JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-21, No.-03, March (2026) pp 47-62 ISSN (Print) 0973-8975

A SYSTEMATIC ANALYSIS OF GASTROINTESTINAL DISEASE CLASSIFICATION USING HYBRID ATTENTION CONVOLUTIONAL NEURAL NETWORK

K. Rajeswari[1], A. S. Salma Banu[2], P. Sunitha[3], P. Arivazhagan[4]

[1,2]Department of Electronics and Communication Engineering, Aalim Muhammed Salegh College of Engineering, Chennai-600055, India.

[3]Department of Computer Science and Engineering, Sriram Engineering College, Thiruvallur-602024, India.

[4]Department of Advanced Computing and Analytics, VELS University, Chennai- 600117, India.

Email: [1]rajeswariams@gmail.com, [2]as.salmabanu@aalimec.ac.in
[3]starsunitha999@gmail.com, [4]stararivu999@gmail.com

Corresponding Author: **K. Rajeswari**

https://doi.org/10.26782/jmcms.2026.03.00004

(Received: December 20, 2025; Revised: February 22, 2026; Accepted : March 09, 2026)

Abstract

Timely intervention of gastrointestinal (GI) diseases, namely polyps, ulcerative colitis, and esophagitis are critical for improving the quality of human life and reducing the mortality rate associated with these conditions. Hence, early detection and diagnosis of GI disease are essential because they can reduce the severity of the disease. Traditional medical imaging techniques are time-intensive, labor-intensive, and susceptible to human error. Recently, deep learning models have been extensively used for image classification tasks, and they are consistently achieving promising results in real-time decision-making. However, the conventional deep learning models struggle with overfitting and poor generalization on medical imaging datasets because of the wide variability in disease types. To address this issue, a Hybrid Attention Convolutional Neural Network (HA-CNN) is proposed in this analysis. This proposed model integrates the strength of the convolutional operation and attention mechanism to focus on discriminative regions and features in medical images. The hybrid model is designed for high variability and complex features in the medical images. This model can accurately recognise lesion regions and detect types of diseases, and avoids overfitting. The effectiveness of the proposed HA-CNN is evaluated using a benchmark dataset, namely the Kvasir dataset, using 5-fold stratified cross-validation. The model achieves a mean classification accuracy of 94.43% ± 0.58, outperforming existing comparative methods. Moreover, the integration of empirical mode decomposition and dynamic scaling enhanced the quality of training data by improving the generalization ability of the model. By overcoming the existing challenges, this framework focuses on improving the

diagnostic process in medical imaging, resulting in the precise detection of GI diseases.

Keywords: Gastrointestinal Disease, Stomach Ulcers, Deep Learning, Convolutional Neural Network, Empirical Mode Decomposition, Endoscopy Imaging.

I. Introduction

Gastrointestinal (GI) diseases severely affect human life, and if left untreated, they may result in death. The highly prevalent disease that causes a considerably increasing mortality rate is gastrointestinal (GI) disease [XXI, XX]. It causes critical health issues; hence, it needs timely intervention. The most common GI diseases are ulcerative colitis, colon cancer, and various forms of gastritis [XXVII]. Patients undergoing regular check-ups, treatment, and surgeries are expensive, which affects their quality of life and increases the overall medical expenses. Globally, cases of colon cancer and other GI diseases are increasing along with several other cancer types. Because of the increased risk of cancer, GI diseases are more prevalent and require timely intervention and accurate classification to support precise treatment [XXII]. Recently, endoscopic images have become sufficient for examining GI disease; however, they require careful analysis and may still lack precise detection. Compared to existing analysis techniques, the techniques currently in practice detect and classify GI disorders with varying degrees of consistency and objectivity [XVII, II]. Among several GI diseases, colorectal cancer is highly prevalent. Both men and women have been severely affected by this disease, and the factors contributing to this development are polyps, ulcers, and bleeding. Among these, polyps are the easiest to detect and diagnose at an early stage. It is formed by the unusual cell formation from the mucous membrane of the digestive system, and it may also be malignant. The most common technique used for identifying the formation is colonoscopy, in which biopsy and polyp removal are performed. Next, an Ulcer is an inflamed lesion, where a stomach ulcer is commonly called a gastric ulcer. In terms of prevalence, it ranks second in women and third in men. As reported by the WHO, colon cancer is the third most prevalent cancer, majorly affecting people and causes 10% of all cancer-related deaths.

The risk factor of colorectal cancer increases as patients get older. Next, heredity is another risk factor, because it affects close relatives and may affect people who are already affected by any form of genetic disorder or gastric ulcer. Moreover, malnutrition, alcohol addiction, and shifts in daily routine result in colorectal cancer. In the early stage, it does not show any symptoms. Hence, regular monitoring is critical for accurate detection and diagnosis. The most common symptoms of this disease are weight loss without any dietary plan change or physical activities, blood in stool, diarrhoea, abdominal cramping, and discomfort. Traditionally, Colonoscopy [VIII], faecal immunochemical Test (FIT) [V], CT Colonography [XIV], Stool DNA Test [I], and Flexible Sigmoidoscopy [XV] are performed to identify the severity of colorectal cancer. However, deep learning models enable the identification of digestive and stomach troubles. CNN and other deep learning models are primarily utilised to address developing training data and detecting fine details in external images [XXVI]. These models do not require manual effort to improve objectivity,

consistency, and accuracy of clinical information. Typically, Artificial Intelligence (AI) is significantly used for detecting and classifying various GI diseases from endoscopic images, resulting in faster, more accurate, and non-invasive diagnosis. The key benefits of using AI in disease diagnosis are reducing diagnostic errors, enabling timely diagnosis, early therapeutic intervention, reducing cost, and achieving precise results. AI models, namely machine learning (ML) and deep learning (DL), are essential in detecting GI disorders, and their contribution to advancing the screening process in medical imaging is significantly increasing [XXIII]. To accurately detect GI diseases, deep learning models like CNN-based models are employed on endoscopic, radiological, and histopathological images. These models can analyse large datasets and detect patterns more effectively. It also makes endoscopies to precisely identify affected regions based on their severity, faster the decision-making process, and optimise time and resource utilisation [XIX]. The on going researches in this field focuses on utilising complex CNN models or integrating multiple models and enhancing feature extraction for more precise detection. The endoscopic images are subject to several limitations and challenges; hence, it is required to improve the image quality. The factors such as volume and power limitations degrade the image quality. This results in complicating disease classification. In spite of these challenges, we designed an automated hybrid technique for the diagnosis of GI disorders combining CNN with channel, spatial, and global attention mechanisms incorporated into the feature extraction layers.

The remaining work is organised as follows: The proposed work is outlined in Section 2. The test results are clarified in Section 3. Section 4 concludes this paper

II. Methodology

II.i. Dataset Description

The Kvasir dataset incorporates gastrointestinal endoscopic images obtained from hospitals associated with the Vestre Viken Health Trust, Norway [XVIII]. This dataset was specifically developed to support research on automated computer-aided detection and diagnosis of intestinal tract diseases. This dataset includes 8000 images of oesophagus, stomach, and colon, along with annotations representing different types of GI diseases, namely polyps, ulcers, and inflammation [XVI]. It includes a total of 8 classes with image resolutions such as 720 × 576, 800 × 600, 1920 × 1072, etc.

To avoid dataset leakage and overly optimistic performance estimation, the dataset was divided before applying any augmentation procedures. A stratified sampling approach was employed to maintain balanced class distribution across all subsets. Data augmentation techniques were applied only to the training folds after splitting, ensuring that no augmented versions of validation or test images were exposed to the model during training.

II.ii. Data Pre-processing

In this stage, several standardisation and enhancement techniques are employed to improve the image quality. During resizing, the images are scaled to 224×224 pixels [XXVIII]. Subsequently, they are normalised to a range of 0 to 1, enabling the

reduction of model training time, thereby improving the numerical stability of the deep learning model. Following this, Empirical Model Decomposition (EMD) is applied to remove noise in the input images. It operates by decomposing the images into a series of intrinsic mode functions (IMFs) with variable frequencies. The high-frequency IMF comprises lesion information, while the low-frequency IMF comprises lesion boundaries. By selecting variable frequency IMFs, the images are reconstructed to obtain heterogeneous characteristics. Its operation is explained as follows: The input image is represented by $f(a,b)$, and the residue of the nth IMF is represented as $x_n(a,b)$. While detecting residue of subsequent IMF, the previous IMF is treated as input. Initially, the input image is represented by $y_{n,t}(a,b)$, where n denotes nth IMF (n=1,…,n) and t denotes tth iteration of the shifting process (t=1,…,t). The variables (a,b) denote two spatial dimensions of the input image. Following this, it extracts local maxima and minima of the input image and develops the maximum surface $y_{max}(a,b)$ and minimum surface $y_{min}(a,b)$ through interpolation. Next, the mean value of these two surfaces is determined by,

$$e_{n,t}(a,b) = \frac{y_{max}(a,b) + y_{min}(a,b)}{2} \tag{1}$$

The input image is updated by subtracting the envelop mean ted from the input data.

$$y_{n,t}(a,b) = y_{n,t-1}(a,b) - e_{n,t}(a,b) \tag{2}$$

Moreover, the obtained result is verified as an IMF or not. In this phase, standard deviation (σ) is formulated as follows:

$$\sigma = \sum_{a=0}^{A-1}\sum_{b=0}^{B-1} \frac{\left|y_{n,t}(a,b) - y_{n,t-1}(a,b)\right|^2}{y_{n,t-1}^2(a,b)} \tag{3}$$

The value of σ should range from 0.2 to 0.3. If the value of σ exceeds the specified range, the steps are repeated from the beginning; if σ is lower than the specified range, the updated input data is selected.

$$m_n(a,b) = y_{n,t}(a,b) \tag{4}$$

Then the residue of nth IMF becomes,

$$x_n(a,b) = f(a,b) - m_n(a,b) \tag{5}$$

During calculation of the next IMF, the preceding residue is utilized as an input, and the steps are repeated from the initial step.

$$y_{n+1,0}(a,b) = x_n(a,b) \tag{6}$$

This calculation continues until the residue reaches its extreme points. Thus, after calculating N number of IMFs and final residue without the extreme point, the original image is defined as:

$$f(a,b) = \sum_{n=1}^{N} m_n(a,b) + x_n(a,b) \tag{7}$$

Typically, the endoscopic images are subjected to dull illumination because of the camera placement. Hence, the images are subjected to brightness adjustment and shadow adjustment, enabling the classifier to capture features at varying conditions. To address orientation and rotation issues, the images undergoes random rotation and flipping technique, allowing precise detection of lesion regions. Furthermore, scaling and zoom augmentation are performed to handle scale variability, facilitating the detection of features at various regions of the image.

II.iii. Segmentation

The Attention U-Net model is used for localising lesion regions in the endoscopic images. This segmentation technique comprises a contracting path that contains a convolutional layer, a max pooling layer, and an expanding path constructed using a convolutional layer and upsampling layers [VI]. The attention gap incorporated within the skip connections improves the model's capability to emphasise relevant features, enabling precise localisation of the lesion region. The attention gate comprises two different feature maps, namely G and X. In the lower layer of the model, the gating signal G is obtained, and its dimensionality is reduced to enhance feature representation. The vector X and vector G undergo strided convolution and 1×1 convolution. Both vectors are integrated using element-wise summation and create a resultant vector [VI]. Next, the resultant vector is fed into a non-linear activation and sigmoid layer to form attention coefficients. In order to preserve the attention coefficients and their dimensions, the trilinear interpolation is applied. Furthermore, the attention coefficient is multiplied by vector X, thereby scaling its features according to their relevance before decoding. This enables precise spatial localisation of lesion structures.

To overcome the methodological decoupling between segmentation and classification, this study employs a unified multi-task learning strategy instead of a sequential pre-processing pipeline. A shared ResNet-50 encoder captures hierarchical feature representations, which are simultaneously forwarded to an Attention U-Net-based segmentation decoder and a classification branch. Both tasks are trained simultaneously using a combined loss function that integrates segmentation Dice loss and classification cross-entropy loss. It is mathematically expressed as,

$$L_{total} = \lambda_1 L_{Dice} + \lambda_2 L_{CE} \tag{8}$$

where L_{Dice} denotes Dice loss for lesion segmentation and L_{CE} for cross-entropy loss for classification. Through joint backpropagation, gradients from the classification branch update the shared encoder and indirectly refine segmentation representations, thereby preventing uncorrected error propagation. This integrated optimisation ensures that lesion localisation actively guides disease classification and formally defines the proposed architecture as a true hybrid segmentation–classification system. It is crucial that segmentation probability maps are not explicitly used for direct masking of classifier inputs. Instead, integration occurs implicitly through shared encoder optimisation and joint gradient propagation. Thus, lesion localisation influences classification via shared feature representation learning rather than functioning as an individual mask-based preprocessing stage.

II.iv. Classification

Traditional CNN-based models fail to capture discriminative lesion regions and long-range contextual information. To overcome these challenges, a hybrid attention convolutional neural network (HA-CNN) is proposed [XXV]. The detailed study of the components is presented in the following sub-sections. Figure 1 presents the architecture diagram of the proposed HA-CNN.

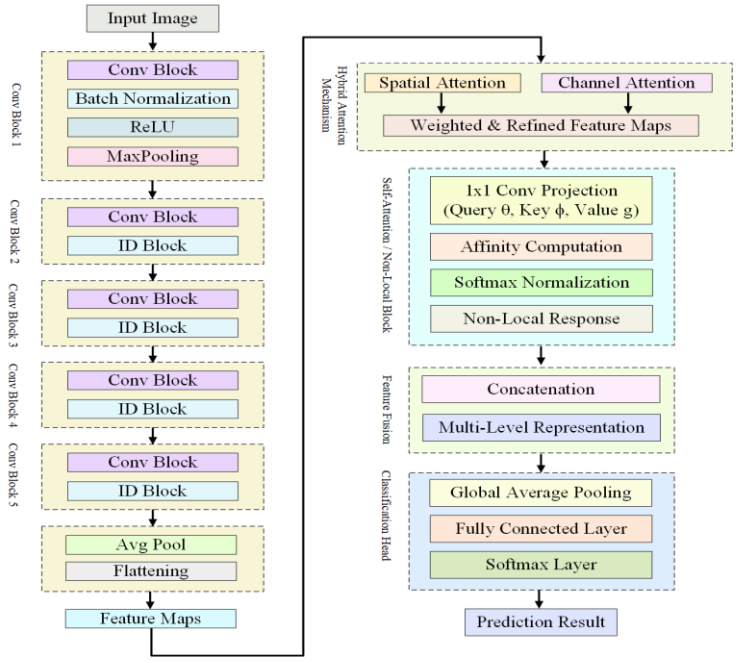

Fig. 1. Architecture diagram of HA-CNN.

A. Local feature extraction

In HA-CNN, ResNet-50 is employed as the backbone and captures local features from the input images. The ResNet-50 backbone described here corresponds to the shared encoder introduced within the segmentation framework, thereby ensuring unified feature learning across both segmentation and classification tasks. This CNN-based model extracts initial target features, and the following feature extraction layers and target detection networks are based on the features captured by it. Hence, the backbone determines the performance of the entire network.

It is developed by combining several bottleneck residual blocks, which comprise a series of 1×1, 3×3, and 1×1 convolutions. This network uses a residual network to overcome the degradation problem that occurs when accuracy reaches a saturation point and affects rapidly as the network depth increases.

Without freezing the entire backbone, a partial fine-tuning strategy is employed to overcome domain adaptation challenges. While early convolutional layers are retained to preserve general low-level features learned from ImageNet, the deeper residual stages are unfrozen and optimised during training. This makes adaptation to

the domain-specific texture statistics, illumination variations, and mucosal structural patterns present in the Kvasir Dataset, which differ substantially from natural images.

This partial fine-tuning enhances feature hierarchy adaptation while preventing catastrophic forgetting of generic visual representations.

B. Channel attention module

This model emphasises identifying the most informative and discriminative features from the endoscopic image. To generate channel attention, the spatial dimensions of the input feature maps are decreased [X]. This is achieved using average-pooling and max-pooling operations. In order to combine spatial information, average-pooling is used, while the max pooling operation captures local features associated with distinctive lesion characteristics. Hence, both average pooling and max pooling features are utilised to infer finer channel-wise attention and improve the representation capability of the model [XIII]. This results in discriminative channel attention maps for GI disease classification.

C. Spatial attention module

The spatial attention maps are formed by utilizing the inter-spatial relationships of features. The spatial attention module emphasises where the distinctive lesion features are located. To determine spatial attention, the operations, namely average pooling and max-pooling, are applied on the channel axis and combined to form an effective feature descriptor. This results in enhancing informative regions. Following this, a convolution layer is utilised on the feature descriptor to form a spatial attention map to represent where the model needs to be emphasised or suppressed [XXIV].

D. Self-attention/Non-local block

This block captures long-range dependencies of the feature maps obtained from both channel attention and the spatial attention module by determining interactions between two positions without considering the distance between them. This facilitates the construction of a rich hierarchy to fuse both local and non-local features. Using 1×1 convolution, the feature embedding, namely query (θ), key (ϕ), and value (g) are generated, and a pairwise similarity operation is computed among these embeddings. Following this, softmax normalises the resultant affinity map and generates the non-local attention response.

E. Feature Fusion layer

The feature fusion combines global and local features extracted from the previous layer to develop a more comprehensive feature representation. The integration of different features extracts multilevel information from the training image and achieves better accuracy. Next, the representation layer captures and combines the fused features to construct a vector representing the legion region. It represents the lesion region as a fixed-length vector; the model converts the lesion region of different sizes into a unified representation, enabling further classification tasks.

F. Classification Head

Initially, the global average pooling layer decreases the spatial dimensions and maintains global information, thereby avoiding the risk of overfitting. Following this,

a fully connected layer captures high-level features from the preceding layer, and the softmax layer outputs the probability distribution across GI disease classes.

II.iv. Training and Validation Protocol

To ensure reliable performance evaluation and minimise the impact of random initialisation, a 5-fold stratified cross-validation strategy was employed. The dataset was partitioned into five mutually exclusive folds while maintaining balanced class distribution in each fold. During each iteration, four folds were used for training, and the remaining fold was reserved for validation. This procedure was repeated five times so that every fold was used once for validation.

Performance metrics are reported as mean ± standard deviation across the five folds to reflect model stability. In addition, experiments were repeated with three different random seeds to further assess robustness. This evaluation framework reduces the risk of overfitting and provides a more dependable estimate of the model's generalisation performance.

III. Result and Discussion

This framework utilises various evaluation metrics that are often used in medical imaging studies to examine the HA-CNN model. The evaluation metrics are accuracy, precision, recall, and F1 score. In this section, the performance of HA-CNN is evaluated against existing models, namely 2D-CNN, VGGNet, and RNN. Mathematically, the evaluation metrics are written as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \times 100 \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \times 100 \tag{11}$$

$$F1\ Score\ = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100 \tag{12}$$

where TP denotes true positive, TN denotes true negative, FP denotes false positive, and FN denotes false negative. Figure 2 shows a sample of classification results generated by the HA-CNN model.

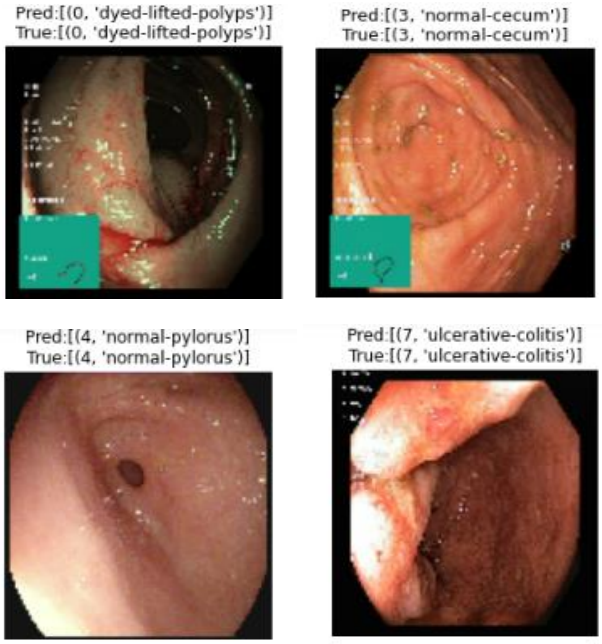

Fig. 2. Classification result.

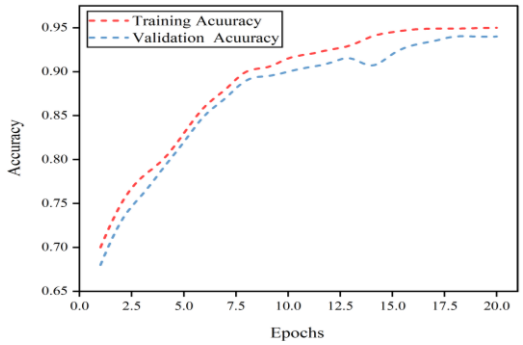

Fig. 3. Accuracy graph.

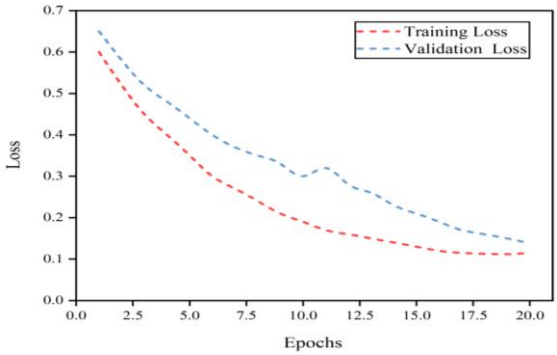

Fig. 4. Loss graph.

Table 1: Performance evaluation of HA-CNN.

Classes	Accuracy	Precision	Recall	F1 score
Dyed lifted polyps	93%	95%	92.7%	93.8%
Dyed resection margins	93%	84.2%	93%	88.4%
Esophagitis	93%	98.1%	92.9%	95.4%
Normal cecum	98%	97.1%	98%	97.5%
Normal pylorus	98%	100%	98.1%	99%
Normal z line	92	98.8%	89.1%	93.7%
Polyps	91%	89.4%	92.5%	90.9%
Ulcerative colitis	95%	93.9%	94.6%	94.3%

Table 1 reports the classification results of the HA-CNN model on different GI diseases. Figure 3 represents the training and validation accuracy obtained by HA-CNN, whereas Figure 4 presents the training and validation loss of the HA-CNN model. The obtained results show that the accuracy of the HA-CNN model was achieved at 20 training epochs by attaining a testing accuracy of 94%. As illustrated in Figure 4, the loss of HA-CNN significantly decreases with an increasing number of epochs.

Figure 5 illustrates the confusion matrix of HA-CNN. The proposed model achieved a mean classification accuracy of 94.43% ± 0.58 across five cross-validation folds. The confusion matrix demonstrates that the HA-CNN model accurately classified GI diseases and significantly reduced the misclassification rate. The evaluation of multiple deep learning models was analysed to visualise the effectiveness and superiority of the HA-CNN model. The proposed Hybrid Attention Convolutional Neural Network was evaluated against existing models, namely 2D-CNN, VGGNet, and RNN. Figure 6 shows that the proposed HA-CNN achieved the highest mean classification accuracy of 94.43% ± 0.58, whereas the RNN model achieved the second-highest mean accuracy of 91.33% ± 0.74. In comparison, VGGNet and 2D-CNN obtained lower mean accuracies of 89.72% ± 0.81 and 87.21% ± 0.93, respectively. Figure 7 demonstrates that HA-CNN achieved superior mean precision and recall values of 93% ± 0.61 and 92% ± 0.65, outperforming the comparative models. Similarly, Figure 8 indicates that HA-CNN attained the highest mean F1 score of 92% ± 0.63, confirming the robustness and consistency of the proposed approach across cross-validation folds.

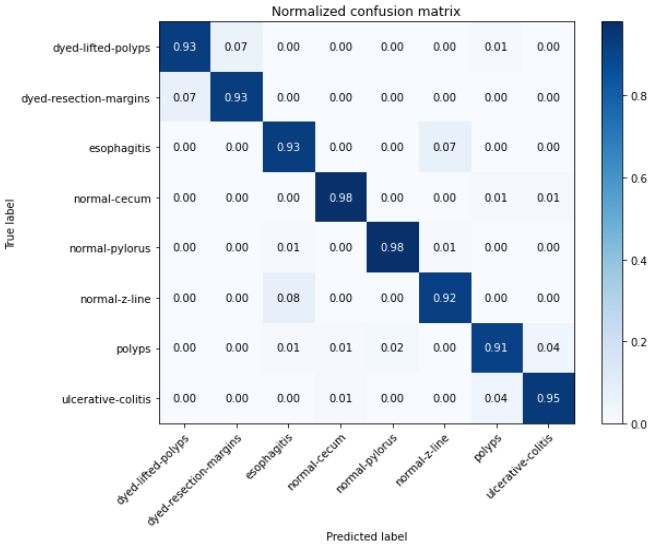

Fig. 5. Confusion matrix.

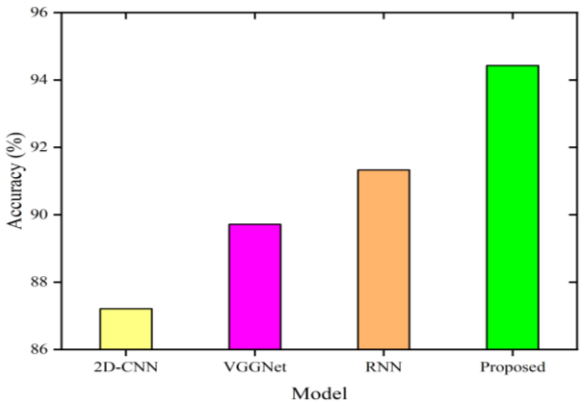

Fig. 6. Model accuracy comparison graph.

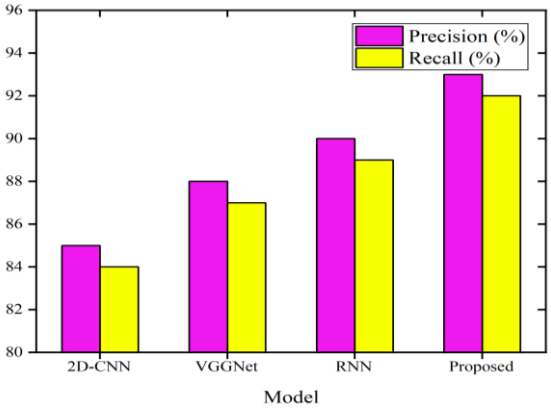

Fig. 7. Model precision and recall comparison graph.

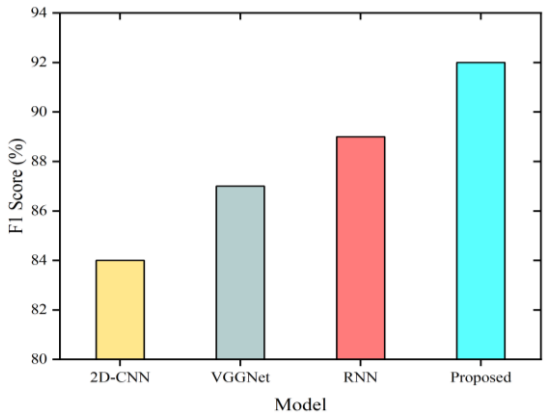

Fig. 8. Model F1 score comparison graph.

Table 2 illustrates the ablation study performed to assess whether lesion localization structurally improves lesion performance under three experimental configurations.

Table 2: Ablation results for segmentation–classification coupling.

Experimental Configuration	Accuracy	Precision	Recall	F1 Score
Classification Only	92.87%	91.9%	91.2%	91.5%
Sequential Segmentation + Classification	93.61%	92.5%	92%	92.2%
Proposed Joint Multi-Task Framework	94.43%	93%	92%	92%

It is observed that the proposed Joint Multi-Task approach achieves the highest mean accuracy across cross-validation folds, confirming that segmentation–classification coupling structurally enhances discriminative learning rather than functioning as an independent pre-processing step.

Table 3 illustrates the assessment of transfer learning adaptation. An additional ablation was conducted using ResNet-50 under two training configurations.

Table 3: Ablation results for backbone adaptation.

Experimental Configuration	Accuracy	Precision	Recall	F1 Score
Frozen ResNet-50	93.61%	92.4%	91.89%	92.1%
Partially Fine-Tuned ResNet-50	94.43%	93%	92%	92%

This analysis demonstrates that backbone adaptation is important when transferring pre-trained ImageNet features to gastrointestinal endoscopic images.

Table 4 reports the assessment of HA-CNN and previous studies on GI disease classification using the Kvasir dataset. The objective of this framework is to identify GI disease more precisely. For analysis, the Kvasir dataset is used. The experimental results are examined using standard evaluation metrics. The class-wise classification

accuracy and the precision and recall scores achieved demonstrate the effectiveness of HA-CNN.

Table 4: Accuracy comparison of proposed model with other models.

Authors	Methods	Accuracy
Huo et al. [IX]	Hierarchical multi-scale feature fusion network	86.12%
D. Gupta et al. [VII]	Voting Classifier	88.18%
İ. Korkmaz et al. [XII]	VGG19	88.6%
Patel et al. [XIX]	EfficientNetB5	92.58%
A.A. Demirbaş et al. [IV]	Spatial-Attention ConvMixer (SAC)	93.37%
D. Mukhtorov et al. [XVI]	ResNet-152 combined with Grad–CAM	93.46%
V.Y. Cambay et al. [III]	DFE based on ResNet50*	94.08%
A. Kamble et al. [XI]	EfficientNetB3	94.25%
Proposed Study	Hybrid Attention Convolutional Neural Network	94.43%

IV. Conclusion

Traditional GI disease diagnosis from endoscopy images is a challenging task; hence, it requires an automated disease diagnosis framework to classify diseases accurately. In this research, a hybrid network is proposed to precisely classify GI diseases. This study primarily aims to enhance the image quality by applying the empirical mode decomposition technique to denoise the image, as these images are affected by motion blur, optical distortions, and other artefacts. Moreover, data augmentation is carried out to improve the generalisation ability of HAA-CNN. Moreover, the attention U-Net segmentation technique localises the lesion regions effectively. Under a 5-fold cross-validation protocol, the model achieves a mean accuracy of 94.43% ± 0.58, demonstrating improved robustness and reduced misclassification compared to existing approaches. Although this model accurately classified GI diseases, it was tested on a single dataset. Future work will aim to implement this model on multiple datasets and validate its performance.

Conflicts of Interest:

The authors report no financial or any other conflicts of interest in this work.

References

I. Abbaszadegan MR, Tavasoli A, Velayati A, Sima HR, Vosooghinia H, Farzadnia M, Asadzedeh H, Gholamin M, Dadkhah E, Aarabi A. (2007). Stool-based DNA testing, a new noninvasive method for colorectal cancer screening, the first report from Iran. World J Gastroenterol. 13(10):1528-1533. 10.3748/wjg.v13.i10.1528.

II. Ajitha Gladis KP, Roja Ramani D, Mohana Suganthi N, Linu Babu. (2024). Gastrointestinal tract disease detection via deep learning based structural and statistical features optimized hexa-classification model. Technol Health Care. 32(6):4453-4473. doi: 10.3233/THC-240603.

III. Cambay VY, Barua PD, Hafeez Baig A, Dogan S, Baygin M, Tuncer T, Acharya UR. (2024). Automated Detection of Gastrointestinal Diseases Using Resnet50*-Based Explainable Deep Feature Engineering Model with Endoscopy Images. Sensors. 24. 10.3390/s24237710

IV. Demirbaş AA, Üzen H, Fırat H. (2024). Spatial-attention ConvMixer architecture for classification and detection of gastrointestinal diseases using the Kvasir dataset. Health Inf. Sci. Syst.12(32). 10.1007/s13755-024-00290-x

V. D'Souza N, Brzezicki A, Abulafi M. (2019). Faecal immunochemical testing in general practice. Br J Gen Pract. 69(679): 60-61. 10.3399/bjgp19X700853.

VI. Gad E, Soliman S, Saeed Darweesh M. (2023). Advancing Brain Tumor Segmentation via Attention-Based 3D U-Net Architecture and Digital Image Processing. 12th International Conference on Model and Data Engineering At: Sousse, Tunisia. 10.1007/978-3-031-49333-1_18.

VII. Gupta D, Anand G, Kirar P, Meel P. (2022). Classification of Endoscopic Images and Identification of Gastrointestinal diseases. 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India. 231-235. doi: 10.1109/COM-IT-CON54601.2022.9850571.

VIII. Hong SM, Baek DH. (2023). A Review of Colonoscopy in Intestinal Diseases. Diagnostics. 13(7):1262. 10.3390/diagnostics13071262

IX. Huo X, Sun G, Tian S, Wang Y, Yu L, Long J, Zhang W, Li A. (2024). HiFuse: Hierarchical multi-scale feature fusion network for medical image classification. Biomed. Signal Process. Control. 87.

X. Jin D, Wen X, Wen Y. (2024). Personalized learning efficiency data analysis based on multi-scale convolution architecture and hybrid loss. Neural Comput & Applic. 36: 9753–9766. 10.1007/s00521-023-09099-3

XI. Kamble A, Bandodkar V, Dharmadhikary S, Anand V, Sanki PK, Wu MX, Jana B. (2025). Enhanced Multi-Class Classification of Gastrointestinal Endoscopic Images with Interpretable Deep Learning Model. arXiv:2503.00780v1.

XII. Korkmaz İ., Soygazi F. (2024). Gastrointestinal Image Classification Using Deep Learning Architectures via Transfer Learning. 2024 Medical Technologies Congress (TIPTEKNO), Mugla, Turkiye. 1-4. 10.1109/TIPTEKNO63488.2024.10755310.

XIII. Li X, Lei L, Sun Y, Li M, Kuang G. (2020). Multimodal Bilinear Fusion Network With Second-Order Attention-Based Channel Selection for Land Cover Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 13: 1011-1026. 10.1109/JSTARS.2020.2975252.

XIV. Mang T, Graser A, Schima W, Maier A. (2007). CT colonography: Techniques, indications, findings. European Journal of Radiology. 61(3): 388-399. 10.1016/j.ejrad.2006.11.019

XV. Medical Advisory Secretariat. (2009). Flexible sigmoidoscopy for colorectal cancer screening: an evidence-based analysis. Ont Health Technol Assess Ser. 9(11):1-23.

XVI. Mukhtorov D, Rakhmonova M, Muksimova S, Cho Y-I. (2023). Endoscopic Image Classification Based on Explainable Deep Learning. Sensors. 23(6). 10.3390/s23063176

XVII. Nouman Noor M, Nazir M, Khan SA, Song O-Y, Ashraf I. (2023). Efficient Gastrointestinal Disease Classification Using Pretrained Deep Convolutional Neural Network. Electronics. 12(7):1557. 10.3390/electronics12071557

XVIII. Pal A, Rai HM, Frej MBH, Razaque A. (2024). Advanced Segmentation of Gastrointestinal (GI) Cancer Disease Using a Novel U-MaskNet Model. Life (Basel), 14(11). doi: 10.3390/life14111488.

XIX. Patel V, Patel K, Goel P, Shah M. (2024). Classification of Gastrointestinal Diseases from Endoscopic Images Using Convolutional Neural Network with Transfer Learning. In Proceedings of the 2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Triunelveli, India. 504–508.

XX. Raut V, Gunjan R, Shete VV, Eknath UD. (2023). Gastrointestinal tract disease segmentation and classification in wireless capsule endoscopy using intelligent deep learning model. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. 11(3): 606–622. 10.1080/21681163.2022.2099298

XXI. Rubab S, Jamshed M, Khan MA, Almujally NA, Damaševičius R, Hussain A, Han N, Nam Y. (2025). Gastrointestinal tract disease classification from wireless capsule endoscopy images based on deep learning information fusion and Newton Raphson controlled marine predator algorithm. Scientific Reports. 15(32180). 10.1038/s41598-025-17204-w

XXII. Saba A, Amin J, Ali MU. (2025). Deep Q-Learning for Gastrointestinal Disease Detection and Classification. Bioengineering. 12(11). 10.3390/bioengineering12111184

XXIII. Shekokar NM, Vasudevan H, Durbha SS, Michalas A, Nagarhalli TP. (2023). Intelligent Approaches to Cyber Security (1st ed.). Chapman and Hall/CRC. 10.1201/9781003408307

XXIV. Wang C, Wang Y, Liu Y, He Z, He R, Sun Z. (2020). ScleraSegNet: An Attention Assisted U-Net Model for Accurate Sclera Segmentation. IEEE Transactions on Biometrics, Behavior and Identity Science. 2(1): 40-54. 10.1109/TBIOM.2019.2962190.

XXV. Xu C, Shu J, Wang Z, Wang J. (2024). A Scene Classification Model Based on Global-Local Features and Attention in Lie Group Space. Remote Sensing. 16(13):2323. 10.3390/rs16132323

XXVI. Yang J, Park K. (2024). Improving Gait Analysis Techniques with Markerless Pose Estimation Based on Smartphone Location. Bioengineering. 11(2). 10.3390/bioengineering11020141

XXVII. Yogapriya J, Chandran V, Sumithra MG, Anitha P, Jenopaul P, Suresh Gnana Dhas C. (2021). Gastrointestinal Tract Disease Classification from Wireless Endoscopy Images Using Pretrained Deep Learning Model. Comput Math Methods Med. 10.1155/2021/5940433.

XXVIII. Zubair Rahman AMJMD, Mythili R, Chokkanathan K, Mahesh TR, Vanitha K, Yimer TE. (2024). Enhancing image-based diagnosis of gastrointestinal tract diseases through deep learning with EfficientNet and advanced data augmentation techniques. BMC Med Imaging. 24(306). 10.1186/s12880-024-01479-y