Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 9 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i09.61119

PAPER

Hybrid Deep Learning and Semantic Segmentation Framework for Colon Cancer Screening in Endoscopic Images

E.N. Srivani[1,2] (✉), G. Seshikala[1]

[1]School of Electronics and Communication, REVA University, Bengaluru, Karnataka, India

[2]Department of Electronics and Communications, SJC Institute of Technology, Chickballapur, Karnataka, India

vani.sri04@gmail.com

ABSTRACT

The rising global population, lifestyle changes, and altered dietary habits have contributed to an increase in gastrointestinal diseases, including colon cancer, necessitating robust computer-aided diagnosis (CAD) systems for early detection and clinical decision-making. This paper proposes a deep semantic segmentation–assisted hybrid deep-ensemble framework (DSH-DeNet) for colon cancer screening using endoscopic images. The proposed framework applies preprocessing and UNet-based region-of-interest (ROI) segmentation to localize salient colorectal regions and enhance feature extraction from endoscopic images. Deep features are extracted using EfficientNet-B7, MobileNet-V2, and DenseNet121, followed by PCA-based feature refinement and z-score normalization. An ensemble-of-ensemble (E2E) learning strategy is employed for robust multi-class endoscopic image classification. With 98.99% accuracy, 99.24% precision, 99.17% recall, and an F-measure of 0.99, the experimental evaluation shows excellent classification performance. The results show that the suggested segmentation-assisted framework for computer-aided colon cancer screening utilizing endoscopic images is successful.

KEYWORDS

colon cancer detection, deep semantic region-of-interest (ROI) segmentation, deep ensemble feature, ensemble-of-ensemble (E2E) ensemble learning

1 INTRODUCTION

Sedentary lifestyles, urbanization, environmental pollution, and dietary changes are some of the causes contributing to the dramatic increase in cancer incidence in recent decades. The unchecked growth of aberrant cells, which can metastasize to other organs, is what defines the disease [1], [2]. The unrestrained development of cancer cells results in the formation of either benign or malignant tumors, in contrast to normal cells that divide only to maintain tissue homeostasis. While

Srivani, E. N., Seshikala, G. (2026). Hybrid Deep Learning and Semantic Segmentation Framework for Colon Cancer Screening in Endoscopic Images. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(9), pp. 59–75. https://doi.org/10.3991/ijoe.v22i09.61119

Article submitted 2026-02-16. Revision uploaded 2026-05-22. Final acceptance 2026-05-22.

malignant tumors can spread to neighboring tissues and frequently recurafter surgical excision, causing serious consequences or even death, benign tumors usually stay confined [3]. The critical need for efficient early diagnostic techniques is highlighted by the World Health Organization's (WHO) prediction of over 20 million new cancer cases and over 9.7 million cancer-related deaths worldwide in 2022 [4]. Lung, breast, and colorectal cancers (CRCs) are among the most common cancers [5]. With around 1.9 million new cases each year and a high mortality rate, CRC is among the most frequently diagnosed cancers worldwide. According to epidemiological data, the incidence of CRC is steadily rising, even in developing countries like India, and its prevalence is rising among younger populations [3]–[6]. Late-stage detection often increases the risk of metastasis and reduces patient survival outcomes. According to studies, about 18% of patients with CRC present with secondary lung involvement, and nearly half of them develop metastatic disease [5, 6]. Early identification is especially difficult because of the colon's complicated anatomical structure and the sluggish growth of precancerous polyps [6]–[9]. A delayed diagnosis makes therapeutic decision-making more difficult and drastically lowers survival rates.

Traditional diagnostic methods are invasive, expensive, time-consuming, and prone to human error. Conversely, non-invasive and trustworthy substitutes for clinical decision assistance are offered by computer-aided diagnosis (CAD) systems [7]. Vision-based CAD systems can now support colon cancer screening using endoscopic and histological images with advancements in medical imaging and artificial intelligence (AI) [1]–[6]. Handcrafted elements, including color, texture, shape, and structural descriptors, were used in earlier CAD techniques [25], [37]. Deep learning architectures are increasingly used in recent techniques to capture intricate feature representations [11]–[37]. The robustness of convolutional neural networks (CNNs) in heterogeneous medical images is limited because they are good at learning local hierarchical features but frequently fall short in capturing global contextual information [10]. The contextual learning capabilities of long short-term memory (LSTM) models are also limited [15], [23]. While hybrid deep ensemble methods aim to improve feature diversity, challenges such as class imbalance and imprecise region-of-interest (ROI) extraction continue to affect screening performance.

This paper presents a hybrid deep-ensemble feature learning model deep semantic segmentation–assisted hybrid deep-ensemble framework (DSH-DeNet) for colon cancer screening using endoscopic images that is supplemented by deep semantic segmentation in order to improve ROI-aware feature extraction. To solve pixel-level imbalance and maintain ROI-specific information, the suggested method combines thorough preprocessing with UNet-based semantic polyp segmentation. EfficientNet-B7, MobileNet-V2, and DenseNet121 are used to extract deep features, which are then improved using principal component analysis (PCA) and z-score normalization before being classified using an ensemble-of-ensemble (E2E) approach.

1.1 Problem definition

The propensity to process whole endoscopic pictures rather than targeted ROIs is a significant drawback of current computer-aided diagnostics. Because of pixel-level imbalances, this "global" technique frequently compromises feature learning and produces biased results. Reliable ROI localization remains challenging because endoscopic imaging often contains heterogeneous textures, illumination variations, and complex anatomical structures. Inaccurate ROI extraction reduces the

usefulness of these instruments in a clinical context and increases false-positive rates. To address these challenges, the proposed framework emphasizes ROI-aware segmentation and heterogeneous deep feature extraction from endoscopic images. Our strategy seeks to improve diagnostic precision and enable more effective, non-invasive CRC screening by overcoming these conventional obstacles.

1.2 Research contributions

The following is a summary of the study's main contributions:

- This paper focuses on proposing a segmentation-assisted hybrid framework for effective computer-aided colon cancer screening based on endoscopic images.
- Pixel-level data imbalance was successfully resolved, and ROI-specific details were preserved by using an automated UNet-based method.
- In order to identify and fuse heterogeneous features, we combined z-score normalization with PCA-based dimensionality reduction using EfficientNet-B7, MobileNet-V2, and DenseNet121.
- According to validation results (98.99% accuracy; 0.99 F-measure), DSH-DeNet performs noticeably better in colon cancer screening.

1.3 Paper organization

The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 details the proposed methodology, Section 4 discusses experimental results, and Section 5 concludes the paper.

2 LITERATURE REVIEW

Existing studies on colon cancer analysis are broadly categorized into histopathology-based approaches and endoscopic image-based CAD systems. The efficiency of deep learning techniques for CRC identification utilizing both histopathological and endoscopic pictures has been demonstrated by recent studies. Several studies have explored histopathological image analysis for CRC assessment. Polyp segmentation before feature extraction greatly improves classification accuracy [11]. VGG-based architectures were used by Yoon et al. [12], who reported better performance than traditional CNNs. In a similar vein, CNN-based models were applied to histopathology datasets by Kather et al. [13] and Wei et al. [14], who achieved accuracies of over 93%.

Additionally, hybrid deep learning frameworks have drawn a lot of interest. High area under the curve (AUC) values were obtained by Iizuka et al. [15] when they used CNNs in conjunction with recurrent neural networks to categorize different types of colorectal tissue. Hamida et al. [17] and Xu et al. [16] used hybrid CNN–ResNet architectures and segmentation-assisted deep learning to show enhanced diagnostic performance. For feature extraction and classification tasks, pre-trained models like Inception-V3, VGG-16, and MobileNet-V2 have been further investigated [18], [19].

Strategies for ensemble learning have demonstrated promise in improving generalization and resilience. CNN-based ensemble models demonstrated good accuracy, according to Sakr et al. [20] and Hasan et al. [21]. While [24] used

DarkNet-based feature extraction in conjunction with support vector machine (SVM) classification, Talukder et al. [22] used hybrid deep-ensemble features. Several studies have attempted ROI-specific analysis through gland or polyp segmentation [29]–[33]; however, most relied on static or handcrafted segmentation techniques that struggle with non-linear endoscopic images. Additionally, handcrafted feature-based approaches [25], [38] lack generalization compared to deep learning models. Overall, challenges related to ROI localization, feature diversity, and robust endoscopic image analysis continue to motivate the development of segmentation-assisted deep ensemble CAD frameworks.

3 PROPOSED WORK

This section discusses the overall proposed model and implementation details. The proposed DSH-DeNet model comprises the following implementation phases. *Phase-1: Data Acquisition, Phase-2: Pre-processing, Phase-3: UNet-based ROI Segmentation, Phase-4: ROI Specific Deep-Ensemble Feature Extraction and Fusion, Phase-5: PCA Based Feature Refinement, Phase-6: Z-score Normalization, Phase-7: E2E Classification. Phase-8: Performance Analysis.* A snippet of the overall proposed DSH-DeNet model is illustrated in Figure 1. The detailed discussion of the proposed DSH-DeNet model for colon cancer screening is given in the subsequent sections.

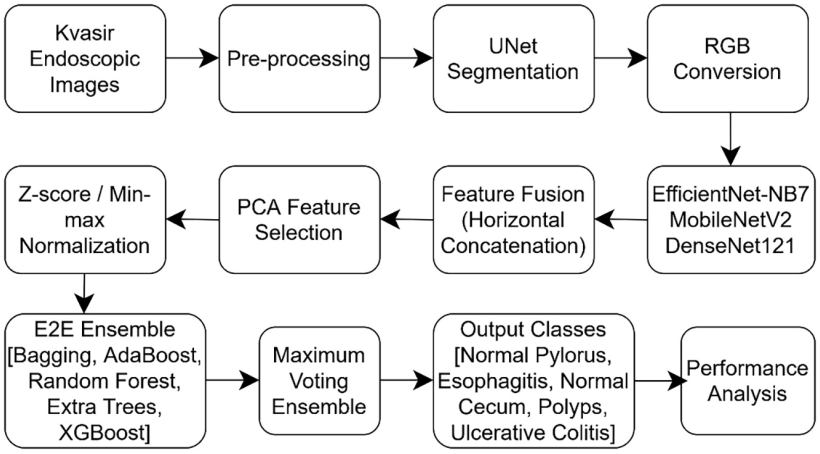

Fig. 1. Proposed DSH-DeNet model for colon cancer detection and classification

3.1 Data acquisition and dataset preparation

The proposed DSH-DeNet framework uses endoscopic pictures instead of invasive colorectal histopathology to increase scalability and patient comfort. Closing the "miss rate" gap (now between 14% and 30% [40]) is crucial because polyp detection rates rise with age and these growths have the potential to develop into malignancy [39]. The proposed framework utilizes separate datasets for segmentation and multi-class endoscopic image classification tasks. Kvasir-Seg [41] and CVC-ColonDB [42] datasets were employed for ROI segmentation due to the availability of pixel-level polyp annotations, while the HyperKvasir [43] classification dataset was used for multi-class colonoscopy image classification involving different gastrointestinal findings and anatomical landmarks. We used a multi-stage preprocessing procedure because of the significant textural and illumination heterogeneity seen in

these gastrointestinal images, as shown in Figure 2. To normalize the data before it moves on to the segmentation and classification stages, this involves shrinking and equalizing the histogram.

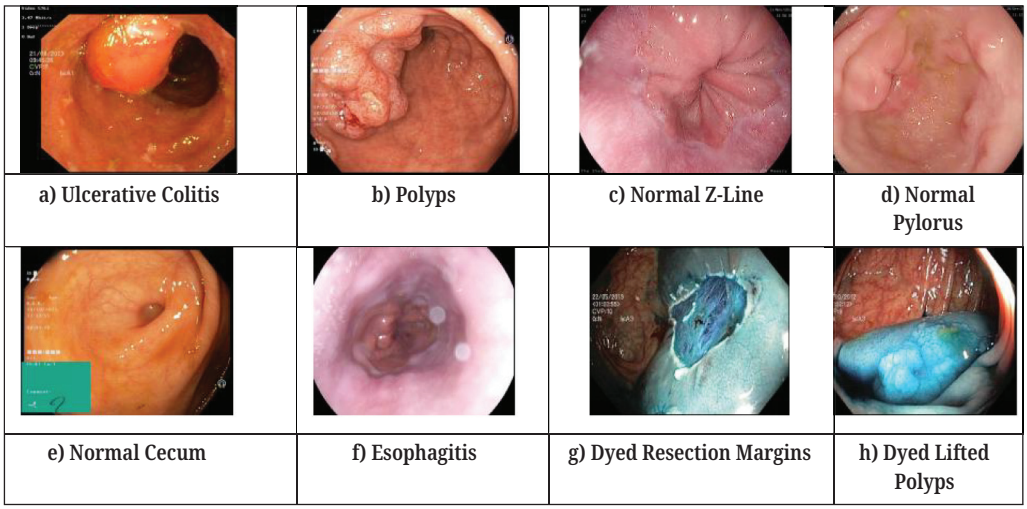

Fig. 2. An illustration of the endoscopic images representing the different gastrointestinal cancer conditions

3.2 Pre-processing

The proposed framework applies intensity equalization, histogram equalization, noise filtering, z-score normalization, and resizing to denoise endoscopic images and ensure uniform pixel distribution for effective feature extraction and learning. Local contrast enhancement and intensity normalization were performed during preprocessing to improve image quality and reduce illumination variations as shown in equation 1.

$$p_n = \left(\frac{\left[u_w(p) - u_w(Min) \right]}{\left[u_w(Max) - u_w(Min) \right]} \right) \tag{1}$$

In equation (1), p refers the pixel number, while Min and Max be the minimum and maximum intensity of the moving window. The parameter u_w refers the targeted minimum and maximum intensity output as shown in equation 2.

$$u_w = \left[1 + exp\left(\frac{f_w - p}{\sigma_w} \right) \right]^{-1} \tag{2}$$

In equation (2), f_w and σ_w are the mean and standard deviation values of each window, which is derived as (3) and (4), respectively.

$$f_w = \frac{1}{N^2} \sum_{(i,j) \in w(x,y)} p(i,j) \tag{3}$$

$$\sigma_w = \frac{1}{N^2} \sum_{(i,j) \in w(x,y)} (p(i,j) - f_w)^2 \tag{4}$$

In equation (4), N be the window size while the pixel positions are defined as x and y over the moving window.In addition to contrast enhancement, intensity and histogram equalization were applied using the imhist() function. Noise removal was performed using median filtering, anisotropic diffusion, and Gabor filtering on each endoscopic image. Subsequently, intensity normalization was applied to the filtered images using the image-wise mean (μ) and standard deviation (σ) to normalize pixel distributions. This normalization maintains a zero mean and near-unit variance, as defined by Equation (5).

$$x^i = \frac{(x - \mu)}{\sigma} \tag{5}$$

Following preprocessing, the proposed DSH-DeNet framework applies a UNet-based segmentation network for ROI localization from endoscopic images. Here, the key motive is to retain Polyp-specific regions from each input colorectal (gastrointestinal tract) endoscopic scanning image. A brief of the UNet segmentation method is given in the subsequent sections.

3.3 UNet-based ROI segmentation

Most existing endoscopic image-based colon cancer detection methods directly apply feature extraction on complete images without considering the proportion of the region of interest (ROI), leading to pixel-level data imbalance since polyp regions are often much smaller than background gastrointestinal areas. This imbalance adversely affects feature learning and classification accuracy.

i) **UNet deep network architecture:**
The proposed work uses an architecture based on UNet [44], which is distinguished by its symmetrical expansion and contracting pathways. During the first contracting phase, the model doubles the feature channels at each down-sampling step using stride-2 max-pooling, extracting deep features using a series of 3×3 convolutions and ReLU activations. The expansion path then replicates this procedure, concatenating cropped feature maps from the previous steps and halving the feature channels through 2×2 up-convolutions to guarantee accurate localization. A 1×1 convolution that directly transfers 64-dimensional features to pixel-wise labels (polyp vs. non-ROI) is the result of this dual-path technique. Overall, the high-fidelity segmentation needed for precise colorectal diagnosis is provided by the 23-layer convolutional layers following the standard UNet architecture.

ii) **UNet Training:**
The input endoscopic images and corresponding segmentation maps were used to train the UNet model using the ADAM optimizer with a learning rate of 0.0001 for 200 epochs. Due to unpadded convolutions, the output size was smaller than the input, and a batch size of 8 was selected to balance segmentation accuracy and computational cost. Training employed a pixel-wise Softmax layer with cross-entropy loss for effective learning and classification. The segmentation loss formulation and weighted cross-entropy strategy follow the standard UNet implementation described in [44]

$$p_k(x) = \frac{exp(a_k(x))}{\left(\sum_{k'}^{k} exp(a_k(x))\right)} \tag{6}$$

In equation (6), $a_k(x)$ represents the activation function for the k-th feature channel at the pixel position of $x \in \Omega$ where $\Omega \in Z^2$. The parameter K states the number of classes, while $p_k(x)$ be the approximation function. The $p_k(x) \approx 1$ for the k having the highest activation $a_k(x)$, and for other values of k, it intends to maintain $p_k(x) \approx 0$. In this work, UNet was trained by using a cross-entropy cost-function so as to penalize at each position the deviation of $p_{l(x)}(x)$ from 1, as per the equation (7).

$$E = \sum_{x \in \Omega} w(x) \log\left(p_{l(x)}(x)\right) \tag{7}$$

In equation (7), l: $\Omega \rightarrow \{1, \ldots, K\}$ represents the label of each pixel, while w: $\Omega \rightarrow \mathbb{R}$ represents the weight map to be applied to measure the pixel's significance during training.

The proposed model computes a class-balanced weight map from the ground truth to compensate for pixel frequency imbalance, enabling improved learning of fine boundary details under non-linear distributions. A morphological operation is applied to estimate separation borders, followed by weight map computation as defined in equation (8).

$$w(x) = w_c(x) + w_0 \cdot exp\left(-\frac{(d_1(x) + d_1(x))^2}{2\sigma^2}\right) \tag{8}$$

In equation (8) w_c: $\Omega \rightarrow \mathbb{R}$ represents the weight map to be applied to balance the class-frequencies, while d_1: $\Omega \rightarrow \mathbb{R}$ and d_1: $\Omega \rightarrow \mathbb{R}$ are the distance to the border of the nearest cell and the second nearest cell, respectively. In this work, to implement UNet $w_0 = 10$ and $\sigma = 5$ pixels were assigned as the fixed values. Moreover, the UNet network was designed with the alternating convolution and ReLU layers. For a 3×3 convolution, a total of 64 channels were used, and therefore, a total number of incoming nodes for each neuron was measured as $N = 9 * 64 = 576$.

3.4 Deep ensemble feature extraction and fusion

To improve feature diversity and representation, accuracy degradation, and insufficient contextual learning in conventional CNN, RNN, and deep architectures, this work employs a deep ensemble feature extraction strategy. Following UNet-based ROI segmentation, ROI-enhanced RGB images are generated for deep feature extraction and processed using EfficientNet-B7, MobileNet-V2, and DenseNet121 to extract complementary features, which are fused to form a deep ensemble representation for learning and classification.

a) EfficientNet-B7

EfficientNet-B7 is an improved version of the native EfficientNet-B0. Here, the input endoscopic images are at first resized to the dimension of 720×576 pixels. The model was designed with the width multiplier factor (w) fixed at 2, though the depth of the multiplier was retained at 3.1.

EfficientNet-B7 is constructed by stacking multiple MBConv (mobile inverted bottleneck convolution) blocks, which differ from classical residual modules by using wider input–output layers and narrower intermediate layers. Each

MBConv block integrates 1×1 convolution, depthwise separable convolution, squeeze-and-attention, dropout, batch normalization, and Swish activation to enhance convergence, non-linearity, and contextual feature representation while reducing overfitting.

$$t_{i,j} = s\left(d_{l_{(i,j)}}\right) = d_{l_{(i,j)}}\left(\frac{1}{1+e^{-\beta\left(d_{l_{(i,j)}}\right)}}\right) \tag{9}$$

In equation (9), t denotes the Swish activation output, and $d_{l_{(i,j)}}$ represents the (i,j)-th entry of matrix D_l; EfficientNet-B7 is employed solely for feature extraction, excluding the Softmax layer. The model utilizes a Squeeze–Excitation mechanism involving global average pooling, non-linear excitation, and channel-wise scaling to assign adaptive weights, producing deep features ($F_{\text{EfficientNet}}$) that are subsequently fused with MobileNet-V2 and DenseNet121 features.

b) MobileNet-V2

A pre-trained backbone was used for the MobileNet-V2 component, but the top-level architecture was substantially altered. In order to stabilize convergence and avoid overfitting, we implemented a modified head that included batch normalization, ReLU activation, and a 0.5 dropout after removing the native GAP, fully connected, and Softmax layers. The strength of the model is its depth-wise separable convolutions, which factorize standard operations into 1×1 point-wise and depth-wise steps. This greatly reduces the computational footprint without compromising the model's capacity to process intricate non-linear inputs. After that, the discriminative features are ready to be fused with the streams of EfficientNet and DenseNet.

c) DenseNet121 Feature Extraction

DenseNet121's dense connectivity pattern, which connects each layer to every one before it, is the reason we integrated it. Throughout its 121 layers, this architecture successfully addresses the "vanishing gradient" issue while promoting widespread feature reuse. DenseNet121 only functions as a feature extractor in our framework. Equation (10) defines a composite function H_l, which concatenates the outputs from its four Dense Blocks:

$$X_l = H_1([X_0, X_1, \ldots, X_{l-1}]) \tag{10}$$

Using a flattening layer and Global Average Pooling (GAP), we were able to construct a learnable feature vector called V_{Dense}. Our unified deep ensemble representation is then produced by horizontally concatenating this vector with features from MobileNet-V2 and EfficientNet-B7:

$$F_{\text{DeepEnsemble}} = [F_{\text{EfficientNetNB7}}, F_{\text{MobileNetV2}}, F_{\text{DenseNet121}}] \tag{11}$$

Feature fusion was performed to combine complementary spatial and semantic representations extracted from heterogeneous deep backbones.

3.5 PCA-based feature refinement and normalization

Despite offering a rich representation, the fused ensemble's large dimensionality may cause overfitting and computing bottlenecks. PCA was applied on the fused

feature vector in equation (11) to reduce feature redundancy while preserving dominant discriminative variance. Feature normalization was subsequently applied to maintain consistent feature scaling prior to classification. The number of retained principal components was selected empirically based on cumulative explained variance. We stabilized the learning process for the E2E ensemble architecture by centering the data at a zero mean (μ) with unit variance (σ). This ensured that the various feature types stay on a consistent scale throughout the classification process.

3.6 E2E learning for multi-class colonoscopy image classification

The proposed E2E learning framework integrates five base classifiers—AdaBoost, Bagging, Random Forest, Extra Trees, and XGBoost—to perform robust multi-class classification of endoscopic images. Each classifier independently assigns labels to the input images across eight gastrointestinal categories, including Normal Pylorus, Esophagitis, Normal Cecum, Polyps, Ulcerative Colitis, Normal Z-line, Dyed Resection Margin, and Dyed Lifted Polyps. Final classification is obtained using a majority voting strategy, improving classification robustness and generalization performance.

The Extra Tree Classifier constructs an ensemble of fully randomized, unpruned decision trees using arbitrary split-point selection and the entire training dataset. Compared to RF, ETC introduces stronger randomization, which reduces variance and enhances generalization. Final predictions are obtained using majority voting across all trees.

XGBoost is an optimized gradient-boosted decision tree ensemble that performs parallel boosting with regularization to prevent overfitting. It models complex non-linear relationships by iteratively minimizing a regularized objective function:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in F \tag{12}$$

The CART space is defined as $F = \{f(x) = wq(x)\}(q: R^m \rightarrow T,\ w \in R^T)$, where q refers to the structure of the tree, and T be the total number of leaves. Each tree f_k states a part of the autonomous value q, while w be the leaf weight. XGBoost algorithm employs decision rules in each tree model, q to classify input data into the leaves, whose output score (i.e., w) is subsequently used to make the final prediction. For ensemble-based classification, the XGBoost algorithm employs a regularization function (13).

$$\ell = \sum_{i=1}^{n} l(\hat{y}_i, y_i) + \sum_{k=1}^{K} \begin{matrix} \Omega(f_k), \\ where\ \Omega(f_k) = \xi T + \frac{1}{2}\xi \|w\|^2 \end{matrix} \tag{13}$$

In equation (13), l be the differential convex loss function that measure the difference between the output $\hat{y}_i$ and the expected class label y_i. $\Omega(f)$ refers the complexity of the tree f_k. In the XGBoost method, ξT and $\xi \|w\|^2$ are applied exclusively to penalize the tree and weights, respectively. It (13) contains a parametric function that is improved by using an optimization technique defined in Euclidean space. Due to

the additive training over t iteration(s), the cost function is updated at the earlier iteration t – 1 by using the improved tree f_k:

$$\ell^{(t)} = \sum_{i=1}^{n} l(\hat{y}_i^{(t-1)} + f_t(x_i), y_i) + \Omega(f_t) \tag{14}$$

The 1st and 2nd order gradients of the cost function.

$$\ell^{(t)} \cong \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{15}$$

In equation (15), $g_i = \delta_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = \delta^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$. As it has been evolved from the native decision tree algorithm, the XGBoost method classifies fixed values within a leaf, $w_q(x)$, where $q(x)$ maps the sample x to a leaf, while w be the score for each leaf, defining the tree $f_k(x)$.

$$\ell^{(t)} \cong \sum_{j=1}^{T} \left[G_j w_i + \frac{1}{2} (H_j + \xi) w_j^2 \right] + \lambda \mathrm{T}, \tag{16}$$

$$\text{where } G_j = \sum_{j \in I_j} g_i, H_j = \sum_{j \in I_j} h_i$$

The cost function reduces $\frac{\delta \ell^{(t)}}{\delta w_i} = G_j + (H_j + \lambda) w_i = 0$, where the optimal weight of the leaf node j is measured as per the equation (17).

$$w^* = -\frac{G_j}{H_j + \xi} \tag{17}$$

The best tree structure is obtained with the objective function defined in (18).

$$\ell^{(t)} = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \xi} + \xi T \tag{18}$$

For a suitable split condition, the XGBoost algorithm applies a greedy technique along with the local and global approximation method on input features (18). Although the classical approximations rely on the direct assessment of the gradient statistics, the XGBoost algorithm functions more efficiently to support both the local as well as the global variant approximation. XGBoost employs regularized gradient boosting to improve classification performance and reduce overfitting during ensemble learning.

$$\ell^{(t)} = \sum_{i=1}^{n} l(\hat{y}_i^{(m-1)} - D_i + \tilde{F}_m, y_i) + \Omega(\tilde{F}_m) \tag{19}$$

Thus, applying the proposed XGBoost algorithm, each of the input endoscopic images was classified into the best possible class (i.e., Normal Pylorus, Esophagitis, Normal Cecum, Polyps, Ulcerates Colitis, Normal z-Line, Dyed Resection Margin, and

Dyed Lifted Polyps classes) and label them accordingly, which is applied later to perform maximum voting-based classification.

Predicted labels from all five base classifiers such as AdaBoost, Bagging, Random Forest, Extra Trees, and XGBoost are aggregated using a maximum voting strategy. The class receiving the highest consensus vote is assigned as the final prediction, supporting robust multi-class endoscopic image classification.

4 RESULTS AND DISCUSSION

Experiments were conducted using CVC-ColonDB and Kvasir-Seg datasets for ROI segmentation analysis, while the HyperKvasir dataset was utilized for multi-class gastrointestinal image classification involving eight categories. Model performance was evaluated using confusion-matrix–based metrics, including accuracy, precision, recall, and F-measure, as given in Table 1. Intra-model assessment analyzed the influence of segmentation, feature models, classifiers, and preprocessing strategies.

Table 1. Performance parameters

Parameter	Mathematical Expression
Accuracy	$\frac{(TN+TP)}{(TN+FN+FP+TP)}$
Precision	$\frac{TP}{(TP+FP)}$
Recall	$\frac{TP}{(TP+FN)}$
F-Measure	$2.\frac{Recall \cdot Precision}{Recall + Precision}$

As shown in Table 2, UNet-based ROI segmentation consistently improved performance over non-segmented inputs, achieving accuracies of 95.63% and 99.32% on CVC-ColonDB and Kvasir-Seg datasets, respectively.

Table 2. Segmentation analysis

Segmentation		Accuracy (%)	Precision (%)	Recall (%)	F-Measure
CVC-ColonDB	Without	94.94	95.01	94.91	0.93
	UNet	95.63	92.99	98.11	0.95
Kvasir-Seg	Without	98.67	99.21	99.16	0.99
	UNet	99.32	99.67	99.29	0.99

The results in Table 3 show a notable improvement in performance when switching from single-model features to our fused ensemble method. The integration of these diverse properties produced the most reliable results, even though DenseNet121 was the strongest backbone when used alone, outperforming both EfficientNet-B7 and MobileNetV2. In particular, the deep-ensemble representation maintained an almost perfect F-measure of 0.99 on both CVC-ColonDB and HyperKvasir, achieving

peak accuracies of 98.67% and 99.21%, respectively. This pattern highlights how ROI-specific feature fusion makes up for the shortcomings of any one design.

Table 3. Feature model assessment

Dataset	Feature Model	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
CVC-ColonDB	EfficientNet-B7	96.79	99.19	95.82	0.96
	MobileNet-V2	94.94	96.32	97.99	0.97
	DenseNet121	97.11	99.27	99.12	0.99
	Deep-Ensemble	98.67	99.21	99.16	0.99
HyperKvasir	EfficientNet-B7	96.77	95.01	96.43	0.95
	MobileNet-V2	95.24	97.69	95.82	0.96
	DenseNet121	98.17	98.21	98.16	0.98
	Deep-Ensemble	99.21	99.67	99.29	0.99

The evaluation in Table 4 shows that the suggested E2E ensemble performed better on both testing datasets than standalone models such as RF, ETC, and XGBoost. We found that the majority-voting approach produced better metrics overall by effectively reducing the individual mistakes of its base learners. In particular, the model achieved 99.32% accuracy on HyperKvasir and 98.67% accuracy on the CVC-ColonDB dataset. These results demonstrate the E2E technique as the effective approach for multi-class endoscopic image classification, surpassing individual classifier benchmarks.

Table 4. Classifier assessment

Dataset	Feature Models	Classifier	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
CVC-ColonDB	Deep-Ensemble	Bagging	94.05	94.37	97.26	0.96
		AdaBoost	94.31	94.17	97.98	0.96
		RF	96.22	97.42	98.12	0.97
		ETC	96.54	97.99	98.47	0.98
		XGBoost	98.21	98.01	98.37	0.97
		E2E-MVE	98.67	99.21	99.16	0.99
HyperKvasir	Deep-Ensemble	Bagging	94.21	96.27	98.17	0.97
		AdaBoost	94.74	96.32	97.99	0.97
		RF	96.17	97.89	98.82	0.98
		ETC	97.66	98.17	98.71	0.98
		XGBoost	98.96	98.72	99.03	0.98
		E2E-MVE	99.32	99.27	99.19	0.99

Anisotropic diffusion filtering offered a marginally consistent advantage over median and Gabor filtering, according to data from our preprocessing evaluation (Table 5). The filter's exceptional ability to reduce noise without sacrificing the crisp edge and texture characteristics necessary for salient ROI identification is probably

the cause of this performance improvement. We obtained a strong 98.99% average accuracy and a 0.99 F-measure by using this improved preprocessing to anchor the DSH-DeNet framework in conjunction with UNet segmentation and E2E classification. These results demonstrate the effectiveness of the integrated segmentation-assisted deep ensemble framework.

Table 5. Pre-processing assessment

Dataset	Feature Model	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
CVC-ColonDB	2D Medium Filter	97.93	97.71	97.24	0.96
	Gabor Filter	97.79	98.01	97.77	0.96
	Anisotropic Diffusion	98.67	99.21	99.16	0.99
HyperKvasir	2D Medium Filter	98.11	98.03	95.33	0.95
	Gabor Filter	98.87	98.11	98.52	0.97
	Anisotropic Diffusion	99.32	99.27	99.19	0.99

Table 6 compares the proposed DSH-DeNet framework with existing endoscopic image-based CAD approaches reported in the literature. The proposed framework demonstrated competitive classification performance with an overall accuracy of 99.32%.

Table 6. Inter-model assessment

Reference	Accuracy (%)
VGGNet [12]	93.48
CNN [13]	94.00
RNNs [15]	93.50
DNN with segmentation [16]	94.90
ResNet and CNN [17]	98.66
Inception-V3, VGG-16, and AlexNet + Bayesian SVM [18]	96.50
CNN and MobileNetV2 [19]	97.49
Bi-LSTM [23]	98.00
DenseNet-121 and RF [25]	98.60
[27]	96.33
[29]	81.00
VGG-16 [31]	88.03
CNN + Active Contour Segmentation [32]	90.00
DSH-DeNet	99.32

5 CONCLUSION

This paper presents a segmentation-assisted hybrid deep learning framework for endoscopic image-based colon cancer screening. The DSH-DeNet model,

which forms the basis of our contribution, extracts deep features by combining DenseNet, MobileNet, and EfficientNet in a diverse architectural blend after first isolating salient ROIs using UNet segmentation. Before transferring the data into a multi-layered E2E learning framework, the proposed work used a PCA for feature refinement to make sure the model remained reliable and effective. The proposed framework demonstrated competitive classification performance across the evaluated endoscopic image datasets, with an F-measure of 0.99 and high overall precision.

6 ACKNOWLEDGMENTS

The authors warmly acknowledge the VGST, Govt. of Karnataka, and REVA University, Bengaluru, Karnataka, for providing the facilities required to carry out the research. They also acknowledge the support extended by VGST, Govt.of Karnataka, in sponsoring this research with software tools and project no. KSTePS/VGST/GRD-684/KFIST(LI)/2018. They also acknowledge the School of ECE, REVA University for supporting this work.

7 REFERENCES

[1] National Cancer Institute, "Cancer statistics," 2025. Available online: https://www.cancer.gov/about-cancer/understanding/statistics [Accessed: June 10, 2024].

[2] R. A. Weinberg, "How cancer arises," *Sci. Am.*, vol. 275, pp. 62–70, 1996. https://doi.org/10.1038/scientificamerican0996-62

[3] National Cancer Institute, "What is cancer?" 2024. Available online: https://www.cancer.gov/about-cancer/understanding/what-is-cancer [Accessed: June 10, 2024].

[4] WHO, "Global cancer burden growing amidst mounting need for services," 2024. Available online: https://www.who.int/news-room/fact-sheets/detail/cancer [Accessed: June 10, 2024].

[5] B. N. Ravi Kumar, N. C. Gowda, B. J. Ambika, H. N. Veena, B. B. Sujitha, and D. R. Ramani, "An efficient breast cancer detection using machine learning classification models," *Int. J. Onl. Eng.*, vol. 20, no. 13, pp. 24–40, 2024. https://doi.org/10.3991/ijoe.v20i13.50289

[6] A. Karthikeyan, S. Jothilakshmi, and S. Suthir, "Colorectal cancer detection based on convolutional neural networks (CNN) and ranking algorithm," *Measurement: Sensors*, vol. 31, p. 100976, 2024. https://doi.org/10.1016/j.measen.2023.100976

[7] World Cancer Research Fund/American Institute for Cancer Research, "Diet, nutrition, physical activity and cancer: A global perspective," Continuous Update Project Expert Report 2018. Available: http://dietandcancerreport.org

[8] T. K. Noah, B. Donahue, and N. F. Shroyer, "Intestinal development and differentiation," *Experimental Cell Research*, vol. 317, no. 19, pp. 2702–2710, 2011. https://doi.org/10.1016/j.yexcr.2011.09.006

[9] S. Gout and J. Huot, "Role of cancer microenvironment in metastasis: Focus on colon cancer," *Cancer Microenvironment*, vol. 1, no. 1, pp. 69–83, 2008. https://doi.org/10.1007/s12307-008-0007-2

[10] H. Alqahtani, E. Alabdulkreem, F. A. Alotaibi, M. M. Alnfiai, C. Singla, and A. S. Salama, "Improved water strider algorithm with convolutional autoencoder for lung and colon cancer detection on histopathological images," *IEEE Access*, vol. 12, pp. 949–956, 2024. https://doi.org/10.1109/ACCESS.2023.3346894

[11] H. N. Veena, K. K. Patil, P. Vanajakshi, A. Ambore, and N. C. Gowda, "An enhanced RNN-LSTM model for fundus image classification to diagnose glaucoma," *SN Comput. Sci.*, vol. 5, no. 5, 2024. https://doi.org/10.1007/s42979-024-02867-5

[12] H. Yoon *et al.*, "Tumor identification in colorectal histology images using a convolutional neural network," *J. Digit. Imaging*, vol. 32, no. 1, pp. 131–140, 2018. https://doi.org/10.1007/s10278-018-0112-9

[13] J. N. Kather *et al.*, "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study," *PLoS Med*, vol. 16, no. 1, p. e1002730, 2019. https://doi.org/10.1371/journal.pmed.1002730

[14] J. W. Wei *et al.*, "Evaluation of a deep neural network for automated classification of colorectal polyps on histopathologic slides," *JAMA Netw Open*, vol. 3, no. 4, p. e203398, 2020. https://doi.org/10.1001/jamanetworkopen.2020.3398

[15] O. Iizuka, F. Kanavati, K. Kato, M. Rambeau, K. Arihiro, and M. Tsuneki, "Deep learning models for histopathological classification of gastric and colonic epithelial tumours," *Sci. Rep.*, vol. 10, no. 1, 2020. https://doi.org/10.1038/s41598-020-58467-9

[16] L. Xu *et al.*, "Colorectal cancer detection based on deep learning," *Journal of Pathology Informatics*, vol. 11, no. 1, p. 28, 2020. https://doi.org/10.4103/jpi.jpi_68_19

[17] A. Ben Hamida *et al.*, "Deep learning for colon cancer histopathological images analysis," *Computers in Biology and Medicine*, vol. 136, p. 104730, 2021. https://doi.org/10.1016/j.compbiomed.2021.104730

[18] T. Babu, T. Singh, D. Gupta, and S. Hameed, "Colon cancer prediction on histological images using deep learning features and Bayesian optimized SVM," *Journal of Intelligent & Fuzzy Systems*, vol. 41, no. 5, pp. 5275–5286, 2021. https://doi.org/10.3233/JIFS-189850

[19] Z. Tasnim *et al.*, "Deep learning predictive model for colon cancer patient using CNN-based classification," *IJACSA*, vol. 12, no. 8, 2021. https://doi.org/10.14569/IJACSA.2021.0120880

[20] A. S. Sakr, N. F. Soliman, M. S. Al-Gaashani, P. Pławiak, A. A. Ateya, and M. Hammad, "An efficient deep learning approach for colon cancer detection," *Applied Sciences*, vol. 12, no. 17, p. 8450, 2022. https://doi.org/10.3390/app12178450

[21] M. I. Hasan, M. S. Ali, M. H. Rahman, and M. K. Islam, "Automated detection and characterization of colon cancer with deep convolutional neural networks," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–12, 2022. https://doi.org/10.1155/2022/5269913

[22] Md. A. Talukder, Md. M. Islam, M. A. Uddin, A. Akhter, K. F. Hasan, and M. A. Moni, "Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning," *arXiv preprint arXiv:2206.01088*, 2022. https://doi.org/10.48550/ARXIV.2206.01088

[23] E. Bostanci, E. Kocak, M. Unal, M. S. Guzel, K. Acici, and T. Asuroglu, "Machine learning analysis of RNA-seq data for diagnostic and prognostic prediction of colon cancer," *Sensors*, vol. 23, no. 6, p. 3080, 2023. https://doi.org/10.3390/s23063080

[24] S. Haddou Bouazza, "Novel framework for robust gene selection and accurate multi-cancer classification," *Int. J. Onl. Eng.*, vol. 21, no. 9, pp. 81–95, 2025. https://doi.org/10.3991/ijoe.v21i09.54669

[25] N. Kumar, M. Sharma, V. P. Singh, C. Madan, and S. Mehandia, "An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images," *Biomedical Signal Processing and Control*, vol. 75, p. 103596, 2022. https://doi.org/10.1016/j.bspc.2022.103596

[26] M. Yildirim and A. Cinar, "Classification with respect to colon adenocarcinoma and colon benign tissue of colon histopathological images with a new CNN model: MA_ColonNET," *Int. J Imaging Syst. Tech.*, vol. 32, no. 1, pp. 155–162, 2021. https://doi.org/10.1002/ima.22623

[27] M. Masud, N. Sikder, A.-A. Nahid, A. K. Bairagi, and M. A. AlZain, "A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework," *Sensors*, vol. 21, no. 3, p. 748, 2021. https://doi.org/10.3390/s21030748

[28] P. Sharma, K. Bora, K. Kasugai, and B. Kumar Balabantaray, "Two stage classification with CNN for colorectal cancer detection," *Oncologie*, vol. 22, no. 3, pp. 129–145, 2020. https://doi.org/10.32604/oncologie.2020.013870

[29] A. Karthikeyan, S. Jothilakshmi, and S. Suthir, "Colorectal cancer detection based on convolutional neural networks (CNN) and ranking algorithm," *Measurement: Sensors*, vol. 31, p. 100976, 2024. https://doi.org/10.1016/j.measen.2023.100976

[30] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2015, pp. 79–83. https://doi.org/10.1109/ISBI.2015.7163821

[31] N. Bayramoglu and J. Heikkilä, "Transfer learning for cell nuclei classification in histopathology images," in *Lecture Notes in Computer Science*, Springer International Publishing, 2016, pp. 532–539. https://doi.org/10.1007/978-3-319-49409-8_46

[32] H. Haj-Hassan, A. Chaddad, Y. Harkouss, C. Desrosiers, M. Toews, and C. Tanougast, "Classifications of multispectral colorectal cancer tissues using convolution neural network," *Journal of Pathology Informatics*, vol. 8, no. 1, p. 1, 2017. https://doi.org/10.4103/jpi.jpi_47_16

[33] P. Kainz, M. Pfeiffer, and M. Urschler, "Segmentation and classification of colon glands with deep convolutional neural networks and total variation regularization," *PeerJ.*, vol. 5, p. e3874, 2017. https://doi.org/10.7717/peerj.3874

[34] J. M. Tomczak, I. Maximilian, and M. Welling, "DeepLearning with permutation-invariant operator for multi-instance histopathology classification," *arXiv preprint arXiv:1712.00310*, 2017. https://doi.org/10.48550/arXiv.1712.00310

[35] H. Yoshida *et al.*, "Automated histological classification of whole slide imagesof colorectal biopsy specimens," *Oncotarget*, vol. 8, no. 53, pp. 90719–90729, 2017. https://doi.org/10.18632/oncotarget.21819

[36] D. Bychkov *et al.*, "Deep learning based tissue analysis predicts outcome in colorectal cancer," *Sci. Rep.*, vol. 8, art. no. 3395, 2018. https://doi.org/10.1038/s41598-018-21758-3

[37] J. Zhang, Y. Xia, Y. Xie, M. Fulham, and D. D. Feng, "Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 5, pp. 1521–1530, 2018. https://doi.org/10.1109/JBHI.2017.2775662

[38] M. Dabass, R. Vig, and S. Vashisth, "Five-grade cancer classification of colon histology images via deep learning," in *Communication and Computing Systems*, CRC Press, 2019, pp. 18–24. https://doi.org/10.1201/9780429444272-3

[39] I. D. Irawati, I. Andrea Larasaty, and S. Hadiyoso, "Comparison of convolution neural network architecture for colon cancer classification," *Int. J. Onl. Eng.*, vol. 18, no. 3, pp. 164–172, 2022. https://doi.org/10.3991/ijoe.v18i03.27777

[40] J. C. van Rijn, J. B. Reitsma, J. Stoker, P. M. Bossuyt, S. J. van Deventer, and E. Dekker, "Polyp miss rate determined by tandem colonoscopy: A systematic review," *Am J. Gastroenterology*, vol. 101, no. 2, pp. 343–350, 2006. https://doi.org/10.1111/j.1572-0241.2006.00390.x

[41] D. Jha *et al.*, "Kvasir-SEG: A segmented polyp dataset," in *Lecture Notes in Computer Science*, Springer International Publishing, 2019, pp. 451–462. https://doi.org/10.1007/978-3-030-37734-2_37

[42] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Trans. Med. Imaging*, vol. 35, no. 2, pp. 630–644, 2016. https://doi.org/10.1109/TMI.2015.2487997

[43] H. Borgli *et al.*, "*HyperKvasir*, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Sci. Data*, vol. 7, no. 1, art. no. 283, 2020. https://doi.org/10.1038/s41597-020-00622-y

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science*, Springer International Publishing, 2015, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

8 AUTHORS

E.N. Srivani is working as an Assistant Professor in the Department of Electronics and Communication Engineering at SJC Institute of Technology and has 13 years of teaching experience. She received her B.Tech degree in Electronics and Communication Engineering from BITIT, JNTUH, Hindupur. She completed her M.Tech degree in Embedded Systems from St. Mary's College of Engineering, JNTUH, Hyderabad. She is currently pursuing her Ph.D. in Electronics and Communication Engineering at REVA University, Bengaluru. She has published papers in refereed international journals and has been granted two patents. She has received several Best Paper Awards at various international conferences. Six KSCST funded projects have been sanctioned to her. Her areas of research include image processing and embedded systems. She is an active member of IETE, IEI and ISTE (E-mail: vani.sri04@gmail.com).

Dr. G. Seshikala is a Professor in the School of Electronics and Communication Engineering at REVA University, Bengaluru, India, with 29 years of teaching and research experience. She holds a Ph.D. and specializes in biomedical signal processing with interests in image processing and signal processing. She has organised FDPs, workshops, conferences and seminars. Dr. Seshikala has authored 15 refereed journal publications and 25 conference papers, contributed 3 book chapters, and holds 4 patents published. She has supervised 40+ UG/PG projects, guided 5 Ph.D. scholars, and is currently guiding 6 Ph.D. candidates. She has also secured one funded research project under VGST of 20L and received the Best Researcher Award in 2018 and 2021.