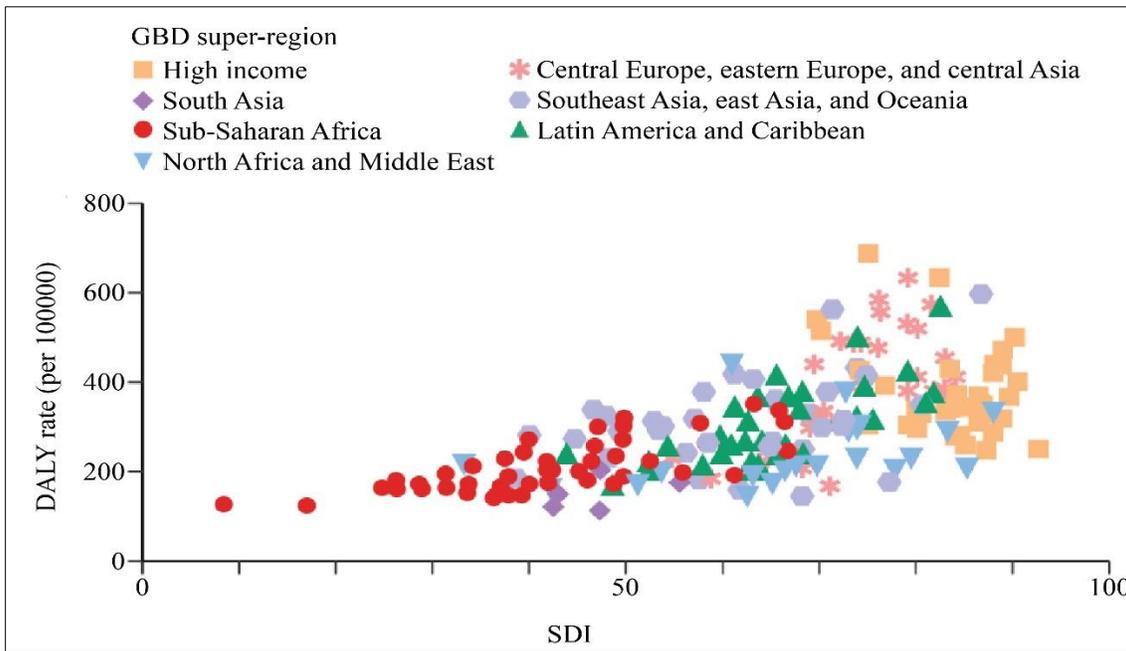


Fig. 5 Age-standardized DALY rates by SDI, summed for both sexes, for each location in 2019

### 3. Problem Definition and Motivation

Illnesses affecting the colon, rectum, and anus—the lower part of the digestive system—are referred to as lower Gastrointestinal (GI) illnesses. Symptoms of these disorders might include bleeding, discomfort, altered bowel habits, and abdominal pain. If left untreated, lower GI disorders can have a major negative influence on a person’s quality of life and provide major health hazards. Approximately 3.6 million new instances of GI malignancies are reported globally each year, of which 1.6 million are associated with cancers of the

stomach and esophagus. These malignancies claim the lives of over 2.7 million individuals annually, of which 1.3 million are attributable to the stomach and esophageal cancers.

Gastroenterologists should write endoscopic procedure reports following every endoscopy; these reports are a crucial component of their profession. The World Endoscopy Organization (WEO) recommends using Minimum Standard Terminology (MST) and Minimal Standard for Reporting (MSR).

The gold standard for GI tract evaluation these days is endoscopy, yet operator performance variability substantially limits its usefulness. Rational screening, better clinical examinations, and enhanced endoscopic performance are essential for reducing morbidity and death associated with GI diseases. Support systems driven by AI have demonstrated potential in arming medical professionals with the resources required to deliver high-quality treatment to a large patient population.

## 4. Public Datasets

### 4.1. KVASIR Dataset

In this paper, the author investigated the field of lower Gastrointestinal (GI) disorders; the KVASIR dataset is an invaluable tool, especially for computer-aided diagnostic research. It is made up of a variety of endoscopic photos and movies that were taken during colonoscopy treatments. The dataset provides a thorough depiction of Lower GI disorders by encompassing a variety of pathological states, including polyps, ulcerative colitis, and colorectal cancer.

Researchers use KVASIR to create and assess machine learning algorithms for automated colon and rectum abnormality detection and classification, particularly in the field of computer vision. The dataset is essential to improving our understanding of these illnesses and enabling the creation of novel diagnostic techniques and tools that will enhance early diagnosis and treatment of lower GI disorders.

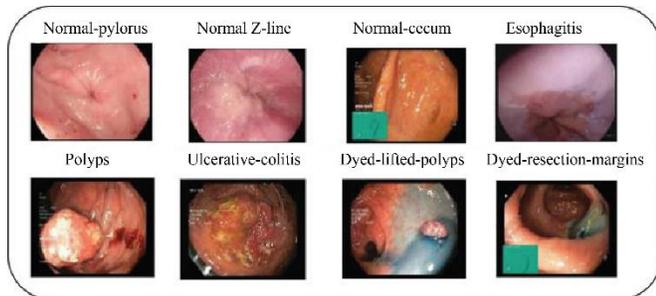


Fig. 6 KVASIR dataset based on eight labels

Figure 6 displays the KVASIR dataset, which comprises hundreds of photos per class that display anatomical landmarks, clinical abnormalities, and GI tract endoscopic operations. Medical professionals with competence in endoscopy have confirmed and commented on the images. There are sufficient photos for a wide range of uses, such as machine learning, deep learning, transfer learning, and visual retrieval. Z-line, pylorus, cecum, and other anatomic landmarks are examples of pathological findings; esophagitis, polyps, ulcerative colitis, and so forth are examples of anatomic landmarks. Furthermore, we provide other sets of photos about lesion removal, such as “lifted and dyed polyp” and “dyed resection margins,” among others. The collection consists of pictures with varying resolutions, ranging from

720x576 to 1920x1072 pixels, arranged and named according to the content.

### 4.2. The Datasets from the MEDICO 2018 and BIOMEDIA 2019 Challenges

Among the tasks involved in this assignment is the evaluation of the techniques used to categorize GI tract normal and regular cases (normal colon mucosa, stool, instrument, etc.) and pathological findings (ulcerative colitis, esophagitis, polyps, lifted and dyed polyps, etc.) as well as anatomical landmarks (e.g., z-line, pylorus, cecum) [7]. GI endoscopy computer-aided tools need to be built with excellent classification accuracy. This takes care of this. The teams are graded according to how well their classification algorithms perform across 16 classes in Figure 7, the GI dataset.

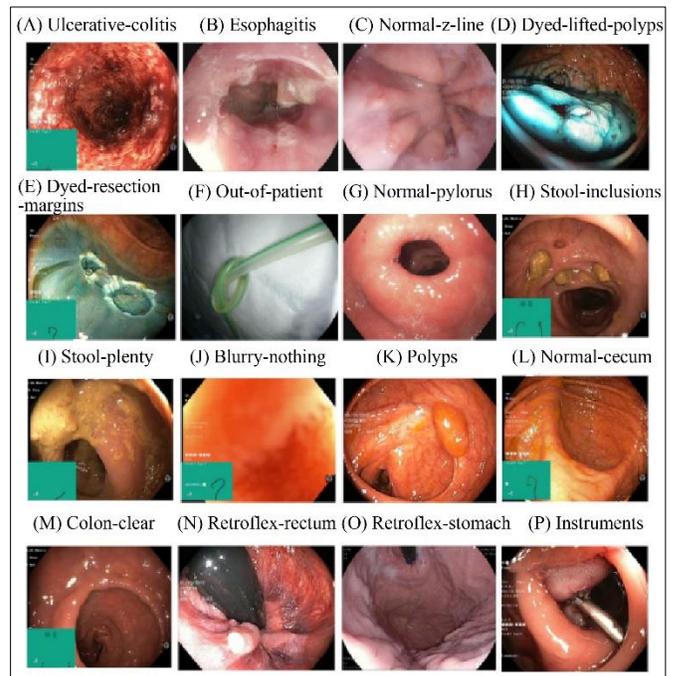


Fig. 7 MEDICO 2018 and BIOMEDIA 2019 challenge datasets with 16 classes

### 4.3. Hyper-KVASIR Dataset

The collection includes 10,662 JPEG-formatted tagged photos in total. The pictures folder contains the images in Figure 8. Every image in the collection belongs to a class that matches the folder in which it is kept (for example, all polyp images are in the “polyp” folder, all Barrett’s esophagus photographs are in the “barretts”, etc.).

Because certain discoveries occur more frequently than others, the medical industry has a common difficulty regarding the unbalanced quantity of photos per class, as shown in Table 2. Because techniques applied to the data must also be able to learn from a little quantity of training data, this presents an extra barrier for researchers.

There are 23 distinct [13] classes shown in Figure 9 of findings represented by the labeled photos. A segmentation mask and a bounding box are applied to each of the 1,000 images in the polyp class. While the backdrop (black) of the mask is devoid of polyp pixels, the foreground (white mask) of the mask depicts the pixels that display the polyp tissue or the region of interest. The bounding box of the found polyp is defined as its outermost pixels. For this segmentation set, we have two files: one containing masks and the other containing images.

Table 2. Hyper-KVASIR dataset specifics

Data Record	Records	Description
Videos	373 Videos	30 different classes
Segmented Images	1,000 pictures	Mask for Polyp discoveries
Labeled Images	10,662 pictures	23 classes
Unlabelled Images	99,417 pictures	Unlabelled

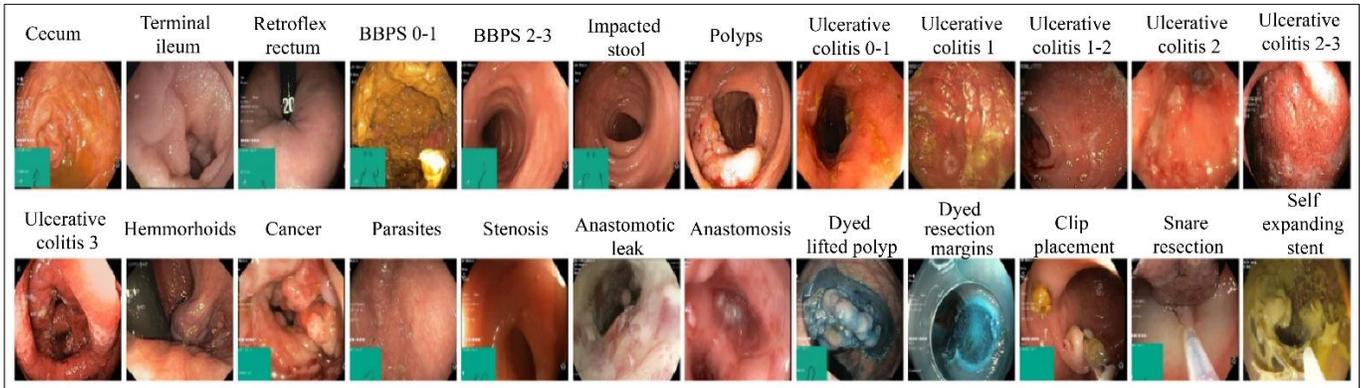


Fig. 8 Hyper-KVASIR dataset with 24 distinct classes

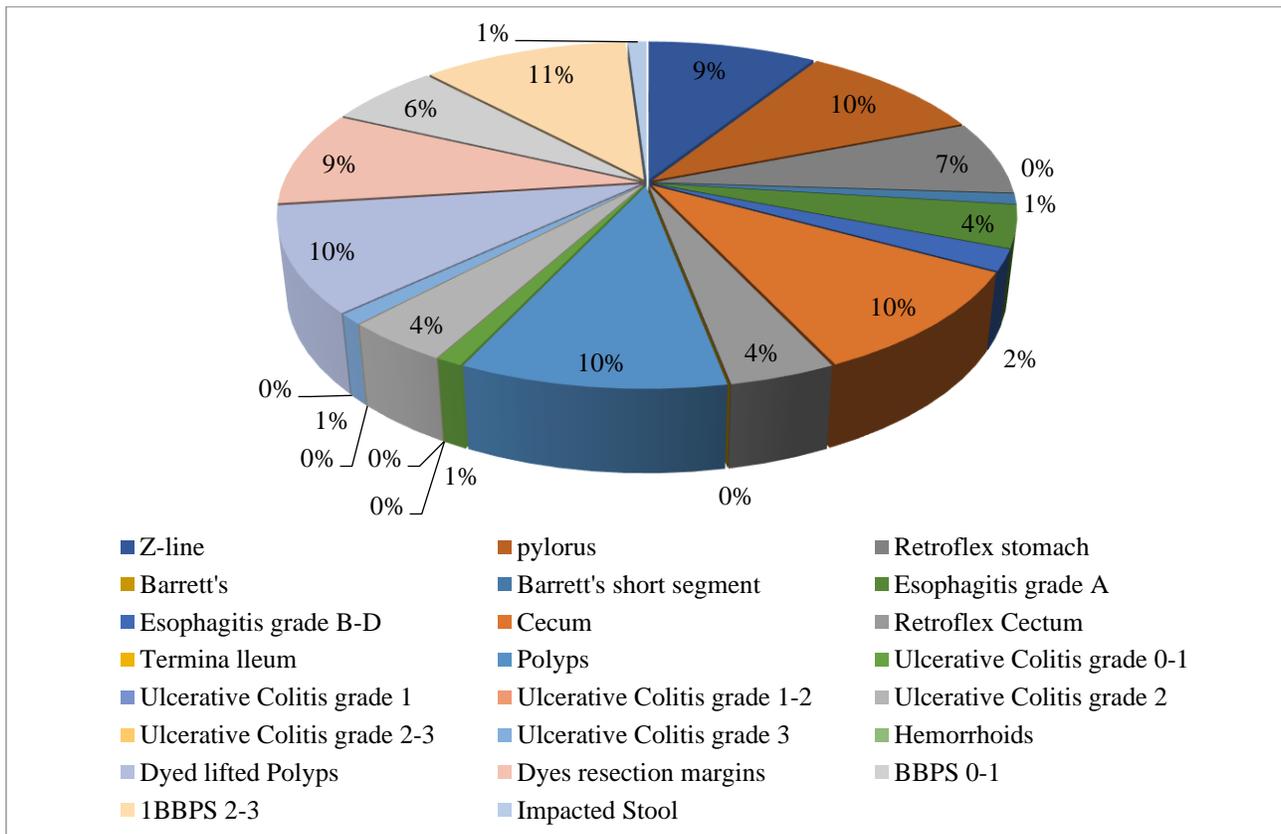
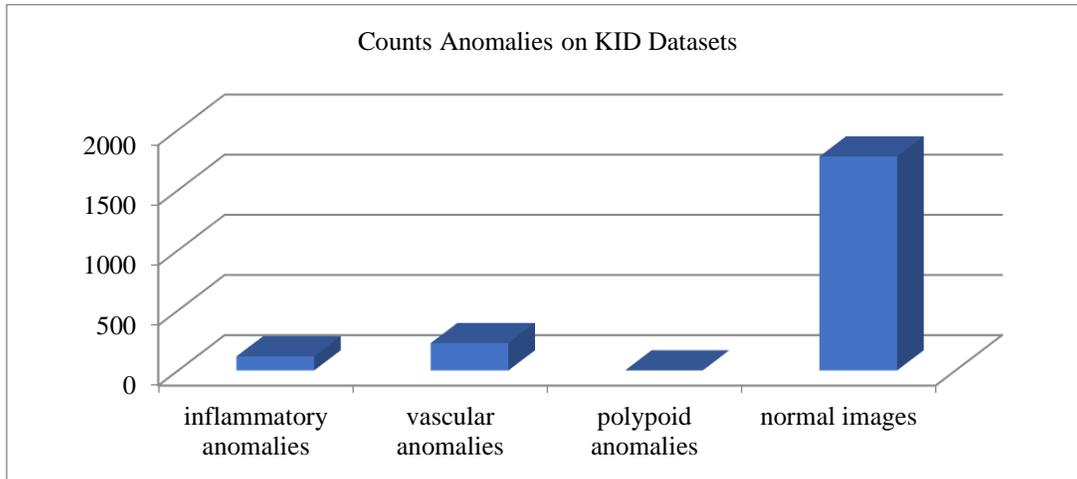


Fig. 9 Amount of Hyper-KVASIR classes

**4.4. KID Dataset**

Pixel-level annotations for WCE photos and videos are made publicly available in the form of the KID dataset. It is made up of 360 x 360-pixel WCE pictures that were taken with a Miro-Cam capsule endoscope all the way down the gastrointestinal system. These consist of 227 images that show inflammatory anomalies like ulcers, aphthae, mucosal breaks with surrounding erythema, cobblestone mucosa, luminal stenoses and/or fibrotic strictures, and mucosal/villous

oedema; 303 images that show vascular anomalies like minor bowel angiectasias, lymphoid nodular hyperplasia, and blood in the lumen; 44 images that represent polypoid anomalies like lymphoid nodular hyperplasia, lymphoid polyplasia, Peutz-Jeghers polyps, and blood in the lumen); and 1,778 images that show normal images from the stomach, small intestine, and colon. In this collection, there are 2,352 images displayed in Figure 10.

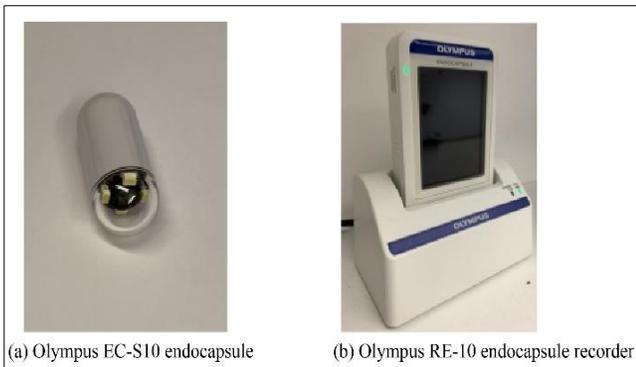


**Fig. 10** Count anomalies on KID dataset

**5. KVASIR Capsule Dataset**

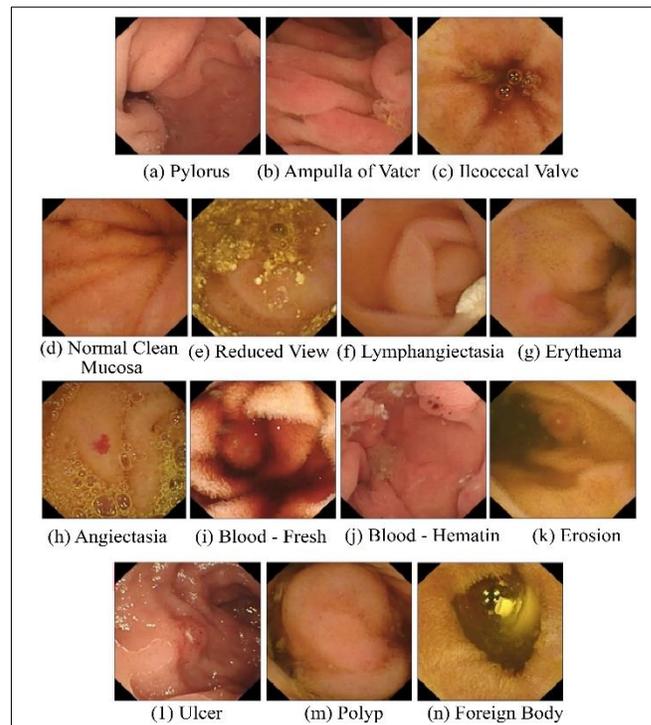
Enhancing anomaly detection while lowering manual labor is possible. However, experienced medical staff seldom have time for the laborious labeling task, and medical data is frequently scant and not accessible to the scientific profession. In this regard, we introduce KVASIR-Capsule, a sizable VCE dataset gathered from assessments conducted in Norwegian hospitals. [8] From the 117 movies that make up KVASIR-Capsule, It is possible to extract 4,741,504 photo frames. Discovered anomalies from 14 different classes have encircled us.

collection also contains 4,694,266 unlabelled frames in addition to these labelled pictures displaced in Table 3.



**Fig. 11** Wireless capsule endoscopy capsule and recorder

Figure 11 shows results with a bounding box in 47,238 frames that have been labeled and medically validated. The



**Fig. 12** Classes of KVASIR capsule dataset

**Table 3. Details of the KVASIR capsule dataset**

Categorized pictures	47,238
Categorized videos	43
Unidentified pictures	46,94,266
Unidentified videos	7

The possible benefits of computer-aided diagnostic devices based on artificial intelligence for VCE are shown by preliminary research. Still, they demonstrate a great deal of room for development, and the KVASIR-Capsule dataset in Figure 12 can be a useful tool for creating more advanced algorithms that will allow VCE technology to realize its full potential.

## 6. A Brief Description of Techniques for Deep Learning

In this section, the author describes an introduction to DL approaches, which belong to the ML branch. Artificial Intelligence (AI) is what DL and ML refer to. Computational models consisting of several processing layers can acquire representations of data with various degrees of abstraction through deep learning.

The state-of-the-art has been significantly enhanced by these techniques in several fields, including drug discovery and genomics, voice recognition, visual object identification, and object detection. By employing the backpropagation technique to suggest changes to a machine’s internal parameters, which are used to compute the representation in each layer based on the representation in the preceding layer, deep learning uncovers complex structures inside massive data sets.

DL approaches are often classified into two primary categories: supervised learning and unsupervised learning. DL architectures widely used in GI image processing are supervised and trained on labeled data. As stated earlier, CNN (supervised learning) is the foundation of almost all deep network research, including Artificial Neural Networks (ANN) and Deep Neural Networks (DNN) utilized in GI image processing. Next, we will cover a detailed overview of CNN and a brief summary of the many DL architectures used in GI image processing.

### 6.1. CNN-Based Neural Network

Deep learning includes the use of Convolutional Neural Networks (CNNs) architecture that is especially intended for structured grid data processing and analysis; it is especially well-suited for image-related applications. Convolution is a basic operation that CNN uses at its foundation. To extract spatial characteristics from an input picture, a filter or kernel must be scanned. Mathematically, the convolution operation is expressed as:

$$(f \cdot g)(x, y) = \sum_{a=1}^m \sum_{b=1}^n f(a, b)g(x - a, y - b)$$

Here, g is the filter or kernel, f is the input image, and (x,y) are the spatial coordinates. The input picture is subjected to several filters by the convolutional layer, which produces feature maps that capture various facets of the spatial hierarchy of the image.

Rectified Linear Unit (ReLU) activation functions frequently add non-linearity to networks. The definition of the ReLU function is:

$$RELU(x) = \max(0, x)$$

This activation function helps the network understand complex patterns and facilitates extracting important features by replacing negative pixel values in the feature maps with zero. One important purpose of pooling layers, like Max Pooling, is to reduce spatial dimensions to minimize overfitting risk and computational expense. Max pooling selects the maximum value within a restricted region. The Max Pooling operation is expressed as follows for a 2x2 area:

$$Max\ Pooling(x, y) = \max \left( \begin{matrix} f(x, y), f(x + 1, y), \\ f(x, y + 1), f(x + 1, y + 1) \end{matrix} \right)$$

The network can generate predictions based on learnt properties thanks to fully linked layers, which link every neuron in one layer to every other layer’s neuron. The following formula is used to calculate the output (y) of a neuron in a fully connected layer:

$$y = \sigma \left( \sum_{i=1}^N w_i \cdot x_i + b \right)$$

Here, N is the number of neurons in the previous layer, σ is the activation function (such as sigmoid or SoftMax), x<sub>i</sub> is the output of neuron. The amount of weight connected to the link from neuron I in the preceding layer is represented by w<sub>i</sub> in that layer, and b is the bias factor.

The output layer frequently uses the SoftMax activation function for multiclass classification problems. The probability P(y = j) that the input belongs to class j is calculated by the SoftMax function in the following way:

$$P(y = j) = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$$

Figure 13 explains that CNNs use pooling layers, activation functions, and convolutional operations to learn hierarchical representations of features. CNNs are essential to

computer vision and image analysis because of these properties, which are subsequently utilized for tasks like segmentation, object recognition, and picture categorization.

The design is especially useful for processing and analyzing complicated visual information since it can be adjusted to the spatial connections in the data.

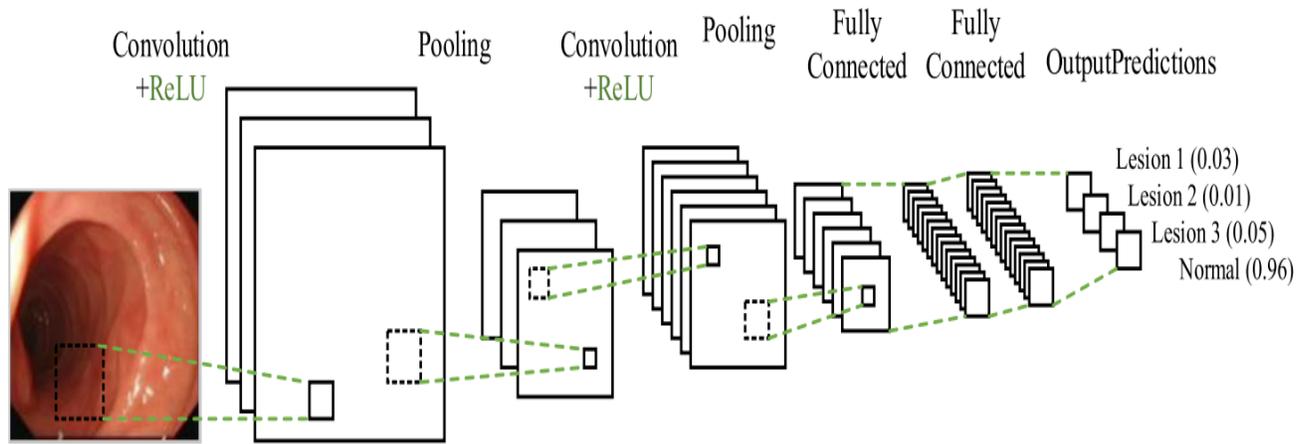


Fig. 13 An elementary illustration of CNN-based GI image categorization. Convolution layers retrieved the characteristics and forwarded them to fully linked layers. Fully linked layers distributed the anticipated categorization results.

## 6.2. Supervised Deep Learning Architectures

An overview of the well-liked deep learning architectures built on the supervised method of GI image analysis is given in this section.

### 6.2.1. Classification Architectures

Yann LeCun and his colleagues created LeNet-5 to recognize handwritten digits, especially in postal codes on letters. LeNet-5 uses convolutional layers to investigate kernels. From the supplied image, local information is extracted. The convolutional technique involves swiping tiny filters, or kernels, across the image. Patterns like edges, corners, and textures are detected by these trained kernels.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton created AlexNet to increase the accuracy of picture categorization. It took first place in the 2012 ImageNet Large Scale Visual Recognition Challenge. AlexNet uses many convolutional layers to investigate kernels. These layers identify both local and global patterns in the input pictures by learning a hierarchy of characteristics [29]. Investigating kernels at various tiers enables network operation.

### 6.2.2. Detection Architectures

F. Yasmin et al. offer a model that applies hyperparameter fine-tuning to Yolov5 [30] to improve optimization and precisely detect coloured lifted polyps and esophagitis. The entire thing is divided into these five sections: prediction, input, backbone, neck, and hyperparameter fine-tuning, illustrating the architecture of the proposed concept.

We used a total of five models to precisely identify and classify targeted polyps and aberrant characteristics such as esophagitis. Out of the four alternative models, the model we proposed had the greatest f1 and mAP values. This is a succinct description of the model-fine-tuning process.

Drive Solid State Lean network technology and a novel depth-by-depth separable convolution one-stage item recognition model are features of MOBILENET V2 SSD (Single Shot MultiBox Detector). MobileNetV2, an architecture for convolutional neural networks, is designed to be mobile device responsive. Its foundation is an inverse residual structure with links between the layers that constitute the bottleneck [31].

The model includes an FPN-lite feature extractor, shared box predictor, and focal loss for SSD Mobilenet V2 object identification. Feature Pyramid Network (FPN) is a fully convolutional feature extractor that can handle any size single-scale image input and produces appropriately sized feature maps at various layers. This process is independent of the foundation’s convolutional models.

Figure 14 displays the SSD Resnet50 v1 FPN model as one object detection model. Single Shot MultiBox Detector (SSD) is a method for locating objects at any location on a feature map with varying aspect ratios and sizes. Multiple preset boxes are separated within the bounding box’s output region. Feature extractor Feature Pyramid Network (FPN) is the brains behind the 50-layer deep convolutional neural network ResNet-50 v1 FPN [32].

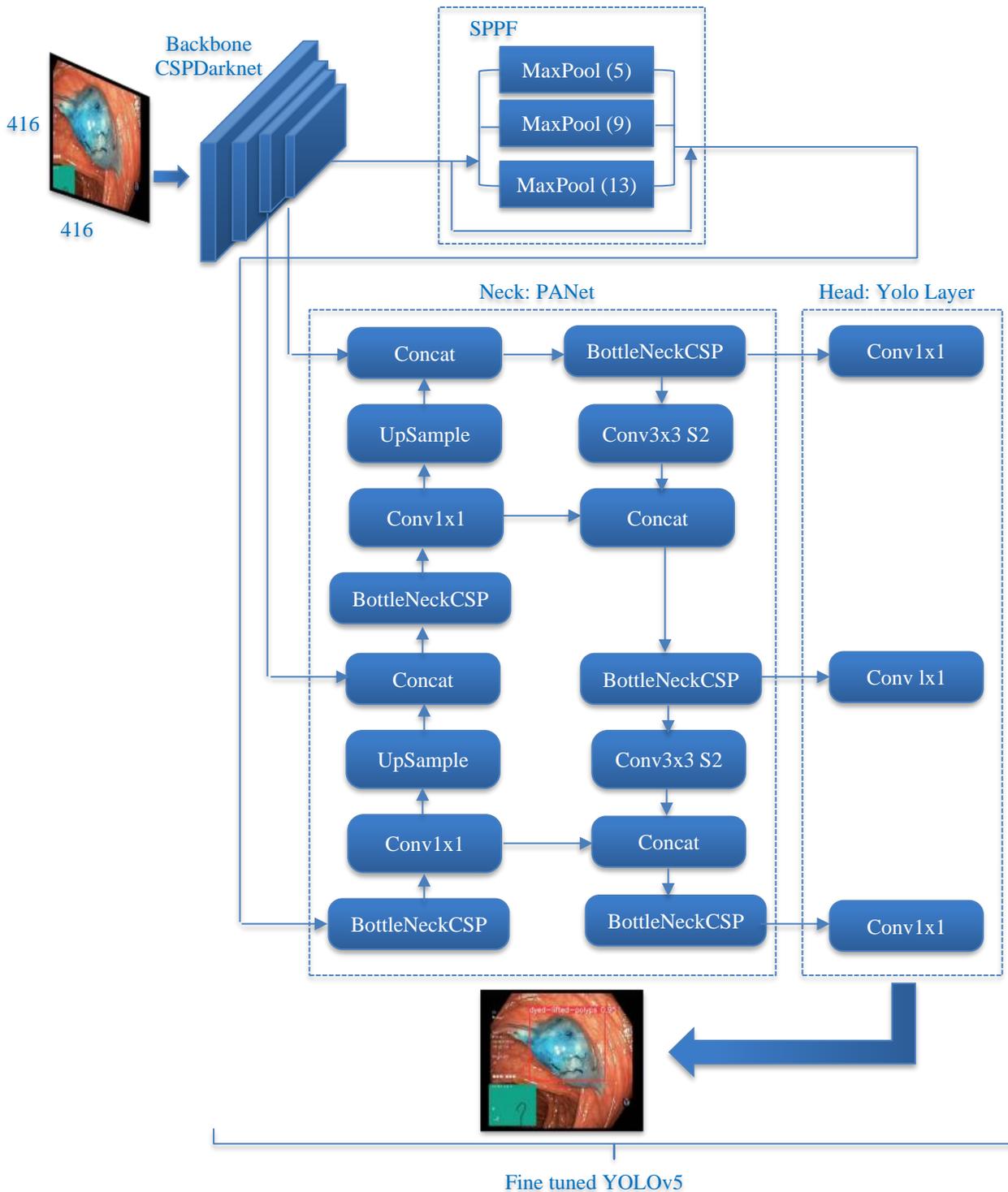


Fig. 14 The suggested model's architecture uses yolo layers as the head, PANet as the neck, and CSPDarknet as the backbone

### 6.2.3. Pre-Processing and Segmentation Architectures

In medical picture denoising, Convolutional Neural Networks (CNNs) and other deep learning architectures are being used more and more. These models are useful for identifying and eliminating noise patterns because they can

understand the correlations between noisy and clean pictures. Getting a picture without losing any important information is a big difficulty in the medical imaging procedure. The noise was present in the acquired photos, and this noise had an impact on the suggested model's classification accuracy.

Gaussian Filter (GF), Median Filter (MF), and Adaptive Median Filter (AMF) are examples of basic digital picture filters.

*Gaussian Filter*

Often employed in image processing, Gaussian filters work by convolving a picture with a two-dimensional Gaussian kernel. A weighted average of the pixel values is produced by this convolution, which efficiently reduces high-frequency noise but introduces significant blurring, especially near edges.

*Median Filter*

Conversely, non-linear median filters function by swiping over the picture and substituting the local window median value for each pixel value. Because this approach is insensitive to high values, it is especially good at reducing impulse noise (salt-and-pepper noise) while keeping edges and fine features intact.

*Adaptive median Filter*

Adaptive Mean Filters (AMF) compute the mean of a neighbourhood for each pixel, providing a localized method of noise reduction. It can adjust to differing noise levels in different picture regions by replacing a pixel with the local mean if The value of the pixel and the local mean diverge more than a certain amount. All things considered, these filtering methods are essential for improving the quality of medical and other photographs by lowering undesired noise and artefacts.

Different segmentation techniques are used in lower Gastrointestinal (GI) illnesses to locate and define regions of interest in medical imaging. Among these methods are:

*Limiting*

In a nutshell, thresholding is the process of classifying picture areas by establishing a pixel intensity threshold. In lower GI pictures, it is utilized to differentiate between various tissues or anomalies according to their intensity levels.

*Growing Region*

In a nutshell, area expansion begins with a seed point and grows the region by appending surrounding pixels that satisfy specific requirements, usually related to similarity in intensity or texture. In lower GI pictures, it helps segment linked entities.

*6.2.4. Post-Processing and Segmentation Architectures*

Medical image analysis, particularly the analysis of Gastrointestinal (GI) pictures, has been investigating unsupervised deep learning architectures more and more. With the goal of learning data representations without labelled samples, these architectures might be useful in situations where getting labelled data is difficult or costly. Several popular unsupervised deep-learning techniques for analyzing gastrointestinal pictures are listed below:

A discriminator trained adversarial and a generator make up a GAN. While the discriminator works to separate created samples from genuine ones, the generator produces realistic data. Data augmentation, domain adaption, and the creation of medical images have all been done with GANs. Similar patterns in GI pictures may be grouped together using clustering algorithms like DBSCAN, K-means, and hierarchical clustering. Optimizing similarity between positive pairings (similar samples) and reducing similarity between negative pairs (dissimilar samples) is the goal of contrastive learning. Applying it to medical imaging applications such as gastrointestinal pictures has allowed for learning meaningful representations. Patterns or abnormalities in the data can be found using these techniques. In order to find similar samples in a lower-dimensional space, deep embedding models learn a mapping from the input space.

*6.2.5. Other Networks*

A basic CapsNet design is displayed in Figure 15. With only two convolutional layers and one fully linked layer, the architecture is simple. Sara et al. [11], Conv1 features ReLU activation and  $256 \times 9 \times 9$  convolution kernels with a stride of 1. This layer translates pixel intensities into the actions of nearby feature detectors, which are then sent into the main capsules. Activating the main capsules is equivalent to reversing the rendering process since they represent the lowest level of multi-dimensional things from an inverted graphics perspective. Capsules are meant to be good at assembling instantiated components into recognizable wholes; this is a fundamentally different kind of computation. The second layer, Primary Capsules, comprises eight convolutional units with a  $9 \times 9$  kernel and a stride of two in each of the major capsules of this 32-channel convolutional 8D capsule layer. All  $256 \times 81$  Conv1 units have their receptive outputs visible to each primary capsule output.

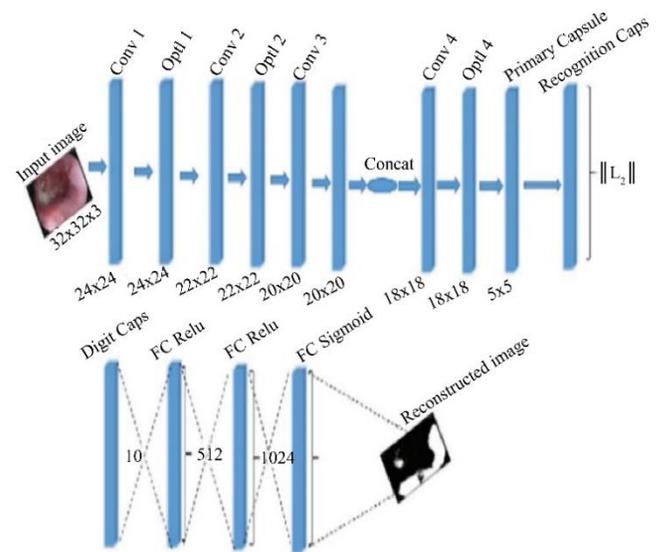


Fig. 15 Proposed architectures on a three-layer convolutional capsule

Many additional effective networks, such as Recurrent Neural Networks (RNNs), Graph Neural Networks (GNNs) [17], Principal Component Analysis Networks (PCANet) [18], and Canonical Correlation Analysis Networks (CCANet) [19], have not yet been applied in GI image analysis in addition to the DL networks. RNNs were created to study discrete sequences. Other applications of medical image processing, such as tissue segmentation, have made use of them [20].

GNNs use the most recent neural network techniques to handle data stored in a graph domain; they were initially proposed in 2009. Few studies have used GNNs for GI pictures and other medical images, even though they are often used for natural or other image-processing tasks.

RNNs are more proficient at processing serialized data and have the ability to map sequences of inputs to generate sequences [21]. For instance, the work in [22] combines CNNs with RNNs, enabling the analysis of all contextual data, regardless of the image size. GNNs allow the DL model some capacity for causal reasoning, which means it can handle extensive relationship information among components, which may be useful in classifying disorders [23].

Effective networks that have been used for the categorization of images of nature include PCANet and CCANet. One distinction between the two is that CCANet can categorize pictures represented by two-view features, but PCANet can only handle data expressed as one-view features.

In the GI image analysis challenge, RNNs, GNNs, PCANet, and CCANet all show promise for the future. Recent years have seen a considerable advancement in the use of Artificial Intelligence (AI) in medical imaging and diagnostics for the early detection and diagnosis of various disorders. Artificial Intelligence (AI) has demonstrated significant promise in enhancing identification precision and overall diagnostic efficacy in relation to Gastrointestinal (GI) issues, including abnormalities and gastrointestinal polyps. Several studies have examined the use of AI algorithms for the detection of various gastrointestinal illnesses [25, 27, 28]. Figure 17 of the paper compares the ResNet50, MobineNetV2, and EfficientNet-B1 encoders with seven traditional semantic segmentation models [26].

An integrated evaluation method incorporating subjective and objective data is recommended to choose the optimal CNN model. Using the MobineNet v2 encoder and UNet++, the automatic polyp-segmentation system is constructed. The semantic segmentation model has significant clinical use in diagnosing stomach polyps, and the assessment method is impartial and objective. As part of Healthcare 4.0, a Masked Graph Neural Network model (MGNN) for real-time polyp detection in gastroscopic images is described in this study.

In order to compensate for manual labelling, the model extracts geographical and semantic information using graph structure and convolution processes. It's been tested using actual gastroscopy pictures.

6.2.6. Techniques Based on Transfer Learning

It might take a while to train and optimize a deep network from scratch since it requires a large amount of labelled data. Acquiring an extensive collection of GI images and having professionals transcribe the accompanying labels is another difficult and error-prone process. Because of this, transfer learning is used in the majority of GI image analysis jobs based on DL techniques, decreasing the requirement for a deep network for training data.

When a deep model is referred to as pre-trained, it has been trained on a sizable picture dataset (like ImageNet). The feature extractor is one approach to transfer learning. The pre-trained model's fully connected layers are replaced with conventional classifiers, like the linear classifier SVM, and the CNN layers are utilized as feature extractors. GI image analysis jobs with limited sample sizes typically use this transfer learning approach.

The process of "fine-tuning" is another technique for transfer learning. Instead of using its input layer, the pre-trained model is trained using new data. One can fine-tune any or all of the layers in the pre-trained deep model. The generic qualities of a picture, such as its color and edge, are frequently extracted by a deep network's first layers and are useful for a wide range of applications.

Usually, the latter layers are the ones that receive fine-tuning since they collect attributes specific to a certain task. The three primary GI analysis tasks-classifying and detecting lesions in GI images are depicted in Figure 16.

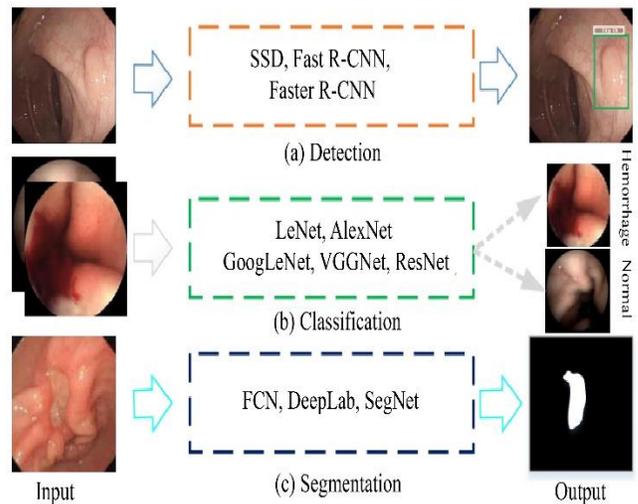


Fig. 16 The three primary GI image analysis tasks

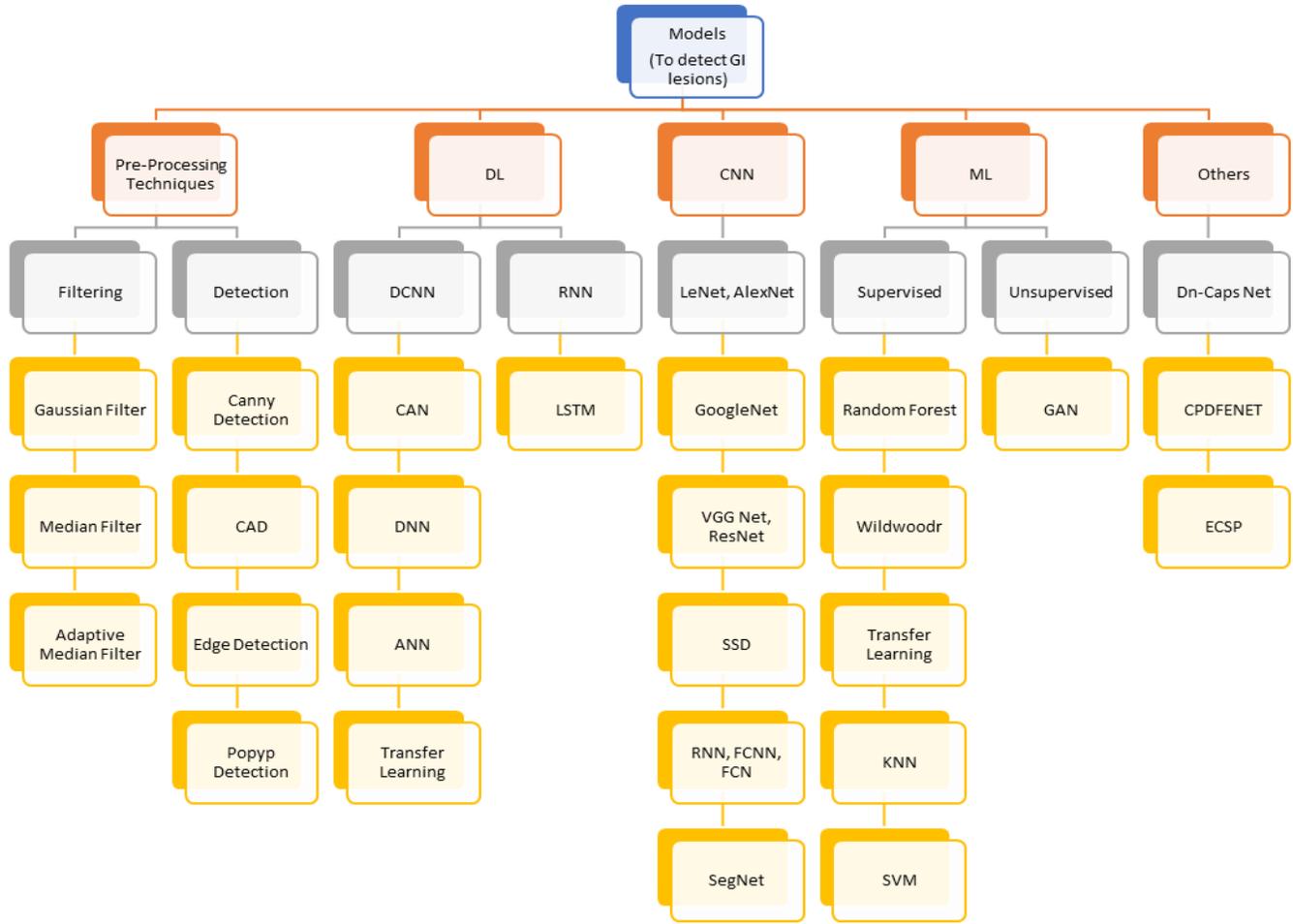


Fig. 17 Model that is frequently used for classification and detection

## 7. Selected Study that Employs in ML Techniques

The author WildWood (WW) proposed a unique ensemble approach for supervised learning of the Random Forest (RF) kind. While traditional RF approaches utilize these samples to compute out-of-bag scores, Dev Gupta, WW uses bootstrap out-of-bag samples to create improved predictions provided by an aggregation of the forecasts of all possible subtrees of each fully grown tree in the forest. This is achieved by aggregating over out-of-bag data using accurately and fast computed exponential weights.

This improvement, along with a histogram-based strategy to accelerate split detection, makes WW fast and competitive when compared to other well-established ensemble methods like extreme gradient boosting algorithms and traditional RF. This RF and WW act in NN models achieved 82.9375 and 82.625, respectively. Using this method, they found the best hyperparameters. Effectiveness-wise, voting is comparable to the state-of-the-art method of stacking with multi-response model trees [5] minus the additional computational load of

meta-training. The study [4] on endoscopic images obtain 88.1875 accuracy with canny detection and 87.4375 without canny edge detection.

The GI concerns brought up and the imaging methods used during the course of the preceding five years were summarized by Jha et al. [7]. Medico 2017, Medico 2018, and Bio-Media 2019 challenges only. Evaluate and consider the practicality and usefulness of Machine and Deep approaches in the context of GI tracks based on WCE pictures.

In Machine Learning (ML), an algorithm takes raw data as input, analyses it for characteristics in a different dataset, and outputs a categorized result based on the requirements. Detection and categorization of images is one of the most popular applications of machine learning in medicine. In classical machine learning, the system is trained using a training set of pictures that include the relevant categories. This leads to improved performance and fewer mistakes. Following several training phases, an independent collection of pictures is used to assess the system's performance. In

classical machine learning, algorithms like Support Vector Machines (SVM) and Multi-Layer Perceptrons (MLP) are frequently used. In GI image analysis, SVM is frequently employed. While MLP, KNN, and random forest are used in eight publications, SVM is used in sixteen. Investigated articles indicate that the range of performance for SVM in detecting bleeding is around 0.87 - 0.98 for accuracy, 0.85 - 0.98 for sensitivity, and 0.93 for specificity. The KNN process research results show competitive performance with 0.96 - 0.99 accuracy, 0.92-0.99 sensitivity, and 0.96 - 0.99 specificity. E.Mossotto [10] Endoscopic, histological, and combined endoscopic and histological data were used to develop three supervised machine learning models. Classification accuracy scores of 71.0%, 76.9%, and 82.7% were generated by the models. The optimal combination model was tested in a statistically independent cohort of 48 PIBD patients from the same clinic. It accurately categorized 83.3% of the patients.

To assess lymph node metastases in gastric cancer, Li et al. (2012) combined machine-learning approaches with gemstone spectral imaging [16]. The kNN classifier was used for 38 lymph node samples from patients with gastric cancer, yielding an overall accuracy of 96.33% in separating lymph node metastasis from non-lymph node metastasis. Feature selection and metric learning methods were used to reduce feature space and data dimension. Wang et al. (2015) created a method for detecting polyps during colonoscopy. During a colonoscopy, it is capable of initiating an alarm and providing prompt feedback. The researchers identified polyp boundaries using a rule-based classifier and visual cues. 97.7% of polyps were detected with the method's [15] accuracy.

**8. Study Selected that Employs DL Techniques**

In Deep Learning (DL), the author described the techniques presently recognized as the most sophisticated AI approaches because of the cutting-edge capabilities of Deep Convolutional Neural Networks (DCNNs). Deep learning has gained traction in two areas: identifying and categorizing images and videos, which has shown encouraging results. The significant progress made in picture and video identification on large-scale annotated training sets has prompted a wide range of businesses, including the medical sector, to create image and video recognition systems. Consequently, recent advancements in medical image analysis have included DL techniques. Using skip connections, ResNet solves the vanishing gradient issue and makes it easier to train extremely deep networks by allowing information to travel through some layers. Thanks to their improved gradient flow and capacity for efficient learning of intricate features, both architectures have pre-trained tasks and acquired 81% in RF and 77.68% in WW using a classification model. Along with Efficient B7 + ResNet50, 85.5625 were acquired in RF and 85.75 in WW. A.Srujan [6] proposed the CNN model for image processing to detect GI images. The collection of images ranges from 0 to 255. In the second step, generate the feature detector 3 \*3

or 7\*7 matrix and convert the entire pooled map into a single column to achieve better accuracy. Introducing the K-number of clusters rule reduces the space between the clusters and follows multiple steps, improving sensitivity, specificity, and accuracy. Yaw Afriyie [9] provides denoising capsule networks (Dn-CapsNets), a Pre-processing technique for identifying endoscopic pictures that is less complex but still effective. Activation Maps (AM) were generated by utilizing feature representations to visualize the results. These evaluations yielded the trained model's accuracy, precision, sensitivity, specificity, F1-score, and Matthew's correlation values of 94.16%, 83.1%, 86.7%, 96.1%, 86.6%, and +0.69. Comparing the proposed technique with the state-of-the-art has shown improved accuracy.

Melaku Bitew Haile [12] proposed combining the features of the VGGNet and InceptionNet networks to form a concatenated neural network model that might be used to diagnose gastrointestinal disorders. Utilizing VGGNet and InceptionNet, two trained deep convolutional neural networks, features are retrieved from the supplied endoscopic pictures. These collected characteristics are then concatenated and classified using machine learning classification techniques (Softmax, k-Nearest Neighbor, Random Forest, and Support Vector Machine). With the provided standard dataset, the Support Vector Machine (SVM) fared better than the other methods. The proposed model's classification accuracy is 98%. Table 4 shows a 97.8% correlation coefficient displacement, significantly improving over previous techniques and different neural network topologies.

**Table 4. Classifier comparison using suggested feature extraction techniques**

Techniques	Method of Noise Filtering	Classifier	Accuracy (%)
VGGNET	AMF	Softmax	93
VGGNET	AMF	KNN	93.2
VGGNET	AMF	RF	93.3
VGGNET	AMF	SVM	93.4
InceptionNet	AMF	Softmax	84
InceptionNet	AMF	KNN	84.5
InceptionNet	AMF	RF	85.7
InceptionNet	AMF	SVM	85
VGGNET - InceptionNet	AMF	Softmax	94
VGGNET - InceptionNet	AMF	KNN	95.4
VGGNET - InceptionNet	AMF	RF	95.7
VGGNET - InceptionNet	AMF	SVM	98
Pre-trained VGGNET16	AMF	-	89
Pre-trained Inceptionv3	AMF	-	91

In a series of tests based on deep convolutional neural networks, Borgli et al. (2020) utilized standard architectures with a little change to identify 23 distinct classes of pictures [13] using Pre-Trained models. The results show that the MCCs for these models are 0.826, 0.898, 0.899, and 0.902, respectively. The results do, however, also highlight the need for improvement because certain classes-such as the inability to distinguish between ulcerative colitis and esophagitis, colored lifted polyps and dyed resection margins, and barrettes from esophagitis or z-line-are harder to identify than others.

A novel method integrating deep CNN with geometric functions was presented by Sharif et al. (2019). First, an advanced technique termed enhanced contrast color features extracts the disease spots from the provided WCE pictures. The delineated area of illness provided the geometric characteristics. Thereafter, a special fusion was carried out using the Euclidean Fisher Vector. [14] of the VGG16 and VGG19 deep CNN algorithms. Once the geometric features and unique features have been combined, the best characteristics are chosen using the conditional entropy approach.

The chosen characteristics were classified by the K-Nearest Neighbor (kNN) algorithm. The proposed technique was evaluated using a privately collected collection of 5500 WCE photos, and the results showed a classification accuracy of 99.42% and a precision rate of 99.51%. But the writers achieved just three classification classesulcers, bleeding, and health. The deep learning-based computer-aided diagnostic system YOLOv4 employed by Durak et al. [24] to identify gastric polyps performed poorly, with a mean average accuracy of 87.95 percent.

F. Yasmin et al. examined the accuracy, efficiency, and detection of polyp and aberrant feature recognition for different types of algorithms and presented the GastrNet model in this study. It is created by hyperparameter fine-tuning YOLOv5 to identify certain polyps and anomalous features, especially esophagitis. Under this method, the entire image is analyzed by a single neural network, after which it is dissected into its parts, and the probability and bounding boxes for each are individually calculated. The purpose of hyperparameter fine-tuning is to increase the overall optimization of the model. A data set with one thousand individual photographs that needed to be labelled was annotated using two alternative techniques. This study employed three different backbone networks: MobileNet v2, MobileNet v2 FPN Lite, and Resnet50 v1 FPN, in addition to applying the fine-tuned SSD model. Furthermore, CSPdarknet53 was utilized in this work to develop the enhanced YOLOv4 model.

The study findings show in Table 8 that the suggested model shown in Table 5, GastroNet, achieved a high mAP (mean Average Precision), F1 score, precision with a value of

0.99, and recall with a value of 1.00 in accurately identifying polyps and aberrant traits. Physicians will greatly benefit from the research's findings in correctly identifying and diagnosing aberrant traits.

## 9. Overview of Papers Using DL & ML Techniques for all Classes

Table 5. Overview of all techniques

Author	Method	End Goal	Classifier	Accuracy/MCC
Farhana Yasmin (2023)	YOLO V5	Early Detection of Gastric Cancer	Multiclass	99%
Dev Gupta (2022)	Efficient B7 + ResNet50	Anatomical Classification of GI Images	Multiclass	88%
Yaw Afriyie, Benjamin A (2022)	Dn-capsNets	Constructed the Activation Maps (AM)	Multiclass	94.16%
Islam et al. (2021)	CPDFENET	Early Detection of Gastric Cancer	Multiclass	93.22%
Sheeraz Ahmad (2020)	YOLO V7	Early Detection of Gastric Cancer	Multiclass	72.00%
Borgli et al. (2020)	ResNet-152 + DenseNet-161+MLP	Experiments Based on GI Diseases	Multiclass	90.20%
Escobar et al. (2020)	VGG16 CNN	Detect the Gastric Abnormalities	Multiclass	94.60%
S. Siwei Chang & Liu (2020)	Denosing Capsule Network	Recognize the Complex Images	Multiclass	90.47%
W. Liu. (2016)	SSD	Early Detection of Gastric Cancer	Multiclass	91.26%
Xiong et al (2019)	Faster R-CNN	Detecting Esophageal cancer	Multiclass	92.00%
Mossotto et al. (2017)	PCA, LDA, MDS	Used Endoscopic Data for Detection	Multiclass	82.70%
Proposed method	VGGNet-InceptionNet + SVM	Classification of GI Diseases	Multiclass	98

## 10. Discussion

GI disorders, however, are varied. Other GI illnesses that should be investigated using the DL approach include intraepithelial neoplasia and invasive mucosal lesions, which are thought to represent significant early stages of cancer but have not yet been covered in the relevant literature. A huge number of labeled training data sets are needed for the DL approach. For instance, there are 1.2 million samples in the AlexNet training dataset. Getting a lot of labeled medical picture data is challenging since labeling by medical specialists is expensive and must consider patient privacy concerns.

In contrast to pictures of the skin, eyes, MR, and CT obtained from the body's surface, images of the gastrointestinal tract must be obtained through an endoscopy, which entails inserting a camera probe inside the patient's body. As a result, obtaining GI image data is more problematic, and using DL for computer-aided GI diagnosis is extremely restricted, laborious, and unproductive.

Furthermore, insufficient training data collection might result in unimpressive analysis from the transfer learning process. Transfer learning based on pre-trained models using medical pictures might produce better outcomes than directly employing pre-trained models from natural photos. This is because, compared to natural photographs, there are fewer distinctions between typical types of medical images. Finally, the suggested concatenated model was used to contrast cutting-edge classifiers.

Using an SVM classifier, our suggested concatenated model attains 98.7% training accuracy, 98.2% validation accuracy, 98% testing accuracy, and 97.8% Matthews' correlation coefficient. Inceptionv3 and VGGNet16, two cutting-edge pre-trained models, and the suggested VGGNet and InceptionNet, were all surpassed by the suggested concatenated model. Based on the research, we have developed a model that, on average, increases the prior work's Matthews' correlation coefficient by 7.6%. This shows that the proposed concatenated model might be a useful diagnostic tool for gastrointestinal disorders when combined with endoscopic images. We provide a comparison of our suggested approach with previous research.

The findings demonstrate that, compared to alternative methods, the suggested procedure yields a high degree of accuracy. Using these varied backbone networks makes it possible to thoroughly assess GastroNet's performance compared to other designs, guaranteeing that the suggested model is flexible and adjustable to a range of medical imaging circumstances. Thus, selecting the pretrained model as GastroNet's backbone network is motivated by factors such as depth, established baseline comparison, computational efficiency, and multi-scale feature extraction. These choices

support GastroNet's cutting-edge ability to identify gastrointestinal abnormalities, which makes it a valuable tool for early detection and diagnosis in gastroenterology.

## 11. Conclusion

As the study concludes, artificial intelligence—and models in particular has significantly advanced the early identification and diagnosis of digestive disorders such as gastrointestinal polyps, etc., using techniques for fine-tuning hyperparameters and cutting-edge algorithms. We suggest the following line of inquiry for further study in light of the experimental results of this investigation:

1. GI image analysis has not used many other effective networks, like RNN, Graph Neural Networks (GNN), PCANet, and Canonical Correlation Analysis Network (CCANet). The "parameter sharing" technique was not used to pre-trained models in transfer learning. Examined for the identification, categorization, and division of polyps, haemorrhages, gastrointestinal cancer, etc. Because GI disorders can take many different forms, it's crucial to identify additional rare illnesses, such as intraepithelial neoplasia and invasive mucosal lesions, as early indicators of malignancy.
2. Cancer detection and clustering are made possible with the help of discrete wave transform.
3. When utilizing computerized systems for the early and precise detection of many illnesses, segmentation is essential. In order to enhance the model's performance, we advise developing a novel simultaneous segmentation technique that uses GRABCUT for segmentation and MASK-RCNN for illness site identification.
4. In future research, it is crucial to assess various feature selection algorithms to identify the lowest subset of characteristics that can support precise categorization of intestinal illness types.

Image processing and computer vision techniques should be improved by future and continuing work in software approaches (i.e., AI) as outlined in this study; nevertheless, additional advances might not be achievable without hardware breakthroughs and the involvement of medical physicians. Software engineers will be able to develop more intelligent software systems with greater capacities and the ability to address open problems with the help of future generations of capsules, which will include higher-quality information.

## Author's Contribution

All authors contributed equally to the manuscript and typed, read, and approved the final manuscript.

## Acknowledgement

We are thankful to the editors and the anonymous reviewers for many valuable suggestions to improve this paper.

## Funding

This research is supported by the Basic Science Research Program through the National Research Foundation of Korea

(NRF), funded by the Ministry of Education (NRF- RS-2023-00237287, NRF-2021S1A5A8062526) and local government-university cooperation-based regional innovation projects (2021RIS-003).

## References

- [1] Hyuna Sung et al., “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA: A Cancer Journal for Clinicians, The Flagship Journal of the American Society*, vol. 71, no. 3, pp. 209-249, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Patsy Gerstner, “The American Society for Gastrointestinal Endoscopy: A History,” *Gastrointestinal Endoscopy*, vol. 37, no. S2, pp. S1-S26, 1991. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] Kenneth K. Wang, and Richard E. Sampliner, “Updated Guidelines 2008 For the Diagnosis, Surveillance and Therapy of Barrett’s Esophagus,” *American Journal of Gastroenterology*, vol. 103, no. 3, pp. 788-797, 2008. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Dev Gupta et al., “Classification of Endoscopic Images and Identification of Gastrointestinal Diseases,” *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, Faridabad, India, pp. 231-235, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas “Effective Voting of Heterogeneous Classifiers,” *15th European Conference on Machine Learning*, Pisa, Italy, pp. 465-476, 2004. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] A. Srujan et al., “Detecting Cancer in Gastrointestinal Images Using MATLAB,” *2021 5th International Conference on Trends in Electronics and Informatics*, Tirunelveli, India, pp. 1517-1521, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Debesh Jha et al., “A Comprehensive Analysis of Classification Methods in Gastrointestinal Endoscopy Imaging,” *Medical Image Analysis*, vol. 70, pp. 1-18, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] Pia H. Smedsrud et al., “Kvasir-Capsule, A Video Capsule Endoscopy Dataset,” *Scientific Data*, vol. 8, pp. 1-10, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Yaw Afriyie, Benjamin A. Weyori, and Alex A. Opoku, “Gastrointestinal Tract Disease Recognition Based on Denoising Capsule Network,” *Cogent Engineering*, vol. 9, no. 1, pp. 1-17, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] E. Mossotto et al, “Classification of Paediatric Inflammatory Bowel Disease Using Machine Learning,” *Scientific Reports*, vol. 7, pp. 1-10, 2017. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton, “Dynamic Routing Between Capsules,” *Advances in Neural Information Processing Systems*, vol. 30, 2017. [[Google Scholar](#)] [[Publisher Link](#)]
- [12] Melaku Bitew Haile et al., “Detection and Classification of Gastrointestinal Disease Using Convolutional Neural Network and SVM,” *Cogent Engineering*, vol. 9, no. 1, pp. 1-23, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [13] Hanna Borgli et al., “HyperKvasir, A Comprehensive Multi-Class Image and Video Dataset for Gastrointestinal Endoscopy,” *Scientific Data*, vol. 7, no. 1, pp. 1-14, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Muhammad Sharif et al., “Deep CNN and Geometric Features-Based Gastrointestinal Tract Diseases Detection and Classification from Wireless Capsule Endoscopy Images,” *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 33, no. 4, pp. 577-599, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Yi Wang et al., “Polyp-Alert: Near Real-Time Feedback During Colonoscopy,” *Computer Methods and Programs in Biomedicine*, vol. 120, no. 3, pp. 164-179, 2015. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Baopu Li, and Max Q.-H. Meng, “Tumor Recognition in Wireless Capsule Endoscopy Images Using Textural Features and SVM-Based Feature Selection,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 3, pp. 323-329, 2012. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Franco Scarselli et al., “The Graph Neural Network Model,” *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61-80, 2009. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Tsung-Han Chan et al., “PCANet: A Simple Deep Learning Baseline for Image Classification?,” *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017-5032, 2015. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Xinghao Yang et al., “Canonical Correlation Analysis Networks for Two-View Image Recognition,” *Information Sciences*, vol. 385-386, pp. 338-352, 2017. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Simon Andermatt, Simon Pezold, and Philippe Cattin, “Multi-Dimensional Gated Recurrent Units for The Segmentation of Biomedical 3D-Data,” *Deep Learning and Data Labeling for Medical Applications*, Athens, Greece, pp. 142-151, 2016. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] Y. Bengio, P. Simard, and P. Frasconi, “Learning Long-Term Dependencies with Gradient Descent Is Difficult,” *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157-166, 1994. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [22] Xinting Gao, Stephen Lin, and Tien Yin Wong, “Automatic Feature Learning to Grade Nuclear Cataracts Based on Deep Learning,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 11, pp. 2693-2701, 2015. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [23] Jie Zhou et al., “Graph Neural Networks: A Review of Methods and Applications,” *AI Open*, vol. 1, pp. 57-81, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [24] Serdar Durak et al., “Deep Neural Network Approaches for Detecting Gastric Polyps in Endoscopic Images,” *Medical and Biological Engineering and Computing*, vol. 59, pp. 1563-1574, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [25] Pradipta Sasmal et al., “An Automated Framework for Detection, Localization, and Classification of Colonic Polyp Using Deep Learning,” *Research Square*, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [26] Tao Yan et al., “Semantic Segmentation of Gastric Polyps in Endoscopic Images Based on Convolutional Neural Networks and an Integrated Evaluation Approach,” *Bioengineering*, vol. 10, no. 7, pp. 1-19, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [27] Iria Varela-Rey et al., “Design and Biopharmaceutical Preclinical Characterisation of a New Thermosensitive Hydrogel for The Removal of Gastric Polyps,” *International Journal of Pharmaceutics*, vol. 635, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [28] Sara Massironi et al., “Exploring the Spectrum of Incidental Gastric Polyps in Autoimmune Gastritis,” *Digestive and Liver Disease*, vol. 55, no. 9, pp. 1201-1207, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [29] Junjun Huang et al., “A Masked Graph Neural Network Model for Real-Time Gastric Polyp Detection in Healthcare 4.0,” *Journal of Industrial Information Integration*, vol. 34, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [30] Farhana Yasmin et al., “GastroNet: Gastrointestinal Polyp and Abnormal Feature Detection and Classification with Deep Learning Approach,” *IEEE Access*, vol. 11, pp. 97605-97624, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [31] Mark Sandler et al., “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 4510-4520, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [32] Tsung-Yi Lin et al., “Focal Loss for Dense Object Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318-327, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]