RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt2547

Development of an Automated Endoscopic Image Segmentation Technique for Accurate Localization of Gastrointestinal Lesions

T. Swetha Kumari [a,*], R. Vasuki [a]

[a] Department of Department of Biomedical Engineering, Bharath Institute of Higher Education and Research, Chennai, India.

* Corresponding Author Email: swetha837@gmail.com

DOI: https://doi.org/10.54392/irjmt2547

Received: 21-03-2025; Revised: 28-06-2025; Accepted: 07-07-2025; Published: 14-07-2025

Abstract: Automatic detection of gastrointestinal (GI) lesions makes endoscopic diagnosis more accurate, but it is difficult because the lesions can look different. A deep learning (DL) method for specific classification makes things more accurate, but for clinical use, a strong model is needed to make sure the location is correct. This study comes up with a better way to divide things up and solves these problems. The goal is to make a custom partition frame with a Residual Swin Transformer Fusion Network (RN-RSTFN) that has been fine-tuned with a Refined Nutcracker. Gaussian Filtering (GF) was used to reduce noise in the images and make them clearer. Then, Z-Score was used to standardize the distribution of pixel intensity. Function Extraction using a histogram of oriented gradients (HOG) helps to get the wound pattern needed for better partition performance. Give the right location for GI ulcers. Endoscopic image datasets that are available to the public from depot and medical institutions include pictures of different GI lesions. Gaussian Filtering (GF) was used to reduce noise in the images and make them clearer. Then, Z-Score was used to standardize the distribution of pixel intensity. The RN-RSTFN that was suggested combines the hierarchical representation and residual learning of the Swin Transformer to improve the boundaries of lesions. When measured using the Dice score and mIOU, Precision, and Recall measures, the RN-RSTFN model showed a lot of improvement in finding GI lesions. The scores were 0.9458, 0.9383, 0.9634, and 0.9596, respectively.

Keywords: Endoscopic, Automatic gastrointestinal (GI), Image Segmentation, Refined Nutcracker-tuned Residual Swin Transformer Fusion Network (RN-RSTFN)

1. Introduction

Gastrointestinal (GI) lesions cause tissue changes in the digestive system that aren't normal, such as tumors, spores, wounds, and areas of inflammation. These wounds can be a sign of many different diseases, from mild inflammation to cancerous tumors like colorectal carcinoma. Finding these deviations early and accurately is important for effective medical intervention. A late diagnosis can have serious effects, like intestinal barriers, cancer progression, and possibly malignant internal bleeding [1]. To stop and treat diseases like Crohn's disease and gastric ulcers, which haven't been treated, can cause long-term health problems and GI lesions. The main way to find GI lesions is through a traditional endoscopic exam, which uses an elastic snake (endoscope) and a camera to take pictures of the patient's digestive system in real time. Manual lesion diagnosis is hard because lesions can look different, mucous, blood, or food residue can make it hard to see them, and the accuracy of the diagnosis depends on the operator [2]. Also, the diagnosis can change because finding lesions is random and depends on the skill of the endoscopist. Also, it might be hard to tell the difference between small or flat tumors and healthy tissue, which means that a diagnosis could be missed. To improve clinical decision-making, decrease human error in endoscopic evaluations, and improve lesion localization, automated, AI-driven procedures must be developed. Automated segmentation can detect GI lesions more accurately, reliably, and efficiently by combining DL and image processing, which will ultimately improve patient outcomes. GI disorder diagnosis requires endoscopic image analysis, yet the subjectivity of assessment combined with the endoscopist's expertise limits its diagnostic effectiveness [3]. Studies indicate there exist both missed lesions and wrong diagnoses because practitioner capability varies in detecting lesions. Manual discovery for GI lesions becomes difficult because their complex nature produces hidden places between small gear abnormities and folds. Researchers focus attention on developing automated image processing equipment that will provide exact objective endoscopic image interpretation to solve existing problems. Progress leads to more uniform medical diagnosis with low dependence on clinical expertise, resulting in the patient's results through rapid specific disease identity [4].

Medical image analysis requires image sharing as an essential stage so that images can be divided into separate areas depending on the properties capable of detecting appropriate wounds. The medical image segmentation technique depends on handmade threshold procedures and craft functions and still shows limited success in complex imaging conditions. Following the emergence of Deep Learning (DL) technology, the performance of the partition was dramatically improved as the Convolution Neural Network (CNN) and transformer -based designs that can reveal complex information patterns from the wider database. Robotic fragmented wounds give them better medical results by distinguishing them from other tissues, and thus doctors help to make more accurate diagnosis of GI lesions when planning treatment. Recent progress of U-Net and attention-based models triggers DL to create a valuable method for clinical purposes in improving the results of the division [5]. The task of segmenting GI lesions presents multiple obstacles to DL despite its capability in medical imaging. The difficulty in lesion segmentation arises from the wide variety of abnormal appearances that present with unique colors, shapes and textures that limit model generalization power. The integration of mucus and blood as well as food particles along with motion artifacts and light variations causes the obtainment of substandard endoscopic images in routine clinical practices [6]. The distribution of lesions across training data is unequal because some rare types appear infrequently, which leads to skewed predictive outcomes. The short supply of databased medical images for segmentation validation remains one of the main barriers limiting the development of accurate segmentation technologies because it requires expert gastroenterologist involvement [7]. Modern approaches that enhance model adaptability along with data assortment solutions and robust performance across different clinical applications must be designed to resolve these problems [8].

1.1 Research gaps

The current success of DL for GI lesion segmentation remains limited by critical unresolved issues. Inappropriately, inability to use different datasets constitutes the leading disadvantage. The training datasets that built existing models contain insufficient variations to adequately capture the complexity of GI lesions during regular medical procedures. The implementation of these systems faces considerable difficulties when used globally because differences in lesion appearance combined with endoscope types and patient demographics lead to substantial performance changes [9]. A major challenge emerges from the inability to properly process photos of poor quality. Uncorrected problem segmentation usually occurs when endoscopic images contain motion blur and the presence of food particles or mucus alongside irregular lighting conditions [10]. The current DL models face challenges when identifying lesions correctly in such situations and thus increase the risk of wrong diagnoses [11]. The development of better lesion detection methods for unreliable imaging situations represents the solution to this problem. Real-time processing speed limitations serve as a significant impediment to clinical use. Several DL models are not suitable for real-time endoscopic treatments because of their excessive computational requirements and slow processing times. Standard medical practice accepts AI-based integration only when combined with precise segmentation accuracy together with fast inference capabilities [12]. The process of training and validating DL models faces constraints because of minimal access to high-quality datasets. A team of expert gastroenterologists is required for GI lesion annotation because this procedure requires expensive resources and takes an extended time. The data imbalance produces class imbalance problems because certain lesion types receive insufficient representation, thus affecting model prediction accuracy [13].

1.2 Objective of this Research

Developing an advanced automatic segmentation structure to accurately locate GI lesions in endoscopic images is the main objective of this research. To progress lesion boundary refinement, the research suggests an optimized segmentation model called the Refined Nutcracker-tuned Residual Swin Transformer Fusion Network (RN-RSTFN), which combines residual learning with the hierarchical representation capabilities of Swin Transformers. By integrating DL-based transformers and feature fusion algorithms [14], this research contributes to the creation of a clinically practical AI-assisted diagnostic tool, increasing real-time GI lesion diagnosis and endoscopic image interpretation [15].

1.3 Contributions of this research

- The RN-RSTFN model was developed in this research to improve lesion localization in GI tract images and increase the accuracy of endoscopic image segmentation.
- In data preprocessing, the Gaussian Filter (GF) was used to smooth GI lesion images and minimize noise for improved segmentation performance, and Z-score normalization was applied to standardize image intensities and ensure uniform input distribution.
- In feature extraction, to improve the approach's capacity to distinguish between ordinary with aberrant GI regions, texture and edge details were captured using the Histogram of Oriented Gradients (HOG).

The following indicates that the research was conducted: The literature review is presented in section 2, the proposed technique is developed in section 3, the findings and discussions are presented in section 4, and the conclusion is illustrated in section 5.

2. Literature Review

The goals, datasets, methodology, and limitations of the current AI-driven systems for GI lesion diagnosis are compiled in Table 1, which also highlights the need for more generalizability and clinical application.

Table 1. Comparative analysis of AI-based methods for GI lesion detection and classification

No. of Ref	Objective	Dataset type	Method	Limitation
[16]	To improve Upper GI Endoscopy (UGIE) by assuring physical completeness and aiding in the identification and categorization of stomach lesions.	It uses UGIE video footage with neoplastic tumors and precancerous situations in the stomach.	UGIE movies are analyzed by AI-based DL models to classify lesions, find anomalies, and locate blind spots.	The existing AI models for UGIE architectures are not interpretable beyond simple saliency techniques and have few task-specific adjustments.
[17]	To reduce reliance on expertise, assist endoscopists in early diagnosis, and enhance the identification of minor stomach lesions.	It consists of endoscopic images that show GI.	To improve the detection of tiny lesions, it incorporates attention processes into YOLOv7.	Its generalizability to other datasets and actual clinical settings is limited because it was trained on a private dataset.
[18]	Create and verify AI models for gastric mucosal lesions utilizing CNNs for lesion identification, differential diagnosis (AI-DDx), and invasion depth (AI-ID).	It includes, as a reference standard, endoscopic images of 1366 individuals with histologic diagnoses.	AI-DDx and AI-ID models were trained and validated on different datasets, and their performance was evaluated with different degrees of experience.	The models could not translate well to a variety of clinical contexts because it was tested on a small dataset.
[19]	It evaluated the effectiveness of several NN performed in classifying Early-stage Gastric Cancer (EGC).	Images of 161 patients with early-stage cancer and 58 patients with gastritis were acquired.	It suggested a way to find the lesion by turning the network's output value into a caloric value, and the algorithm was enhanced by applying the concepts of transfer learning and fine-tuning.	The model's generalizability could be limited by the comparatively small dataset it used.
[20]	To create a real-time segmentation model that improves accuracy in a variety of color modes for identifying Gastric Intestinal Metaplasia (GIM) in endoscopic images.	It consists of 1239 non-GIM photos and 940 histologically verified GIM images.	A model based on a Bilateral Segmentation Network (BiSeNet) with an auxiliary head and extra loss functions, employing preprocessing methods, such as label smoothing, jigsaw augmentation, and location-wise negative sampling.	Its performance could differ depending on the endoscopic device and the clinical context, and it could have trouble with extremely complicated lesions and loud labeling.
[21]	Examine and contrast AI techniques for identifying upper GIcancers, improving CAD systems for both retrospective and real-time analysis.	65 investigations with a range of models, validation techniques, image counts, and endoscopic modalities.	While CNN-based supervised learning models predominate in DL approaches, Support Vector Machines (SVM) is commonly utilized in Machine Learning (ML)-based GI image processing.	For high-quality research, interdisciplinary collaboration is necessary because clinicians lack technical AI skills.
[22]	Create an automated method that uses DL and optimization techniques to identify stomach lesions in endoscopic images.	Endoscopic images that include both intra-class and inter-class variability in lesion types.	Deep Belief Networks (DBNs) extract important features for classification; GA-based morphological operations optimize	Lesion size and shape variability, DBN training challenges, and the requirement for big

			segmentation, and Maximally Stable Extremal Regions (MSER) segment images.	datasets for optimal performance are among the obstacles.
[23]	Create a DL automated method to divide up healthy GI tract organs so that patients with GI cancer can receive more accurate radiation treatment.	UW-Madison's GI Tract Image Segmentation collection, which includes training and testing medical images.	DenseNet201 and InceptionResNetv2 are combined in ensemble learning to increase segmentation accuracy.	For expanded clinical relevance, the model could need more validation on a variety of datasets and a lot of processing resources.
[24]	Create a framework that combines segmentation and thresholding to accurately extract Gastric Polyps (GP) from endoscopic images.	Four datasets of benchmark endoscopic images were used.	It combines automated and semi-automated segmentation approaches and applies tri-level thresholding using the Aquila Optimization Algorithm.	To effectively generalize, the method could require a lot of processing power and adjustment for various dataset variances.
[25]	Improve the diagnosis of GI diseases with the use of AI, enhanced imaging, and controlled motility in Video Capsule Endoscopy (VCE).	Analysis of the use of VCE in the verdict of esophageal/colon diseases, Crohn's disease, celiac disease, malignancies, and GIhemorrhage.	It enables focused navigation and diagnostics through onboard/externally controlled locomotion, enhanced imaging for better viewing, and AI for automated diagnosis.	There are still issues with real-time image processing, localization, and broad clinical acceptance, and locomotion control is in the experimental stage.
[26]	Establish GINet, a DL-based system, to increase the precision, efficiency, and economy of diagnosing GI disorders.	Images from wireless capsule endoscopy (3658) address GI disorders, such as ulcers, GI hemorrhage, lymphangiectasia, and Angi ectasia.	With the use of GradCAM and Guided-GradCAM for lesion localization and domain adaption to lessen bias, a 13-layer CNN (GINet) was trained and evaluated for the classification of GI diseases.	Reliance on high-quality endoscopic images and the requirement for additional validation in various clinical settings is potential obstacles.
[27]	Implement an automatic EGC diagnosis technique based on DL.	Annotated endoscopic images were gathered from a single healthcare facility.	A U-Net model with attention and multi-scale feature extraction for segmentation, and a guided-attention deep network based on ResNet-50 for EGC classification.	Lower IoU scores could result from the U-Net-based segmentation model's inability to precisely define small or irregular tumor areas.
[28]	Construct a DL system that uses White-Light Endoscopy (WLE) to identify stomach lesions and forecast neoplasms.	Endoscopic images and videos, comprising both prospective and retrospective data were collected.	Using archived movies, a DL-based system was trained on WLE images, verified both internally and externally, and contrasted with skilled endoscopists.	False positive results (3.04 ± 3.04 per gastroscopy) could result in unnecessary follow-ups.
[29]	Develop a precise and effective DL model that uses Wireless Capsule Endoscopy (WCE) images to classify GI tract illnesses in real-time.	Unbalanced GI tract images can be found in the Kvasir-Capsule image dataset.	It suggested GI Disease Detection Network (GIDD-Net), a customized CNN model with fewer parameters for the low-cost identification of GI diseases.	Additional optimization could be necessary to handle real-time analysis with massive patient data.
[30]	To improve the identification of gastric cancer, create a model that can recognize anatomical areas in esophagogastroduodenoscopy images.	2,054 endoscopic images of 13 different stomach areas from 96 individuals.	Improve the accuracy of stomach region categorization by utilizing the ConvNeXT architecture with information fusion and depth map estimation.	A small sample size could affect the ability to generalize to larger populations.

3. Methodology

For precise lesion segmentation, this research employs a multi-step methodology using endoscopic GI lesion imaging data. Using Z-score normalization for standardization and a GF for noise reduction are two examples of preprocessing. The RN-RSTFN model is used for segmentation, guaranteeing accurate lesion location, and the HOG is used for feature extraction. Figure 1 represents the research flow.

3.1 Data collection

The GI tract Kvasir SEG (KSEG) images used for illness identification are part of the publicly accessible Kvasir dataset [31]. Three clinically relevant findings, two categories about polyp removal, and images of three anatomical landmarks are all included. To guarantee correctness, skilled endoscopists annotated the dataset. Kvasir-SEG Dataset Splitting Strategy clearly state how the 800 Kvasir-SEG images were divided into training, validation, and independent test cohorts. Include: Exact Split Ratios "The 800 Kvasir-SEG images were randomly split into training, validation, and test sets with a ratio of 70%, 15%, and 15% respectively."

Cross-Validation used cross-validation, specify the number of folds and how the data was partitioned in each fold. It proposes to fund studies in automatic reporting, classification, and computer-aided disease detection. Researchers using AI and multimedia can use the dataset to create automated endoscope evaluation tools that will increase healthcare productivity. Kvasir encourages improvements in GI disease identification and classification while addressing the dearth of repeatable medical datasets. Figure 2 represents the image samples in the dataset.

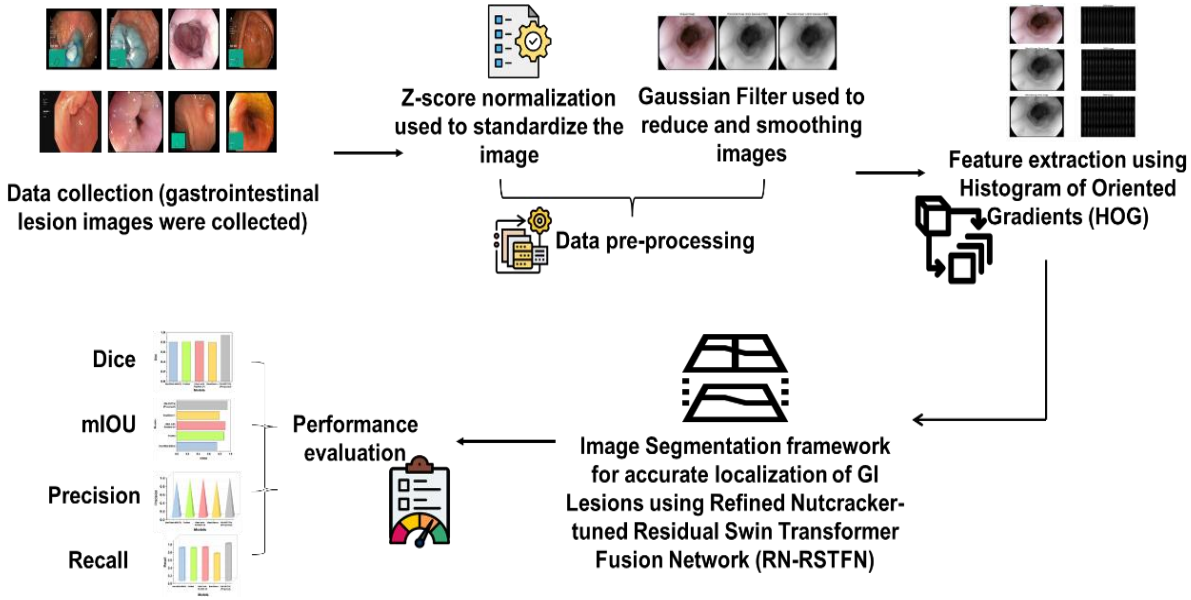

Figure 1. AI-Based GI Lesion Detection and Classification Processes

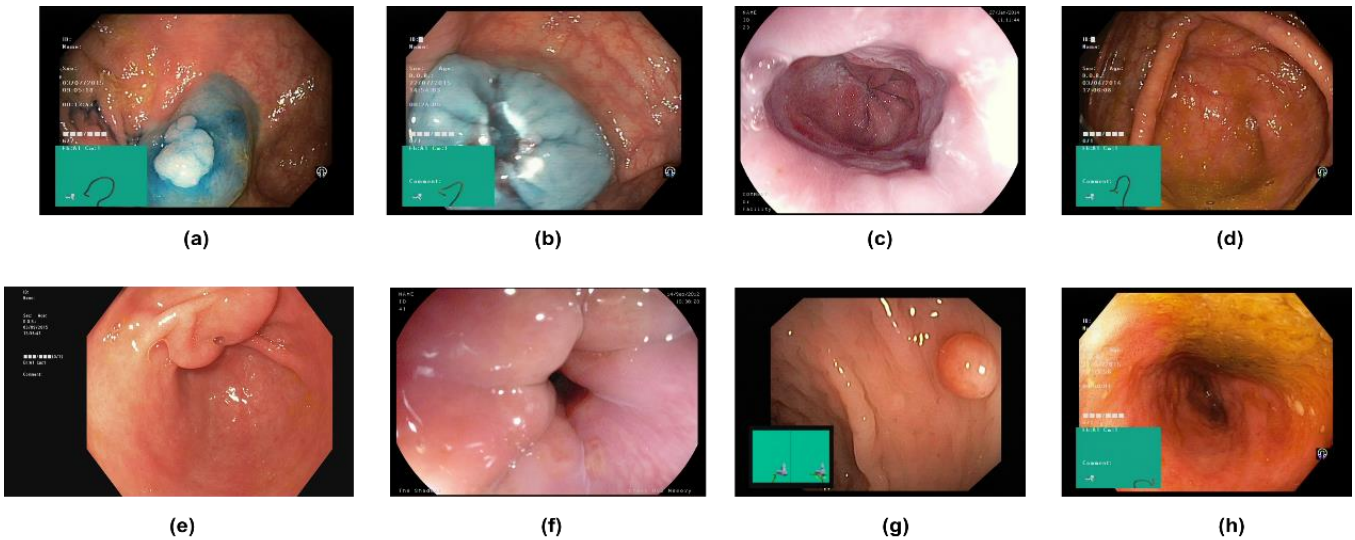

Figure 2. Sample images of **(a)**-Dyed lifted polyps, **(b)**-Dyed resection margins, **(c)**-Esophagitis, (**d)**-Normal cecum, **(e)**-Normal pylorus, **(f)**-Normal z-line, **(g)**-Polyps, **(h)**-Ulcerative colitis

3.2 Data preprocessing

The preprocessing stage of the data guarantees reliable and consistent input for the segmentation of GI lesions. Z-score normalization is used for standardization, which improves feature consistency across photos by reducing variances in brightness and contrast. By eliminating distortions and maintaining crucial structural information, GF improves segmentation accuracy by lowering noise and smoothing lesion boundaries.

3.2.1 Reducing noise and smoothing GI lesion images using Gaussian Filter (GF)

The research utilizes GF to refine GI lesion images while it filters additional noise. The combination of sensor limitations and motion disturbances coupled with uneven lighting produces unwanted noise that covers up lesion boundaries in endoscopic images. The imaging of lesions benefits from GF because it reduces noise at high frequencies without damaging structural elements. The smoothing procedure helps the model focus on essential lesion characteristics rather than image artifacts, therefore enhancing segmentation results. The GF enhances the source image's details and removes noise. This filter lessens the brightness variation surrounding a pixel. The entire contour of the image is smoothed out (Figure 3). While the elliptical GF could improve features by specifying σ_w and σ_z separately, the circular GF ($\sigma_w = \sigma_z$) is commonly used for preprocessing images for smoothness and noise reduction. Equation (1) defines the filtering function.

$$H(w,z:\sigma) = \exp\{\frac{w^2}{\sigma_w^2} - \frac{z^2}{\sigma_z^2}\} \quad (1)$$

Equation (2) demonstrates that the Gaussian initial derivation filtering is a variation of the second derivative of the GF filtering method, $\sigma_w = \sigma_z$.

$$sdg(w,z:\emptyset) = \frac{(w_\emptyset^2-\sigma_\emptyset^2)(z_\emptyset^2-\sigma_\emptyset^2)}{2\pi\sigma_\emptyset^{10}}\exp[-\frac{w_\emptyset^2+z_\emptyset^2}{\sigma_\emptyset^2}] \quad (2)$$

To varied degrees, the GF with KSEGD images can make ridges bigger and valleys smaller. Change the cosine value as shown in Equation (3) to increase the filter's effectiveness or directional sensitivity.

$$sdg(w,z:\emptyset,e) = \frac{(w_\emptyset^2-\sigma_\emptyset^2)(z_\emptyset^2-\sigma_\emptyset^2)}{2\pi\sigma_\emptyset^{10}}\exp[-\frac{w_\emptyset^2+z_\emptyset^2}{\sigma_\emptyset^2}]\cos(2\pi e w_\emptyset) \quad (3)$$

The constraints of $\emptyset, \sigma_\emptyset, w_\emptyset$, $and z_\emptyset$. In form, the modified GF is quite similar to the widely used Gabor filter.

3.2.2 Standardization using Z-score normalization

The KSEG data is analyzed using Z-score normalization in this research to standardize pixel intensity distributions among images. Model performance could be impacted by the brightness and contrast variations in endoscopic images caused by various acquisition situations. The Z-score normalization process guarantees consistent intensity levels by transforming pixel data to have a standard deviation (σ) of 1 and a mean (μ) of 0, as presented in Equation (4). The model's capacity to generalize across various GI lesions is improved by this standardization, which also lessens biases and increases feature consistency.

$$I'(x,y) = \frac{I(x,y)-\mu}{\sigma} \quad (4)$$

3.3 Feature extraction using Histogram of Oriented Gradients (HOG)

HOG feature extraction is used to improve GI lesion segmentation after preprocessing the KSEGD. By examining gradient orientations, HOG highlights form and texture characteristics that are vital for differentiating lesions in KSEGD images. By strengthening edge detection and maintaining structural details, this technique enhances segmentation performance. By extracting strong feature representations, HOG helps the model differentiate between normal and pathological GI zones. The HOG feature descriptor states that the density distribution of gradients can be used to characterize the local object's look and shape inside an image. By segmenting the image into tiny areas known as cells, this description can be implemented. For each pixel inside the cell, a gradient direction histogram is created.

Figure 3. Preprocessed KSEG image of **(a)** Original, **(b)** Processed in 5mm, **(c)** Processed in 10mm

The HOG approach extracts the item in four phases. The first step is to compute the rise standards using a 1-D arrangement to find the gradient of the derived cover in both horizontal and vertical directions, as indicated in Equations (5 & 6).

$$C_w = [-1\ 0\ 1] \quad (5)$$

$$C_z = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad (6)$$

If the object image is J, it could use the convolution method to get the x and y derivatives as defined in Equation (7):

$$J_w = J_w * C_w \quad \text{and} \quad J_z = J_z * C_z \quad (7)$$

Equations (8&9) are used to determine the gradient's magnitude.

$$|H| = \sqrt{J_w^2 + J_z^2} \quad (8)$$

$$\theta = \arctan \frac{J_z}{J_w} \quad (9)$$

Spatial orientation binning is the next phase. The purpose of this stage is to provide the cell histogram result through a voting process. A weighted vote for orientation is cast by each pixel in the texture image, according to the nearest bin between 0 and 180 degrees. Using the L2 norm, the final step, block normalization, is carried out using Equation (10):

$$a = \frac{a}{\sqrt{||a||^2 + \varepsilon^2}} \quad (10)$$

The windows descriptor is required to gather descriptors from every block and convert them into vector form following the HOG normalization method. Figure 4 represents the feature-extracted image of KSEGD.

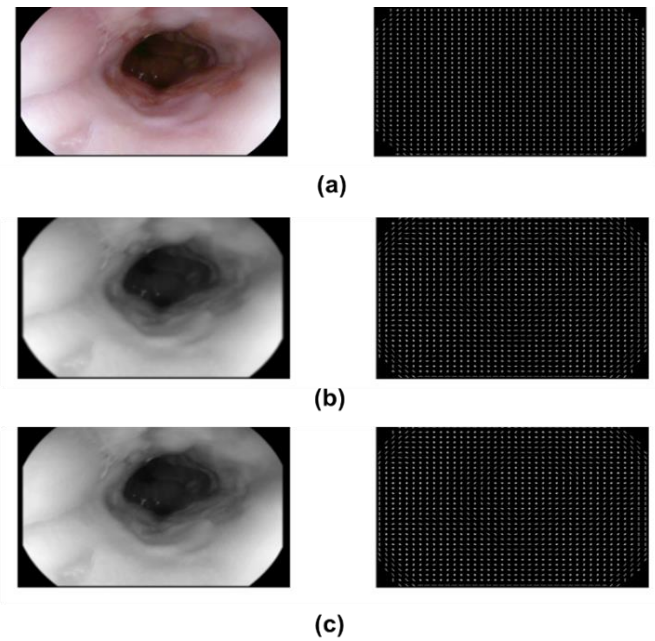

Figure 4. Extracted image analysis of **(a)** Original, **(b)** Filtered 10mm image, **(c)** Filtered 5mm image

Following the extraction of HOG features, gradient-based feature maps that depict the crucial shape and texture aspects of GI lesions will be present in the processed KSEGD. The RN-RSTFN will be trained using these extracted features to achieve precise lesion segmentation. With improved lesion representations in the dataset, the model will be better able to discriminate among typical and pathological GI areas.

3.4 Image Segmentation framework for accurate localization of GI Lesions using Refined Nutcracker-tuned Residual Swin Transformer Fusion Network (RN-RSTFN)

The image segmentation framework implementing RN-RSTFN operates post-HOG feature extraction because it performs precise localization for GI lesions in the KSEGD. Feature fusion is improved and lesion border refinement occurs in KSEGD images through the implementation of RN-RSTFN, which integrates residual learning with Swin Transformer hierarchical representation. An accurate segmentation results from the network, which utilizes the gradient-based feature maps from KSEGD as inputs for acquiring both local and global contextual information. The adjustment of model parameters in the RN system strengthens segmentation performance for diverse types of lesions in KSEGD imaging. The advanced system enables medical professionals to make accurate endoscopic diagnoses through highly precise GI lesion detection within KSEGD.

3.4.1 Residual Swin Transformer Fusion Network (RN-RSTFN)

The RSTFN is used for precise GI lesion segmentation in the KSEGD following the feature extraction procedure. While residual learning maintains crucial lesion characteristics and stops gradient vanishing, the Swin Transformer records multi-scale spatial dependencies. To improve segmentation accuracy, the fusion process combines global contextual information with local fine-grained features. By enhancing lesion border refinement, the network's hierarchical attention mechanism guarantees precise localization. RSTFN helps with clinical diagnostics by achieving excellent segmentation accuracy and resilience in KSEGD images through Nutcracker tune optimization.The three primary components of Swin Fuse are feature reconstruction, a fusion layer, and global feature extraction. Equations (11- 13) defines the initial features $\Phi^K \in Q^{G \times X \times D}$. Where the m^{th} RSTB is denoted by G_{RSTB_n}. The global characteristics of visible and infrared images are retrieved using these methods.

$$\Phi^K = G_{Pos}(J^s) \quad (11)$$

$$\Phi_{GF}^K = G_{RSTB_n}(\Phi_{SV}^K) \quad (12)$$

$$\Phi_E = G_{Norm}(\Phi_{GF}^K) \quad (13)$$

Where the fusion operation is indicated by G_{Norm}. Lastly, it transforms the fused global features from $Q^{GM\times D}$. Equation (14), which defines a fusion image J_E, is reconstructed using a convolutional layer and $Q^{G\times M\times D}$.

$$J_E = G_{Conv}(\Phi_E) \quad (14)$$

Where the feature reconstruction is denoted by G_{Conv}. This convolutional layer has a Tanh activation layer, a $1x1$ kernel, and a padding of 0.

3.4.1.1 Residual Swin Transformer Block (RSTB)

The construction of the RSTB, which consists of several Swin Transformer Blocks (STLs) and a residual connection, is shown in Figure 5(a). Equation (15) calculates the final output of RSTB after applying mSTL to obtain overall traits that are intermediate $\Phi^K_{n,m-1}$ given the input sequence vectors $\Phi^K_{n,0}$.

$$\Phi^K_{n,m} = G_{STLn,m}\cdot\left(\Phi^K_{n,m-1}\right) + \Phi^K_{n,0} \quad (15)$$

Where the m^{th} Swin Transformer layer is indicated by $G_{STLn,m}$. The multi-layer Swin Transform design is capable of efficiently modeling comprehensive structures, and the enduring assembly can aggregate the various feature levels, much like the CNN architecture. The Swin Transformer Layer (STL), seen in Figure 5 (b). Equation (16) calculates the matrices $R, L, and U$ for the local window feature Φ_y.

$$R = \Phi_y X_R, L = \Phi_y, X_l, U = \Phi_y, X_U \quad (16)$$

Where d is the dimension of (R, L) and X_R, X_l, and $X_U \in Q^{M^2\times c}$ are the learnable parameters of 3 layers of linear projection that are shared across many windows. In the meantime, Equation (17) formulates the categorization environments of the self-attention process.

$$\text{Attention(R, L, U)} = \text{Sof tM ax}\left(\frac{RL^S}{\sqrt{c}} + o\right) \quad (17)$$

3.4.1.2 Fusion strategy

The fusion layer is shown in Figure 6, uses columns and row vector extents to assess the activity level of the visible and infrared image sequence matrices and develops a new fusion method based on the K_1norm. It first computes their row vector weights by the K_1 norm for their respective global features, referred to as $\Phi^K_{GF}(j,i) and \Phi^{vis}_{GF}(j,i)$. Then, it uses softmax to determine their activity level, which is represented by Equations (18 & 19) and is denoted as $\varphi^{jq}_{row}(j)$ and$\varphi^{vis}_{row}(j)\varphi^{jq}_{row}(j)$.

$$\varphi^{jq}_{row}(j) = \frac{\exp(|\left|\Phi^{jq}_{GF}(j)\right|_1)}{\exp(|\left|\Phi^{jq}_{GF}(j)\right|_1)+exp\exp(|\left|\Phi^{vis}_{GF}(j)\right|_1)} \quad (18)$$

$$\varphi^{vis}_{row}(j) = \frac{\exp(|\left|\Phi^{vis}_{GF}(j)\right|_1)}{\exp(|\left|\Phi^{jq}_{GF}(j)\right|_1)+exp(|\left|\Phi^{vis}_{GF}(j)\right|_1)} \quad (19)$$

where the K_1 norm computation is shown by $||.||_1$. Equation (20) is utilized to formulate this.

$$\varphi^{row}_E(j,i) = \varphi^{jq}_{row}(j,i)\times\varphi^{jq}_E(j) + \Phi^{vis}_{GF}(j,i)\times\varphi^{vis}_E(j) \quad (20)$$

Then assess their activity level from the column vector dimension, which is called $\varphi^{jq}_{col}(i) and \varphi^{vis}_{col}(i)$, and is expressed by Equations (21 & 22). This process is identical to the operations mentioned previously.

$$\varphi^{jq}_{col}(i) = \frac{\exp(|\left|\Phi^{jq}_{GF}(i)\right|_1)}{\exp(|\left|\Phi^{jq}_{GF}(i)\right|_1)+\exp(|\left|\Phi^{vis}_{GF}(i)\right|_1)} \quad (21)$$

$$\varphi^{vis}_{col}(i) = \frac{\exp(|\left|\Phi^{vis}_{GF}(i)\right|_1)}{\exp(|\left|\Phi^{jq}_{GF}(i)\right|_1)+\exp(|\left|\Phi^{vis}_{GF}(i)\right|_1)} \quad (22)$$

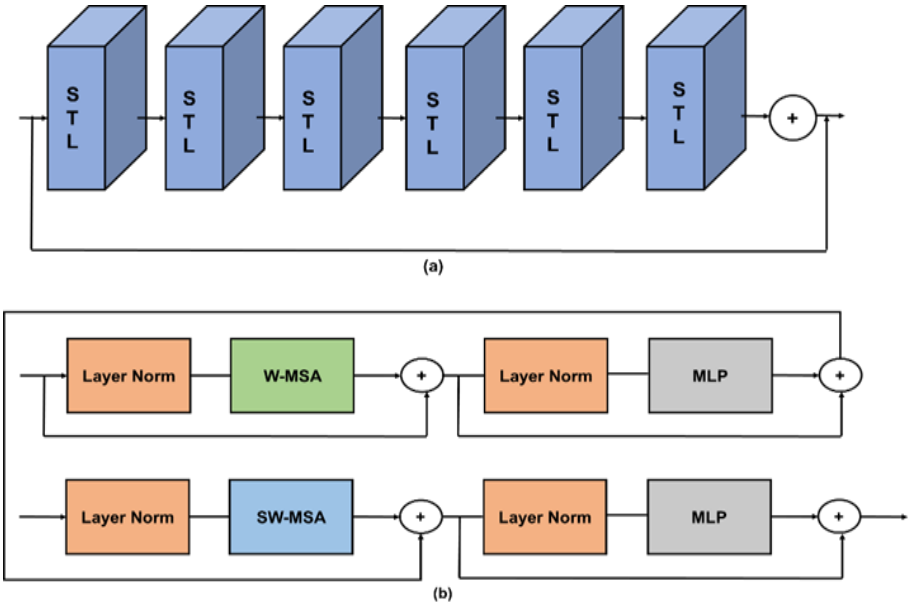

Figure 5. Analysis of (a) RSTB and (b) STL

After that, it could get the attached comprehensive structures with support vector measurement, which are defined by Equations (23& 24) and are called $\varphi_E^{col}(j,i)$.

$$\Phi_E^{col}(j,i) = \varphi_{GF}^{jq}(j,i) \times \varphi_{col}^{jq}(i) + \Phi_{GF}^{vis}(j,i) \times \varphi_{GF}^{vis}(i) \quad (23)$$

$$\varphi_E(j,i) = \Phi_E^{row}(j,i) + \Phi_E^{col}(j,i) \quad (24)$$

The assembled image is reconstructed by a convolutional layer using the final fused global features. It is crucial that the fusion layer is only destroyed during training and retained during testing.

3.4.2 Refined Nutcracker Optimization

By adjusting the hyperparameters of RN-RSTFN, the RNO improves GI lesion segmentation in KSEGD. To increase segmentation accuracy and convergence speed, it improves transformer block layouts, learning rates, and attention weights. RNO improves hierarchical attention methods to guarantee balanced feature extraction. For accurate identification of lesion boundaries, it reduces loss functions. Overall, RNO improves segmentation robustness, which supports accurate clinical diagnosis.

3.4.2.1 Initialization and evaluation

The suggested approach, including population-based algorithms, uses Equation (25) to randomly initialize M solutions $\vec{W}_j(j \in M)$ in the search space of the optimization problem, where each solution has dimensions that need to be optimized.

$$\vec{W}_j = \vec{K} + (\vec{V} - \vec{K}).\vec{q} \quad (25)$$

Where $\vec{q}$ is a vector between 0 and 1 that is created at random. Since the multi-thresholding image segmentation problem is discrete and these initialization solutions are continuous, it cannot be applied to it. To make these solutions pertinent to this issue, they will be transformed into discrete solutions, as explained below. The 8-bit representation of the greyscale level has minimum values ($\vec{K}$) of 0 and maximum values ($\vec{V}$) of 2^8-1=255. Suppose the individual have ten different objects (L) in an image that need to be separated by determining the threshold values that are close to the ideal.

3.4.2.2 Convergence Improvement Strategy (CIS)

This technique is intended to reduce the problem of local minima, while accelerating the NOA's rate of convergence. There are three main parts to the strategy. It is conceivable that in comparison to the best solution thus far, the randomly chosen option could be closer to the near-optimal solution. The goal is to avoid becoming stuck in local minima and to speed up convergence. Equations (26&27) provide a mathematical formulation of CIS characteristics.

$$\vec{H}_j(s=1) = \vec{W}_j(s) + (q * (\vec{u}_1) + (1-q) * (\vec{u}_2) \quad (26)$$

$$\vec{u}_1 = \begin{cases} \vec{W}_j^s - \vec{W}_b^s, if e(\vec{W}_j^s) \le e(\vec{W}_b^s) \\ \vec{W}_b^s - \vec{W}_j^s, \text{otherwise} \end{cases} \quad (27)$$

where $b /= a /= d /= j.q$ are three randomly chosen numbers from the solutions. A random number between 0 and 1 is called q. Equations (28&29), respectively, formulate the second and third folds to avoid becoming trapped in local minima and to speed up the conjunction speed in the track of the best-so-far explanation (see equation (30&31).

$$\vec{u}_2 = \begin{cases} \vec{W}_a^s - \vec{W}_d^s, if e(\vec{W}_a^s) \le e(\vec{W}_d^s) \\ \vec{W}_d^s - \vec{W}_a^s, \text{otherwise} \end{cases} \quad (28)$$

$$\vec{W}_j(s+1) = \frac{(\vec{W}_j(s)+\vec{W}_b(s)+\vec{W}_{best})}{3} + \beta * \left(-\vec{W}_{best}^s + \vec{W}_j(s)\right) + q_1 * (\vec{W}_a^s - \vec{W}_d^s) \quad (29)$$

$$\vec{W}_j(s+1) = \frac{(\vec{W}_j(s)+\vec{W}_b(s)+\vec{W}_{best})}{3} + q * \left(\vec{W}_{best}^s + \vec{W}_j(s)\right) + \vec{V} + q_1 * \left(\vec{W}_a^s - \vec{W}_d^s\right) * \vec{V}_1 \quad (30)$$

$$\beta = \frac{1}{1+f^s} \quad (31)$$

where l is a variable that includes values that decrease linearly from -1 to -2, and σ is a constant value that is decided later in the research sector. Binary vectors $\vec{V}$ and $\vec{V}_1$ are produced using Equations (32 &33), respectively.

$$\vec{V} = \begin{cases} 1\ \vec{q}_2 \ge \vec{q}_3 \\ 0, otherwise \end{cases} \quad (32)$$

$$\vec{V}_1 = \begin{cases} 1\ \vec{q}_2 \ge q \\ 0, otherwise \end{cases} \quad (33)$$

To determine the tradeoff between these three folds, utilize Equation (34).

$$\vec{W}_j(s+1) \begin{cases} \text{q.}(52) \text{ if q } < \delta \\ E.(53) if\, q_1 < \delta_1 \\ \text{q.}(49) \text{ Else} \end{cases} \quad (34)$$

Where the trials' controlling parameters, δ and δ_1, are used to calculate the likelihood of each fold. To capture spatial and contextual data, the RN-RSTFN uses Swin Transformer blocks with residual connections to extract features (Algorithm 1). The information obtained from KSEGD includes accurately segmented GI lesion regions with clearly defined boundaries, lesion size, shape, and localization characteristics following the Refined Nutcracker-tuned RN-RSTFN segmentation procedure. In addition to offering improved visual outputs for additional clinical assessment and diagnosis, this data aids in the comprehension of lesion distribution patterns.

Algorithm 1. Refined Nutcracker-tuned Residual Swin Transformer Fusion Network (RN-RSTFN)

$importtorch$

$importtorch.nnasnn$

$\#\ SimpleSwinTransformerBlock$

$classSwinBlock(nn.Module)$:

$def\ \_init\_(self, dim)$:

$super().\_init\_()$

$self.attn = nn.MultiheadAttention(embed\_dim = dim, num\_heads = 4)$

$self.mlp = nn.Sequential(nn.Linear(dim, 4 * dim), nn.GELU(), nn.Linear(4 * dim, dim))$

$Defforward\ (self, x)$:

$x = x + self.attn(x, x, x)[0]\ \#\ ResidualSelf - Attention$

$returnx + self.map(x)\ \#\ ResidualMLP$

$\#\ RN - RSTFNModel$

$classRN - RSTFN(nn.Module)$:

$def\ \_init\_(self, dim = 96, blocks = 3)$:

$super().\_init\_()$

$self.layers = nn.ModuleList([SwinBlock(dim)for\_nrange(blocks)])$

$defforward(self, x)$:

$forlayerinself.layers$:

$x = layer(x)$

$returnx$

$model = RN - RSTFN()$

$input\_tensor = torch.rand(10, 96)\ \#\ DummyInput$

$output = model(input\_tensor)$

$print(output.shape)$

4. Results and Discussion

The effectiveness of the RN-RSTFN model in segmenting GI lesions is assessed in this section. Windows 7, an Intel i7-7700 CPU, a GeForce GTX 960 GPU, 16GB of RAM, and 512GB of storage make up the experiment's setup.

- Perform Repeated Experiments: Rerun your experiments on the Kvasir-SEG dataset using either:
- At least five random splits: Divide your dataset into training, validation, and test sets at least five times, with each split using a different random seed.
- K-fold cross-validation: Typically, 5-fold or 10-fold cross-validation is used. This involves dividing the dataset into K equally sized folds, training the model K times (each time using K-1 folds for training and one fold for testing/validation), and averaging the results.
- Present Mean ± Standard Deviation (SD): After running these multiple experiments, calculate the mean and standard deviation for each reported metric (e.g., Dice, mIoU, etc.).
- Update Results Section: In your manuscript's "Results" section, replace the single-point values with the mean ± SD. For example, instead of "Dice 0.9458," you would report "Dice: 0.9458 ± 0.0025."

4.1 Model evaluation and classification performance

The training loss and validation accuracy trends for the RN-RSTFN model throughout epochs are displayed in Figure 6. A consistent decrease in training loss is shown in Figure 6(a), indicating successful learning and RN-RSTFN convergence. The model's strong generalization is demonstrated by Figure 6(b), which displays validation accuracy rising quickly and stabilizing at 90%. These outcomes validate the effectiveness of the RN-RSTFN model in maximizing classification accuracy and loss reduction.

The confusion matrix in Figure 7 shows the classification performance of the RN-RSTFN model in five categories: Normal, Esophagitis, Polyp, Ulcerative Colitis, and Others. Correctly classified cases are indicated by diagonal values, and misclassified cases are indicated by off-diagonal values. The high diagonal values demonstrate the RN-RSTFN model's great performance, particularly in detecting normal, esophagitis, and polyps. There are certain incorrect categories, particularly when it comes to differentiating Ulcerative Colitis from other kinds.

The RN-RSTFN model's capacity to discriminate across classes is assessed by the ROC-AUC curve (Figure 8(a)). The AUC shows the strength of the classification, and each curve represents a class. All classes have AUC values that are almost 1.0, indicating strong discrimination ability. The RN-RSTFN model's strong classification performance is confirmed by AUC values of 0.98 for ulcerative colitis and 0.99 for the majority of classes.

For every RN-RSTFN class, the precision-recall curve calculates the effect within recall with precision (Figure 8(b)). High recall and accuracy scores indicate that the RN-RSTFN model successfully captures genuine positives while lowering false positives. The curves show great prediction accuracy, since it stays high across a range of memory levels. Minor difficulties in borderline classifications, but overall RN-RSTFN high reliability, are indicated by a rapid drop-off near the end.

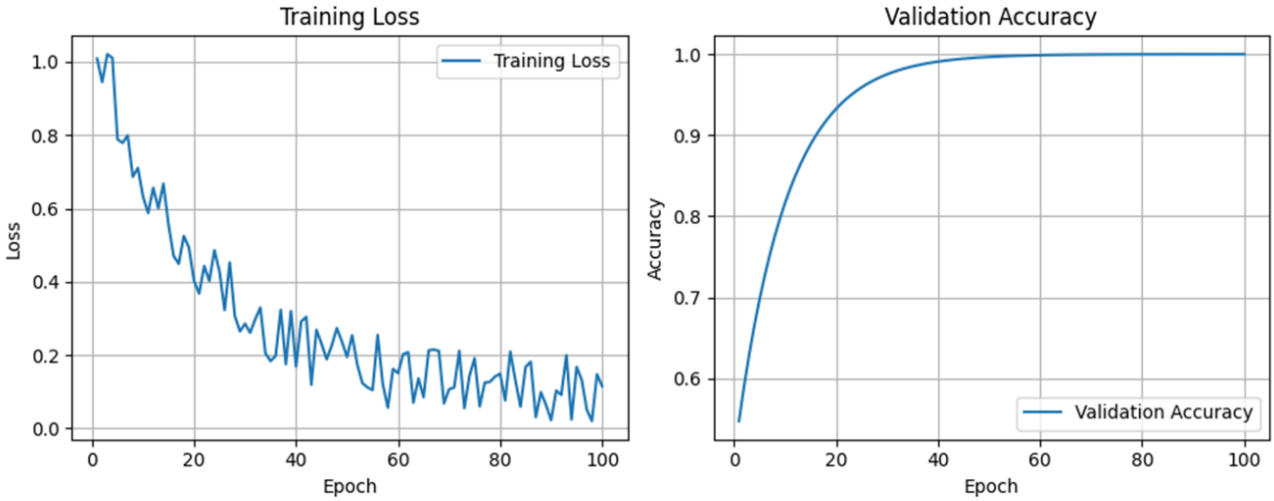

Figure 6. Evaluation of (a) Training loss and (b) Validation accuracy

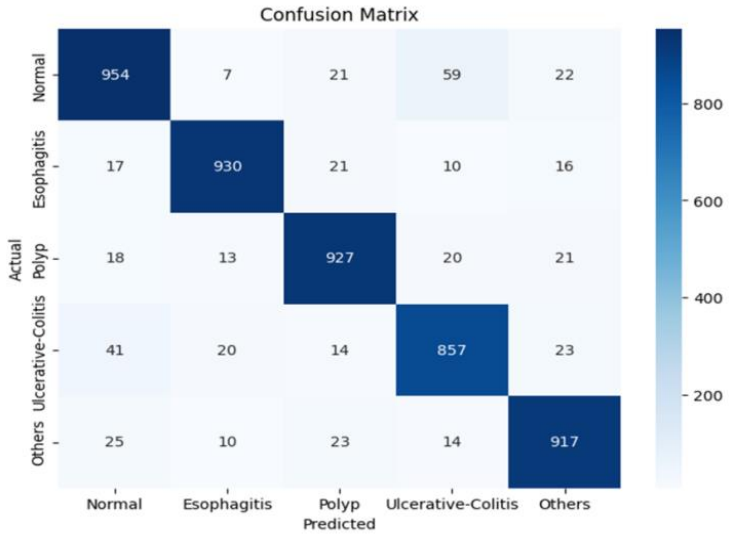

Figure 7. Confusion matrix evaluation

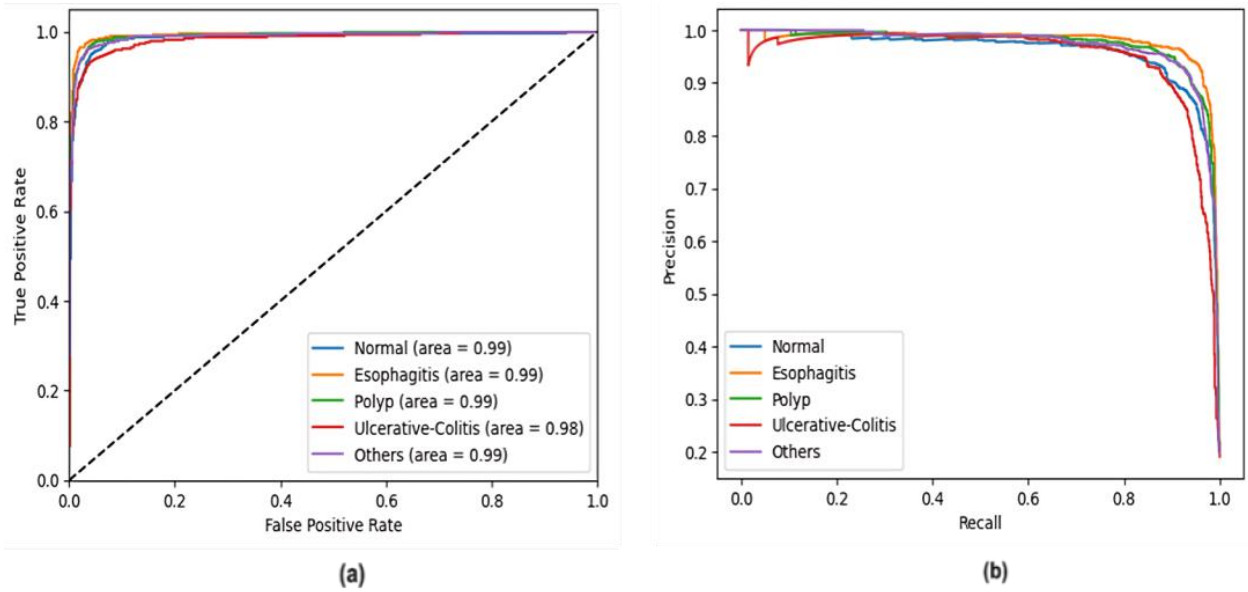

Figure 8. Analysis of **(a)** ROC-AUC and **(b)** precision-recall curve

4.2 Comparative Analysis

The suggested RN-RSTFN technique was contrasted with several other cutting-edge segmentation techniques, such as Harmonic Densely Connected Network for Medical SEGmentation (HarDNet-MSEG) [26], Parallel Reverse Attention Network (PraNet) [26], U-Net with Residual Network-34 as Backbone (UNet with ResNet-34) [26], and Residual U-Net++ (ResUNet++) [27] to precisely segment GI lesions in endoscopic images. Dice, mIOU, Precision, and Recall were among the performance indicators employed in the assessment.

Dice: Frequently referred to as an overlap evaluation, this is the most commonly used metric for assessing and determining how well medical image segmentation works as represented in Equation (35).

$$Dice = \frac{2*M}{2*M+N+O} \quad (35)$$

An important metric for assessing segmentation accuracy in GIlesion diagnosis is the Dice score, which calculates the overlap in the forecast with ground truth segmentations. Better model performance is shown (Figure 9) by a higher Dice score; the suggested RN-RSTFN achieved 0.9458, which is much better than current models and guarantees more accurate lesion localization.

mIOU: The overall count of pixels in both images are divided by the doubled overlap zone between the ground truth and the predicted segmented image. Medical segmentation is typically checked using mIOU, or mean intersection over union, as shown in Equation (36). The IOU is calculated by dividing the total number of pixels by the predicted segmentation overlap over the ground truth.

$$mIOU = \frac{M}{M+N+O} \quad (36)$$

To ensure accurate border delineation in GI imaging, the mIOU metric measures the overlap between anticipated and actual lesion locations. Greater segmentation quality is indicated by higher mIOU values; the suggested RN-RSTFN achieved 0.9383, which is far better than the current models for improved lesion localization presented in Figure 10.

Precision: The attribute of accuracy is known as a precision measure. It evaluates the precision of these research forecasts, which are shown in Equation (37).

$$Precision = \frac{M}{M+N} \quad (37)$$

Precision minimizes false positives for an accurate diagnosis by calculating the percentage of accurately diagnosed lesion pixels among all expected lesion pixels. With an RN-RSTFN of 0.9634, which surpasses current models and guarantees more accurate localization of GIlesions, a higher precision denotes better model reliability, as shown in Figure 11.

Recall: It is a metric that is used to determine the positive ground truth points that a model predicts using Equations (38).

$$Recall = \frac{M}{M+O} \quad (38)$$

A comparison of several DL models for automated GI lesion segmentation is shown in Table 2. It shows that the suggested RN-RSTFN model performs noticeably better than current methods in terms of Dice score, mIOU, precision, and recall, guaranteeing more precise lesion localization. Recall evaluates whether the model can detect all pertinent GI lesions, reducing false negatives and guaranteeing that no important areas are overlooked. Improved lesion localization, which is necessary for a precise diagnosis, is indicated by a higher recall.

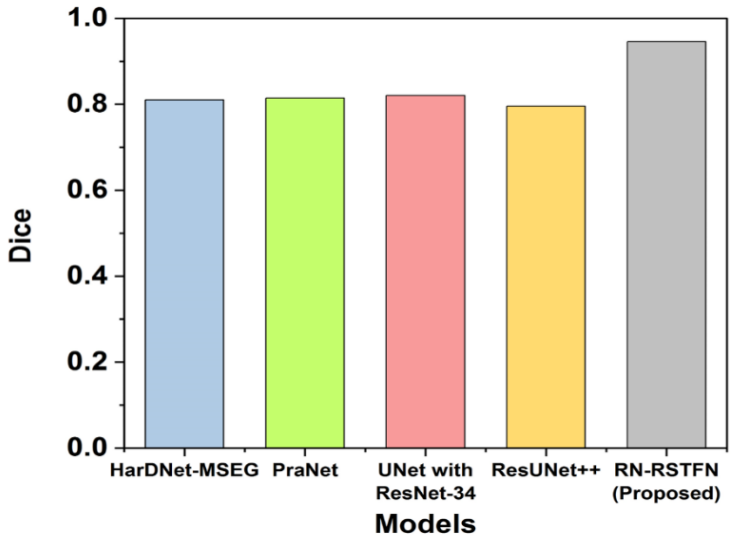

Figure 9. Comparing Dice Scores for Models of GI Lesion Segmentation

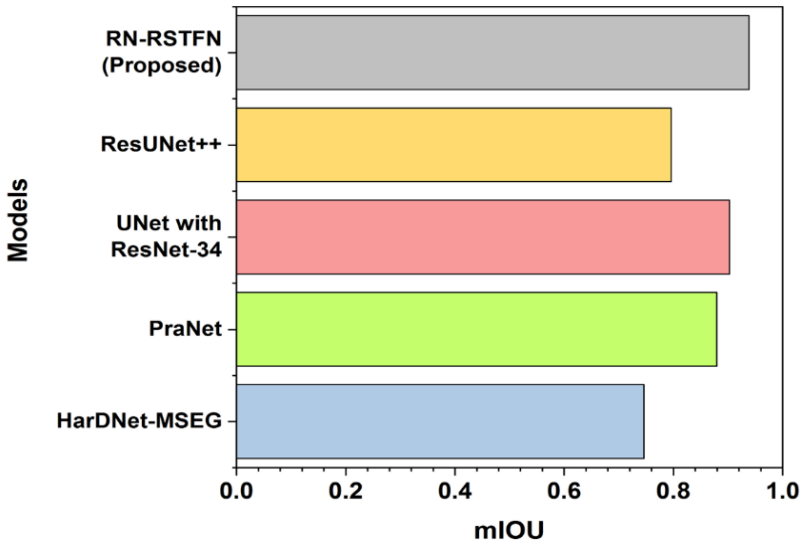

Figure 10. Comparing Mean Intersection over Union (mIOU) for Segmenting GI Lesion

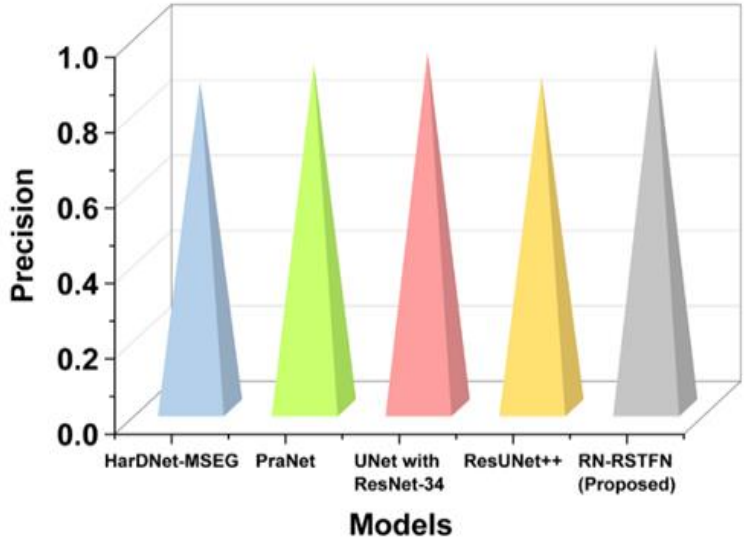

Figure 11. Accurate GI Lesion Detection by Precision Evaluation

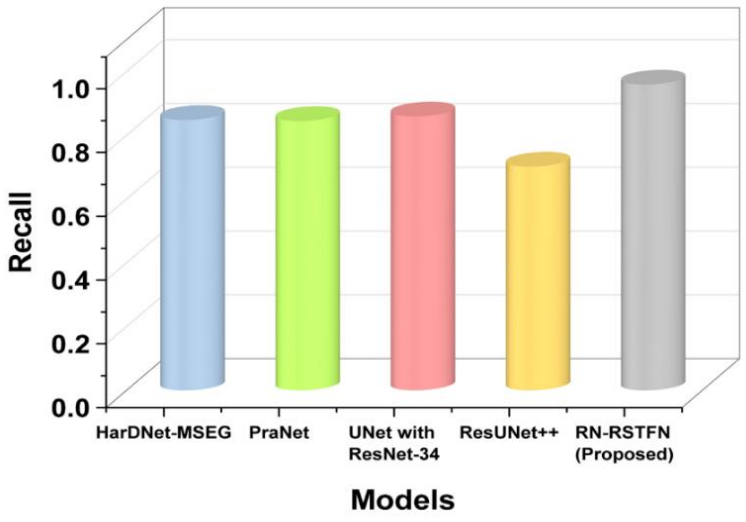

Figure 12. Evaluation of Recall for Detailed GI Lesion Identification

Table 2. Performance comparison of GI Lesion segmentation model

Models	Dice	mIOU	Precision	Recall
HarDNet-MSEG [26]	0.8102	0.7459	0.8652	0.8485
PraNet [26]	0.8142	0.8796	0.9126	0.8453
UNet with ResNet-34 [26]	0.8208	0.9030	0.9435	0.8597
ResUNet++ [27]	0.7955	0.7962	0.8785	0.7022
RN-RSTFN (Proposed)	0.9458	0.9383	0.9634	0.9596

With a sensitivity of 0.9596 (Figure 12), the RN-RSTFN model outperforms other approaches and shows superiority in lesion segmentation.

Developing an automated endoscopic image segmentation method for accurate GI lesion location is the aim of this research. By utilizing the RN-RSTFN model and refining feature extraction methods, the research seeks to increase segmentation accuracy. Due to its high computational complexity, HarDNet-MSEG [32] is not as well suited for segmenting GI lesions in real-time. It could mis-classify lesion boundaries, since it has trouble catching fine-grained information in endoscopic images. Complex GI lesion boundaries are difficult for PraNet [33] to handle, which results in inaccurate segmentation of irregularly shaped lesions. Using low-contrast endoscopic images hampers the detection capability of subtle lesion locations by this system [34]. The segmentation outcomes of UNet with ResNet-34 [35] decrease when GI lesions exhibit major shape and textural inconsistencies. The skip connections in this system prevent essential background information from being eliminated and thus reduce its segmentation precision. Safety of GI lesion identification proves to be a difficult task for ResUNet++ [36] because this model fails to discriminate between overlapping structures combined with irregular boundaries [37]. Real-time endoscopic applications do not benefit from using this deep model because it comes with an increased computational expense [38]. Research findings introduce RN-RSTFN as a suitable solution to these challenges because it enhances GI lesion segmentation accuracy through specialized border detection [39]. The model uses both spatial attention techniques together with improved feature extraction to enhance segmentation results while reducing unneeded feature propagation [40].

5. Conclusion

To develop automated endoscopic image segmentation that would provide accurate GI lesion placement was the aim of this research. The proposed RN-RSTFN model for medical imaging purposes enhances boundary detection of lesions while minimizing erroneous segmentations to provide superior accuracy results. The experimental outcomes demonstrate that RN-RSTFN achieves better results across all essential evaluation metrics when compared to HarDNet-MSEG, PraNet, and both the ResNet-34-based UNet and ResUNet++ models. The RN-RSTFN model exhibited significant progress in detecting GI lesions when evaluated using Dice score and mIOU, Precision, and Recall measures, which reached 0.9458, 0.9383, 0.9634, and 0.9596 respectively. The model maintains robust performance in detecting GI lesions because its high precision and recall values eliminate false detections while ensuring essential lesions remain identified. The measured data confirms that RN-RSTFN functions exceptionally well as an important DL approach for GI lesion segmentation.

5.1 Limitation and Future Scope

The high accuracy of the RN-RSTFN model presents processing challenges that raise uncertainties about its real-time clinical deployment uncertain due to computing system constraints. Future research should expand the model's diagnostic utility by integrating multi-modal medical imaging while adding real-time segmentation functionality and improving computational efficiency.

References

[1] A.C. Vasconcelos, M. Dinis-Ribeiro, D. Libânio, Endoscopic resection of early gastric cancer and pre-malignant gastric lesions. Cancers, Cancers, 15(12), (2023) 3084. https://doi.org/10.3390/cancers15123084

[2] M. Sekine, T. Asano, H. Mashima, The diagnosis of small gastrointestinal subepithelial lesions by endoscopic ultrasound-guided fine needle aspiration and biopsy. Diagnostics, 12(4), (2022) 810. https://doi.org/10.3390/diagnostics12040810

[3] W. Wang, X. Yang, X. Li, J. Tang, Convolutional-capsule network for gastrointestinal endoscopy image classification. International Journal of Intelligent Systems, 37(9), (2022) 5796-5815. https://doi.org/10.1002/int.22815

[4] S.M. Bitar, M. Moussa, The risk factors for the recurrent upper gastrointestinal hemorrhage

among acute peptic ulcer disease patients in Syria: A prospective cohort study. Annals of Medicine and Surgery, 74, (2022) 103252. https://doi.org/10.1016/j.amsu.2022.103252

[5] Y. Liu, S. Zuo, Self-supervised monocular depth estimation for gastrointestinal endoscopy. Computer Methods and Programs in Biomedicine, 238, (2023) 107619. https://doi.org/10.1016/j.cmpb.2023.107619

[6] R. Zhang, B. Peng, Y. Liu, X. Liu, J. Huang, K. Suzuki, Y. Nakajima, D. Nemoto, K. Togashi, X. Zhu, Localization of Capsule Endoscope in Alimentary Tract by Computer-Aided Analysis of Endoscopic Images. Sensors, 25(3), (2025) 746. https://doi.org/10.3390/s25030746

[7] Y. Okagawa, S. Abe, M. Yamada, I. Oda, and Y. Saito, Artificial intelligence in endoscopy. Digestive diseases and sciences, 67(5), (2022) 1553-1572. https://doi.org/10.1007/s10620-021-07086-z

[8] A. Nagahara, A. Shiotani, K. Iijima, T. Kamada, Y. Fujiwara, K. Kasugai, M. Kato, K. Higuchi, The role of advanced endoscopy in the management of inflammatory digestive diseases (upper gastrointestinal tract). Digestive Endoscopy, 34(1), (2022) 63-72. https://doi.org/10.1111/den.13982

[9] P.H. Conze, G. Andrade-Miranda, V.K. Singh, V. Jaouen, D. Visvikis, Current and emerging trends in medical image segmentation with deep learning. IEEE Transactions on Radiation and Plasma Medical Sciences, 7(6), (2023) 545-569. https://doi.org/10.1109/TRPMS.2023.3265863

[10] F. Renna, M. Martins, A. Neto, A. Cunha, D. Libânio, M. Dinis-Ribeiro, M. Coimbra, Artificial intelligence for upper gastrointestinal endoscopy: a roadmap from technology development to clinical practice. Diagnostics, 12(5), (2022) 1278. https://doi.org/10.3390/diagnostics12051278

[11] S. Ahmad, J.S. Kim, D.K. Park, T. Whangbo, Automated detection of gastric lesions in endoscopic images by leveraging attention-based yolov7. IEEE Access, 11, (2023) 87166-87177. https://doi.org/10.1109/ACCESS.2023.3296710

[12] J.Y. Nam, H.J. Chung, K.S. Choi, H. Lee, T.J. Kim, H. Soh, E. Ae Kang , S.J. Cho, J. Chul Ye, J. Pil Im, S.G. Kim, J.S. Kim, H. Chung, J.H. Lee, Deep learning model for diagnosing gastric mucosal lesions using endoscopic images: development, validation, and method comparison. Gastrointestinal Endoscopy, 95(2), (2022) 258-268. https://doi.org/10.1016/j.gie.2021.08.022

[13] S. Wang, Q. He, P. Zhang, X. Chen, S. Zuo, Toward Automatic Detection of Gastric Lesion for Upper Gastrointestinal Endoscopy with Neural Network. Journal of Medical Robotics Research, 7(01), (2022) 2141003. https://doi.org/10.1142/S2424905X21410038

[14] P. Pornvoraphat, K. Tiankanon, R. Pittayanon, P. Sunthornwetchapong, P. Vateekul, R. Rerknimitr, Real-time gastric intestinal metaplasia diagnosis tailored for bias and noisy-labeled data with multiple endoscopic imaging. Computers in Biology and Medicine, 154, (2023) 106582. https://doi.org/10.1016/j.compbiomed.2023.106582

[15] M. Vania, B.A. Tama, H. Maulahela, S. Lim, Recent advances in applying machine learning and deep learning to detect upper gastrointestinal tract lesions. IEEE Access, 11, (2023) 66544-66567. https://doi.org/10.1109/ACCESS.2023.3290997

[16] M. Alhajlah, Automated lesion detection in gastrointestinal endoscopic images: leveraging deep belief networks and genetic algorithm-based Segmentation. Multimedia Tools and Applications, (2024) 1-15. https://doi.org/10.1007/s11042-024-20439-w

[17] R. Mantri, R.A.H. Khan, D.T. Mane, An Efficient System for Detection and Classification of Acute Lymphoblastic Leukemia Using Semi-Supervised Segmentation Technique. International Research Journal of Multidisciplinary Technovation, 7(2), (2025) 121-134. https://doi.org/10.54392/irjmt25210

[18] V. Rajinikanth, S. Aslam, S. Kadry, O. Thinnukool, Semi/Fully-Automated Segmentation of Gastric-Polyp Using Aquila-Optimization-Algorithm Enhanced Images. Computers, Materials & Continua, 70(2), (2022). http://dx.doi.org/10.32604/cmc.2022.019786

[19] M. Hanscom, D.R. Cave, Endoscopic capsule robot-based diagnosis, navigation and localization in the gastrointestinal tract. Frontiers in Robotics and AI, 9, (2022) 896028. https://doi.org/10.3389/frobt.2022.896028

[20] D. Bajhaiya, S.N. Unni, Deep learning-enabled detection and localization of gastrointestinal diseases using wireless-capsule endoscopic images. Biomedical Signal Processing and Control, 93, (2024) 106125. https://doi.org/10.1016/j.bspc.2024.106125

[21] L. Ma, X. Su, L. Ma, X. Gao, M. Sun, Deep learning for classification and localization of early gastric cancer in endoscopic images. Biomedical Signal Processing and Control, 79, (2023) 104200. https://doi.org/10.1016/j.bspc.2022.104200

[22] L. Wu, M. Xu, X. Jiang, X. He, H. Zhang, Y. Ai, Q. Tong , P. Lv, B. Lu, M. Guo, M. Huang, L. Ye, L. Shen, H.Yu, Real-time artificial intelligence for detecting focal lesions and diagnosing neoplasms of the stomach by white-light endoscopy (with videos). Gastrointestinal Endoscopy, 95(2), (2022) 269-280. https://doi.org/10.1016/j.gie.2021.09.017

[23] S. Mahmood, M.M.S. Fareed, G. Ahmed, F. Dawood, S. Zikria, A. Mostafa, S.F. Jilani, M. Asad, M. Aslam, A robust deep model for classification of peptic ulcer and other digestive tract disorders using endoscopic images. Biomedicines, 10(9), (2022) 2195. https://doi.org/10.3390/biomedicines10092195

[24] D. Bravo, J. Ruano, M. Jaramillo, S. Medina, M. Gómez, F.A. González, E. Romero, (2024) Automatic endoscopy classification by fusing depth estimations and image information. IEEE International Symposium on Biomedical Imaging (ISBI), IEEE, Greece. https://doi.org/10.1109/ISBI56570.2024.10635452

[25] J. Chen, G. Wang, Y. Ding, Z. Zhang, K. Xia, L. Xu, X. Xu, Development of an AI-Assisted System for Automatic Recognition and Localization Marking of Colonic Polyps (With Video). Journal of Gastroenterology and Hepatology, 40(7), (2025) 1797-1808. https://doi.org/10.1111/jgh.16980

[26] M.N. Noor, M. Nazir, S.A. Khan, I. Ashraf, O.Y. Song, Localization and classification of gastrointestinal tract disorders using explainable AI from endoscopic images. Applied Sciences, 13(15), (2023) 9031. https://doi.org/10.3390/app13159031

[27] D. Jha, P.H. Smedsrud, M.A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, H.D. Johansen, (2019). Resunet++: An advanced architecture for medical image segmentation. In 2019 IEEE international symposium on multimedia (ISM), IEEE, USA.

[28] Y. Qin, J. Chang, L. Li, M. Wu, Enhancing gastroenterology with multimodal learning: the role of large language model chatbots in digestive endoscopy. Frontiers in Medicine, 12, (2025) 1583514. https://doi.org/10.3389/fmed.2025.1583514

[29] E.M. El-Gammal, W. El-Shafai, T.E. Taha, A.S. El-Fishawy, F.E. Abd El-Samie, A survey of artificial intelligence models for wireless capsule endoscopy videos for superior automatic diagnosis: problems and solutions. Multimedia Tools and Applications, (2025) 1-35. https://doi.org/10.1007/s11042-024-18300-1

[30] R. Zhang, B. Peng, Y. Liu, X. Liu, J. Huang, K. Suzuki, Y. Nakajima, D. Nemoto, K. Togashi, X. Zhu, Localization of Capsule Endoscope in Alimentary Tract by Computer-Aided Analysis of Endoscopic Images. Sensors, 25(3), (2025) 746. https://doi.org/10.3390/s25030746

[31] Kaggle, (2023) Kvasir-SEG Dataset, Kaggle. Available: https://www.kaggle.com/datasets/tanmaydebnath/kvasir-seg-dataset?select=dyed-lifted-polyps

[32] A. Garbaz, Y. Oukdach, S. Charfi, M. El Ansari, L. Koutti, M. Salihoun, Gsac-uformer: Groupwise self-attention convolutional transformer-based unet for medical image segmentation. Cognitive Computation, 17(69), (2025) 1-14. https://doi.org/10.1007/s12559-025-10425-1

[33] N. Rochmawati, C. Fatichah, B. Amaliah, A.B. Raharjo, F. Dumont, E. Thibaudeau, C. Dumas, Deep Learning-Based Lesion Detection in Endoscopy: A Systematic Literature Review. IEEE Access, 13, (2025) 43532 – 43556. https://doi.org/10.1109/ACCESS.2025.3548167

[34] A. Dhali, V. Kipkorir, R. Maity, B.S. Srichawla, J. Biswas, R.B. Rathna, H.R. Bharadwaj, I. Ongidi, T. Chaudhry, G. Morara, M. Waithaka, C. Rugut, M. Lemashon, I. Cheruiyot, D. Ojuka, S. Ray, G. K. Dhali, Artificial intelligence–assisted capsule endoscopy versus conventional capsule endoscopy for detection of small bowel lesions: a systematic review and meta-analysis. Journal of Gastroenterology and Hepatology, 40(5), (2025) 1105-1118. https://doi.org/10.1111/jgh.16931

[35] X. Guo, L. Pang, P. Chen, Q. Jiang, Y. Zhong, Deep ensemble framework with Bayesian optimization for multi-lesion recognition in capsule endoscopy images. Medical & Biological Engineering & Computing, (2025) 1-16. https://doi.org/10.1007/s11517-025-03380-4

[36] C. Hu, Y. Xia, Z. Zheng, M. Cao, G. Zheng, S. Chen, J. Sun, W. Chen, Q. Zheng, S. Pan, Y. Zhang, J. Chen, P. Yu, J. Xu, J. Xu, Z. Qiu, T. Lin, B. Yun, J. Yao, W. Guo, C. Gao, X. Kong, K. Chen, Z. Wen, G. Zhu, J. Qiao, Y. Pan, H. Li, X. Gong, Z. Ye, W. Ao, L. Zhang, X. Yan, Y. Tong, X. Yang, X. Zheng, S. Fan, J.Cao, C. Yan, K. Xie, S. Zhang, Y. Wang, L. Zheng, Y. Wu, Z. Ge, X. Tian, X. Zhang, Y. Wang, R. Zhang, Y. Wei, W. Zhu, J. Zhang, H. Qiu, M. Su, L. Shi, Z. Xu, L. Zhang, Cheng, X. (2025). AI-based large-scale screening of gastric cancer from noncontrast CT imaging. Nature Medicine, 1-9. https://doi.org/10.1038/s41591-025-03785-6

[37] S. Siddiqui, J.A. Khan, T. Akram, M. Alharbi, J. Cha, D.A. AlHammadi, SNet: A novel convolutional neural network architecture for advanced endoscopic image classification of gastrointestinal disorders. SLAS technology, 33, (2025) 100304. https://doi.org/10.1016/j.slast.2025.100304

[38] P. Patil, V. Narawade, Radiology Image Data Augmentation and Image Enhancement in Respiratory Disease Infection Detection Using Machine Learning Approach. International Research Journal of Multidisciplinary Technovation, 6(2), (2024) 133-155. https://doi.org/10.54392/irjmt24211

[39] T.T. Habe, K. Haataja, P. Toivanen, Precision enhancement in wireless capsule endoscopy: a novel transformer-based approach for real-time video object detection. Frontiers in Artificial Intelligence, 8, (2025) 1529814. https://doi.org/10.3389/frai.2025.1529814

[40] W.A. Mustafa, H. Alquran, Editorial for the Special Issue Advances in Medical Image Processing, Segmentation, and Classification. Diagnostics, 15(9), (2025) 1114. https://doi.org/10.3390/diagnostics15091114

Authors Contribution Statement

T. Swetha Kumari: Conceptualization, Methodology, Experimental Work, Data Curation, Writing–Original Draft, Investigation, Validation, Visualization. R. Vasuki: Supervision, Writing–Review & Editing, Resources, Investigation, Project Administration, Writing–Review. Both the authors read and approved the final version of the manuscript.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Has this article screened for similarity?

Yes

About the License