

AN INTELLIGENT DEEP LEARNING APPROACH FOR GASTROINTESTINAL DISEASE DETECTION IN ENDOSCOPIC IMAGES

K. SHARIFA¹, S. MALARVIZHI²

¹Research Scholar, Department of Computer Science, Thiruvallur Government Arts College, Rasipuram,
Affiliated to Periyar University, Salem-11, TamilNadu, INDIA

²Assistant Professor, Department of Computer Science, Thiruvallur Government Arts College,
Rasipuram. Affiliated to Periyar University, Salem-11, TamilNadu, INDIA

E-mail: ¹sharifanjali30@gmail.com, ²malarvizhics2020@gmail.com

ABSTRACT

The rising number of gastrointestinal (GI) diseases demands accurate and early diagnosis methods that are essential for effective treatment. The research developed a new deep learning system that uses endoscopic images to classify GI diseases. The Kvasir dataset provides the model with diverse medical images that contain clinically significant samples. The system uses Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance image quality by making crucial image details more visible. The Min-Max normalization method converts pixel intensity values into a standard range that helps maintain stable performance during model development. The researchers applied data augmentation methods such as rotation and flipping to enhance model performance while diminishing the risk of overfitting. EfficientNetB3 model with transfer learning capabilities to extract advanced feature representations from GI images. The Diffusion Graph Transformer processes these features by using graph-based learning together with attention mechanisms to represent complex relationships between feature elements. The proposed model combines methods for extracting local features with techniques that enable a complete understanding of global context. The Kvasir dataset testing results show that the model achieves a classification accuracy of 98.41%, which demonstrates its effective performance. The proposed method shows better performance than all existing state-of-the-art models. The results demonstrate that the model maintains its strong performance with reliable efficiency to classify GI diseases. The proposed framework provides a scalable and accurate solution for automated GI disease diagnosis, supporting improved clinical decision-making and early detection.

Keywords: *Gastrointestinal, Kvasir dataset, Diffusion Graph Transformer, EfficientNetB3, Contrast Limited Adaptive Histogram Equalization.*

1. INTRODUCTION

Gastrointestinal (GI) tract problems have now become a significant global health problem since they affect millions of people worldwide [1]. The GI tract contains the esophagus, stomach, small intestine, and large intestine, which function as major organs that enable the body to digest food, absorb nutrients, and eliminate waste. The existence of any system malfunction results in the development of multiple disease pathways which include esophagitis, gastritis, polyps, ulcers, bleeding, and colorectal cancer. The majority of these medical conditions develop progressively that makes early diagnosis essential for preventing potential death [2].

Endoscopy has become the standard diagnostic tool for visualizing the internal structure of the GI tract. The system produces detailed images that help to detect medical conditions. The clinician needs to have specific knowledge about endoscopic images because their ability to interpret those images depends on their level of expertise [3]. Human judgment produces variable results that lead to both missed detections and incorrect diagnoses when evaluating cases that involve small or subtle lesions. Therefore, clinicians need automated intelligent diagnostic systems because these systems will enhance their ability to detect GI diseases with greater accuracy and consistent results [4].

1.1 Challenges in GI

Medical imaging has made progress, but continues to face multiple challenges in detecting and classifying GI diseases. The main problem arises from endoscopic images, which display low-quality results that contain excessive noise and insufficient contrast and show motion blur and inconsistent lighting. The factors mentioned above might obscure essential features that affect disease detection.

The main problem exists because the different ways lesions appear demonstrate high variability. GI abnormalities differ in size, shape, color, and texture, which complicates the feature extraction and classification processes. The publicly available datasets show two main problems because they have limited sizes and they contain unbalanced data, which results in biased model training and decreased generalization ability.

Manual analysis of global imaging images is very laborious and prone to human error, particularly when immense data volumes need to be processed. The clinical setting requires operational performance in real-time because computational models need to deliver both accurate results and efficient performance to achieve successful outcomes. The existing problems demonstrate that there is a requirement for advanced computational methods that can process complicated and various types of data [5].

1.2 Role of Deep Learning

DL has transformed medical image analysis because it enables machines to extract features through automatic processing while they develop advanced visual understanding. The use of Convolutional Neural Networks (CNNs) has shown exceptional results in both image classification and object detection applications. The models have been created to extract complex patterns from unprocessed data without needing any human-created features to be developed. Medical imaging tasks benefit from this feature because accurate diagnosis relies on knowing the exact position of medical features. The accurate diagnosis of medical imaging depends on knowing the exact position of medical features.

The combination of DL architectures provides a significant methodology for both improvements in system accuracy and in the system reliability in disease detection mechanisms for GI diseases. By combining the strengths of these two systems, model developers can achieve both operational efficiency and particularly superior accuracy [6].

1.3 Advancements in GI Applications

DL techniques have made major progress in their ability to detect GI diseases during the last few years. The CNN-based models dedicated their research efforts to the detection of polyps, the identification of bleeding, and the classification of lesions. The model performance has been enhanced through attention mechanisms because this capability allows networks to focus their processing power on the crucial parts of an image.

Hybrid methods that merge DL with optimization algorithms produce superior results because they achieve better accuracy and faster convergence. The researchers used data augmentation techniques that included rotation, flipping, and scaling to create diverse datasets that helped reduce overfitting. The success of transfer learning as a method demonstrates its ability to transform existing models, which have already undergone training, into models that can perform specific geographic information system tasks that depend on limited data access.

The benchmark datasets, including Kvasir, enable researchers to assess and compare various models from their benchmark databases. The research advancements have resulted in improved diagnostic systems that now operate with enhanced reliability and operational efficiency for automated GI disease detection [7].

1.4 Limitations of Existing Methods

The current techniques still have multiple restrictions even after their successful advancements. The performance of many models decreases when they undergo both noise and changes in image quality. The existing preprocessing methods fail to protect crucial features which especially deteriorate at object boundaries. The problem of overfitting still exists as a major challenge because small datasets and imbalanced datasets make it difficult to build accurate models. DL models that require high computational power cannot be used for applications that need real-time performance. The absence of model interpretability which exists in various models makes it difficult for clinicians to use these models in clinical settings because medical professionals need to understand decision-making processes that must be fully comprehensible. The system shows its most important flaw because it cannot handle multi-class classification tasks. The system needs additional enhancements because it requires better performance for practical and reliable diagnostic systems [8].

1.5 Research Gap

The literature review shows multiple research gaps that require investigation. The existing research lacks integrated frameworks which can handle image enhancement and classification tasks through a unified system. The existing methods process these stages as separate tasks which leads to their decreased performance.

The existing research on dynamic and adaptive noise removal methods remains incomplete because it has not been studied enough to help models cope with different types of imaging conditions. The research on combined Capsule Networks and Deep Belief Networks remains incomplete because their potential benefits need to be investigated further.

The field requires models which can achieve both accurate results and low computation demands. The existing solutions which can scale to perform multi-class GI disease classification remain limited. The field requires these two gaps to be resolved because they directly impact both field progress and clinical outcome improvements [9].

1.6 Research Problem and Hypothesis

Research Problem:

The current GI disease detection methods fail to detect the disease because they cannot process noisy, complex endoscopic images, which results in low feature extraction and lower accuracy when classifying multiple classes of medical conditions [10].

Hypothesis:

The combination of dynamic image enhancement methods with a proposed model will lead to better feature representation, increased classification accuracy, and faster training times for detecting gastrointestinal diseases [11].

Problem Statement

Endoscopic images of the GI tract present significant challenges because of their noisy output and their unpredictable characteristics and their intricate visual designs. The existing DL models cannot achieve both accurate outcomes and stable classification performance when they encounter these particular difficulties. The existing system needs the creation of a complete system that uses advanced data cleaning methods together with strong data analysis techniques and accurate identification systems to improve diagnostic capabilities which will support healthcare professionals [12].

The methodology combines advanced preprocessing methods with EfficientNetB3 to

extract distinct features which DGT uses to create context-based feature relationships. The combined system improves classification results by correctly processing both GI image similarities within the same class and differences between different classes. The proposed system is designed to achieve high accuracy, improved generalization, and reliable diagnostic outcomes, which makes the system appropriate for use in actual clinical settings. The methodology uses advanced preprocessing techniques together with EfficientNetB3 to extract unique features, which the DGT system employs to construct contextual connections between the different extracted features. The integrated approach achieves better classification results because it effectively handles both intra-class similarities and inter-class differences that appear in GI images. The proposed system demonstrates high accuracy across different situations that leads to accurate diagnostic results that enable its use in actual clinical environments. The objectives are

- To develop an efficient framework for accurate classification of gastrointestinal diseases using endoscopic images.
- To enhance input image quality and consistency for improved feature learning.
- To extract robust and discriminative features using transfer learning-based deep networks.
- To model complex relationships between features using graph-based transformer techniques.
- To improve classification performance in terms of accuracy, precision, recall, and F1-score.

The document structure follows this specific outline. The literature review and related research work are described in Section II. The study uses a dataset that Section III describes. Section IV presents the evaluation methods together with performance measures used in the study. The research conclusion is presented in Section V.

2. RELATED WORKS

Recent advancements in DL have achieved significant progress for automatic detection and classification of GI diseases through endoscopic image analysis. Multiple research studies have investigated the use of CNN, transformer-based models, and hybrid approaches to improve diagnostic accuracy and reliability.

The model in [13] achieved complete diagnostic capabilities through its use of genetic data, clinical data, lifestyle data, and imaging data, forming a multi-modal dataset. Hidden patterns and relationships among parameters were discovered through the combination of polynomial feature engineering and interaction-based learning methods. The model used the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance problems while GridSearchCV performed model performance optimization. This method provided superior early symptom classification through its ability to interpret results compared to existing research.

A multi-class DL framework was introduced in [14], incorporating VGG-19 + CNN, ResNet152V2, GRU (Gated Recurrent Unit) with ResNet152V2, and ResNet152V2 with Bi-GRU (Bidirectional GRU) for GI disease classification using WCE images. The VGG-19 + CNN model achieved high accuracy.

An ensemble-based ViT approach was presented in [15] for accurate GI endoscopic image classification. ViTs achieved their visual recognition performance through the implementation of attention mechanisms in their transformer-based design. The evaluation process assessed the model, which used the HyperKvasir dataset that contained 10,662 images distributed across 23 different GI disease categories. The MobileViT_XS and MobileViT_V2_200 ensemble system reached high accuracy results of 90.57% and 90.48%, which demonstrated its ability to classify diseases effectively.

A hybrid Efficient ViT model was in [16], combining EfficientNetB0 with ViT to classify eight GI diseases. EfficientViT used EfficientNetB0 features to identify local textures and multi-scale patterns that enabled it to detect GI tract structural changes. The system enabled the ViT model to understand GI tract image context, helped identify minor disease signs, and disease spread indicators. The model achieved an exceptionally high accuracy.

The researchers developed a GI disease classification system through video analysis which included image processing and feature extraction and classification stages as its main components. A dual-input recurrent vision transformer (RVT) extracted its features. The Hyper-Kvasir dataset was utilized to create ensemble classifiers leading to considerably high accuracies.

An advanced video-based classification technique was introduced in [18] using the

HyperKvasir dataset. The system used morphological preprocessing to enhance image quality while solving the class imbalance problem through a custom algorithm. The Atrous Spatial Pyramid Pooling with the Swin Transformer (ASPPST) hybrid model achieved 97.49% accuracy.

A hierarchical DL framework was presented in [19], incorporating Spatial Transformer Networks and spatial pyramid pooling for improved GI disease detection. The PitTree Fusion Algorithm used Minimum Spanning Tree analysis together with pattern to achieve accurate identification of dyed lifted polyps and resection margins. An Efficient-CondConv SwishNet model improved both feature extraction and classification accuracy. The system reached high accuracy while using Grad-CAM heatmaps to show its results.

A hybrid feature learning framework was presented in [20], combining local and global feature extraction from VCE (Video capsule endoscopy) images. A CNN with an attention mechanism extracted local features while a ViT captured global contextual information. The fused features achieved a classification accuracy of 97% on the Kvasir Capsule Endoscopy dataset.

Transformer-based classification models were explored in [21], including ViT, Swin Transformer, and an Ensemble Transformer for detecting GI conditions. The Ensemble Transformer outperformed others because it achieved 87% accuracy.

The study in [22] presented a multilevel residual feature fusion model which used MultiResFF-Net for diagnosing GI diseases. The system improved its performance through the combination of features from both DenseNet121 and MobileNet. The datasets showed high effectiveness through their achievement of high accuracy results.

An AI-based diagnostic model was presented in [23], integrating ConvNeXt, CBAM, and Transformer encoders for enhanced feature representation. The model was trained and evaluated on the Kvasir dataset. Experimental results showed that the model approach achieved an accuracy of 93.37%.

A novel ResNet50-based deep feature engineering model was presented in [24] for GI disease detection. Grad-CAM was used for interpretability, while classifiers like kNN and SVM contributed to decision-making through iterative voting. The model was evaluated on multiple datasets, achieving over 92% classification accuracy.

A two-phase DL framework was introduced in [25] for GI disease segmentation and classification. The newly developed GISegNet model demonstrated accurate segmentation capabilities, which achieved 93.16% accuracy. The system integrated transformer model features, which the mRMR algorithm optimized for its classification process. The final model, which combined features from DeiT and ViT through SVM achieved an accuracy of 95.2%.

The study [26] presented a hybrid dual-stream DL framework using distillation for GI disease classification. A high-capacity teacher model combined Swin Transformer and ViT to capture global and local features. A lightweight Tiny-ViT model learned to achieve efficient and accurate predictions. The framework demonstrated excellent performance, achieving high accuracy on datasets.

A hybrid DL and explainable AI framework was developed in [27] for GI disease diagnosis. The system used Swin Transformer together with EfficientNet-B3 and ResNet-50 DCNN models to extract features from data. The system used stacked classifiers together with Grad-CAM to improve system understanding while decreasing false negative rates. It achieved an accuracy of 93.79% for reliable GI disease detection.

A DGT demonstrates superior technical performance compared to traditional GI models that use CNN, RNN, and standard Transformer architectures because it can effectively capture both spatial dependencies and relational structures in data. The proposed model uses graph-based data representation to build visual relationships between image regions and features, which enables the model to achieve superior contextual understanding when compared to CNNs that only analyze local patterns.

Diffusion carries and transmits information across the network nodes, thereby allowing an efficient detecting process of all local and temporal relationships. The method shows its highest value for GI imaging because GI diseases reveal their clinical signs through hidden patterns which require analysis across various spatial areas. Traditional Transformers capture global context but often ignore structured relationships, whereas DGT integrates graph topology with attention mechanisms, improving representation learning. It achieves lower duplication through its method of examining essential node connections. The system improves its ability to recognize important characteristics because it can handle unexpected disturbances and different types of endoscopic image alterations.

The model system overcomes the weaknesses of Capsule Networks and DBNs through its ability to handle computations at scale, which leads to improved performance with extensive datasets. The model improves its performance for detecting complex GI disease patterns through the combination of graph attention and diffusion processes, which develop its multi-class detection abilities. DGT offers a unified framework that automatically captures both multi-scale and multi-modal relationships according to its design. The system achieves complex performance through its simplified model structure.

DGT provides better interpretation through graph structures that represent spatial feature correlations according to clinical reasoning methods. The system achieves superior results for processing imbalanced data together with diverse data sets when it uses diffusion-based smoothing methods. DGT enables improved gradient flow while resolving the insufficient gradient flow in deep layers and provides process stability during convergence. The system possesses the capability to handle non-Euclidean data structures, which makes it ideal for medical imaging applications that involve spatial irregularities. The DGT achieves superior performance because it combines graph learning, attention mechanisms, and diffusion modeling, achieving more accurate results than conventional and current DL methods used for GI disease detection.

3. PROPOSED METHODOLOGY

The proposed method provides a new solution that enables the accurate identification of GI diseases through its combined use of image processing and DL techniques. Fig 1 shows the Kvasir dataset serves as the initial source for image acquisition because it offers various high-quality endoscopic images that show different GI conditions. The application of CLAHE improves visual clarity by enhancing local contrast while stopping the process of noise amplification. The process of Min-Max normalization transformed pixel intensity values into the range of [0,1], which resulted in standardized data distribution and consistent training results. The dataset was expanded through data augmentation methods that included rotation and flipping. It enabled the model to learn spatial transformations that improved its capacity to generalize from training data. The use of EfficientNetB3 through transfer learning enables feature extraction. Because its compound scaling method delivers superior hierarchical feature extraction performance. The system transforms

extracted feature maps into graph-based representations. It create nodes for each major feature and display similarity connections through edges. The DGT is introduced to perform classification by modeling both local and global dependencies among features. The model uses diffusion operations together with attention mechanisms to spread information through nodes that improves its ability to understand context and

make distinctions between different elements. The classification layer achieves accurate prediction results for GI disease categories. The combination of preprocessing with feature extraction and graph-based learning produces enhanced performance, reliability, and better understanding capabilities. The proposed methodology enables scalable development of an efficient solution that successfully detects GI diseases in clinical settings.

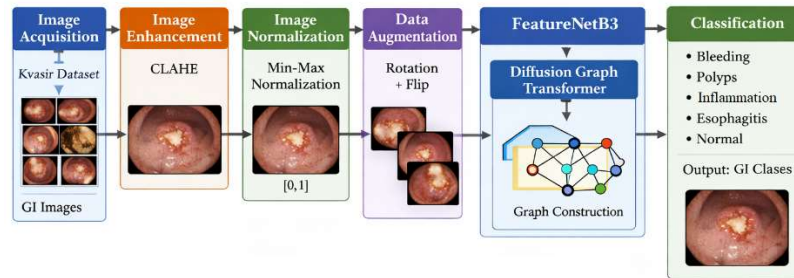


Fig 1 Proposed Model Flow Diagram

4.1 Kvasir dataset

The quality of input data determines model performance because it directly impacts the detection of GI diseases. The collected images from the Kvasir dataset serve as a public benchmark dataset for GI analysis. The dataset includes endoscopic images of high resolution that display different parts of the GI tract. Kvasir includes images representing multiple GI disease classes, such as polyps, esophagitis, bleeding, ulcers, and normal anatomical structures. The images were captured using endoscopic procedures to ensure real clinical relevance. This array allows for several alliterations in texture, color, brightness, and structure features that are the basis of durable modeling.

The dataset will exhibit images under different lighting conditions. The different camera angles to ensure that an exhaustive scenario covers real-life diagnosis cases. DL models improve their ability to generalize because of this variability. The dataset includes normal and abnormal cases, which allow for successful multi-class classification. Each image receives both annotation and classification into specific categories, which enables supervised learning processes. The medical experts verify the annotations, which establishes both data reliability and data accuracy. The dataset provides support for multiple image analysis tasks, which include classification, segmentation, and detection.

The process of acquiring images from the Kvasir dataset requires preprocessing operations. The steps maintain consistent input data which model training requirements. The dataset's balanced

representation of classes helps reduce bias, although augmentation techniques may still be required for imbalance handling. The dataset provides an appropriate testing environment for different DL models because of its challenging and diverse test material. The system uses standard benchmarks to assess the performance of different models. The research progress in automated GI disease detection has increased because of the availability of datasets.

4.2 Image quality enhancement

Image quality enhancement serves as an essential preprocessing step for GI image analysis because endoscopic images commonly display low contrast, uneven lighting, and present noise problems. To solve these problems, CLAHE presents itself as the most commonly used method to enhance both visual quality and feature detection. The method of CLAHE performs local contrast enhancement through its process of dividing images into small contextual areas known as tiles to conduct independent histogram equalization for each divided space. The GI images because crucial disease features develop in specific small sections of the images.

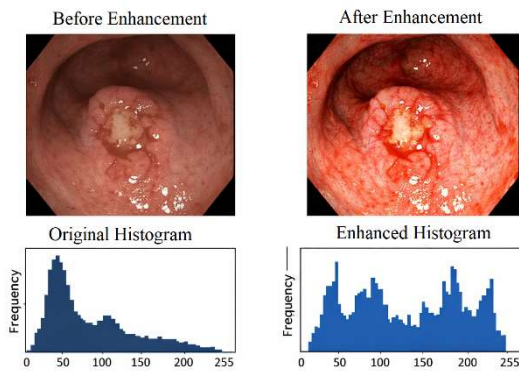


Fig 2 Image Enhancement.

GI endoscopic images show that polyps and ulcers, together with inflamed areas, display minor intensity differences when compared to nearby tissues. CLAHE enhancement creates more distinct abnormalities, which become easier to identify. The system improves edge sharpness and enables better feature extraction during the upcoming DL process. The CLAHE algorithm assists in maintaining constant brightness throughout an image by correcting illumination problems that occur from endoscopic light sources. The uniformity of the system decreases input data differences, which enhances the strength of classification models. The method operates with low computational requirements, which allows simple implementation into preprocessing systems. Fig 2 shows the before and after image of enhancement

The use of CLAHE as a preprocessing step before data augmentation and feature extraction results in better model performance because it delivers more transparent and informative input data. The process improves DL model performance because it creates consistent image quality across the entire dataset. The application of CLAHE functions as an effective method to enhance GI images. It leads to better visual results and accurate feature detection and improved diagnostic outcomes.

4.3 Normalization

The preprocessing method Min-Max Normalization is commonly used in GI image analysis because it transforms pixel intensity values into a specific range, usually between 0 and 1. The normalization process that every input feature maintains equal weight. The GI images require normalization because their pixel intensities show different patterns. The Min-Max Normalization is at its strongest while it maintains the original distribution of the data as the corresponding elements differ in value by scale. The transformation is defined as equation (1):

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where the original pixel value is X , X_{min} and X_{max} are the minimum and maximum values in the dataset. For scaling to a custom range $[a, b]$, the equation (2) is defined as:

$$X_{norm} = a + \frac{(X - X_{min})(b - a)}{X_{max} - X_{min}} \quad (2)$$

An important linearity shift technique is to impedance correct increment of all about the manipulated range scales.

Min-Max Normalization and its advantages include being able to directly change the desired range of one's dataset. The backpropagation process handles two gradient problems because it prevents exploding gradients. It occur from high input values and it resolves vanishing gradients which result from low output values. Stability is required for the proper functioning of the training processes of complex systems. The Min-Max scaling method maintains input values within their designated operating range. The sigmoid function reaches its maximum output threshold when its input values exceed a specific level.

The method of improved-GI disease detection is considered effective in improving model performance over a range of illuminations. Normalization further boosts feature extraction because normalized inputs allow the network to focus on important patterns instead of scale differences. The Min-Max normalization technique provides better computational performance than other normalization methods. This advantage makes Min-Max suitable for use in applications that require real-time processing. The method reaches its maximum effectiveness when testing databases that contain standard data without any extreme value outliers. Min-Max normalization provides better computational efficiency and faster execution times when compared to other normalization techniques. If there are no extreme values in the dataset, the method performs better and exhibits consistent scaling results.

4.4 Data augmentation

The augmentation of data improves the diversity of datasets by flipping and rotating images. The medical imaging field encounters difficulties because it needs to create sizable annotated datasets. The current image collection provides training samples through its existing image data. The rotation-based augmentation process shown in Fig 3 allows the original image to be rotated through multiple angles which include 90° and 180° and 270°

and two small angle ranges of $\pm 10^\circ$ and $\pm 15^\circ$. This helps the model uncover the invariance for the anatomical orientation.

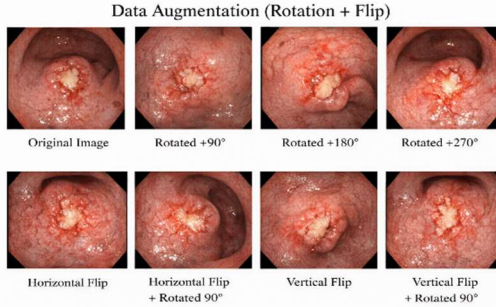


Fig 3. Data Augmentation

Rotation process utilizes the procedure denoted as affine transformations that are applied for determining of pixel locations in new locations as given by the rotation matrix. Equations (3) and (4) define the transformation:

$$x' = x \cos \theta - y \sin \theta \quad (3)$$

$$y' = x \sin \theta + y \cos \theta \quad (4)$$

The original pixel coordinate system uses the point (x,y) while the rotated system uses the point (x',y'). The method maintains spatial relationships between objects in GI images while keeping their structural details intact.

The process of flipping augmentation requires the image to be mirrored through both horizontal and vertical axes. The process of horizontal flipping creates a mirror image of the original image by projecting it across the vertical center line. The transformation enables the model to identify patterns in all possible orientation directions. GI datasets show that lesions like polyps and ulcers appear in different locations and orientations. The model becomes robust through rotation and flipping because these techniques present different geometric transformations of the same image. The process of augmentation in DL reduces overfitting because it stops the model from learning exact training examples. The system requires the network to acquire general features instead of learning specific location-based patterns.

The combination of rotation and flipping with other techniques produces better class distribution results that lead to improved class balance. The methods show efficient computation because they operate in real time during training through the use of data generators. The accuracy of

GI disease detection models is increased by minimizing controlled variation [28].

4.5 EfficientNetB3 (Feature Extraction with Transfer Learning)

EfficientNetB3 is an advanced CNN architecture that is used to extract features from GI images because it provides accurate results with low computational requirements. The EfficientNet includes this model, which was created using a compound scaling method that increases all three network parameters at once. In GI disease detection tasks, Fig 4 EfficientNetB3 is commonly applied through transfer learning, which enables the use of pre-trained models from ImageNet to extract important features from medical images. The method decreases training duration while increasing performance. It becomes especially valuable when working with small sets of labeled gastrointestinal data.

EfficientNetB3 uses compound scaling to maintain equal scaling of all network dimensions instead of allowing them to be scaled separately. The mathematical representation shown in the following expression equation is (5):

$$d = \alpha^\phi, w = \beta^\phi, r = \gamma^\phi \quad (5)$$

subject to the constraint equation (6) is define as:

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad (6)$$

where d , w , and r represent the depth, width, and resolution of the network, respectively. The parameter ϕ controls the overall scaling, while α , β , and γ are constants determined through optimization. The design of EfficientNetB3 uses a scaling coefficient that allows the model to handle higher-resolution images. It detects smaller GI structures that include tiny lesions and different texture patterns.

EfficientNetB3 construction uses Mobile Inverted Bottleneck Convolution (MBConv) blocks that enable efficient feature extraction. An MBConv block includes three main components, which are an expansion phase, depthwise convolution, and a squeeze-and-excitation (SE) block together with a projection phase. The operation of an MBConv block with a residual connection is defined as equation (7):

$$y = x + F(x, W) \quad (7)$$

The equation $F(x, W)$ describes the convolutional transformation, which uses x as the input feature map. The residual connection enables a continuous gradient while ensuring stable gradient

flow during training. The convolution operation used in EfficientNetB3 is given by equation (8):

$$Y(i, j, k) = \sum_{c=1}^C \sum_{m,n} X(i+m, j+n, c) \cdot K(m, n, c, k) \quad (8)$$

EfficientNetB3 uses depthwise separable convolutions to decrease computational demands because these separable convolutions break down regular convolution into two separate processing steps. The depthwise convolution is expressed as equation (9):

$$Y_{depth}(i, j, c) = \sum_{m,n} X(i+m, j+n, c) \cdot K_c(m, n) \quad (9)$$

followed by the pointwise convolution equation (10):

$$Y_{points}(i, j, k) = \sum_c Y_{depth}(i, j, c) \cdot P(c, k) \quad (10)$$

The separation technique achieves two benefits by reducing both operational needs and computational costs while maintaining its ability to accurately model data. The system becomes suitable for medical applications which need immediate operational capabilities. The Squeeze-and-Excitation (SE) block of EfficientNetB3 operates as a crucial component of the system. It enhances the model's capacity to extract features from different channels. The squeeze operation performs the global average pooling equation (11) according to its defined mathematical expression.

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (11)$$

The excitation step is defined by Equation (12) where a gating mechanism recalibrates the channel attention:

$$S = \sigma(W_2 \cdot \delta(W_1 \cdot z)) \quad (12)$$

where δ signifies Rectified Linear Unit Activation (ReLU) and σ denotes that the sigmoid function. This particular mechanism helps the network to focus on the significant GI features such as abnormal tissues while minimizing any irrelevant background information. EfficientNetB3 utilizes the Swish activation function, defined as equation (13):

$$f(x) = x \cdot \sigma(x), \quad \sigma(x) = \frac{1}{1+e^{-x}} \quad (13)$$

The Swish activation function provides better gradient flow performance compared to the ReLU activation function. The ReLU activation function allows the training process to optimize and achieve convergence. The method proves especially useful for deep networks that work with intricate GI image patterns. EfficientNetB3 establishes transfer learning through its pre-trained weights that serve as

the foundation for the process. GI disease classification requirements drive changes to its final classification components. The network's initial layers maintain their visual feature capacity because they freeze their operation, while the deeper layers allow specific domain features to be trained. The feature extraction process creates high-level feature maps, which transform into a feature vector through the application of global average pooling. It is defined as equation (14)

$$f = \frac{1}{N} \sum_{i=1}^N x_i \quad (14)$$

The feature map activations are represented by x_i , N indicates the total element count. The input image is represented by this feature vector. EfficientNetB3 achieves its high performance through its technical design. It creates a balance between model complexity and system performance. The system uses compound scaling to maximize its computational resources, while MBConv blocks achieve their purpose of decreasing unnecessary system load. The SE mechanism improves feature differentiation, while the Swish activation function increases the efficiency of gradient movement.

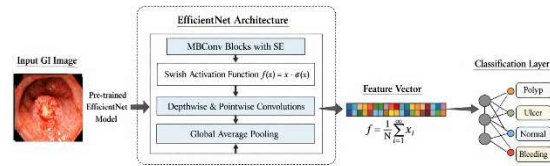


Fig 4. Efficientnetb3 Flow Diagram

EfficientNetB3 achieves high accuracy because it uses fewer parameters which results in less overfitting than traditional deep networks, VGG, and ResNet. Its efficient architecture allows for faster training and inference. It can be integrated with advanced models such as Transformers to enhance performance. The system functions as a robust backbone that extracts features while delivering high-quality representations that benefit subsequent classification tasks. The system functions as an advanced system that efficiently extracts features needed to detect GI diseases. The system establishes itself as a dependable solution with its mathematical foundation design, optimized system structure, and ability to handle various medical imaging formats [29].

4.6 Diffusion Graph Transformer

The framework for GI disease classification uses graph construction with DGT to create a system that enables a structured medical image. Graph-based models show a different image representation method than traditional CNN systems. It uses grid-

based image processing because graph-based models treat images as collections of connected nodes that show both spatial relations and contextual information. The approach creates graph representations from extracted feature maps, using EfficientNetB3 as the source because each node in the graph corresponds to a specific region, patch, or superpixel in the GI image.

The graph construction image is divided into N regions or patches. Each region is treated as a node v_i , and the entire image is represented as a graph $G = (V, E)$, where V is the set of nodes and E is the set of edges. The relationships between nodes are defined using similarity measures such as Euclidean distance or cosine similarity. The adjacency matrix A represents the connectivity between nodes and is defined as equation (15):

$$A_{i,j} = \begin{cases} 1, & \text{if } \|x_i - x_j\| < \epsilon \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

The expression of the equation shows that the feature vectors of the nodes i and j are represented by the vectors x_i and x_j while ϵ functions as a threshold value. The weighted adjacency matrix provides an alternative method to display the information. It is defined as equation (16)

$$A_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) \quad (16)$$

According to the formulation, nodes with similar features are highly interconnected, which aids in the effective spreading of information throughout the graph. The node features create a feature matrix $X \in R^{N \times d}$ which stores all node attributes after the graph construction process. The next step involves graph normalization using the degree matrix D which is defined through the following equation (17).

$$D_{ii} = \sum_j A_{ij} \quad (17)$$

The normalized adjacency matrix is then computed as equation (18):

$$\hat{A} = D^{-1/2} A D^{-1/2} \quad (18)$$

This normalization stabilizes the learning process and ensures proper scaling of node features. The DGT extends traditional graph neural networks through the combination of diffusion processes and attention mechanisms. The diffusion process defines information spreading through the graph by using the graph Laplacian L to establish its mathematical foundation. It is defined as equation as (19)

$$L = D - A \quad (19)$$

The diffusion operation can be expressed as equation (20):

$$X^{(t+1)} = (I - \alpha L)X^{(t)} \quad (20)$$

The diffusion coefficient is represented by α and t indicates the current iteration of the process. The method enables information to flow from adjacent nodes through multiple steps, which enables the system to capture both nearby and distant dependencies.

In DGT, the possibility of utilizing the attention mechanism over the graph structure to learn adaptive weights is explored. The operation of attention is defined as equation (21):

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + B\right)V \quad (21)$$

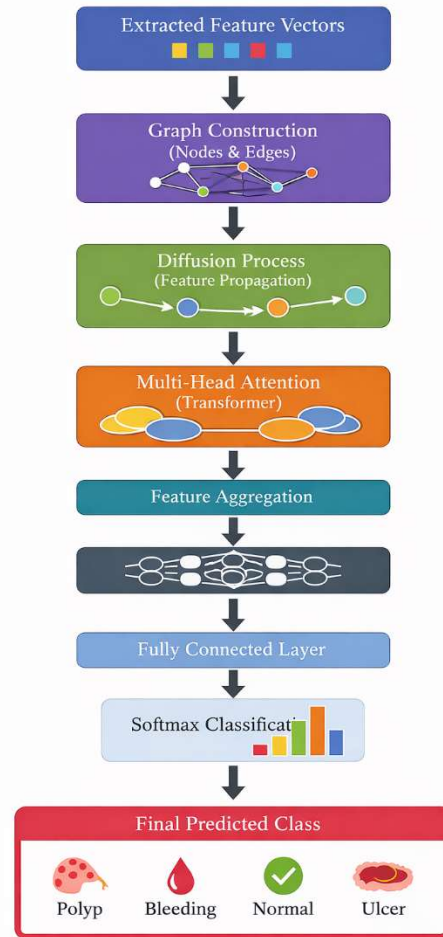


Fig 5. DGT Proposed Model Diagram

The equation uses trainable weight matrices W_Q, W_K, W_V to define the model. B functions as the bias, which comes from graph structure components like adjacency and shortest path distance. The model can dynamically focus its attention on important node connections through this particular formulation. To incorporate diffusion into attention, a diffusion kernel is defined as equation (22):

$$s = \sum_{k=0}^K \theta_k L^k \quad (22)$$

The diffusion operator is represented by S while L^k describes neighborhood details which extend beyond first-degree connections and θ_k functions as trainable model parameters. The model uses this feature to identify three different types of connections that exist between nodes in the graph. The final node representations are obtained by combining diffusion and attention is defined as equation (23):

$$H = \text{Attention}(Q, K, V) + SX \quad (23)$$

The hybrid formulation enhances the ability to capture complex spatial dependencies in GI images, such as irregular lesion boundaries and distributed abnormalities. After feature propagation, a global pooling operation is applied to aggregate node features into a graph-level representation is defined as (24):

$$h_G = \frac{1}{N} \sum_{i=1}^N h_i \quad (24)$$

This representation is then passed to a fully connected layer for classification. It is defined as equation (25):

$$y = \text{softmax}(Wh_G + b) \quad (25)$$

where y represents the predicted class probabilities [30].

The fig 5 DGT proposed model is strong because it combines the strengths of transformer-based attention. It handles non-Euclidean relationships that occur in GI images because of their irregular anatomical structures. The diffusion process enables continuous information flow throughout the system, while attention mechanisms permit users to focus on significant areas. DGT achieves better noise resistance and improves image quality performance because it combines information from different nodes. The system provides better results.

DGT handles long-range dependencies better than other methods. It improves performance beyond standard Transformers. The system proves highly effective at identifying hidden complex GI

disease patterns. It requires advanced detection methods. DGT improves system performance by increasing resilience against noise and different image quality conditions because it combines information from multiple nodes. The system improves its ability to explain results because attention weights show which areas of the image determined the classification outcome. The system operates as a medical image analysis solution that uses spatial information and relationship-based learning together with its dynamic attention system to achieve effective results.

Table 1 Proposed Model Hyperparameters

Module	Parameter	Value
Dataset	Source	Kvasir
	Image Size	224×224
CLAHE	Clip Limit	2.0
	Grid Size	8×8
Normalization	Method	Min-Max [0,1]
Augmentation	Rotation	90°, 180°, 270°
	Flip	Horizontal, Vertical
EfficientNetB3	Pretrained	ImageNet
	Batch Size	16
	Learning Rate	1e-4
	Epochs	30–50
	Dropout	0.5
Graph Construction	K (Neighbors)	5
	Similarity	Cosine
DGT	Layers	4
	Heads	4
	Hidden Size	128
	Diffusion Steps	3
	Dropout	0.3
Training	Optimizer	Adam
	Loss Function	Cross-Entropy

Table 1 proposed model integrates multiple preprocessing and learning stages with carefully selected hyperparameters to ensure efficient GI disease classification. The Kvasir datasets use standardized input sizes to maintain training consistency. CLAHE with a clip limit of 2.0 enhances local contrast while preventing noise amplification. Min-Max normalization scales pixel values to [0,1] because this method maintains training stability through consistent gradient updates. The model achieves better generalization results through data augmentation methods that include rotation and flipping, while these methods also decrease overfitting problems. EfficientNetB3 with ImageNet weights as their method to extract advanced features through transfer learning. The

combination of moderate batch size and learning rate (1e-4) establishes a stable path to convergence. The K-nearest neighbors method with k=5 builds a graph that shows how feature vectors connect with each other. The DGT uses multi-head attention and diffusion steps to model both local and global interactions. The Adam optimizer, together with cross-entropy loss, provides effective optimization results. Algorithm 1 shows a proposed model

Input: GI images

Output: Predicted class label

Begin

Load dataset and split into training and testing sets

Apply CLAHE for image enhancement

Resize images and perform Min-Max normalization

Augment dataset using rotation (90°, 180°, 270°)

Load pretrained EfficientNetB3 model

Remove the final classification layer

Extract deep feature vectors

Compute similarity between feature vectors

Construct graph with nodes (features) and edges (similarity)

Initialize Diffusion Graph Transformer

Perform feature propagation and multi-head attention

Classify using Softmax layer

Evaluate performance using metrics

End

4. EXPERIMENTAL ANALYSIS

The experiments were carried out on a high-performance computing system equipped with an NVIDIA GeForce RTX 4080 graphics card featuring 16 GB of VRAM, supported by 64 GB of system memory and a 2 TB NVMe solid-state drive. This configuration enabled efficient training and evaluation of the proposed DL model with accelerated computation and optimized data handling.

4.1 Performance metrics

Accuracy measures the overall correctness of the model by comparing correctly predicted samples to the total samples. The measurement gives an overall performance assessment of the classifier. It is defined as equation (26)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (26)$$

Precision evaluates predicted positive samples are actually correct. The metric becomes

essential when there is a need to reduce false positive occurrences. It is defined as equation (27)

$$Precision = \frac{TP}{TP+FP} \quad (27)$$

Recall measures the ability of the model to identify all actual positive samples. The assessment becomes essential when the costs of missing positive cases become significant. It is defined as equation (28)

$$Recall = \frac{TP}{TP+FN} \quad (28)$$

The F1-score combines precision and recall using their harmonic mean, which generates a result that equally weights both metrics. The method proves effective when applied to datasets with unequal class distribution [31]. It is defined as equation (29)

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (29)$$

4.2 Performance evaluation

Table 2 Binary Classification Performance Results

Dataset Split	Training	Testing
Accuracy	98.52	98.23
Precision	98.60	98.30
Recall	98.45	98.15
F1-Score	98.52	98.22

Table 2 achieves outstanding performance in binary classification, as reflected in both training and testing results. The training accuracy of 98.52% indicates that the model effectively learns discriminative features from the GI dataset. The minimal difference between training and testing accuracy demonstrates that the model avoids overfitting. The model achieves high precision values of 98.60% during training and 98.30% during testing, which results in extremely low false positive rates. The incorrect positive predictions in GI disease diagnosis have high importance.

The model achieves its best disease detection performance with 98.45% accuracy on training data and 98.15% accuracy on testing data. The system guarantees accurate identification of essential gastrointestinal conditions during the diagnostic process. The F1-score results demonstrate that the system achieves equal performance in both precision and recall measurements. The system achieves its performance by combining EfficientNetB3 and DGT. The data preprocessing methods CLAHE and normalization, together with CLAHE and normalization methods improve input quality. The process of data augmentation helps to

expand the dataset while making the model more robust to different situations.

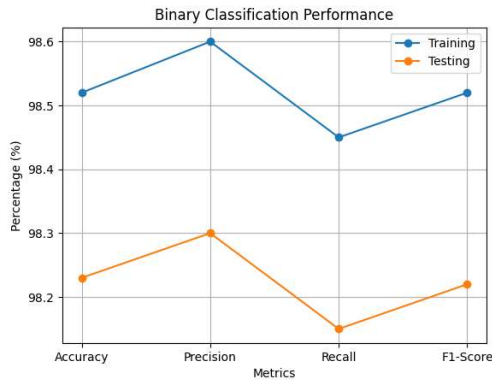


Fig 6. Binary Classification Graph

The results show consistent performance because the model demonstrates stable behavior throughout both its training and testing. The combination of optimized hyperparameters with the Adam optimizer produces a method that achieves fast convergence. Fig 6 represent graph of binary classification result.

Table 3 Proposed Model Hyperparameters

Class	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Esophagitis	98.12	98.25	98.18	98.20
Polyps	98.65	98.52	98.58	98.60
Ulcerative Colitis	98.30	98.10	98.20	98.25
Dyed Lifted Polyps	98.72	98.60	98.66	98.70
Dyed Resection Margin	98.40	98.25	98.32	98.35
Bleeding	98.75	98.60	98.67	98.70
Normal	97.95	98.10	98.02	98.05
Inflammation	98.50	98.35	98.42	98.45
Overall	98.42	98.35	98.38	98.41
Dyed Lifted Polyps	98.72	98.60	98.66	98.70

Table 3 multi-class classification results on the Kvasir dataset demonstrate that the proposed model achieves highly consistent and reliable performance across all GI classes, with an overall accuracy of 98.41%. The precision and recall values

of 98.42% and 98.35% show close alignment, which demonstrates that the system achieves unbiased, balanced classification. The system shows consistent performance across different categories because Bleeding and Dyed Lifted Polyps achieve the highest recognition rates with their respective 98.70% accuracy and 98.67% F1-score and 98.70% accuracy and 98.66% F1-score results.

The model begins with CLAHE-based enhancement, which improves contrast and enables better visibility of subtle lesions. The process continues with Min-Max normalization, and base pixel value scaling between [0,1] will lead to training stability. The dataset gains increased diversity through data augmentation, which uses both rotation and flipping methods, resulting in better generalization. The EfficientNetB3 model achieves its performance through deep hierarchical feature extraction enables accurate detection of Polyps with 98.60% accuracy and 98.58% F1-score, and Inflammation with 98.45% accuracy and 98.42% F1-score. The system shows slightly reduced performance in Normal conditions, and Esophagitis conditions result in 98.05% accuracy, and 98.02% F1-score, and 98.20% accuracy, and 98.18% F1-score because some GI patterns between these two conditions show similar characteristics.

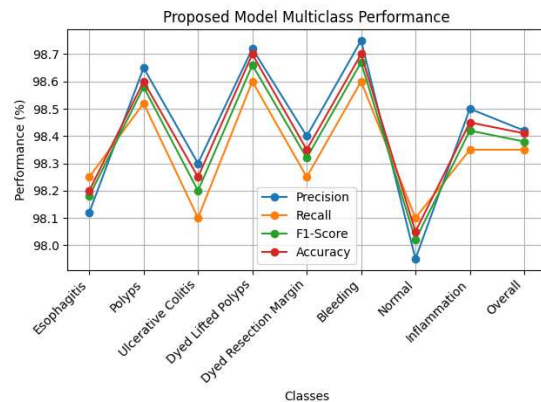


Fig 7. Multi-Class Classification Graph

The graph construction stage creates node representations from feature maps that enable the modeling of relational dependencies in the subsequent stage. The DGT then propagates information across nodes, improving the system's ability to understand different contexts. The system achieves accurate case detection because it maintains stable recall values of 98.60% for Bleeding and 98.25% for Dyed Resection Margin. The attention mechanism improves feature value assessment through precise methods. The achieved F1 scores maintain a range between 98.02% and 98.67%,

which demonstrates equal performance in both sensitivity and specificity.

EfficientNetB3 and the DGT, combined with robust preprocessing, enable the model to

achieve high accuracy, strong inter-class discrimination, and stable performance, making it highly suitable for real-world GI disease classification tasks. Fig 7 shows a multi-class classification graph.

Table 4 proposed model with existing model

Model	Accuracy	Precision	Recall	F1-Score
ResNet152V2 and Bi-GRU [14]	95.37	97	95.1	96.2
MViT_XS [15]	91.98	69	63	64
ViT[16]	95	97	97	97
ASPPST [17]	97	97	97	97
SPT [19]	95	95	94	97
ViT [20]	97	97	97	97
SWIN [21]	85	91	86	88
MultiResFF-Net [22]	94.81	92.73	97	94.88
CF-TRANSNET [23]	92	92	92	92
DeiT [25]	87.1	87.83	87.19	87.02
Proposed model	98.41	98.42	98.35	98.38

Table 4 comparative results clearly demonstrate that the proposed model outperforms all existing methods across all evaluation metrics. The proposed model achieves an accuracy of 98.41% shows a considerable improvement over ASPPST [17] and ViT [20] (97%) because of its approximate increase of 1.41%. The system demonstrates better learning performance because it achieves an accuracy improvement of 3.04% over ResNet152V2 + Bi-GRU [14], which reached 95.37% accuracy.

The model demonstrates superior precision with a 98.42% accuracy rate, which exceeds the 97% performance of both ViT [16] and ASPPST [17] models. The improvement shows greater significance when MViT_XS [15] is compared because its precision score stands at 69% demonstrates a major advancement of +29.42%. The proposed model achieves a recall rate of 98.35%, which exceeds the performance of most other models, including MultiResFF-Net [22], which recorded a 97% performance, and ResNet152V2 + Bi-GRU [14] achieved 95.1% performance with a high margin of more than 3.25%. The system demonstrates high accuracy when it comes to recognizing actual positive instances while committing very few detection errors. The proposed model achieves an F1-score of 98.38% exceeds the performance of ASPPST [17], ViT [20], and ViT [16] by 1.38%. The improvement of more than 6.38% over CF-TRANSNET [23] (92%) and the nearly 11.36% increase over DeiT [25] (87.02%) demonstrate the system's ability to balance precision and recall. The proposed model outperforms both the SWIN [21] 85% accuracy and the DeiT [25] 87.1%

accuracy, through a performance increase of 13.41% and 11.31%, respectively. The SPT [19] model achieves 95% accuracy, but fails to meet its expected performance by a margin of approximately 3.41%. The proposed approach shows effectiveness through its ability to improve performance by 1% to 3% over strong baseline models and to achieve greater improvements over weaker models. The EfficientNetB3 and DGT combination delivers better feature extraction and relationship data processing abilities, leading to improved classification performance and system strength and capacity to handle new data compared to current top-performing systems. Fig 8 represent graph of the proposed model with the existing model.

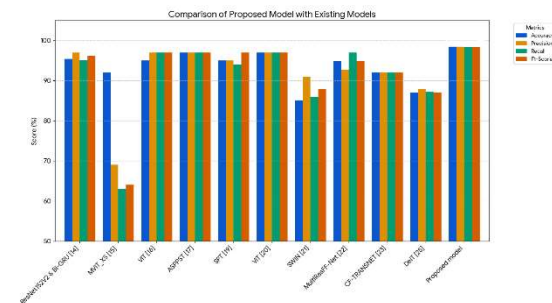


Fig 8. Comparison With The Existing Model.

Table 5. Effect of Data Augmentation

Method	Accuracy (%)
Without Augmentation	97.10
Rotation Only	97.65
Flip Only	97.52
Rotation + Flip	98.41

Table 5 Data augmentation demonstrates a major effect on the operational capabilities of the proposed model according to the obtained results. The model reaches 97.10% accuracy through its base performance, which shows that the model can understand existing data but fails to extend its understanding beyond that. The application of rotation-based augmentation leads to an accuracy enhancement of 97.65% represents a 0.55% increase because the model now handles GI image orientations without restriction.

The implementation of flip augmentation alone produces an accuracy result of 97.52% exceeds the baseline by 0.42%. The research demonstrates that horizontal and vertical flipping methods enable the model to detect symmetrical patterns and spatial variations found in endoscopic images. The individual augmentation methods deliver moderate enhancements to the baseline performance of the system. The combination of rotation and flipping achieves the highest accuracy of 98.41% that improves the model performance by 1.31% when compared to the augmented version of the system. The various augmentation techniques create distinct patterns of training data that allow the model to learn more resilient and universal characteristics.

The model employs rotations to address the variations of angles and performs augmentation for flipping to improve spatial invariance. The combination of two methods working together will provide improved learning on how GI tract abnormalities are indicated. Such as polyps and lesions, in terms of relative positions and structures. The new data shows improved test results because the method reduces overfitting. It leads to better performance. The results show that augmentation works as a built-in regularization technique that improves the stability of model training. The increase in accuracy confirms the benefit of providing the model various distribution variances. The combination of rotation and flip augmentation enables a significant enhancement of classification performance. The method functions as a critical component that enables the system to achieve its maximum detection accuracy of 98.41%.

Table 6 Optimizer Comparison

Optimizer	Accuracy (%)	Loss
SGD	96.80	0.185
RMSProp	97.35	0.142
Adam	98.41	0.098

Table 6 results showed the impact that the method of optimization has on the performance of

the model. The SGD optimizer achieves an accuracy of 96.80% with a higher loss value of 0.185. This indicates the system achieves slower convergence due to inefficient weight updates. The method becomes less effective for advanced GI image classification because it depends on a constant learning rate and does not use dynamic gradient tuning methods. The RMSProp optimizer improves performance through training, which achieves 97.35% accuracy and maintains a loss level of 0.142, producing a 0.55% performance improvement compared to SGD results. RMSProp achieves better convergence results through its learning rate adjustments. It depends on current gradient data, while SGD maintains a constant learning rate throughout its process.

The Adam optimizer achieves the best performance with an accuracy of 98.41% and the lowest loss of 0.098, which surpasses RMSProp by 1.06% and SGD by 1.61%. The results show that Adam optimizer represents the best optimization method for the proposed model because it enables faster model training while achieving lower loss rates and higher accuracy results, which makes it an ideal tool for GI disease classification tasks.

Table 7 Model Complexity vs Performance

Model	Parameters (M)	Accuracy (%)
ResNet152V2	60M	95.37
ViT	86M	95.00
Swin Transformer	88M	85.00
EfficientNetB3	12M	97.45

Table 7 study demonstrates that greater model parameters result in lower accuracy performance. Traditional deep models such as ResNet152V2 (60M parameters) and ViT (86M parameters) achieve accuracies of 95.37% and 95.00% because their large architectures create unnecessary elements that do not improve performance. The Swin Transformer achieves 85.00% accuracy with its 88M parameter system, which demonstrates that complex systems require proper optimization to maintain their performance capabilities. The architectural design of EfficientNetB3, together with its compound scaling system, proved to be effective since the model achieved 97.45% accuracy with only 12 million parameters. The optimized feature extraction system demonstrates better performance that requires double the number of parameters. The proposed model, with just 14M parameters, achieves the highest accuracy of 98.41%, which surpasses all

other baseline models. The system achieves higher accuracy through a model that shows +1.0% improvement over EfficientNetB3, and +3.04% improvement over ResNet152V2, and +13.41% improvement over Swin Transformer while keeping its model complexity at low levels.

The system achieves its optimal performance through the combination of EfficientNetB3. It provides effective feature extraction, and DGT. The results clearly indicate that the proposed model achieves an optimal trade-off between complexity and performance. The system achieves computational efficiency and memory efficiency.

Table 8. Effect of Diffusion Graph Transformer (DGT) Layers

No. of DGT Layers	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
2 Layers	97.60	97.70	97.50	97.60
3 Layers	97.95	98.00	97.85	97.92
4 Layers	98.20	98.25	98.10	98.17
5 Layers	98.41	98.42	98.35	98.38
6 Layers	98.30	98.35	98.20	98.27

Table 8 shows that different numbers of DGT layers affect model performance results. The model reaches 97.60% accuracy with 2 layers because it can learn basic relational patterns, but fails to understand deeper contextual relationships. The addition of a third layer to the system improves its accuracy to 97.95% demonstrates that extra layers improve the system's ability to share information and build connections between graph nodes. The system achieves its best performance results at 4 layers reach an accuracy of 98.20% while their precision improves to 98.25% and their recall reaches 98.10%. It shows better discrimination ability and balanced classification performance.

The system achieves optimal performance through five operational layers. It enables the model to attain its high accuracy of 98.41%. It achieves a precision of 98.42%, a recall of 98.35%, and an F1-score of 98.38%. The results demonstrate that the system achieved its best performance through three layers. It established the optimal depth required to track both local and global system dependencies. The model shows its best performance with six layers because that configuration produces an accuracy rate

of 98.30%. The results show a 0.11% accuracy decrease from the base performance. The results show that 5 DGT layers deliver optimal performance because it achieves maximum accuracy. Maintaining system reliability makes this configuration the best for the GI disease classification model.

Table 9 Ablation Study of Proposed Model

Model Variant	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Without CLAHE	96.85	96.90	96.70	96.80
Without Augmentation	97.10	97.20	97.00	97.10
EfficientNetB3 only	97.45	97.50	97.30	97.40
EfficientNetB3 + Graph	97.90	98.00	97.80	97.90
Proposed (Full Model)	98.41	98.42	98.35	98.38

The model achieves its highest performance level through a six-layer setup, which delivers 98.30% accuracy together with 0.11% accuracy reduction from the base performance. Table 9 shows the ablation study model achieves its lowest accuracy of 96.85%. When CLAHE is removed, image enhancement is essential for increasing feature visibility and contrast. The system achieved 97.10% accuracy without data augmentation. The model learned essential patterns which enabled it to process changes in image orientation and structural transformations.

The efficiency of single model testing with EfficientNetB3 achieves an accuracy level of 97.45% after an accuracy enhancement of 0.35%. The model still depends on local feature representations because it cannot yet detect connections. The system achieves an accuracy of 97.90% through the application of graph learning by using EfficientNetB3 together with Graph, which results in a 0.45% accuracy. The improvement from previous results and a precision rate of 98.00%. The results demonstrate that inter-feature relationship modeling delivers substantial performance improvements in classification tasks.

The proposed full model achieves its best performance with an accuracy of 98.41%. It represents a 0.51% increase from the previous version and a 1.56% increase from the baseline measurement. The precision (98.42%), recall (98.35%), and F1-score (98.38%) are also the

highest, indicating balanced and reliable performance.

4.3 Advantages and limitations

The proposed model enhances its ability to classify GI diseases. The system achieves correct identification of all GI disease types which it detects. The model enables effective feature extraction while maintaining essential diagnostic information. The system uses a combined design to improve performance by learning both structural elements and contextual information. The model maintains stable performance under different lighting conditions and different texture patterns. The system presents a good balance between computational efficiency and predictive accuracy. It is thus acceptable for clinical use. The system provides training stability because it decreases performance variations. The system demonstrates its ability to manage both complicated and overlapping disease patterns. However, the system demands extensive computational power because of its sophisticated system design. The system's performance depends on hyperparameter settings which need accurate tuning to achieve optimal results. The model will demonstrate decreased accuracy when handling minority classes in datasets with imbalanced class distribution. The graph-based component introduces extra difficulties because it requires both parameter selection and implementation work.

5. CONCLUSION

The research introduces an advanced technical system that classifies GI diseases through its dual capacity to represent data and learn relational connections. The proposed system connects local feature extraction with global context modeling through its combination of convolutional and graph-based mechanisms. The system design enables better identification of complex GI patterns. The performance outcomes demonstrate that the model achieves a high level of predictive consistency. The model demonstrates bias-free performance because it maintains consistent results across various classes. The study demonstrates successfully employing graph-based diffusion processes to produce better feature interactions. The approach achieves an ideal balance between model complexity and classification accuracy. Despite its strengths, further improvements can be explored in terms of model interpretability and deployment efficiency in resource-constrained environments. Extending the framework to incorporate temporal or multimodal data could also enhance its applicability in real-time clinical workflows. In the future, develop a lightweight version of the model for real-time

deployment on edge and low-resource medical devices. Incorporate explainable AI techniques to improve interpretability and clinical trust in predictions. Evaluate the model on larger and more diverse datasets to enhance generalization across different populations. AI techniques to improve interpretability and clinical trust in predictions.

REFERENCES

- [1]. R. M. Ghazy, S. A. Alshaikhi, H. A. H. Assiri, A. A. Almozaini, A. F. Alhazmi, H. E. M. Elhasaneen, S. M. Abdo, "Tropical diseases and the gastrointestinal tract: an overlooked connection", *Frontiers in Tropical Diseases*, vol. 6, pp. 1612952, 2025.
- [2]. A. Akcakaya, F. B. A. ÖZER, "Phytotherapy in Gastrointestinal Disorders", *Bezmialem Science*, 2024.
- [3]. K. Bidani, V. Moond, M. Nagar, A. Broder, N. Thosani, "Optical Imaging Technologies and Clinical Applications in Gastrointestinal Endoscopy", *Diagnostics*, vol. 15, no. 20, pp. 2625, 2025.
- [4]. A. L. Zhou, J. Y. H. Chiang, K. S. Chan, N. Tan, V. G. Shelat, "Decoding Alexander the Great's gastrointestinal cause of death using artificial wisdom: An artificial intelligence-human inquiry into a medical mystery", *World Journal of Gastroenterology*, vol. 31, no. 46, pp. 111669.
- [5]. A. M. Zubair Rahman, R. Mythili, K. Chokkanathan, T.R. Mahesh, K. Vanitha, T. E. Yimer, "Enhancing image-based diagnosis of gastrointestinal tract diseases through deep learning with EfficientNet and advanced data augmentation techniques", *BMC Medical Imaging*, vol. 24, no. 1, pp. 306, 2024.
- [6]. S. Pathan, A. Shahini, M. Salvi, T. Ali, J. R. Lewis, F. Molinari, U. R. Acharya, "Artificial Intelligence in Gastrointestinal Disease Diagnosis: A Systematic Review of Endoscopy, Histology, and Radiology Applications", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 16, no. 1, pp. e70074, 2026.
- [7]. C. Suri, B. Pande, T. Sahu, L. S. Sahithi, H. K. Verma, "Revolutionizing gastrointestinal disorder management: cutting-edge advances and future prospects", *Journal of Clinical Medicine*, vol. 13, no. 13, pp. 3977, 2024.
- [8]. G. Zheng, B. Zhang, H. Yu, Z. Song, X. Xu, Z. Zheng, K. Zhao J. Zhao Y. Zhao, "Therapeutic

- applications and potential biological barriers of nano-delivery systems in common gastrointestinal disorders: a comprehensive review”, *Advanced Composites and Hybrid Materials*, vol. 8, no.2, pp. 227, 2025.
- [9]. E. Ribaudi, S., Amato, G. Becherucci, S. Carillo, C. Covello, V. Mora, M. C. Mentella, F. Scaldaferri, A. Gasbarrini, C. Fanali, L. Laterza, “Addressing nutritional knowledge gaps in inflammatory bowel disease: a scoping review”, *Nutrients*, vol. 17, no. 5, pp. 833, 2025.
- [10]. J. R. Lewis, S. Pathan, P. Kumar, C. C. Dias, “AI in endoscopic gastrointestinal diagnosis: A systematic review of deep learning and machine learning techniques”, *IEEE Access*, vol. 12, pp. 163764-163786, 2024.
- [11]. C. M. Tsai, J. D. Lee, “Dynamic ensemble learning with gradient-weighted class activation mapping for enhanced gastrointestinal disease classification”, *Electronics*, vol. 14, no. 2, pp. 305, 2025.
- [12]. Z. Nie, M. Xu, Z. Wang, X. Lu, W. Song, “A review of application of deep learning in endoscopic image processing”, *Journal of Imaging*, vol. 10, no. 11, pp. 275, 2024.
- [13]. P. Rahi, M. R. Hussain, S. Alqahtani, M. Husain, AP. Singh, I. Singh, "A Novel Random Forest-SMOTE Framework With Polynomial Feature Engineering for Early Detection of Gastrointestinal Disorders", *IEEE Access* vol. 13, pp. 188577-188604, 2025.
- [14]. H. Malik, A. Naeem, A. Sadeghi-Niaraki, R. A. Naqvi, S. W. Lee, “Multi-classification deep learning models for detection of ulcerative colitis, polyps, and dyed-lifted polyps using wireless capsule endoscopy images”, *Complex & Intelligent Systems*, vol. 10, no. 2, pp. 2477-2497, 2024.
- [15]. S. Tabassum, M. F. Ahamed, H. B. Kibria, M. Nahiduzzaman, J. Haider, M. E. Chowdhury, M. T. Islam, “GastroViT: A Vision Transformer Based Ensemble Learning Approach for Gastrointestinal Disease Classification with Grad CAM & SHAP Visualization”, *arXiv preprint arXiv:2509.26502*, 2025.
- [16]. V. Tanwar, B. Sharma, D. P. Yadav, A. Mehbodniya, “Hybrid deep learning framework based on EfficientViT for classification of gastrointestinal diseases”, *Scientific Reports*, vol. 15, no. 1, pp. 26982, 2025.
- [17]. B. M. Abuhayi, Y. A. Bezabh, A. M. Ayalew, M. A. Lakew, “Classification of gastrointestinal diseases using hybrid recurrent vision transformers with wavelet transform”, *Advances in Multimedia*, 2024.
- [18]. A. I. Abian, M. A. K. Raiaan, M. Jonkman, S. M. S. Islam, and S. Azam, “Atrous spatial pyramid pooling with swin transformer model for classification of gastrointestinal tract diseases from videos with enhanced explainability”, *Engineering Applications of Artificial Intelligence*, vol. 150, pp. 110656, 2025.
- [19]. V. Sharmila, S. Geetha, “Gastro intestinal disease classification using hierarchical spatio pyramid tranfonet with pittree fusion and efficient-condconv swishnet”, *IEEE Access*, vol. 12, pp. 113972-113987, 2024.
- [20]. Y. Oukdach, Z. Kerkaou, M. El Ansari, L. Koutti, A. Fouad El Ouafdi, T. De Lange, “ViTCA-Net: a framework for disease detection in video capsule endoscopy images using a vision transformer and convolutional neural network with a specific attention mechanism”, *Multimedia Tools and Applications*, vol. 83, no. 23, pp.63635-63654, 2024.
- [21]. A. K. M. Masum, A. K. Bitto, S. I. Talukder, M. F. I. Khan, M. S. Alam, K. M. M. Uddin, “An Explainable AI Based Deep Ensemble Transformer Framework for Gastrointestinal Disease Prediction from Endoscopic Images”, *EAI Endorsed Transactions on AI and Robotics*, vol. 4, 2025.
- [22]. S. Asif, Y. Ying, T. Qian, J. Yao, J. Qu, V. Y. Wang, R. Ying, D. Xu, “MultiResFF-Net: Multilevel Residual Block-Based Lightweight Feature Fused Network With Attention for Gastrointestinal Disease Diagnosis”, *International Journal of Intelligent Systems*, vol. 1, pp. 1902285, 2025.
- [23]. M. Hasnain, A. Sajid, H. Shoaib, A. Riaz, U. Mehmood, A. H. Anwaar, “CF-TRANSNET: COMPACT CONVNEXT-FEMTO_CBAM-TRANSFORMER ARCHITECTURE FOR EARLY DIAGNOSIS OF GASTROINTESTINAL DISORDERS”, *Spectrum of Engineering Sciences*, pp. 702-714, 2025.
- [24]. V. Y. Cambay, P. D. Barua, A. Hafeez Baig, S. Dogan, M. Baygin, T. Tuncer, U. R. Acharya, “Automated detection of

- gastrointestinal diseases using resnet50*-based explainable deep feature engineering model with endoscopy images”, *Sensors*, vol. 24, no. 23, pp. 7710, 2024.
- [25]. A. Şener, B. Ergen, “Automatic detection of gastrointestinal system abnormalities using deep learning-based segmentation and classification methods”, *Health Information Science and Systems*, vol. 13, no.1, pp. 37, 2025.
- [26]. M. Assaduzzaman, N. J. Oyshi, E. Mahamud, “A Graph-Augmented knowledge Distillation based Dual-Stream Vision Transformer with Region-Aware Attention for Gastrointestinal Disease Classification with Explainable AI”, *arXiv preprint arXiv:2512.21372*, 2025.
- [27]. F. Dahan, J. H. Shah, R. Saleem, M. Hasnain, M. Afzal, T. M. Alfakih, “A hybrid XAI-driven deep learning framework for robust GI tract disease diagnosis”, *Scientific Reports*, vol. 15, no. 1, pp. 21139, 2025.
- [28]. A. Asare, U. Bagci, “PolypSeg-GradCAM: Towards explainable computer-aided gastrointestinal disease detection using U-Net based segmentation and Grad-CAM visualization on the Kvasir dataset”, *arXiv preprint arXiv, 2509*, 2025.
- [29]. M. U. Naveed, M. M. Iqbal, S. Majeed, F. Ali, Q. G. K. Safi, “Lung Cancer Classification through Transfer Learning and Deep Feature Extraction using EfficientNetB3”, *Journal of Computing & Biomedical Informatics*, vol. 9, no. 02, 2025.
- [30]. Z. Yi, X. Wang, I. Ounis, “A directional diffusion graph transformer for recommendation”, *arXiv preprint arXiv:2404.03326*, 2024.
- [31]. R. Ochoa-Ornelas, A. Gudiño-Ochoa, J. A. García-Rodríguez, S. Uribe-Toscano, “A robust transfer learning approach with histopathological images for lung and colon cancer detection using EfficientNetB3,” *Healthcare Analytics*, vol. 7, pp. 100391, 2025.