Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 1, January 2026, pp: 240-256 e-ISSN: 2656-8632

RESEARCH ARTICLE

OPEN ACCESS

Improving the Segmentation of Colorectal Cancer from Histopathological Images Using a Hybrid Deep Learning Pipeline: A Case Study

Fahima IDIRI[1], Farid MEZIANE[2], Hakim BOUCHAL[3]

[1] Université de Bejaia, Faculté des Sciences Exactes, Laboratoire d'Informatique Médicale et des Environnements Dynamiques et Intelligents (LIMED), Bejaia 06000, Algeria
[2] University of Derby, Data Science Research Center, United Kingdom
[3] Université de Bejaia, Faculté de Technologie, Laboratoire d'Informatique Médicale et des Environnements Dynamiques et Intelligents (LIMED), Bejaia 06000, Algeria

Corresponding author: Fahima IDIRI (e-mail: fahima.idiri@univ-bejaia.dz, ORCID), **Author(s) Email**: Farid MEZIANE (e-mail: F.Meziane@derby.ac.uk, ORCID), Hakim BOUCHAL (e-mail hakim.bouchal@univ-bejaia.dz)

Abstract **Early and precise diagnosis of colorectal cancer plays a crucial role in enhancing patients' outcomes. Although histopathological assessment remains the reference standard for diagnosis, it is often lengthy and subject to variability between pathologists. This study aims to develop and evaluate a hybrid deep learning-based approach for the automated segmentation of Hematoxylin and Eosin-stained colorectal histopathology images. The work investigates how preprocessing strategies and architectural design choices influence the model's ability to identify meaningful tissue patterns while preserving computational efficiency. Furthermore, it demonstrates the integration of a deep learning-based segmentation module into colorectal cancer diagnostic workflows. Several deep learning–based segmentation models with varying architectural configurations were trained and evaluated using a publicly available endoscopic biopsy histopathological hematoxylin and eosin image dataset. Preprocessing procedures were applied to generate computationally efficient image representations, thereby improving training stability and overall segmentation performance. The best-performing configuration achieved a segmentation accuracy of 0.97, reflecting consistent and reliable performance across samples. It accurately delineated cancerous tissue boundaries and effectively distinguished benign from malignant regions, demonstrating sensitivity to fine morphological details relevant to diagnosis. Strong agreement between predicted and expert-annotated regions confirmed the model's reliability and alignment with expert assessments. Minimal overfitting was observed, indicating stable training behavior and robust generalization across different colorectal tissue types. In comparative evaluations, the model maintained high accuracy across all cancer categories and outperformed existing state-of-the-art approaches. Overall, these findings demonstrate the model's robustness, efficiency, and adaptability, confirming that careful architectural and preprocessing optimization can substantially enhance segmentation quality and diagnostic reliability. The proposed approach can support pathologists by providing accurate tissue segmentation, streamlining diagnostic procedures, and improving clinical decision-making. This study underscores the value of optimized deep learning models as intelligent decision-support tools for efficient and consistent colorectal cancer diagnosis.**

Keywords **Colorectal cancer; Histopathological Hematoxylin and Eosin Images; Deep Learning; intelligent decision-support tools.**

I. Introduction

Colorectal cancer (CRC), also referred to as bowel cancer, constitutes a significant global oncological burden, with high incidence and mortality rates worldwide [1]. Recent estimates suggest that, in 2024 alone, approximately 2.2 million new cases of CRC were diagnosed worldwide, with an associated 1.1 million deaths [2]. These figures highlight the pressing need for improved diagnostic, prognostic, and treatment approaches to reduce the global burden of CRC. Colorectal cancer is primarily detected through colonoscopic examination followed by a histopathological evaluation. Pathological assessment is regarded as the gold standard for confirming whether a lesion is benign or malignant. In conventional practice, pathologists examine stained tissue

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1158

specimens under a microscope to identify abnormal cellular structures that are indicative of incipient neoplastic lesions within limited histological regions [3]. Digital pathology has emerged as a prominent biomedical research field, largely supported by the development of whole slide imaging (WSI) technology. WSI enables digitization of entire histological slides at high resolution, facilitating detailed examination of tissue samples [4]. This approach has become increasingly valuable in colonoscopy-based pathological analysis [5]. The high resolution of WSIs enables visualization of intricate morphological features. However, their large file sizes make manual inspection by pathologists both time-consuming and labor-intensive. Moreover, accurate diagnosis demands significant expertise, posing a challenge for healthcare facilities in resource-limited settings, particularly in rural areas of developing countries, where there is often a shortage of experienced pathologists [5], [6]. To address these challenges, deep learning-based methods have emerged as promising tools in medical image analysis. These models have demonstrated notable performance in many tasks [5], [7]. It is important to recognize that extensive prior research has examined the use of diverse deep learning methods for CRC segmentation. The present literature review prioritizes studies closely aligned with our research objectives, notably [4], [8], [9],[10], and [11], to compare and evaluate the performance of different deep learning approaches for diagnosing malignancies of the colon using H&E-stained histological images.

Cheng et al. [8] developed a publicly available dataset, EBHI-Seg, consisting of 4,456 histopathological images of colorectal tissue covering different stages of tumor progression. The dataset classified the images into six types: normal tissue, polyp, low-grade intraepithelial neoplasia, high-grade intraepithelial neoplasia, serrated adenoma, and adenocarcinoma. Each image was annotated with a corresponding ground-truth segmentation mask, allowing precise evaluation of segmentation models. The study evaluated the performance of several machine learning (ML) and deep learning (DL) approaches using this dataset. Conventional ML methods achieved a maximum Dice coefficient of 0.65, with Precision and Recall values of 0.70 and 0.90, respectively. In contrast, DL-based models outperformed traditional approaches, with the best-performing model reaching a Dice score of 0.95 and both Precision and Recall attaining 0.90. Liu et al. [11] provided a recent review of DL methods used for segmenting colorectal cancer histopathology images. The review examines various types of models, including conventional convolutional neural networks (CNNs), U-Net-based architectures, and attention-enhanced networks. The authors identified several key challenges in this field, such as variations in tissue staining, complex tissue structures, and class imbalance in the data. They further emphasize the advantages of employing pre-trained networks and data augmentation strategies to enhance segmentation performance. In addition, the review notes the growing interest in transformer-based models, which can capture more global image features and may improve results in future research. Sengupta et al. [4] systematically evaluated several U-Net variants to delineate colorectal adenocarcinoma regions in H&E-stained histopathological images derived from the EBHI-Seg dataset. The study compared six architectures: U-Net, Attention U-Net, and U-Net models incorporating ResNet50, MobileNetV2, EfficientNet-B0, and DenseNet121 backbones. In extensive experiments on 795 images for binary segmentation of cancerous versus non-cancerous tissue, the authors reported that U-Net models with DenseNet121 and ResNet50 backbones achieved the best performance, attaining testing accuracies of 90.21 and 89.81, with corresponding Dice coefficients of 94.42 and 94.17, respectively. Their findings indicated that the choice of backbone architecture substantially influenced segmentation outcomes and provided a reliable benchmark for subsequent CNN-driven adenocarcinoma research. Xuan et al. [10] proposed MASK2TASKS, a DL approach that combines segmentation and classification tasks to improve performance in colorectal cancer histopathology image analysis. Their approach leverages segmentation masks to guide attention toward the most informative regions during the classification process. This joint learning strategy leads to better results in both segmentation and classification, as shown by improved accuracy and F1-score on their test datasets. The study demonstrated that combining these two tasks can help models better handle the complexity and variability of histopathology images. Sun and Sheng [9] proposed a Double-Level Fusion Domain Adapter Vision Transformer (DDViT) that integrates CNNs and ViTs through a hierarchical encoder-decoder architecture. DDViT introduces a Double-Level Fusion (DLF) module and employs a plug-in domain adapter within the transformer branch of a TransFuse-based encoder. The domain adapter uses domain-aware attention to modulate multi-head self-attention outputs, improving robustness across domains. Furthermore, DDViT leverages mutual knowledge distillation between a universal network and domain-specific branches, enhancing segmentation performance. Extensive experiments on the EBHI-Seg dataset demonstrated that DDViT achieved superior results compared with CNN-only and transformer-only models, confirming the effectiveness of this hybrid domain-adaptive approach.

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1158

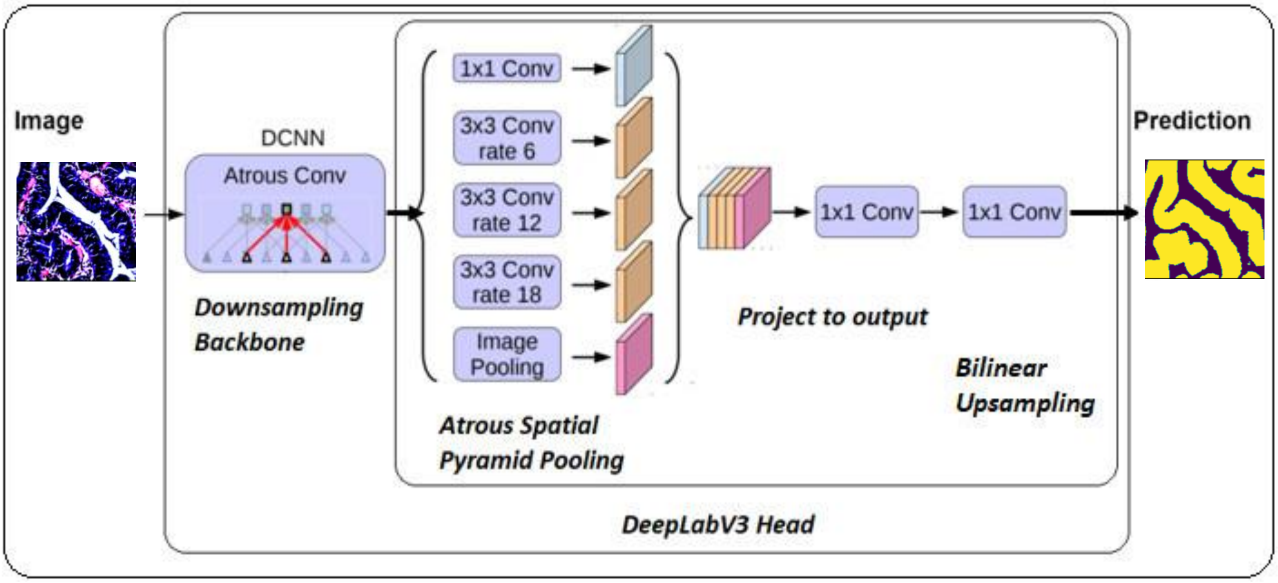


Fig. 1. **DeepLabV3 architecture**

To build on these advancements, the core contributions of this paper can be stated as follows: 1) We conduct a systematic comparison of ResNet-50 [12], ResNet-101 [12], and MobileNetV3_Large [13] backbones within a DeepLabV3 [14] segmentation framework on the EBHI-Seg dataset to identify the most effective feature extractor for CRC segmentation. 2) We propose an efficient data preprocessing and augmentation strategy designed for histopathology images to enhance training stability, model robustness, and generalization. 3) We conduct an extensive binary segmentation evaluation across multiple tissue categories, providing quantitative metrics to validate the performance of the proposed approach. 4) We introduce a feasible workflow that incorporates segmentation techniques powered by deep learning into colorectal cancer diagnostic pipelines, supporting pathologists and enhancing diagnostic decision processes.

This study is structured as follows: Section II presents the proposed method, including the dataset, data splitting, and data augmentation strategies. Section III outlines the assessment criteria, reports the accuracy of our approach, compares it with existing methods, and illustrates representative segmentation results. Section IV introduces a workflow that integrates deep learning–based segmentation models into critical stages of the histopathological diagnostic process. Section V addresses the study's limitations and outlines avenues for future work. Finally, Section VI concludes the study by restating the objectives and summarizing the key findings.

II. Method

A. Architecture

The DeepLabV3 architecture comprises two key components: a backbone that produces high-resolution feature maps via atrous convolutions, and a DeepLabV3 head that captures multi-scale features, maps them to the desired number of segmentation classes, and upsamples them to the original image resolution [14]. Fig. 1 illustrates this architecture.

The modular nature of DeepLabV3 enables flexible combination of its blocks to achieve the desired performance. In our experiments, we used three different pretrained backbones, namely: ResNet-50 [12], ResNet-101 [12] and MobileNetV3_Large [13]. We obtained different performance metrics, and ResNet-101 achieved the best performance.

B. Theoretical Background

Let $\boldsymbol{D} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1,\dots,N}, \quad (1)$

Eq. (1) denotes the training dataset, where $\boldsymbol{x}_i \in \mathbb{R}^{H\times W\times C}$ represents the i-th input histopathology image patch with height H, width W, and C channels (with C=3 for RGB images).

$\boldsymbol{y}_i \in \{0,1\}^{H\times W}$ is the corresponding binary ground truth segmentation mask, where each pixel takes the value 1 for the target region and 0 for the background. The variable N denotes the total number of (image, mask) pairs in the dataset. The objective is to learn a mapping, defined in Eq. (2):

$$\boldsymbol{M}_\theta : \mathbb{R}^{H\times W\times C} \to [0,1]^{H\times W}, \quad (2)$$

that predicts a segmentation probability map $\hat{\boldsymbol{y}}$ for a given input image $\boldsymbol{x}$, where θ represents the learnable parameters.

The DeepLabV3 segmentation model is formulated as a composition of three main components, as in Eq. (3):

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1158

$$\hat{\boldsymbol{y}} = \boldsymbol{M}_{\theta}(\boldsymbol{x}) = \boldsymbol{U}(\mathrm{ASPP}(\boldsymbol{E}_b(\boldsymbol{x},\theta_E),\theta_A),\theta_U), \tag{3}$$

where Eq. (4)

$$\boldsymbol{E}_b : \mathbb{R}^{H\times W\times C} \rightarrow \mathbb{R}^{H'\times W'\times D}, \tag{4}$$

is the Deep Convolutional Neural Network (DCNN) encoder backbone network indexed by b responsible for extracting multi-scale feature representations. The ASPP module is defined in Eq. (5),

$$\mathrm{ASPP} : \mathbb{R}^{H'\times W'\times D} \rightarrow \mathbb{R}^{H'\times W'\times D'}, \tag{5}$$

the Atrous Spatial Pyramid Pooling module, applies parallel atrous convolutions to capture contextual information at multiple receptive field sizes. According to Eq. (6), the decoder

$$\boldsymbol{U} : \mathbb{R}^{H'\times W'\times D'} \rightarrow [0,1]^{H\times W}, \tag{6}$$

performs feature refinement followed by bilinear upsampling to recover the original spatial resolution. The reduced feature-map height and width are given by Eq. (7), Eq. (8) and Eq. (9)

$$H' = H/s\,, \tag{7}$$

And

$$W' = W/s\,, \tag{8}$$

where s is the output stride, and

$$\theta = (\theta_E, \theta_A, \theta_U), \tag{9}$$

represents the complete set of learnable parameters associated with the encoder, ASPP module, and decoder, respectively. Three pre-trained backbone architectures are evaluated:

$$b \in \{\mathrm{ResNet50}, \mathrm{ResNet101}, \mathrm{MobileNetV3\_Large}\},$$

Residual networks ResNet-50 and ResNet-101 use *identity (skip) connections* to ease training in very deep networks by learning residual mappings. A generic residual block is written as in Eq. (10):

$$y_l = R(x_l\,, \boldsymbol{\theta}_l) + S(x_l), \tag{10}$$

where x_l and y_l are the input and output feature maps of the block l, $R(\cdot,\ \boldsymbol{\theta}_l)$ is the residual function (a small stack of convolution, normalization and activation layers) parameterized by $\boldsymbol{\theta}_l$, S $(\cdot)$ is the skip mapping (usually identity, or a learned linear projection when shapes differ). After the addition, an activation may be applied as in Eq. (11):

$$x_{l+1} = \sigma(y_l)\,, \tag{11}$$

The Bottleneck residual block used in ResNet-50/101 reduces parameter cost while preserving representational power. For block ℓ the residual function is defined in Eq. (12):

$$R(x) = W_{3,l} * \ \sigma(BN(W_{2,l} * \ \sigma(BN(W_{1,l} * x)))), \tag{12}$$

where $(W_{1,l} * .)$ is a 1 × 1 convolution that reduces dimensionality, $(W_{2,l} * .)$ is a 3 × 3 convolution (spatial processing), and $(W_{3,l} * .)$ is a 1 × 1 convolution that restores dimensions. $BN(.)$ is batch normalization and $\sigma(.)$ is a ReLU. If the input and output channel-counts or spatial sizes differ (e.g., due to stride), a projection $S\,(.)$ uses a 1 × 1 convolution as in Eq. (13):

$$S\,(x) = W_{s,l} * x\,, \tag{13}$$

is applied. The difference between ResNet-50 and ResNet-101 is the number of bottleneck blocks in the deeper stages:

ResNet-50 uses {3, 4, 6, 3} bottleneck blocks across the four major stages, and ResNet-101 uses {3, 4, 23, 3} bottleneck blocks. Increasing the number of blocks increases the network's representational depth and capacity for hierarchical feature extraction:

$$x^{(L)} = R_L \circ R_{L-1} \circ \ldots \circ R_1\ (x^{(0)})\ , \tag{14}$$

Eq. (14), Means starting from the input $x^{(0)}$, apply block 1, then block 2, ..., up to block L, where each R_L $(\cdot)$ denotes a residual block as in Eq. (10). According to Eq. (15), MobileNetV3_Large encodes the input image x through a sequence of inverted residual blocks that use depthwise separable convolutions and a squeeze-and-excitation (SE) attention mechanism.

$$f_l = BN(Conv(\sigma(BN(Conv^{dw}_{k,k}\ (f_{l-1}))))), \tag{15}$$

In this formulation, depthwise convolution with kernel size k is denoted as $(Conv^{dw}_{k,k})$, k specifies the spatial support of the depthwise convolution, d represents the channel depth of the input tensor, and w denotes the width applied by the pointwise 1×1 convolutions within the inverted residual block. BN is a batch normalization, σ is a non-linear activation function, and f_{l-1} is the input feature map to the block. In some blocks, MobileNetV3_Large integrates SE attention to improve channel sensitivity. It helps the network learn which channels are more important. The SE module works like in Eq. (16) and Eq. (17) :

$$\mathrm{S} = \sigma(W_2.\sigma(W_1.GAP(f_l))), \tag{16}$$

$$f'_l = \ S\ \odot (f_l), \tag{17}$$

where GAP is the global average pooling operator, $W_1$and W_2 are learnable parameters, σ is the ReLU activation, and $\odot$ denotes channel-wise multiplication.

Finally, the encoded feature representation is expressed as in Eq. (18):

$$E_{MobileNetV3_{Large}}(x) = \ f_L\ , \tag{18}$$

where $E_{MobileNetV3_{Large}}$represents the entire encoder function, $f_L$is the feature map of the last block, which passes to the segmentation head. The key innovation in the DeepLabV3 model is the atrous (dilated) convolution, which expands the receptive field without increasing parameters. For a standard 2D convolution, we use Eq. (19) [14] to compute the output at spatial location (i,j) as :

$$\boldsymbol{y}(i,j) = \sum_m \sum_n \boldsymbol{w}(m,n) \cdot \boldsymbol{x}(i+m, j+n)\,, \tag{19}$$

where $\boldsymbol{w}$ denotes the convolution kernel and $\boldsymbol{x}$ is the input feature map. In contrast, atrous convolution introduces a dilation rate r, and the operation becomes as in Eq. (20) [14]:

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1158

$$\boldsymbol{y}(i,j) = \sum_{m}\sum_{n} \boldsymbol{w}(m,n) \cdot \boldsymbol{x}(i + rm, j + rn), \tag{20}$$

where $\boldsymbol{w}$ is the convolution kernel and r controls the spacing between the kernel elements. m and n represent the spatial indices of the convolution kernel along the height and width directions, respectively. According to Eq. (21), Eq. (22), Eq. (23) and Eq. (24), ASPP captures multi-scale contextual information by applying parallel atrous convolutions with different dilation rates [14]:

$$\boldsymbol{f}_{\text{ASPP}} = \boldsymbol{f}_0 \oplus \boldsymbol{f}_{r1} \oplus \boldsymbol{f}_{r2} \oplus \boldsymbol{f}_{r3} \oplus \boldsymbol{f}_{\text{GAP}}, \tag{21}$$

where

$$\boldsymbol{f}_0(p) = \sum_q w_0(q) \cdot \boldsymbol{f}_E(p), \tag{22}$$

is a 1×1 convolution,

$$\boldsymbol{f}_{ri} = \sum_q w_i(q) \cdot \boldsymbol{f}_E(p + r_i q), \tag{23}$$

for $i = \{1,2,3\}$ is the atrous convolution with rates $r_1 = 6$, $r_2 = 12$, $r_3 = 18$,

$$\boldsymbol{f}_{\text{GAP}} = \frac{1}{H'W'} \sum_p \boldsymbol{f}_E(p), \tag{24}$$

is the global pooling that provides image-level features, and $\boldsymbol{f}_E$ is the encoder output feature map. q indexes the spatial positions of the convolution kernel, and p denotes the current spatial location in the feature map at which the convolution is being evaluated. The concatenated features are then processed through an 1×1 convolution to reduce dimensionality. A Dice loss adapted for class imbalance is deployed in this work, is shown in Eq. (25) and Eq. (26) [15]:

$$\mathcal{L}_{\text{Dice}}(\hat{\boldsymbol{y}}, \boldsymbol{y}) = 1 - \frac{2\sum_p \hat{y}_p y_p + \varepsilon}{\sum_p \hat{y}_p^2 + \sum_p y_p^2 + \varepsilon}, \tag{25}$$

where $\hat{\boldsymbol{y}} = \sigma(\boldsymbol{M}_\theta(\boldsymbol{x}))$, (26)

is the predicted probability map after sigmoid activation σ, p indexes all pixels in the image, and $\varepsilon = 10^{-6}$ is a smoothing constant to prevent division by zero. This formulation uses squared terms in the denominator to better handle class imbalance compared to standard Dice loss. The parameters are optimized using the Adam optimizer with the update rule provided in Eq. (27) [16]:

$$\begin{cases} \boldsymbol{m}_t = \beta_1 \boldsymbol{m}_{t-1} + (1 - \beta_1)\boldsymbol{g}_t, \\ \boldsymbol{v}_t = \beta_2 \boldsymbol{v}_{t-1} + (1 - \beta_2)\boldsymbol{g}_t^2, \\ \theta_t = \theta_{t-1} - \dfrac{n\hat{\boldsymbol{m}}_t}{\sqrt{\hat{\boldsymbol{v}}_t + \varepsilon}}, \end{cases} \tag{27}$$

where $\boldsymbol{g}_t = \nabla_\theta \mathcal{L}_{\text{Dice}}$ is the gradient at iteration t, $\beta_1 = 0.9$ and $\beta_2 = 0.999$ are momentum parameters, $n = 10^{-2}$ is the internal learning rate, and $\hat{\boldsymbol{m}}_t$ and $\hat{\boldsymbol{v}}_t$ are bias-corrected moment estimates. The learning rate n is dynamically adjusted using a two-phase strategy and calculated based on Eq. (28) [15], [16]. The first phase (epochs $t \le T_{\text{switch}}$) has cosine annealing warm restarts

$$n_t = n_{\min} + \frac{n_{\max} - n_{\min}}{2} \cdot \left(1 + \cos\left(\pi \frac{t_{\text{cur}}}{T_0}\right)\right), \tag{28}$$

where t_{cur} is the current epoch with a restart cycle and $T_0 = 15$. The second phase (epochs $t > T_{\text{switch}}$) has plateau-based reduction $n_t = n_{t-1}\gamma$ if validation loss stagnates, where $\gamma = 0.1$ is the reduction factor. The switching epoch T_{switch} is determined by monitoring validation loss stagnation [15], [16]. At inference time, the final binary prediction is obtained by evaluating the decision rule specified in Eq. (29):

$$\hat{M}(x_{i,j}) = \begin{cases} 1 & \text{if } \sigma\left(M_\theta(x_{i,j})\right) \ge \tau; \\ 0 & \text{otherwise;} \end{cases} \tag{29}$$

where $\tau = 0.5$ is the classification threshold, $M_\theta(x)$ is the raw model output before activation for image x, σ is the sigmoid function, which converts the raw model output to probabilities in [0,1]. $x_{i,j}$ is the pixel at position (i, j) in the input image. $\hat{M}(x_{i,j})$ is the final binary prediction for that pixel (either 0 or 1).

C. The Dataset

1. Description of the dataset

The EBHI-Seg (Enteroscope Biopsy Histopathological H&E Image Dataset for Image Segmentation Tasks, available here), developed in 2022 by the Cancer Hospital of China Medical University in Shenyang, includes 2,228 histopathological images with corresponding ground-truth segmentation masks, covering six colorectal tumor stages [8]. The dataset is categorized into six histological classes: Normal, Polyp, Low-grade Intraepithelial Neoplasia (IN), High-grade IN, Adenocarcinoma, and Serrated Adenoma [8]. All images are stored in PNG format with a resolution of 224 × 224 pixels, and are uniformly categorized based on the histopathological characteristics described below:

1. Normal: Colorectal tissue sections exhibiting well-organized tubular structures with no evidence of pathological alterations, as observed under light microscopy [8].
2. Polyp: These images display redundant mucosal growths that maintain some structural resemblance to normal tissue but exhibit unique histopathological features [8].
3. Low-grade Intraepithelial Neoplasia (IN): A significant precancerous lesion characterized by increased glandular branching, dense cellular arrangements, and mild irregularities in luminal morphology. Architectural disruption and nuclear enlargement are moderate [8].
4. High-grade Intraepithelial Neoplasia (IN): A severe precancerous lesion exhibiting pronounced glandular distortion, marked nuclear

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i1.1158

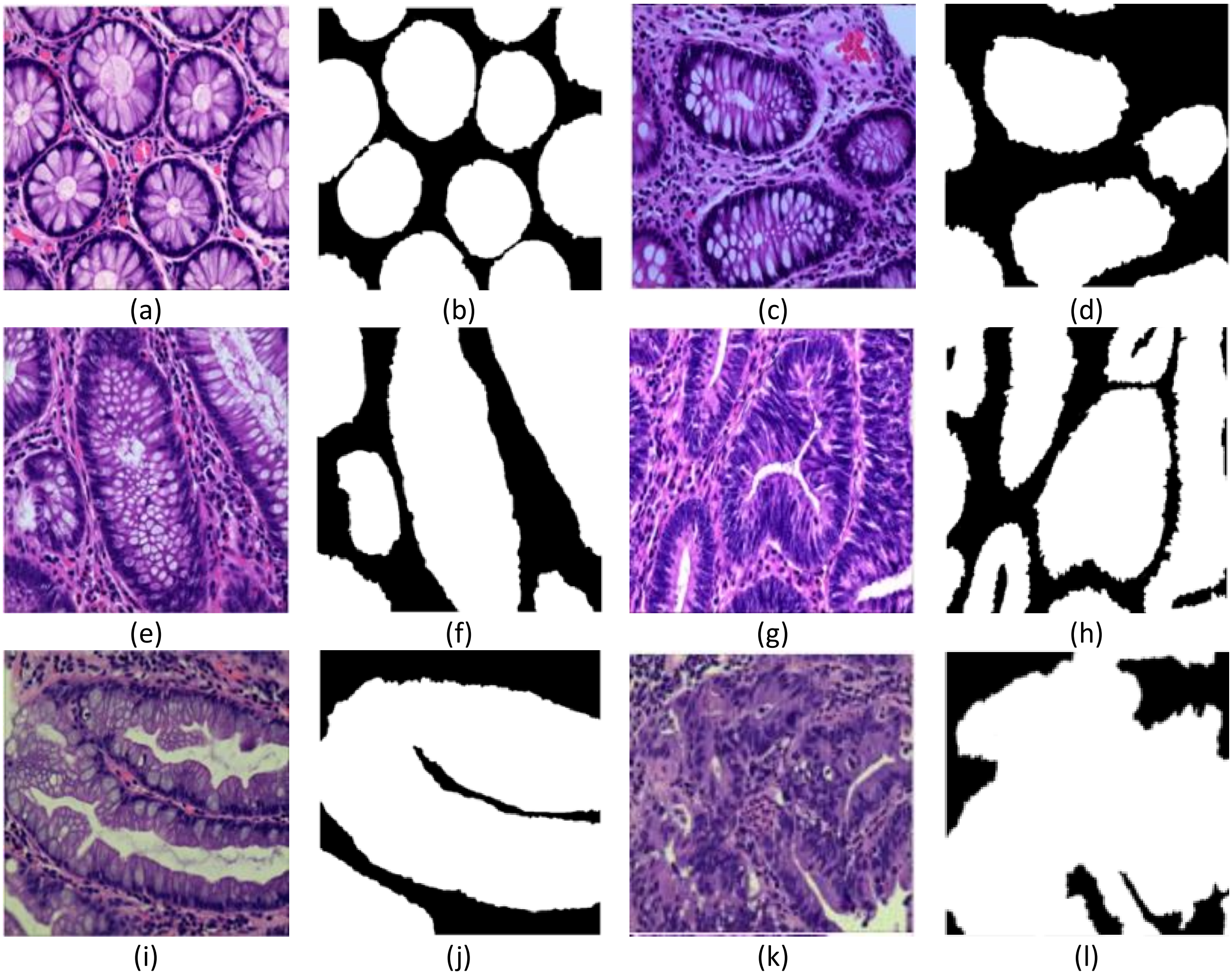


Fig. 2. **Samples from the EBHI-Seg dataset. (a) Normal class; (b) corresponding mask of (a); (c) Polyp class; (d) corresponding mask of (c); (e) Low-grade IN; (f) corresponding mask of (e); (g) High-grade IN; (h) corresponding mask of (g); (i) Serrated adenoma; (j) corresponding mask of (i); (k) Adenocarcinoma; (l) corresponding mask of (k)**

enlargement, and more extensive cellular atypia compared to low-grade IN [8].

5. Adenocarcinoma: A malignant neoplasm of the digestive tract, adenocarcinoma is typified by irregular glandular structures, poorly defined borders, and notably enlarged nuclei, complicating histopathological assessment [8].
6. Serrated Adenoma: An uncommon lesion representing roughly 1% of colonic polyps, serrated adenomas are defined by their distinctive serrated architectural patterns [8]

Representative sample images illustrate these classes in Fig. 2.

2. Data splitting

During the experiments, the dataset was divided into training, validation, and testing subsets with a 4:4:2 proportion. After splitting, the distribution of samples in each class is presented in Table 1, where **T. set** and **V. set** refer to the training and validating sets, respectively.

3. Data augmentation

The limited availability of real-world histopathology data poses a significant hurdle for training deep learning models for colorectal cancer segmentation.

This limitation can lead to overfitting, where the model memorizes specific training samples rather than learning robust features for precise classification of unseen images [17], [18]. To address this issue, data augmentation techniques are applied to artificially expand the training dataset and improve model performance [19], [20]. Data augmentation involves

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1158

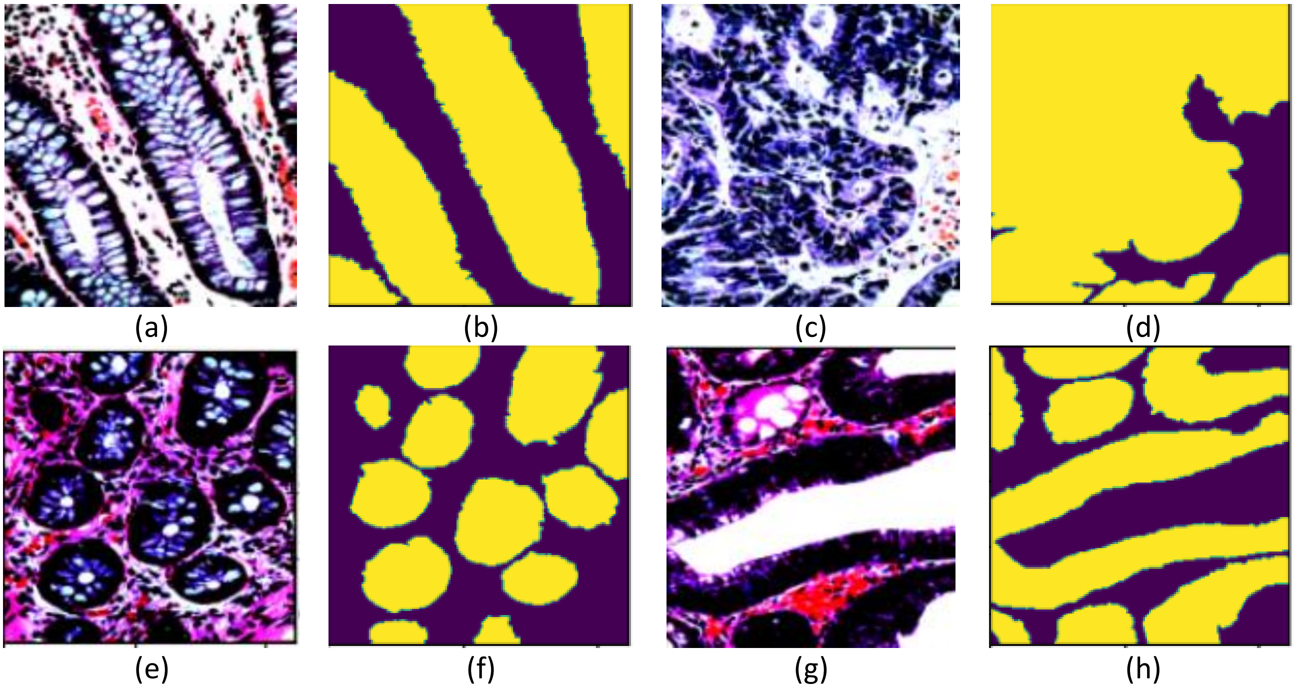


Fig. 3. **Augmented samples. (a) Augmented image of the serrated adenoma class; (b) corresponding mask of (a); (c) augmented image of the adenocarcinoma class; (d) corresponding mask of (c); (e) augmented image of the low-grade IN class; (f) corresponding mask of (e); (g) augmented image of the high-grade IN class; (h) corresponding mask of (g).**

generating new variations of existing images within the dataset, simulating the natural variability observed in real-world histopathology images [21], [22]. Data augmentation in medical imaging must be applied conservatively to preserve critical histological structures. In this study, we employ carefully selected augmentation techniques to enhance image variability while maintaining diagnostic integrity across 1,200 histopathological images. We simulate lighting variations by randomly adjusting brightness, contrast, and saturation levels. Digital images are represented as tensors of shape (height × width × color channels). Augmenting in the color channel space offers an efficient way to introduce realistic illumination variability.

Table 1. **Detailed information of the EBHI-Seg dataset**

Class Types	T. set	V. set	Test set	Total
Normal	30	30	16	76
Polyp	189	190	95	474
High-grade IN	74	74	38	186
Low-grade IN	254	255	130	639
Adenocarcinoma	318	318	159	795
Serrated Adenoma	23	23	12	58

Basic transformations involve isolating a single-color channel (R, G, or B) by retaining its matrix and setting the others to zero, or applying linear intensity operations to uniformly adjust brightness [23], [24], [25]. To simulate focus variations, we apply a Gaussian blur [21] with a kernel size randomly selected between 3 and 7 pixels, and a sigma value ranging from 0.1 to 3. Additionally, we randomly adjust image sharpness to reflect variations in tissue texture and scanner quality. These augmentations improve model robustness and generalization without compromising the morphological fidelity of the tissue. Representative augmented samples are shown in Fig. 3.

D. Research implementation procedure

The proposed research methodology for CRC segmentation using histopathology images comprises the phases shown in the flowchart in Fig. 4: Collect Dataset from Selected Sources. The first step involves obtaining histopathology images. Partnerships with hospitals or pathology laboratories is important to obtain anonymized image data from colorectal cancer cases and healthy controls. Alternatively, publicly available datasets may be used to provide diversity and meticulous annotation by experienced pathologists. In our case, the EBHI-Seg publicly available dataset was used. Separate the Dataset into Vital Sets. Three important subsets should be created from the dataset:

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1158

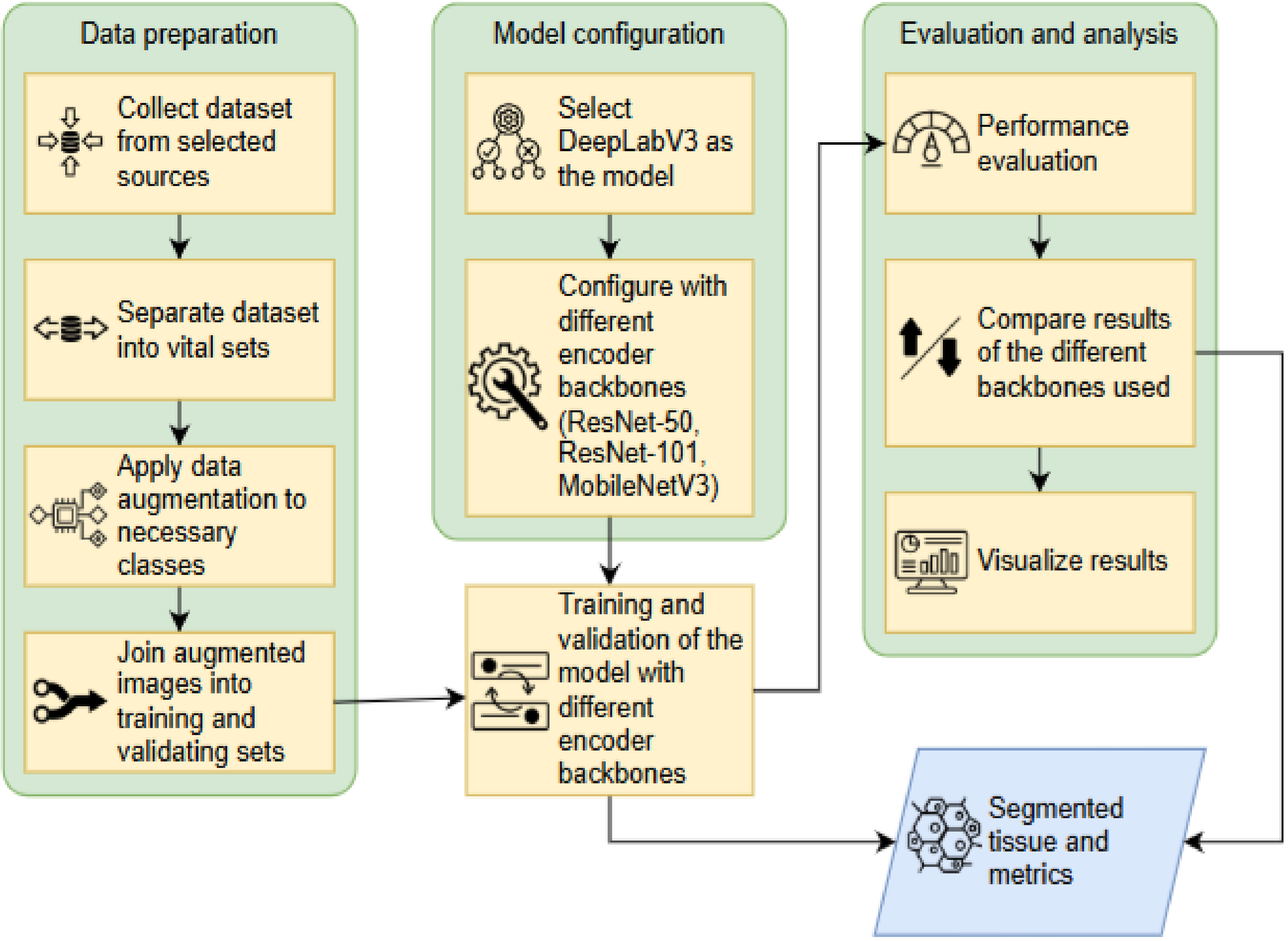

Data preparation
Collect dataset from selected sources
Separate dataset into vital sets
Apply data augmentation to necessary classes
Join augmented images into training and validating sets
Model configuration
Select DeepLabV3 as the model
Configure with different encoder backbones (ResNet-50, ResNet-101, MobileNetV3)
Training and validation of the model with different encoder backbones
Evaluation and analysis
Performance evaluation
Compare results of the different backbones used
Visualize results
Segmented tissue and metrics

Fig. 4. **The implementation procedure flowchart**

a) Training Set: This portion constitutes the majority of the dataset and is utilized to teach the model, enabling it to recognize and extract distinguishing patterns and features associated with colorectal cancer from the images.
b) Validation Set: The validation set is employed during training to tune model parameters and monitor performance. It helps mitigate overfitting and ensures the model generalizes well to unseen samples.
c) Testing Set: This subset is held out for the final assessment of the model, allowing evaluation of its ability to generalize to completely new data.

Apply Data Augmentation to Necessary Classes. To mitigate limited data availability and class imbalance, data augmentation procedures are selectively employed to underrepresented classes. This strategy increases sample diversity and supports more effective learning across all classes [21]. Join Augmented Images into Training and Validating Sets. The newly generated images are included in both the training and validation sets, which increases the dataset size and strengthens the model's generalization and resilience.

Select DeepLabV3 as the model. DeepLabV3 is selected as the baseline model due to its proven effectiveness in semantic segmentation tasks [26], [27]. Training and Validation of the Model with Different Encoder Backbones. Training is performed on the augmented dataset, and validation data are used to assess performance and control overfitting. Pre-trained ResNet-50 [12], ResNet-10 [12], and MobileNetv3_Large [13] models are used as initial backbone networks for the DeepLabV3 model. These models were originally trained on the COCO train2017 dataset, allowing the encoder to benefit from previously learned feature representations obtained from generic image recognition tasks. These features are then fine-tuned to adapt the model to the colorectal cancer histopathology segmentation task. The mathematical formulation of these machine learning models was given in Section II.A.1.

After training is completed, the model is assessed using a held-out test set that was not seen during training. Segmentation performance is evaluated using accuracy and the Jaccard index to measure the model's effectiveness in identifying colorectal cancer

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1158

Table 4. **Performance comparison of segmentation models on Adenocarcinoma class across various backbone architectures**

Model	Dice	Jaccard	Precision	Recall	Accuracy
DeepLabV3 + ResNet-50	0.9645	0.9328	0.9720	0.9683	0.9599
DeepLabV3 + ResNet-101	**0.9721**	**0.9465**	**0.9826**	0.9681	**0.9671**
DeepLabV3 + MobileNetV3_Large	0.9468	0.9023	0.9643	0.9441	0.9390
U-Net	0.887	0.808	0.850	0.950	--
Seg-Net	0.865	0.775	0.792	**0.977**	--
MedT	0.735	0.595	0.662	0.864	
Mask2Task	--	0.830	--	--	--
DDViT	0.901	--	--	--	--
U-Net + Attention U-net	0.9463	0.8906	0.9071	0.9112	0.8687
U-Net + ResNet-50	0.9417	0.8363	0.9207	0.9389	0.8981
U-Net + MobileNet-V2	0.8895	0.7758	0.7477	0.9733	0.7456
U-Net + EfficientNet-B0	0.9011	0.7827	0.8884	0.9377	0.8716
U-Net + DenseNet21	0.9442	0.8373	0.9071	0.9366	0.9251

regions. This evaluation is performed across different encoder backbones, including ResNet-50, ResNet-101, and MobileNetv3_Large. Finally, the results are visualized and analyzed by comparing overall performance in terms of accuracy, Jaccard index, precision, and recall.

III. Result

A. Evaluation Metrics

The objective evaluation of digital pathological image segmentation algorithms is essential for validating their robustness and ensuring their safe deployment in clinical diagnostic settings. A diverse set of statistical metrics is used to ensure a robust evaluation of model performance, including accuracy, precision, recall, Jaccard index, and Dice similarity coefficient.

1. Accuracy. It represents the proportion of pixels correctly classified by the model out of the total, serving as an indicator of overall performance. In contrast, the loss function captures the deviation between predictions and ground-truth labels, with lower values indicating better convergence and more reliable predictions [28]. Pixel-level predictions are categorized into four groups: true positives (TP), representing correctly identified cancerous pixels, true negatives (TN), corresponding to correctly classified non-cancerous pixels, false positives (FP), where non-cancerous pixels are incorrectly labeled as cancerous, and false negatives (FN), denoting cancerous pixels that are misclassified as non-cancerous [8]. These quantities are then used to compute the evaluation metrics defined in Eq. (30), Eq. (31), Eq. (32):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{30}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{31}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{32}$$

2. The Jaccard coefficient, commonly known as Intersection over Union (IoU), is a standard metric used to evaluate segmentation quality by measuring the overlap between predicted regions and ground truth annotations. As expressed in Eq. (33), it is computed as the ratio of the shared area to the total combined area of the two regions. Higher IoU values indicate greater consistency between the predicted segmentation and the reference mask. This metric is particularly suitable for imbalanced datasets, as it focuses on region overlap rather than being influenced by dominant background pixels [8].

$$\text{Jaccard} = \frac{P \cap G}{P \cup G} \tag{33}$$

3. Dice Coefficient, or Sørensen–Dice Index [4], quantifies the similarity between predicted segmentation masks and the reference annotations [8]. Calculated via Eq. (34) and measures the overlap between predicted and reference regions by comparing twice the intersection to the total number of pixels in both. A higher score denotes better segmentation performance. The Dice Similarity Coefficient (DSC) is a spatial overlap index that is generally more robust to class imbalance than accuracy-based metrics, as it evaluates the shared proportion of pixels rather than overall counts [8].

$$\text{Dice} = \frac{2 \times |P \cap G|}{|P| + |G|} \tag{34}$$

where P is the predicted mask, and G is the ground truth mask.

B. Experiments and Results

This research evaluated a binary semantic segmentation task using DeepLabV3 with three backbones: ResNet-50, ResNet-101, and MobileNetV3_Large. Each model underwent training for 40 epochs, employing a batch size of 4 for training, 8 for validation, and optimized using the Adam algorithm with

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1158

a learning rate of $1 \times e^{-4}$. The experimental outcomes for the entire dataset, with all classes combined, are presented inTable 2 and Table 3.

Table 2. Assessment metrics with respect to the validation set

Metrics	ResNet-50	ResNet-101	MobilNet-V3_Large
Epochs	40	40	40
Jaccard	**0.9035**	0.8985	0.8758
Dice	**0.9478**	0.9442	0.9321
Training Time	1h 34m 11s	1h 34m 11s	14m 09s

Table 3. Assessment metrics with respect to test set

Metrics	ResNet-50	ResNet-101	MobilNet-V3_Large
Epochs	40	40	40
Jaccard	0.9406	**0.9464**	0.9095
Dice	0.9693	**0.9722**	0.9525

Table 2 and Table 3 present the quantitative results of the DeepLabV3 model with three different backbone architectures, ResNet-50, ResNet-101, and MobileNetV3_Large, evaluated on both the validation and test sets. On the validation set (Table 2) ResNet-50 achieved the highest Dice coefficient of 0.9478 and a Jaccard index of 0.9035, slightly outperforming ResNet-101 (Dice = 0.9442, Jaccard = 0.8985). MobileNetV3_Large exhibited lower performance (Dice = 0.9321, Jaccard = 0.8758), but with significantly shorter training time (14 minutes vs. over 1 hour for the others). This demonstrates that MobileNetV3_Large offers a lightweight alternative when computational efficiency is prioritized, even though there is some sacrifice in segmentation accuracy. For the test set (Table 3) ResNet-101 delivered the best overall performance, achieving a Dice coefficient of 0.9722 and a Jaccard index of 0.9464. ResNet-50 followed closely (Dice = 0.9693, Jaccard = 0.9406), while MobileNetV3_Large scored lower but still acceptable (Dice = 0.9525, Jaccard = 0.9095). These results suggest that ResNet-101, owing to its deeper architecture, generalizes slightly better on unseen data.

C. Performance Comparison with Existing Works

This section presents an evaluation of DeepLabV3 using ResNet-50, ResNet-101, and MobileNetV3_Large backbones, alongside a comparison with existing studies conducted on the same dataset, including U-Net [8], SegNet [8], MedT [8], Mask2Tasks [10], DDViT [9], U-Net + Attention U-Net [4], U-Net + ResNet-50 [4], U-Net + MobileNet-V2 [4], U-Net + EfficientNet-B0 [4], U-Net + DenseNet21 [4]. Table 4 compares the proposed method with other deep learning approaches for CRC segmentation evaluated with the EBHI-Seg dataset. The symbol " -- " denotes unavailable results, and the highest values in each column are highlighted in bold. Additional per-class segmentation results are provided in the appendix.

The performance metrics in Table 4 demonstrate the effectiveness of the proposed approach in segmenting adenocarcinoma tissue using a ResNet-101 backbone. The method achieves the highest performance across nearly all metrics, with a Dice coefficient of 0.9721, a Jaccard index of 0.9465, a precision of 0.9826, and an accuracy of 0.9671, significantly outperforming the baseline and competing models. The elevated precision score (0.9826) indicates the model's strong ability to correctly identify true positive regions with minimal false positives. Additionally, the recall (0.9681) is competitively high, suggesting good sensitivity for detecting adenocarcinoma areas. Although Seg-Net achieves the highest recall (0.977), it does so at the cost of lower Dice and Jaccard scores. Among the U-Net variants, Attention U-Net and U-Net with DenseNet21 show strong results, with Dice scores of 0.9463 and 0.9442, respectively. However, both fall short of our model in terms of overlap-based metrics and precision. U-Net with a ResNet-50 encoder performs decently but demonstrates lower Jaccard and Dice indices than the proposed ResNet-101-based model, highlighting the benefits of deeper residual representations.

Transformer-based approaches yield mixed outcomes. MedT exhibits weak performance (Dice = 0.735, Jaccard = 0.595), likely due to its high data demands and lack of strong inductive biases. DDViT shows moderate capability (Dice = 0.901).

The Mask2Tasks model, which focuses on multi-task learning, reports only the Jaccard index (0.830). While this value is superior to that of some CNN baselines, it still lags behind our model by a significant margin. Furthermore, this table reinforces the influence of the encoder backbone. U-Net with MobileNet-V2, despite achieving high recall (0.9733), exhibits low precision (0.7477). U-Net with EfficientNet-B0 provides balanced results but still falls short of the proposed method’s overall performance. The model utilizing the ResNet-101 backbone exhibits superior segmentation performance for adenocarcinoma, benefiting from both deeper feature extraction and better generalization, outperforming both traditional CNN architectures and

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i1.1158

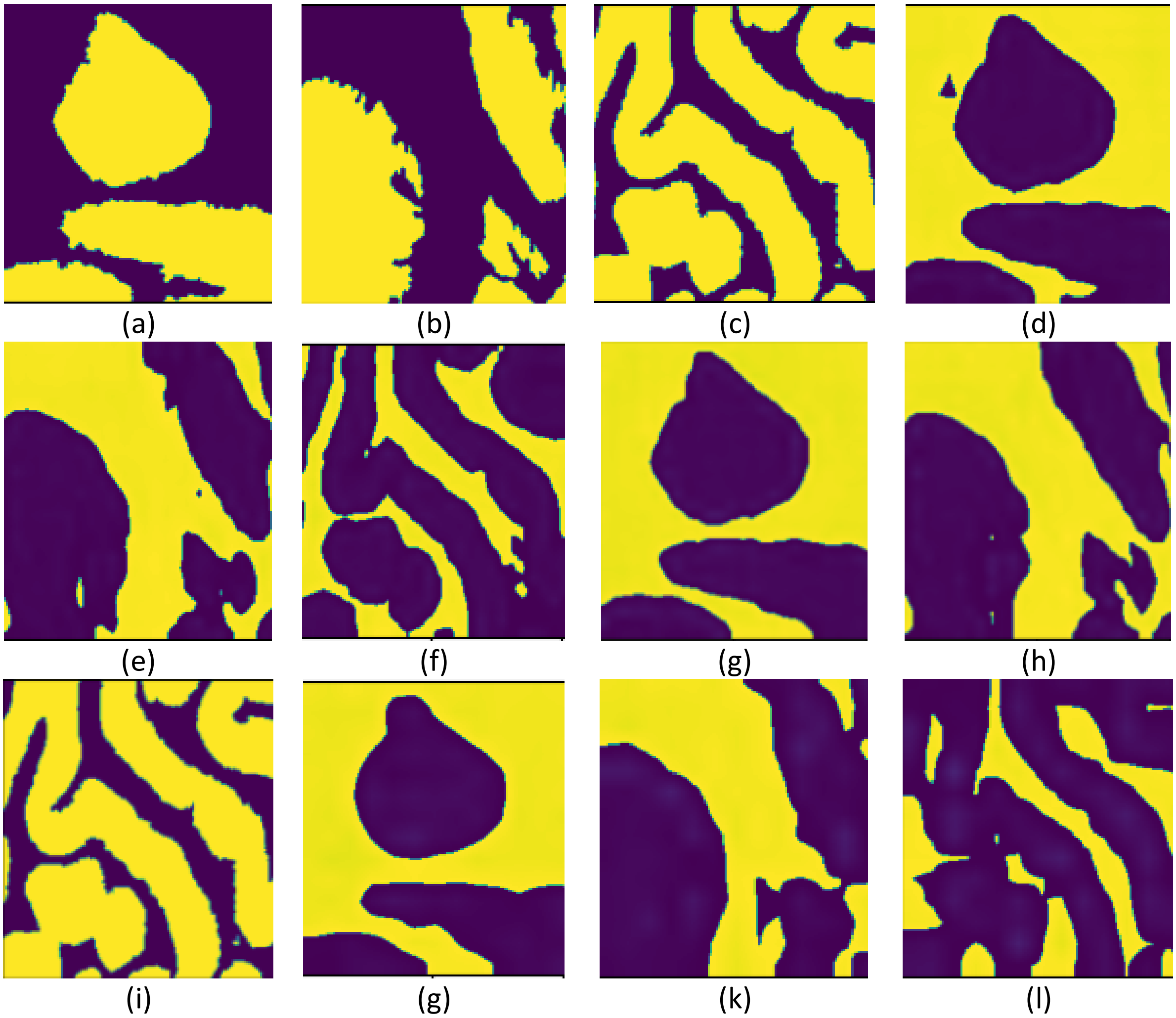


Fig. 5. Ground truth and segmentation results for some colorectal tissue classes using DeepLabV3 with different backbones. (a–c) ground truth masks for low-grade IN, high-grade IN, and serrated adenoma; (d–f) results with ResNet-50; (g–i) results with ResNet-101; (j–l) results with MobileNetV3-Large backbones.

recent transformer-driven approaches across nearly all metrics.

Our proposed approach demonstrates a strong balance of accuracy, robustness, and clinical relevance. It precisely delineates cancerous tissue boundaries and effectively distinguishes benign from malignant regions with high sensitivity. The strong agreement between predicted regions and those delineated by experts confirms its reliability, while minimal overfitting reflects stable training behavior and solid generalization across diverse colorectal tissue types. By outperforming leading approaches for all cancer categories, the model demonstrates both efficiency and adaptability. These strengths highlight its suitability as a reliable decision-support tool for improving diagnostic accuracy in colorectal cancer histopathology.

D. Segmentation Results Images

Qualitative results presented in Fig. 5 demonstrate that all DeepLabV3-based architectures effectively delineate cancerous and non-cancerous regions in colorectal histopathological images. The outputs produced by the three variants exhibit strong visual agreement with expert-labeled regions, confirming the capacity of the models to capture relevant tissue morphology. Among the evaluated models, the DeepLabV3-ResNet-101 configuration exhibited the most precise boundary localization and preserved fine glandular structures more effectively than the other versions. The DeepLabV3-ResNet-50 model achieved comparable results. However, it occasionally produced smoother boundaries, while the MobileNetV3_Large backbone, despite its faster inference, showed slight degradation in boundary precision, particularly in areas with complex glandular morphology. These visual outcomes align with the quantitative results reported in Tables 2 and 3. The strong performance of

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i1.1158

DeepLabV3-ResNet-101 stems from its deep hierarchical architecture and the Atrous Spatial Pyramid Pooling (ASPP) module, which together enable multi-level contextual understanding and precise discrimination of epithelial, stromal, and glandular regions, effectively capturing the heterogeneous textures and glandular variability characteristic of colorectal tissue. In contrast, the lightweight MobileNetV3_Large backbone demonstrates that efficient models can still achieve acceptable segmentation quality while substantially reducing computational cost, rendering it suitable for immediate processing scenarios or resource-limited applications.

IV. Discussion

This study investigated how different backbone networks within the DeepLabV3 framework affect segmentation performance in colorectal histopathological images. The models were tested using ResNet-50, ResNet-101, and MobileNetV3_Large encoders to analyze differences in accuracy and boundary delineation for various tissue classes. With DeepLabV3-ResNet-101 as the baseline, all alternative models showed varying degrees of performance decline**.** Accordingly, Tables 5 and 6 present the percentage changes of the other backbones relative to the DeepLabV3-ResNet-101 baseline, providing a quantitative comparison of their evaluation on the validation and test sets.

Table 5. Percentage differences in validation set metrics compared to the DeepLabV3-ResNet-101 baseline

Metrics (%)	ResNet-50	MobilNetV3_Large
Jaccard	+0.56%	–2.53%
Dice	+0.38%	–1.28%

On the validation set, ResNet-50 shows slight improvements in some metrics, with a Jaccard gain of +0.56% and a Dice gain of +0.38%, suggesting it generalizes reasonably well. In contrast, MobileNetV3_Large exhibits decreases (Jaccard –2.53%, Dice –1.28%), reflecting its limited capacity to capture complex features. Test-set results further highlight differences in generalization. ResNet-50 shows minor declines (Jaccard –0.61%, Dice –0.30%), indicating slightly reduced performance on completely unseen images compared to the baseline. MobileNetV3_Large experiences more substantial drops (Jaccard –3.90%, Dice –2.03%), confirming that while it is computationally efficient, it struggles to maintain segmentation accuracy on more challenging data. Overall, the results emphasize that DeepLabV3-ResNet-101 remains the most robust backbone**,** achieving consistently high performance across both validation and test sets. ResNet-50 offers a reasonable trade-off with slight reductions in some metrics, whereas MobileNetV3_Large demonstrates that efficiency gains may come at the expense of accuracy in histopathological segmentation.

Table 6. Percentage differences in test set metrics compared to the DeepLabV3-ResNet-101 baseline

Metrics (%)	ResNet-50	MobilNetV3_Large
Jaccard	-0.61%	–3.90%
Dice	-0.30%	–2.03%

A. Integration of Deep Learning-Based Segmentation into Clinical Workflow for Colorectal Cancer Diagnosis

The integration of DL into digital pathology workflows holds transformative potential for enhancing colorectal cancer diagnosis and segmentation. DL-based models, notably those employing convolutional neural networks (CNNs) [29], [30], [31] and transformer architectures [32], [33], enable automated, accurate segmentation of histopathological whole-slide images (WSIs), as demonstrated in the Experiments and Results section. However, to achieve real-world clinical impact, these models must be embedded within comprehensive diagnostic pipelines that support and augment the expertise of pathologists. Incorporating DL systems into the clinical workflow yields several critical benefits. First, they streamline efficiency by significantly shortening slide interpretation time, thus accelerating the diagnostic workflow [34]. Second, they enhance accuracy by detecting subtle morphological features that may be overlooked by human observers, contributing to more precise and reproducible assessments [35]. Third, DL integration facilitates scalability, enabling high-throughput analysis of large datasets, which is particularly advantageous in large-scale clinical studies and screening programs [36]. Finally, these systems increase consistency by minimizing inter- and intra-observer variability, supporting standardized diagnostic outcomes [37]. Beyond colorectal cancer, the proposed pipeline demonstrates strong generalizability. Although initially validated using the EBHI-Seg dataset, its modular and adaptable architecture supports retraining or fine-tuning with other histopathological datasets. This flexibility allows for application across a broad range of cancer types and tissue structures, enhancing the utility of DL-based tools in diverse clinical contexts.

B. Proposed Workflow Integration

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1158

Inspired by [34], [38] [39] and [40], we propose a workflow, illustrated in Fig. 6, that incorporates DL-based segmentation models at critical stages of the histopathological diagnostic process:

1. Slide Acquisition and Preprocessing. Histopathological slides are digitized using high-resolution scanners. Basic preprocessing steps such as stain normalization, artifact removal, and tiling into manageable patches are performed to prepare the data for automated analysis [41].
2. Deep Learning-Based Segmentation. A DL segmentation model (e.g., DeepLabV3 based on ResNet-101) is applied to delineate cancerous tissue from surrounding normal or dysplastic regions. This identifies regions of interest (ROIs), such as glandular structures, tumor boundaries, and stromal invasion zones.
3. Pathologist-Guided Review with CAD Assistance. The segmented regions are overlaid on the original WSI and displayed within a computer-aided diagnosis (CAD) interface. Pathologists can interact with predictions, verify or modify masks, and make informed decisions based on visual and quantitative cues.
4. Quantitative Reporting and Decision Support. The segmentation output is further processed to derive metrics such as tumor burden, gland density, or invasion depth. These outputs aid in staging, prognosis, and treatment planning.
5. Post-Diagnosis Archiving and Model Feedback. Verified annotations and reports are archived. Pathologists' corrections can be used for continual learning to improve model performance.

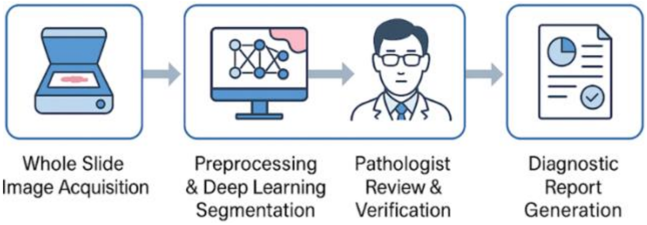

Whole Slide Image Acquisition
Preprocessing & Deep Learning Segmentation
Pathologist Review & Verification
Diagnostic Report Generation

Fig. 6. **Integration of the proposed deep learning–based segmentation model into the colorectal cancer diagnostic workflow.**

While the proposed approach demonstrates strong segmentation performance and clear clinical potential, it is important to acknowledge its current limitations and highlight avenues for future research. Although the EBHI-Seg dataset provides a valuable benchmark for colorectal cancer segmentation, its relatively limited size and single-institution origin may restrict the generalizability of our findings. Variations in staining protocols, scanner types, and patient demographics across different clinical centers can influence model performance. Expanding training and validation to include multi-center datasets will be essential to ensure broader applicability.

The best-performing configuration in this study, DeepLabV3 with a ResNet-101 backbone, requires substantial computational resources, which may pose challenges for deployment in resource-limited clinical environments. While we explored more efficient architectures such as MobileNetV3, further optimization and model compression techniques are needed to balance accuracy with practicality.

Our current work focuses primarily on colorectal cancer segmentation. The model's performance on other cancer types or more challenging pre-cancerous lesions remains untested. Extending the evaluation to a broader range of pathological conditions would yield a better understanding of the reliability and versatility of the proposed pipeline.

To overcome these limitations, future research will concentrate on:

1. Conducting extensive, multi-institutional validation studies to assess the model's robustness across different patient populations and clinical settings.
2. Integrating the model into forward-looking clinical studies to evaluate its impact in actual clinical settings on diagnostic accuracy, pathologist workflow, and reporting times.
3. exploring lightweight and efficient network architectures and compression strategies to support deployment in varied healthcare environments.

V. Conclusion

This study aimed to evaluate the effectiveness of deep learning-based semantic segmentation models for computer-aided diagnosis of colorectal cancer (CRC) in histopathological images, with particular focus on evaluating backbone architectures and their impact on performance. Our results show that the DeepLabV3 architecture, particularly when paired with ResNet-50 or ResNet-101 backbones, achieves a segmentation accuracy of 0.97, effectively delineating glandular structures and reliably distinguishing benign from malignant tissue regions. The model demonstrated consistent performance across evaluation metrics, confirming its robustness, stability, and strong generalization capability. Rigorous dataset preprocessing and targeted data augmentation further enhanced segmentation accuracy and model convergence. Moreover, the multi-class evaluation across key colorectal tissue categories provided a more clinically relevant and fine-grained analysis than previous studies, highlighting the model's potential utility in supporting diagnostic processes. Upcoming studies will concentrate on large-scale validation across multi-center datasets and exploration

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i1.1158

of advanced architectures to further optimize performance. Additionally, prospective validation through real-world clinical trials will be undertaken to strengthen the practical utility and clinical relevance of the proposed method. Ultimately, the integration of deep learning into CRC diagnostic workflows could support more consistent, efficient, and timely clinical decision-making in pathology practice.

Acknowledgment

The authors would like to acknowledge that this work was carried out independently. We have no specific individuals or institutions to thank for direct contributions to the completion of this research. Nonetheless, we remain grateful for the general academic environment and resources that have supported our learning and development.

Funding

No specific financial support was provided for this research by public, commercial, or non-profit organizations.

Data Availability

The EBHI-Seg dataset can be freely accessed at: https://figshare.com/articles/dataset/EBHI-SEG/21540159/1

Author Contribution

Fahima IDIRI designed the study, led the conceptualization, implemented the methodology, conducted data collection, and wrote the manuscript. Farid MEZIANE supervised, validated the study and assisted in drafting the manuscript and revisions. Hakim BOUCHAL contributed to the conceptualization and implementation of the methodology.

All authors reviewed and approved the final manuscript and gave their consent for publication.

Declarations

Ethical Approval

This study did not require ethical approval, as it utilized publicly available datasets containing no personally identifiable information. All data used were obtained in accordance with the terms and conditions set by the original data providers.

Consent for Publication Participants.

All participants provided consent for publication.

Competing Interests

The authors declare that they have no competing interests.

References

[1] J.-M. Bokhorst *et al.*, “Deep learning for multi-class semantic segmentation enables colorectal cancer detection and classification in digital pathology images,” *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-35491-z.

[2] A. S. N. Raju *et al.*, “Colorectal cancer detection with enhanced precision using a hybrid supervised and unsupervised learning approach,” *Sci. Rep.*, vol. 15, no. 1, p. 3180, 2025, doi: 10.1038/s41598-025-86590-y.

[3] B. Á. Pataki *et al.*, “HunCRC: annotated pathological slides to enhance deep learning applications in colorectal cancer screening,” *Sci. Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1038/s41597-022-01450-y.

[4] S. Sengupta *et al.*, “Assessment of different U-Net backbones in segmenting colorectal adenocarcinoma from H&E histopathology,” *Pathol. Res. Pract.*, vol. 266, p. 155820, Feb. 2025, doi: 10.1016/j.prp.2025.155820.

[5] R. Feng, X. Liu, J. Chen, D. Z. Chen, H. Gao, and J. Wu, “A deep learning approach for colonoscopy pathology WSI analysis: accurate segmentation and classification,” *IEEE J. Biomed. Health Inform.*, vol. 25, no. 10, pp. 3700–3708, Oct. 2021, doi: 10.1109/JBHI.2020.3040269.

[6] S. R. Alotaibi, M. A. Alohali, M. Maashi, H. Alqahtani, M. Alotaibi, and A. Mahmud, “Advances in colorectal cancer diagnosis using optimal deep feature fusion approach on biomedical images,” *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-024-83466-5.

[7] S. Qiu, H. Lu, J. Shu, T. Liang, and T. Zhou, “Colorectal cancer segmentation algorithm based on deep features from enhanced CT images,” *Comput. Mater. Contin.*, vol. 80, no. 2, pp. 2495–2510, Aug. 2024, doi: 10.32604/CMC.2024.052476.

[8] J. Cheng *et al.*, “EBHI-Seg: a novel enteroscope biopsy histopathological hematoxylin and eosin image dataset for image segmentation tasks,” *Front. Med.*, vol. 10, Jan. 2023, doi: 10.3389/fmed.2023.1114673.

[9] L. Sun and V. S. Sheng, “DDViT: double-level fusion domain adapter vision transformer,” in *Proc. AAAI Conf. Artif. Intell.*, Mar. 2024, pp. 23661–23663. doi: 10.1609/aaai.v38i21.30516.

[10] H. Le Xuan, M. H. Le, V. V. Truong, H. P. Quang, and H. V. Huy, “MASK2TASKS: leveraging segmentation to enhance classification performance in histopathological colorectal images,” in *Proc. 2nd Tiny Papers Track at ICLR*, 2024. [Online]. Available: https://openreview.net/forum?id=WMxXuVTzFm

[11] T. Liu, H. Yang, J. Huang, and Y. Zhou, “A review of colorectal cancer histopathology image segmentation using deep learning Methods,” in *Proc. 2024 6th Int. Conf. Control

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1158

Robot. (ICCR), 2024, pp. 281–291. doi: 10.1109/ICCR64365.2024.10927570.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[13] A. Howard *et al.*, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324. doi: 10.1109/ICCV.2019.00140.

[14] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," in *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Apr. 2018, pp. 834–848. doi: 10.1109/TPAMI.2017.2699184.

[15] M. Yeung, L. Rundo, Y. Nan, E. Sala, C. B. Schönlieb, and G. Yang, "Calibrating the dice loss to handle neural network overconfidence for biomedical image segmentation," *J. Digit. Imaging*, vol. 36, no. 2, p. 739, Apr. 2022, doi: 10.1007/S10278-022-00735-3.

[16] S. A. Aula and T. A. Rashid, "Foxtsage vs. Adam: revolution or evolution in optimization?," *Cogn. Syst. Res.*, vol. 92, p. 101373, Sep. 2025, doi: 10.1016/J.COGSYS.2025.101373.

[17] M. Xiao *et al.*, "Addressing overfitting problem in deep learning-based solutions for next generation data-driven networks," *Wirel. Commun. Mob. Comput.*, Aug. 2021, doi: 10.1155/2021/8493795.

[18] K. Lee *et al.*, "Deep learning of histopathology images at the single cell level," *Front. Artif. Intell.*, vol. 4, Sep. 2021, doi: 10.3389/frai.2021.754641.

[19] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, Jul. 2019, doi: 10.1186/s40537-019-0197-0.

[20] M. Abbas, M. Arslan, R. Abid Bhatty, F. Yousaf, A. Ahmad Khan, and A. Rafay, "Enhanced skin disease diagnosis through convolutional neural networks and data augmentation techniques," *J. Comput. Biomed. Inf.*, vol. 7, pp. 87–106, Jun. 2024, doi: 10.56979.

[21] A. Mumuni and F. Mumuni, "Data augmentation: a comprehensive survey of moderne approaches," *Array*, vol. 16, p. 100258, Dec. 2022, doi: 10.1016/j.array.2022.100258.

[22] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A comprehensive survey of image augmentation techniques for deep learning," *Pattern Recognit.*, vol. 137, p. 109347, May 2023, doi: 10.1016/j.patcog.2023.109347.

[23] T. Kumar, R. Brennan, A. Mileo, and M. Bendechache, "Image data augmentation approaches: a comprehensive survey and future directions," *IEEE Access*, vol. 12, pp. 187536–187571, 2024, doi: 10.1109/ACCESS.2024.3470122.

[24] E. Goceri, "Medical image data augmentation: techniques, comparisons and interpretations," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 12561–12605, Mar. 2023, doi: 10.1007/s10462-023-10453-z.

[25] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," *J. Med. Imaging Radiat. Oncol.*, vol. 65, no. 5, pp. 545–563, 2021, doi: 10.1111/1754-9485.13261.

[26] R. J and P. A, "Analysis of lung cells with a novel segmentation methodology using FCN and Deeplab V3," in *Proc. 2024 3rd Int. Conf. Distributed Comput. Electr. Circuits Electron. (ICDCECE)*, 2024, pp. 1–8. doi: 10.1109/ICDCECE60827.2024.10548209.

[27] S. Pak *et al.*, "Application of deep learning for semantic segmentation in robotic prostatectomy: comparison of convolutional neural networks and visual transformers," *Investig. Clin. Urol.*, vol. 65, no. 6, pp. 551–558, Nov. 2024, doi: 10.4111/icu.20240159.

[28] R. Archana and P. S. E. Jeevaraj, "Deep learning models for digital image processing: a review," *Artif Intell Rev*, vol. 57, no. 1, p. 11, 2024, doi: 10.1007/s10462-023-10631-z.

[29] T. Jo, "Convolutional neural networks," in *Deep Learning Foundations*, Springer Int. Publ., 2023, pp. 303–326. doi: 10.1007/978-3-031-32879-4_12.

[30] S. S. Kshatri and D. Singh, "Convolutional neural network in medical image analysis: a review," *Arch. Comput. Methods Eng.*, vol. 30, no. 4, pp. 2793–2810, 2023, doi: 10.1007/s11831-023-09898-w.

[31] D. R. Sarvamangala and R. V Kulkarni, "Convolutional neural networks in medical image understanding: a survey," *Evol. Intell.*, vol. 15, no. 1, pp. 1–22, 2022, doi: 10.1007/s12065-020-00540-3.

[32] G. Hille, P. Tummala, L. Spitz, and S. Saalfeld, "Transformers for colorectal cancer segmentation in CT imaging," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 19, pp. 2079--2087, 2024, doi: 10.1007/s11548-024-03217-9.

[33] H. Xiao, L. Li, Q. Liu, X. Zhu, and Q. Zhang, "Transformers in medical image segmentation: a review," *Biomed. Signal Process. Control*, vol. 84, p. 104791, 2023, doi: 10.1016/j.bspc.2023.104791.

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeemi.v8i1.1158

[34] Z. Yin, C. Yao, L. Zhang, and S. Qi, "Application of artificial intelligence in diagnosis and treatment of colorectal cancer: a novel Prospect," *Front. Med. (Lausanne)*, vol. 10, p. 1128084, 2023, doi: 10.3389/fmed.2023.1128084.

[35] P. C. Neto *et al.*, "An interpretable machine learning system for colorectal cancer diagnosis from pathology slides," *NPJ Precis. Oncol.*, vol. 8, no. 1, Dec. 2024, doi: 10.1038/s41698-024-00539-4.

[36] M. Cooper, Z. Ji, and R. G. Krishnan, "Machine learning in computational histopathology: challenges and opportunities," Sep. 01, 2023, *John Wiley and Sons Inc*. doi: 10.1002/gcc.23177.

[37] A. Santhoshi and A. Muthukumaravel, "Texture and shape-based feature extraction for colorectal tumor segmentation," in *Proc. 2024 10th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, IEEE, 2024, pp. 315–320. doi: 10.1109/ICACCS60874.2024.10717222.

[38] C. Ho *et al.*, "A promising deep learning-assistive algorithm for histopathological screening of colorectal cancer," *Sci. Rep.*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-06264-x.

[39] H. M. Selby, Y. A. Son, V. R. Sheth, T. H. Wagner, E. L. Pollom, and A. M. Morris, "AI-ready rectal cancer MR imaging: a workflow for tumor detection and segmentation," *BMC Med. Imaging*, vol. 25, no. 1, Dec. 2025, doi: 10.1186/s12880-025-01614-3.

[40] M. F. Avram, D. C. Lazăr, M. I. Mariş, and S. Olariu, "Artificial intelligence in improving the outcome of surgical treatment in colorectal cancer," *Front. Oncol.*, vol. 13, Jan. 2023, doi: 10.3389/fonc.2023.1116761.

[41] K. S. Wang *et al.*, "Accurate diagnosis of colorectal cancer based on histopathology images using artificial intelligence," *BMC Med.*, vol. 19, no. 1, Dec. 2021, doi: 10.1186/s12916-021-01942-5.

Author Biography

Fahima IDIRI is currently a doctoral researcher affiliated with the Laboratory of Medical Informatics and Intelligent and Dynamic Environments (LIMED) at the University of Bejaia, Algeria. She obtained her Bachelor's degree in Computer Science in 2014, followed by a Master's degree in Networks and Distributed Systems in 2016, both from the same institution. Her doctoral research focuses on medical image segmentation using deep learning, with applications in computational pathology and clinical decision support. Her work further explores whole-slide image analysis, optimization of image segmentation architectures, and the development of robust evaluation pipelines designed to enhance diagnostic accuracy and support real clinical workflows.

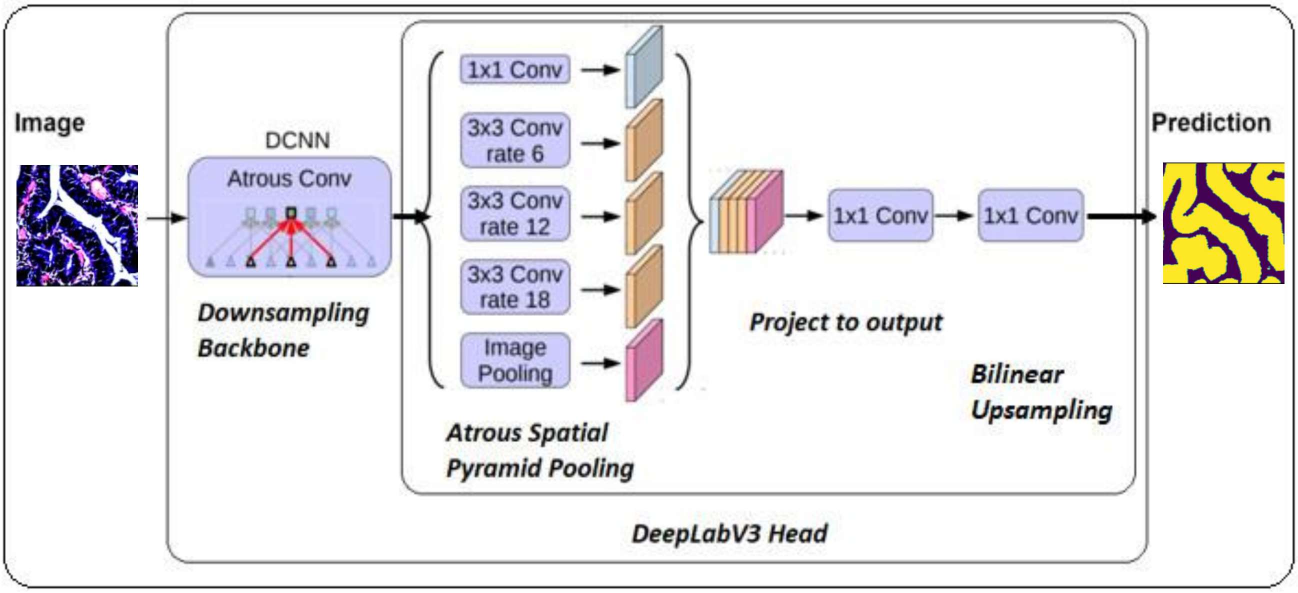
Farid MEZIANE is a professor of Data Science, Head of the Data Science Research Centre, the University's lead for the Data Science academic research theme at the University of Derby, UK. He obtained a PhD in Computer Science from the University of Salford, UK, for his work on producing formal specifications from Natural Language requirements. The work was considered pioneering at the time and paved the way for significant interest in automating the production of software specifications from informal requirements.

He has authored over 200 scientific papers and participated in many national and international research projects. He is the co-chair of the International Conference on Application of Natural Language to Information Systems; co-chair of the International Conference on Information Science and Systems. He is serving on the programme committee of over ten international conferences. He is an associate editor for the Data and Knowledge Engineering (Elsevier) journal and the managing editor of the International Journal of Information Technology and Web Engineering. He was awarded the Highly Commended Award by the Literati Club in 2001 for his paper on Intelligent Systems in Manufacturing: Current Development and Future Prospects. His research expertise includes Natural Language processing, semantic computing, data mining, big data, and knowledge Engineering.

Webpage: https://www.derby.ac.uk/staff/farid-meziane/

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1158

Hakim BOUCHAL is currently a doctoral researcher affiliated with the Laboratory of Medical Informatics and Intelligent and Dynamic Environments (LIMED) at the University of Bejaia, Algeria. He obtained his Bachelor's degree in Electronics in 2014, followed by a Master's degree in Telecommunications in 2016, both from the same institution. His PhD work lies at the intersection of pattern recognition and natural language processing, with a dedicated emphasis on the automatic recognition of historical Arabic manuscripts. His research interests include deep learning–based handwriting recognition, document image analysis, dataset construction and annotation, and segmentation methodologies for complex handwritten archival documents. His ongoing work contributes to digital heritage preservation.



Colorectal Cancer Segmentation from Histopathological Images Using a Hybrid Deep Learning Pipeline

Manuscript received 5 August 2025; Revised 12 November 2025; Accepted 5 January 2026; Available online 14 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1158