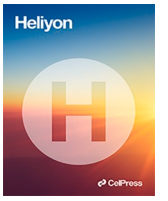
Heliyon 10 (2024) e33655

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Efficient colorectal polyp segmentation using wavelet transformation and AdaptUNet: A hybrid U-Net

Devika Rajasekar [a], Girish Theja [a], Manas Ranjan Prusty [b,*], Suchismita Chinara [c]

[a] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India
[b] Centre for Cyber Physical Systems, Vellore Institute of Technology, Chennai, India
[c] Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India

ARTICLE INFO

Keywords:
Deep learning
U-Net
Colorectal polyps
Polyp segmentation
Wavelet transformation
Data augmentation

ABSTRACT

The prevalence of colorectal cancer, primarily emerging from polyps, underscores the importance of their early detection in colonoscopy images. Due to the inherent complexity and variability of polyp appearances, the task stands difficult despite recent advances in medical technology. To tackle these challenges, a deep learning model featuring a customized U-Net architecture, AdaptUNet is proposed. Attention mechanisms and skip connections facilitate the effective combination of low-level details and high-level contextual information for accurate polyp segmentation. Further, wavelet transformations are used to extract useful features overlooked in conventional image processing. The model achieves benchmark results with a Dice coefficient of 0.9104, an Intersection over Union (IoU) coefficient of 0.8368, and a Balanced Accuracy of 0.9880 on the CVC-300 dataset. Additionally, it shows exceptional performance on other datasets, including Kvasir-SEG and Etis-LaribDB. Training was performed using the Hyper Kvasir segmented images dataset, further evidencing the model's ability to handle diverse data inputs. The proposed method offers a comprehensive and efficient implementation for polyp detection without compromising performance, thus promising an improved precision and reduction in manual labour for colorectal polyp detection.

1. Introduction

A new era of medical diagnostics and patient care has begun with the advent of technology in healthcare. Recent advancements in machine learning and computer vision have yielded promising results for automated disease detection, a previously unexplored field. This newly discovered territory is especially important for the early detection of conditions such as colorectal polyps, abnormal tissue growths in the bowel that may progress to colorectal cancer if left undiagnosed and untreated.

The morphologies of colorectal polyps range from flat to pedunculated, and their colours and surface patterns are distinct. Their subtle appearances frequently result in oversights during manual examination, highlighting the need for efficient automatic detection tools. One of the pivotal advancements within the medical diagnostics domain is the application of deep learning techniques, such as the U-Net architecture, for automated disease detection. Polyp segmentation is the process of identifying and delineating these polyps in medical images, which is a challenging task due to their varied appearances. Specifically, the term "U-Net" refers to a convolutional neural network (CNN) architecture that is particularly well-suited for biomedical image segmentation due to its ability to capture both

* Corresponding author.
E-mail address: manas.iter144@gmail.com (M.R. Prusty).

https://doi.org/10.1016/j.heliyon.2024.e33655
Received 22 February 2024; Received in revised form 6 March 2024; Accepted 25 June 2024
Available online 26 June 2024

local and global information in images [1,2].

In this research study, the authors have introduced a novel deep learning approach utilizing a customized U-Net framework enhanced with attention mechanisms and skip connections. Wavelet transformation is a mathematical tool used to decompose an image into its frequency components, allowing for the extraction of essential features that might be overlooked in standard image processing. In the context of this study, wavelet transformations are employed to enhance the feature representation of colonoscopy images. This combination improves the accuracy of polyp segmentation by preserving spatial details and facilitating information flow throughout the network. This inventive approach, which combines deep learning with wavelet transformations and data augmentation, has the potential to advance polyp detection, resulting in earlier interventions and better patient outcomes [3,4].

The research content can be summarized as follows: Section 2 provides an insight into the motivations and unique contributions of the study. Section 3 delves into existing methods and relevant studies in the field. The datasets used for model training and testing are elucidated in Section 4. Methodological details, including data pre-processing and the deployment of the AdaptUNet model, are covered in Section 5. Section 6 is dedicated to the presentation and in-depth discussion of the results across various datasets. The paper concludes with Section 7, where a summary and implications of the findings are articulated.

2. Motivation and objectives

Deep learning methodologies have emerged as potent tools, exhibiting profound potential in the realm of image-guided medical diagnostics. These methods have found significant applicability in the field of colorectal cancer detection, where they examine colonoscopy images with precision. By teaching these complex artificial intelligence-driven models to distinguish and segment polyps from the milieu of background tissues, they improve detection accuracy while simultaneously reducing the amount of manual labour required for polyp identification. As a result, these developments may pave the way for decreased screening costs, reduced patient distress, and increased patient compliance with examinations.

This research is motivated by the critical need to identify and eliminate colorectal polyps at an early stage, aiming to prevent the development of cancer in the colon. Despite its efficacy, the current gold standard for screening, colonoscopy, is hampered by its time-consuming nature, invasiveness, and susceptibility to human error, resulting in occasional missed polyps and increased cancer risks. The incorporation of deep learning techniques for automated polyp detection and segmentation appears to be a promising countermeasure against these obstacles.

3. Related works

Over the recent years, the rapid advancements in machine learning techniques have displayed immense potential in transforming healthcare outcomes, particularly in the field of colorectal cancer detection and diagnosis. Researchers have explored various techniques and methodologies to develop increasingly precise and efficient models, including convolutional neural networks (CNNs), ensemble classifiers, hybrid models, and the integration of deep learning with classical machine learning approaches.

Tharwat et al. outlined the overall potential of the recent advancements in the application of machine learning techniques for the early detection of colon cancer [5]. Specifically, various research has verified the efficiency of CNNs in this domain. For instance, Collins et al. showed the effectiveness of a 3D CNN in conjunction with Support Vector Machines (SVM) in detecting colon and esophagogastric cancer tissues, achieving an impressive ROC-AUC of 0.93 [6]. Similarly, González-Bueno Puyal et al. introduced a hybrid 2D/3D CNN architecture for enhancing polyp detection in colonoscopy videos [7]. Nisha et al. proposed a Dual-Path CNN (DP-CNN) for polyp detection, showing higher precision, recall, and F1-score on two databases [8]. Yeung et al. built upon the U-Net architecture, introducing the Focus U-Net, a dual attention-gated deep neural network, to enhance polyp segmentation with mean DSC and IoU scores of 0.878 and 0.809, respectively [9].

Several other studies introduced unique deep learning architectures for the detection and classification of colon cancer. The concept of model fusion was manifested by Sharma et al. who championed an ensemble classifier approach integrating various classification models, yielding impressive performance metrics [10]. This approach was mirrored by Talukder M. et al., who developed a hybrid model for lung and colon cancer detection, utilizing preprocessing, cross-validation, transfer learning, and ensemble learning techniques [11]. Ho et al. developed a composite algorithm combining deep learning with a Faster Region-Based Convolutional Neural Network (Faster-RCNN) architecture, paired with a ResNet-101 feature extraction backbone for glandular segmentation [12]. Escorcia-Gutierrez et al. combined Galactic Swarm Optimization (GSO) with deep transfer learning, achieving an accuracy of 95 % on a test set of histopathological colorectal cancer images [13].

Dealing with challenges such as low contrast, blurred images, and noisy data, numerous studies employed pre-processing techniques alongside deep learning. Murugesan et al. used the YOLOv3 Multi-Scale Framework (YOLOv3-MSF) for effective detection and classification of various stages of colon cancer [14]. Similarly, Khan et al. developed an AI-based screening method for lymph node metastases in CRC, combining a segmentation model for lymph node tissues with a CNN using Xception architecture and Vision Transformer (ViT16) [15]. Chen et al. employed a self-attention-based faster R–CNN for polyp detection from colonoscopy images, achieving an accuracy of 93.4 % [16].

Notable strides were made in detecting and classifying colon adenocarcinomas. Hasan et al. applied a deep convolutional neural network (DCNN) and achieved an impressive accuracy of 99.80 % [17]. Xu et al. utilized the Inception V3 model to a similar effect [18]. Tanwar et al. used SSD for colorectal polyp detection and classification, obtaining an accuracy of 92 % [19]. S. Hosseinzadeh Kassani and colleagues investigated the efficacy of numerous deep learning architectures for the automatic segmentation of colorectal tumor tissue samples. They found that a shared DenseNet and LinkNet architecture outperformed other methods, achieving a dice

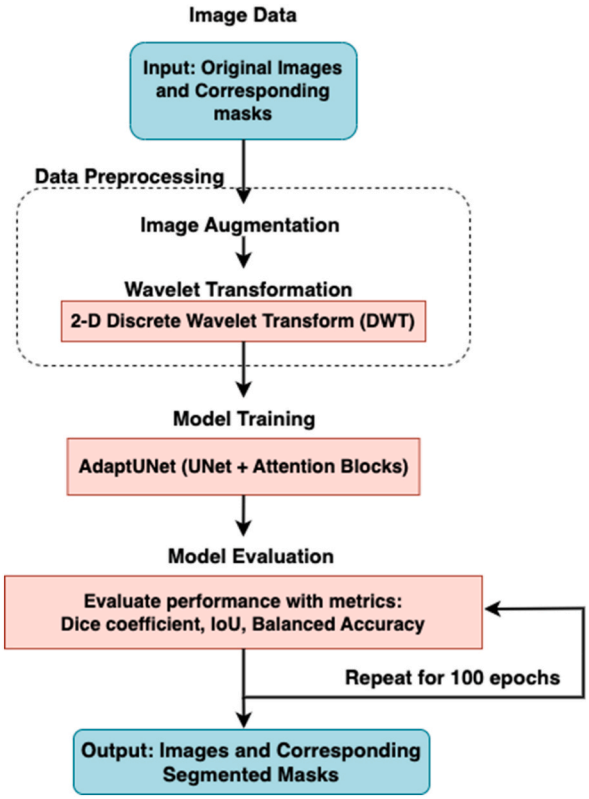


Fig. 1. Proposed workflow.

similarity index of 82.74 % ± 1.77, accuracy of 87.07 % ± 1.56, and an f1-score value of 82.79 % ± 1.79 [20].

Zhang et al. developed a method for label-free colorectal cancer screening that combines spatial light interference microscopy and AI. They manually segmented the images and used the VGG16 network for classifying cancerous and benign tissue, yielding an accuracy of 97 % [21]. The AMNet architecture, as developed by Song. P. et al., combines advanced feature fusion and attention mechanisms for improved polyp segmentation. Employing the Res2Net backbone, it focuses on high-level features to enhance performance and efficiency. The network features a multi-scale fusion model, utilizing bilinear upsampling and 2 × 2 convolution for feature integration. Central to its design are the Contextual Attention Module, Polarised Self-Attention (PSA), and Reverse Context Fusion (RCF) modules. These elements collaboratively enhance segmentation by optimizing multi-scale feature extraction, adjusting weight distributions, and fusing features and contextual guidance. Specifically, the CFP module captures varied contextual information through dilation convolution, and the PSA module refines features by redistributing weights in both channel and spatial dimensions. The RCF module further improves segmentation by merging PSA outputs with prior contextual insights, showcasing the network's modular and efficient approach to medical image segmentation [22].

Billah et al. (2017) have implemented a methodological approach that applies a three-level, two-dimensional discrete wavelet transformation to each color channel of RGB images, with a particular focus on the middle wavelet detail images for the purpose of textural analysis. Additionally, they incorporate a co-occurrence matrix to analyze spatial relationships within these images. This approach is rooted in the statistical examination of texture feature distributions by analyzing the spatial relationships of pixels, specifically how pixel values co-occur within an image at predetermined orientations and distances. The utilization of a co-occurrence matrix is especially effective in capturing both texture and structural information within the images, offering a nuanced understanding of texture and spatial relationships [23].

Overall, the integration and augmentation of deep learning techniques are propelling advancements in the early detection and diagnosis of colorectal cancer. These methods enhance detection accuracy, streamline disease detection, and improve patient outcomes. This promising field could potentially lead to major breakthroughs in colorectal cancer diagnosis and treatment.

4. Datasets

In the study, the HyperKvasir Segmented Images dataset, consisting of 1000 images focused on the polyp class, was employed for training the machine learning model. This dataset was divided into training, validation, and testing sets in a 70:15:15 ratio. Specifically, 70 % of the images were used for training the model, while 15 % were set aside for validation, and the remaining 15 % for testing. This split ensures a comprehensive assessment of the model's performance across different subsets of the data. This dataset

provides original images, their corresponding segmentation masks, and bounding boxes for each polyp image. The segmentation masks are binary images that differentiate the region of interest, which in this case is the polyp tissue (highlighted in white), from the background (depicted in black). This detailed demarcation of the polyp regions assists in accurate segmentation and serves as a crucial tool in training this model.

For the purpose of testing and evaluating the effectiveness of AdaptUNet, a set of four datasets have been employed:

1. CVC-300 [24]: This dataset comprises 60 endoscopic images depicting various gastrointestinal (GI) tract diseases, including polyps, ulcers, and inflammatory conditions. It serves as a robust benchmark for evaluating the efficacy of algorithms and models in identifying and classifying GI tract diseases.
2. ETIS-LaribDB [25]: The ETIS-LaribDB dataset has a particular focus on colorectal lesions, especially polyps. It contains 196 endoscopic images along with ground truth annotations, including binary masks for polyp segmentation. This dataset has been widely utilized for benchmarking computer-aided diagnosis systems and evaluating polyp detection and segmentation algorithms.
3. Kvasir-SEG [26]: The Kvasir-SEG dataset comprises 100 gastrointestinal endoscopy images that include annotations specifically designed for semantic segmentation. It includes a variety of anatomical structures and pathological conditions, providing pixel-level annotations for organs, lesions, and background regions. It has proven instrumental in the development and evaluation of segmentation models in the field of gastroenterology.
4. CVC-ColonDB [27]: The CVC-ColonDB dataset focuses specifically on colonoscopy images and provides a collection of 380 images with corresponding ground truth annotations. It is designed for evaluating the performance of algorithms and models in detecting and segmenting polyps in the colon. The dataset includes various types of polyps, along with normal tissue and other abnormalities found in colonoscopy images.

During the experiments, the model trained on the HyperKvasir Segmented Images dataset were tested on the CVC-300, ETIS-LaribDB, and Kvasir-SEG datasets, by leveraging the weights obtained from the training process. This approach allowed us to assess the adaptability and robustness of the trained model on different datasets.

5. Proposed methodology

The proposed methodology as shown in Fig. 1 for polyp segmentation encompasses two key components: pre-processing of images and a modified U-Net model architecture with attention mechanisms. The workflow begins with the input of original images and their corresponding masks, which are subjected to a Data Preprocessing stage. Within this stage:

1. Image Augmentation is performed to artificially increase the dataset's size and variability, enhancing the model's ability to generalize.
2. The images then undergo a Wavelet Transformation, specifically a 2-D Discrete Wavelet Transform (DWT) where they are processed using the 'bior1.3' biorthogonal wavelet type. This transformation is conducted with the PyWavelets library, which performs a single level of decomposition. This operation divides the image into four sets of coefficients: the approximation coefficient (LL) and detail coefficients (LH - horizontal, HL - vertical, HH - diagonal). Each set of coefficients captures different aspects of the image's structure at various resolutions and orientations. After decomposition, the coefficients are resized back to 256×256 pixels to ensure uniformity. The resized components are then concatenated back into a single image forming a multi-channel image where each channel corresponds to one set of coefficients. This concatenated image is further normalized to have values between 0 and 1 to ensure that the model receives inputs within a standardized range. The transformation enhances edges and textures along with retaining the original image's salient features that are important for the segmentation task.

Following preprocessing, the images are fed into the Model Training stage, where the AdaptUNet model is employed. AdaptUNet is a hybrid model that combines the architecture of U-Net with Attention Blocks. These attention blocks help the model focus on specific regions of the image, enhancing its segmentation accuracy.

After training, the model enters the Model Evaluation stage. For this stage, the trained model was tested on both the validation subset (15 % of HyperKvasir dataset) and the external datasets (CVC-300, ETIS-LaribDB, Kvasir-SEG, and CVC-ColonDB). Here, its performance is assessed using various metrics, including the Dice coefficient, Intersection over Union (IoU), and Balanced Accuracy. This evaluation process is repeated for 100 epochs to ensure the model's robustness and accuracy over multiple iterations.

The final output of the workflow is a set of images paired with their corresponding segmented masks, showcasing the model's ability to accurately identify and delineate colorectal polyps.

This section introduces these components and provides a brief overview of their significance in achieving accurate and robust polyp segmentation.

5.1. Pre-processing

Pre-processing plays a crucial role in preparing the dataset and enhancing the input images for the subsequent segmentation task. The pre-processing techniques employed by the authors include data augmentation and wavelet transformation. These techniques enable the model to learn from a more diverse and informative dataset, enhancing its ability to accurately segment polyps from colonoscopy images.

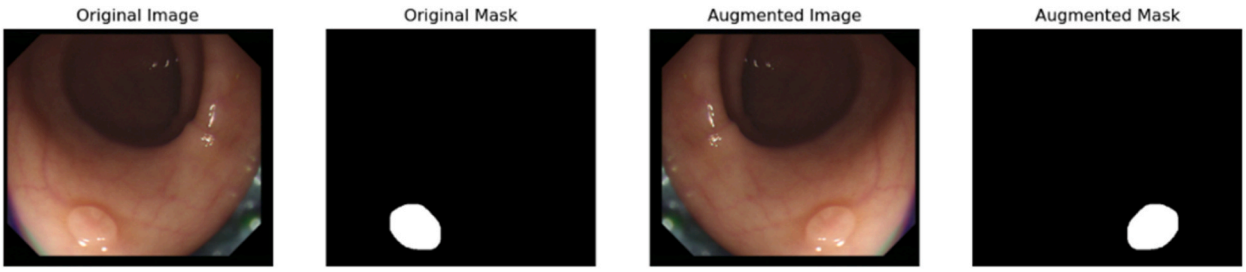


Fig. 2. Visual comparison of original and augmented images and masks.

The following discussion delves deeper into the specific techniques utilized for data augmentation. These components work in synergy to improve the accuracy and robustness of polyp segmentation, contributing to more effective diagnosis of colorectal diseases.

5.1.1. Augmentation techniques

In this study, the authors employed the Albumentations library, which offers a wide range of powerful and efficient augmentation techniques. A combination of geometric and image-only transformations was utilized to augment the training dataset, allowing the model to effectively deal with diverse variations in polyp images. Geometric transformations, such as horizontal and vertical flips, as well as random rotations, were employed to introduce spatial variations, simulate natural distortions, and replicate real-life scenarios.

These transformations enable the model to learn from different polyp orientations, shapes, and appearances, making it more robust and adaptable to diverse input data. Additionally, image-only transformations such as random brightness contrast, random gamma adjustment, and CLAHE were utilized to improve the model's performance under different lighting conditions, noise levels, and contrast variations. By augmenting the dataset with these transformations, the model becomes more capable of accurately segmenting polyps in endoscopic images, ultimately leading to improved performance and generalization.

Fig. 2 showcases two example images from the training dataset, highlighting the impact of various augmentation techniques discussed. The original images and masks represent the raw input data used for training the model. They depict real-world scenarios with inherent variations in lighting, perspective, and object placement. The augmented images exhibit increased diversity in terms of object positions, orientations, and backgrounds, providing a more comprehensive representation of real-world scenarios. This augmented dataset facilitates better training of the model, enabling it to handle a wider range of variations and improve its overall performance in object recognition and localization tasks. Note: The images in Fig. 2 are for illustrative purposes only and do not represent the full range of augmented images and masks used in this research.

5.1.2. Wavelet transformation

In this study, the authors employed a comprehensive data preprocessing approach to prepare the dataset for the polyp segmentation task. The preprocessing pipeline consisted of several key operations aimed at improving the feature representation of the images:

1. Image resizing: Each input image (I) is adjusted to a standard size of (256, 256, c), ensuring consistent dimensions without altering the color channels c. This is essential for ensuring consistent input dimensions across the dataset.
2. Image normalization: The pixel values of each image (I) are rescaled to a common range. This involves dividing each pixel value by the maximum pixel value (255 for 8-bit images). Equation (1) represents the above:

$$I_{normalized} = I/255 \tag{1}$$

3. Grayscale conversion: To convert RGB images to grayscale, a weighted sum of the color channels is calculated. When dealing with an image containing red (R), green (G), and blue (B) channels, the grayscale image ($I_{grayscale}$) is computed as shown in Equation (2):

$$I_{grayscale} = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B \tag{2}$$

4. Wavelet transformation: Grayscale images undergo a 2D discrete wavelet transform using the 'bior1.3′ biorthogonal wavelet where they are divided into two sets of coefficients: approximation coefficients (A) and detail coefficients (H - *horizontal, V - vertical, D - diagonal*). By utilizing the impulse response of a low-pass filter denoted as $h[n]$, the approximation coefficients can be computed as:

$$A(x,y) = \sum\sum f(u,v) \times h(x-u) \times h(y-v) \tag{3}$$

5. Image concatenation: The original images ($I_{normalized}$) are concatenated with the wavelet-transformed images (A) to create a comprehensive image representation (I_{final}).
6. Mask preprocessing: The corresponding masks are resized and binarized in a similar manner to the images. If M is the original mask and T is the chosen threshold (usually 0.5 for binary masks), the binarization process can be formulated as:

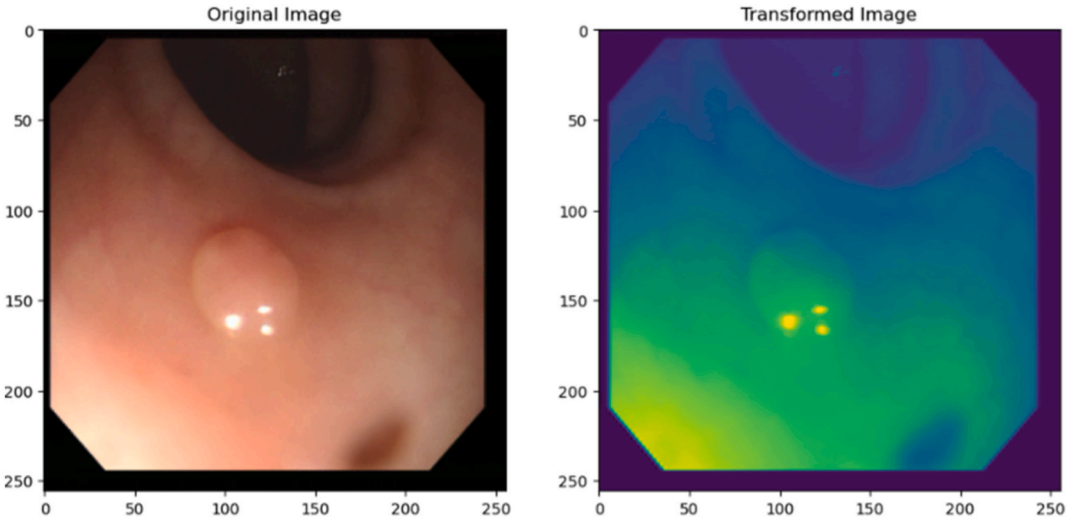


Fig. 3. Comparison of images before and after wavelet transformation.

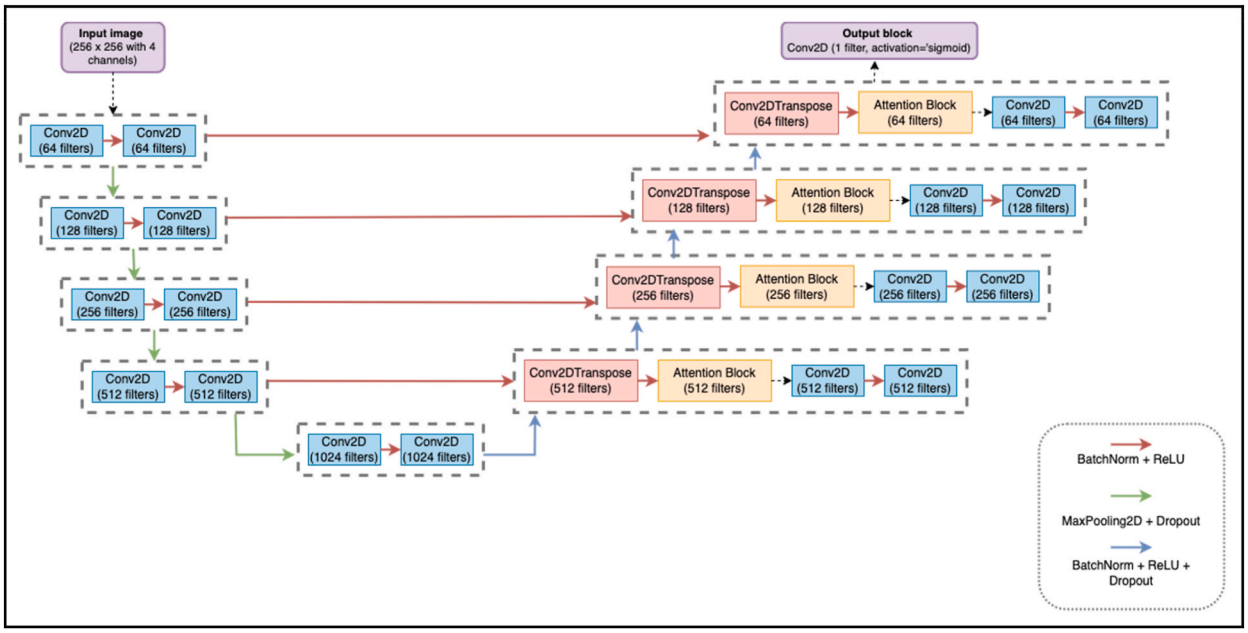


Fig. 4. Proposed AdaptUNet architecture.

$$M_{binarized} = 1 if M > Telse 0 \tag{4}$$

By incorporating a combination of image transformations and wavelet transforms, the preprocessing pipeline effectively extracted informative features, providing a strong foundation for training the subsequent polyp segmentation model.

Fig. 3 illustrates a visual comparison of images before and after the application of a 2D discrete wavelet transform (DWT). The left side of the figure displays the original images extracted from the training dataset, while the right side showcases the corresponding images after undergoing the wavelet transformation.

Furthermore, the specificity of the 'bior1.3' biorthogonal wavelet in capturing the nuanced textures and edges peculiar to polyp structures in colonoscopy images represents a novel approach within the preprocessing methodology. This wavelet was selected based on its proven effectiveness in medical image analysis for retaining critical spatial frequency characteristics, which are pivotal for the precise delineation of polyp boundaries [28]. Additionally, the subsequent concatenation of wavelet-transformed images with the original dataset is a distinctive strategy that enriches the model's input with a multifaceted representation of polyp features. This enriched set of features is specially crafted to improve the model's ability to distinguish between different elements, offering a novel

Table 1
Hyperparameter table of the suggested AdaptUNet model.

Hyperparameters	Value
Image input shape	(256, 256, 4)
Mask input shape	(256, 256, 1)
Batch size	16
Epochs	100
Optimizer	SGD
Loss function	DiceBCELoss

contribution to the task of identifying polyps. This approach marks a notable improvement on the usual methods of preparing images for analysis.

5.2. AdaptUNet model

The model architecture employed, AdaptUNet is a customized variant of the U-Net architecture, a widely recognized framework for semantic segmentation tasks. This architecture has proven effective in various image analysis tasks, including polyp segmentation.

The proposed model, referred to as AdaptUNet in Fig. 4, adopts an encoder-decoder architecture that takes advantage of skip connections. These connections effectively preserve spatial details while facilitating information flow throughout the network. By incorporating both low-level and high-level features, the model strives to capture intricate patterns and structures in the polyp images. Setting AdaptUNet apart from traditional U-Net architectures is the integration of adaptive spatial and channel attention blocks. These novel components dynamically refine the model's focus on the most informative features within the feature maps, accounting for both spatial and channel-wise relevance. This advanced strategy is designed to yield a more nuanced and accurate segmentation by enhancing the network's sensitivity to critical features in the image data.

During the initial phase, the input images which have a resolution of 256×256 pixels and consist of four channels, are subjected to a series of convolutional layers. These layers, each comprising a 3×3 kernel, perform feature extraction operations boosting the model's proficiency to identify relevant patterns. To ensure stable training and introduce non-linearity within the network, subsequent steps involve the utilization of batch normalization and rectified linear unit (ReLU) activation functions. To enable down sampling and preserve the most salient information, the authors employ MaxPooling2D layers with a pool size of 2×2. Additionally, dropout regularization (rate $= 0.1$) is introduced to mitigate overfitting during training. The network employs spatial attention mechanism that computes attention weights based on the features' spatial distribution. Utilizing a dynamically sized convolution kernel, the network adaptively adjusts its focus to different spatial scales, which is critical for capturing the varying sizes and shapes of polyps in colonoscopy images. Similarly, the channel attention mechanism selectively enhances the most informative channels, ensuring a comprehensive feature representation for accurate segmentation.

The bridge section of the model utilizes a set of two convolutional layers with 1024 filters and a 3×3 kernel size. These layers focus on capturing high-level features that are crucial for accurate polyp segmentation.

The decoder part of the model is crucial for recovering the spatial resolution of the segmentation masks. To achieve this, Conv2DTranspose layers are employed, which perform upsampling by a factor of 2. Subsequently, concatenation operations are applied to combine the upsampled feature maps with the corresponding feature maps from the encoder pathway. This integration of multi-scale information aids in precise localization and context-aware segmentation.

To selectively attend to informative regions, attention blocks are introduced in the decoder pathway. These blocks consist of attention mechanisms, including spatial and channel attention. The spatial attention block adaptively weighs the importance of different spatial regions, while the channel attention block focuses on relevant channels for improved feature representation. These attention mechanisms enhance the model's discriminative power and enable it to concentrate on essential polyp regions. Within each decoder block, the authors utilize convolutional layers with decreasing numbers of filters (512, 256, 128, 64), batch normalization, ReLU activation functions, and dropout regularization (with rates of 0.4, 0.3, 0.2, and 0.1, respectively). This hierarchical decoding process allows for the extraction of increasingly abstract features and aids in the reconstruction of the image. These adaptive attention blocks are integrated at each level of the decoder, ensuring that the upsampled feature maps are refined with both the gating signal and the inter signal. This results in a feature map that is optimized for the subsequent convolutional operations. The spatial attention blocks operate on 2D feature maps using 2D convolutions, while the channel attention blocks process 1D feature maps with 1D convolutions. Consequently, at each stage of the decoder, the feature maps are attentively adjusted, both spatially and channel-wise, before being merged, leading to a more discerning reconstruction of the segmentation mask.

Ultimately, the model generates results through the utilization of a single-filter convolution and a kernel size of 1, employing a sigmoid activation function in the subsequent step. This generates the predicted segmentation mask, where each pixel value represents the probability of it belonging to the polyp class. In summary, the customized U-Net architecture, enriched with attention mechanisms and skip connections, facilitates accurate polyp segmentation by effectively capturing detailed information and contextual cues.

In this study, crucial model hyperparameters, including the cyclic learning rate (CLR) and the loss function utilized, Dice binary cross-entropy (BCE), played instrumental roles in improving the model's performance. The CLR, implemented using a custom CyclicLR class, optimized the learning rate during model training. It dynamically modulated the learning rate throughout the process, enabling the model to converge faster and potentially achieve enhanced performance. The CLR's rate was updated at the inception of each

batch, fluctuating within each cycle based on the current iteration. This allowed the model to probe different learning rates and potentially discover an optimal range for improved convergence.

Complementing this was the Dice BCE loss function, which guided the polyp segmentation training process. This loss function combined the BCE loss, which assessed pixel-wise binary classification error, and Dice loss, quantifying the similarity between the predicted and target masks. By minimizing the Dice BCE loss during training, the model aimed to augment the precision of polyp segmentation, considering both binary classification and overlap extent between predicted and target masks. Consequently, the judicious employment of CLR and Dice BCE loss as key model hyperparameters amplified the training process but also facilitated achieving superior results in polyp segmentation.

Table 1 provides an overview of the key hyperparameters used in the development and training of the proposed AdaptUNet model.

5.3. Computational requirements

An essential aspect of our model's design is its computational efficiency, which we quantitatively assess by the total number of parameters and the memory footprint. The model comprises a total of 28,621,413 parameters, divided into 28,609,637 trainable parameters and 11,776 non-trainable parameters. The non-trainable parameters primarily reside in layers such as Batch Normalization, which are utilized in a frozen state during the inference phase to stabilize the network's predictions.

The memory requirement for storing the model parameters is a critical factor, especially when deploying the model on hardware with limited resources. Given that each parameter is represented as a 32-bit floating-point number, the total memory footprint of the model is approximately 109.18 MB. This calculation is based on the assumption that each parameter requires 4 bytes of storage:

Total memory footprint $= 28{,}621{,}413 \times 4$ bytes ≈ 109.18 MB.

This compact memory footprint allows our model to be deployed in cloud-based environments and in edge devices, facilitating real-time application.

5.4. Performance measures

The performance measures incorporated in the evaluation encompass the Dice coefficient, IOU, and Balanced Accuracy. These metrics offer quantitative assessments to gauge the precision and efficacy of image segmentation algorithms and models.

1. **Dice Coefficient:** The Dice coefficient is utilized to assess the similarity or overlap between two sets, particularly in image segmentation. It serves as a metric to evaluate the agreement between the predicted segmentation and the ground truth segmentation. With a scale from 0 to 1, the Dice coefficient signifies the degree of overlap, with a value of 1 representing a complete match between the predicted and ground truth segmentations. The formula presented in equation (5) calculates the Dice coefficient as follows:

$$DICE = \frac{2TP}{2TP + FP + FN} \tag{5}$$

2. **Intersection over Union**: *IoU* commonly referred to as Jaccard Index represents another similarity measure used in image segmentation tasks. It calculates the ratio of the intersection to the union of two sets. In the context of image segmentation, it measures the overlap between the predicted segmentation and the ground truth segmentation. Like the Dice coefficient, the *IoU* ranges from 0 to 1, with 1 indicating a perfect overlap between the predicted and ground truth segmentations. The formula presented in equation (6) calculates the *IoU* as follows:

$$IoU = \frac{TP}{TP + FP + FN} \tag{6}$$

3 **Balanced Accuracy (BAcc)**: The Balanced Accuracy is a performance measure that considers both true positive and true negative rates. It provides an overall assessment of a model's performance across different classes or categories, considering imbalanced datasets. Balanced Accuracy is particularly useful in imbalanced datasets because it gives an equal weight to the performance on each class, regardless of its frequency in the dataset which is why we have chosen this metric over F1 score to calculate the performance of our proposed architecture. The *BAcc* is calculated as the average of sensitivity (true positive rate) and specificity (true negative rate). The *BAcc* ranges from 0 to 1, with a higher value indicating a higher level of overall accuracy in the segmentation task. The formula presented in equation (7) calculates the Balanced Accuracy as follows:

$$BAcc = 0.5 \times (Sensitivity + Specificity) \tag{7}$$

6. Results and discussions

This section presents a comparative analysis of several models, including UNet, SFA, PraNet, UACANet-L, EU-Net, MSNet, BDG-Net, SANet, MFBGR, and AdaptUNet. Each model is evaluated using various metrics such as Dice coefficient, Intersection over Union (IoU), and Balanced Accuracy across multiple datasets, such as CVC-300, CVC-ColonDB, Kvasir, and ETIS-LaribDB. Their respective

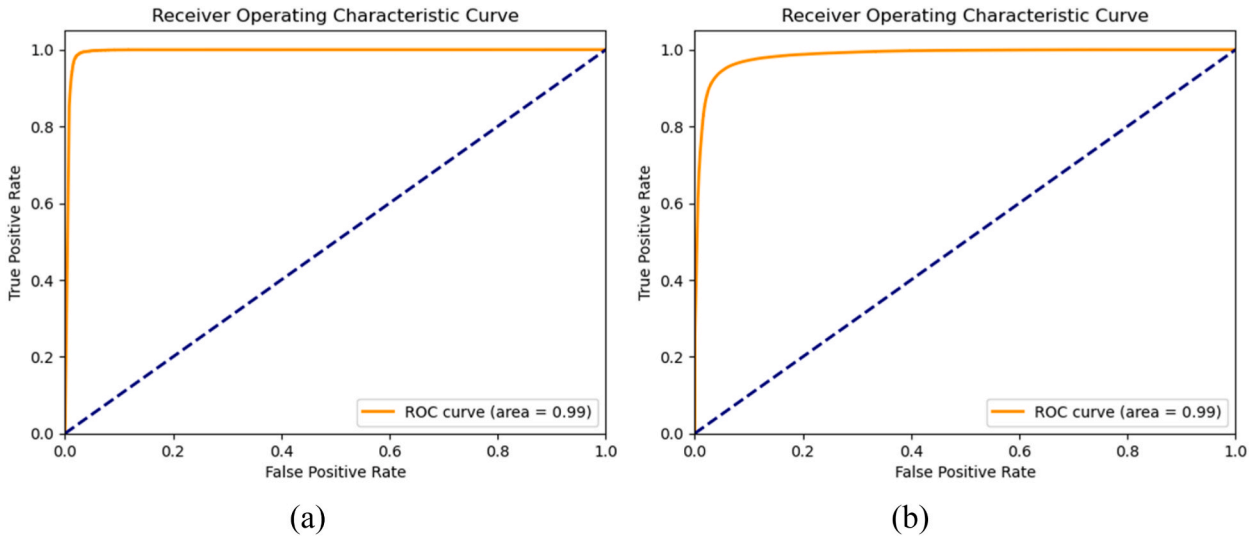


Fig. 5. ROC Curves of the model training on (a) CVC-300 and (b) CVC-ColonDB dataset.

Table 2
Evaluation metrics of AdaptUNet on the CVC-300 and CVC-ColonDB datasets.

Network	CVC-300			CVC-ColonDB		
	Dice ↑	IoU ↑	BAcc ↑	Dice ↑	IoU ↑	BAcc ↑
UNet [29]	0.5042	0.4401	0.7521	0.7176	0.6390	0.8790
SFA [30]	0.4607	0.3401	0.8211	0.4714	0.3316	0.9157
PraNet [31]	0.7153	0.6449	0.8591	0.8767	0.8039	0.9672
UACANet-L [32]	0.7532	0.6803	0.8717	0.9135	0.8526	0.9683
SANet [33]	0.7591	0.6787	0.8951	0.8990	0.8294	0.9598
MSNet [34]	0.7558	0.6773	0.8815	0.8727	0.8077	0.9633
EU-Net [35]	0.7584	0.6839	0.9130	0.8404	0.7688	0.9677
BDG-Net [36]	0.8012	0.7276	0.9103	0.9070	0.8425	**0.9771**
MFBGR [37]	0.8031	0.7281	0.9081	**0.9174**	**0.8550**	0.9769
AdaptUNet	**0.9104**	**0.8368**	**0.9880**	0.8075	0.7215	0.9687

performances are concisely summarized in tabular form for easy comparison.

Furthermore, a detailed visual representation illustrating the performance metrics of the AdaptUNet model throughout the training process across the four datasets is provided. This includes the Loss Curve, Accuracy Curve, IoU Evolution, Dice Evolution, and Receiver Operating Characteristic (ROC) curve, offering valuable insights into the model's progressive performance enhancement.

Additional figures showcase the evaluation results of the AdaptUNet model on each of the four datasets (CVC-300, CVC-ColonDB, Kvasir, and ETIS-Larib). These include the wavelet-transformed input image, the true mask (ground truth), and the mask predicted by the AdaptUNet model. These depictions underscore the proficiency and precision of the model's polyp detection and segmentation capabilities in comparison to the ground truth. Each subsection emphasizes a particular dataset and offers an in-depth analysis of the model's performance.

6.1. Results on the CVC-300 and CVC-ColonDB datasets

This subsection displays the results drawn from an exhaustive analysis that utilizes several metrics to gauge how the model perform on the CVC-300 and CVC-ColonDB datasets. It is accompanied by a comprehensive interpretation of the findings.

Fig. 5 provides a comprehensive visual representation of the model's performance metrics throughout the training process on the CVC-300 dataset and CVC-ColonDB dataset using the Receiver Operating Characteristic (ROC) curve. The ROC curves visualizes the trade-off between sensitivity and specificity, providing a comprehensive measure of the model's performance across various threshold settings. The ROC curve demonstrates a value of 0.99 for both the CVC-300 and CVC-ColonDB datasets. This indicates the model's exceptional performance in distinguishing between positive (presence of polyps) and negative instances, specifically in polyp detection. The high Area under ROC curve (AUC value) highlights the model's ability to accurately rank instances and assign higher probabilities to positive cases. It means that the model is exceptionally effective at assigning higher probability scores to actual cases of polyps while giving lower scores to non-polyp instances. Therefore, the model is reliable and effective in aiding early detection or screening for polyps, as evidenced by the strong discriminatory power exhibited in Fig. 5.

Table 3
Ablation study on the CVC-300 dataset.

Model Configuration	Dice Coefficient	IoU Coefficient	Balanced Accuracy
Original Model (Complete Network)	0.9104	0.8368	0.9880
Without Wavelet Transform	0.0815	0.0429	0.7185
Without Extra Decoder Blocks	0.0649	0.0338	0.6522
Without Both Components	0.0635	0.0331	0.6487

Table 4
Evaluation metrics of AdaptUNet on the Kvasir and ETIS-Larib datasets.

Network	Kvasir			ETIS-Larib		
	Dice ↑	IoU ↑	BAcc ↑	Dice ↑	IoU ↑	BAcc ↑
UNet [29]	0.8252	0.7559	0.9131	0.4024	0.3430	0.7273
SFA [30]	0.7311	0.6187	0.8869	0.2996	0.2187	0.7681
PraNet [31]	0.9035	0.8475	0.9484	0.6367	0.5757	0.8311
UACANet-L [32]	0.9147	0.8621	0.9546	0.7690	0.6912	0.9018
SANet [33]	0.9086	0.8532	0.9503	0.7633	0.6699	0.9308
MSNet [34]	0.9145	0.8621	0.9505	0.7332	0.6651	0.8920
EU-Net [35]	0.9115	0.8587	0.9566	0.6904	0.6127	0.9028
BDG-Net [36]	**0.9197**	**0.8682**	0.9541	0.7576	0.6877	0.9050
MFBGR [37]	0.8031	0.7281	0.9081	0.7850	0.7044	0.9327
Proposed AdaptUNet	0.8749	0.7883	**0.9601**	**0.8075**	**0.7215**	**0.9687**

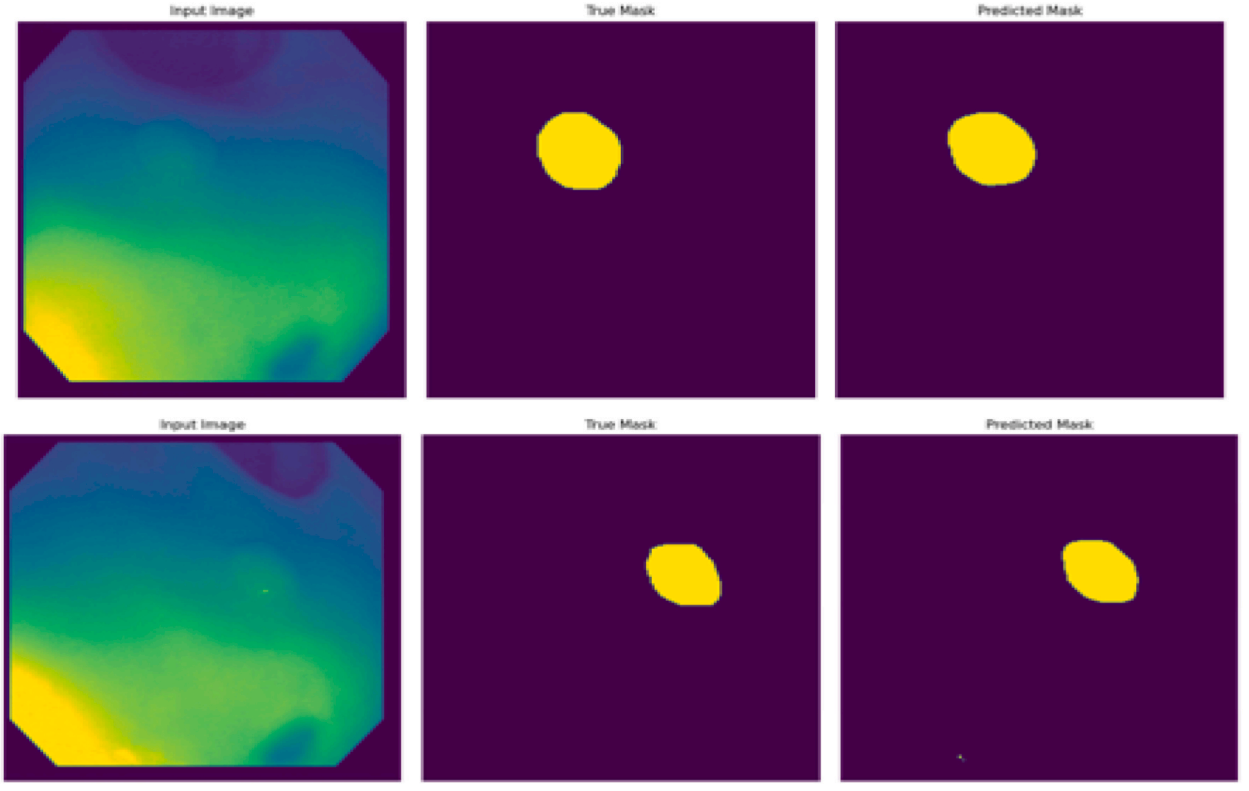


Fig. 6. Visual evaluation results for CVC-300 Dataset.

In Tables 2 and 4, the upward arrow (↑) next to each performance measure indicates that higher values correspond to better results. Table 2 provides a comparative analysis of the AdaptUNet model's performance on the CVC-300 and CVC-ColonDB datasets.

In the context of the CVC-300 dataset, the AdaptUNet model excelled, achieving superior results compared to all other models evaluated, including state-of-the-art methodologies like UNet, SFA, PraNet, UACANet-L, SANet, MSNet, EU-Net, BDG-Net, and MFBGR. The compared models in these tables use different pre-processing and data partitioning strategies. As indicated by a Dice

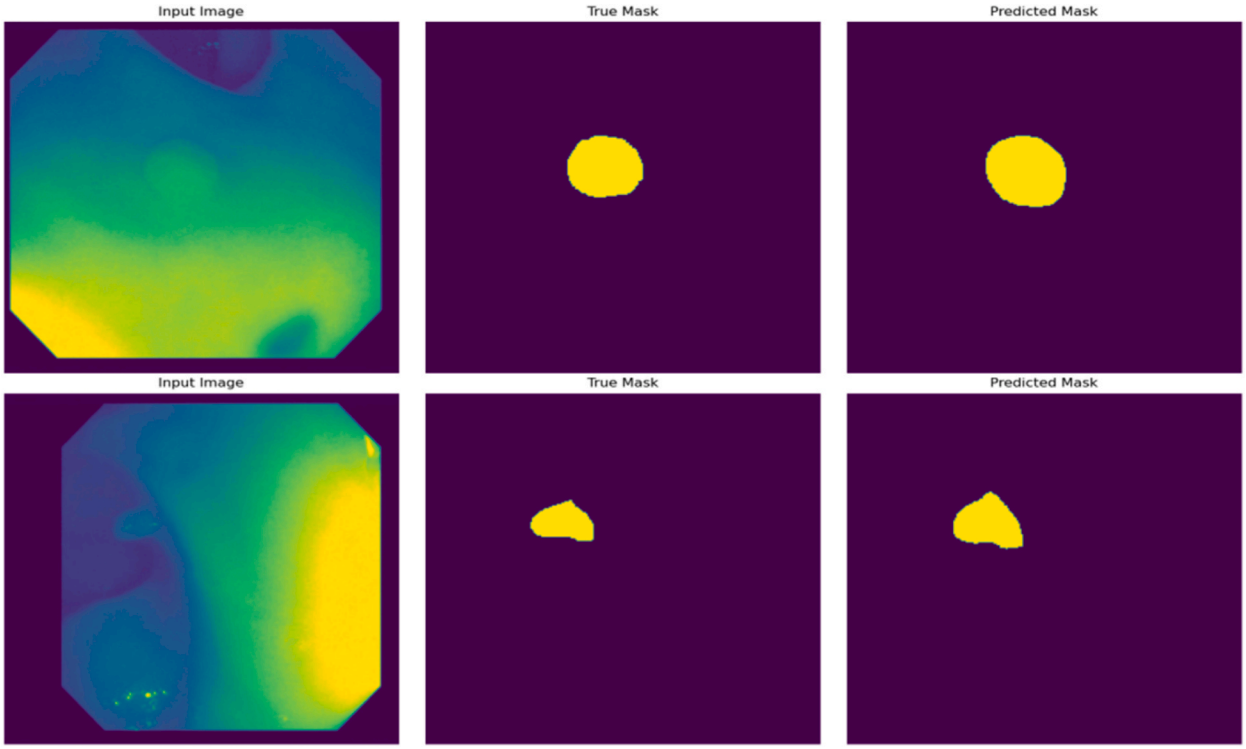


Fig. 7. Visual evaluation results for CVC-ColonDB dataset.

coefficient of 0.9104 and an IOU score of 0.8368, this model demonstrated extraordinary accuracy in polyp segmentation. The Balanced Accuracy score of 0.9880 further underlines the model's exceptional performance. These results illuminate the potential of the AdaptUNet model for high-precision tasks in polyp detection and diagnosis on the CVC-300 dataset.

Moving on to the CVC-ColonDB dataset, the model also delivered competitive performance. Despite not surpassing all models on Dice and IOU indicators, it displayed reliable segmentation capabilities, achieving a Dice coefficient of 0.8075 and an IOU score of 0.7215. A Balanced Accuracy score of 0.9687 further indicates its overall high accuracy. Although models like UACANet-L, BDG-Net, and MFBGR scored slightly higher on some indicators, the balance of computational efficiency, model complexity, and generalization capabilities position this model as a compelling option.

These results, together, offer valuable insights into the strengths and weaknesses of different models, guiding researchers in choosing suitable methodologies for polyp segmentation on the CVC-300 and CVC-ColonDB datasets. Opportunities for future research and performance enhancements of the model on these specific datasets are also suggested.

Fig. 6 presents the evaluation results derived from testing the model on the CVC-300 Dataset. The figure is composed of three main components.

1. Firstly, it shows the input image which has undergone wavelet transformation, providing a visual representation of how the input data is prepared for the model.
2. Secondly, it displays the true mask corresponding to the input image, serving as a benchmark for the polyp segmentation that the model aims to achieve.
3. Lastly, the figure includes the mask as predicted by the model, demonstrating the effectiveness and accuracy of the model's polyp detection and segmentation capabilities in comparison to the true mask.

This triptych of images provides a comprehensive visualization of the model's performance. Fig. 7 showcases the evaluation of the model on the CVC-ColonDB dataset providing a concise overview of the model's segmentation capabilities.

6.2. Ablation study

In medical image segmentation, understanding the contribution of each component to a model's overall performance is crucial for optimization. This section presents an ablation study designed to dissect the impact of specific elements within our proposed AdaptUNet architecture for colorectal polyp segmentation. The CVC-300 dataset, known for its challenging and diverse set of images, was selected as the focal point for this study due to the exemplary results achieved with our complete model. The components subjected to ablation include the wavelet transform and extra decoder blocks, chosen for their presumed significance in enhancing model

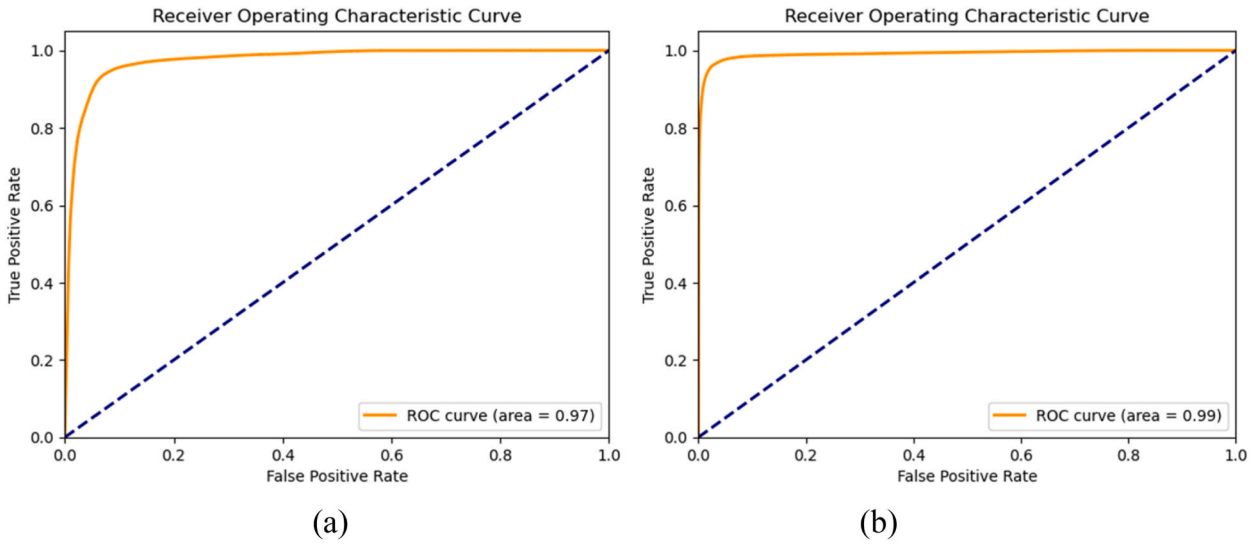


Fig. 8. (a) ROC Curve of Kvasir dataset, (b) ROC Curve of ETIS Datasets.

performance. This study aims to elucidate their individual and combined contributions towards the model's accuracy and generalization capabilities.

7. Methodology

Our ablation study adheres to a consistent experimental setup across all tests to ensure comparability of results. Each variant of the AdaptUNet model—without wavelet transform, without extra decoder blocks, and lacking both—was trained on the CVC-300 dataset for 1000 epochs. This consistency in training duration underscores our commitment to a fair and rigorous evaluation of each component's impact. Performance metrics including the Dice coefficient, Intersection over Union (IoU), and Balanced Accuracy (BAcc) serve as the benchmarks for assessment, providing a multifaceted view of each ablation scenario's effects on segmentation quality.

To present the findings, the authors employ a comparative format that juxtaposes the performance metrics of the original model against those of its ablated variants. The table below encapsulates these results, offering a clear visualization of the impact exerted by the wavelet transform and extra decoder blocks on the proposed model's efficacy. Table 3 elucidates the profound impact of both the wavelet transform and extra decoder blocks on the model's performance. The substantial differences in performance metrics observed in the ablation study can be explained by the critical roles played by both the wavelet transform and the decoder blocks in the proposed model's architecture. These components significantly influence the model's performance.

7.1. Wavelet transform

The wavelet transform decomposes an image into various frequency components, capturing both the approximation (low-frequency details) and detailed coefficients (high-frequency details). This operation enriches the feature space available to the model, allowing it to capture and utilize both global structures and fine details within the images more effectively. Removing the wavelet transform leads to a model that relies solely on raw pixel intensities, which might not be as informative or discriminative for complex segmentation tasks.

By incorporating wavelet transforms, the proposed model can analyze the image content at multiple scales, improving its ability to recognize patterns and structures of varying sizes. This is particularly important in medical or detailed imagery where objects of interest can appear at different scales. The absence of this component forces the model to operate at a single scale, significantly limiting its segmentation capabilities.

7.2. Decoder blocks

Decoder blocks in U-Net-like architectures are essential for recovering spatial resolution after the encoding (downsampling) phase. They gradually upsample the feature maps and integrate skip connections from the encoder, which provide rich contextual information. This process is crucial for accurate pixel-wise classification needed in segmentation tasks. Without extra decoder blocks, the proposed model loses a significant amount of spatial context, leading to poorer reconstruction of the segmented objects.

The proposed model utilizes attention mechanisms within the decoder blocks to focus on relevant features by weighting the importance of different spatial locations and channels. This targeted approach enhances the model's ability to distinguish between relevant and irrelevant patterns in the image, leading to more precise segmentation. Removing these blocks reduces the model's

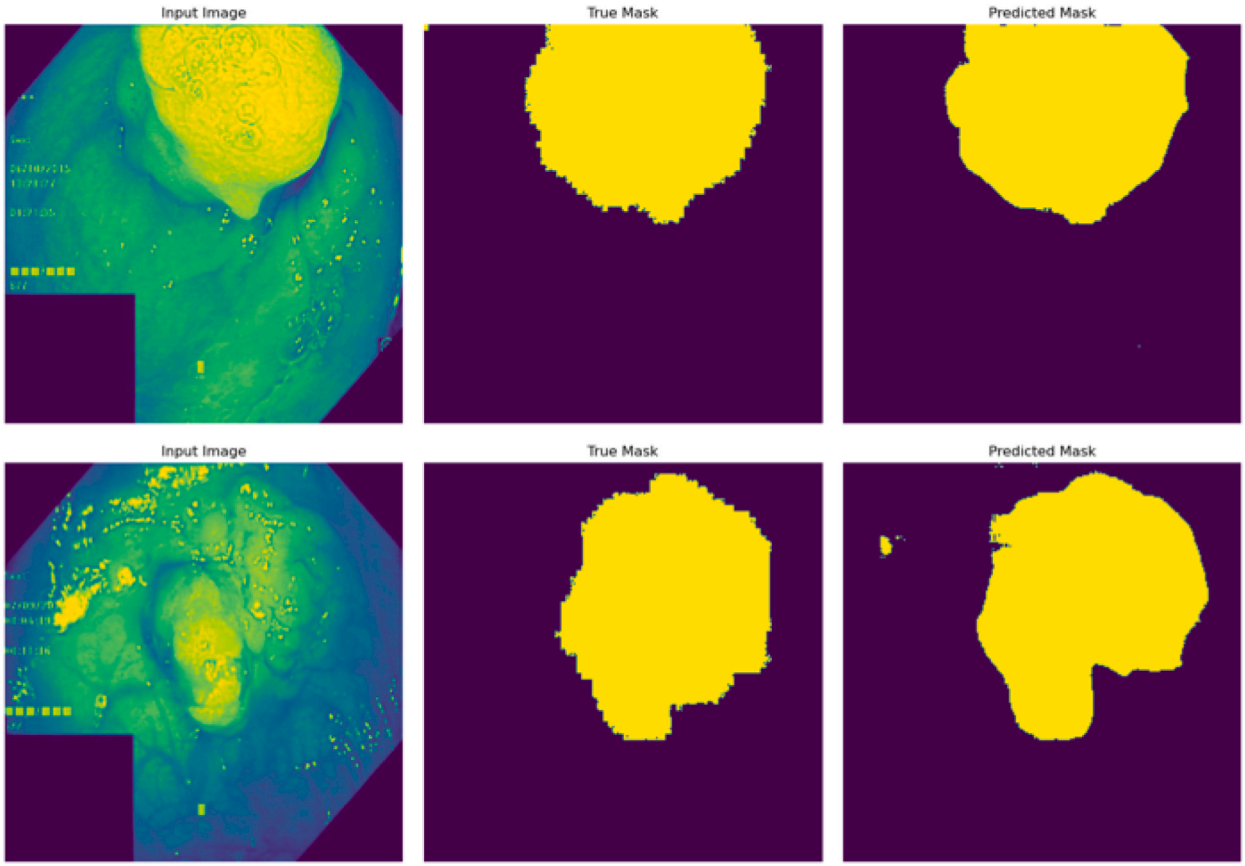


Fig. 9. Visual evaluation results on the Kvasir dataset.

capacity and its ability to selectively process information, further degrading performance.

7.3. Combined effect of Removing both components

The combined removal of both the wavelet transform and extra decoder blocks drastically reduces the proposed model's ability to capture and utilize critical information necessary for segmentation. The wavelet transform's multi-scale, enriched feature representation and the decoder blocks' spatial context recovery and attention-guided feature selection are both pivotal for achieving high accuracy in segmentation tasks. Without these, the model is significantly handicapped, relying on a much simpler, less informative feature set and lacking the mechanisms to effectively reconstruct detailed segmentations, as reflected in the drastically lower performance metrics observed.

Through this ablation study, it becomes evident that both the wavelet transforms and the extra decoder blocks are integral to the AdaptUNet model's superior performance, highlighting the importance of these components in enhancing edge detection, texture analysis, and the model's focus on relevant image regions for accurate polyp segmentation.

7.4. Results on the Kvasir and ETIS-Larib datasets

In this subsection, results obtained from the extensive analysis using various metrics are presented to assess how to model performs on the Kvasir dataset, accompanied by a detailed discussion of the findings. Fig. 8 showcases the model's training performance metrics on the Kvasir dataset and ETIS dataset, including changes in ROC curve, which collectively illustrate the model's performance evolution.

In Fig. 8, the ROC curves for the Kvasir and ETIS datasets are displayed, showing AUC values of 0.97 and 0.99, respectively. These high AUC values indicate that the models trained on these datasets perform exceptionally well in distinguishing between positive and negative instances. The AUC of 0.97 for Kvasir suggests a strong discriminatory power, while the AUC of 0.99 for ETIS suggests an even higher level of accuracy in polyp detection and classification. These results demonstrate the effectiveness of the models in accurately predicting polyp presence, making them valuable tools for diagnostic and screening purposes.

Table 4 provides a detailed comparison of the AdaptUNet model's performance on the Kvasir and ETIS datasets, benchmarked against several other models including UNet, SFA, PraNet, UACANet-L, SANet, MSNet, EU-Net, BDG-Net, and MFBGR. Focusing on the

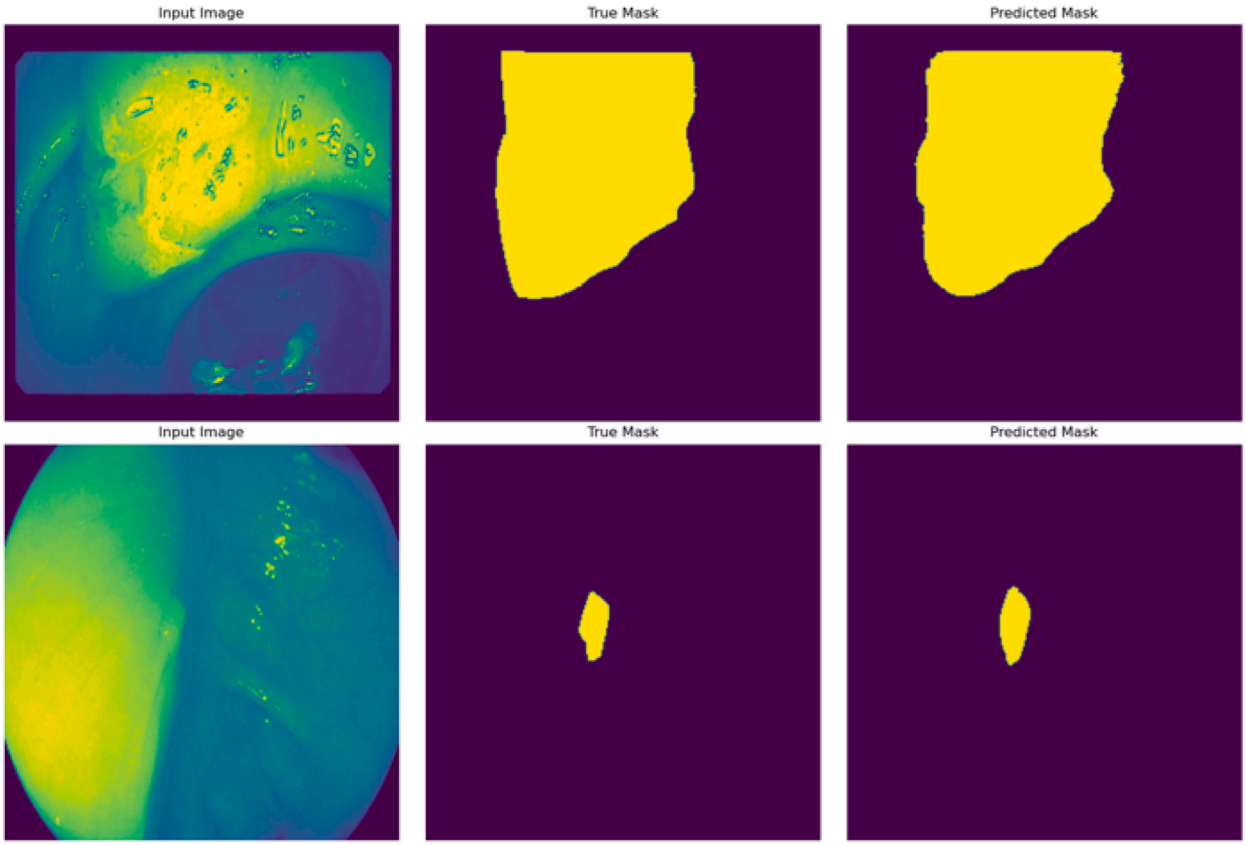


Fig. 10. Visual evaluation results on the ETIS Dataset.

Kvasir dataset, the AdaptUNet model demonstrated strong accuracy and effectiveness in polyp segmentation, as shown by its Dice coefficient of 0.8749 and an IOU score of 0.7883. Although UACANet-L, SANet, and MSNet achieved similar or slightly higher scores, the AdaptUNet model proved its robustness, given its high Balanced Accuracy score of 0.9601. Additional factors, such as computational efficiency, model complexity, and generalization capabilities, further underline the competitiveness of the AdaptUNet model. The DICE score for our method on the Kvasir dataset is lower compared to the other methods mentioned, likely due to the limited number of images (only 100) and the unbalanced nature of the dataset. However, as evidenced by our results on the ETIS-Larib dataset, where a larger number of images were available, our polyp segmentation model performs better when the sample size is increased. This suggests that with a more extensive and balanced dataset, our method has the potential to achieve higher DICE scores comparable to or even surpassing those of the compared methods.

Turning attention to the ETIS dataset, the AdaptUNet model outperformed all other models, with a Dice coefficient of 0.8075 and an IoU score of 0.7215. Furthermore, the high Balanced Accuracy score of 0.9687 indicates exceptional overall accuracy, thereby further confirming the model's effectiveness and reliability. These results collectively underscore the potential of the AdaptUNet model for achieving highly accurate and effective polyp segmentation across diverse datasets. Further enhancements and application of this model may yield significant advancements in polyp detection and diagnosis, thereby enriching the body of research in this field. Fig. 9 showcases the evaluation of the model on the Kvasir dataset providing a concise overview of the model's segmentation capabilities. Fig. 10 displays evaluation results from testing the model on the ETIS dataset.

8. Conclusion

This study introduces AdaptUNet, a pioneering deep learning architecture meticulously tailored for precise detection and segmentation of colorectal polyps in colonoscopy images. Its exceptional performance, demonstrated through the high Dice, IoU scores, and Balanced Accuracy on benchmark datasets, including CVC-300, Kvasir-SEG, and ETIS-LaribDB, underscores its superiority over existing models. The model's seamless integration of attention mechanisms, skip connections, wavelet transformations, and data augmentation techniques delivers a well-rounded solution for efficient and accurate polyp segmentation. The efficiency of AdaptUNet is marked by its accurate segmentation and its ease of implementation, offering a distinct advantage. Compared to existing models, AdaptUNet offers a straightforward implementation process without compromising efficiency or accuracy. A notable feature of this study is the innovative application of wavelet transformation, which ensures comprehensive feature extraction. The model's

performance is further optimized by strategically employing the cyclic learning rate (CLR) and the Dice binary cross-entropy (BCE) loss function. Demonstrating its robustness, AdaptUNet exhibits outstanding performance across various datasets, a testament to its effective generalization capabilities. The meticulous selection and tuning of hyperparameters, as detailed in Tables 2 and 3, contribute to the model's superior results in polyp segmentation. The significance of this research lies in its contribution to the field of automated disease detection, specifically in the context of colorectal polyps. The model's performance and potential for real-world application highlight its importance in advancing colorectal cancer screening and improving patient outcomes. Further research and validation can be conducted to assess its performance in clinical settings and facilitate its integration into medical practice. In conclusion, AdaptUNet presents a promising new horizon for detecting and segmenting colorectal polyps. Its distinctive features and impressive performance serve as a valuable guide for researchers seeking to tailor polyp segmentation tasks to specific requirements and datasets. This research paves the way for future advancements in automated disease detection, emphasizing the necessity for continuous innovation in this critical field.

CRediT authorship contribution statement

Devika Rajasekar: Writing – review & editing, Writing – original draft, Methodology, Data curation. **Girish Theja:** Writing – review & editing, Writing – original draft, Methodology, Data curation. **Manas Ranjan Prusty:** Supervision, Investigation, Conceptualization, Validation, Writing – review & editing. **Suchismita Chinara:** Writing – review & editing, Visualization, Supervision, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Ahmed, Colon cancer: a clinician's perspective in 2019, Gastroenterol. Res. 13 (1) (Feb. 2020) 1–10, https://doi.org/10.14740/gr.v13i1.1239.
[2] A.M. Godkhindi, R.M. Gowda, Automated detection of polyps in CT colonography images using deep learning algorithms in colon cancer diagnosis, in: 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Aug. 2017, pp. 1722–1728, https://doi.org/10.1109/ICECDS.2017.8389744.
[3] J. Petrova, E. Hostalkova, Edge detection in medical image using the Wavelet transform, Rep. Res. Dep. Comput. Control Eng. Czech Public 78 (1) (2011) 1–5.
[4] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, A. Haworth, A review of medical image data augmentation techniques for deep learning applications, J. Med. Imaging Radiat. Oncol. 65 (5) (2021) 545–563, https://doi.org/10.1111/1754-9485.13261.
[5] M. Tharwat, N.A. Sakr, S. El-Sappagh, H. Soliman, K.-S. Kwak, M. Elmogy, Colon cancer diagnosis based on machine learning and deep learning: modalities and analysis techniques, Sensors 22 (23) (Nov. 2022) 9250, https://doi.org/10.3390/s22239250.
[6] T. Collins, et al., Automatic recognition of colon and esophagogastric cancer with machine learning and hyperspectral imaging, Diagnostics 11 (10) (Sep. 2021) 1810, https://doi.org/10.3390/diagnostics11101810.
[7] J. González-Bueno Puyal, et al., Polyp detection on video colonoscopy using a hybrid 2D/3D CNN, Med. Image Anal. 82 (Nov. 2022) 102625, https://doi.org/10.1016/j.media.2022.102625.
[8] J.S. Nisha, V.P. Gopi, P. Palanisamy, Automated colorectal polyp detection based on image enhancement and dual-path CNN architecture, Biomed. Signal Process Control 73 (Mar. 2022) 103465, https://doi.org/10.1016/j.bspc.2021.103465.
[9] M. Yeung, E. Sala, C.-B. Schönlieb, L. Rundo, Focus U-Net: a novel dual attention-gated CNN for polyp segmentation during colonoscopy, Comput. Biol. Med. 137 (Oct. 2021) 104815, https://doi.org/10.1016/j.compbiomed.2021.104815.
[10] P. Sharma, B.K. Balabantaray, K. Bora, S. Mallik, K. Kasugai, Z. Zhao, An ensemble-based deep convolutional neural network for computer-aided polyps identification from colonoscopy, Front. Genet. 13 (Apr. 2022) 844391, https://doi.org/10.3389/fgene.2022.844391.
[11] M.A. Talukder, M.M. Islam, M.A. Uddin, A. Akhter, K.F. Hasan, M.A. Moni, Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning, Expert Syst. Appl. 205 (117695) (2022) 1–37.
[12] C. Ho, et al., A promising deep learning-assistive algorithm for histopathological screening of colorectal cancer, Sci. Rep. 12 (1) (Feb. 2022) 2222, https://doi.org/10.1038/s41598-022-06264-x.
[13] J. Escorcia-Gutierrez, et al., Galactic swarm optimization with deep transfer learning driven colorectal cancer classification for image guided intervention, Comput. Electr. Eng. 104 (Dec. 2022) 108462, https://doi.org/10.1016/j.compeleceng.2022.108462.
[14] M. Murugesan, R. Madonna Arieth, S. Balraj, R. Nirmala, Colon cancer stage detection in colonoscopy images using YOLOv3 MSF deep learning architecture, Biomed. Signal Process Control 80 (Feb. 2023) 104283, https://doi.org/10.1016/j.bspc.2022.104283.
[15] A. Khan, et al., Computer-assisted diagnosis of lymph node metastases in colorectal cancers using transfer learning with an ensemble model, Mod. Pathol. 36 (5) (May 2023) 100118, https://doi.org/10.1016/j.modpat.2023.100118.
[16] B.-L. Chen, J.-J. Wan, T.-Y. Chen, Y.-T. Yu, M. Ji, A self-attention based faster R-CNN for polyp detection from colonoscopy images, Biomed. Signal Process Control 70 (Sep. 2021) 103019, https://doi.org/10.1016/j.bspc.2021.103019.
[17] M.I. Hasan, M.S. Ali, M.H. Rahman, M.K. Islam, Automated detection and characterization of colon cancer with deep convolutional neural networks, J. Healthc. Eng. 2022 (Aug. 2022) e5269913, https://doi.org/10.1155/2022/5269913.
[18] L. Xu, et al., Colorectal cancer detection based on deep learning, J. Pathol. Inf. 11 (1) (Jan. 2020) 28, https://doi.org/10.4103/jpi.jpi_68_19.
[19] S. Tanwar, S. Vijayalakshmi, M. Sabharwal, M. Kaur, A.A. AlZubi, H.-N. Lee, Detection and classification of colorectal polyp using deep learning, BioMed Res. Int. 2022 (Apr. 2022) e2805607, https://doi.org/10.1155/2022/2805607.
[20] S. Hosseinzadeh Kassani, P. Hosseinzadeh Kassani, M.J. Wesolowski, K.A. Schneider, R. Deters, Deep transfer learning based model for colorectal cancer histopathology segmentation: a comparative study of deep pre-trained models, Int. J. Med. Inf. 159 (Mar. 2022) 104669, https://doi.org/10.1016/j.ijmedinf.2021.104669.
[21] J.K. Zhang, Y.R. He, N. Sobh, G. Popescu, Label-free colorectal cancer screening using deep learning and spatial light interference microscopy (SLIM), APL Photonics 5 (4) (Apr. 2020) 40805, https://doi.org/10.1063/5.0004723.
[22] M. Billah, S. Waheed, M.M. Rahman, An automatic gastrointestinal polyp detection system in video endoscopy using fusion of color wavelet and convolutional neural network features, Int. J. Biomed. Imag. (2017) 2017.
[23] P. Song, J. Li, H. Fan, Attention based multi-scale parallel network for polyp segmentation, Comput. Biol. Med. 146 (2022) 105476.

[24] D. Vázquez, et al., A benchmark for endoluminal scene segmentation of colonoscopy images, J. Healthc. Eng. 2017 (Jul. 2017) e4037190, https://doi.org/10.1155/2017/4037190.
[25] J. Silva, A. Histace, O. Romain, X. Dray, B. Granado, Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer, Int. J. Comput. Assist. Radiol. Surg. 9 (2) (Mar. 2014) 283–293, https://doi.org/10.1007/s11548-013-0926-3.
[26] D. Jha, et al., Kvasir-SEG: a segmented polyp dataset, in: MultiMedia Modeling, Y.M. Ro, W.-H. Cheng, J. Kim, W.-T. Chu, P. Cui, J.-W. Choi, M.-C. Hu, W. De Neve (Eds.), Lecture Notes in Computer Science, Springer International Publishing, Cham, 2020, pp. 451–462, https://doi.org/10.1007/978-3-030-37734-2_37.
[27] J. Bernal, J. Sánchez, F. Vilariño, Towards automatic polyp detection with a polyp appearance model, Pattern Recogn. 45 (9) (Sep. 2012) 3166–3182, https://doi.org/10.1016/j.patcog.2012.03.002.
[28] P.M.K. Prasad, D.Y.V. Prasad, G.S. Rao, Performance analysis of orthogonal and biorthogonal wavelets for edge detection of X-ray images, Procedia Comput. Sci. 87 (2016) 116–121.
[29] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science, 9351, Springer International Publishing, Cham, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.
[30] Y. Fang, C. Chen, Y. Yuan, K. Tong, Selective feature aggregation network with area-boundary constraints for polyp segmentation', in medical image computing and computer assisted intervention – miccai 2019, in: D. Shen, T. Liu, T.M. Peters, L.H. Staib, C. Essert, S. Zhou, P.-T. Yap, A. Khan (Eds.), Lecture Notes in Computer Science, Springer International Publishing, Cham, 2019, pp. 302–310, https://doi.org/10.1007/978-3-030-32239-7_34.
[31] X. Fang, P. Yan, Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction, IEEE Trans. Med. Imag. 39 (11) (Nov. 2020) 3619–3629, https://doi.org/10.1109/TMI.2020.3001036.
[32] T. Kim, H. Lee, D. Kim, UACANet: uncertainty augmented context attention for polyp segmentation, in: Proceedings of the 29th ACM International Conference on Multimedia, Oct. 2021, pp. 2167–2175, https://doi.org/10.1145/3474085.3475375.
[33] J. Wei, Y. Hu, R. Zhang, Z. Li, S.K. Zhou, S. Cui, Shallow attention network for polyp segmentation', in medical image computing and computer assisted intervention – miccai 2021, in: M. de Bruijne, P.C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, C. Essert (Eds.), Lecture Notes in Computer Science, Springer International Publishing, Cham, 2021, pp. 699–708, https://doi.org/10.1007/978-3-030-87193-2_66.
[34] X. Zhao, L. Zhang, H. Lu, Automatic polyp segmentation via multi-scale subtraction network', in medical image computing and computer assisted intervention – miccai 2021, in: M. de Bruijne, P.C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, C. Essert (Eds.), Lecture Notes in Computer Science, Springer International Publishing, Cham, 2021, pp. 120–130, https://doi.org/10.1007/978-3-030-87193-2_12.
[35] K. Patel, A.M. Bur, G. Wang, Enhanced U-net: a feature enhancement network for polyp segmentation, in: 2021 18th Conference on Robots and Vision (CRV), May 2021, pp. 181–188, https://doi.org/10.1109/CRV52889.2021.00032.
[36] Z. Qiu, Z. Wang, M. Zhang, Z. Xu, J. Fan, L. Xu, BDG-net: boundary distribution guided network for accurate polyp segmentation, in: Medical Imaging 2022: Image Processing, Apr. 2022, p. 105, https://doi.org/10.1117/12.2606785.
[37] F. Liu, Z. Hua, J. Li, L. Fan, MFBGR: multi-scale feature boundary graph reasoning network for polyp segmentation, Eng. Appl. Artif. Intell. 123 (Aug. 2023) 106213, https://doi.org/10.1016/j.engappai.2023.106213.