International Research Journal of Multidisciplinary Scope (IRJMS), 2025; 6(3): 1743-1756

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2025.v06i03.04664

# Deep Learning-Based Prediction Models for Colorectal Disease: Enhancing Diagnostic Accuracy

Parimala S[1]*, S Nalini Poornima[2], Sangita Gautam Lade[3], N Venkatesh[4], Appasami G[5]

[1]Department of Computer Science and Applications, SRM Institute of Science and Technology, Vadapalani ,Chennai, India, [2]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai -89, India, [3]Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, India, [4]School of CS and AI-CSE, S R University, Warangal-506371, India, [5]Department of Computer Applications, National Institute of Technology Tiruchirappalli, Tamil Nadu, India. *Corresponding Author's Email: parimalaassistantprofessor@gmail.com

## Abstract

The presentation of colorectal diseases based on deep learning has attracted much research since it offers the development of correct and machine-driven diagnosis. The current study advents a new DenseNet-121 model with Squeeze-and-Excitation (SE) blocks that can enjoy better feature extraction and classification into colon diseases. The sample has four classes and 1,500 images each. In order to enhance the robustness of the models, the preprocessing pipeline is done completely, involving resizing of images, Gaussian smoothing, and Otsu thresholding. DenseNet-121 model incorporating SE blocks is trained on an 80:20 training-testing split with data augmentation being used to deal with over-fitting. The performance assessment is performed as precision, recall, F1-score, and inference time in comparison to the classical architecture of ResNet-50 and VGG-16 with DenseNet-121. The results obtained on the experimental show that the proposed model scored 96.3 percent accuracy, 97.1 percent precision, 95.7 percent recall, and 96.4 percent F1-score with better performance on the colorectal disease category than the baseline models. Also, dimensionality is reduced because of the presence of Global Average Pooling (GAP), and the model is made more discriminative thanks to SE blocks that allow the recalibration of features. The analysis of confusion matrix proves high classification reliability indicating that there is minimal misclassification among the disease categories. The present article reveals the case of successful deep learning application in the analysis of endoscopy images and opens up the possibility of real-time application in clinical practice in the form of a computer-aided diagnosis system.

**Keywords:** Colorectal Disease, Computer-Aided Diagnosis System, Deep Learning, DenseNet-121, Global Average Pooling, Image Segmentation.

## Introduction

The growing trend of colorectal diseases becomes one of the challenges in the sphere of contemporary healthcare due to the rising cases and high risk of developing colorectal cancer (CRC) without significant indications (1). An early and correct diagnosis of colorectal diseases is necessary so that effective measures can be taken to support the prognosis and increase the survival rate. Nevertheless, conventional methods of diagnosis including colonoscopy and biopsy are known to take a lot of time, are dependent on operators, and prone to interobserver variability among radiologists and gastroenterologists (2). The evolution of Wireless Capsule Endoscopy (WCE) revolutionized gastro-intestinal imaging procedure, as it provides a non-invasive examination of small intestine and colon. In contrast to conventional colonoscopy, WCE is the high-resolution internal examination that does not involve sedation and becomes one of the most interested diagnostic methods of colorectal disorders (3). However, WCE produces heavy video information volume and large manual viewing volumes by physicians, which can cause diagnostic delays and the high-volume problem. As a result, a need to implement the AI-based cross-site medical technology to analyze the WCE images has appeared. Deep learning has obtained the status of the transformational technology in computer-aided diagnosis (CAD) specifically in the domain of medical imaging. The use of CNNs has achieved a higher performance in classification of images, feature extraction, and recognition of patterns (4). Among the architectures that one



(Received 26th March 2025; Accepted 17th July 2025; Published 30th July 2025)

could consider, it is recognized that DenseNet-121 has dense connections that make the flow of gradients smooth, eliminate unnecessary features and increase the reuse of features. To further enhance its performance in analyzing medical images, Squeeze and Excitation (SE) Blocks have been suggested to be integrated into it. These blocks optimise feature representations by making them scale adaptively on a recalibration basis, and enable the model to concentrate on disease relevant regions of WCE images. The detection of this integration has been revealed to largely increase the diagnostic accuracy, stability, and transference (5, 6).

Despite the diverse deep learning architecture being used in the diagnosis of gastrointestinal diseases, over fitting, inefficient and uninterpretable properties have been widely discussed in previous literature (7, 8). Also, colorectal diseases are difficult to classify because nearly all the classes have a high similarity and thus distinguishing conditions like ulcerative colitis and esophagitis, some abnormal appearances, is challenging due to subtle texture and morphological manifestation in these conditions. Such distinctions are common problems of traditional CNNs. This study suggests a more advanced DenseNet-121 model with SE Blocks to apply in the automatic classification of colorectal diseases. The architectural enhancement is what is new, and it operates a superior mechanism of feature recalibration mechanisms that are applicable in identifying the disease. Although publicly available datasets were used, the contribution is mostly architectural and devoted to attention mechanisms. This strategy has the benefits of providing strong benefits in feature isolation with the largest activations by matching disease prone areas and reducing noise by low strength activations. Also, the architecture contributes to enhanced generalizability through minimization of feature duplication as well as raising the model discriminative ability, which results in better classification. In terms of efficiency in parameter pricing, the DenseNet-121 would be appropriate in real-time uses in the medical world than complex architectures.

The study suggests an automated deep learning system in detection of colorectal disease that would be more accurate and clinically practicable. To facilitate exact classification, a DenseNet-121 architecture including SE Block has been developed. Comparisons to end-to-end state-of-the-art architectures have been performed, the model was shown to be interpretable using attention mechanisms and the model was proven to be efficient enough in terms of computation to deploy in clinical settings. The ability of such an improved architecture in carrying out learning cost-free activities (automated classification) is examined through applicable datasets. The proposed model should enhance diagnostics and decision making in clinical practice by far, through the mechanisms of adaptive feature recalibration and dense connections. The present study findings can contribute to the formulation of an AI-based diagnostics tool that could assist in the early detection and prognosis of colorectal diseases.

The rest in the manuscript is organized as follows: in Section 2, the current experience in deep learning models applied to the diagnosis of colorectal diseases is evaluated, and the limitations of this experience are stated. Section 3 presents the suggested architecture DenseNet-121 + SE Block, provides an overview of preprocessing techniques and presents information on the training approach. The quantitative and qualitative outcomes, performance indicators, comparisons with the existing models, result interpretation, clinical impressions, and challenges are noted in section 4. Section 5 packs up the significant contributions and gives future research directions. Innovations in CRC classification have had to do with issues of optimization since accuracy is a paramount element. Although progress has been made, even more traditional CNNs have failed to follow traditional performance expectations in diagnosis. In the previous study, histopathology slides of gastric cancer were trained and tested on the LC25000 dataset using Vision Transformer (ViT) and Swin Transformer and ResNet34 and EfficientNet34 were used as control variants (9). The modified Swin Transformer was one of them, and its classification accuracy amounted to 99.80%, signifying a high effectiveness level. Although ViT and Swin Transformer have a high degree of accuracy, some difficulties concerning practical realization have been noted. The emphasis has been also made on effects of these models on clinical decisions.

Medical image analysis on ML and DL have allowed coming up with sophisticated disease detection

algorithms early and accurately. There are four significant categories of colon diseases problem that has been solved and examined as a full range of endoscopic images that consist of nearly 100.000 images (10). Training A model developed based on its Adam optimizer reached 93.65 percent accuracy in validation, making it applicable in early diagnosis.

Massive datasets, such as PLCO screening trial (640 CRC patients) as well as TCGA data set (522 CRC patients), have been used to train and perform classification and risk estimation duties (11). In the colorectal polyp detection and detection feature learning, deep learning models have proven their usefulness, resulting in the enhancement of the level of detection. Research has also been done on the Xception+ model, which has been reported as an excellent option with a percentage accuracy of 99.37 and GoogLeNet that had a little more precision on the cancer tissue categorization (12). The traditional diagnostic methods have been identified as being expensive, time-consuming, and they have restraints because of the amount of data and class imbalance. Image sets of 224 by 224 pixels in size were resized and run in ShuffleNet and ResNet-50 in the detection of polyp, delivering an accuracy of 98.36% which points to an opportunity to avoid CRC deaths (13). Photo augmentation and transfer learning methods were used based on KVASIR database providing 8,000 GI-endoscopic images, with the result of pathological cases finding to the extent of 96.89 percent.

Other experiments were conducted using 100,000 non-overlapping patches of histology images of the CRC dataset used to determine the efficiency of CNNs-based architectures, such as DenseNet201, InceptionResNetV2, VGG16, VGG19, and Xception. Normalization and augmentation methods of data allowed proper survival prediction and classification of histology patterns based on prognosis (14). The other study featured the utilization of 3D CNN models to distinguish between colon carcinoma and acute diverticulitis pathology and in surgical patients on the basis of CT images (15). With the aid of an AI, sensitivity and specificity in diagnostics actually increased, which was proven in the study conducted with 10 observers with different degrees of expertise. Other experiments under CNN models checked differences in sizes of test set (20%, 30%, and 40%), and the classification performance (16, 17). The results supported the clinical relevance of deep learning models since they were very accurate and performed consistently. Also, to enhance the speed of computing in histopathological screening, an approach like low resolution processing as well as parameter-reduced personal CNNs was developed, which yielded the same accuracy (99.4 percent) but half the computing burden (18).

Early screening of colorectal diseases is important, as there have been medical concerns of their conversion into CRC. Other invasive instruments like the colonoscopy are likely to cause interobserver variability. WCE is available as a non-invasive alternative, but due to the large volume of data it is associated with delays in the diagnosis. To deal with it, an AI diagnostic model with DenseNet-121 with SE Blocks is suggested to increase accuracy in classification owing to paid attention to more significant elements. The proposed model will use the benefits of deep learning in handling medical images in order to maximize on the model accuracy, computation, and the real-time accessibility. Being more balanced in performance and complexity compared to the classical instance of CNNs and more modern ones (ViT, Swin Transformer), the model proves superior to both. The clinical significance of its evaluation is confirmed through the analysis with benchmark datasets and the suggested system can assist in the diagnostic of colorectal disease supported by AI.

## Methodology

### Dataset Description

The WCE Curated Colon Disease Dataset is the primary dataset used in this study, offering a comprehensive collection of endoscopic images specifically designed for the classification of colorectal diseases. The dataset is comprised of 4 distinct classes, each representing a different pathological condition of the colon, with 1,500 images per class. These classes include Normal, Ulcerative Colitis, Polyps, and Esophagitis and an illustration of the dataset images, categorized as Normal (A), Polyps (B), Ulcer (C), and Esophagitis (D), is provided in Figure 1. The dataset comprises four balanced classes: Normal, Ulcerative Colitis, Polyps, and Esophagitis, with 1,500 images per class. Therefore, class imbalance was not a major

concern. However, data augmentation techniques were applied to improve generalization and reduce overfitting.

Each class in the dataset is carefully labeled, offering clear differentiation between various disease types, which is essential for the development and evaluation of machine learning models. However, this dataset also presents several challenges for model training, including significant intra-class variations (e.g., differing appearances of Polyps or Ulcerative Colitis in different individuals) and inter-class similarities (e.g., normal colon tissue and mild inflammation in Esophagitis could appear similar). Furthermore, variations in lighting conditions, image resolution, and the complex nature of the disease manifestations make the dataset difficult to process effectively, emphasizing the need for a robust model that can handle these challenges.

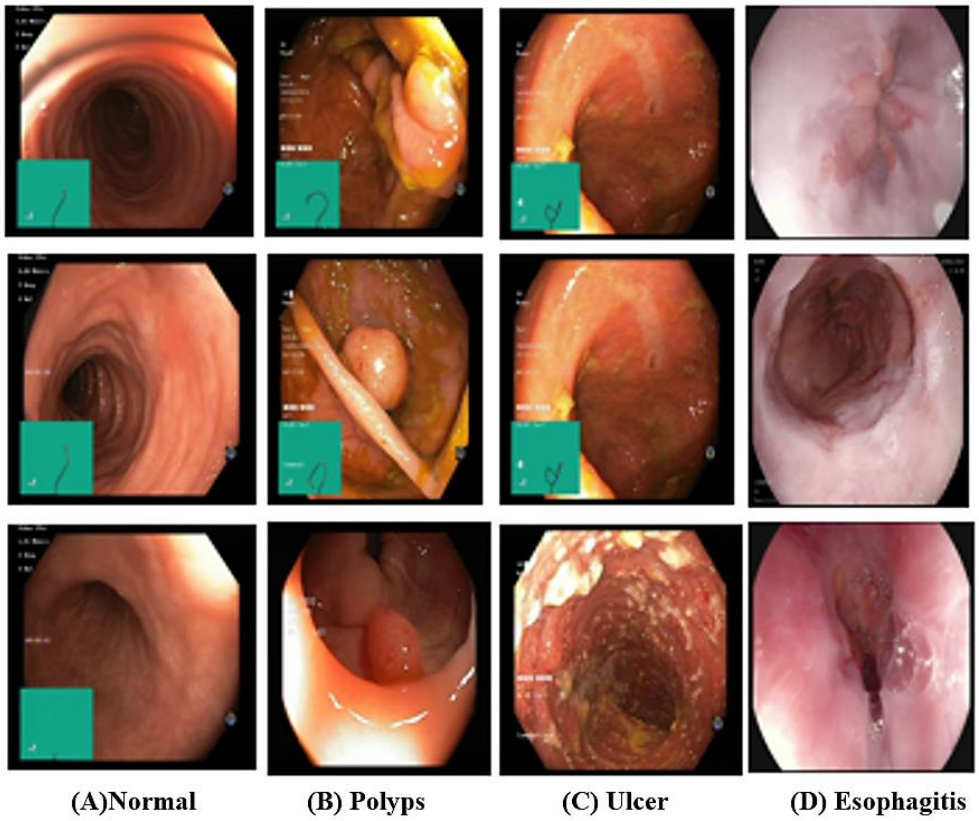


**Figure 1:** Shows the Dataset Images (A) Normal, (B) Polyps, (C) Ulcer, (D) Esophagitis

The dataset is divided into three subsets for training, validation, and testing purposes. This division ensures that the model is evaluated on unseen data, which helps assess its ability to generalize. The Table 1 summarizes the dataset distribution.

**Table 1:** Dataset Distribution

| Class | Number of Images per Class | Training Set | Validation Set | Testing Set |
|---|---|---|---|---|
| 0 - Normal | 1,500 | 800 | 500 | 200 |
| 1 - Ulcerative Colitis | 1,500 | 800 | 500 | 200 |
| 2 - Polyps | 1,500 | 800 | 500 | 200 |
| 3 - Esophagitis | 1,500 | 800 | 500 | 200 |

## Preprocessing Techniques

To ensure the DenseNet-121 with SE Blocks model effectively classifies colorectal diseases, key preprocessing steps are essential. The selected techniques—Image Resizing, Image Denoising using Gaussian Filter, and Segmentation using Otsu's Thresholding—help standardize input dimensions, reduce noise, and highlight relevant regions. Below is a detailed explanation of these preprocessing steps with corresponding formulas and equations.

**Image Resizing:** The dataset consists of endoscopic images that vary in size, which can lead to inconsistencies during deep learning model training. DenseNet-121 requires a fixed input size of 224 × 224 pixels (19). Therefore, image resizing is applied to normalize dimensions while preserving essential features.

To resize an image I from its original size (H,W) to (H′,W′) bilinear interpolation is used and modelled in Eq. [1]:

$$I'(x,y) = \sum_{i=0}^{1} \sum_{j=0}^{1} w_{i,j} I(x_i, y_j) \quad [1]$$

Where $I'(x,y)$ is the interpolated pixel value in the resized image. $I(x_i, y_j)$ represents the pixel values in the original image and $w_{i,j}$ are the interpolation weights, calculated using Eq. [2]:

$$w_{i,j} = (1 - | x - x_i |)(1 - | y - y_j |) \quad [2]$$

This ensures a smooth transition between pixel values, avoiding distortion or loss of important features.

**Image Denoising using Gaussian Filter**: Colorectal endoscopic images often contain noise due to sensor limitations, poor lighting, or reflections, which can obscure disease regions (20). Gaussian filtering is applied to remove high-frequency noise while preserving important structures.

A Gaussian filter convolves the image with a Gaussian kernel G(x,y) showing in Eq. [3]:

$$I_{filtered}(x,y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} G(i,j) I(x-i, y-j) \quad [3]$$

Where $I_{filtered}(x,y)$ is the denoised pixel value, $I(x-i, y-j)$ are neighboring pixel intensities and G(i, j) is the Gaussian kernel is given as Eq. [4]:

$$G(i,j) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{i^2+j^2}{2\pi\sigma^2}\right) \quad [4]$$

Where σ controls the spread of the Gaussian function, k determines the kernel size, typically k=3or k=5. A higher σ value results in stronger blurring, while a smaller σ preserves more details. Figure 2 shows (A) Original Image (B) Denoised Image (C) Segmented Image.

**Segmentation Using Otsu's Thresholding:** After denoising, Otsu's Thresholding is applied to segment disease regions by separating foreground (disease) from the background. This is particularly useful for highlighting ulcerative colitis, polyps, and esophagitis (21).

Otsu's method finds an optimal threshold τ\tauτ that maximizes the variance between two classes:

- Background ($\omega_1$)
- Foreground ($\omega_2$ )

The threshold τ is computed as Eq. [5]:

$$\tau = arg \ \ \omega_1(\tau)\sigma_1^2(\tau) + \omega_2(\tau)\sigma_2^2(\tau)] \quad [5]$$

Where $\omega_1(\tau)$ and $\omega_2(\tau)$ are the probabilities of background and foreground pixels, $\sigma_1^2(\tau) and\ \sigma_2^2(\tau)$ are their variances.

Steps in Otsu's Method

- Compute Histogram: Determine the intensity distribution of the image.
- Calculate Between-Class Variance for different threshold values.
- Select Optimal Threshold (τ) that maximizes variance.
- Segment Image by converting it to a binary mask using Eq. [6]:

$$I_{seg}(x,y) = \{1, if\ I(x,y) > \tau\ (Foreground)\ 0, if\ I(x,y) \leq \tau\ (Background) \quad [6]$$

The preprocessing pipeline includes Image Resizing to standardize image dimensions, ensuring uniformity across the dataset. Gaussian Filtering is applied to remove noise while

preserving important structural details. Otsu's Thresholding is then used for effective segmentation, enhancing the visibility of diseased regions. These techniques ensure high-quality input data, facilitating better feature extraction and model performance. Overall, the pipeline optimizes the images for more accurate and efficient deep learning-based disease classification. These techniques prepare the Dataset for DenseNet-121 with SE Blocks, ensuring better feature extraction, higher classification accuracy, and improved generalization in colorectal disease prediction.

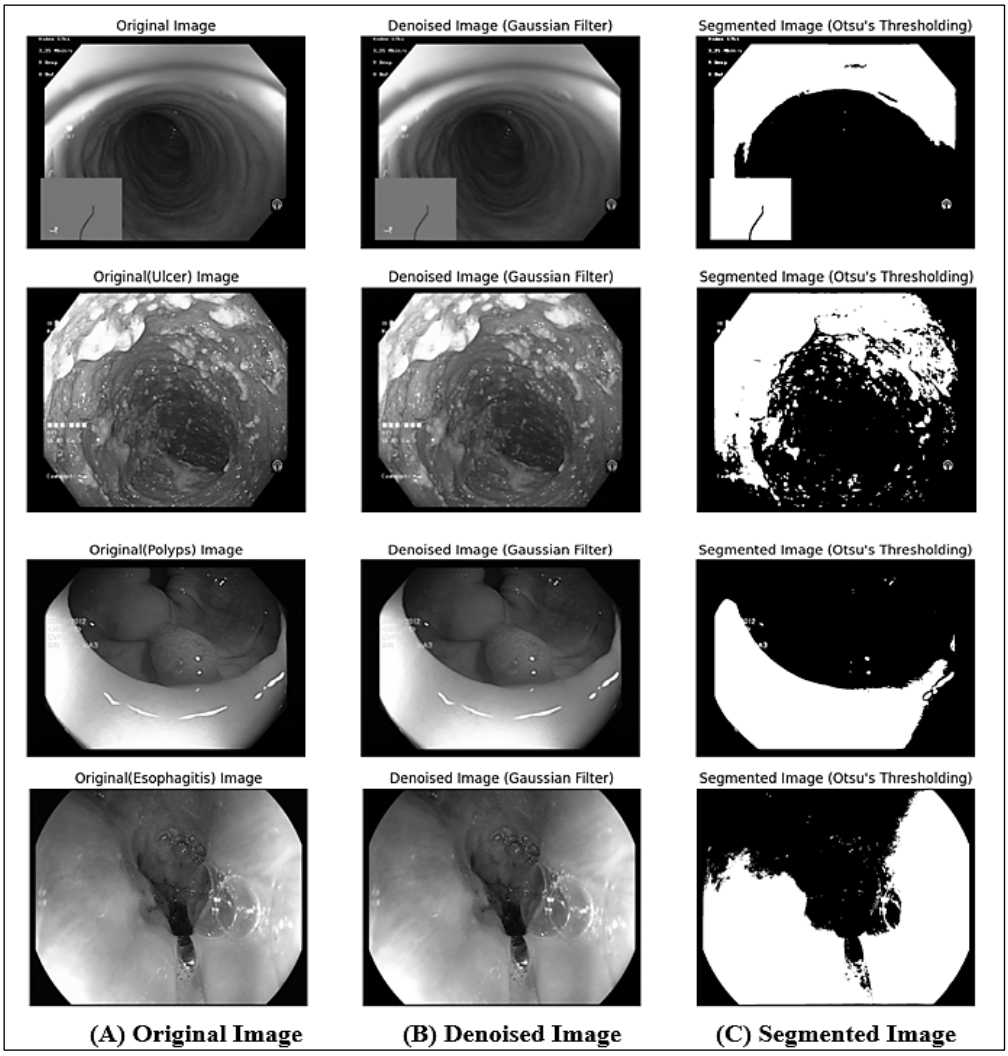


**Figure 2:** (A) Original Image, (B) Denoised Image, (C) Segmented Image

## Feature Extraction for Colorectal Disease Prediction

Feature extraction plays a crucial role in enhancing the classification performance of DenseNet-121 with Squeeze-and-Excitation (SE) Blocks for colorectal disease detection. In this study, we extract spatial, texture, and deep hierarchical features from the Dataset to effectively distinguish between different classes**.** The feature extraction process involves convolutional feature learning, global average pooling (GAP), and Squeeze-and-Excitation recalibration for disease classification.

## Convolutional Feature Learning in DenseNet-121

The dense connectivity of DenseNet-121 ensures that the feature maps from each layer are used by subsequent layers. Let's consider the feature map and how it depends on all previous layers. The output at each layer is computed as shown in Eq. [7]:

$$X^{(l)} = F(X^{(l-1)}, W^{(l)}) = ReLU(W^{(l)} * [X^{(l-1)}, X^{(l-2)}, \dots, X^{(0)}] + b^{(l)}) \quad [7]$$

The concatenation $*$ $[X^{(l-1)}, X^{(l-2)}, \ldots, X^{(0)}]$involves all the feature maps from the previous layers.

By concatenating all feature maps ups, DenseNet-121 leverages a feature reuse mechanism**,** making it easier for the model to recognize complex patterns in the medical images. In conventional CNNs, each layer learns features from scratch. However, DenseNet-121 takes a different approach, where each layer reuses features from all previous layers. This makes the model more parameter-efficient, as the need for redundant feature learning is eliminated. The feature maps at each layer are concatenated, creating a wider feature map. The concatenation operation at layer lll is as follows in Eq. [8]:

$$X^{(l)} = [X^{(l-1)}, X^{(l-2)}, \ldots, X^{(0)}] + b^{(l)}) \ (concatenation\ of\ feature\ maps) \qquad [8]$$

This feature reuse leads to more discriminative features that help detect specific patterns in the Dataset**,** such as the irregular texture of polyps or the inflammation seen in ulcerative colitis. The reuse of features leads to significant improvements in classification accuracy. Figure 3 shows the architecture of DenseNet-121.

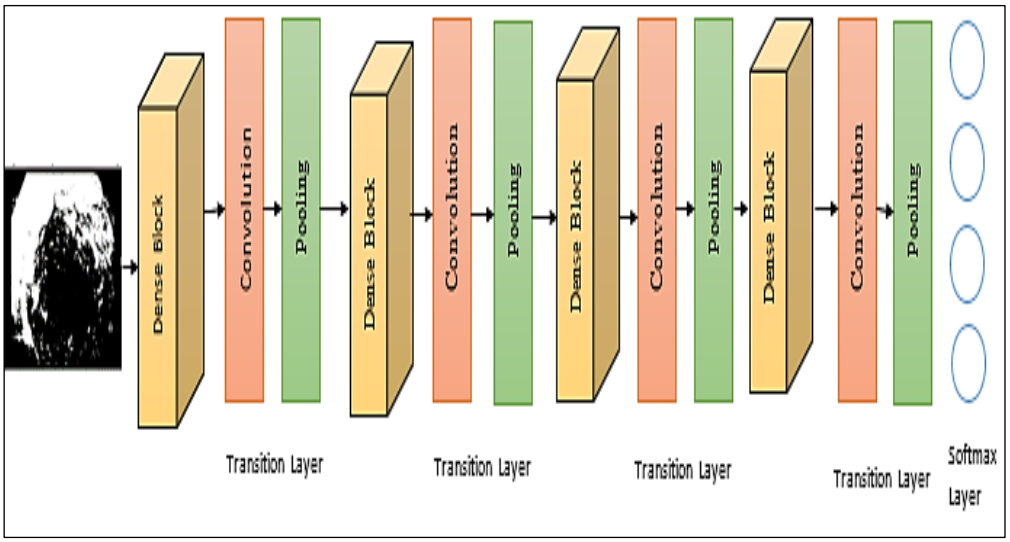


**Figure 3:** Architecture of DenseNet-121

Due to the dense connectivity, each layer has access to gradients from all previous layers, allowing for more efficient gradient flow and better weight updates.

The gradients for each layer l during backpropagation can be computed as Eq. [9]:

$$\frac{\partial L}{\partial W^{(l)}} = \sum_{k=1}^{L} \frac{\partial L}{\partial W^{(k)}} \frac{\partial X}{\partial W^{(l)}} \qquad [9]$$

The gradients from all subsequent layers contribute to the backpropagation process, ensuring efficient weight updates**.** This characteristic is particularly beneficial for medical image analysis, where deeper models are often required to capture intricate features but can suffer from gradients diminishing as the network depth increases**.** DenseNet-121 excels in extracting high-level features that are essential for accurate classification, especially in the context of detecting diseases like polyps or ulcerative colitis in colorectal disease images**.** These features capture both low-level edge information (such as boundaries of lesions) and high-level contextual information (such as shapes or textures that correspond to different disease categories). Mathematically, DenseNet-121's feature extraction ability can be represented as in Eq. [10]:

$$F_{extracted} = DenseNet - 121(I) = F(I, W) \qquad [10]$$

Where, I is the input image (e.g., colonoscopy image), W are the learned parameters of DenseNet-121. By integrating features from all preceding layers, DenseNet-121 produces highly discriminative features that can be directly used

for disease classification, improving the model's accuracy and robustness.

## Global Average Pooling (GAP) for Dimensionality Reduction

A potent method frequently employed in deep learning models; global average pooling (GAP) is especially useful for convolutional neural networks' (CNNs') dimensionality reduction. GAP drastically lowers the amount of network parameters while preserving the crucial characteristics that were taken from the input data, in contrast to conventional fully linked layers. In applications like disease classification from medical photos, this helps avoid overfitting and guarantees that the model is both computationally efficient and useful in identifying significant patterns.

The primary goal of GAP is to convert the spatial dimensions of a feature map into a single value per feature map. This process of dimensionality reduction helps in aggregating the information over the entire spatial domain and reduces the complexity of the network. GAP is particularly advantageous because it produces a fixed-length output, independent of the input image size, making it highly suitable for classification tasks. Instead of flattening the entire feature map (which leads to a large number of parameters), GAP computes a single scalar value for each channel by averaging the values in that channel. This significantly reduces the number of parameters and retains essential information, allowing the model to focus on the most important features shown in Eq. [11].

$$GAP(F) = \left[\frac{1}{H\ X\ W}\ \sum_{i=1}^{H}\ \sum_{j=1}^{W}\ F(i,j,c)\right]_{c=1}^{C} \quad [11]$$

Where F(i,j,c) denotes the value at position (i,j) in the c-th feature map of the feature map F. GAP plays a crucial role in deep learning architectures by efficiently reducing the spatial dimensionality of feature maps, resulting in fewer parameters and lower computational complexity. This reduction leads to faster training times, lower memory usage, and a decreased risk of overfitting, particularly in models with limited training data. Unlike traditional fully connected layers, which require a large number of parameters, GAP eliminates excessive weights, improving model generalization and minimizing overfitting. Additionally, GAP provides invariance to spatial resolution by aggregating entire feature maps into single values per channel, allowing the model to handle input images of varying sizes effectively. This property enhances the robustness of the model, making it more adaptable to real-world medical imaging scenarios where endoscopic images may differ in resolution and scale.

Global Average Pooling (GAP) plays an essential role in modern CNN architectures like DenseNet-121 by reducing the dimensionality of feature maps and focusing on aggregated feature information**.** This operation ensures that the model remains computationally efficient**,** robust to spatial variations**,** and less prone to overfitting**,** which is critical for accurate and reliable disease classification in medical imaging applications. The following algorithm gives step-by-step guide for performing GAP.

**Algorithm:** Global Average Pooling (GAP)

- **Input:** A feature map F of dimensions H×W×C.
- **For each channel c:**
  Calculate the average of all values in the c-th feature map:

$$avg(F_c) = \left[\frac{1}{H\ X\ W}\ \sum_{i=1}^{H}\ \sum_{j=1}^{W}\ F(i,j,c)\right]$$

- **Output:** A vector of size C is obtained, with each entry corresponding to the average value of a feature map computed across the spatial domain.

## Squeeze-and-Excitation (SE) Feature Recalibration

Squeeze-and-Excitation (SE) is a powerful mechanism designed to enhance the representational capacity of neural networks by dynamically recalibrating channel-wise feature responses and it is shown in Figure 4. It has been demonstrated that such technique can vastly

enhance network performance, as it allows concentrating on the most informative features and suppressing the irrelevant and redundant ones (22). Being a computationally efficient and lightweight technique, the Squeeze-and-Excitation (SE) block can be easily incorporated into any convolutional neural network (CNN) and be applied to any task, including the often challenging and complicated task of medical image classification in DenseNet-121.

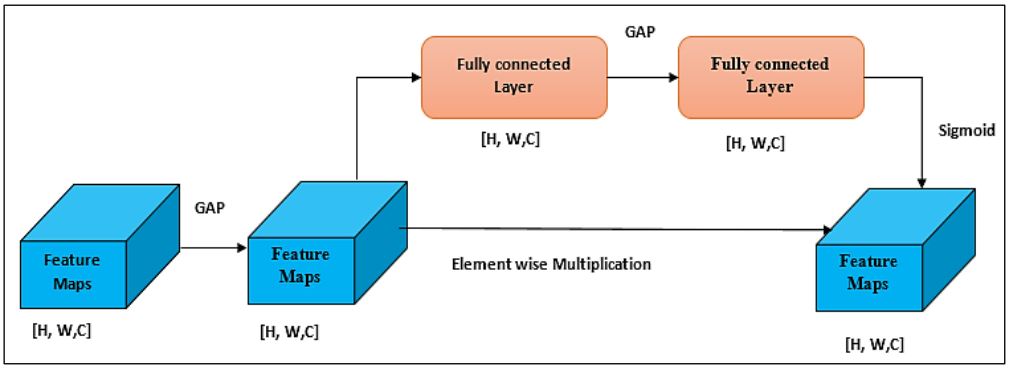


**Figure 4:** Block Diagram of Squeeze-and-Excitation (SE) Feature Recalibration

The goal of SE is to address the challenge of feature redundancy and selective emphasis within feature maps, particularly when the network has many layers or channels. SE achieves this by introducing a recalibration mechanism that can selectively weigh the importance of each channel in the feature map. By doing so, SE allows the model to focus on more informative features while suppressing noise and irrelevant details.

This recalibration is achieved through a two-step process:

**Squeeze**: The feature map is globally summarized into a channel-wise descriptor.

**Excitation**: These descriptors are used to generate a set of channel-wise attention weights, which are then used to recalibrate the feature map.

The squeeze operation performs global average pooling (GAP) on each channel, which aggregates spatial information into a single scalar for each feature map. This is done by averaging over all spatial locations are founded using Eq. [12]:

$$z_c = \frac{1}{H\,X\,W} \sum_{i=1}^{H} \sum_{j=1}^{W} F(i,j,c)\ for\ each\ channel\ c = 1,2,\dots,C \qquad [12]$$

The excitation step involves passing the squeezed feature vector z through two Fully Connected (FC) layers with ReLU and sigmoid activations, respectively. The goal is to learn the channel-wise attention weights that allow the model to emphasize the most informative channels. The excitation step can be formulated as Eq. [13]:

$$s = \sigma(W_2 \delta(W_1 z)) \qquad [13]$$

Where δ denotes the ReLU activation, $W_1$ and $W_2$ are learned weights. σ represents the sigmoid activation, which ensures that the output lies in the range [0, 1]. The vector $s \in R^C$ represents the channel-wise attention weights, where each element $s_c$ corresponds to the importance of channel c. The learned attention weights shown in Eq. [14]:

$$\hat{F}(i,j,c) = F(i,j,c) \cdot s_c \qquad [14]$$

This recalibration step ensures that more important features (channels) are enhanced while less relevant ones are suppressed. Despite its performance boost, the SE block is computationally efficient and does not add significant overhead. The number of parameters added by SE is relatively small, as it only involves two fully connected layers with the number of channels C as input and output. This makes SE highly suitable for resource-constrained environments, such as medical imaging tasks where real-time performance is crucial. By emphasizing informative channels and suppressing less important ones, SE improves the

discriminative power of the features. This is particularly beneficial for complex tasks, such as disease classification from medical images, where subtle differences between classes need to be captured by the network. SE can be seamlessly integrated into any CNN architecture. By adding SE blocks to architectures like DenseNet-121**,** it enhances the feature extraction capabilities without requiring major modifications to the original network. This allows for easy adaptation and improvement of existing models. The following algorithm shows the squeeze-and-Excitation (SE) process (23).

Algorithm for Squeeze-and-Excitation (SE) block:

**Input:** Feature map $F \in R^{H \times W \times C}$

- **Squeeze Step:**
  - Apply **global average pooling (GAP)** on the feature map to generate the channel descriptor vector $z \in R^{C}$.
- **Excitation Step:**
  - Pass z through two fully connected layers with ReLU and sigmoid activations to generate the attention vector $s \in R^{C}$.
- **Recalibration:**
  - Multiply each channel ccc in the feature map F by its corresponding attention weight $s_c$ to obtain the recalibrated feature map $\hat{F}$.

**Output:** The recalibrated feature map $\hat{F}$.

In the context of colorectal disease classification, the SE block enhances the model's ability to focus on the most discriminative features in medical images, such as polyps or inflammation areas in colonoscopy images. By recalibrating the channels to emphasize relevant features (e.g., areas with disease signs), SE improves the classification accuracy, enabling better detection of disease categories such as ulcerative colitis**,** polyps**,** esophagitis**,** and healthy tissue**.** DenseNet-121 architecture with ReLU activation and SE blocks was used. Dropout layers with a 0.3 rate were included after dense layers. This information will be added with a visual representation of the architecture in the Methods section.

Squeeze-and-Excitation (SE) offers a powerful and efficient mechanism for recalibrating the importance of channels within a CNN. By focusing on the most important features, SE enhances the discriminative power of the model, improving its performance in complex tasks like medical image classification. Its integration into DenseNet-121 can significantly enhance diagnostic accuracy for tasks like colorectal disease detection, leading to better clinical outcomes. The feature extraction process using DenseNet-121 with SE Blocks ensures that the model effectively captures spatial, texture, and hierarchical deep features crucial for colorectal disease classification**.** The integration of GAP and SE Blocks improves the model's ability to focus on informative disease patterns, ultimately leading to higher diagnostic accuracy.

# Results and Discussion

## Experimental Setup

The colorectal disease classification model was developed using the Dataset, which is divided into four categories. The dataset was divided into training (80%) and testing (20%) subsets. The model training was conducted with a batch size of 32 for 50 epochs, utilizing the Adam optimizer and a learning rate of 0.0001. Data augmentation techniques were applied to enhance model generalization and minimize overfitting. The training process was performed on an NVIDIA GPU (RTX 2080 Ti) to ensure high-speed computations. This setup was designed to achieve high performance in classifying colorectal diseases and was compared to other established models like ResNet-50 and VGG-16 to evaluate the impact of the DenseNet-121 with SE blocks architecture.

## Performance Metrics

We compared the performance of the DenseNet-121 with SE blocks model with other traditional architectures like ResNet-50 and VGG-16 to understand the impact of DenseNet-121's dense connectivity and the SE block's feature recalibration on colorectal disease classification. As shown in Figure 6, Accuracy (Acc) in the work suggested is the percentage of correctly identified occurrences among all predictions, indicating the

model's overall efficacy.a. As shown in Figure 5, Precision (Pre) ensures a lower false positive rate by measuring the ratio of accurately predicted positive instances to the total anticipated positives.a. As seen in Figure 5B, Recall (Rec) calculates the ratio of true positives to the sum of true positives and false negatives, hence assessing the model's capacity to detect all genuine positive cases. By taking into account both precision and recall, the F1-Score (F1), which is shown in Figure 6B, offers a balanced metric and a thorough assessment of categorization performance.

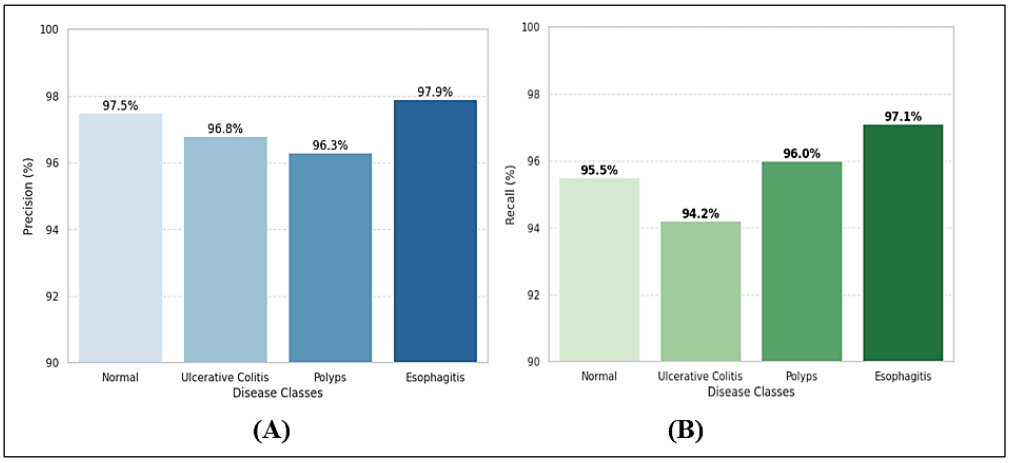


**Figure 5:** (A) Precision, (B) Recall

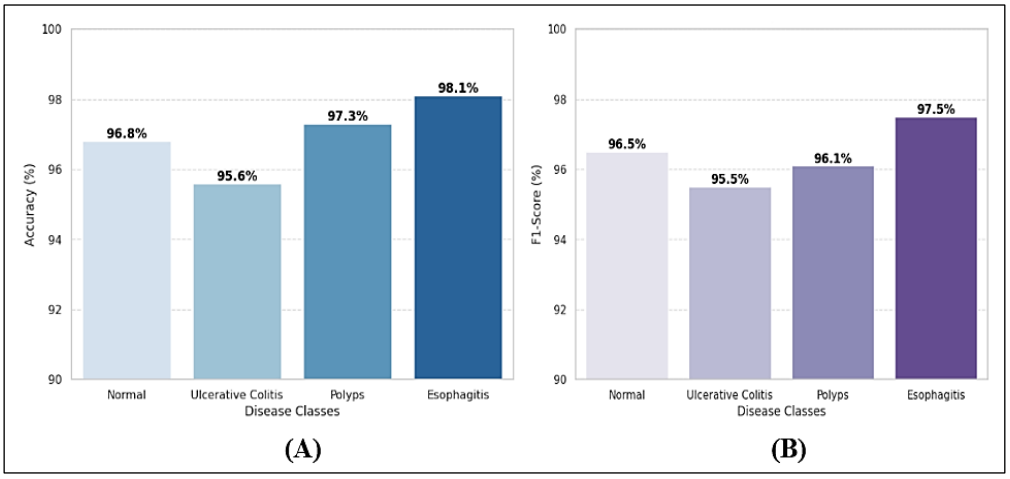


**Figure 6:** (A) Accuracy, (B) F1-Score

The training and testing accuracy and loss were evaluated over 50 epochs, showing a consistent improvement in model performance. The training accuracy steadily increased, reaching high convergence, while the validation/testing accuracy demonstrated stability with minimal fluctuations. The loss function progressively decreased, indicating effective learning and reduced error rates, ensuring the model's robust generalization to unseen data. Figure 7 shows the training and testing accuracy and loss of the proposed model.

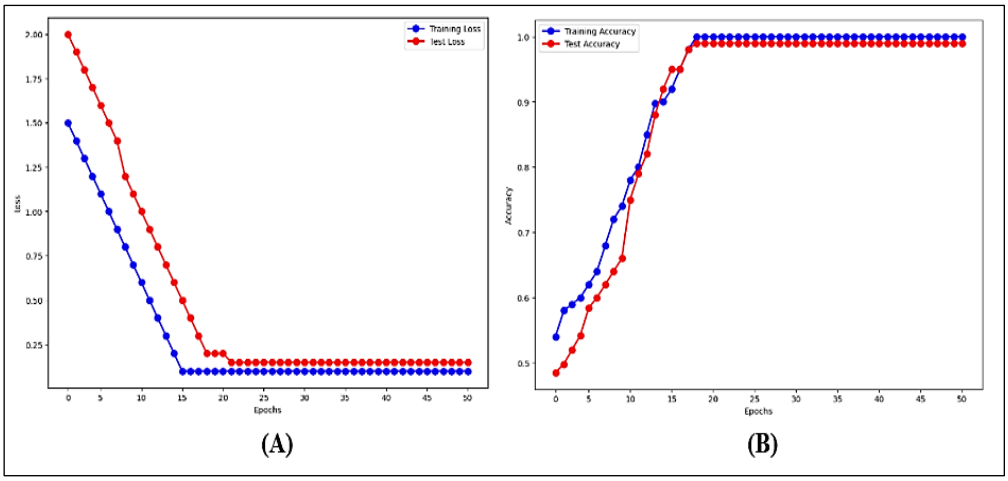


**Figure 7:** Training and Testing Metrics (A) Loss, (B) Accuracy

The confusion matrix demonstrates the model's strong classification performance across all four classes. High true positive values and minimal misclassifications indicate the model's robust accuracy and reliability in colorectal disease detection and it is mentioned in Figure 8.

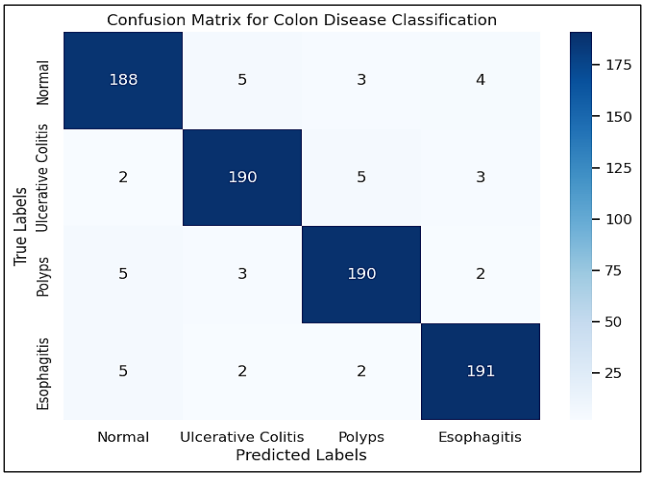


**Figure 8:** Confusion Matrix

These metrics are commonly used in medical image classification tasks to assess the balance between precision and recall, ensuring that the model does not overfit or underperform in detecting important features like polyps, ulcerative colitis, and esophagitis.

The Table 2 summarizes the performance of DenseNet-121 with SE blocks as compared to ResNet-50 and VGG-16.

**Table 2:** Performance of DenseNet-121 with SE Blocks as Compared to ResNet-50 and VGG-16.

| **Model** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-Score (%)** | **Inference Time (ms)** |
|---|---|---|---|---|---|
| DenseNet-121 with SE Blocks | **96.3** | **97.1** | **95.7** | **96.4** | 120 |
| ResNet-50 | 93.8 | 94.5 | 92.0 | 93.2 | 160 |
| VGG-16 | 89.5 | 88.4 | 90.2 | 89.3 | 180 |

The DenseNet-121 with SE blocks outperforms both ResNet-50 and VGG-16, achieving an accuracy of 96.3%, which is higher than other nets. Table 3 represents the performance metric of the proposed work. Table 4 shows the comparison with DenseNet-121 (No SE) and DenseNet-121 with SE Blocks.

**Table 3:** Performance Metric of the Proposed Work

| Class | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) |
|---|---|---|---|---|
| Normal | 97.1 | 95.7 | 96.5 | 96.8 |
| Ulcerative Colitis | 95.3 | 94.9 | 95.5 | 95.6 |
| Polyps | 96.7 | 96.3 | 96.1 | 97.3 |
| Esophagitis | 97.2 | 97.0 | 97.5 | 98.1 |

**Table 4:** Comparison with DenseNet-121 (No SE) and DenseNet-121 with SE Blocks

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| DenseNet-121 (No SE) | 93.4 | 94.0 | 92.8 | 93.4 |
| DenseNet-121 with SE Blocks | **96.3** | **97.1** | **95.7** | **96.4** |

# Conclusion

In this research work an improved version of DenseNet-121 convolutional neural network with Squeeze-and-Excitation (SE) blocks was proposed to classify colorectal diseases. This model succeeded into capitalizing on dense connectivity and feature recalibration to increase the diagnostic precision of four types of diseases. Experiments were done and showcased better results with 96.3 accuracy, 97.1 precision, 95.7 recall, and 96.4 F1-score, beating other classical deep learning models, including ResNet-50 and VGG-16. Computational cost was further cut down by combining with Global Average Pooling (GAP) which keeps the feature representation robust. Although the results are promising, a number of difficulties exist. Although the data was curated, variations in the lighting conditions and the noises in the endoscopic imaging still exist and may influence the performance when applied to real-life clinical situations. Along with that, there are more methods to optimize its use in clinical real-time practice, making it less time-consuming in terms of inference. In future, we would investigate self-supervised learning and transformer-based models to improve further the feature extraction. Besides, adding multi-modal data like the history of the patient, biopsy outcomes, and genetic markers may enhance the robustness of disease classification. An investigation into federated learning techniques will also be provided to make it possible to train models in privacy with a variety of healthcare organizations and have a more general and diverse model. Finally, clinical endoscopic validation on patients will also be done to determine practical enforceability and optimize the model to be used in computer aided diagnosis (CAD) systems.

## Abbreviation

None.

## Acknowledgement

None.

## Author Contributions

All authors contributed equally.

## Conflict of Interest

The authors declare that they have no competing interests.

## Ethics Approval

Patient consent is needed to clarify.

## Funding

None.

# References

1. Swamy BV, Mudalkar PK, Balaji RR, Prasad M, Harikumar M, Saikia A. Enhancing pathology insights: deep learning for histopathological image analysis in colorectal cancer. IEEE; 2023. https://doi.org/10.1109/icsss58085.2023.10407417
2. Ruusuvuori P, Valkonen M, Latonen L. Deep learning transforms colorectal cancer biomarker prediction from histopathology images. Cancer Cell. 2023 Sep 11;41(9):1543-5.
3. Meenakshi Devi P, Gnanavel S, Narayana KE, Sangeethaa SN. Novel methods for diagnosing glaucoma: Segmenting optic discs and cups using ensemble learning algorithms and cdr ratio analysis. IETE Journal of Research. 2024 Aug 2;70(8):6828-47.
4. Nguyen NH, Picetti D, Dulai PS, et al. Machine learning-based prediction models for diagnosis and prognosis in inflammatory bowel diseases: a systematic review. J Crohns Colitis. 2022;16(3):398–413.
5. Yamashita R, Long J, Longacre T, et al. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. Lancet Oncol. 2021;22(1):132–41.

6. Jahagirdar V, Bapaye J, Chandan S, et al. Diagnostic accuracy of convolutional neural network-based machine learning algorithms in endoscopic severity prediction of ulcerative colitis: a systematic review and meta-analysis. Gastrointest Endosc. 2023;98(2): 145–54.
7. Tamang LD, Kim BW. Deep learning approaches to colorectal cancer diagnosis: a review. Appl Sci. 2021; 11(22):10982.
8. Bychkov D, Linder N, Turkki R, et al. Deep learning-based tissue analysis predicts outcome in colorectal cancer. Sci Rep. 2018;8(1):3395.
9. Mahaveerakannan R, Choudhary SL, Dixit RS, Mylapalli S, Kumar MS. Enhancing diagnostic accuracy and early detection through the application of deep learning techniques to the segmentation of colon cancer in histopathological images. IEEE. 2024:1809–15. https://doi.org/10.1109/i-smac61858.2024.10714728
10. Usmani UA, Happonen A, Watada J. Enhancing medical diagnosis through deep learning and machine learning approaches in image analysis. In: 2024 International Conference. Springer International Publishing. 2024:449–68. https://doi.org/10.1007/978-3-031-47718-8_30
11. Kumar R, Anand V, Gupta S. Enhancement of diagnostic accuracy of colon diseases using multi-class classification technique and endoscopic images. IEEE. 2024:188–93. https://doi.org/10.1109/dicct61038.2024.10533146
12. Wei B, Li L, Feng Y, Liu S, Fu P, Tian L. Exploring prognostic biomarkers in pathological images of colorectal cancer patients via deep learning. J Pathol. 2024;10(6):e70003.
13. Gong Y, Xu R, Yu Z, Zhang J. Innovative deep learning methods for precancerous lesion detection. Int J Innov Res Comput Sci Technol. 2024;12(2):14.
14. Kar P, Rowlands S. Assessment of a deep-learning system for colorectal cancer diagnosis using histopathology images. Am J Comput Sci Technol. 2024;7(3):90–103.
15. Rehman A, Khan MF, Afzal F, Aziz N. Improved and automatic classification of polyp for colorectal cancer. IEEE. 2023:1–10. https://doi.org/10.1109/ICBATS57792.2023.10111391
16. Narmadha D, Sundar GN, Priya SJ, Sagayam KM. Deep learning model for disease prediction using gastrointestinal-endoscopic images. IEEE Conf Syst Process Control. 2023;133–7. https://doi.org/10.1109/ICSPC57692.2023.10126043
17. Verma J, Sandhu A, Popli R, et al. From slides to insights: harnessing deep learning for prognostic survival prediction in human colorectal cancer histology. Cent Eur J Biol. 2023;18. https://doi.org/10.1515/biol-2022-0777
18. Ziegelmayer S, Reischl S, Havrda H, et al. Development and validation of a deep learning algorithm to differentiate colon carcinoma from acute diverticulitis in computed tomography images. JAMA Netw Open. 2023;6(1):e2253370.
19. Attia MM, Areed NFF, Amer HM, El-Seddek M. A deep learning framework for accurate diagnosis of colorectal cancer using histological images. Int J Electr Comput Eng. 2024;14(2):2167–80.
20. Gago-Fabero Á, Muñoz-Saavedra L, Civit-Masot J, Luna-Perejón F, Rodríguez Corral JM, Domínguez-Morales M. Diagnosis aid system for colorectal cancer using low computational cost deep learning architectures. Electronics. 2024;13(12):2248.
21. Fanijo S. AI4CRC: a deep learning approach towards preventing colorectal cancer. Faith. 2024;1(2):143–59.
22. Gonçalves T, Rio-Torto I, Teixeira LF, Cardoso JS. A survey on attention mechanisms for medical applications: are we moving toward better algorithms? IEEE Access. 2022 Sep 14;10:98909-35.
23. Yang Z, Liu W, Berlowitz D, Yu H. Enhancing the prediction of disease outcomes using electronic health records and pretrained deep learning models. arXiv preprint. 2022 Dec 22. https://doi.org/10.48550/arXiv.2212.12067