IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 February 2023, accepted 10 February 2023, date of publication 14 February 2023, date of current version 22 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3244789

RESEARCH ARTICLE

ConvSegNet: Automated Polyp Segmentation From Colonoscopy Using Context Feature Refinement With Multiple Convolutional Kernel Sizes

AYOKUNLE OLALEKAN IGE[1,3], NIKHIL KUMAR TOMAR[2], FELIX OLA ARANUWA[3], OLUWAFEMI ORIOLA[3], ALABA O. AKINGBESOTE[3], MOHD HALIM MOHD NOOR[1], MANUEL MAZZARA[4], AND BENJAMIN SEGUN ARIBISALA[5,6]

[1]School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang 11800, Malaysia
[2]School of Computer and Information Sciences (SOCIS), Indira Gandhi National Open University, New Delhi 110068, India
[3]Department of Computer Science, Adekunle Ajasin University, Akungba-Akoko, Ondo, Nigeria
[4]Institute of Software Development and Engineering, Innopolis University, 420500 Innopolis, Russia
[5]Department of Computer Science, Lagos State University, Lagos 102101, Nigeria
[6]Department of Medicine, The University of Chicago, Chicago, IL 60637, USA

Corresponding author: Ayokunle Olalekan Ige (ayo.ige@aaua.edu.ng)

ABSTRACT Colorectal cancer occurs in the rectal of humans, and early detection has been proved to reduce its mortality rate. Colonoscopy is the standard used in detecting the presence of polyps in the rectal, and accurate segmentation of the polyps from colonoscopy images often provides helpful information for early diagnosis and treatment. Although existing deep learning models often achieve high segmentation performance when tested on the same dataset used in model training; still, their performance often degrades when applied to out-of-distribution datasets, leading to low model generalization or overfitting. This challenge is often associated with the quality of the features learnt from the input images. In this work, a novel Context Feature Refinement (CFR) module is proposed to address the challenge of low model generalization and segmentation performance. The CFR module is built to extract contextual information from the incoming feature map by using multiple parallel convolutional layers with progressively increasing kernel sizes. Using multiple parallel convolutions with different kernel sizes helped to extract more efficient multi-scale contextual information and thus enabled the network to effectively identify and segment small and fine details, as well as larger and more complex structures in the input images. Extensive experiments on three public benchmark datasets in CVC-ClinicDB, Kvasir-SEG, and BKAI-NeoPolyp showed that the proposed ConvSegNet model achieved jaccard, dice and F2 scores of 0.8650, 0.9177, and 0.9328 on CVC-ClinicDB, 0.7936, 0.8618, and 0.8855 on Kvasir-SEG, and 0.8045, 0.8747 and 0.8909 on BKAI-NeoPolyp datasets respectively. Also, an improved generalization performance was achieved by the ConvSegNet model, compared to the benchmark polyp segmentation models. Code is available at https://github.com/AOige/ConvSegNet.

INDEX TERMS Biomedical, colonoscopy, image, kernel, polyp, segmentation.

I. INTRODUCTION

Image segmentation is an essential task in biomedical imaging, and it has seen its application across various clinical areas. Image segmentation is a technique for splitting images into easily analyzable and interpretable Regions of Interest (ROI). In recent times, deep neural networks, especially convolutional neural network (CNN), have improved image segmentation compared to shallow networks [1], [2]. This improvement has seen its applicability in numerous segmentation areas, such as brain tumors [3], skin cancers [4], covid-19 [5], and lung cancer [6], among other areas.

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

Recently, researchers have intensified their focus on polyp segmentation from colonoscopy due to the mortality tendency of colorectal cancer [7], [8]. Colorectal cancer is the second most common cancer type among women and the third most common among men [9]. Generally, polyps indicate the presence of colorectal cancers in the rectum, and early detection and removal are essential to mitigate mortality. Polyps are abnormal tissues generated from the mucus membrane, and they have been discovered to be present in 50% of individuals that undergo colonoscopy screening, and the frequency often increases as age increases [10]. However, detecting polyps from colonoscopy manually is quite laborious, and the miss rate is between 14% to 30% when trying to detect the presence of polyps in the rectal manually, with the type and size being the determining factor [11].

In most cases, polyps may be hidden from the line of vision during manual inspection and, sometimes, might be present in the operator's range of view, but remain undetectable [12]. These challenges have prompted the development of real-time Artificial Intelligence (AI) algorithm, as seen in [13]. The polyp segmentation technique in this scenario strives to accurately delineate the polyp border from the surrounding mucosa and detect polyps. Also, various forms of noises such as shadow, blurriness, reflection, and others can be present in colonoscopy, which can also affect the detection of the presence of polyps [14]. Recently, several deep learning models have been proposed to effectively extract cogent features to aid the segmentation of polyps from colonoscopy [15], [16], [17], [18]. However, a general limitation of image segmentation models is the quality of features extracted from input images, and the low segmentation performance achieved when tested on out-of-distribution datasets, which leads to low model generalization [19]. Due to the varying size and types of polyps, several models have been proposed to address the issues of low-quality feature extraction and low model generalization in polyp segmentation when tested on new colonoscopy images. However, it remains a challenging area in polyps' segmentation from colonoscopy. In this work, we propose ConvSegNet, an image segmentation model which uses a novel Context Feature Refinement (CFR) module to address low model generalization and segmentation performance. The novel CFR module is built to extract quality features by applying multiple parallel convolutional layers with different kernel sizes in the decoder block. This unique structure enables the network to effectively capture multi-scale context features. This is important, as it will enable the network to effectively identify and segment small and fine details, as well as larger and more complex structures in the image. This is crucial for achieving high-quality segmentation results.

Specifically, our main contributions are in four folds:

- The CFR module leverages progressively increasing kernel sizes to extract contextual information from feature maps.
- The proposed model improved the quality of extracted features, and is efficient in terms of speed and size, as it achieved improved segmentation performance with few parameters and standard Frames Per Second.
- Comprehensive experiments were done on three datasets to evaluate the ConvSegNet model against other benchmark polyp segmentation methods using six standard performance metrics.
- Lastly, we have expanded the standard benchmarks for polyp segmentation, which can be used to create clinically useful procedures.

The rest of this paper is organized as follows. Section II discusses the state-of-the-art polyp and biomedical image segmentation models. Section III describes the architecture of the proposed ConvSegNet model, Section IV describes the datasets, performance evaluation metrics, and experimental results, and Section V concludes.

II. RELATED WORKS

Various traditional methods have been adopted for polyp segmentation. For example, Hwang et al. [20] presented a polyp detection approach based on the elliptical shape of virtually all small polyps. Segmentation was done based on watersheds image segmentation and ellipse fitting method, then matching curvature and contour distance were used to separate the ellipses of the polyp and non-polyp zones. Ameling et al. [21] considered textural cues and local binary patterns for polyp segmentation. In [22], these two methods were combined by considering shape and texture for polyp segmentation. The shape feature was utilized to accurately identify polyps with curving borders, while the texture information was used to discriminate polyps from non-polyp structures. Also, various Machine Learning (ML) models have been proposed, as seen in [23], [24], and [25], among others. As shown in Figure 1, ML techniques involve data pre-processing, handcrafted feature extraction, and feature selection before the classification phase. In contrast, deep learning models ignore most of these phases and achieve better results. The introduction of deep learning models, especially CNN, has been prompted by several limitations of machine learning models, including issues with automated segmentation of biomedical images, considerable changes in form, size, texture, and in some cases, the colour of ROI between patients, and poor contrast between areas [26], [27].

CNN architectures have improved semantic image segmentation, with most of the existing architectures based on U-Net [28], a modified architecture developed for biomedical image segmentation. The U-Net comprises an encoding network that captures image context and a symmetrical decoding network that allows the localization of salient regions. Several other models have been proposed based on the U-Net architecture. UNet++ [29] expands the U-Net by including skip connections to close the semantic gap between the encoder and decoder's feature maps before fusion. In [30], ResUNet architecture was proposed based on a semantic segmentation

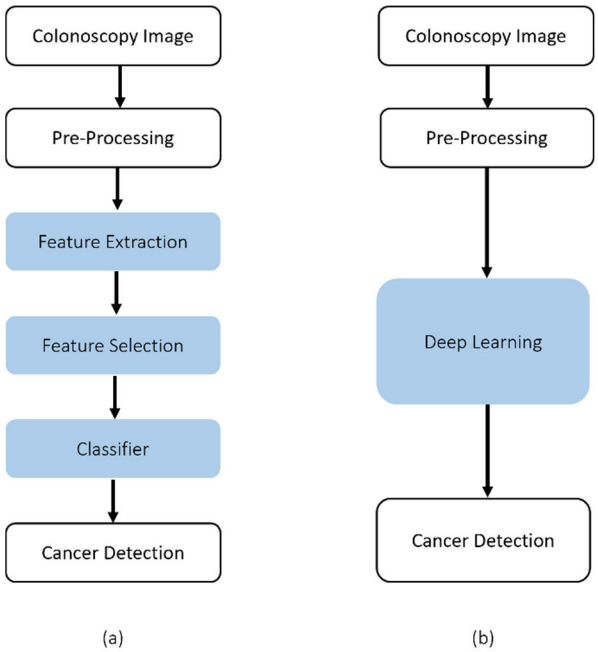


FIGURE 1. Classifier approach (a) Machine learning (b) Deep learning.

neural network. It integrated U-Net with residual neural network strengths. This combination allowed the residual unit to facilitate network training. The skip connections within the residual unit and between low and high levels of the network facilitated information propagation without degradation. This allowed the architecture to be designed with few parameters while still achieving comparable semantic segmentation performance. Based on these architectures, various image segmentation models have been proposed. Polyps come in a variety of shapes and sizes.

A small colorectal polyp may lack distinguishing textural characteristics in its early stages, making it easy to confuse with normal intestinal tissue. Therefore, some biomedical image segmentation models might perform well on other biomedical image datasets but perform below par when applied to polyp segmentation, as seen in [16], [31], [32], [33], and [34]. For this reason, more polyp specific segmentation architectures are being proposed. The following section presents some existing state-of-the-art polyp segmentation from colonoscopy architectures.

A. POLYP SEGMENTATION ARCHITECTURES

In Jha et al. [35], the ResUNet++ model was developed to integrate residual units with the Atrous Spatial Pyramidal Pooling (ASPP) and squeeze-and-excitation block based on channel attention. Yeung et al. [12] proposed Focus U-Net; a dual-attention gated model that combined spatial and channel-based attention and used the hybrid focal loss to address class imbalance in polyp's datasets. The model was tested on some polyps benchmarking datasets and evaluated against U-Net and Attention U-Net [36], and the model showed improved performance. However, the model ignored computational and generalization efficiency and focused solely on segmentation performance. In Kim et al. [37], they presented a U-Net-based architecture with extra encoder and decoder modules called UACA-Net. Foreground, background, and uncertain region maps are calculated for each representation using saliency maps computed by a prediction module in the UACA-Net. The next prediction module computes the relationship between each representation and employs it.

In Mahmud et al. [38], PolypSegNet architecture was proposed based on encoder-decoder architecture. The aggregate feature into each unit layer included several successive depths dilated inception blocks. Rather than connecting different levels of encoder and decoder separately, different scales of contextual information from all encoder unit layers were fed through the PolypSegNet's deep fusion skip module to generate skip interconnection with each decoder layer. This addressed computational efficiency. However, the generalization performance of the model is quite low, due to the plain skip interconnection. This is because the plain skip connections tend to combine semantically diverse low- and high-level convolutional features, resulting in hazy feature maps.

Zunair and Hamza [32], presented Sharp U-Net without plain skip connections. Before merging the encoder and decoder features, a depth wise convolution of the encoder feature map with a sharpening kernel filter was used instead of the simple skip connection. They were able to create a sharpened intermediate feature map of the same size as the encoder map. The model was also able to smooth out artefacts throughout the network layers during the early training phases by applying the sharpening filter layer. Experiments were done on polyp datasets, covid-19 datasets, lung, and three other datasets. Even though the model achieved higher Jaccard and Dice Scores on the five other datasets, results on polyp segmentation from colonoscopy were relatively low, as the Jaccard was 83.98% and Dice was 90.05%. The low performance can be attributed to the varying sizes and shapes of polyps that the sharpened kernel filter might ignore due to the semantic gap between the encoder and decoder features.

In Zhao et al. [39], MSNet architecture was proposed to segment polyps from colonoscopy images. They combined lower-order and higher-order cross-level complementary information with level-specific information to increase multi-scale feature representation by pyramidally concatenating numerous subtraction units. Even though the model achieved high segmentation performance when trained and tested on the same polyp dataset, the model generalization performance was low, and the number of parameters was relatively high. Also, in [40], a Context Extractor Module was proposed, which consists of DAC block and the RMP block. The DAC block utilized three different dilated convolutions with a fixed 3×3 kernel size, while the RMP block used a multiple pooling strategy with different pooling windows $[2 \times 2, 3 \times 3, 5 \times 5, 6 \times 6]$ and then performed upsampling to have equal spatial dimensions for concatenation. However, by using this approach, positional information is

lost, which automatically influences the quality of features learnt.

Some methods based on inception module have also been proposed to increase the segmentation of polyps from colonoscopy. For example, Qadir et al. [41], used two ensemble models based on inception, which was benchmarked on the CVC-ColonDB dataset [22], and the model achieved a recall of 72.59%, precision of 80%, Jaccard of 61.24%, and a Dice score of 70.42%. However, the architecture of the nception module uses 1×1, 3×3, 5×5, and some pooling operations, which also caused positional information loss over a broad range of features from the input, thereby affecting the segmentation performance and generalization performance of such models. Tomar et al. [42] proposed FANet, for polyp segmentation from colonoscopy. The FANet which is an attention feedback model that unifies past epoch mask with the feature map of the current training epoch, which then uses the mask of the previous epoch to provide hard attention to the learned feature maps at several convolutional layers. Even though the model achieved high performance on some other datasets, results on polyp datasets were not SOTA, due to the maxpooling which was done on the input mask before scaling. Also, the FANet model is parameter heavy.

In a bid to develop lightweight segmentation models, Valanarasu and Patel [43], proposed UNeXt, a convolutional multilayer perceptron based network for image segmentation. The model was designed with an early convolutional stage and a MLP in the latent stage. Experiments showed that the UNeXt model was able to improve segmentation performance with minimal model parameters. Also, Li-SegPNet was proposed in [44]. The model utilized a unique encoder block with modified triplet attention to harness cross-dimensional interaction in feature maps. To solve the issue of segmenting objects at various sizes, the authors employed spatial pyramid pooling, and used an attention gating-based modified skip connection to overcome the semantic discrepancy between the encoder and decoder. The model was evaluated on CVC-ClinicDB and Kvasir-SEG, and the result showed that the model performed best on medium-sized polyps, with below par performance on smaller polyps. This limitation can be attributed to the pooling operations in their architecture [38].

In this paper, we propose the ConvSegNet model which uses progressively increasing kernel sizes starting from 1×1, 3×3, 7×7 and 11×11 in the CFR module, without pooling operations. By doing this, a broader range of features can be extracted progressively from the input, which can help to capture more discriminative features. The novelty of the proposed ConvSegNet model lies in the Context Feature Refinement (CFR) module used in the decoder block. The parallel architecture of the CFR module, consisting of four parallel convolutional layers with progressively increasing kernel sizes. This unique structure will enable the network to effectively capture multi-scale context features. A detailed description of the proposed ConvSegNet model is presented in the next section.

III. METHODOLOGY

To address the challenges of the existing architectures, we propose a novel Context Feature Refinement (CFR) module to extract contextual information from the incoming feature map by applying multiple parallel convolutional layers with different kernel sizes. This section presents the data processing method, model block diagram, and the architecture of the proposed ConvSegNet model.

A. DATA PRE-PROCESSING

The images and masks were resized to 256 x 256 pixels, followed by the pixel value normalization. Data augmentation techniques such as random rotation, horizontal flipping, vertical flipping, and coarse dropout were used to improve the robustness of the input data.

B. CONTEXT FEATURE REFINEMENT MODULE

The Context Feature Refinement (CFR) module in the decoder block is shown in Figure 2. The CFR module is built to extract contextual information from the incoming feature map by applying multiple parallel convolutional layers with progressively increasing kernel sizes, as shown in Equation 1, where feature map O_x is given as:

$$O_x = b_x + \sum_r F_{xr} * I_r \tag{1}$$

where F_{xr} is the convolutional kernel, I_r is the input, b_x is the bias term, and $*$ is the convolutional operation. Then we concatenate the output of these layers and pass them through a 1×1 convolution to refine these features. The CFR module begins with four parallel convolutional layers with 1×1, 3×3, 7×7 and 11×11 as their respective kernel sizes, as shown in Figure 2.

Using different kernel sizes helped increase the receptive field during the convolution operation, which helps to better capture the contextual feature from the input feature map. Zero padding was used in our module to ensure that all the feature maps have the same spatial dimensions, allowing easy concatenation into a single feature map. After that, each convolutional layer is followed by batch normalization and a ReLU activation function, which is given in Equation N.

$$f(x) = \max(0, I_x) \tag{2}$$

Next, we concatenate the output of the four ReLU layers along the channel axis and pass them through a 1×1 convolutional layer which is again followed by batch normalization and ReLU layer.

C. CONVSEGNET

Our model uses the novel Context Feature Refinement (CFR) module to extract contextual information from the incoming feature map. The proposed architecture is fed with an RGB image passed to the encoder, consisting of a pre-trained ResNet50. The ResNet50 is used to extract different level features from different blocks with varying resolutions. Each feature map is then passed into a 3×3 convolutional layer

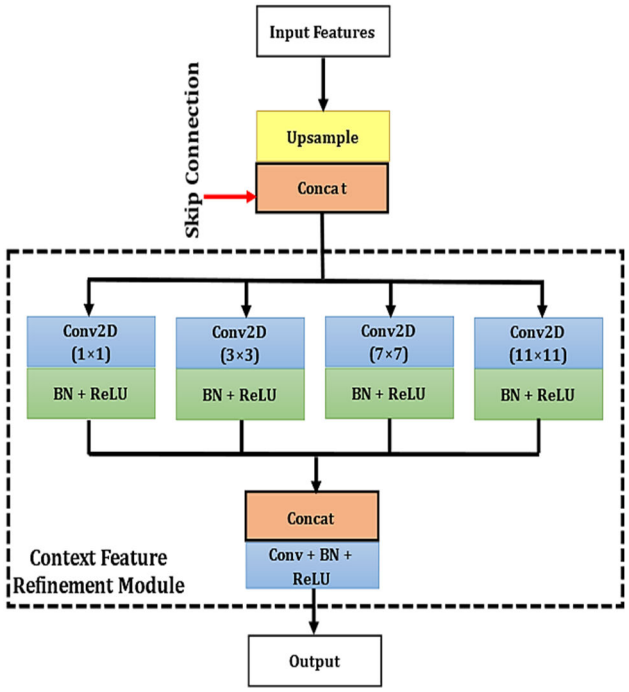


FIGURE 2. Decoder block with context feature refinement module.

to reduce the number of feature channels to 64. The convolutional layer is further followed by a batch normalization layer and ReLU activation function. Next, the network is followed by four decoder blocks, each taking the previous feature maps as the main input and a skip connection (indicated by a red colour arrow in Figure 2 and Figure 3.

Each decoder block begins with an upsampling layer where the spatial dimensions (height and width) of the incoming feature map are increased by a factor of two using bilinear interpolation method. Next, it is followed by a concatenation of upsampled feature map with the feature map from the skip connection. Using these skip connections helped to provide additional information to the decoder, to generate better semantic features, while providing additional paths for the better flow of gradients during the backpropagation.

The concatenated feature maps are then passed through the novel context feature refinement module, which uses four convolutional layers with varying kernel sizes to extract contextual information from the input feature. Next, we concatenate the contextual information and pass it through a convolutional layer for refinement. The refined feature acts as the output of the decoder block, which is further passed to the next set of decoders. In the last decoder, we used the low-level features from the input image, then passed the input image through a convolutional layer, then used it as the skip connection. By doing this, we were able to take advantage of the low-level features to generate high-quality semantic features. The output from the last decoder block is then passed through a 1×1 convolutional layer followed by a sigmoid activation function which generates a binary segmentation mask.

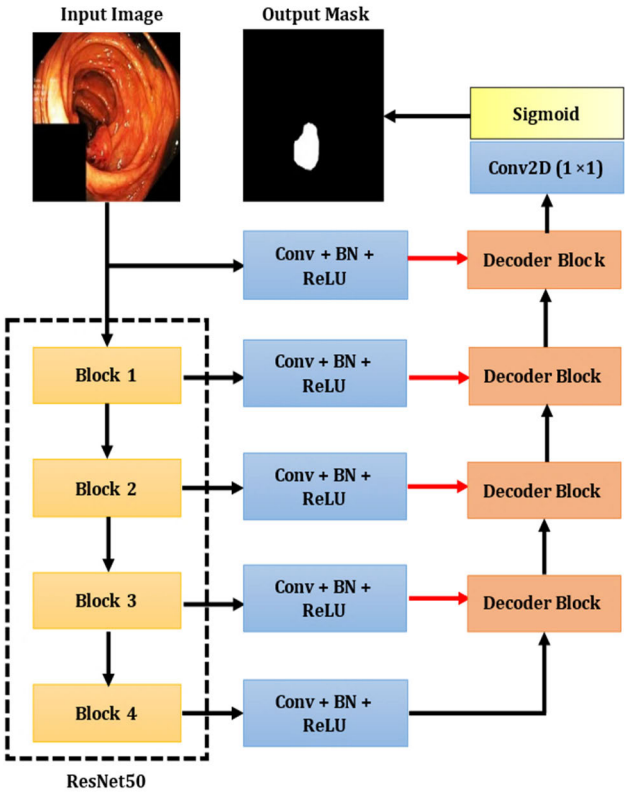


FIGURE 3. Architecture of the ConvSegNet.

IV. EXPERIMENTAL RESULTS

This section presents the dataset description, the details of our implementation, and the standard performance metrics considered in the evaluation of the proposed model against benchmarking architectures. We then presented a detailed comparison of the performance of the proposed model based on quantitative experiments and generalization experiments. Also, ablation studies were presented, and experiments on two other non-polyp datasets, to demonstrate the extensibility of the proposed ConvSegNet model.

A. DATASETS

According to the literature, there are six (6) publicly available datasets that have been used for polyp segmentation from colonoscopy model benchmarking. Kvasir-SEG [45], CVC-ClinicDB [46], ETIS-Larib [44], CVC-ColonDB [22], BKAI-NeoPolyp [47] and CVC-300 [48]. Out of these six, only CVC-ClinicDB, Kvasir-SEG, BKAI-NeoPolyp and ETIS-Larib contain manually labelled ground truth masks. Among these four, studies have shown that Kvasir-SEG and CVC-ClinicDB are the most used datasets for fair generalization evaluation since they are both in standard definition. While Etis-Larib is in high definition and has only 196 images, BKAI-NeoPolyp has 1200 images. For this reason, we have chosen the Kvasir-SEG with 1000 images, CVC-ClinicDB with 612 images, and BKAI-NeoPolyp with 1200 images as the benchmark datasets to evaluate the proposed ConvSegNet model.

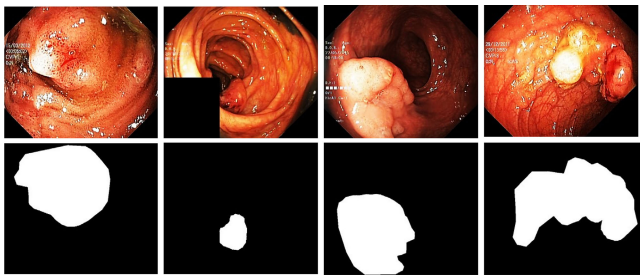


FIGURE 4. Sample images with ground truth in Kvasir-SEG dataset.

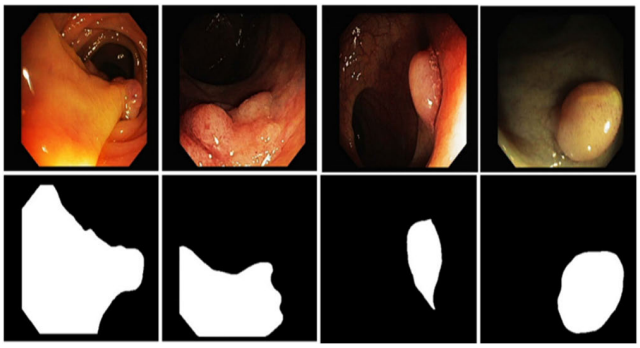


FIGURE 5. Sample images with ground truth mask in CVC-ColonDB dataset.

1) KVASIR-SEG DATASET

This dataset was extracted from the polyp class in the Kvasir dataset [49]. Kvasir-SEG contains 1000 polyp images, their accompanying masks, and bounding box information taken by electromagnetic imaging devices. The segmentation task can be done with the images and their ground truths, whereas the detection task can be done with the bounding box information. The images in this dataset range in resolution from 332×487 to 1920×1072 pixels. Samples of the images and the annotated masks from this dataset are shown in Figure 4.

2) CVC-ClinicDB

The CVC-ClinicDB dataset is an open-access dataset consisting of 612 images with a resolution of 384×288 from colonoscopy sequences. Samples of the images and the annotated masks from this dataset are shown in Figure 5.

3) BKAI-NEOPOLYP

The BioKinesiology Association of Ireland-NeoPolyp dataset consists of 1200 polyp images, with 1000 images for training and 200 for testing. Samples of the dataset and the ground truth mask is shown in Figure 6.

B. IMPLEMENTATION DETAILS

For a fair comparison, the state-of-the-art benchmark architectures and the proposed ConvSegNet were implemented using the PyTorch framework and trained on RTX 3090. To train the model for polyp segmentation, we selected three polyp datasets: Kavsir-SEG, CVC-ClinicDB, and BKAI-NEOPolyp. First, we split the dataset for proper training. For the Kvasir-SEG, we followed the official split of 880/120. Here 880 images and their masks were used for training the model, while the remaining 120 images and their masks were used for testing. In the case of CVC-ClinicDB, we split the dataset with the ratio of 80-10-10, where 80% of images and masks were used for training, and the rest were used for validation and testing, and in BKAI-NEOPOLYP, 1000 images were used for training and 200 for testing. For a fair comparison, we have used the same set of hyperparameters to train all the models; we have used the Adam optimizer with a learning rate of 1e-4 (0.0001). An epoch of 200 was set, and early stopping mechanism was used to stop the training once the model stops improving. A combination of dice loss and binary cross-entropy were used as the loss function, with a batch size of 16.

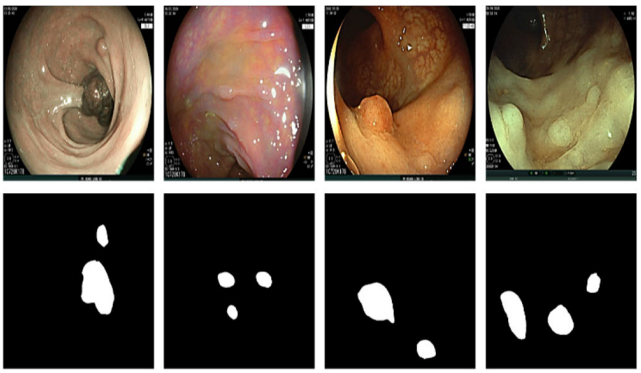


FIGURE 6. Sample images with ground truth mask in BKAI-NeoPolyp dataset.

C. PERFORMANCE METRICS

For model evaluation, six (6) standard metrics were used to compare the performance of the proposed ConvSegNet model to the existing state-of-the-art. Jaccard, Dice score, Recall, Precision, Accuracy and F2-Measure were considered.

The Jaccard index (JI, Equation 3), is the ratio of the overlapping area between the predicted and ground truth to the area of union between the predicted and ground truth segmentation, where S denote segmentation.

$$Jaccard = \frac{S_{Groundtruth} \cap S_{Automated}}{S_{GroundTruth} \cup S_{Automated}} = \frac{TP}{TP + FP + FN} \tag{3}$$

The Dice Score (DSC, Equation 4), also called F1 measure, measures the boundary matching between predicted and ground truth segmentation, as shown in equation 4.

$$DSC = \frac{2 \times TP}{2 \times TP + FN} \tag{4}$$

As shown in equation 5, Precision is important in biomedical segmentation because it considers the ratio of the correctly predicted disease pixels to the total number of ground truth pixels.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

Equation 6 shows the Recall, which considers the ratio of disease pixels in the ground truth that the segmentation model

TABLE 1. Comparison of our model with benchmark models on CVC-ClinicDB.

Model	Jaccard	Dice	Recall	Precision	Accuracy	F2
U-Net	0.8428	0.8978	0.9001	0.9209	0.9861	0.8981
ResU-Net	0.7892	0.8648	0.8836	0.8804	0.9793	0.8722
U-Net++	0.8337	0.8913	0.9129	0.8988	0.9859	0.9026
HardDNet-MSEG	0.8388	0.8967	0.8929	**0.9216**	0.9871	0.8938
FANet	0.7958	0.8625	0.8570	0.9151	0.9772	0.8569
UNeXt	0.6676	0.7673	0.7546	0.8617	0.9722	0.7563
ConvSegNet	**0.8650**	**0.9177**	**0.9518**	0.9048	**0.9881**	**0.9328**

is able to segment correctly.

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

As shown in equation 7, the model's accuracy considers the percentage of the image pixels that are correctly classified.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + FP} \tag{7}$$

The $F2_{measure}$ as shown in equation 8 is more focused on the recall than precision, and it is suitable when it is more important to classify correctly as many positive samples as possible.

$$F2_{measure} = \frac{5 \times Precision \times Recall}{4 \times Precision + Recall} \tag{8}$$

D. EXPERIMENTS

This section presents the results of the experiments conducted on the proposed ConvSegNet model and the existing methods. For a fair comparison, we considered six standard state-of-the-art deep learning architectures as benchmark models in U-Net [28], ResU-Net [30], U-Net++ [29], HardDNet-MSEG [50], FANet [42] and UNeXt [43]. We trained all models with the same hyperparameters and hardware. We performed quantitative and generalization experiments to evaluate the performance of the proposed ConvSegNet model against the benchmark architectures.

The quantitative experiments focused on training and testing on the same dataset, while the generalization experiments considered training on one dataset and testing on another. Also, recent state-of-the-art polyp segmentation models were used to evaluate the generalization performance of our ConvSegNet model.

1) QUANTITATIVE RESULTS

The results from training and testing of CVC-ClinicDB dataset on U-Net, ResU-Net, U-Net++, HardDNet-MSEG, FANet and UNeXt and the proposed ConvSegNet are presented in Table 1. The proposed ConvSegNet model achieved a Jaccard score of 0.8650, which outperformed the U-Net, ResU-Net, U-Net++, HardDNet-MSEG, FANet and UNeXt architectures which recorded 0.8428, 0.7892, 0.8337, and 0.8388 Jaccard scores, respectively. Likewise, a much better Dice score of 0.9177 was achieved by the ConvSegNet model.

Recall of 0.9518 was also achieved by ConvSegNet, which was better than the benchmark architectures, with U-Net++ achieving the closest recall at 0.9129, proving an improvement of 0.0389 recall score over existing architectures. A Precision of 0.9216 was achieved on the CVC-ClinicDB dataset by the HardDNet-MSEG benchmark, as against the proposed ConvSegNet model by a difference of 0.0168. This is because the foreground (positive) regions are similar to the negative regions in the whole colonoscopy images. Hence, the benchmark models were exhibiting asymmetric errors by having more false positives than false negatives. However, a state-of-the-art accuracy and F2-measure of 0.9881 and 0.9328, respectively, were achieved by the proposed ConvSegNet model, which outperformed the U-Net by 0.002 and 0.0347, ResU-Net by 0.0088 and 0.0606, U-Net++ by 0.0022 and 0.0302, HardDNet-MSEG by 0.001 and 0.039 respectively. Visual representation of the input colonoscopy images, and the segmented polyp regions obtained using the benchmark architecture and ConvSegNet on CVC-ClinicDB is presented in Figure 7.

Results on the Kvasir-SEG dataset is presented in Table 2. Similar to the results on the CVC-ClinicDB dataset, the proposed ConvSegNet model achieved a better Jaccard score of 0.7936 compared to the U-Net, ResU-Net, U-Net++, HardDNet-MSEG, FANet and UNeXt architectures, which recorded 0.7472, 0.6634, 0.7419, 0.7459, 0.6941 and 0.6284 respectively. The dice score, achieved by the ConvSegNet model also outperformed the benchmark architectures by a difference of 0.0354, 0.0976, 0.039, 0.0358, 0.0803 and 0.1300 respectively. Recall showed that ConvSegNet also outperformed the benchmarks with a recall of 0.9124 compared to 0.8504 on U-Net, 0.8025 on ResU-Net, 0.8437 on U-Net++, 0.8485 on HardDNet-MSEG, 0.8452 on FANet and 0.7840 on UNeXt.

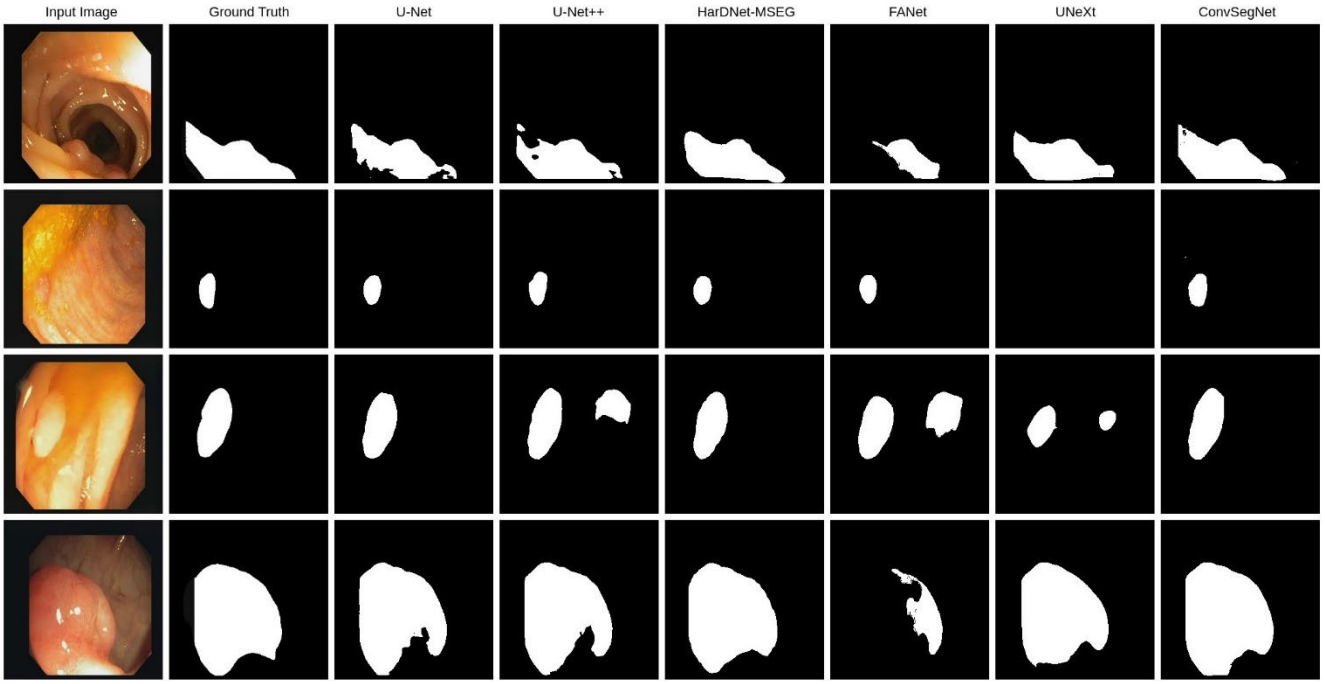


FIGURE 7. Visual representation of the input colonoscopy images, and the segmented polyp regions obtained using the benchmark architectures and ConvSegNet on CVC-ClinicDB.

TABLE 2. Comparison of our model with benchmark models on Kvasir-SEG.

Model	Jaccard	Dice	Recall	Precision	Accuracy	F2
U-Net	0.7472	0.8264	0.8504	**0.8703**	0.9510	0.8353
ResU-Net	0.6634	0.7642	0.8025	0.8200	0.9341	0.7740
U-Net++	0.7419	0.8228	0.8437	0.8607	0.9491	0.8295
HardDNet-MSEG	0.7459	0.8260	0.8485	0.8652	0.9492	0.8358
FANet	0.6941	0.7815	0.8452	0.8159	0.9220	0.8002
UNeXt	0.6284	0.7318	0.7840	0.7656	0.9208	0.7507
ConvSegNet	**0.7936**	**0.8618**	**0.9124**	0.8692	**0.9617**	**0.8855**

Except for U-Net architecture which had a precision score of 0.8703, due to the foreground positive regions that are similar to the negative regions in the colonoscopy images, the ConvSegNet performed better than ResU-Net, U-Net++ and HardDNet-MSEG. ConvSegNet also outperformed the four benchmark architectures in terms of Accuracy and F2-measure with an improvement of 0.0107 and 0.0502 above U-Net, 0.0276 and 0.1115 above ResU-Net, 0.0126 and 0.056 above U-Net++, 0.0125 and 0.8855 above HardDNet-MSEG, 0.0397 and 0.0853 above FANet and 0.0409 and 0.1348 above UNeXt architectures, respectively. Visual representation of the input colonoscopy images, and the segmented polyp regions obtained using the benchmark architectures and ConvSegNet on Kvasir-SEG is presented in Figure 8.

Table 3 shows the result of the experiment on BKAI-NeoPolyp dataset. As shown, the ConvSegNet outperformed the benchmarks by achieving a Jaccard of 0.8045, dice score of 0.8747, recall of 0.9068, accuracy of 0.9922, and F2 of 0.8909. This outperformed the U-Net by 0.0446, 0.0461, 0.0773, 0.0019, and 0.0645. However, the U-Net architecture had the highest precision at 0.8999. The results on ResU-Net had a Jaccard of 0.6589, 0.7433 dice, 0.7447 recall, 0.871 precision, 0.9843 accuracy and 0.7387 F2. U-Net++ had a performance of 0.7563 Jaccard, 0.8275 dice, 0.8388 recall, 0.8942 precision, 0.9895 accuracy and an F2 measure of 0.8308. HardDNet-MSEG was also outperformed by the ConvSegNet at 0.6734 jaccard, 0.8305 dice and 0.7528 F2. Also, FANet had 0.7578 jaccard, 0.8305 dice and 0.7528 F2, while UNeXt had 0.4680 jaccard, 0.5622 dice and 0.5692 F2

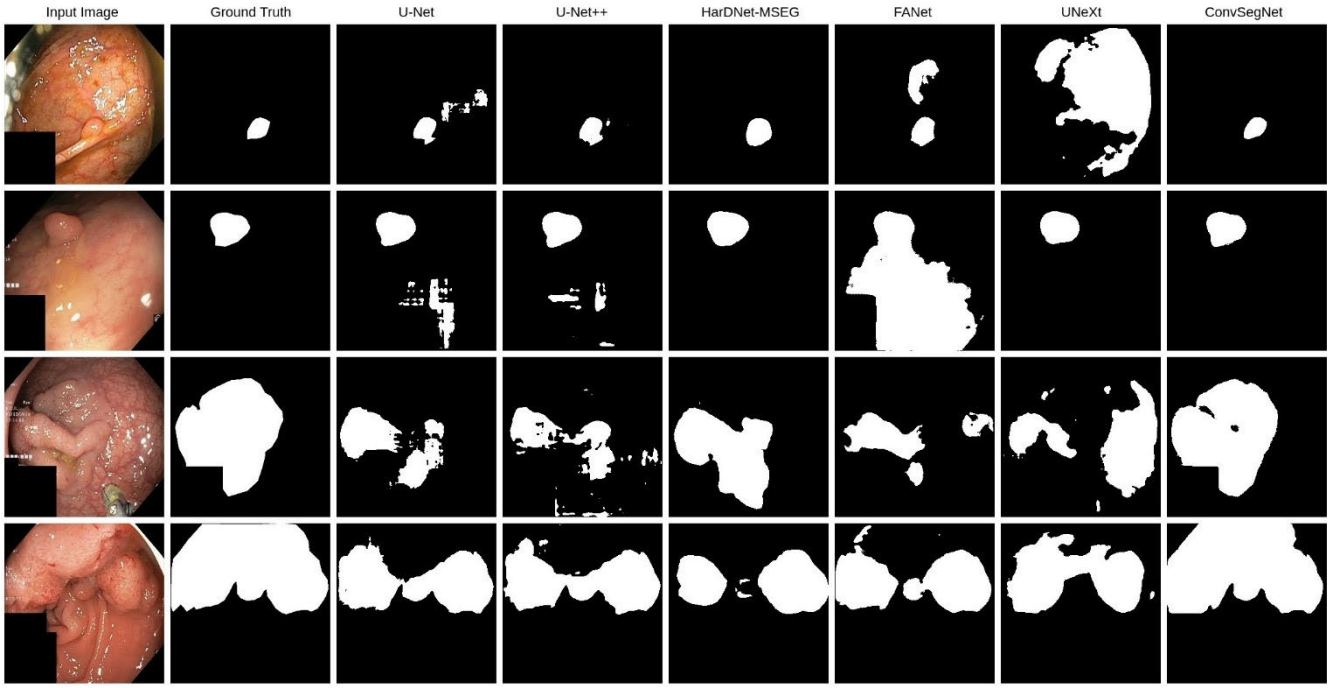


FIGURE 8. Visual representation of the input colonoscopy images, and the segmented polyp regions obtained using the benchmark architecture and ConvSegNet on Kvasir-SEG.

TABLE 3. Comparison of our model with benchmark models on Bkai-NeoPolyp.

Model	Jaccard	Dice	Recall	Precision	Accuracy	F2
U-Net	0.7599	0.8286	0.8295	**0.8999**	0.9903	0.8264
ResU-Net	0.6580	0.7433	0.7447	0.8711	0.9843	0.7387
U-Net++	0.7563	0.8275	0.8388	0.8942	0.9895	0.8308
HardDNet-MSEG	0.6734	0.7627	0.7532	0.8344	0.9863	0.7528
FANet	0.7578	0.8305	0.8285	0.9169	0.9863	0.7528
UNeXt	0.4680	0.5622	0.5926	0.7366	0.9740	0.5692
ConvSegNet	**0.8045**	**0.8747**	**0.9068**	0.8702	**0.9922**	**0.8909**

scores. Visual results of the segmentation performance are shown in Figure 9.

2) GENERALIZATION RESULTS

We conducted two experiments to test the generalization ability of the proposed ConvSegNet and compare it against the benchmarking architectures. In the first experiment, the whole Kvasir-SEG dataset was deployed for training, while the models were tested on the CVC-ColonDB. While in the second generalization experiment, the total CVC-ColonDB dataset was used for training, and Kvasir-SEG was used as the test set. The first experiments on the five models are shown in Table 4, while Table 5 shows the results of the second generalization experiment.

Results on the first generalization experiment, where we trained the models with Kvasir-Seg dataset and tested on CVC-clinicDB showed that our model outperformed all the benchmarking architectures. The Jaccard score of 0.7003 showed that ConvSegNet outperformed the benchmarking models with a difference of 0.157, 0.2036, 0.1528, 0.0946, 0.1658, and 0.3102 to U-Net, ResU-Net, U-Net++, HardDNet-MSEG, FANet and UNeXt respectively. Also, the dice score of 0.7764 achieved with the proposed model was better than the dice score of other models. Likewise, recall of 0.8078, Precision of 0.8625, Accuracy of 0.9685, and F2 measure of 0.7891 achieved with the ConvSegNet model outperformed the other benchmark architectures.

Similarly, as shown in Table 5, the second generalization experiment which trained on CVC-ClinicDB and tested on Kvasir-SEG showed that the proposed ConvSegNet model outperformed the benchmarking architectures. The Jaccard score of 0.6139 achieved by the proposed ConvSegNet is

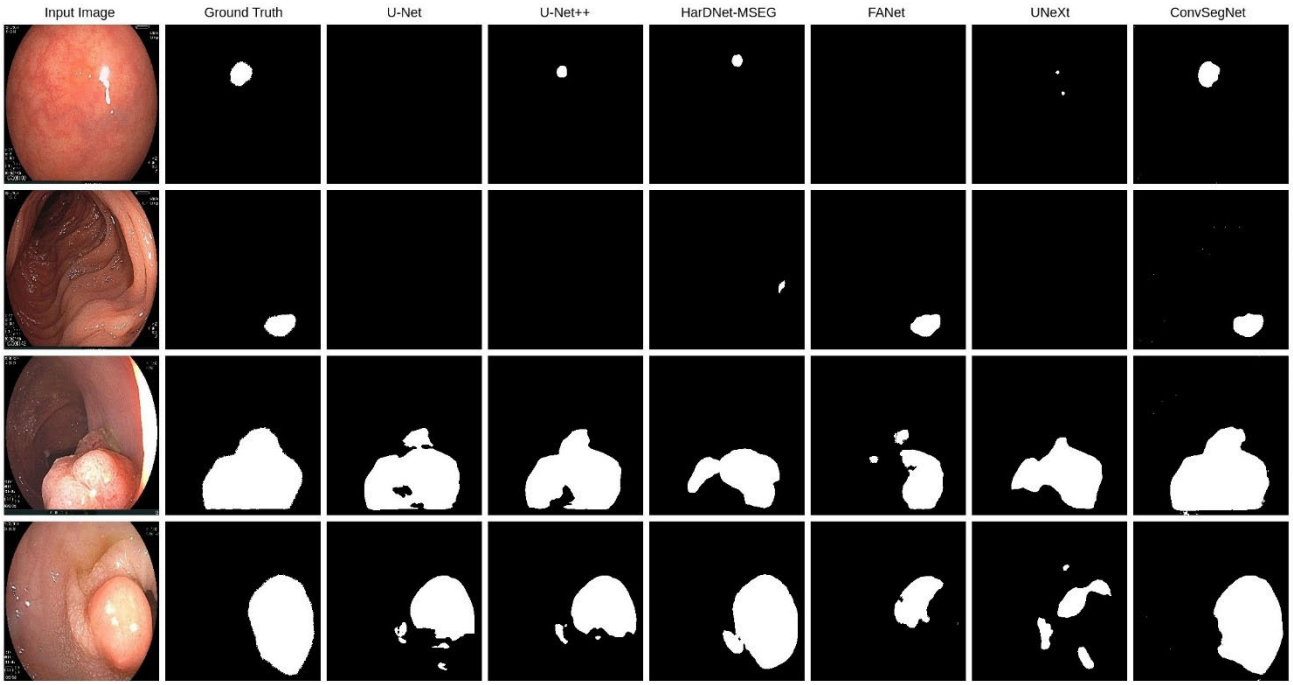


FIGURE 9. Visual representation of the input colonoscopy images, and the segmented polyp regions obtained using the benchmark architecture and ConvSegNet on BKAI-NeoPolyp dataset.

TABLE 4. Generalization comparison of ConvSegNet and benchmark architectures (train: Kvasir-SEG, test - CVC-ClinicDB).

Model	Jaccard	Dice	Recall	Precision	Accuracy	F2
U-Net	0.5433	0.6336	0.6982	0.7891	0.9484	0.6563
ResU-Net	0.4967	0.5970	0.6210	0.8005	0.9465	0.5991
U-Net++	0.5475	0.6350	0.6933	0.7967	0.9504	0.6556
HardDNet-MSEG	0.6057	0.6960	0.7173	0.8528	0.9592	0.7010
FANet	0.5345	0.6306	0.7707	0.6957	0.9283	0.6762
UNeXt	0.3901	0.4915	0.6125	0.6609	0.9216	0.5318
ConvSegNet	**0.7003**	**0.7764**	**0.8078**	**0.8625**	**0.9685**	**0.7891**

almost a 30% improvement over HardDNet-MSEG, which achieved the best performance of 0.4338 among the benchmarking models. Also, the ConvSegNet's Dice score, Recall, Precision, Accuracy, and F2 measure were relatively better than the benchmarking architectures.

3) ABLATION STUDIES

Ablation studies to investigate the effect of the CFR module were done. Experiments were carried out on a baseline model with the same hyperparameters as the ConvSegNet, but with the exclusion of the novel CFR module which we have introduced. The result of our ablation studies is shown in Table 6 and Table 7.

As shown in Table 6, the baseline model without the CFR module trained and tested on the CVC-ClinicDB dataset achieved a jaccard of 0.7560, dice of 0.8463, 0.9586 recall, 0.7662 precision, 0.9786 accuracy and 0.9074 F2 measure. This result showed that the CFR module introduced allowed more contextual information to be learnt. The results of the ablation studies on Kvasir-SEG dataset is presented in Table 7.

As shown in Table 7, the CFR module in the ConvSegNet model improved the performance of the baseline model. The baseline model however, had the same F2 measure as the ConvSegNet model, and had increased recall when compared to the 0.9124 achieved by the ConvSegNet on Kvasir-SEG dataset.

4) COMPUTATIONAL EVALUATION

A comparison of the size of the segmentation models, number of Flops and Frames per second was also done. We evaluated

TABLE 5. Generalization comparison of ConvSegNet and benchmark architectures (train: CVC-ClinicDB, test: Kvasir-Seg).

Model	Jaccard	Dice	Recall	Precision	Accuracy	F2
U-Net	0.3904	0.5126	0.8280	0.4628	0.7641	0.6112
ResU-Net	0.2789	0.4000	0.8801	0.3087	0.6293	0.5399
U-Net++	0.3489	0.4692	0.8294	0.4095	0.7143	0.5772
HardDNet-MSEG	0.4338	0.5521	0.7585	0.5479	0.8142	0.6128
FANet	0.4110	0.5189	0.8656	0.4762	0.7138	0.6163
UNeXt	0.3163	0.4365	0.7203	0.4175	0.7475	0.5204
ConvSegNet	**0.6139**	**0.7205**	**0.9069**	**0.6767**	**0.8928**	**0.7880**

TABLE 6. Ablation studies on CVC-ClinicDB.

Model	Jaccard	Dice	Recall	Precision	Acc	F2
Baseline w/o CFR	0.7560	0.8463	**0.9586**	0.7662	0.9786	0.9074
ConvSegNet (Proposed)	**0.8650**	**0.9177**	0.9518	**0.9048**	**0.9881**	**0.9328**

TABLE 7. Ablation studies on Kvasir-SEG.

Model	Jaccard	Dice	Recall	Precision	Acc	F2
Baseline w/o CFR	0.7414	0.8344	**0.9389**	0.7871	0.9546	0.8855
ConvSegNet (Proposed)	**0.7936**	**0.8618**	0.9124	**0.8692**	**0.9617**	0.8855

TABLE 8. Computational comprison of ConSegNet and benchmark models.

Model	Parameters (Million)	Flops (GMac)	FPS
U-Net	31.04	54.75	156.83
ResU-Net	8.22	45.42	196.85
U-Net++	9.16	34.65	126.14
HardDNet-MSEG	33.34	6.02	42
FANet	7.72	94.75	44
UNeXt	1.47	569.56	88.89
ConvSegNet	15.58	135.98	64

the benchmark models and the proposed ConvSegNet model and presented the results in Table 8.

As shown in Table 8, the U-Net architecture had a parameter value of 31.04M, HardDNet-MSEG had a parameter value of 33.34M, while ResU-Net and U-Net++ had 8.22M and 9.16M respectively. Also, FANet and UNeXt had 7.72M and 1.47M model parameters. The 15.58M parameter value achieved by the proposed ConvSegNet is minimal, compared to the segmentation performance achieved using the model, making the ConvSegNet model less bulky than U-Net and HardDNet-MSEG. The benchmark of U-Net, ResU-Net, U-Net++, HardDNet-MSEG, FANet and UNeXt had Flops of 54.75, 45.72, 34.65, 6.02, 94.75 and 569.56 respectively, while the ConvSegNet had 135.98 Flops. Likewise, the frames per second of the U-Net benchmark model was 156.83, ResU-Net had 196.85, U-Net++ had 126.14, HardDNet-MSEG had 42, FANet had 44 and UNeXt had 88.89 FPS, while ConvSegNet recorded 64 FPS. Showing that the ResU-Net model is faster than the other benchmarks and the ConvSegNet model.

E. DISCUSSION

We proposed a novel segmentation model called ConvvSegNet for the segmentation of polyp from colonoscopy. The Qualitative and generalization experiments showed that the proposed ConvSegNet architecture outperformed the benchmark architectures on which we performed experiments. Thereby proving that the proposed ConvSegNet model can extract a broad range of features progressively from input images, which enabled more significant features to be captured, making our network more robust. The generalization experiments also showed that the ConvSegNet model was able to generalize better than the benchmark models, by achieving improved performance scores over the benchmark models.

Ablation studies were also carried out to investigate the effect of the CFR module on the network by excluding the CFR module from the segmentation model. A model which we termed the Baseline, and the results were presented in Table 6 and Table 7. The baseline model was trained and tested on the same benchmark datasets, and the results showed that the inclusion of the CFR module improved the segmentation performance on CVC-ClinicDB dataset, with an improvement of 12.60% on the jaccard, 7.78% on dice index, 15.31% in precision, 0.96% on accuracy, and 2.72% improvement on F2 score. On the Kvasir-SEG dataset, an improvement of 6.57% jaccard was recoreded, 3.17% improvement in dice index, 9.44% in precision, and 0.73% in accuracy. This proved that the model achieved better performance when the CFR module is included in the segmentation network.

The comparison of the computation cost and efficiency of the ConvSegNet model with the benchmark showed that the ConvSegNet has less computational complexity than two of the benchmark models (U-Net and HardDNet-MSEG). However, the complexity of ResU-Net, U-Net++, FANet and UNeXt is far less than the proposed ConvSegNet, but their segmentation performance was outperformed by the ConvSegNet model. Also, the processing speed of the proposed ConvSegNet model was relatively low when compared to four of the benchmark models. However, the 64 FPS achieved by the ConvSegNet model is standard, and outperformed the recent benchmark of HardDNet-MSEG, which recorded 42FPS.

V. CONCLUSION

In this work, ConvSegNet architecture based on context feature refinement with multiple kernel sizes is trained for polyp segmentation from colonoscopy. The novel Context Feature Refinement module is proposed to address low model generalization and segmentation performance. The module is built to extract contextual information from the incoming feature map by applying multiple parallel convolutional layers with different kernel sizes. The outputs of these layers were then concatenated and passed through a 1×1 convolution for feature refinement. This way, we were able to take advantage of the low-level features to generate high-quality semantic features. Using different kernel sizes, we increased the receptive field during the convolution operation, which helped to better capture the contextual feature from the input feature map. The application of the proposed ConvSegNet model saw an improvement in polyp segmentation in terms of quantitative and generalization performances compared to the benchmark models in the study. The method proposed in this work can be further improved in terms of speed, segmentation performance and robustness.. Even though the ConvSegNet model was trained for polyp segmentation, the architecture can easily be extended for other biomedical image segmentation tasks. For future work, we plan to explore transformer models to guide the segmentation model in extracting more contextual information and explore ways to increase processing speed of the segmentation model.

REFERENCES

[1] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 4, 2012, pp. 2843–2851.

[2] P. Brandao, E. Mazomenos, G. Ciuti, R. Calió, F. Bianchi, A. Menciassi, P. Dario, A. Koulaouzidis, A. Arezzo, and D. Stoyanov, "Fully convolutional neural networks for polyp segmentation in colonoscopy," *Proc. SPIE*, vol. 10134, Mar. 2017, Art. no. 101340F, doi: 10.1117/12.2254361.

[3] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017, doi: 10.1016/j.media.2016.05.004.

[4] S. M. Thomas, J. G. Lefevre, G. Baxter, and N. A. Hamilton, "Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer," *Med. Image Anal.*, vol. 68, Feb. 2021, Art. no. 101915, doi: 10.1016/j.media.2020.101915.

[5] L. Zhou, Z. Li, J. Zhou, H. Li, Y. Chen, Y. Huang, D. Xie, L. Zhao, M. Fan, S. Hashmi, and F. Abdelkareem, "A rapid, accurate and machine-agnostic segmentation and quantification method for CT-based COVID-19 diagnosis," *IEEE Trans. Med. Imag.* vol. 39, no. 8, pp. 2638–2652, Aug. 2020, doi: 10.1109/TMI.2020.3001810.

[6] D. Bouget, A. Jørgensen, G. Kiss, H. O. Leira, and T. Langø, "Semantic segmentation and detection of mediastinal lymph nodes and anatomical structures in CT data for lung cancer staging," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 6, pp. 977–986, Jun. 2019, doi: 10.1007/s11548-019-01948-8.

[7] M. Akbari, M. Mohrekesh, E. Nasr-Esfahani, S. M. R. Soroushmehr, N. Karimi, S. Samavi, and K. Najarian, "Polyp segmentation in colonoscopy images using fully convolutional network," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 69–72, doi: 10.1109/EMBC.2018.8512197.

[8] N.-Q. Nguyen, D. M. Vo, and S.-W. Lee, "Contour-aware polyp segmentation in colonoscopy images using detailed upsampling encoder–decoder networks," *IEEE Access*, vol. 8, pp. 99495–99508, 2020, doi: 10.1109/ACCESS.2020.2995630.

[9] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA, Cancer J. Clin.*, vol. 65, no. 2, pp. 87–108, 2015, doi: 10.3322/caac.21262.

[10] A. G. Rundle, B. Lebwohl, R. Vogel, S. Levine, and A. I. Neugut, "Colonoscopic screening in average-risk individuals ages 40 to 49 vs 50 to 59 years," *Gastroenterology*, vol. 134, no. 5, pp. 1311–1315, May 2008, doi: 10.1053/j.gastro.2008.02.032.

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

[12] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Focus U-Net: A novel dual attention-gated CNN for polyp segmentation during colonoscopy," *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104815, doi: 10.1016/j.compbiomed.2021.104815.

[13] S. Thakkar, N. M. Carleton, B. Rao, and A. Syed, "Use of artificial intelligence-based analytics from live colonoscopies to optimize the quality of the colonoscopy examination in real time: Proof of concept," *Gastroenterology*, vol. 158, no. 5, pp. 1219–1221, 2020, doi: 10.1053/j.gastro.2019.12.035.

[14] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Automatic polyp segmentation using convolutional neural networks," in *Advances in Artificial Intelligence* (Lecture Notes in Computer Science), vol. 12109, C. Goutte and X. Zhu, Eds. Cham, Switzerland: Springer,2020, doi: 10.1007/978-3-030-47358-7_29.

[15] J. Zhong, W. Wang, H. Wu, Z. Wen, and J. Qin, "PolypSeg: An efficient context-aware network for polyp segmentation from colonoscopy videos," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 12266, 2020, pp. 285–294, doi: 10.1007/978-3-030-59725-2_28.

[16] J. Bernal, J. M. Núñez, F. J. Sánchez, and F. Vilariño, "Polyp segmentation method in colonoscopy videos by means of MSA-DOVA energy maps calculation," in *Proc. Workshop Clin. Image-Based Procedures*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 8680, 2014, pp. 41–49, doi: 10.1007/978-3-319-13909-8_6.

[17] J. G. B. Puyal, K. K. Bhatia, P. Brandao, O. F. Ahmad, D. Toth, R. Kader, L. Lovat, P. Mountney, and D. Stoyanov, "Endoscopic polyp segmentation using a hybrid 2D/3D CNN," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 12266, 2020, pp. 295–305, doi: 10.1007/978-3-030-59725-2_29.

[18] L. Zhang, S. Dolwani, and X. Ye, "Automated polyp segmentation in colonoscopy frames using fully convolutional neural network and textons," in *Proc. Annu. Conf. Med. Image Understand. Anal.*, in Communications in Computer and Information Science, vol. 723, 2017, pp. 707–717, doi: 10.1007/978-3-319-60964-5_62.

[19] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, and X. Chen, "CPFNet: Context pyramid fusion network for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3008–3018, Oct. 2020, doi: 10.1109/TMI.2020.2983721.

[20] S. Hwang, J. Oh, W. Tavanapong, J. Wong, and P. C. de Groen, "Polyp detection in colonoscopy video using elliptical shape feature," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2, Sep. 2007, pp. 465–468, doi: 10.1109/ICIP.2007.4379193.

[21] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilarino, "Texture-based polyp detection in colonoscopy," in *Bildverarbeitung für die Medizin* (Informatik aktuell), H. P. Meinzer, T. M. Deserno, H. Handels, and T. Tolxdorff, Eds. Berlin, Germany: Springer, 2009, doi: 10.1007/978-3-540-93860-6_70.

[22] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 630–644, Feb. 2015, doi: 10.1109/TMI.2015.2487997.

[23] D. You, S. Antani, D. Demner-Fushman, and G. R. Thoma, "An MRF model for biomedical image segmentation," in *Proc. IEEE 27th Int. Symp. Comput.-Based Med. Syst.*, May 2014, pp. 539–540, doi: 10.1109/CBMS.2014.128.

[24] A. van Opbroek, M. A. Ikram, M. W. Vernooij, and M. de Bruijne, "Transfer learning improves supervised image segmentation across imaging protocols," *IEEE Trans. Med. Imag.*, vol. 34, no. 5, pp. 1018–1030, May 2015, doi: 10.1109/TMI.2014.2366792.

[25] A. Norouzi, M. S. M. Rahim, A. Altameem, T. Saba, A. E. Rad, A. Rehman, and M. Uddin, "Medical image segmentation methods, algorithms, and applications," *IETE Tech. Rev.*, vol. 31, no. 3, pp. 199–213, 2014, doi: 10.1080/02564602.2014.906861.

[26] X. Zhou and G. B. Thompson, "Influence of solute partitioning on the microstructure and growth stresses in nanocrystalline Fe(Cr) thin films," *Thin Solid Films*, vol. 648, pp. 83–93, Feb. 2018, doi: 10.1016/j.tsf.2018.01.007.

[27] X. Zhou, K. Yamada, T. Kojima, R. Takayama, S. Wang, X. Zhou, T. Hara, and H. Fujita, "Performance evaluation of 2D and 3D deep learning approaches for automatic segmentation of multiple organs on CT images," *Proc. SPIE*, vol. 10575, Jan. 2018, Art. no. 105752C, doi: 10.1117/12.2295178.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[29] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Int. Workshop Multimodal Learn. Clin. Decis. Support*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11045, Jun. 2018, pp. 3–11, doi: 10.1007/978-3-030-00889-5_1.

[30] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018, doi: 10.1109/LGRS.2018.2802944.

[31] F. Condessa and J. Bioucas-Dias, "Segmentation and detection of colorectal polyps using local polynomial approximation," in *Proc. ICIAR*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 7325, 2012, pp. 188–197, doi: 10.1007/978-3-642-31298-4_23.

[32] H. Zunair and A. B. Hamza, "Sharp U-Net: Depthwise convolutional network for biomedical image segmentation," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104699, doi: 10.1016/j.compbiomed.2021.104699.

[33] A. Srivastava, D. Jha, S. Chanda, U. Pal, H. D. Johansen, D. Johansen, M. A. Riegler, S. Ali, and P. Halvorsen, "MSRF-Net: A multi-scale residual fusion network for biomedical image segmentation," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 5, pp. 2252–2263, Dec. 2021, doi: 10.1109/JBHI.2021.3138024.

[34] W. Weng and X. Zhu, "INet: Convolutional networks for biomedical image segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, 2021, doi: 10.1109/ACCESS.2021.3053408.

[35] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. D. Lange, P. Halvorsen, and H. D. Johansen, "ResUNet++: An advanced architecture for medical image segmentation," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2019, pp. 225–230, doi: 10.1109/ISM46123.2019.00049.

[36] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Ruecker, "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019, doi: 10.1016/j.media.2019.01.012.

[37] T. Kim, H. Lee, and D. Kim, "UACANet: Uncertainty augmented context attention for polyp segmentation," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 2167–2175, doi: 10.1145/3474085.3475375.

[38] T. Mahmud, B. Paul, and S. A. Fattah, "PolypSegNet: A modified encoder–decoder architecture for automated polyp segmentation from colonoscopy images," *Comput. Biol. Med.*, vol. 128, Jan. 2021, Art. no. 104119, doi: 10.1016/j.compbiomed.2020.104119.

[39] X. Zhao, L. Zhang, and H. Lu, "Automatic polyp segmentation via multi-scale subtraction network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 12901, 2021, pp. 120–130, doi: 10.1007/978-3-030-87193-2_12.

[40] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019, doi: 10.1109/TMI.2019.2903562.

[41] H. A. Qadir, Y. Shin, J. Solhusvik, J. Bergsland, L. Aabakken, and I. Balasingham, "Polyp detection and segmentation using mask R-CNN: Does a deeper feature extractor CNN always perform better?" in *Proc. 13th Int. Symp. Med. Inf. Commun. Technol. (ISMICT)*, May 2019, pp. 1–6, doi: 10.1109/ISMICT.2019.8743694.

[42] N. K. Tomar, D. Jha, M. A. Riegler, H. D. Johansen, D. Johansen, J. Rittscher, P. Halvorsen, and S. Ali, "FANet: A feedback attention network for improved biomedical image segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 25, 2022, doi: 10.1109/TNNLS.2022.3159394.

[43] J. M. J. Valanarasu and V. M. Patel, "UNeXt: MLP-based rapid medical image segmentation network," in *Medical Image Computing and Computer Assisted Intervention—MICCAI* (Lecture Notes in Computer Science), vol. 13435, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham, Switzerland: Springer, 2020, doi: 10.1007/978-3-031-16443-9_3.

[44] P. Sharma, A. Gautam, P. Maji, R. B. Pachori, and B. K. Balabantaray, "Li-SegPNet: Encoder–decoder mode lightweight segmentation network for colorectal polyps analysis," *IEEE Trans. Biomed. Eng.*, early access, Oct. 21, 2022, doi: 10.1109/TBME.2022.3216269.

[45] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-SEG: A segmented polyp dataset," in *Proc. Int. Conf. Multimedia Modeling*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11962, 2020, pp. 451–462, doi: 10.1007/978-3-030-37734-2_37.

[46] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Computerized Med. Imag. Graph.*, vol. 43, pp. 99–111, Jul. 2015, doi: 10.1016/j.compmedimag.2015.02.007.

[47] P. N. Lan, N. S. An, D. V. Hang, D. V. Long, T. Q. Trung, N. T. Thuy, and D. V. Sang, "NeoUNet: Towards accurate colon polyp segmentation and neoplasm detection," in *Proc. 16th Int. Symp. Adv. Vis. Comput. (ISVC)*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 13018, Oct. 2021, pp. 15–28.

[48] D. Vázquez, J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, A. M. López, A. Romero, M. Drozdzal, and A. Courville, "A benchmark for endoluminal scene segmentation of colonoscopy images," *J. Healthcare Eng.*, vol. 2017, pp. 1–9, Jan. 2017, doi: 10.1155/2017/4037190.

[49] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P. T. Schmidt, M. Riegler, and P. Halvorsen, "KVASIR: A multi-class image dataset for computer aided gastrointestinal disease detection," in *Proc. 8th ACM Multimedia Syst. Conf.*, Jun. 2017, pp. 164–169, doi: 10.1145/3083187.3083212.

[50] C.-H. Huang, H.-Y. Wu, and Y.-L. Lin, "HarDNet-MSEG: A simple encoder–decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 FPS," 2021, *arXiv:2101.07172*.

AYOKUNLE OLALEKAN IGE received the B.Sc. (Hons.) and M.Sc. degrees in computer science from Adekunle Ajasin University, Akungba-Akoko, Nigeria, in 2014 and 2019, respectively. He is currently a Ph.D. Researcher with the School of Computer Sciences, Universiti Sains Malaysia. His research interests include image segmentation, deep learning, and computer vision.

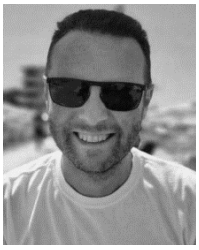
NIKHIL KUMAR TOMAR is currently pursuing the master's degree in computer application with Indira Gandhi Open University, New Delhi, India. He has worked on different deep learning-based biomedical image analysis problems in close collaboration with researchers from different universities. His research interests include computer vision, artificial intelligence, parallel processing, and medical image segmentation.

FELIX OLA ARANUWA received the Ph.D. degree from the Malaysia University of Science and Technology, in 2015. He is a currently a Senior Faculty Member with the Department of Computer Science, Adekunle Ajasin University, Akungba-Akoko, Ondo, Nigeria. His research interests include data analytics, machine learning, and biometrics/image processing.

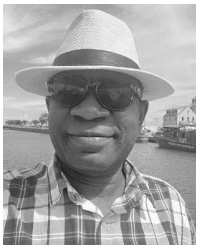
OLUWAFEMI ORIOLA received the B.Sc. degree (Hons.) in computer science from Adekunle Ajasin University, Akungba-Akoko, Nigeria, in 2006, and the M.Sc. and Ph.D. degrees from the University of Ibadan, Nigeria, in 2010 and 2015, respectively. He is currently a Senior Lecturer with the Department of Computer Science, Adekunle Ajasin University. He was a Doctoral Fellow with the CSCAN, University of Plymouth, U.K., in 2014. He was also a Postdoctoral Fellow with the Department of Computer Science and Informatics, University of the Free State, South Africa, from 2019 to 2020. His research interests include machine learning, deep learning, predictive analytics, and cybersecurity.

ALABA O. AKINGBESOTE received the B.Sc. degree (Hons.) from Ogun State University, Nigeria, the M.Tech. degree in computer science from the Federal University of Technology, Akure, Nigeria, and the Ph.D. degree in computer science from the University of Zululand, South Africa, in 2015. He is currently a Senior Faculty Member with the Department of Computer Science, Adekunle Ajasin University, Akungba-Akoko, Ondo, Nigeria. He has published his research in many conference proceedings and various reputable journals. His research interests include artificial intelligence, cloud e-market, and machine learning.

MOHD HALIM MOHD NOOR received the B.Eng. degree (Hons.) in computer and information engineering, the M.Sc. degree in electrical and electronic engineering, and the Ph.D. degree in computer systems engineering from the University of Auckland, New Zealand. He is currently a Senior Lecturer with the School of Computer Sciences, Universiti Sains Malaysia. His research interests include machine learning, deep learning, computer vision, and pervasive computing.



MANUEL MAZZARA received the Ph.D. degree in computing science from the University of Bologna, Italy. He is currently a Professor of computer science with Innopolis University, Russia, with a research background in software engineering, service-oriented architecture, concurrency theory, formal methods, and software verification. Currently, he is the Dean of the Faculty of Computer Science and Engineering and the Head of the International Cooperation Office at Innopolis University. He has published many relevant and highly cited papers, in particular in the field of service engineering and software architectures. He has collaborated with European and U.S. industries and governmental and intergovernmental organizations, such as the United Nations, always at the edge between science and software production. The work conducted by him and his team in recent years focuses on the development of theories, methods, tools, and programs covering the two major aspects of software engineering, such as the process side, related to how we develop software; and the product side, concerning the results of this process.



BENJAMIN SEGUN ARIBISALA received the Ph.D. degree in computer science from the University of Birmingham, U.K. He became a Professor of computer science, in 2013. He has served at several leadership positions, such as the Head of the Department of Computer Science, the Dean of the Faculty of Science, the Director of ICT, and a member of the Governing Council in Lagos State University (Nigeria). He was worked as an Academic Staff with Newcastle University (England, U.K.) and the University of Edinburgh (Scotland, U.K.). He is currently a Professor of computer science at Lagos State University and a Fulbright Visiting Professor of computer science at The University of Chicago (USA). His research interests include medical image analysis, data science, machine learning, deep learning, and artificial intelligence. His research focuses on improving medical diagnosis, quality of life, and life expectancy.