Journal of Theoretical and Applied Information Technology
15st September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DYNAMICALLY STACKED ENDOSCOPY IMAGE QUALITY ENHANCED CAPSULE DEEP BELIEF NETWORK FOR GASTROINTESTINAL TRACT DETECTION

K. SHARIFA [1] , S. MALARVIZHI[2]

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

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

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

ABSTRACT

The growing pervasiveness of gastrointestinal (GI) tract disorders globally heightens the intense requirement for accurate diagnosis, as these diseases considerably influence human life and bring about high mortality rates. An effective treatment mechanism is indispensable for addressing this critical health issue. Early diagnosis with accurate treatment of gastrointestinal diseases is crucial for reducing mortality and boosting quality of life. In this study, a novel Dynamically Stacked Capsule Deep Belief Network (DS-CDBN) is proposed for comprehensive gastrointestinal disease detection. They are image quality enhancement and classification for gastrointestinal disease detection. First, an image quality enhancement model is applied to different categories of the Kvasir dataset by applying distance vector field and Dynamically Stacked Median Filter not only preserves the edge but also dynamically applies median filters to detect noise candidates and process accordingly. Next, rotation-based data augmentation is performed with varied magnitudes aids in detecting the disease for different categories. Finally, Capsule Proximal Greedy Layer-wise Deep Belief Network-based Gastrointestinal Tract Disease Detection is performed. Initially, features are learnt using a vector activation function called squashing through a capsule network. Then, to the learnt features proximal policy greedy layer is applied to generate diagnostic predictions for gastrointestinal tract diseases in an accurate and precise manner. Experimental evaluation on the publicly available Kvasir dataset achieved 27% accuracy and significantly lower training time by 37% compared to state-of-the-art methods. The results confirm the method's efficiency for accurate classification of eight gastrointestinal tract diseases, offering a scalable and interpretable solution for enhanced clinical decision-making in gastroenterology.

Keywords: *Gastrointestinal, Dynamically Stacked Median Filter, Capsule Proximal Greedy Layer-wise, Deep Belief Network.*

1. INTRODUCTION

GI diseases stand as a highly prevalent disease that impacts millions of people throughout the world every year. The undiagnosed and untreated conditions, which include ulcers and polyps, lead to severe health problems that decrease patients daily living conditions. Deep learning (DL) algorithms have made computer-aided diagnosis (CAD) and significant advancements in medical imaging, which now enables accurate disease diagnosis and classification. The system still faces difficulties because multiple hyperparameters create two problems, which include class imbalance and intra-class variability.

A Shallow Convolutional Neural Network (SNet) emerged as a solution to create a dependable system that could handle intricate classification tasks in [1]. The researchers first used data augmentation techniques on endoscopic images to create new images, which they used for subsequent feature extraction work. The research implemented minimum redundancy maximum relevance function to address the problem of the curse of dimensionality. It employed Convolutional Neural Network technology to enhance their classification accuracy results. The absence of class balancing in the system prevented accurate classification results despite the system achieving better classification accuracy. A novel DL framework integrating CNN-

GRU (Convolutional Neural Network and Gated Recurrent Units) was proposed in [2] to address the class imbalance issue and improve computational efficiency. In this approach, data augmentation was first performed to handle class imbalance. Subsequently, an improved version of sparse convolution layers with self-attention was designed to reduce computational complexity. Finally, the integration leveraged both sequential dependencies and spatial, ensuring disease classification.

The process of analyzing gastrointestinal endoscopic images encounters significant difficulties because the imaging conditions within the body create minor variations in image quality. The research study in [3] demonstrates the application of advanced CNNs and Transformer-based systems to detect neoplasia in Barrett's esophagus. The existing methods use traditional DL techniques, which provide effective results but continue to experience the problem of overfitting. The research in [4] introduced a new depthwise separable convolution layer, which enhances classification performance. The research study in [5] developed a strong method that uses EfficientNet to provide precise disease intervention at the correct time.

Gastrointestinal diseases are one of the major reasons for mortality globally and hence early with accurate diagnosis. Traditional methods like endoscopy consume time. A novel method in [6] employing concatenated recurrent vision transformer to boost automated diagnosis of disease at an early stage. Yet another hybrid DL technique combining an efficient vision transformer to achieve structural changes in gastrointestinal disease [7] to ensure accurate diagnosis. A method to focus on accuracy employing convolutional neural networks with Q-learning [8] to segment the lesion regions precisely.

Complicated artificial intelligence (AI) methods like deep neural networks have manifested notable potentialities to detect early-stage polyps and tumors. Three distinct explainable artificial intelligence methods in [9] were used through deep neural network to detect gastrointestinal abnormalities. Yet another inspired algorithm with DL using a nature-inspired algorithm [10] to demonstrate results. However, the computational cost was not focused. To concentrate on this aspect, statistical and structural features [11] utilizing a tuned hexa classification architecture.

A holistic review exploring the application of a CNN to focus on polyp and tumor detection [12], highlighting their advantages in accuracy and real-time decision making therefore reducing the diagnostic errors. Yet another method focusing on execution time employs a sparse transformer along with the principal gradient function [13] to not only ensure accuracy but also interpretability. A detailed review on the application of DL techniques to classify different categories of gastrointestinal tract detection in [14]. Though accuracy and computational costs were focused the class imbalance issue was not handled efficiently. To address this aspect, DL based segmentation was applied in [15] therefore ensuring accurate class-balanced results.

1.1 Contributing remarks

The major problems related to the above existing methods, such as high training time, ineffective performance, reduced recall rate, and minimized PSNR, to design a method for image quality enhancement and gastrointestinal tract detection. For this, the proposed work intends to develop a DL method called the DS-CDBN for image quality enhancement and gastrointestinal tract detection. The major contributions of the proposed work are as follows:

- First, based on the DL DS-CDBN method for ensuring image quality enhancement with improved gastrointestinal tract detection.
- Second, based on the Dynamically Stacked Median Filter model, the proposed DS-CDBN method not only preserves the edge and fine details but also detects noisy candidates by processing by dynamically applying median filters, therefore minimizing training time and improving PSNR accordingly.
- Third, Capsule Proximal Greedy Layer-wise Deep Belief Network employing first learnt the relevant features using vector activation function called squashing and generate diagnostic predictions for gastrointestinal tract diseases employing proximal policy greedy layer ensures precise and accurate results.
- Finally, the quantitative analysis results of performance evaluation represent that the training time along with PSNR and precision, recall, accuracy of the proposed DS-CDBN method are better than existing methods.

1.2 Roadmap

This paper is as follows. Section 2 describes the preliminaries employed in our work for ensuring image quality enhancement and detecting the gastrointestinal tract. Then, the proposed work methodology, along with pseudo code representation called DS-CDBN, is detailed in Section 3. A case analysis and inferences are presented in Section 4, with detailed experiment settings and discussion included in Section 5 along with table and graphical representation and concluding remarks of this work are given in Section 6.

2. RELATED WORKS

Early diagnosis and accurate classification of cancer are pivotal for better patient results. a DL system which uses convolutional neural networks to identify whether gastrointestinal endoscopic tissue images show malignant or benign. A novel gastrointestinal disease Detection and Classification method employing Hybrid Rice Optimization with Deep Learning (GDDC-HRODL) model [16] to accomplish enhanced gastrointestinal cancer classification. Nevertheless, the efficiency of endoscopy laboriously depended on the competence of the endoscopist. An automated method incorporating Squeeze and Excitation attention block in [17], clearing the way for early diagnosis and minimizing dependency on the endoscopist's competence.

As an influential analytic tool as far as medical image analysis was concerned, specifically, the deep convolutional neural network (DCNN) received extraordinary awareness owing to its potential in providing results comparable to those of medical experts. Identification of GI tract abnormalities through automated processing of endoscopic images remained an extremely difficult challenge for advanced gastroenterologists which could lower their time requirements and their costs of investigation. A specialized DCNN [18] was facilitating and marshalling the human GI abnormalities, along with minimizing cost and time. An autonomous method employing DL based polyps segmentation was presented in [19] to achieve improved classification accuracy. Though the efficiency of diagnosis was ensured, the reliability of model prediction was not ensured. A vision transformer-based explainable AI was in [20].

A computer-aided diagnosis method integrating a capsule network along with the DL technique was [21] to ensure improved precision and accuracy. Current research works strove to ascertain the feasibility and efficiency. Deep transfer learning was used in [22] to accomplish the image quality control task. A holistic study of the application of multimodal DL for gastrointestinal tract detection was conducted in [23]. A plethora of DL techniques were presented in [24] to classify the input images into five different categories, ensuring early detection.

The limited access to medical imaging data which was available to created an essential problem that caused their classifiers to perform poorly. The study [25] developed two solutions to solve their problem through the combination of geometric color augmentation with synthetic data generation techniques. Therefore, demonstrating an improvement in image classification. However the method did not use the full potentiality of data augmentation and hence a robust method was in [26] employing EfficientNet resulting in more accurate and timely disease interventions.

A bidirectional fusion model integrating recurrent neural network and CNN was presented in [27] with superior detection and lesion segmentation capabilities. Nevertheless the computational cost was not reduced. To minimize the computational cost along with the accurate detection of gastrointestinal tract disease transformer based classification model was designed in [28]. Yet another heterogeneous CNN and vision transformer technique was presented in [29] for rapid disease identification.

Based on the aforementioned mechanisms and algorithms discussed and drawbacks identified in this work, a DL based method called DS-CDBN for image quality enhancement and gastrointestinal tract detection is proposed. The elaborate description of the DS-CDBN PRH-LVH method is provided in the following sub-sections.

3. MATERIALS AND METHODOLOGY

3.1 Dataset Collection and Description

The Kvasir dataset [30] extracted from https://www.kaggle.com/datasets/meetnagadia/kvasir-dataset contains images from inside the gastrointestinal tract. The images show three main anatomical landmarks together with three important medical findings. The collection contains two types of images, which show the process of endoscopic polyp removal. The data collection used endoscopic equipment at Vestre Viken Health Trust (VV) in Norway, which serves 470.000 residents, to create the disease detection dataset. The anatomical landmarks consist of the Z-line, the pylorus, cecum. The pathological finding includes three conditions, which are esophagitis, polyps, and ulcerative colitis.

Journal of Theoretical and Applied Information Technology
15st September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The dataset includes images that have different resolutions that range from '720*576' to '1920*1072' pixels and they are organized into different folders that match their specific content.

3.2 Dynamically Stacked Capsule Deep Belief Network based gastrointestinal tract disease detection

The evolution in application of deep learning algorithms and computer aided diagnosis (CAD), helps radiologists to diagnose diseases more accurately even in early stages. The process of identifying GI abnormalities through endoscopic imaging presents significant challenges which even experienced gastroenterologists struggle to overcome. These challenges have the potential to revolutionize medical diagnostics while they decrease the duration and expenses of subsequent investigative procedures. Once the acquisition of samples is accomplished, image quality enhancement is an indispensable stage that minimizes training time as well as error rate. Nevertheless, as far as medical diagnosis is concerned, GI tract detections in humans require manual detection methods which depend on the skills of gastrointestinal endoscopists. The proposed methodology for the classification of gastrointestinal tract diseases is composed of two major phases. They are image quality enhancement and classification for gastrointestinal tract disease detection. Image quality enhancement is performed to improve the gastrointestinal tract image quality. Figure 1 shows the architecture of the Dynamically Stacked Capsule Deep Belief Network (DS-CDBN) method.

As illustrated in Figure 1 above, the DS-CDBN method consists of two main steps. Firstly, the raw multi-class gastrointestinal image quality is enhanced using the dynamically stacked median filter where the noisy candidates are detected and processed to generate robust output image free of noise. Then the quality enhanced image is used as a guide image to perform rotation-based augmentation. Finally, features relevant to gastrointestinal tract disease detection is learnt using Capsule Proximal Greedy Layer-wise model. Deep Belief Network based classifier for gastrointestinal Tract Disease Detection. Details of each step are elaborated below.

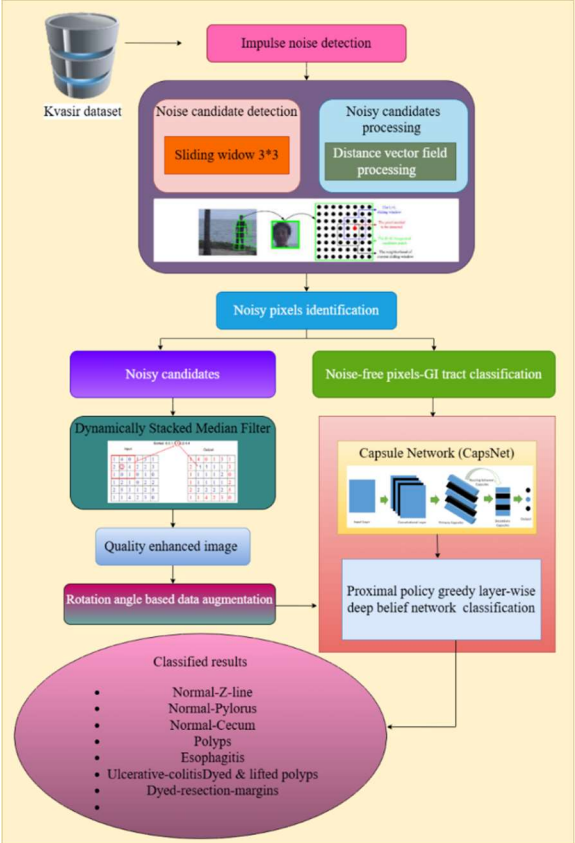


Figure 1 Architecture of DS-CDBN

3.2.1 Dynamically Stacked Median Filter Image Quality Enhancement model

Dynamically Stacked Median Filter a nonlinear image filtering technique has been extensively utilized in several image processing applications. Dynamically Stacked Median Filter has received popularity owing to its potentiality in efficiently removing impulse noise. Dynamically Stacked Median Filter is its potentiality to adjust to distinct magnitude of noise present in the image. The Dynamically Stacked Median Filter uses adaptive window size adjustments through its system because it needs to handle different noise levels for effective image processing. The system uses larger windows in areas with dense noise to achieve precise median values, while smaller windows operate in low-noise areas to maintain delicate details. The system provides better edge detection results than standard linear filters because it protects edges while maintaining fine details, which makes it an ideal solution for diagnosing gastrointestinal tract diseases. Figure 2 shows the structure of the Dynamically Stacked Median Filter Image Quality Enhancement model.

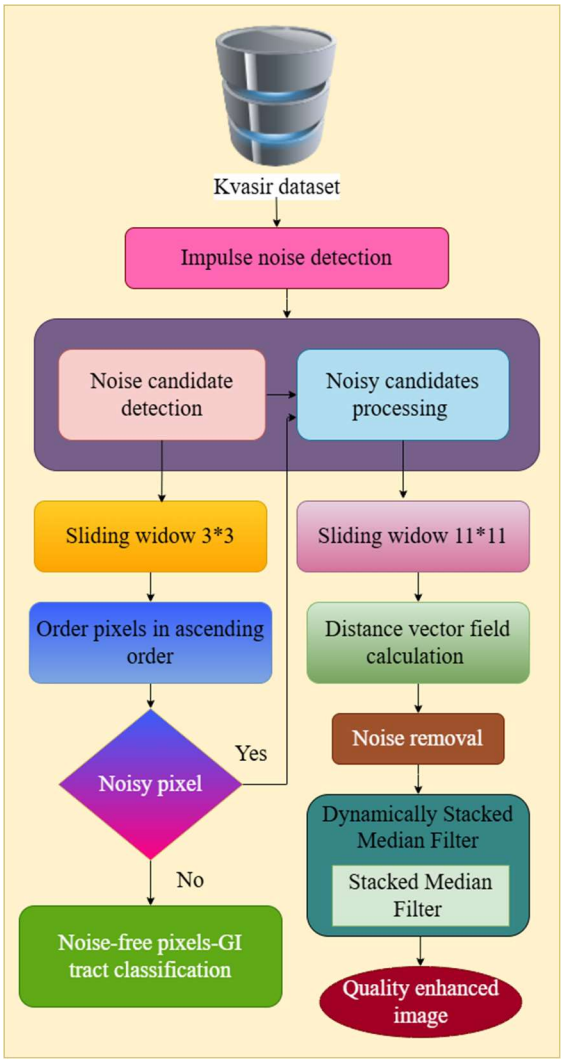


Figure 2 Structure of Dynamically Stacked Median Filter Image Quality Enhancement model

As shown in the above figure, the proposed image quality enhancement model for gastrointestinal tract disease detection. The method includes two stages which consists of impulse noise detection and noise removal. The process starts with noise detection through a sliding window method which leads to adaptive filtering that uses window sizes of 3×3 and 5×5 and 7×7 according to the detected noise levels. The Dynamically Stacked Median Filter improves image quality by processing all identified noisy pixels. The noise detection process begins with the sliding window method to locate noise candidates, which leads to subsequent image quality improvement work. Following which the noise candidates are processed by means of Distance Vector Field.

3.2.2 Noise Candidate Detection

To start with, across every pixel '$p(i,j)$', a '$3*3$' window is moved, positioned at '(i,j)'. Followed by which for each sliding window '$Win^{3}_{(i,j)}$' the pixels are ordered in ascending order as given below.

$$R(i,j) = [r_1(i,j), r_2(i,j), \dots . r_9(i,j)] \quad (1)$$

Next, based on the above ordering of pixel as given in equation (1) noisy pixel or noise-free pixel results are arrived at as given below in equation (2),

$$\varepsilon = \{p(i,j) = r_1(i,j)\ or\ p(i,j) = r_9(i,j), then\ p(i,j)\ is\ a\ noisy\ candidates\ p(i,j) \neq r_1(i,j)\ or\ p(i,j) \neq r_9(i,j), then\ p(i,j)\ is\ a\ noise\ free\ candidates \quad (2)$$

Based on the above results the noisy candidates are used for further processing whereas the noise free pixel is subjected to classification for gastrointestinal tract detection.

3.2.3 Distance Vector Field based Noisy Candidates Processing

The second step is based on the final noisy pixels selection by processing noisy candidates using the Distance Vector Field. First the sliding window '$11 * 11$' is moved '$W^{11}_{(i_1,j_1)}$' across all noisy candidates at position '(i_1, j_1)' and is represented as given below equation (3).

$$R(i_1,j_1) = [r_1(i_1,j_1), r_2(i_1,j_1), \dots, r_{121}(i_1,j_1)] \quad (3)$$

Based on the above noisy candidates position, the Distance Vector Field is arrived at as given below equations (4 and 5).

$$Dis(i_1,j_1) = [dis_1(i_1,j_1), dis_2(i_1,j_1), \dots, dis_{120}(i_1,j_1)] \quad (4)$$

$$dis_i(i_1,j_1) = r_{(n+1)}(i_1,j_1) - r_{(n)}(i_1,j_1) \quad (5)$$

According to the above position value, the four largest distances are obtained with the intent of preserving the edge, and it is mathematically formulated as given below equation (6).

$$dis_{max1} = dis_a(m_1,n_1); dis_{max2} = dis_b(m_1,n_1); dis_{max3} = dis_c(m_1,n_1); dis_{max4} = dis_c(m_1,n_1) \quad (6)$$

With the above obtained four largest distances noisy pixels finally to generate image quality enhanced results are defined as given below equation (7).

Journal of Theoretical and Applied Information Technology
15st September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$Res = \{if\ p(m_1, n_1) < w_{min}\ or\ p(m_1, n_1) > w_{max}, then\ noisy\ pixel\ if\ p(m_1, n_1) > w_{min}\ or\ p(m_1, n_1) < w_{max}, then\ noise\ free\ pixel \quad (7)$$

3.2.4 Dynamically Stacked Median Filter

For improved image quality, The filter procedure is consequently predicated on the number of noisy pixels. Formally, given a pixel 'm' in a sample image 'S', then for each pixel '$n \in R(p)$' the Stacked Median Filter is defined as given below.

$$w_{mn} = G[f(m)f(n)] \quad (8)$$

From the above equation (8) 'w_{mn}' denotes the weight based on similarity of pixel 'm' and 'n' in the subsequent feature map 'f'. Moreover, '$f(m)$' and '$f(n)$', represent the feature at pixels 'm' and 'n' in 'f' respectively with Gaussian function 'G' influencing between adjacent pixels. Finally by sorting all pixels values Dynamically Stacked Median Filter is obtained to remove noisy pixels as given below.

$$S(m') = k\ s.t\ \sum_{n=1}^{k} w_{mn} \geq \frac{1}{2}\sum_{n=1}^{N} w_{mn} \quad (9)$$

From the above equation (9) Dynamically Stacked Median Filtered quality enhanced image '$S(m')$' is obtained with reference to minimum value of 'k' in the sampled images (according to the pixel positions)

3.2.5. Rotation Angle-based Data Augmentation

DL has been widely used for medical image classification, delivering improved performance. The study used rotational transformations to increase the dataset size because the study needed to reach 47,398 images. The rotation function creates different image variations by selecting random angles from its specified angle range.

$$DA = Res[\alpha\ \beta\ (1-\alpha).centre.x - \beta.centre.y\ -\beta\ \alpha\ \beta.centre.x + (1-\alpha).centre.y\ ] \quad (10)$$

From the above equation (10) the rotation-angle based data augmentation result 'DA' is arrived at based on the scaling with respect to '$cos\ cos\ \theta$' (i.e. '$\alpha = scale * cos\ cos\ \theta$') and scaling with respect to '$sin\ sin\ \theta$' (i.e. '$\alpha = scale * sin\ sin\ \theta$') with '$\theta$' denoting the rotation angle. The pseudo code representation of Dynamically Stacked Median Filter Image Quality Enhancement is given below.

1: **Initialize** '$m = 4000$'
2: **Begin**
3: **For** each Dataset 'DS' with Sample Images 'S'
4: Order pixels in ascending order for sliding window according to (1)
5: Generate noisy pixel or noise free pixel results according to (2)
6: Move sliding window across all noisy candidates according to (3)
7: Generate Distance Vector Field results according to (4) and (5)
8: Obtain four largest distance results according to (6)
9: Generate image quality enhanced results according to (7)
10: **Return image quality enhanced result** 'Res'
11: Fine-tune the weight according to (8)
12: Generate Dynamically Stacked Median Filtered according to (9)
13: Generate data augmented result according to (10)
14: **Return data augmented result** 'DA'
15: **End for**
16: **End**

Algorithm 1 Dynamically Stacked Median Filter Image Quality Enhancement

The entire process has been split into two distinct segments so as to ensure the enhancement of image quality. They are detecting noise candidates and processing of noisy candidate pixels. First noise candidates present in the sample images are detected based on the sliding window concept. Followed by which noise free pixels are discarded from further processing whereas the noisy pixels are processed using Distance Vector Field. Adoption of the concept of distance vector field not only reduces the training times but also retains the sharp edge. Finally, Dynamically Stacked Median Filter is applied to fine-tune the weight while processing the noisy candidates that by applying dynamic results for varied classes (like dyed-lifted-polyps, normal-cecum, normal-pylorus, etc.) in turn improves the overall peak signal to noise ratio. The Dynamically Stacked Median Filter being highly efficient in discarding impulse noise that appear as random black and white spots in endoscopic images. By eliminating this noise the proposed method reduces the mean squared error between the original and pre-processed image, therefore increasing the overall peak signal to noise ratio. Finally by applying the Rotation Angle-based Data augmentation efficiently increase the diversity and size of the training dataset. This in turn helps prevent the model from overfitting to limited original data and enhances its performance on unseen pixel data while performing gastrointestinal tract.

3.3. Capsule Proximal Greedy Layer-wise Deep Belief Network based Gastrointestinal Tract Disease Detection

Input: Dataset 'DS', Sample Images '$S = \{S_1, S_2, \ldots, S_M\}$'
Output: Robust image quality enhancement

In this section with the image quality enhanced and data augmented results as input are

subjected to gastrointestinal tract disease detection. To achieve better classification accuracy and lesion detection in to respective categories, Capsule Proximal Greedy Layer-wise Deep Belief Network based Gastrointestinal Tract Disease Detection on endoscopy images is proposed (figure 3).

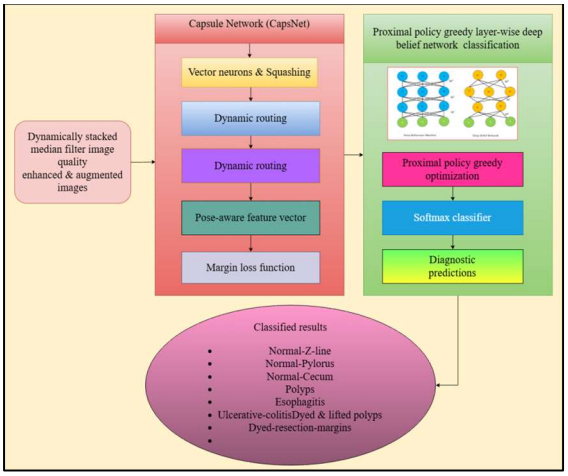


Figure 3 Structure of Capsule Proximal Greedy Layer-wise Deep Belief Network based Gastrointestinal Tract Disease Detection

As illustrated in the above figure, the Capsule Deep Belief Network based Gastrointestinal Tract Disease Detection on endoscopy images are split into two parts. In the first part, better features are learnt with the augmented images as input using Capsule Network (CapsNet). The Proximal Policy Greedy Layer-wise Deep Belief Network classifier is applied to the extracted features to improve classification accuracy and precision. The Capsule Network (CapsNet) uses dynamic routing to produce vector outputs while it maintains spatial details and hierarchical relationships from augmented images, which enables the creation of better feature representations (i.e. to preserve critical pose information) owing to the reason that the capsule consists of numerous neurons. Moreover using vector activation function called squashing with intent of preserving critical pose vector representation for full set of instantiation parameters for specific class is obtained as given below.

$$CO_j = \frac{||DA_j||^2}{1+||DA_j||^2} \frac{DA_j}{||DA_j||} \quad (11)$$

From the above equation (11) 'CO_j' denotes the capsule output whereas the 'DA_j' represent the total capsule input or the data augmented results to which simulation is to be performed. The calculation of the 'DA_j' capsules total input values requires the weighted sum of predictive vector '$V_{j|i}$' which the lower layer capsule presents as its established value. The evaluation of '$V_{j|i}$' requires the multiplication of the lower layer capsule with the weight matrix 'WM_{ij}' and the output 'Out_i'. This is mathematically defined as given below.

$$DA_j = \sum_i Coeff_{ij} v_{j|i} \quad (12)$$

$$v_{j|i} = WM_{ij} Out_i \quad (13)$$

From the above equations (12) and (13) '$Coeff_{ij}$' represents the coefficient matrix defined by dynamic routing to model spatial correlations between learned features in medical images. This dynamic routing assists in accurately differentiating fine-drawn structural dissimilarities connected with various types of gastrointestinal tract diseases. This dynamic routing function is defined as given below.

$$Coeff_{ij} = \frac{expexp\,(\tau_{ij})}{\sum_k (\tau_{ik})} \quad (14)$$

From the above equation (14) 'τ_{ij}' denotes the log likelihood function for the corresponding coefficient results '$Coeff_{ij}$' of the subsequent learned features. Next to generate optimal results margin loss function is used to make certain that the magnitude of the distance vector for the correct class is close to one and the magnitudes of the distance vector for incorrect classes are close to zero with a significant margin in between.

$$L_k = \alpha_k(Out, \beta^+ - ||CO_k||)^2 + \lambda\big(1 - \alpha_k\big(Out, ||CO_k|| - \beta^-\big)\big)^2 \quad (15)$$

From the above equation (15) the margin loss function result 'L_k' is arrived at by substituting the values of 'α_k' to '1'. Moreover the value of 'β^+' and 'β^-' is assigned to '0.9' and '0.1' with which ensures optimal results with respect to the capsule output 'CO_k' respectively.

Finally using the learnt feature results generated with optimal results in the proposed method, and Greedy Layer-wise Deep Belief Network based Classification is used. The Proximal Policy Greedy Optimization (PPGO) for classification makes locally optimal choice at each stage instead of optimizing entire deep network all at once. The PPGO dynamically fine-tunes the prioritization of features during classification. This in turn aids in generating the results of classification accurate and precisely for different types of categories in gastrointestinal tract. Let us assume that the Deep Belief Network (DBN) encompasses of '$m-th$' hidden layer, followed by which layer is initialized by greedy strategy and weight from '$m-l$' to hidden layers '$m-th$' its general structure is formulated as given below.

$$\sum_{i=1}^{m} \blacksquare \gamma_i g(W_i Out_{m-1} + b_i) = Pre_j, j = 1,2,..,l \quad (16)$$

$$Out_m \gamma = H \quad (17)$$

From the above equations (16) and (17) 'Out_n', denote the output from '$m - lth$' to '$m - th$' layer that is generated as given below.

$$Out_m = \left[g\left(W_1 . Out_{n-1,l} + b_1\right), g\left(W_2 . Out_{n-1,l} + b_2\right), \dots g\left(W_m . Out_{n-1,l} + b_m\right) \right] \quad (18)$$

From the above equation (18) once the 'W_i' and 'b_i' values are selected arbitrarily the classification output 'H' is generated as given below.

$$\gamma' = Out_m^T H \quad (19)$$

The PPGO-based DBN classifier uses an adaptive decision-making strategy which changes its operational procedures to match different classification requirements and various noise conditions. The system achieves stable convergence with reduced misclassification errors by treating classification as a sequential process that optimizes a sparse kernel function. The dynamic feature reweighting process proves essential for medical diagnostics because medical data frequently exhibits contradictory patterns. The refined visual feature vector is fed into softmax classifier to generate diagnostic predictions for gastrointestinal tract diseases '$\{Out_1, Out_2, \dots, Out_C\}$' as given below.

$$Prob(Out_c|FV) = \frac{exp\ exp\ (W_c^T FV + b_c)}{\sum_{c=1}^{C} \boxdot\ exp\ exp\ (W_c^T FV + b_c)} \quad (20)$$

From the above equation (20) 'C' represents the total number of diseased categories (i.e. died-lifted-polyps, dyed-resection-margin, normal-cecum and so on), 'W_c' and 'b_c'are learnable parameters for category 'C' respectively. The pseudo code representation of Capsule Proximal Greedy Layer-wise Deep Belief Network based Gastrointestinal Tract Disease Detection is given below.

Input: Dataset 'DS', Sample Images '$S = \{S_1, S_2, \dots, S_M\}$'
Output: Precise and accurate gastrointestinal tract detection

1: **Initialize** quality enhanced result 'Res', data augmented result 'DA'
2: **Begin**
//Feature learning
3: **For** each Dataset 'DS' with quality 'DA'
4: Generate squashing based vector activation function according to (11)
5: Generate total input values of the 'DA_j' according to (12) and (13)
6: Generate dynamic routing function according to (14)
7: Evaluate marginal loss function results according to (15)
8: **Return feature learnt results** 'L_k'
9: **End for**
//Gastrointestinal tract detection
10: **For** each Dataset 'DS' with quality data augmented 'DA' and feature learnt results'L_k'
11: Generate proximal policy greedy strategy to (16) and (17)
12: Generate actual output result according to (18)
13: Generate arbitrarily the classification output according to (19)
14: Generate diagnostic predictions for gastrointestinal tract diseases
16: **End for**
17: **End**

Algorithm 2 Capsule Proximal Greedy Layer-wise Deep Belief Network based Gastrointestinal Tract Disease Detection

As given in the above algorithm the entire process of gastrointestinal tract detection is split into two processes. First, the quality enhanced data augmented results for the corresponding categories are obtained as input. Second squashing based vector activation function via capsule network is generated with the intent of with intent of preserving critical pose vector for specific class or category. Followed by which dynamic routing is applied to the critical pose preserved vector results for a specific category that in turn differentiates between fine-drawn structural dissimilarities associated with different types of gastrointestinal tract diseases, therefore improving the overall precision and recall. Third, with the learnt feature results generated optimally where using proximal policy greedy strategy is applied make certain that each layer learns meaningful features that construct upon the previous layer, imparting a strong foundation for the final gastrointestinal tract detection. The system achieves successful diagnosis of eight gastrointestinal disease classes because it maintains dynamic modality

balance while its components remain easy to understand.

4. CASE STUDY AND INFERENCES

In this section case analysis of image quality enhancement and gastrointestinal tract detection using the 90Kvasir dataset dataset are simulated by applying deep learning method called DS-CDBN method. Figure 4 given below shows 3 sample images for eight different classes.

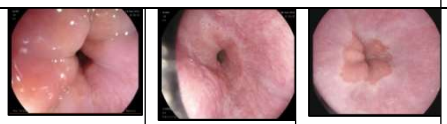
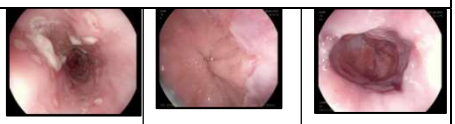
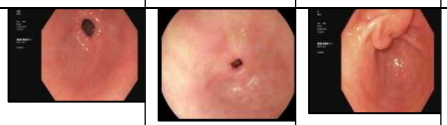
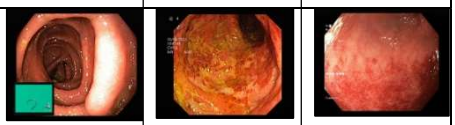
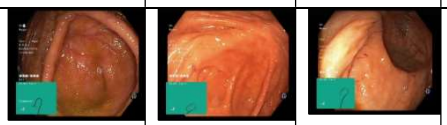
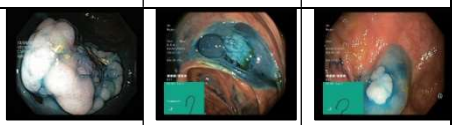
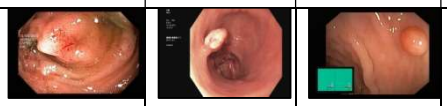
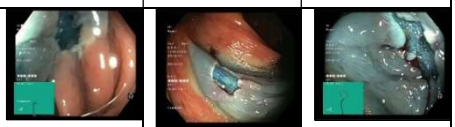
Class name	Images		Images
Normal-Z-line		Esophagitis	
Normal-Pylorus		Ulcerative-colitis	
Normal-Cecum		Dyed & lifted polyps	
Polyps		Dyed-resection-margins	

Figure 4 Sample images

To the above sample images of eight different classes Dynamically Stacked Median Filter was applied as image quality enhancement. Here initially noisy candidates were detected and followed by which the noisy candidate pixels were processed via Dynamically Stacked Median Filter with the intent of preserving the edge and fine details for further processing. Figure 5 shows the image quality enhanced results.

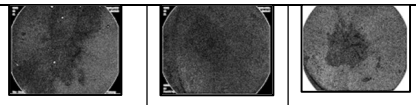
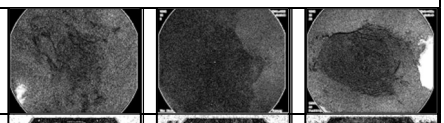
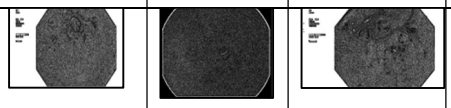
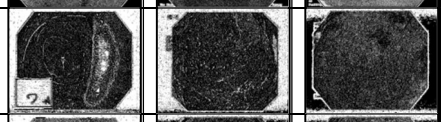
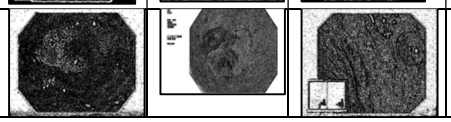
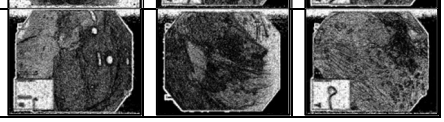
Classes Name	**Images**	**Classes Name**	**Images**
Normal-Z-line		Esophagitis	
Normal-Pylorus		Ulcerative-colitis	
Normal-Cecum		Dyed & lifted polyps	
Polyps		Dyed-resection-margins	

Figure 5 Image quality enhanced results

As illustrated in the above figure by applying the image quality enhancement algorithm employing distance vector field and stacked median filter in an adaptive fashion aids in reducing the training time

along with improved PSNR. Also to increase the diversity of images, rotation angle based data augmentation was applied and figure 6 shows its results.

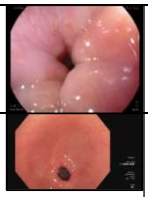
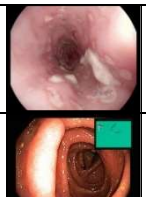
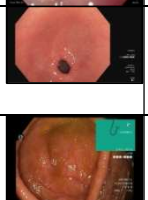
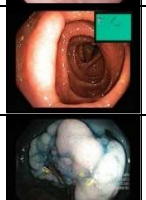
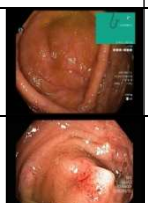
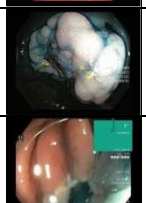
Classes Name	Images	Classes Name	Images
Normal-Z-line		Esophagitis	
Normal-Pylorus		Ulcerative-colitis	
Normal-Cecum		Dyed & lifted polyps	
Polyps		Dyed-resection-margins	

Figure 6 Rotation-angle based data augmentation results

As shown in the above figure after applying the rotation-angle based data augmentation each category frequencies were increased to 1000 number of images and as a result an overall of 8000 data augmented samples were generated for further processing. With the data augmented sample results feature learning was performed by applying vector activation function called squashing and to preserve critical pose vector. Moreover dynamic routing was performed with the intent of differentiating fine-drawn structural dissimilarities associated with different types of gastrointestinal tract diseases. Figure 7 shows the results of two different classes polyps and ulcerative-colitis.

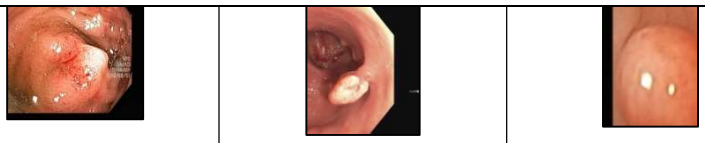
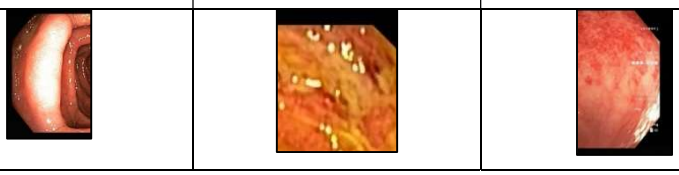
Class name	Sample images
Polyps	
Ulcerative-colitis	

Figure 7 Feature learnt results

With the above feature learnt results finally the gastrointestinal tract detection using Greedy Layer-wise Deep Belief Network is finally performed. The Proximal Policy Greedy Optimization (PPGO) for classification dynamically fine-tunes prioritization of features during classification ensuring accurate and precise results. Figure 8 shows the classified results.

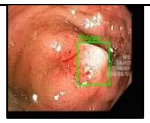
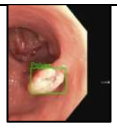
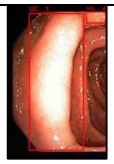
Class name	Sample images		
Polyps			
Ulcerative-colitis			

Figure 8 Classified Results For Gastrointestinal Tract Detection

As shown in the above figure with green bar represents the detection results for polyps class and red bar denoting the detection results for ulcerative-colitis ensure stable convergence. The quantitative results are provided in the following sub-sections.

5. EXPERIMENTAL STUDY AND DISCUSSION

DS-CDBN is proposed for comprehensive gastrointestinal disease detection is explored and tested with other significant methods, namely Shallow Convolutional Neural Network (SNet) [1] and CNN-GRU [2]. The evaluation metric shows its ability to distinguish between results from various DL methods which Python developers created using the Kvasir dataset that they retrieved from https://www.kaggle.com/datasets/meetnagadia/kvasir-dataset [30]. The DL method shows its efficiency through two assessment methods which include execution measures and performance evaluations that use multiple assessment benchmarks together with training duration and peak signal to noise ratio precision, recall, and accuracy. The Kvasir dataset evaluation method assesses medical image classification performance through multiple metrics which evaluate different aspects of the model's predictive performance. The method uses different classification metrics to show how well it classifies gastrointestinal diseases which helps researchers ensure system reliability during clinical testing.

5.1 Performance analysis of training time

The training time involved in the process of image quality enhancement and gastrointestinal tract detection is discussed. A significant portion of energy needs to be utilized during the detection process and quality enhancement process which researchers refer to as training time. The training time is mathematically defined as given below.

$$TT = \sum_{i=1}^{m} DA_i * Time\left(Prob(Out_c|FV)\right) \quad (21)$$

From equation (21) the 'TT' training time, is measured based on the total augmented samples 'DA_i' and the consumed time in obtaining the probabilistic output '$Time\left(Prob(Out_c|FV)\right)$'The measurement system uses seconds as its base. The table below presents the training duration results which were evaluated through the DS-CDBN an SNet [1], CNN-GRU [2].

Table 1 Training time results

Samples	Training time (sec)		
	DS-CDBN	SNet [1]	CNN-GRU [2]
700	17.5	20.3	22.15
1400	20.35	26.25	30.35
2100	25.15	31.55	35.15
2800	45.35	51.25	55.25
3500	60.15	67.35	71.35
4200	69.15	76.55	80.55
4900	75.35	82.85	86.35
5600	82.55	90.55	94.55
6300	90.15	97.35	101.55

7000	95.35	102.55	108.35

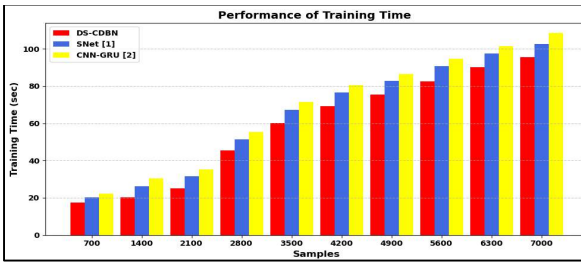


Figure 9 Training time versus samples

Figure 9 illustrates the graphical representation of training time using the proposed DS-CDBN and two existing methods SNet [1] and CNN-GRU [2]. The horizontal axis represents the augmented samples involving different categories ranging between 700 and 7000 whereas the vertical axis represents the training time results by substituting the values in equation 21. An increase in training time was found using all the three methods. This in turn causes an increase in the training time. Bust comparative analysis show minimal training time using DS-CDBN method upon comparison to [1] and [2]. This was because by applying the Dynamically Stacked Median Filter Image Quality Enhancement model the edge details and fine details of the images were retained. Moreover fine-tuning was even made for different magnitude of noise present in different categories. With this the overall training time of DS-CDBN method was reduced by 14% compared to [1] and 23% compared to [2].

5.2 Performance analysis of peak signal to noise ratio

The term peak signal-to-noise ratio (PSNR) is a performance metric evaluating the ratio between the maximum possible value and the power of distorting noise that influence the quality of image enhancement representation. PSNR is mathematically denoted as given below.

$$PSNR = 20log_{10}\left(\frac{Max_f}{\sqrt{MSE}}\right) \quad (22)$$

$$MSE = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(DA(i,j) - Out\ (i,j)\right)^2 \quad (23)$$

From the above equations (22) and (23) the mean square error 'MSE' and peak signal to noise ratio '$PSNR$' results are arrived at based on the matrix data of original data augmented result 'DA', predicted result 'Out', 'm' number of rows, 'n' number of columns and '$Max_f = 255$' respectively. Table 2 given below lists the PSNR analyzed using the DS-CDBN and SNet [1], CNN-GRU [2].

Table 2 PSNR results

Samples	PSNR (dB)		
	DS-CDBN	SNet [1]	CNN-GRU [2]
700	11.35	10.25	9.15
1400	13.15	11	9.85
2100	15.35	13.25	11.15
2800	16.85	14.55	12.25
3500	18.25	16.15	14.35
4200	20	18.35	16.25
4900	22.35	20.15	18.55
5600	24.15	22.55	20.15
6300	26	24.35	22.25
7000	28.35	26.45	24.35

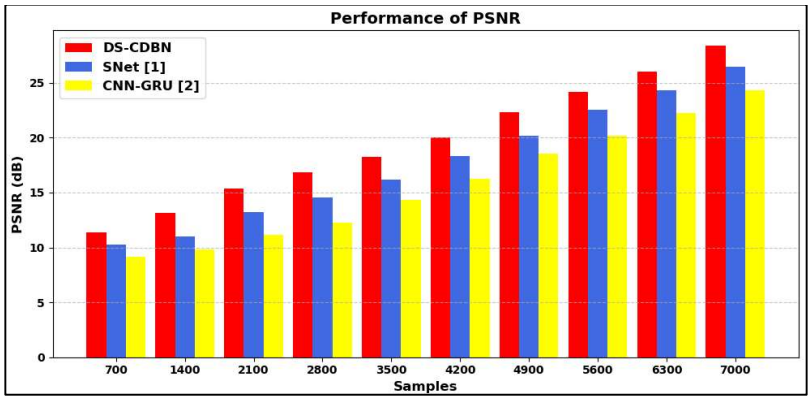


Figure 10 PSNR versus samples

Figure 10 illustrates the graphical representation of PSNR. The results of PSNR using the three methods were arrived at by substituting the values in equations (22) and (23) and simulations performed over ten runs. To ensure fair comparison similar set of data augmented samples were applied for all the three methods and accordingly PSNR was measured and plotted in the above graph. The PSNR results of the proposed DS-CDBN method showed better performance than [1] and [2] according to the results presented in the above figure. The proposed DS-CDBN method showed improved PSNR results because it used Dynamically Stacked Median Filter Image Quality Enhancement algorithm. By applying this algorithm first noisy candidates were identified and followed by which noisy candidates were processed by applying the Distance Vector Field. Finally based on the different categories of data augmented samples, Dynamically Stacked Median Filter was applied according to pixel positions. This in turn improved the overall PSNR of proposed DS-CDBN method by 10% compared to [1] and 20% compared to [2].

5.3 Performance analysis of precision, recall and accuracy

The positive predictive value which measures precision assesses how correctly positive results are predicted. For example, high precision in detecting polyps is indispensable as false positives could result in lead to irrelevant invasive process like biopsies. The precision rate is mathematically defined as given below.

$$Pre = \frac{TP}{TP+FP} \quad (24)$$

Recall establishes its capacity to determine every significant sample present in each category. The conditions of ulcerative colitis and esophagitis require high recall because any missed detection would result in essential treatment delays. The recall rate is defined as follows.

$$Rec = \frac{TP}{TP+FN} \quad (25)$$

Finally accuracy is the most commonly used performance metric. High accuracy is representative of a method's overall efficiency across all classes. Accuracy is mathematically defined as given below.

$$Acc = \frac{TP+TN}{TP+FP+TN+F} \quad (26)$$

The equations between Eqs. (24) and Eqs. (26) show that true positive results show esophagitis samples which were identified correctly as esophagitis. True negative results show that polyps were successfully identified as polyps. False positive results show that esophagitis samples were wrongly identified as polyps. False negative results show that polyps were wrongly identified as esophagitis. The entire sample set includes all samples used for classification. Table 3 shows the results which compare the precision and recall and accuracy of DS-CDBN against SNet [1] and CNN-GRU [2].

Table 3 Precision, Recall and Accuracy results

Samples	Precision			Recall			Accuracy		
	DS-CDBN	SNet [1]	CNN-GRU [2]	DS-CDBN	SNet [1]	CNN-GRU [2]	DS-CDBN	SNet [1]	CNN-GRU [2]
700	0.95	0.92	0.89	0.99	0.98	0.98	0.95	0.91	0.88
1400	0.93	0.83	0.75	0.96	0.9	0.8	0.94	0.84	0.79
2100	0.9	0.8	0.72	0.95	0.89	0.79	0.92	0.82	0.77
2800	0.88	0.78	0.7	0.93	0.87	0.77	0.9	0.8	0.75
3500	0.85	0.75	0.67	0.91	0.85	0.75	0.88	0.78	0.73
4200	0.82	0.72	0.64	0.89	0.83	0.73	0.85	0.75	0.7
4900	0.8	0.7	0.62	0.85	0.79	0.69	0.82	0.72	0.67
5600	0.84	0.74	0.66	0.88	0.82	0.72	0.86	0.76	0.71
6300	0.86	0.76	0.68	0.92	0.86	0.76	0.89	0.79	0.74
7000	0.88	0.78	0.7	0.95	0.89	0.79	0.93	0.83	0.78

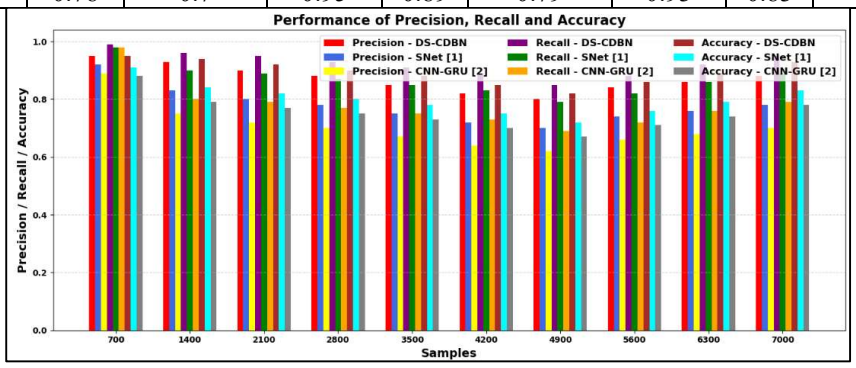


Figure 11 Precision, Recall and Accuracy results

Figure 11 illustrates the graphical representation of precision, recall and accuracy. From the above figure with x axis representing the data augmented samples the corresponding precision, recall and accuracy analysis were made in the y axis for the three methods, DS-CDBN and SNet [1], CNN-GRU [2]. Though a decreasing results were observed for all the three methods with respect to first seven iterations whereas the rest of three iterations were found to be high. This corroborates the objective that increasing the sample size does not have influence on the results. Also the precision, recall and accuracy of DS-CDBN method were found to be comparatively better than [1] and [2]. The reason was first by applying vector activation function called squashing via Capsule Deep Belief Network based Gastrointestinal Tract Disease Detection on endoscopy images critical pose vector were preserved. Moreover better features relevant for gastrointestinal tract detection were learnt, therefore improving the precision of DS-CDBN method by 11% and 19% compared to [1] and [2]. In addition by applying dynamic routing aided in differentiating fine-drawn structural dissimilarities connected with different types of gastrointestinal tract diseases therefore improving the recall of DS-CDBN method by 6% compared to [1] and 16% compared to [2]. Finally by using Greedy Layer-wise Deep Belief Network based Classification Proximal Policy Greedy Optimization was applied that dynamically fine tuned the feature prioritization during classification, making certain that decisions demonstrate the most current and relevant diagnostic result. This in turn improved the overall accuracy rate of DS-CDBN method by 11% and 16%.

Table 4 `Ablation Study of DS-CDBN Components

	Image Quality Enhancement	Capsule Feature Learning	Greedy DBN Layer	Accuracy (%)	Training Time (min)
Baseline CNN	✗	✗	✗	89.4	98
+ Dynamic Median Filtering	✓	✗	✗	91.6	95
+ Capsule Feature Extraction	✓	✓	✗	94.2	88
Full DS-CDBN	✓	✓	✓	96.5	74

The ablation study results presented in Table 4. The baseline CNN model achieved an accuracy of 89.4% with a training time of 98 minutes, which shows that conventional architectures are able to extract general visual features, but they have difficulty capturing the intricate spatial patterns that exist in endoscopic images. The accuracy improves to 91.6% when the dynamic median filtering-based image quality enhancement module is added to the system, which also results in a minor reduction of training time to 95 minutes. The model has improved because the filtering system successfully eliminates unwanted elements while maintaining the essential boundary details in gastrointestinal images, which helps the model learn more distinctive attributes.

The accuracy of the system increases to 94.2% through an additional capsule feature extraction layer. The system achieves optimal performance after 88 minutes of training time. The spatial and hierarchical feature relationships in capsule networks maintain their structure through vector activations and routing functions, which enable better identification of complex anatomical patterns in endoscopic images. The traditional convolutional filters fail to match the ability of capsules to detect orientation and positional changes in lesions. The complete DS-CDBN framework, which uses the proximal greedy layer-wise deep belief network, achieved its best performance with a 96.5% accuracy rating and required 74 minutes for training. The greedy DBN learning mechanism enables hierarchical feature learning, which optimizes performance through reduced redundant parameter updates that increase system efficiency. The results demonstrate that the combination of image enhancement with capsule-based feature representation and greedy deep belief learning creates a robust technical framework that performs accurate gastrointestinal disease classification while using fewer computational resources.

Table 5 Computational Complexity Analysis

Model	Parameters (Millions)	FLOPs (G)	Training Time (min)	Inference Time (ms/image)
CapsNet	18.5	8.2	110	42
EfficientNet-B4	19.3	9.1	118	39
Vision Transformer	24.8	10.7	140	46
Proposed DS-CDBN	15.2	6.4	74	28

Table 5 provides a computational complexity analysis that shows how effectively the DS-CDBN model performs when compared to various advanced DL models. The ViT exhibits the highest computational complexity with 24.8 million parameters and 10.7 GFLOPs, resulting in the longest training time of 140 minutes and inference time of 46 ms per image. The self-attention mechanism needs extensive matrix operations to create global feature models, which makes this system processing complex. The CapsNet system needs 18.5 million parameters to operate; EfficientNet-B4 requires 19.3 million parameters, according to their extended training times, which exceed 110 minutes.

The proposed DS-CDBN model demonstrates significantly lower computational requirements with only 15.2 million parameters and 6.4 GFLOPs, while achieving the fastest training time of 74 minutes and inference time of 28 ms per image. The dynamically stacked feature learning method achieves reduced complexity through its

design, which eliminates unnecessary convolutional processing to concentrate on crucial feature extraction areas. The capsule-based representation system uses its vector activations to encode spatial relationships more effectively than deep convolutional stacks. The proximal greedy layer-wise learning mechanism reduces computational requirements by updating parameters through its sequential hierarchical process, rather than using full backpropagation in deep networks within the system. The DS-CDBN architecture establishes an optimal balance between its predictive capabilities and computational performance.

Table 6 Noise Robustness Evaluation

Noise Level	CapsNet Accuracy (%)	EfficientNet Accuracy (%)	DS-CDBN Accuracy (%)
No Noise	92.1	93.8	96.5
5% Noise	89.3	90.4	94.8
10% Noise	85.7	87.2	92.6
15% Noise	82.5	84.1	90.3

The evaluation of noise robustness shown in Table 6 performs various noise levels on endoscopic images. The proposed DS-CDBN model achieves its maximum accuracy of 96.5% under conditions of no noise, which gives it an advantage over both CapsNet and EfficientNet. DS-CDBN achieves better preservation of accuracy at 94.8% while both CapsNet and EfficientNet show decreased performance accuracy of 89.3% and 90.4%, respectively, at 5% noise. The conventional models show increased performance degradation when noise intensity reaches 10% and 15% additional testing. CapsNet accuracy drops to 85.7% and 82.5%, and EfficientNet decreases to 87.2% and 84.1%. The proposed model demonstrates superior performance through its ability to achieve 92.6% and 90.3% accuracy under noisy conditions.

The robustness of the system results from its image quality enhancement module, which uses dynamically stacked median filtering to reduce impulse noise while maintaining essential structural edges in gastrointestinal images. The capsule-based feature representation further improves robustness by encoding spatial relationships and orientation information of tissue structures through vector-based activations. The greedy deep belief network establishes its ability to extract reliable, unique features from images through its capacity to learn multiple levels of image characteristics. DS-CDBN demonstrates its capability to preserve diagnostic data while achieving stable classification results throughout various noisy imaging situations that occur in endoscopic procedures.

Table 7 Hyperparameter Sensitivity Analysis

Capsule Dimension	Learning Rate	Batch Size	Accuracy (%)
8	0.001	32	94.1
16	0.001	32	96.5
32	0.001	32	95.2
16	0.0005	32	95.6
16	0.001	64	95.9

The hyperparameter sensitivity analysis presented in Table 7. The model reaches 94.1% accuracy when its capsule dimension is set to 8 because the reduced feature representation reduces its ability to detect complex spatial patterns found in gastrointestinal images. The accuracy of the model reaches 96.5% when the capsule dimension increases to 16 because vector representations improve the system's ability to learn features. The system loses accuracy to 95.2%, it increases the capsule dimension to 32, because of both additional parameter requirements and potential duplicate patterns in the data.

The experiment demonstrates that changes in learning rate produce direct effects on experimental outcomes because decreasing the rate from 0.001 to 0.0005 leads to a 95.6% accuracy drop. It shows that moderate learning rates provide faster and more reliable convergence results. The accuracy increases to 95.9%, and the batch size is increased from 32 to 64 because larger batch sizes deliver more consistent gradient updates, though slightly change optimization patterns. The proposed model reaches its best performance when using a capsule dimension of 16. This configuration produces the highest accuracy of 96.5%.

The capsule dimension controls vector-based feature representations, which extend lengthwise and show directional movement and track spatial relationships between gastrointestinal tissue structures. A moderate dimension allows the capsule network to capture discriminative spatial information without introducing unnecessary complexity. The learning rate controls the magnitude of gradient updates during backpropagation, which affects both the convergence speed and stability of the greedy deep belief learning mechanism. The batch size controls how many samples are used to calculate gradient updates, which affects both

training stability and generalization ability of the DS-CDBN architecture.

Table 8 Training Convergence Analysis

Model	Epochs to Convergence	Final Loss	Training Stability
CapsNet	65	0.142	Moderate
EfficientNet-B4	72	0.135	Moderate
ViT	80	0.128	High
DS-CDBN	48	0.102	High

The training convergence analysis presented in Table 8 evaluates stably different models that reach optimal performance during the training process. The ViT model needs 80 epochs to reach its final performance because it requires the most time among the tested models, but achieves a 0.128 final loss, which shows the model needs extended training time due to its complex self-attention system. EfficientNet-B4 requires 72 epochs to reach its final performance with a 0.135 loss value, while CapsNet achieves its performance at 65 epochs with a 0.142 loss value. The two models display only moderate training stability because their optimization process experiences irregularities in how gradients are updated during training. The DS-CDBN model achieves faster convergence during its first 48 epochs while reaching a final loss of 0.102. It demonstrates better ability to learn discriminative features from gastrointestinal images. The proposed architecture shows a strong ability to capture essential patterns during the initial phases of training because it demonstrates faster convergence times. The model demonstrates high training stability because its loss progresses to decrease steadily while the optimization process continues without experiencing significant fluctuations.

The dynamic image enhancement stage of the system delivers cleaner input features, which decrease learning difficulties associated with noise-based learning problems. The gastrointestinal tissue structures exhibit spatial relationship preservation through capsule-based feature learning, which uses vector-based activations to create detailed feature representations. The proximal greedy layer-wise deep belief learning method enables easier parameter optimization because it learns hierarchical features in a step-by-step manner instead of updating the entire network at once. The system achieves both computational efficiency and training stability through DS-CDBN reduces gradient instability while speeding up the training process.

6. CONCLUSION

The study developed a new medical diagnostic system called DS-CDBN that used endoscopic images from the Kvasir dataset to detect gastrointestinal tract diseases. The proposed framework integrated image quality enhancement, capsule-based feature learning, and greedy layer-wise deep belief classification to improve diagnostic performance. The dynamically stacked median filtering mechanism effectively enhanced image quality by suppressing noise while preserving important structural edges. Capsule networks provided the model with strong spatial feature representation abilities because they enabled the model to understand complex anatomical relationships that existed in gastrointestinal tissues. The proximal greedy deep belief learning strategy enabled better extraction of hierarchical features while it decreased training time requirements through its reduced computing needs. The experimental results showed that DS-CDBN outperformed all existing DL models in terms of classification accuracy. The system showed higher computational efficiency while it achieved faster training progress and better stability during the learning process. The analysis of noise robustness showed that the model maintained its ability to achieve high accuracy results under conditions of reduced image quality. The architectural design proved to be stable and dependable based on the results of hyperparameter sensitivity testing. The DS-CDBN framework delivered an effective system that automated gastrointestinal disease diagnosis while enhancing clinical decision-making processes within gastroenterology. Future research will focus on extending the proposed DS-CDBN framework to larger and multi-center gastrointestinal datasets to improve generalization capability. The combination of attention-based mechanisms with explainable AI techniques will improve the capacity of the system to provide better results, which doctors need to use in their work. The system can also be adapted for real-time deployment in endoscopic diagnostic platforms.

REFERENCES

[1]. S. Siddiqui, J. A. Khan, T. Akram, M. Alharbi, J. Cha, and D. A. AlHammadi, “SNet: A novel convolutional neural network architecture for advanced endoscopic image classification of gastrointestinal disorders”, *SLAS technology*, vol. *33*, pp. 100304, 2025.

[2]. M. A. Khan, U. Shafiq, A. Hamza, A. M. Mirza, J. Baili, D. A. AlHammadi, and B. Chang, “A

novel network-level fused deep learning architecture with shallow neural network classifier for gastrointestinal cancer classification from wireless capsule endoscopy images" *BMC Medical Informatics and Decision Making*, vol. *25*, no. 1, pp. 150, 2025.

[3]. C. H. Kusters, T. J. Jaspers, T. G. Boers, M. R. Jong, J. B. Jukema, K. N. Fockens, and P. H. De With, "Will Transformers change gastrointestinal endoscopic image analysis? A comparative analysis between CNNs and Transformers, in terms of performance, robustness and generalization," *Medical Image Analysis*, vol. 99, pp. 103348, 2025.

[4]. M. F. Ahamed, M. Nahiduzzaman, M. R. Islam, M. Naznine, M. A. Ayari, A. Khandakar, and J. Haider, "Detection of various gastrointestinal tract diseases through a deep learning method with ensemble ELM and explainable AI," *Expert Systems with Applications*, vol. 256, pp. 124908, 2024.

[5]. A. M. Zubair Rahman, R. Mythili, K. Chokkanathan, T. R. Mahesh, K. Vanitha, and T. E. Yimer, T. E. "Enhancing image-based diagnosis of gastrointestinal tract diseases through deep learning with EfficientNet and advanced data augmentation techniques," *BMC Medical Imaging*, vol. 24, no. 1, pp. 306, 2024.

[6]. B. Mulugeta Abuhayi, Y. Agegnehu Bezabh, A. Melese Ayalew, and M. Alemayehu Lakew, "Classification of gastrointestinal diseases using hybrid recurrent vision transformers with wavelet transform", *Advances in Multimedia*, vol. 1, pp. 8334358, 2024.

[7]. V. Tanwar, B. Sharma, D. P. Yadav, and A. Mehbodniya, "Hybrid deep learning framework based on EfficientViT for classification of gastrointestinal diseases", *Scientific Reports*, vol. 15, no. 1, pp. 26982, 2025.

[8]. A. Saba, J. Amin, and M. U. Ali, "Deep Q-Learning for Gastrointestinal Disease Detection and Classification", *Bioengineering*, vol. 12, no. 11, pp. 1184, 2025.

[9]. A. M. Storås, M., Dreyer, F. Pahde, S. Lapuschkin, W. Samek, P. Halvorsen, and M. A. Riegler, "Exploring the clinical value of concept-based AI explanations in gastrointestinal disease detection", *Scientific Reports*, vol. 15, no. 1, pp. 28860, 2025.

[10]. A. Alruban, E. Alabdulkreem, M. M. Eltahir, A. R. Alharbi, I. Issaoui, I., and A. Sayed, "Endoscopic image analysis for gastrointestinal tract disease diagnosis using nature inspired algorithm with deep learning approach", *IEEE Access*, vol. 11, PP. 130022-130030, 2023.

[11]. K. P. Ajitha Gladis D. Roja Ramani N. Mohana Suganthi and P. Linu Babu, "Gastrointestinal tract disease detection via deep learning based structural and statistical features optimized hexa-classification model", *Technology and Healthcare*, vol. *32*, no. 6, pp. 4453-4473, 2024.

[12]. Y. Y. Wang, B. Liu, and J. H. Wang, "Application of deep learning-based convolutional neural networks in gastrointestinal disease endoscopic examination", *World Journal of Gastroenterology*, vol. 31, no. 36, pp. 111137, 2025.

[13]. V. Sharmila, and S. Geetha, "A recurrent multimodal sparse transformer framework for gastrointestinal disease classification", *Scientific Reports*, vol 15, no. 1, pp. 24206, *15*(1), 24206, 2025.

[14]. M. Ramzan, M. Raza, Z. F. Khan, M. A. Khan, N. Bačanin-Džakula, R. Damaševičius, and Y. Nam, "A review on computer-aided diagnostic system to classify the disorders of the gastrointestinal tract", *European Journal of Medical Research*, vol. 30, no. 1, pp. 674, 2025.

[15]. S. Ali, M. Dmitrieva, N. Ghatwary, S. Bano, G. Polat, A. Temizel, A. Krenzer, A. Hekalo, Y. B. Guo, B. Matuszewski, and M. Gridach, "Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy", *Medical image analysis*, vol. 70, pp.102002, 2021.

[16]. O. M. Mirza, A. Alsobhi, T. Hasanin, M. K. Ishak, F. K. Karim, and S. M. Mostafa, "Computer aided diagnosis for gastrointestinal cancer classification using hybrid rice optimization with deep learning', *IEEE Access*, Vol. 11, pp.76321-76329, 2023.

[17]. S. Ahmad, J. S. Kim, D. K. Park, and T. Whangbo, "Automated detection of gastric lesions in endoscopic images by leveraging attention-based yolov7", *IEEE Access*, vol. 11, pp. 87166-87177, 2023.

[18]. I. Iqbal, K. Walayat, M. U. Kakar, and J. Ma, "Automated identification of human gastrointestinal tract abnormalities based on deep convolutional neural network with endoscopic images", *Intelligent Systems with Applications*, vol. 16, pp.200149, 2022.

[19]. M. Shakhawat Hossain, Mahbubul Syeed, Mahmudur Rahman, Mohammad Faisaluddin, Mahady Hasanim Aulad Hossain, Amel Ksibi, Monam Jamjoom, Zahidullah, Abdus Samad,

"DeepPoly: Deep Learning-Based Polyps Segmentation and Classification for Autonomous Colonoscopy Examination", *IEEE Access*, Vol. 11, 2023

[20]. D. Varam, R. Mitra, M. Mkadmi, R. A. Riyas, D. A. Abuhani, S. Dhou, and A. Alzaatreh, "Wireless capsule endoscopy image classification: an explainable AI approach", *IEEE Access*, vol. 11, pp. 105262-105280, 2023.

[21]. S. Rubab, M. Jamshed, M. A. Khan, N. A. Almujally, R. Damaševičius, A. Hussain, and Y. Nam, "Gastrointestinal tract disease classification from wireless capsule endoscopy images based on deep learning information fusion and Newton Raphson controlled marine predator algorithm", *Scientific Reports*, vol. 15, no. 1, pp. 32180, 2025.

[22]. Y. Zhang, K. Zhang, Y. Ding, S. Liu, M. Wang, X. Wang, Z. Qin, X. Zhang, T. Ma, F. Hu, and P. Li, "Deep transfer learning from ordinary to capsule esophagogastroduodenoscopy for image quality controlling", *Engineering Reports*, vol. 6, no. 5, pp.e12776, 2024.

[23]. H. Shi, H. Wang, X. Ding, and Z. Dang, "An in-depth study of personalized anesthesia management models in gastrointestinal endoscopy based on multimodal deep learning", *IEEE Access*, vol. 13, pp. 15410-15429, 2025.

[24]. M. Hmoud Al-Adhaileh, E. Mohammed Senan, F. W. Alsaade, T. H. H. Aldhyani, N. Alsharif, A. Abdullah Alqarni, M. I. Uddin, M. Y. Alzahrani, E. D. Alzain, and M. E. Jadhav, "Deep learning algorithms for detection and classification of gastrointestinal diseases", *Complexity*, vol. 1, pp.6170416, 2021.

[25]. A. Wrona, M. M. H. Atanasious, and F. Delli Priscoli, "Deep Reinforcement Learning-based Automatic Augmentation for Gastrointestinal Disease Classification", *ACM Transactions on Computing for Healthcare*, vol. 6, no. 3, pp.1-19, 2025.

[26]. A. M. Zubair Rahman, R. Mythili, K. Chokkanathan, T. R. Mahesh, K. Vanitha, T. K. and T. E. Yimer, "Enhancing image-based diagnosis of gastrointestinal tract diseases through deep learning with EfficientNet and advanced data augmentation techniques", *BMC Medical Imaging*, vol. 24, no. 1, pp.306, 2024.

[27]. K. Zhang, H. Wang, Y. Cheng, H. Liu, Q Gong, Q. Zeng, T. Zhang, G. Wei, Z. Wei, and D. Chen, "Early gastric cancer detection and lesion segmentation based on deep learning and gastroscopic images", *Scientific Reports*, vol. 14, no.1, pp.7847, 2024.

[28]. A. Şener, and B. Ergen, "Automatic detection of gastrointestinal system abnormalities using deep learning-based segmentation and classification methods. *Health Information Science and Systems*", vol. 13, no. 1, pp.37, 2025.

[29]. S. Wu, R. Zhang, J. Yan, C. Li, Q. Liu, L. Wang, and H. Wang, "High-speed and accurate diagnosis of gastrointestinal disease: learning on endoscopy images using lightweight transformer with local feature attention", *Bioengineering*, vol. 10, no. 12, pp. 1416, 2023.

[30]. [30] https://www.kaggle.com/datasets/meetnagadia/kvasir-dataset