IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 August 2023, accepted 24 August 2023, date of publication 31 August 2023, date of current version 8 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3310541

RESEARCH ARTICLE

DeepPoly: Deep Learning-Based Polyps Segmentation and Classification for Autonomous Colonoscopy Examination

MD SHAKHAWAT HOSSAIN[1,2], (Member, IEEE), MD. MAHMUDUR RAHMAN[3], M. MAHBUBUL SYEED[1,2], (Member, IEEE), MOHAMMAD FAISAL UDDIN[1,2], (Member, IEEE), MAHADY HASAN[1,2], MD. AULAD HOSSAIN[4], AMEL KSIBI[5], MONA M. JAMJOOM[6], ZAHID ULLAH[7], (Member, IEEE), AND MD ABDUS SAMAD[8], (Member, IEEE)

[1]Department of Computer Science and Engineering, Independent University, Bangladesh, Dhaka 1229, Bangladesh
[2]RIoT Research Center, Independent University, Bangladesh, Dhaka 1229, Bangladesh
[3]Department of Computer Science, American International University-Bangladesh, Dhaka 1229, Bangladesh
[4]Department of Urology, Bangabandhu Sheikh Mujib Medical University Hospital, Dhaka 1000, Bangladesh
[5]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[6]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[7]Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[8]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan-si, Gyeongsangbuk-do 38541, South Korea

Corresponding author: Md Abdus Samad (masamad@yu.ac.kr)

This work was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R104), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Independent University, Bangladesh, under Application No. 2023-SETS-0223.

ABSTRACT Colorectal cancer (CRC) is the third most common cause of cancer-related deaths in the United States and is anticipated to cause another 52,580 deaths in 2023. The standard medical procedure for screening and treating colorectal disease is a colonoscopy. By effectively examining the colonoscopy to identify precancerous polyps early and remove them before they become cancerous, CRC mortality can be lowered significantly. Manual colonoscopy examination for precancerous polyps detection is time-consuming, tedious, and prone to human error. Automatic segmentation and analysis could be fast and practical; however, existing automated methods fail to attain adequate accuracy in polyps segmentation. Moreover, these methods do not assess the risk of detected polyps. In this paper, we proposed an autonomous CRC screening method to detect polyps and assess their potential threats. The proposed method utilized DoubleU-Net for polyps segmentation and Vision Transformer (ViT) for classifying them based on their risks. The proposed method has achieved a mean dice-coefficient of 0.834 and 0.956 in segmentation for the Endotech challenge and Kvasir-SEG dataset, accordingly outperforming the existing state-of-the-art polyps segmentation. Then, this method classified the segmented polyps as hyper-plastic or adenomatous with 99% test accuracy.

INDEX TERMS Colorectal cancer, colonoscopy, polyps segmentation, polyps classification, EndoTect challenge, Kvasir.

I. INTRODUCTION

The human digestive tract is susceptible to anomalies ranging from minor health concerns to life-threatening cancer. The digestive system mainly consists of the gastrointestinal (GI) tract and digestive organs such as the liver, gallbladder, and pancreas. Although the GI is a part of the digestive system, it is also commonly known as the digestive tract. The GI tract includes the organs that process foods, such as mouth, throat, esophagus, stomach, small intestine (i.e., duodenum,

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

jejunum, and ileum), large intestine or colon and rectum. The GI tract, mainly the large intestine or colon, can be affected by mucosal abnormalities such as ulcers or polyps caused by different issues, including infections, bacteria, viruses, or parasites. Some of these polyps may be fatal and significantly threaten the patient's health, including cancer development, commonly known as colorectal cancer (CRC). However, if such polyps are detected early, they can be healed completely, and the chances of CRC can be eliminated. A polyp is a growth of tissue that protrudes from a bodily surface, most commonly a mucous membrane. They look like small clumps or tiny mushroom-like stalks. Polyps in the colon are the most prevalent, although they can also form in the uterus, ear canal, cervix, stomach, nose, and throat. Polyps are harmless or benign. However, over time, they can develop into cancer and become invasive, which may be fatal when found in their later stages. Polyps form on the lining of the colon or rectum as noncancerous adenomas and then progress to malignancy over the years, as shown in Fig. 1. These phases can be typically divided into three categories: 1) hyperplastic stage, 2) adenomatous or pre-cancerous stage and 3) malignant or CRC stage.

Hyper-proliferation is considered the hyperplastic stage. Hyperplastic polyps are often tiny and found at the end of the colon, such as the rectum or sigmoid colon. They have no malignant potential and are not threatening. Adenomatous polyps are the most prevalent, accounting for almost two-thirds of all colon polyps. Small polyps, large polyps and severe dysplasia come under adenomatous polyps, also known as adenomas. Most of these polyps do not progress to cancer, even though they have the potential to do so. Polyps are categorized according to size, general appearance, and specific features visible under a microscope. The larger the polyps, the more probable it is to develop into cancer. As a result, urologists advise removing polyps larger than 5 mm to avoid future CRC. After the adenomatous stage, the polyp cells become malignant CRC and then advance to invasive CRC. Adenocarcinoma and invasive CRC are both considered serious threats. Fig. 2 shows the images of hyperplastic and adenomatous polyps. Hyperplastic polyps can be characterized by their vessels and surfaces. They have lighter vessels in comparison to their surroundings. A circular pattern surface with a small dot pattern surrounded by lighter mucosa can occasionally be seen. The adenomatous polyps have dark brown vessels with a lighter area in the center. Mostly, they have oval or tubular surfaces surrounded by brown vessels.

Polyps develop slowly over time [1]. Colorectal malignancies develop from precancerous polyps, such as adenomatous polyps or severe dysplasia polyps, which form over time and eventually become cancer. The progression from hyperplastic polyps to malignant polyps can take as short as five years or as long as 20 years. As of 2016, the Canadian Task Force on Preventive Health Care recommends screening adults aged 50 to 74 for CRC every two years. Anyone can develop colon or rectum polyps. Usually, people who are overweight, older than 50, have a smoking habit, or have a family history of colon polyps or cancer are at a higher risk. However, the risk of CRC can be prevented by detecting the polyps of the colon and rectum in the early stage, such as hyperplastic and then removing them safely and completely. Detecting polyps at the precancerous stage also prevents excessive treatment costs by diagnosing before CRC invades other organs [2]. According to one study [3], a patient has a 90% chance of 5-year survival if the CRC is detected at a local stage, which means the cancer is restricted to its original site. If the CRC is detected at a metastatic stage, when it has spread to nearby organs, the chance of survival is reduced to 70%. When detected at an invasive stage, the chances are 10%.

Polyps often do not cause symptoms [1], [2]; thus, the best prevention for CRC is the regular screening and examination of the colon and rectum. Consequently, colonoscopy is performed routinely for screening and detecting pre-cancerous polyps. However, distinguishing hyperplastic polyps from adenomatous polyps with colonoscopy vision is challenging and urologists frequently rely on microscopic analysis of colon biopsies. Colonoscopy is a form of endoscopy widely accepted and the most effective method for examining the gastrointestinal (GI) tract, encompassing the esophagus, stomach, duodenum, colon, and rectum. The endoscopy can be further classified into two types: Gastroscopy or upper endoscopy and colonoscopy or lower endoscopy. Colonoscopy is a nonsurgical procedure in which a thin and flexible tube called an endoscope is inserted into the colon for examination. The endoscope contains a powerful light and a tiny camera that visually examines an organ without making large incisions. A colonoscopy is performed through the rectum to examine the rectum, large intestine, and colon. In current clinical practice, the examination is performed manually based on a naked observation by an expert gastroenterologist. This procedure is time-consuming, vulnerable to fatigue and subject to inter and intra-observer variability. Moreover, such expert-dependent tests delay healthcare when there is a severe scarcity of health professionals [4] worldwide. This signifies the need for automatic examination for time efficient, consistent, and accurate colonoscopy.

This paper proposes a polyps segmentation and classification method for automated colonoscopy examination. Firstly, this method segmented the polyps from the colonoscopy image and then classified the segmented polyps as hyperplastic and adenomatous. Currently, in colonoscopy, the abnormalities are identified by an expert. This examination depends on the expert's skill and experience and occasionally results in misidentification. Borgli et al. [5] reported 20% misidentification on average in colonoscopy. Automated polyps segmentation and classification can aid experts by detecting and tracing polyps from colonoscopy images or videos. This has drawn the attention of computer vision experts, resulting in abundant studies of autonomous polyp detection. These methods, however, could not attain adequate accuracy. Another issue is that methods proposed by different

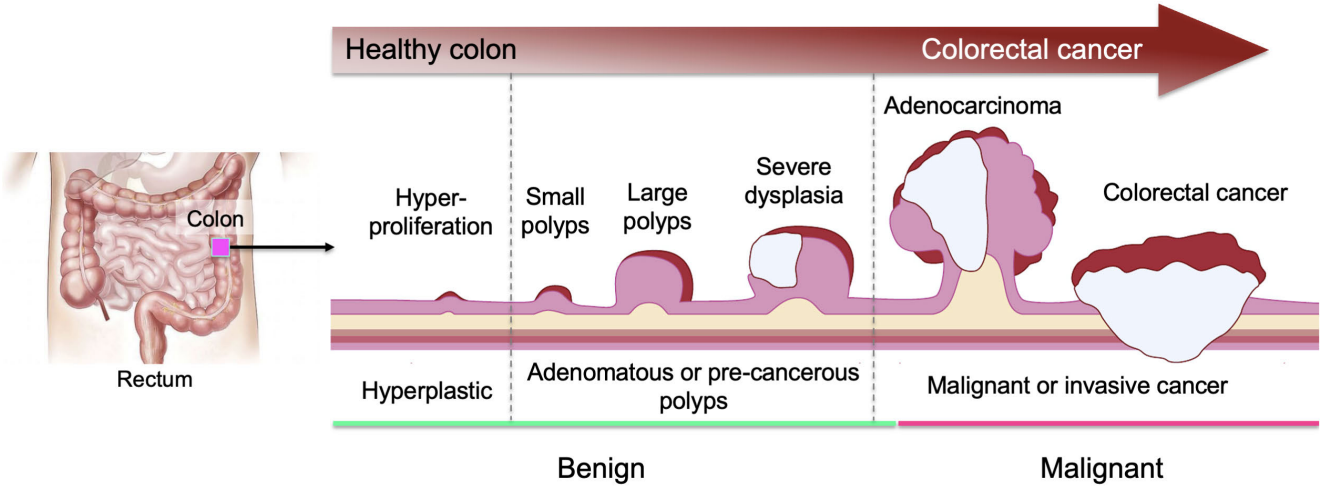

FIGURE 1. Different stages of colorectal cancer development.

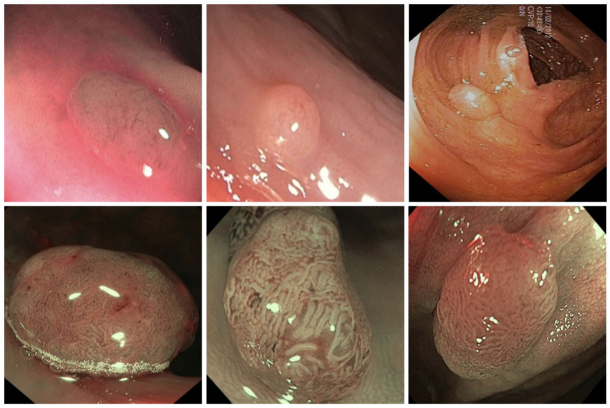

FIGURE 2. Colonoscopy images of hyperplastic (top-row) and adenomatous polyps (bottom-row).

groups were trained and tested on different datasets, making it difficult to compare.

In 2020 a dataset was released by Hicks et al. called the Endotect challenge dataset [6], which provides an opportunity to benchmark the polyps detection method. The segmentation network of the proposed system was trained using the Hyper-Kvasir dataset and tested using two different datasets: Endotect [6] Challenge test images and Kvasir-SEG [7] dataset. The segmentation network of the proposed system outperformed the previous polyps segmentation methods when tested on the Endotect challenge and Kvasir dataset. Moreover, the proposed system incorporates the classification of polyps with the polyps segmentation. The colonoscopy examination is incomplete without the classification or risk assessment of polyps. Existing automated methods are limited to the segmentation of polyps and the classification of polyps based on its severity was not considered. An autonomous colonoscopy examination system must incorporate polyps segmentation with classification.

The major contributions of this paper are listed as follows: 1) the development of an automated colonoscopy examination system that incorporates polyps classification with the polyps segmentation, 2) the improvement of polyps segmentation accuracy, 3) validation of the proposed method on heterogeneous dataset to ensure its resiliency and 4) subjective evaluation of the system by experts for clinical use.

II. RELATED WORKS

As the number of deaths from colorectal cancer has increased over the years, computer vision experts have paid close attention to autonomous polyp examination for colonoscopy screening. More than 100 publications have been published since 2018 for segmenting, identifying, or categorizing polyps in colonoscopy images. Several CADx systems have recently been proposed for polyp segmentation by multiple groups [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24]. These methods, however, were trained and evaluated on separate datasets. As a result, comparing their performance is difficult. To address this issue, in 2020, a dataset called Endotect challenge [6] was created to benchmark and assess various segmentation methods, allowing them to be compared using the same dataset. In 2021, a polyp segmentation method based on DDAnet was proposed by Tomar et al. [13]. This method was tested using both Kvasir-SEG [7] and Endotect challenge dataset [6]. However, this method achieved a dice coefficient of 0.78 for Endotect and 0.85 for the Kvasir-SEG dataset, which is not efficient. Some other methods were proposed for segmenting polyps; however, none were tested for the Endotect Challenge dataset.

Nguyen-Mau et al. suggested a method for polyps segmentation that integrates the Multi Kernel Positional Embedding block (MPE) with the ConvNeXt backbone to extend the receptive field and obtain multi-scale information [14]. This method achieved a dice coefficient of 0.88 on the Kvasir-SEG

dataset. Another method called MSRF-Net [15] capitalized the multi-scale features of varying receptive fields was proposed by Srivastava et al. to improve polyps segmentation to 0.921 [15]. Two more methods termed COLONFORMER [16] and GMSRF-Net [17] were developed to improve segmentation utilizing the multilevel features. These approaches produced dice coefficients comparable to the MSRF-Net. Sanderson and Matuszewski presented the FCN-Transformer [18] approach, which combined a fully convolutional network with a transformer. This method achieved the dice coefficient of 0.938 for the Kvasir-SEG dataset. Chang et al. introduced an architecture for object segmentation from medical images employing pre-trained transformer encoders and a stage-wise feature pyramid decoder [19]. This architecture was primarily designed for segmenting bronchoscopy lesions and was effective for polyps segmentation. The dice coefficient obtained with this method was 0.931. Some other techniques such as CaraNet [20], DuAT [21], HarDNet-DFUS [22], SEP [23] and SSFormer-L [24] were also proposed but they failed to achieve dice coefficient higher than 0.941 for the Kvasir-SEG dataset.

Other polyps segmentation algorithms [25], [26], [27] have been trained and evaluated on private datasets, making it difficult to generalize and compare their performance. Zhang et al. utilized a single-shot multi-box detector that reused shifted information through max-pooling layers to achieve an accuracy of 90.4% [25]. Bagheri et al. achieved 97.7% accuracy for polyps segmentation using color information and sophisticated preprocessing [26]. Yuan and Meng proposed a stacked sparse autoencoder-based method for detecting polyps, which yielded an accuracy of 98% [27]. However, it is not possible to generalize these methods' performance for public datasets. As a result, we only compared the proposed system to those methods tested on public datasets. All the methods discussed earlier are limited to the segmentation of polyps and do not consider classifying the grade of it.

Some studies incorporated the polyps classification with the segmentation for autonomous colonoscopy examination. Ribeiro et al. presented a CNN-based method to classify polyps as healthy or abnormal in which the accuracy was 90% [28]. Another method was proposed by Zheng et al. in which they integrated CNN-based features with a Support Vector Machine (SVM) to classify polyps as hyperplastic or adenomatous [29]. This method obtained 86% accuracy. Bryne et al. also classified polyps as hyperplastic or adenomatous, in which they used only narrow-band imaging (NBI) [30] and the accuracy was 94%. Lui et al. classified polys as endoscopically curable lesions or incurable lesions [31]. This method also used NBI images and achieved 85.5% accuracy. NBI is a sophisticated endoscopic technique that evaluates surface patterns and microvascular architecture using a narrower spectrum of light. However, not all laboratories have access to the NBI imaging facility.

Unlike the NBI image-based method, Hsu et al. used grayscale images to classify polyps as neoplastic or hyperplastic with 82.8% accuracy [32]. In 2022 Chung-Ming et al. presented a method to classify polyps as hyperplastic or adenomatous. They relied on the traditional feature selection methods such as the Gabor filter to extract polyps features, which a CNN-based model then used to classify the polyps. This method obtained 96.4% accuracy. Another recent work was proposed by Krenzer et al. to classify polyps according to the Paris classification criteria [34]. Paris classification scheme characterizes the potentially high-risk polyps according to their shape [35]. This method first detected and cropped the polyps on the image and then classified the cropped area using a transformer, similar to our approach. This method was tested on Showa University and Nagoya University (SUN) dataset [36] in which, it achieved 89.35% accuracy. Few more studies proposed to classify polyps using CNN; however, the accuracy of these methods [37], [38], [39], [40] was lower than 85%.

In this study, we used an approach like that used by Krenzer et al. [34]. In the first stage, the proposed system segmented the polyps in the images and then generated ROIs for the segmented polyps. A ViT model was then used to classify each segmented ROI. However, unlike prior methods, this system was evaluated on heterogeneous datasets and outperformed the existing methods with 95.6% segmentation accuracy and 99.6% classification accuracy. Furthermore, we aimed to evaluate the functional efficacy of the system in order to facilitate its implementation in hospitals.

III. MATERIALS AND METHOD

A. DATASET

In this investigation, we used five different datasets: 1) Kvasir, 2) Kvasir-SEG, 3) Hyper-Kvasir, 4) Endotect Challenge and 5) BSM-DU dataset for our experiments. These images were generated using different colonoscopy devices and prepared in different laboratories. We trained, validated, and tested the proposed system using these heterogeneous images to ensure the robustness of the proposed system. The distribution of the dataset is given in Table 1. In total, 2000 images were used for the segmentation experiment and 1800 for classification.

Three experts reviewed the images and their associated segmentation masks. Then, the experts assessed the grade of polyps and determined their class as hyperplastic or adenomatous independently. After that, images for which 100% concordance was achieved were used in this study. Images of malignant polyps or images for which 100% concordance was not achieved were eliminated. The image dimensions vary in the datasets. Therefore, we have resized the images to fit our experiments. This study utilized deidentified human specimens and we obtained ethical approval from the Institutional Review Board of Independent University, Bangladesh (approval code: 2023-SETS-0223).

TABLE 1. Training, validation and test data distribution for segmentation and classification experiment.

	Segmentation			Classification	
	Hyper-Kvasir	Endotect	Kvasir-SEG	Kvasir	Hyper-Kvasir
Training	600	-	-	800	-
Validation	200	-	-	200	-
Test	-	200	1000	-	800

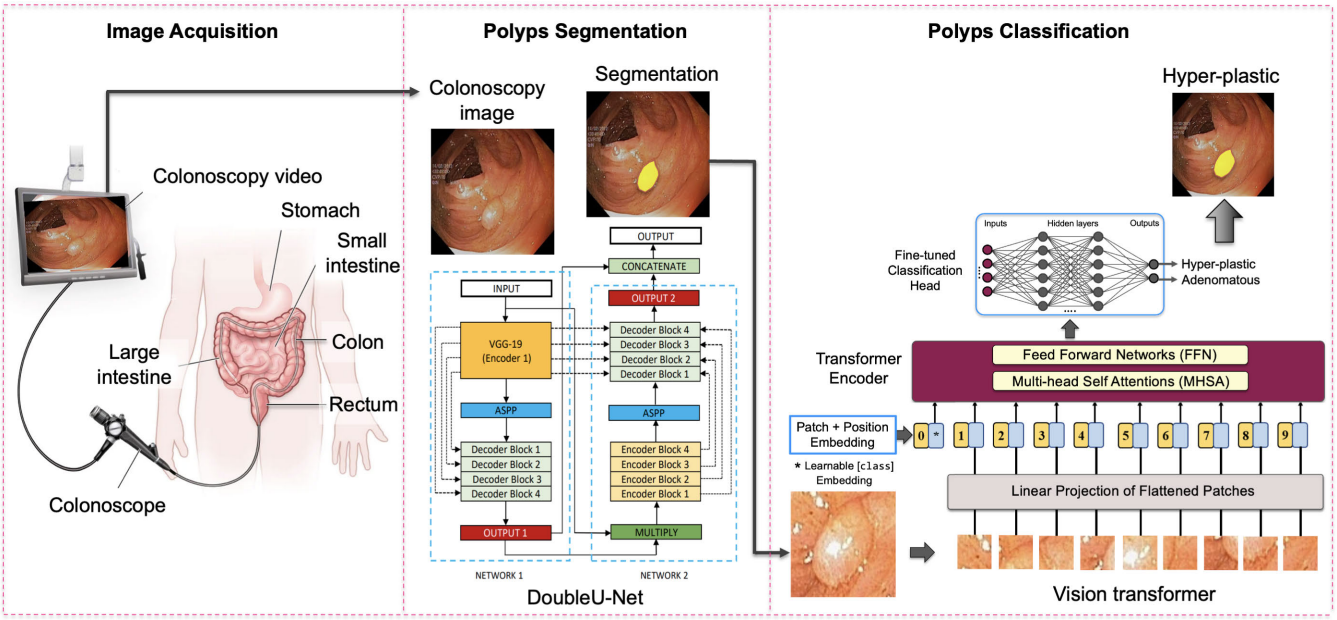

FIGURE 3. Architecture of the proposed system.

1) KVASIR DATASET

The Kvasir [41] dataset consists of 4,000 images, including 500 polyps images. Later, it was extended to 8000 images containing 1000 polyps images.

2) KVASIR-SEG DATASET

The Kvasir-SEG dataset [7] was generated from the Kvasir dataset [41], which was initially developed for classification tasks. The Kvasir-SEG dataset was prepared in Vestre Viken Health Trust in Norway. It is an open-access dataset that contains 1000 images and corresponding 1000 masks for polyps segmentation. A medical doctor prepared these masks and verified them by an experienced gastroenterologist in Vestre Viken Health Trust. Further, these masks were verified by our experts for this study.

3) HYPER-KVASIR DATASET

The Hyper-Kvasir [5] is another publicly available dataset that contains 110,079 images and 374 videos. Olympus and Pentax imaging devices were used for developing this dataset at Vestre Viken Hospital Trust, Norway. Vestre Viken Hospital, the Cancer Registry of Norway and Karolinska University Hospital in Sweden verified these images. Our study used 800 polyps class images from this dataset, which our experts then reviewed to categorize them as hyperplastic and adenomatous polyps. The Hyper-Kvasir dataset does not contain the subclass information of the polyps images. Hyper-Kvasir also includes segmentation masks for 1,000 images from the polyp class. Our experts verified these masks.

4) ENDOTECT CHALLENGE DATASET

The Endotect challenge dataset includes [6] the Hyper-Kvasir dataset for training and validation. Additionally, it contains 200 challenge images to test the methods.

5) BMS-DU PRIVATE DATASET

BSM-DU is a private dataset prepared for this study to evaluate the performance of the proposed system for routine clinical applications in the hospitals of Bangladesh. This dataset consists of 40 colonoscopy images generated from the colonoscopy video of 10 patients for the evaluative criticism of the proposed system. This dataset is not publicly available now.

B. ARCHITECTURE OF THE PROPOSED SYSTEM

The architecture of the proposed system consists of three modules: 1) image acquisition module, 2) polyps

segmentation module and 3) polyps classification module, as shown in Fig. 3. Firstly, in the image acquisition, colonoscopy RGB images are extracted from the colonoscopy video. In our experiment, we used an Olympus colonovideoscope machine for colonoscopy. This device was combined with the brighter narrow-band imaging to provide a clear view of the mucosal surface and capillary networks; however, we did not utilize the NBI images. We automatically extracted the RGB images from the colonoscopic video and resized them to 720 × 570 pixels. Then, this RGB image is further pre-processed in the segmentation module for polyps segmentation, which includes quality evaluation, black pixel estimation and color correction. The algorithm 1 explains the proposed method step by step.

The segmentation module evaluates the RGB image for black pixels and image quality. If an image contains more than 50% of black pixels or the quality degradation score is higher than 5, it is rejected for further processing. A pixel whose intensity is less than 50 is considered black or specimen-less area. If an image contains mostly (50%) black pixels, it is considered useless. Again, if the quality degradation index of the image is higher than 5, it is regarded as a poor-quality image and not valid for diagnosis. The quality degradation score of the image is determined based on its sharpness using Eq. (1) as:

$$\text{Quality degradation score} = \frac{1}{N}\sum_{i=1}^{N}\text{Diff}(i) \tag{1}$$

where N is the number of total edges, and Diff(i) represents the difference between the local maximum and minimum of edge i.

The sharpness of an edge can be estimated based on the difference between its local maxima and minima [46]. Thus, the proposed system estimates the local maxima-minima differences for all edges and calculates the sum of differences. Then the average difference is calculated by dividing it by the number of edges. A sharp edge produces a spike-line intensity plot; thus, the local maxima-minima difference will be lower for a sharp edge, usually less than 3. Alternatively, an unsharp edge will have a higher local maxima-minima difference, usually higher than 5. Thus, an image with less than 5 degradation value is considered sharp or good. Image quality is considered a crucial factor for automated analysis using medical images [47]; thus, the proposed system only uses those images whose sharpness is less than 5. After that, the proposed system transforms the RGB into sRGB space to compensate for the color variations of the image. Then, the image is fed into the DoubleU-Net-based segmentation network to segment the polyps, if any. Finally, the segmentation module generates an image ROI for each polyp-detected region, which serves as the input for the classification module.

The classification modules resize the polyp-ROIs to 224 × 224 pixels for the ViT-based classifier network. This network classifies each ROI as hyperplastic or adenomatous polyps.

Algorithm 1 Polyps Segmentation and Stage Classification

Input: I_{Frame}, $\text{PolypSeg}_{\text{Net}}$, $\text{PolypClass}_{\text{Net}}$
I_{Frame}: Colonoscopy Image Frame
Q_{th}: quality threshold
Black_{th}: Black pixel threshold
$\text{PolypSeg}_{\text{Net}}$: Trained DoubleU-Net for colorectal polyps segmentation with parameters
$\text{PolypClass}_{\text{Net}}$: Trained ViT-Net for colorectal polyps classification with parameters
Initialization:
Load segmentation network $\text{PolypSeg}_{\text{Net}}$
Load classification network $\text{PolypClass}_{\text{Net}}$
$Q_{\text{th}} \leftarrow 5$
$\text{Black}_{\text{th}} \leftarrow 50$
while $I_{\text{Frame}}! = \text{NIL}$ **do**
  $\text{Black}_I = \text{Percentage of black pixels in } I_{\text{Frame}}$
  if $\text{Black}_I \leq \text{Black}_{\text{th}}$ **then**
    $\text{QualityScore}_I = \text{Quality index of } I_{\text{Frame}}$
    if $\text{QualityScore}_I \leq Q_{\text{th}}$ **then**
      Estimate C_{Linear} from I_{Frame}
      if $C_{\text{Linear}} \leq 0.0031$ **then**
        $I_{\text{sRGB}} = 12.92 \times C_{\text{Linear}}$
      else
        $I_{\text{sRGB}} = 1.0552 \times C_{\text{Linear}}^{\frac{1}{2.4}}$
      end if
      Apply $\text{PolypSeg}_{\text{Net}}$ on I_{sRGB}
      $\text{Polyp}_{\text{ROI}} = \text{Polyps segmented ROIs}$
      if $\text{Polyp}_{\text{ROI}}! = \text{NIL}$ **then**
        Resize $\text{Polyp}_{\text{ROI}}$ to 224 × 224
        $\text{Polyp}_{\text{class}} = \text{PolyClass}_{\text{Net}}(\text{Polyp}_{\text{ROI}})$
        if $\text{Polyp}_{\text{class}}==1$ **then**
          $\psi \leftarrow$ Hyperplastic
        else if $\text{Polyp}_{\text{class}}==2$ **then**
          $\psi \leftarrow$ Adenomatous
        end if
      end if
    end if
  end if
end while
return ψ

The details of polyps segmentation and classification are explained in Section III-C and III-D.

C. POLYPS SEGMENTATION METHOD

The proposed method utilized DoubleU-Net [42] network for segmenting polyps from colonoscopy images. DoubleU-Net is a deep convolutional neural network that utilizes the encoder-decoder-based approach like the Vanilla U-net network. However, it stacks one U-net network over another. This enables more efficient use of semantic information by utilizing two encoders and two decoders and is found very effective in segmenting different objects from medical images [43].

TABLE 2. Hyperparameter optimization for fine tuning the DoubleU-Net to segment polyps.

Criteria	Search Space
Epoch	[25, 50, 100, 150, 200]
Batch size	[8, 16, 32]
Learning rate	[0.0001, 0.001, 0.01, 0.03]
Optimizer	[Adam, Adamax, RMSprop, SGD]
Loss function	[Dice Coefficient Loss]

In this study, we optimized the DoubleU-net network to segment polyps from a RGB colonoscopy image of 720 × 570 pixels. The network was trained using 1000 images taken from the Hyper-Kvasir dataset. The RGB image was fed to the VGG19 [44] coupled encoder of the first U-Net of DoubleU-Net. This encoder was pre-trained using the ImageNet dataset [45]. The first U-Net produced a prediction mask for the RGB image. This mask was multiplied by its original RGB image and then fed to the encoder of the second U-Net. The second U-Net then resulted in another prediction mask. The output of the DoubleU-Net network was the final prediction mask (created by concatenating the masks of both U-Nets). The final prediction mask is further prepossessed to mark the segmented polyps on the RGB images, as shown in Fig. 3.

The DoubleU-Net network extracted high-resolution feature maps using Atrous Spatial Pyramid Pooling (ASPP). Data augmentation was employed during the training process by implementing various transformations on the images, including vertical flipping, horizontal flipping, and rotation at 15-degree intervals. During the training, different values of hyperparameters were utilized to achieve the best segmentation accuracy, as shown in Table 2. The combination of best hyperparameters was identified through the grid-based searching of hyperparameters. The DoubleU-Net model achieved superior performance when trained with a batch size of 16, a learning rate of 0.0001, the RMSprop optimizer, the Dice Coefficient loss function, and the Sigmoid output function over a span of 100 epochs.

The proposed polyps segmentation network was tested using two different datasets: 200 images from the Endotect challenge dataset and 1000 images from the Kvasir-SEG dataset. These images were not used for training the network.

D. POLYPS CLASSIFICATION METHOD

The proposed system is designed to detect pre-cancerous polyps to facilitate the treatment of colon and reduce the death rate of colorectal cancer patients. Thus, this system was trained to classify polyps as hyperplastic or adenomatous, both in the benign stage. Although the adenomatous polyps could be of three different types, the subclassification of adenomatous polyps, which requires the estimation of polyps size, was not considered in this study.

The proposed system utilized ViT to classify the polyps using the ROI image from the segmented polyps. The ViT model employed transformer blocks, including feed-forward networks and self-attention layers, to interpret the position information in a structure. Understanding positional information is extremely useful in natural language processing for determining the relevance of a word in a sentence. This process was recently applied to images to understand the positional relationship between pixels and was found very effective, which is not possible using Convolutional Neural Networks (CNNs). The de facto operation of CNN is convolution, which treats all pixels equally and fails to relate spatial distance as it only uses a small portion of an image at a time. ViT was reported to achieve higher classification accuracy for many medical image analysis applications [48].

Thus, in this study, we utilized the ViT model and trained it by fine-tuning the transformer's multi-head self-attention layers (MHSA) and customized the classification (MCTN) head. This approach was motivated by the work of Hossain et al. [48]. In this study, the ViT-B/32 models were used with pre-trained weights to keep the training time short. The ViT-B/32 model uses a 32 × 32 pixels patch as input, making the training faster. In our experiment, we trained only the MHSA layers of the transformer encoded except the feed-forward network (FFN) layers to reduce the training time. Training the FFN layers is time-consuming due to its high number of parameters. Touvron et al. [49] reported that fine-tuning only the MHSA layers improves classification accuracy and reduces training time. We also customized the MCTN head by adding more dense layers to gradually reduce the output layers to two neurons. Then, we trained both the MHSA layers and the customized classification head using the training dataset. The models were trained for 100 epochs using different combinations of hyperparameters. Table 3 shows the values of the hyperparameters explored while training the ViT models. We used a grid-based search to deploy the hyperparameter values.

For the classification experiment, a set of 1800 images were used; 800 images were used for training, 200 for validating and 800 for testing the models. The ViT models generated in the grid searching were compared based on accuracy. The ViT network with the customized MTCN head $1000 - 128 - 64 - 32 - 16 - 8 - 4 - 2$ generated by gradually reducing the neurons of dense layers achieved higher accuracy on the test data.

IV. RESULTS

A. POLYPS SEGMENTATION RESULTS

We generated different versions of the DoubleU-Net model by varying the values of the hyperparameters using grid search, as shown in Table 2. Then, the network with the highest mean dice coefficient in validation was selected for the proposed system (Table 4). Fig. 4 shows the training and validation accuracies for the specified network with RMSprop optimizer, 0.0001 learning rate and 16 batch sizes.

After that, the selected network was applied to the polyps segmentation test dataset. Fig. 7 shows the polyps

TABLE 3. Hyperparameter optimization for training the ViT networks to classify polyps.

Criteria	Search Space
Epoch	[20, 40, 60, 80, 100]
Batch size	[16, 32, 64]
Learning rate	[0.0001, 0.001, 0.01]
Optimizer	[Adam, AdamW, SGD]
Loss function	[Sparse Categorical Cross entropy]

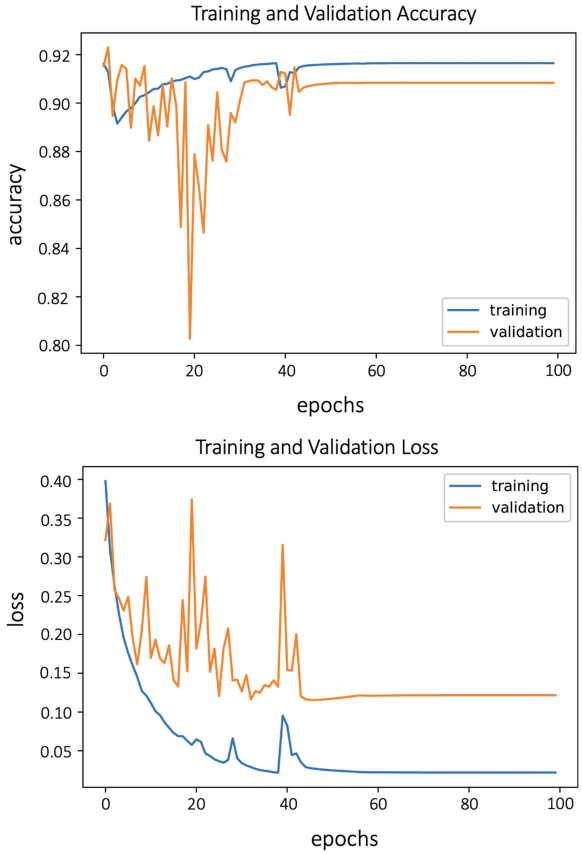

FIGURE 4. DoubleU-Net Model accuracy and loss during training and validation.

TABLE 4. Top six segmentation networks generated from DoubleU-Net.

Network	Optimizer	Learning Rate	Batch Size	Loss	Mean Dice Coef.
1	RMSprop	0.0001	16	0.132	0.910
2	RMSprop	0.001	16	0.140	0.871
3	RMSprop	0.0001	8	0.142	0.865
4	RMSprop	0.001	8	0.170	0.862
5	RMSprop	0.01	16	0.226	0.731
6	RMSprop	0.01	8	0.201	0.677

segmentation result by the proposed method for Kvasir-SEG and Endotect test images. Finally, we compared the segmentation outcomes of the proposed method and other existing methods using the Kvasir-SEG dataset. This comparison is illustrated in Fig. 8. The proposed method outperformed the existing methods regarding the mean dice coefficient and

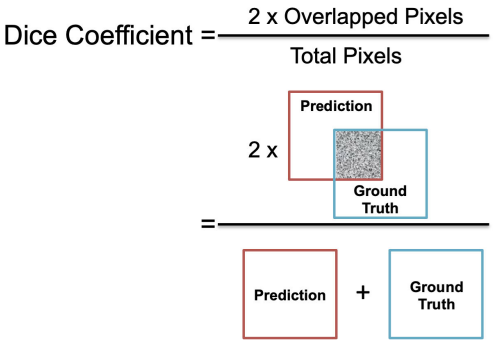

FIGURE 5. Dice coefficient estimation.

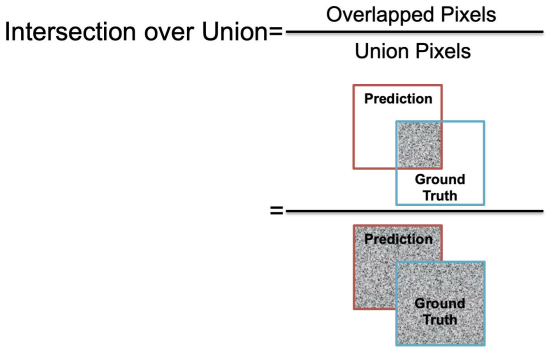

FIGURE 6. Intersection over union estimation.

mean IoU for the Kvasir dataset. The dice coefficient was computed by dividing 2×intersection by the total number of pixels in both images, as shown in Fig. 5. The mean dice coefficient was the mean of dice coefficients for all images. The intersection over union (IoU) was estimated as shown in Fig. 6. The mean IoU is the mean of IoU scores for all images. The Hausdorff distance and average symmetrical surface distance have recently gained popularity in medical image segmentation for measuring the distance between two sets of points belonging to ground truth and predicted output. However, they are computationally expensive. In this study, we used IoU and dice coefficients, which are computationally faster and allowed us to compare the proposed method's results with existing methods. The proposed method was trained using only the Hyper-Kvasir dataset. The Endotect challenge and Kvasir-SEG images were unseen to the network. Thus, the high dice coefficient and IoU scores indicate the robustness of the proposed system.

Further, we compared the performance of the proposed method with the existing methods for the Endotect dataset. However, only the DDAnet [13] presented by Tomar et al. evaluated their method on the Endotect challenge dataset. The proposed method outperformed the DDAnet [13] method, as shown in Fig. 9. We have also experimented with colonoscopy images with no polyps. The proposed method resulted in no false positives in the absence of polyps in the image. This ensured the reliability of the proposed method.

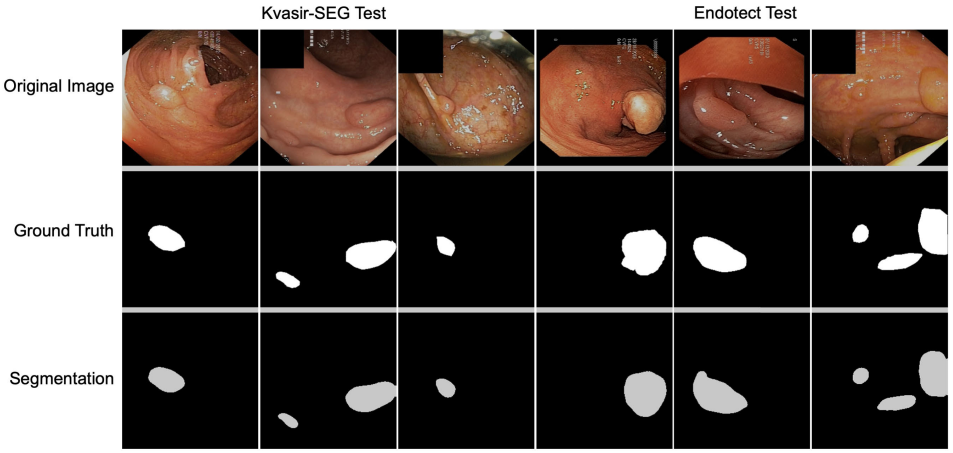

FIGURE 7. Polyps segmentation by proposed method for Kvasir-SEG and Endotect test data.

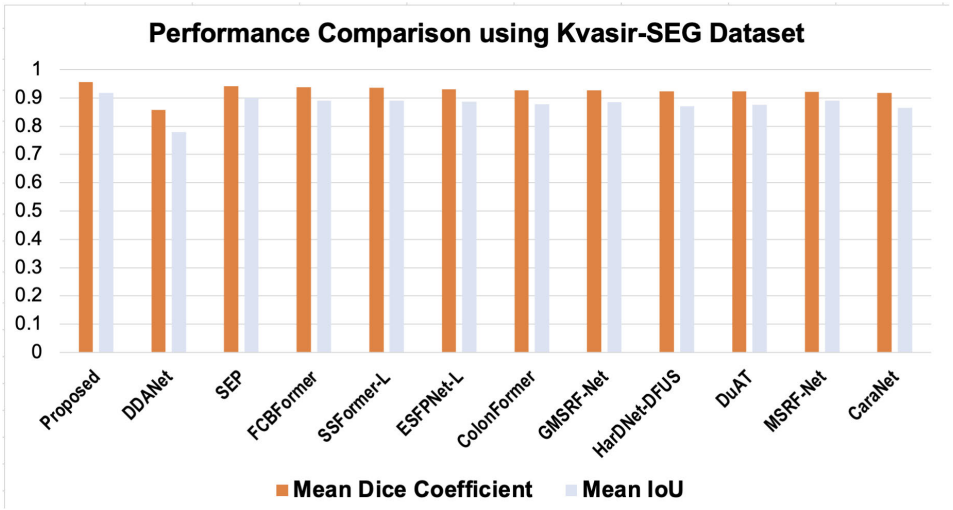

FIGURE 8. Comparison of proposed polyps segmentation methods with existing method using Kvasir-SEG dataset.

B. POLYPS CLASSIFICATION RESULTS

After the segmentation, polyps were classified as hyperplastic or adenomatous to indicate their potential risks. The classification of polyps was achieved by the ViT network, selected from our experiment. The ViT network $1000 - 128 - 64 - 32 - 16 - 8 - 4 - 2$ was chosen as it achieved the highest validation accuracy in the training, as shown in Table 5. The results of Table 5 indicate that gradual reduction in the dense layers is significant in achieving good performance using ViT model. In our experiment, we also found that the AdamW optimizer with decay 0.0001 is more suitable for training ViT models compared to Adam and SGD optimizers. ViT networks achieved the best accuracy using batch size 32. Fig. 10 depicts the training and validation curves of the chosen network. Then, we evaluated the classifier using the unseen test data. Fig. 11 shows the confusion matrix for the test experiment using the Hyper-Kvasir dataset.

The proposed method yielded an accuracy of 99% for the test dataset. The true positive rate (sensitivity) and true negative rate (specificity) were 99.3% and 98.6%, respectively. The proposed method resulted in only a few false positives for both classes. Finally, a comparison was conducted between the classification outcomes of the proposed methodology and those of existing methods. The results are shown in Table 6. The method proposed in this study demonstrated superior performance in terms of accuracy compared to the existing methods. Nevertheless, the proposed method's accuracy was assessed using the public dataset, while the existing techniques were evaluated based on their private dataset.

C. SUBJECTIVE EVALUATION OF THE SYSTEM

Practical usability is a major challenge to the success of computerized autonomous diagnosis systems. Most automated diagnosis systems were found efficient in the validation

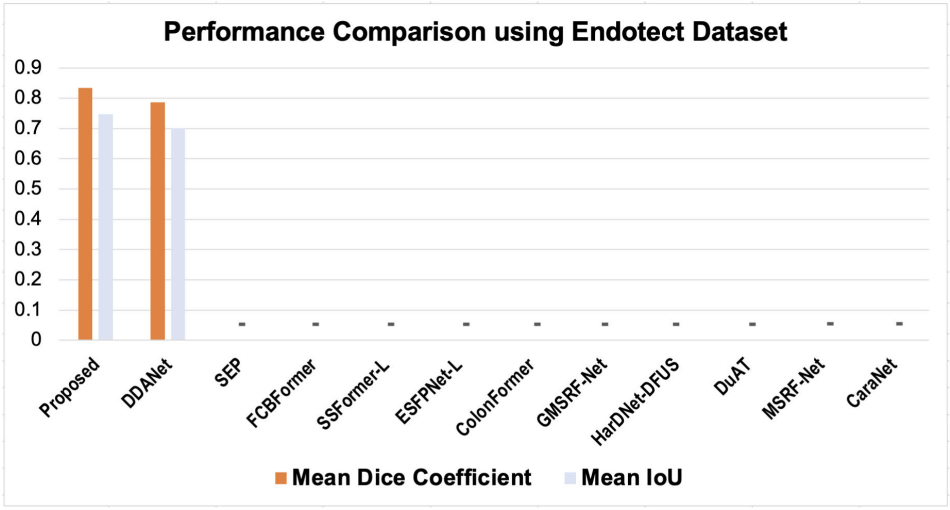

FIGURE 9. Comparison of proposed polyps segmentation method with existing methods using Endotect Challenge dataset.

TABLE 5. Top ten classification networks generated from ViT.

Network	Optimizer	Learning Rate	Batch Size	Dense Layers	Loss	Mean Dice Coef.
1	AdamW	0.0001	32	1000−128−64−32−16−8−4−2	0.203	0.917
2	AdamW	0.001	32	1000−128−64−32−16−8−4−2	0.203	0.907
3	AdamW	0.0001	32	1000−128−64−32 − 16 − 4 − 2	0.244	0.907
4	AdamW	0.001	32	1000−128−64−32 − 16 − 8 − 2	0.260	0.907
5	AdamW	0.001	32	1000−128−64−32 − 16 − 4 − 2	0.241	0.906
6	AdamW	0.0001	32	1000−128−64−32 − 16 − 8 − 2	0.245	0.902
7	AdamW	0.0001	32	1000−128−64−32 − 8 − 4 − 2	0.251	0.902
8	AdamW	0.001	32	1000−128−64−32 − 8 − 4 − 2	0.247	0.899
9	Adam	0.0001	32	1000−128−64−32−16−8−4−2	0.259	0.899
10	Adam	0.001	32	1000−128−64−32−16−8−4−2	0.255	0.894

TABLE 6. Comparison of the proposed method with existing polyp classification methods.

Method	Data	Classes	Accuracy
Ribeiro *et al.* [28]	Private	Healthy or Abnormal	90.96%
Zhang *et al.* [29]	Private	Hyperplastic or Adenomatous	85.9%
Bryne *et al.* [30]	Private	Hyperplastic or Adenomatous	94%
Hsu *et al.* [32]	Private	Neoplastic or Hyperplastic	82.8%
Chung-Ming *et al.* [33]	Private	Hyperplastic or Adenomatous	96.4%
Krenzer *et al.* [34]	SUN dataset (not accessible currently)	Paris classification (Not Dangerous, Dangerous or Cancer)	89.35%
Komeda *et al.* [40]	Private	Adenomatous or Non-adenomatous	75.1%
Lui *et al.* [31]	Private	Curable or Non-curable	85.5%
Bour *et al.* [37]	Private	Not Dangerous, Dangerous or Cancer	87.1%
Tanwar *et al.* [39]	Private	Benign, Non-malignant or Malignant	84%
Ozawa *et al.* [38]	Private	Hyperplastic or Adenomatous	83%
Proposed	**Public (Kvasir, Hyper-Kvasir, Endotect Challenge) and Private**	**Hyperplastic or Adenomatous**	**99%**

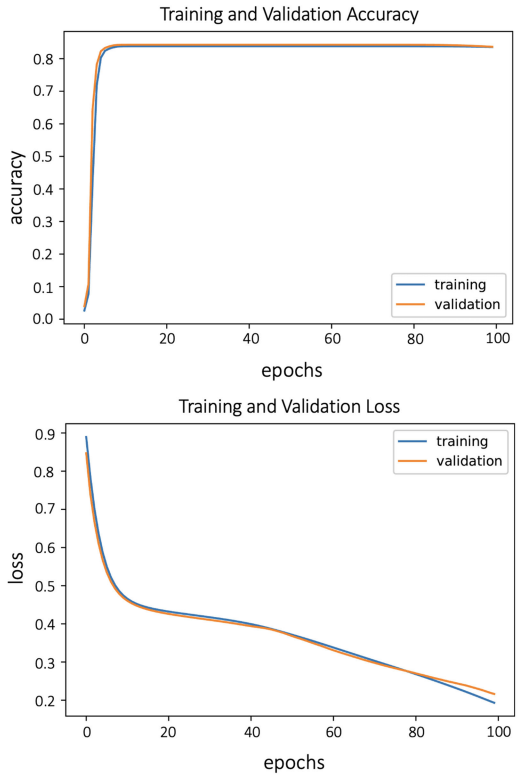

FIGURE 10. Accuracy and loss for the selected ViT network during training and validation.

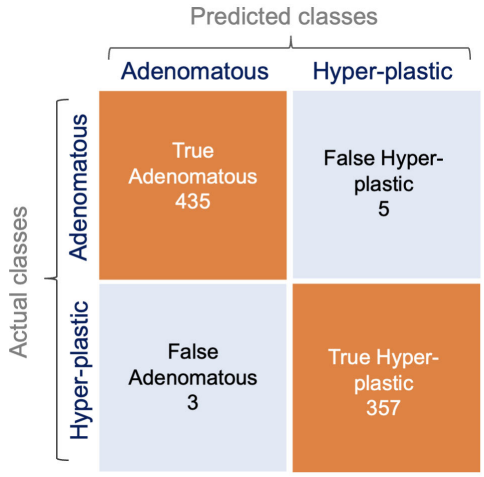

FIGURE 11. Confusion of polyps classification for Hyper-Kvasir data.

experiment; however, they failed to achieve the desired accuracy and reliability when demonstrated for practical application in hospitals. Therefore, in this study, beyond the objective evaluation, we have ensured the practical usability of the proposed system based on the subjective assessment of practitioners by deploying the system in the clinical environment and allowing them to use it. In this demonstration, a gastroenterologist captured colonoscopy videos using an Olympus colonovideoscope machine for 10 selected patients. Then, these videos were used to extract colonoscopic images manually, which was then used by the proposed system. The proposed system then segmented and classified the polyps. Then 3 expert gastroenterologists evaluated the polyps as a team. The results of the proposed system were compared with those obtained by gastroenterologists. The Cohen's Kappa value was 1, suggesting that the experts' score and the proposed system were in complete agreement.

V. DISCUSSION

The proposed system enables automated polyps segmentation and classification using deep learning technology for autonomous colonoscopy examination. This system is designed to detect polyps with potential threats at early stage; thus, this system is expected to improve colorectal disease diagnosis and care. The proposed system outperformed the existing method in terms of polyps segmentation and classification accuracy. It achieved a high accuracy when tested on the unseen polyps dataset, prepared in different labs using different imaging devices. The segmentation method was trained using Hyper-Kvasir data and tested using Endotect challenge and Kvasir-SEG data. The proposed segmentation method outperformed the existing methods in both test cases using unseen public data. The classification method achieved 99% accuracy for the Hyper-Kvasir data compared to the other methods that were tested using private data and accuracy was lower than the proposed method. This demonstrates the generalization ability of the proposed system. Moreover, this system was implemented in a local hospital in Bangladesh and demonstrated using their colonoscopy images. The proposed system resulted in 100% agreement with the experts this ensured practical usability. For automated medical image analysis systems, practical usability is critical [47]. In most cases, method proposed for one environment fails to achieve sufficient accuracy or usability when demonstrated in different settings. The demonstration of the proposed system using a heterogeneous dataset also ensured its robustness. This system relies on a series of pre-assessment techniques such as the black pixel estimation, image quality evaluation and color transformation to ensure robustness.

VI. CONCLUSION

Manual colonoscopy examination takes time, is biased and is vulnerable to fatigue. Automated colonoscopy examination can potentially improve polyp examination and colon care. The proposed method achieved high precision in polyp segmentation and classification, rendering autonomous colonoscopy examination. Demonstrating the proposed system in the clinical setting ensured its routine use in hospitals. In the future, this system can be extended to detect polyps from Gastroscopy or other Endoscopy images.

ACKNOWLEDGMENT

The authors are grateful to Dr. Aulad Hossain, Dr. Rubina Bari, and Dr. Sahria Bakar for their contribution for

reviewing the colonoscopy images and evaluating the proposed system for its routine use in the hospitals.

REFERENCES

[1] F. Stracci, M. Zorzi, and G. Grazzini, "Colorectal cancer screening: Tests, strategies, and perspectives," *Frontiers Public Health*, vol. 2, p. 210, Oct. 2014.

[2] V. Balchen and K. Simon, "Colorectal cancer development and advances in screening," *Clin. Intervent. Aging*, vol. 11, pp. 967–976, Jul. 2016.

[3] W. R. Zipfel, R. M. Williams, and W. W. Webb, "Nonlinear magic: Multiphoton microscopy in the biosciences," *Nature Biotechnol.*, vol. 21, no. 11, pp. 1369–1377, Nov. 2003.

[4] M. S. Hossain, M. M. M. Syeed, K. Fatema, and M. F. Uddin, "The perception of health professionals in Bangladesh toward the digitalization of the health sector," *Int. J. Environ. Res. Public Health*, vol. 19, no. 20, p. 13695, Oct. 2022.

[5] H. Borgli, V. Thambawita, P. H. Smedsrud, S. Hicks, D. Jha, S. L. Eskeland, K. R. Randel, K. Pogorelov, M. Lux, D. T. D. Nguyen, D. Johansen, C. Griwodz, H. K. Stensland, E. Garcia-Ceja, P. T. Schmidt, H. L. Hammer, M. A. Riegler, P. Halvorsen, and T. D. Lange, "*HyperKvasir*, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Sci. Data*, vol. 7, pp. 1–14, Aug. 2020.

[6] S. A. Hicks, D. Jha, V. Thambawita, P. Halvorsen, H. L. Hammer, and M. A. Riegler, "The EndoTect 2020 challenge: Evaluation and comparison of classification, segmentation and inference time for endoscopy," in *Proc. Int. Conf. Pattern Recognit.*, 2021, pp. 263–274.

[7] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. D. Lange, D. Johansen, and H. D. Johansen, "Kvasir-SEG: A segmented polyp dataset," in *Proc. Int. Conf. Multimedia Modeling*, 2020, pp. 451–462.

[8] S. Ali et al., "Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 102002.

[9] S. Ali et al., "An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy," *Sci. Rep.*, vol. 10, no. 1, pp. 1–15, Feb. 2020.

[10] J. Bernal et al., "Comparative validation of polyp detection methods in video colonoscopy: Results from the MICCAI 2015 endoscopic vision challenge," *IEEE Trans. Med. Imag.*, vol. 36, no. 6, pp. 1231–1249, Jun. 2017.

[11] Y. Guo, J. Bernal, and B. J. Matuszewski, "Polyp segmentation with fully convolutional deep neural networks—Extended evaluation study," *J. Imag.*, vol. 6, no. 7, p. 69, Jul. 2020.

[12] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "PraNet: Parallel reverse attention network for polyp segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2020, pp. 263–273.

[13] N. K. Tomar, D. Jha, S. Ali, H. D. Johansen, D. Johansen, M. A. Riegler, and P. Halvorsen, "DDANet: Dual decoder attention network for automatic polyp segmentation," in *Proc. Int. Conf. Pattern Recognit.*, 2021, pp. 307–314.

[14] T.-H. Nguyen-Mau, Q.-H. Trinh, N.-T. Bui, M.-T. Tran, and H.-D. Nguyen, "Multi kernel positional embedding ConvNeXt for polyp segmentation," in *Proc. RIVF Int. Conf. Comput. Commun. Technol. (RIVF)*, Dec. 2022, pp. 731–736.

[15] 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, May 2022.

[16] N. T. Duc, N. T. Oanh, N. T. Thuy, T. M. Triet, and V. S. Dinh, "ColonFormer: An efficient transformer based method for colon polyp segmentation," *IEEE Access*, vol. 10, pp. 80575–80586, 2022.

[17] A. Srivastava, S. Chanda, D. Jha, U. Pal, and S. Ali, "GMSRF-Net: An improved generalizability with global multi-scale residual fusion network for polyp segmentation," in *Proc. 26th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2022, pp. 4321–4327.

[18] E. Sanderson and B. Matuszewski, "FCN-transformer feature fusion for polyp segmentation," in *Proc. Annu. Conf. Med. Image Understand. and Anal.*, 2022, pp. 892–907.

[19] Q. Chang, D. Ahmad, J. Toth, R. Bascom, and W. E. Higgins, "ESFPNet: Efficient deep learning architecture for real-time lesion segmentation in autofluorescence bronchoscopic video," *Proc. SPIE*, vol. 12468, Apr. 2023, Art. no. 1246803.

[20] A. Lou, S. Guan, and M. Loew, "CaraNet: Context axial reverse attention network for segmentation of small medical objects," *J. Med. Imag.*, vol. 10, no. 1, Feb. 2023, Art. no. 014005.

[21] F. Tang, Q. Huang, J. Wang, X. Hou, J. Su, and J. Liu, "DuAT: Dual-aggregation transformer network for medical image segmentation," 2022, *arXiv:2212.11677*.

[22] T.-Y. Liao, C.-H. Yang, Y.-W. Lo, K.-Y. Lai, P.-H. Shen, and Y.-L. Lin, "HarDNet-DFUS: An enhanced harmonically-connected network for diabetic foot ulcer image segmentation and colonoscopy polyp segmentation," 2022, *arXiv:2209.07313*.

[23] L. Zhou, "Spatially exclusive pasting: A general data augmentation for the polyp segmentation," 2022, *arXiv:2211.08284*.

[24] J. Wang, Q. Huang, F. Tang, J. Meng, J. Su, and S. Song, "Stepwise feature fusion: Local guides global," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2022, pp. 110–120.

[25] X. Zhang, F. Chen, T. Yu, J. An, Z. Huang, J. Liu, W. Hu, L. Wang, H. Duan, and J. Si, "Real-time gastric polyp detection using convolutional neural networks," *PLoS ONE*, vol. 14, no. 3, 2019, Art. no. e0214133.

[26] M. Bagheri, M. Mohrekesh, M. Tehrani, K. Najarian, N. Karimi, S. Samavi, and S. M. R. Soroushmehr, "Deep neural network based polyp segmentation in colonoscopy images using a combination of color spaces," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 6742–6745.

[27] Y. Yuan and M. Q.-H. Meng, "Deep learning for polyp recognition in wireless capsule endoscopy images," *Med. Phys.*, vol. 44, no. 4, pp. 1379–1389, Apr. 2017.

[28] E. Ribeiro, A. Uhl, and M. Häfner, "Colonic polyp classification with convolutional neural networks," in *Proc. IEEE 29th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2016, pp. 253–258.

[29] R. Zhang, Y. Zheng, T. W. C. Mak, R. Yu, S. H. Wong, J. Y. W. Lau, and C. C. Y. Poon, "Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 41–47, Jan. 2017.

[30] M. F. Byrne, N. Chapados, F. Soudan, C. Oertel, M. L. Pérez, R. Kelly, N. Iqbal, F. Chandelier, and D. K. Rex, "Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model," *Gut*, vol. 68, no. 1, pp. 94–100, Jan. 2019.

[31] T. Lui, K. Wong, L. Mak, M. Ko, S. Tsao, and W. Leung, "Endoscopic prediction of deeply submucosal invasive carcinoma with use of artificial intelligence," *Endoscopy Int. Open*, vol. 7, no. 4, pp. E514–E520, Apr. 2019.

[32] C.-M. Hsu, C.-C. Hsu, Z.-M. Hsu, F.-Y. Shih, M.-L. Chang, and T.-H. Chen, "Colorectal polyp image detection and classification through grayscale images and deep learning," *Sensors*, vol. 21, no. 18, p. 5995, Sep. 2021.

[33] C.-M. Lo, Y.-H. Yeh, J.-H. Tang, C.-C. Chang, and H.-J. Yeh, "Rapid polyp classification in colonoscopy using textural and convolutional features," *Healthcare*, vol. 10, no. 8, p. 1494, Aug. 2022.

[34] A. Krenzer, S. Heil, D. Fitting, S. Matti, W. G. Zoller, A. Hann, and F. Puppe, "Automated classification of polyps using deep learning architectures and few-shot learning," *BMC Med. Imag.*, vol. 23, no. 1, p. 59, Apr. 2023.

[35] Endoscopic Classification Review Group, "Update on the Paris classification of superficial neoplastic lesions in the digestive tract," *Endoscopy*, vol. 37, no. 6, pp. 570–578, 2005.

[36] M. Misawa, S.-E. Kudo, Y. Mori, K. Hotta, K. Ohtsuka, T. Matsuda, S. Saito, T. Kudo, T. Baba, F. Ishida, H. Itoh, M. Oda, and K. Mori, "Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video)," *Gastrointestinal Endoscopy*, vol. 93, no. 4, pp. 960–967, Apr. 2021.

[37] A. Bour, C. Castillo-Olea, B. Garcia-Zapirain, and S. Zahia, "Automatic colon polyp classification using convolutional neural network: A case study at Basque country," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2019, pp. 1–5.

[38] T. Ozawa, S. Ishihara, M. Fujishiro, Y. Kumagai, S. Shichijo, and T. Tada, "Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks," *Therapeutic Adv. Gastroenterol.*, vol. 13, Mar. 2020, Art. no. 1756284820910659.

[39] S. Tanwar, P. M. Goel, P. Johri, and M. Divan, "Classification of benign and malignant colorectal polyps using pit pattern classification," in *Proc. 4th Int. Conf., Innov. Adv. Eng. Technol. (IAET)*, 2020, p. 6.

[40] Y. Komeda, H. Handa, T. Watanabe, T. Nomura, M. Kitahashi, T. Sakurai, A. Okamoto, T. Minami, M. Kono, T. Arizumi, M. Takenaka, S. Hagiwara, S. Matsui, N. Nishida, H. Kashida, and M. Kudo, "Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: Preliminary experience," *Oncology*, vol. 93, no. 1, pp. 30–34, 2017.
[41] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. D. Lange, D. Johansen, C. Spampinato, D. T. D. Nguyen, M. Lux, P. T. Schmidt, M. A. Riegler, and P. Halvorsen, "Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection," in *Proc. 8th ACM Multimedia Syst. Conf.*, 2017, pp. 164–169.
[42] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, "DoubleU-Net: A deep convolutional neural network for medical image segmentation," in *Proc. IEEE 33rd Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jul. 2020, pp. 558–564.
[43] Md. S. Hossain, G. M. Shahriar, M. M. M. Syeed, M. F. Uddin, M. Hasan, M. S. Hossain, and R. Bari, "Tissue artifact segmentation and severity assessment for automatic analysis using WSI," *IEEE Access*, vol. 11, pp. 21977–21991, 2023.
[44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[45] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[46] H. M. Shakhawat, T. Nakamura, F. Kimura, Y. Yagi, and M. Yamaguchi, "[Paper] automatic quality evaluation of whole slide images for the practical use of whole slide imaging scanner," *ITE Trans. Media Technol. Appl.*, vol. 8, no. 4, pp. 252–268, 2020.
[47] M. S. Hossain, T. Nakamura, F. Kimura, Y. Yagi, and M. Yamaguchi, "Practical image quality evaluation for whole slide imaging scanner," *Proc. SPIE*, vol. 10711, pp. 203–206, Apr. 2018.
[48] M. S. Hossain, G. M. Shahriar, M. M. M. Syeed, M. F. Uddin, M. Hasan, S. Shivam, and S. Advani, "Region of interest (ROI) selection using vision transformer for automatic analysis using whole slide images," *Sci. Rep.*, vol. 13, no. 1, p. 11314, Jul. 2023.
[49] H. Touvron, M. Cord, A. El-Nouby, J. Verbeek, and H. Jégou, "Three things everyone should know about vision transformers," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 497–515.

MD SHAKHAWAT HOSSAIN (Member, IEEE) received the Ph.D. degree from the Tokyo Institute of Technology, Japan. He was a Research Scientist with the Memorial Sloan Kettering Cancer Center, USA. He is an Assistant Professor with the Computer Science and Engineering Department, Independent University, Bangladesh. His research focuses on using machine learning and whole slide image analysis techniques to gain insight into the treatment of cancer patients. He received an offer to join Oxford University, U.K., as a Senior Researcher of machine learning in medical imaging. Recently, his work has been focused on unraveling the role of the human epidermal growth factor receptor 2 (HER2) gene in the growth of different cancers, such as breast, colon, and gastric. He is a member of the Association of Pathology Informatics, USA.

MD. MAHMUDUR RAHMAN received the B.Sc. and M.Sc. degrees in computer science and engineering from American International University-Bangladesh, Bangladesh. His research interests include deep learning, computer vision, and machine learning. His current research interests include image classification, object detection, segmentation, and bio-inspired computer vision.

M. MAHBUBUL SYEED (Member, IEEE) received the Ph.D. degree. He is currently an Associate Professor with the Department of CSE, IUB. He is the Project Coordinator for "Coding-for-All" Initiative, IUB, and the Deputy Director of the RIoT Research Center, IUB. He is also leading the OBE Curriculum Development Committee, CSE, IUB. In the recent past, he has completed his tenure as the Head of the Department of CS, AIUB. He has a long track of teaching and research experience at the tertiary level. During the past 18 years, he served as a Distinguished Faculty and a Researcher in multiple renowned universities both in Bangladesh and abroad. His work experience includes working with the Tampere University of Technology, Finland, as a Researcher and a Faculty Member, with the European Space Agency as a Senior Research Scientist, and as an Architect with Orange Telecom, France. His research work has produced 30 scholarly publications which are published in international ranked forums. During his research career, he achieved several research project fundings, including three European research projects. His research interests include software engineering, software project development and management, the IoT/NB-IoT for intelligent system development, outcome-based education (OBE) curriculum design, development, and automation, and image processing. During his Ph.D. degree, he was awarded twice by Nokia Research Foundation for excellent research.

MOHAMMAD FAISAL UDDIN (Member, IEEE) received the B.Sc. degree in electrical and electronics engineering from the Islamic University of Technology (IUT), Dhaka, Bangladesh, in 2002, and the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from Concordia University, Montreal, QC, Canada, in 2006 and 2012, respectively. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Independent University, Bangladesh (IUB), Dhaka. He is also leading the RIoT Research Centre, IUB, as the Director. Before joining IUB in January 2018, he was working with Tata Communication Ltd., Matawan, NJ, USA, as a Principal Ethernet Engineer, where he worked on designing Ethernet networks. His research interests include optimization theory and its applications.

MAHADY HASAN received the bachelor's degree in computer science from Independent University, Bangladesh (IUB), in 2001, the master's degree in information technology from The University of Sydney, in 2004, and the Ph.D. degree in spatial database from the Computer Science and Engineering Department, University of New South Wales (UNSW), in 2010. Prior to joining IUB as an Assistant Professor, he was with UNSW, Australian Pacific College, and James Cook University, Australia. He has been working at IUB, since 2004. He is currently the Head of the Department of Computer Science and Engineering, as an Associate Professor. He is also the Director of the FAB Laboratory and leading the industrial automation wing at the RIoT Center, IUB. He has published several book chapters, journals, and conference papers in many renowned venues. His research interests include software engineering, big data, data analytics, and engineering education.

MD. AULAD HOSSAIN received the M.B.B.S. degree from the Dhaka Medical College, Bangladesh, and the master's degree in urology. He has completed the Fellow of College of Physicians and Surgeons (FCPS) in surgery. He is currently a Consultant with Bangabandhu Sheikh Mujib Medical University Hospital, Bangladesh.

AMEL KSIBI received the B.S., M.S., and Ph.D. degrees in computer engineering from the National School of Engineering of Sfax (ENIS), University of Sfax, Tunisia, in 2008, 2010, and 2014, respectively. She spent three years at ENIS as a Teaching Assistant, before joining the Higher Institute of Computer Science and Multimedia Gabes (ISIMG) as a permanent Lecturer, in 2013. She joined the Computer Science Department, Umm Qura University (UQU), as an Assistant Professor, in 2014. She joined Princess Nourah Bint Abdulrahman University, in 2018, where she is currently an Assistant Professor with the Department of Information Systems, College of Computer Sciences and Information. Her research interests include computer vision, image processing, deep learning, information retrieval, lifelogging and wellbeing, smart education, smart agriculture, and sustainable environment.

MONA M. JAMJOOM received the Ph.D. degree in computer science from King Saud University. She is currently an Assistant Professor with the Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia. Her research interests include artificial intelligence, cognitive computing, and machine learning. She has published several research articles in her research field.

ZAHID ULLAH (Member, IEEE) received the master's degree in computer science from the University of Peshawar, Pakistan. He was a Researcher with King Saud University, Riyadh, Saudi Arabia, for five years. He is currently a Lecturer with the Information Systems Department, Faculty of Computing and Information Technology (FCIT), King Abdulaziz University, Jeddah, Saudi Arabia. His research interests include cognitive computing, business-IT alignment, IT business values, customer relationship management (CRM), and enterprise resource planning (ERP). He has published several papers in his research areas.

MD ABDUS SAMAD (Member, IEEE) received the Ph.D. degree in information and communication engineering from Chosun University, Gwangju, South Korea. He was an Assistant Professor with the Department of Electronics and Telecommunication Engineering, International Islamic University Chittagong, Chattogram, Bangladesh, from 2013 to 2017. He has been a Research Professor with the Department of Information and Communication Engineering, Yeungnam University, South Korea. His research interests include signal processing, antenna design, electromagnetic wave propagation, applications of artificial neural networks, deep learning, and millimeter-wave propagation by interference and/or atmospheric causes for 5G and beyond wireless networks. He won the prestigious Korean Government Scholarship (GKS) for his Ph.D. study.