Artificial Intelligence Review (2024) 57:7
https://doi.org/10.1007/s10462-023-10621-1

A systematic review of deep learning based image segmentation to detect polyp

Mayuri Gupta[1] · **Ashish Mishra**[1]

Accepted: 1 October 2023 / Published online: 5 January 2024

Abstract
Among the world's most common cancers, colorectal cancer is the third most severe form of cancer. Early polyp detection reduces the risk of colorectal cancer, vital for effective treatment. Artificial intelligence methods such as deep learning have emerged as leading techniques for polyp image segmentation that have gained success in advancing medical image diagnosis. This study aims to provide a review of the most recent research studies that have used deep learning methods and models for polyp segmentation. A comprehensive review of deep learning-based image segmentation models is provided based on existing research studies that are essential for polyp segmentation. Convolutional neural networks, encoder–decoder models, recurrent neural networks, attention-based models, and generative models were the most popular deep learning models which play an essential role in detecting and diagnosing polyp at an early stage. Additionally, this study also aims to provide a detailed classification of prominently used polyp image and video datasets. The evaluation metrics for assessing the effectiveness of different methods, models, and techniques are identified and discussed. A statistical analysis of deep learning models based on polyp datasets and performance metrics is presented, with a discussion of future research trends and limitations.

Keywords Colorectal cancer · Polyp · Deep learning · Medical image segmentation · Kvasir-SEG · Review

1 Introduction

Recent advances in artificial intelligence (AI), particularly deep learning, have enabled automated image analysis to improve its diagnostic performance. Deep Learning (DL), a novel area of research in the field of AI (Spring et al. 2022), is based on the training and uses of the healthcare domain. With the advancement in AI (Castiglioni et al. 2021),

✉ Mayuri Gupta
mayurigupta2010@gmail.com

Ashish Mishra
ashishmishra81@gmail.com

[1] Department of Computer Science & Engineering and Information Technology, Jaypee Institute of Information Technology, Sector-62, Noida-201309, Uttar Pradesh, India

new approaches are coming in the healthcare domain to increase accuracy and efficiency (Palanisamy and Thirunavukarasu 2019). Already, AI has achieved great heights in different sectors of the healthcare domain like cancer, mental health, heart diseases, genetic diseases, COVID-19, and so on (Gupta and Sinha 2022; Gupta et al. 2022). Currently, cancer is a prominent area of research, and it is a complex disease characterized by a variety of genetic and epigenetic variants. Worldwide, it ranks as the second most common cause of death. Every year, about 10 million individuals die due to cancer (Lewandowska et al. 2019). It is estimated that 70% of deaths caused by cancer occur in countries with low to middle incomes. Approximately 3.7M lives could be protected every year through the implementation of appropriate strategies for prevention, early detection, and treatment. There is an estimate that the total financial cost of cancer is US $1.16 trillion per year (Eloranta et al. 2021). There are various types of cancer including colorectal cancer, lung cancer, breast cancer, skin cancer, kidney cancer etc., and in the maximum type of cancers, either research work started or is in the nascent stage. It is estimated that more than one third of all cancers can be prevented if timely measures are taken. Developed countries are doing progress in the segmentation and prediction of cancer.

The methods and techniques of AI for the identification of mutated genes and abnormal protein interactions have the potential to identify these diseases at an early stage. In modern biomedical research, artificial intelligence technologies are also being brought to clinicians safely and ethically (Messina et al. 2022). Likely, AI-aided pathologists and physicians will substantially improve the diagnosis, prognosis, and therapeutic prediction of disease (Fernando et al. 2022). Machine learning (ML) and AI applications in cancer treatment and diagnosis will be the upcoming therapeutic guidance as they will enable the rapid mapping of a new treatment for each patient. Using an AI-based system approach, researchers may communicate in real-time and exchange data online, to accelerate the knowledge of readers (Iqbal et al. 2021).

Nowadays, colorectal cancer (MBiostat et al. 2022) is one such type of cancer in which the research work in current years has increased. Colorectal cancer (CRC) is the result of the growth of this type of tumor in the colon or rectum (Singh et al. 2022). Polyps, which are noncancerous growths, arise from CRCs that form in the mucosa of the colon or rectum. There are two types of polyps: sessile polyps with a broad base and pedunculated polyps with a narrow base. A polyp may occur anywhere in the human body, but it is most common in the nose, urinary bladder, and gastrointestinal tract, particularly the rectum and colon. It is estimated that half of the individuals who undergo colonoscopy who have an average risk of 50 years or older will have polyps (Cufí Jou 2022), with the prevalence being higher in older individuals and among men in comparison to women. However, less than 10% of polyp growth to invasive cancer, a method that normally occurs slowly throughout 10 to 20 years and becomes more likely as the polyps grow.

As of 2022, there are 106,180 new colon cancer cases and 44,850 new colorectal cancer cases in the United States (Gangcuangco et al. 2022). Adults over the age of 50 are most affected by polyps. A total of 17,930 cases are detected in people younger than 50 years of age (Segev et al. 2020), which is equivalent to 49 new cases being diagnosed every day. According to their growth pattern, polyps are classified as either adenomatous (i.e., adenoma), which is the most common cancer precursor, or serrated due to their saw-tooth appearance under a microscope. Cancerous polyps may invade the blood vessels and lymph vessels removing fluid and waste from the colon and rectum if they turn into cancerous polyps.

It is important to conduct screening tests like colonoscopy to detect polyps before they develop into cancer. Additionally, these tests can detect colorectal cancer at an early stage, when there is a good chance of recovery (Yu and Helwig 2022). Though colonoscopy has been successful in reducing cancer rates, the predictable adenoma miss rate is approximately 6–27%, and this rate is expected to increase for other colorectal cancer types. Another reason for the missing polyps is that either the polyp is not in the visible area or it is not known despite being in the visual area due to the fast extraction of the colonoscope. Due to different sizes, shapes, colors, and appearances, it becomes difficult to detect the polyps. Polyp segmentation is crucial in the segmentation of polyps and removing the polyps if it is cancerous (Rahim et al. 2021). Robust use of technology in the segmentation and prediction of polyps can solve this issue. If properly implemented, the accuracy and efficiency of the segmentation of the polyp can be increased. Deep learning is one such field that can make tedious tasks easy for us.

Image segmentation is a challenging technique in polyp segmentation. It is the process of grouping elements of an image that correspond to a similar object class (Chou and Chen 2022). This is also known as pixel-level categorization. Alternatively, it involves the division of images and video frames into many parts or objects. Deep Learning-based Image Segmentation has been effectively applied in the field of remote sensing to segment satellite pictures, including approaches for urban planning and precision agriculture (Dong et al. 2022; Ghosh et al. 2020). Additionally, images captured by drones (UAVs) have been fragmented using Deep Learning-based models, giving the ability to solve vital environmental issues linked to weather change. There are various deep learning image segmentation models like Recurrent-Neural- Networks (RNN), Generative-Adversarial-Networks (GANs), and Convolution-Neural-Networks (CNN) with different efficiency and accuracy, which also can be used for the segmentation of the polyp from the image and video clip. It may not only help in accurate polyp segmentation but also help in increasing efficiency (Pacal and Karaboga 2021).

1.1 Scope and motivation

The study's scope including the nature of feature representation, the feasibility of various deep learning image segmentation techniques and models used in polyp segmentation, standard datasets, and performance metrics in polyp image and video datasets.

The motivation for this study comes from analyzing research studies on image segmentation techniques, and DL-based models. This study used these techniques for a better understanding of the readers in polyp segmentation. Polyp segmentation is already a challenging domain because there is a limited review of the existing problems that includes a comparison analysis of retrieved publications. Most of the review papers on polyps do not properly explain the segmentation techniques in terms of the models and the available datasets on polyps. Some of the review studies discussed the endoscopy domain and some of the papers only represent the summarized study which does not show any significant understanding of polyp. According to our knowledge, there are not properly defined papers that discuss the detailed understanding of the DL-based image segmentation models with a detailed description of image and video datasets. Our approach analyses how the different image segmentation techniques are used in a polyp, a detailed comparative study of all image and video datasets available on a polyp, what type of DL-based segmentation models were used, and shows the performance metrics mathematically in detail. This study motivates researchers to analyze all available datasets on polyp, analyze different performance metrics, and understand a variety of image segmentation techniques. This paper is helpful for those researchers who work in the healthcare domain using AI technologies, especially in colorectal cancer and polyp segmentation. The above factors serve as part of the motivation for this study, and in the following section, the research contribution was discussed.

In this review study, over 100 papers have been reviewed from the last 6 years (2018–2023) on polyp segmentation using deep learning image segmentation models shown in Fig. 1. It provides a comprehensive review including the model architectures, datasets descriptions, and models performance with their key contributions.

This study contributes to the following areas:

- The generalized image segmentation techniques have been discussed including semantic segmentation, instance segmentation, and panoptic segmentation.
- A detailed analysis of deep learning-based image segmentation models on polyp, including convolutional neural networks, recurrent neural networks, attention-based models, encoder–decoder models, and generative models, is presented in this study.
- This study presents a detailed classification of all the image and video polyp datasets.
- According to the existing literature review, this article discusses various evaluation metrics for judging the effectiveness of various methods on polyp segmentation.
- Future research trends and challenges for polyp segmentation have been suggested.

In the remainder of this survey, each section is presented as follows: Sect. 2 represents Literature Search and Review Procedures. Section 3 discusses the image segmentation techniques. Section 4 classifies all deep learning-based image segmentation models used in polyp segmentation. Section 5 shows all image and video datasets used for polyp segmentation. Section 6 reviews the study of the performance of models with their metrics. Section 7 represents the review outcomes and the comparative analysis of the review study. Section 8 discusses the main limitations and future research contributions based on polyp segmentation using DL-based image segmentation models, polyp

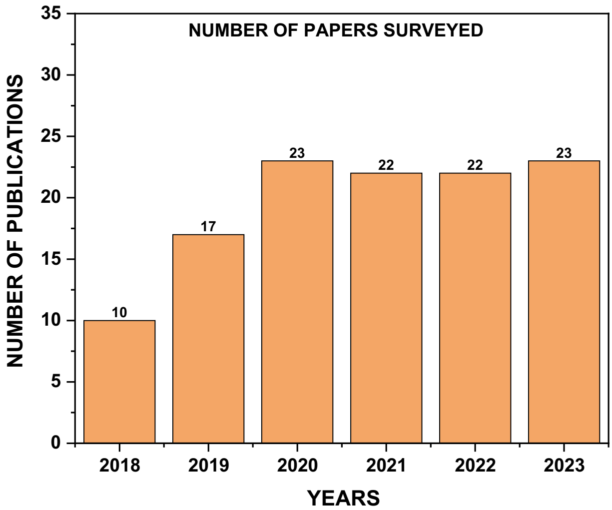


Fig. 1 The publication surveyed data from (2018–2023) years

datasets and performance metrics. Section 9 shows the conclusion of the whole review study.

2 Literature search and review procedures

The purpose of this study was to develop an in-depth understanding of the literature review process encountered by the researchers when working with polyp segmentation and exploration of review study methodologies, based on insights obtained from three perspectives: existing literature review studies on polyp segmentation, identifying and synthesizing prior research approach, and research papers selection criteria.

2.1 Related reviews

There have been many survey papers published in the field of colonoscopy that cover image and video datasets in current years. This section presents a wide variety of previous survey methodologies developed for polyp segmentation. There are limited review papers on polyp segmentation but available in different categories.

In this review paper, Barua et al. presented polyp segmentation technologies based on AI during colonoscopy aiming to improve lesion detection and colonoscopy quality (Barua et al. 2021). To establish the utility of AI-based polyp segmentation approaches for polyp and colon cancer detection, authors presented a systematic survey study and meta-analysis of prospective studies and conducted systematic research in the databases i.e., MEDLINE, EMBASE, and Cochrane CENTRAL. Through the calculation of relative and absolute risks and the difference in means, they compared colonoscopy with and without the use of AI for the segmentation of polyps, adenomas, and colorectal cancer. Gopakumar (2020) reviewed some current works for polyp detection and segmentation in-depth. In the field of healthcare, segmenting polyps from colonoscopy images remains a difficult task. Numerous experiments are used in this situation, and deep learning has demonstrated effective performance over other approaches. Each of the models covered in this study has its own merits and demerits. Additionally, each model has been shown to perform well in a variety of image segmentation tasks. In this paper, Mi et al. evaluate the efficiency of WCE in detecting polyps using AI models like DL, a meta-analysis was carried out because the full studies had small-size samples and various algorithms (Mi et al. 2022). For possibly suitable research published up to December 8, 2021, the search was conducted by two independent reviewers using Embase, the Cochrane Library, PubMed, and, the Web of Science. These papers were then individually evaluated for each image. This meta-analysis was performed with STATA RevMan and Meta-DiSc. Prasath (2017) reviewed several polyp segmentation methodologies for VCE imaging and provided a systematic analysis with the limitations that standard image processing analysis and computer vision systems confront. In this paper, the study was to determine endoscopy imaging which is different from our study on polyp segmentation. Many groups of researchers have been focused on automatic polyp identification throughout the previous decade and M. Luca et al. give a summary of these recent studies (Luca et al. 2019). Deep learning has just begun to be applied in various fields, and try to highlight the relevant applications for colonoscopy in this study. It does not show any relevant work on the polyp except the summary of papers and the applications. In this research (Pacal et al. 2020), the first section of this paper provides an outline of prominent deep learning frameworks used in the study of colon cancer. The studies that relate to colorectal cancer are then grouped under the domain of deep learning and colon cancer. These studies are further subdivided into five groups: detection, survival prediction, segmentation, classification, and inflammatory bowel disease. The research was gathered under each section and then thoroughly summarised and listed. This paper represents the colonoscopy so it does not highlight the polyp in detail as compared to this study. In this review paper, Sánchez-Peralta et al. (2020) show the summary study of 35 research papers. The current systematic review analyses these techniques, indicating advantages and drawbacks for the various categories applied; analyzes performance metrics for seven publicly accessible datasets of colonoscopy images, and discusses future limitations and recommendations. It shows a good study of polyps but it primarily emphasizes the research gaps of the polyp and does not give proper solutions to that research gaps. It indicates only the drawbacks and advantages of the polyp. According to researchers (Wittenberg and Raithel 2020), AI-based detection of adenomas and polyps during colonoscopy has evolved significantly over the last 35 years, beginning with "handcrafted geometrical features" and simple classification systems, progressing to the development and use of "texture-based features" and machine learning algorithms, and concluding with recent growths in the area of deep learning using CNN. This study only shows the review of deep learning-based models which only used CNN but our study shows all the DL-based models in detail including CNN for polyp segmentation. The purpose of this review by Sánchez-Montes et al. (2020) is to provide gastroenterologists with computational methodologies and endoscopic imaging specificities that affect the image segmentation analysis. Currently, the Segment Anything Model (SAM) is receiving considerable attention in both the field of natural and medical image segmentation. In several image benchmark tests, SAM demonstrated superior performance and shows significant potential for the segmentation of medical images. Li et al. (2023) present Poly-SAM, a fine-tuned SAM model for polyp segmentation, and compare its performance with that of several state-of-the-art models. The two transfer learning strategies of SAM with and without finetuning its encoders are also compared. This study (Macháček et al. 2023) presents a conditional diffusion probabilistic model (DPMs) framework for producing synthetic GI polyp images based on generated segmentation masks. According to the experimental results, the system can produce an unlimited number of high-fidelity synthetic polyp images along with the corresponding ground truth masks.

There were several review studies on polyp segmentation mentioned above, but they differ from our work in the following ways:

- Different authors target different organs, but not exclusively polyps.
- Several reviews have been published in other domains that use the biological field to explain their research, which differs from ours.
- Many studies have focused on gastroenterologists or endoscopic imaging for polyp segmentation, but this study reviews the use of colonoscopy for polyp segmentation.
- In most research on polyps, deep learning-based image segmentation models, polyp datasets or performance metrics are not adequately described.
- The classification and comparative studies relevant to polyp datasets have not been properly reviewed in the existing literature.
- A limited literature study discusses performance metrics on polyp segmentation in detail, but our research work correctly determines the commonly used performance metrics by analyzing datasets and deep learning-based image segmentation models.

Our research aims to provide a comprehensive study of image segmentation methods, and techniques, the DL-based image segmentation models, the types of polyp datasets classified as image and video polyp datasets, as well as prominent polyp segmentation performance metrics.

Accordingly, the four key steps for conducting literature reviews were followed in this study as suggested by Chu et al. (2020), Patel and Tyagi (2022). (1) define the search strategy, (2) The data to be extracted from each primary study should be identified by screening, (3) maintain lists of included and excluded studies, (4) final review and the selection of research papers for the review study. An overview of identifying and synthesizing prior research approach is presented in Fig. 2.

2.2 Identifying and search of research studies

The preliminary search process began by identifying papers by keywords and narrowing the search space by including papers published within the last 6 years, i.e., 2018–2023 shown in Step-1 in Fig. 2. To find the relevant papers, a search space strategy was developed. The papers searched were all popular indexed articles, such as Scopus, SCI/SCIE, PubMed, etc., from reputable journals and conferences. The search platform selected was Google Scholar, ACM digital library, Wiley, MDPI, Elsevier, Arxiv, and so on that claim to possess the largest database of research studies and citations. Over a wide range of disciplines, it covers many peer-reviewed conference papers and journal articles. Keywords used for the search were Polyp Segmentation, Deep Learning, and Review study. In Step-1, an initial search yielded 1210 records as shown in Fig. 2.

2.3 Screening of research studies

After Step-1 was completed, in Step-2 the results obtained were filtered to provide a more accurate representation of the research. The studies were chosen further after duplicated

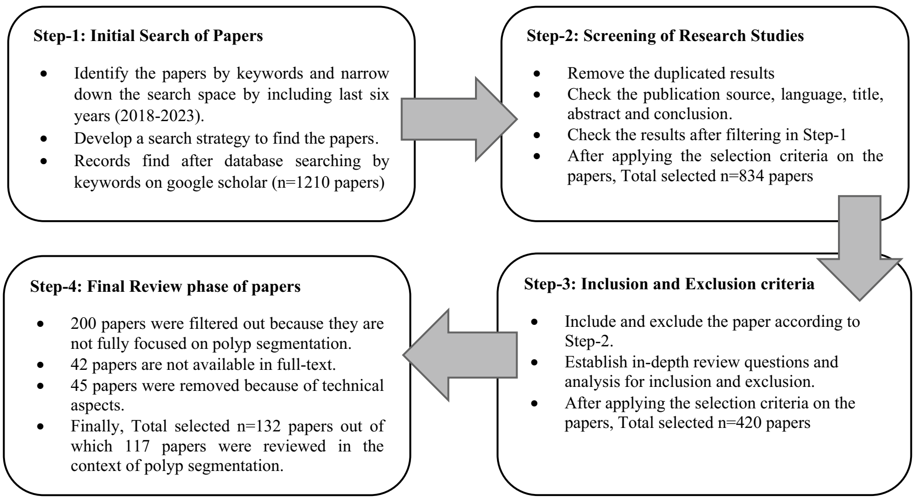


Fig. 2 Identifying and Synthesizing Prior Research Approach

results were removed. During the screening process, the research studies were screened based on the publication source, the title of the paper, the abstract of the paper, as well as the conclusion of the research study. A total of 834 research studies were selected for further analysis after applying the screening process in Step 2 in Fig. 2.

2.4 Inclusion and exclusion criteria

First, most of the papers are excluded by applying the Step-2 in Fig. 2. Now in Step 3, a set of in-depth review questions were developed and used it for data extraction from 834 research studies for inclusion and exclusion criteria. These questions covered information such as the research question, study focus, research method, and key results. The in-depth review questions for data extraction were:

- How does the study approach the proposed research question and what is the main research question it intends to address?
- What is the overall research methodology or the models used in the paper?
- What is the focus and the motivation of the study?
- What are the key findings and the analysis of the study?
- Do the findings of the research study suggest any suggestions or recommendations regarding polyp segmentation models and their performance analysis?
- Which evaluation metrics are commonly used in polyp segmentation like Dice Coefficient, Intersection over Union etc.?

After applying the above in-depth review questions, the 420 research papers were selected in step-3 in Fig. 2 for further analysis of this review study.

2.5 Final review phase of research study

After applying the step-3, from 420 research papers, 200 papers were filtered out because all the papers were not focused fully on polyp segmentation, 72 of the papers were on image classification and localization, 63 papers were on colon cancer and contain polyp but also discuss the other organs with polyp like haemorrhoids, ulcerative-colitis-grade and so on and 65 papers were discussed the polyp segmentation but not clearly understand their

Table 1 Papers collection count and their associated resources

Resources of papers collection	Numbers of papers collected/cited from Journals/Conferences
IEEE	Journals-8, Conferences-23, Transactions-5
Springer	Journals-21, Conferences-7
Elsevier	Journals-25
Arxiv	Journals-10
ACM	Journals-5, Conferences-1
Nature	Journals-6
MDPI	Journals-5
Wiley	Journals-2
Hindawi	Journal-1
Others	Journals-8, Conferences-3, Book Chapters-2

approaches. The 42 papers are not available in full text due to which these papers were removed. Afterward, 45 papers were removed due to technical aspects like they explain methodology and improvements but their technical approach or models were not explained properly with their advantages, issues, and performance metrics. At last, only 132 papers were final for this review study discussed in Table 1.

In Table 1, all publication resources are listed with the number of papers that have been collected from transactions, journals, conferences, and book chapters. IEEE considers to be one of the top publishing sources. This review study only considered papers published in reputable journals and some good conference papers. A total of 132 papers were collected and cited, through which 117 papers were reviewed.

Figure 3 illustrates the distribution of the selected articles for this study across different journal databases. The collection shows that IEEE leads the way with 36 articles. This is followed by Springer, Elsevier, Arxiv, ACM, Nature, MDPI, Wiley, Hindawi etc. with 36, 28, 25, 10, 6, 6, 5, 2 and so on.

One of the challenges of conducting a literature review is to keep track of the sources and citations of the papers that are relevant to the polyp segmentation. To overcome this challenge, a reference management software is used to organize, annotate, and cite the papers easily. It also helps to generate a bibliography of the sources in the format that need for this review paper and save time and effort for managing the citations.

3 Image segmentation techniques

Image segmentation involves the division of an image into parts based on its properties and features. Image segmentation could be divided into two sub-tasks: semantic segmentation, instance segmentation, and panoptic segmentation shown in Fig. 4. Semantic segmentation is used to classify each pixel in the image, whereas instance segmentation classifies each object in the image and panoptic segmentation is a combination of both semantic and instance segmentation. In this section, a detailed description for semantic, instance and panoptic segmentation is given.

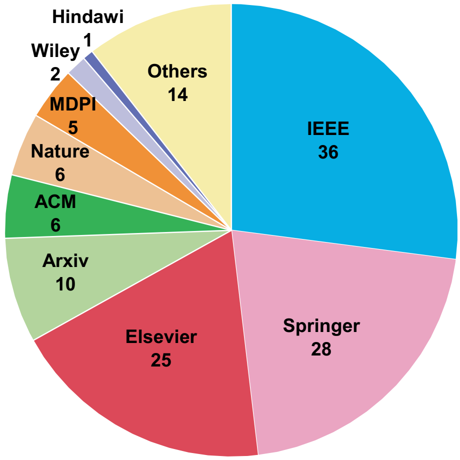


Fig. 3 Distribution of papers by Publication Resources

Fig. 4 Various image segmentation methods (Chen et al. 2021)

3.1 Semantic segmentation

In computer vision, semantic segmentation assigns a class label to every pixel of an input image, which might seem to be an elementary but difficult task. Semantic segmentation is capable of producing class information at the pixel level of images, so several real-world applications are taking advantage of it, including self-driving (Xu et al. 2018), pedestrian detection (Hsu and Yang 2023) and computer-aided diagnosis (Ruan et al. 2023). There have been several methods proposed for semantic segmentation before the advent of deep learning (Chuang et al. 2023) that are characterized using handcrafted features and bottom-up approaches. As deep learning has become more advanced, that have shown remarkable performance over traditional methods. Several semantic segmentation techniques are commonly used, including SegNet (Badrinarayanan et al. 2017), U-Net (Safarov and Whangbo 2021; Al Jowair et al. 2023), DeconvNet (Mukherjee et al. 2019), and FCNs (Yun Guo and Matuszewski 2019; Wen et al. 2023).

3.2 Instance segmentation

The term instance segmentation refers to the assignment of different labels to different instances of an object belonging to the same class. Thus, instance segmentation may be considered a technique for dealing with both object detection (Padilla et al. 2020) and semantic segmentation problems (Yanming Guo et al. 2018) simultaneously. It has emerged as one of the most significant, sophisticated, and challenging topics in computer vision research. Instance segmentation has a significant impact on robotics, autonomous driving, surveillance, and many other areas (Alfred Daniel et al. 2023). The introduction of deep learning led to the development of several frameworks for instance segmentation, notably CNNs, in which segmentation accuracy increased rapidly. There are three types of instance segmentation methods: multi-stage methods, single-stage methods, and semi-supervised or weakly supervised methods. Some of the most commonly used instance segmentation techniques are Mask R-CNNs (Cong et al. 2023; Gupta et al. 2023; Manshadi et al. 2023), Faster R-CNNs (ELKarazle et al. 2023; Manshadi et al. 2023), PANets (Quan et al. 2023), and YOLACTs (Hafiz and Bhat 2020).

3.3 Panoptic segmentation

The panoptic segmentation technique is proposed as a solution to the lack of scene understanding in semantic segmentation and instance segmentation (Chuang et al. 2023). According to its basic concept, panoptic segmentation involves identifying and segmenting the stuff and things within an image. The segmentation results of stuff and things are distinguished by different colors (Elharrouss et al. 2021). Comparatively, to semantic and instance segmentation, panoptic segmentation examines each component of the entire scene globally, significantly improving the ability to perceive the scene, making it a suitable tool for automating driving, video surveillance, and analyzing medical images. Due to the perceived defects in both semantic segmentation and instance segmentation, early panoptic segmentation algorithms are direct combinations of the semantic segmentation model and the instance segmentation model. Mask R-CNN is the most used method for panoptic segmentation (Li and Chen 2022). The UPSNet (Lei et al. 2023; Mao et al. 2023), FPSNet (Nie et al. 2023; Simeth et al. 2023), and VPSNet (Gopinath et al. 2023; Choudhuri et al. 2023) are some of its backbone architectures.

4 Deep learning-based polyp image segmentation models

There are various forms of deep learning models used in polyp segmentation in previous research. From these researches, there are various DL models were found used in polyp whereas Machine learning models were hardly used in the polyp. Afterward, all techniques, methods, and models' data were gathered. The polyp-based deep learning models categorized into various forms including Convolution based models which consist of Fully convolutional networks (FCNs), Region-based CNN, and DeepLab-or Dilation CNN, Encoder–Decoder-based models consisting of UNet and its variations, Recurrent neural networks-based models consist of RNN and Long-Short-Term-Memory (LSTM), Attention-based models and GAN models.

4.1 Convolutional neural networks-based models

It is a kind of artificial neural network known as a Convolutional-Neural-Network commonly used to recognize and classify images and objects. In image processing, CNNs are widely used for tasks such as segmentation and localization, video analysis, obstacle recognition in autonomous vehicles, and recognition of speech for natural language processing (NLP) (Guo et al. 2019).

In this section, this research work describes the variations of convolution neural networks consisting of Fully convolution networks (FCNs), Region-based CNN, and DeepLab or Dilation CNN elaborately which were used in prior research studies to detect the polyp.

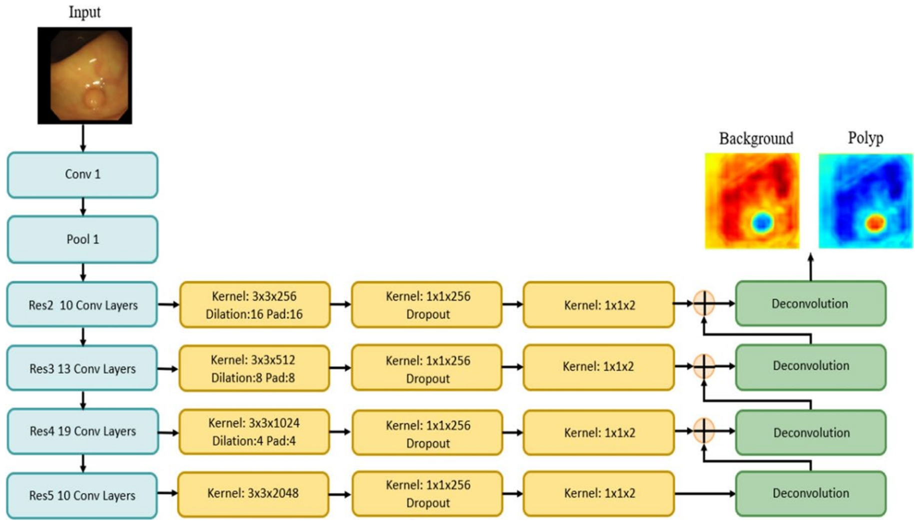


Fig. 5 Polyp segmentation network for Dilated ResFCN. Frome left to right, Blue color shows the Feature extraction part; Yellow shows the Dilation convolution; Green shows the Skip connection (Yun Guo and Matuszewski 2019)

4.1.1 Fully convolution networks

Fully-Convolutional-Networks (FCNs), a significant advancement in DL-based semantic image segmentation models, was introduced by Shelhamer et al. (2017). The FCN with only convolutional layers can provide a segmentation map that is similar in size to the input layer. In this study (Yun Guo and Matuszewski 2019; Yunbo Guo et al. 2020), a deep fully convolutional neural network called Dilated ResFCN is described and represented the architecture in Fig. 5. It was created specifically for polyps' segmentation in colonoscopy images. The benchmark methods FCN8s and ResFCN have been used to compare the Dilated ResFCN model. It has been observed that properly selected dilation kernels can greatly enhance polyp segmentation performance on several evaluation measures. The Dilated ResFCN approach has been proved to be the best method for polyp segmentation with the largest value of the Dice coefficient. It is also effective at matching the polyp's shape to the smallest and most consistent Hausdorff distance value. M. Akbari et al. provided a new polyp segmentation approach based on FCN and Otsu thresholding (Akbari et al. 2018). FCN was used because of its strong capability in semantic segmentation, and it was implemented on the CVC ColonDB database using the Caffe framework, achieving an 81% dice score in this database. A new method for the automatic segmentation of polyps in colonoscopy frames is presented in this study (Wen et al. 2023) which integrates an atrous FCN with a ResNet50 backbone for the proposal of regions and a classification CNN for the refinement of regions.

4.1.2 Region-based CNN

It is well known that region-based CNNs, such as those depicted in Fig. 6, employ a selective search method to identify feasible object regions and then classify these regions on the cropped window to determine the most suitable location based on the probability

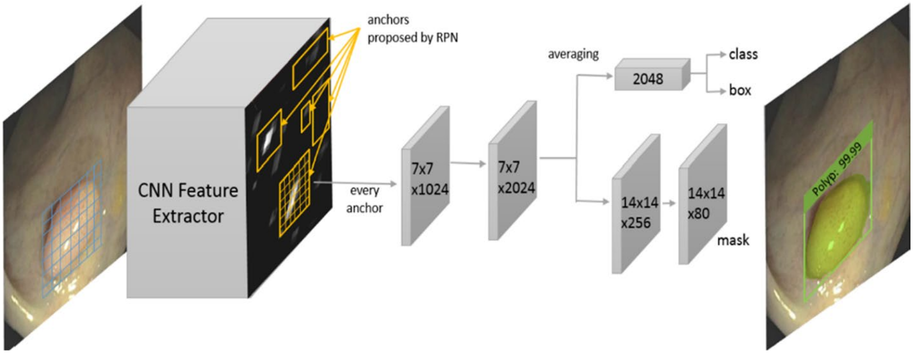


Fig. 6 A framework for R-CNN mask. Resnet-50, Resnet-101, and Resnet-Inception v2 should be evaluated as feature extractors for detecting and segmenting polyps. In the second stage, the confidence value is predicted, as well as offsets and masks within the proposed box (Qadir et al. 2019)

distribution of the output (Girshick et al. 2014). When using the selective search technique (Uijlings et al. 2013), pixels are clustered into objects by analyzing several characteristics such as texture, color, or intensity. Qadir et al. developed and analyzed instance segmentation technique called Mask RCNN (Qadir et al. 2019) for the segmentation of polyp using three new CNN feature extractors, namely Resnet-50, Resnet-101, and Inception-Resnet (v2). In this research, R-CNN was trained to detect and locate polyps in colonoscopy videos using masked region-based convolutional neural networks (Mazumdar et al. 2023).

Bernal et al. divides off-the-shelf polyp segmentation algorithms into three categories: hybrid, end-to-end learning, and hand-crafted (Jorge Bernal et al. 2017). Mo et al. developed an extremely powerful model for polyp instance segmentation (Mo et al. 2018), namely the Faster R-CNN. In compared to the informed findings of state-of-the-art algorithms on polyp segmentation, the experiments determine that the Faster RCNN produces extremely competitive outcomes and is an effective method for medical studies.

4.1.3 DeepLab or dilation CNN

X. Sun et al. provide an innovative end-to-end DL system for polyp segmentation (Sun et al. 2019). The model comprises a decoder to extend the feature maps into a polyp

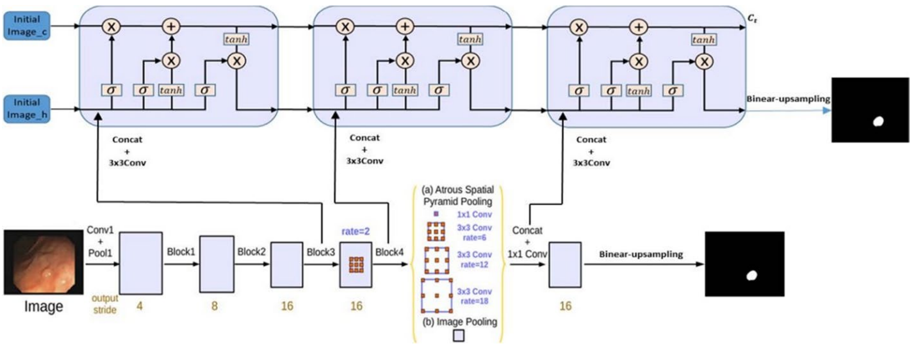


Fig. 7 DeepLab_v3 and LSTM parallel modules (Xiao et al. 2018)

segmentation map and an encoder to retrieve multi-scale information. The authors use dilated-convolution to learn high-level semantic features without degrading the resolution of the encoder, thus improving the encoder's feature representation abilities. Authors (Tomar et al. 2023) present a novel network named DilatedSegNet for the segmentation of polyps in this study. A pre-trained ResNet50 architecture is used along with a dilated convolution pooling block to capture a wider range of reliable and diverse features for better delineation. In this study, authors (Xiao et al. 2018) proposed the DeepLabv3 deep neural network for polyp segmentation in colonoscopy images. The location of polyps may not be efficiently transferred or kept due to their enormous structure. To solve the problem, the DeepLabv3 and LSTM networks were merged to improve the signal of the polyp's position shown in Fig. 7.

4.2 Encoder decoder-based models

In encoder–decoders, data points are transferred from an input class to an output class through a two-stage network.

By using an encoding function, the encoder represents:

$$z = g(x) \tag{1}$$

By compressing the input x, the encoder produces a latent-space representation called z. The decoder, on the other hand, uses a decoding function to:

$$y = f(z) \tag{2}$$

forecasts the output y from z. The latent, or feature (vector), structure captures semantic data from the input image that is used to predict the output.

As a result of using such models, various applications in Natural Language Processing (NLP) and image translation can be derived, such as the de-blurring of the resolution of images, or the segmentation of images.

Mulliqi et al. show several ways for polyp segmentation, but encoder–decoder-based models have recently shown considerable performance (Mulliqi et al. 2020). Colorectal polyp segmentation is performed in encoder–decoder-based models by applying typical CNN architectures in the contracting path with repetitive down-sampling layers such as pooling layers. Down sampling layers enable local translation invariance, minimize representation spatial size, and extract reduced-level feature maps. There are various models of encoder–decoder based on image segmentation in which U-Net (Liu et al. 2021a; Minaee et al. 2022; Ghosh et al. 2020) is the base model for image segmentation in biomedical analysis. Currently, there are limited studies that discuss the classification of encoder–decoder-based models and all variations of U-Net that are used to polyp segmentation in detail. In this literature study, the various models of U-Net are shown to be incredibly efficient and outperform the original U-Net. The various types of models are:

4.2.1 U-Net

Initially, Weng and Zhu (2015) proposed that if an image is sent through an encoder that constantly decreases the spatial size of the feature block, the network will specialize in storing just the most important features and reject the others. Figure 8 represents

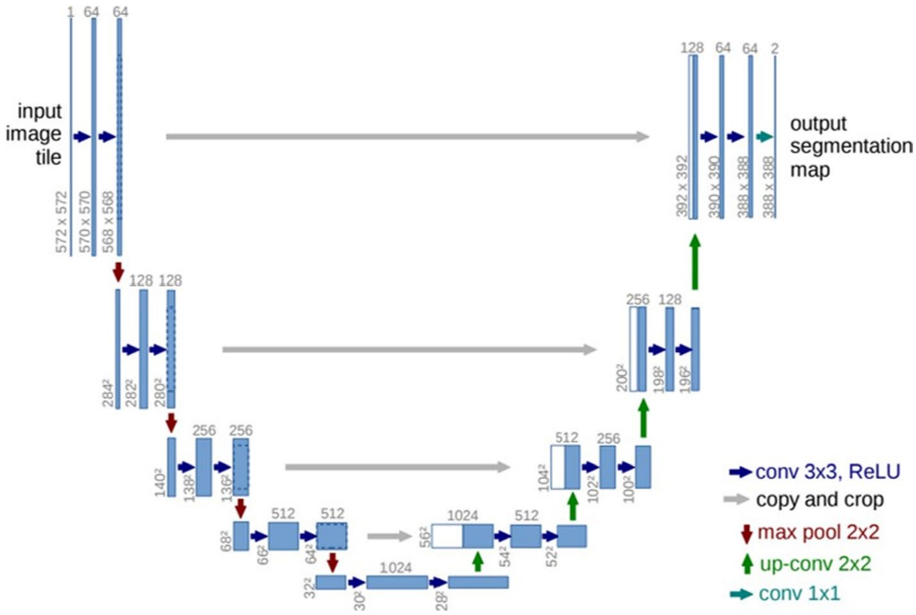


Fig. 8 U-Net Architecture (Weng and Zhu 2015)

the architecture of the U-Net. Tran et al. (2022b) propose a novel model in this study that is based on the UNet model and is known as TDC-Unet. Segmentation of nuclei, polyp, left atriums, and skin lesions are one of the implementations in this research. Tashk et al. implement a novel U-Net framework (Tashk et al. 2019). This framework is applied to the CVC-ClinicDB, the CVC-ColonDB, and the ETIS-Larib, which are international standard optical colonoscopy datasets. In terms of accuracy, precision, recall, and F-Score, the proposed U-Net gives existing competitive techniques for automatic polyp segmentation based on the results of the implementation and evaluation. Specialists and physicians can use it to locate polyps more accurately and efficiently.

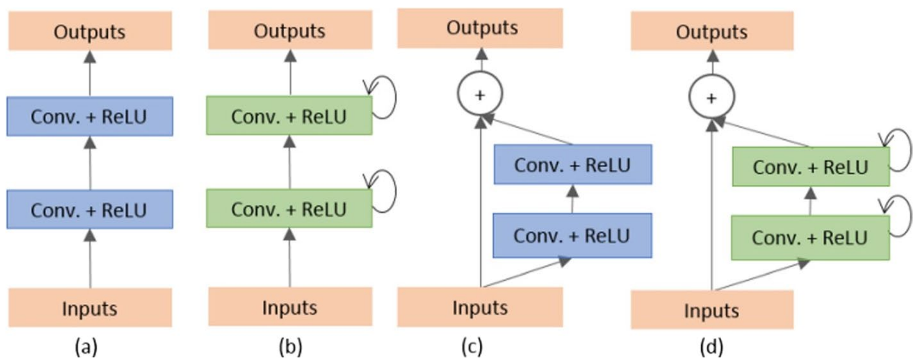


Fig. 9 Various versions of convolutional and recurrent convolutional networks **a** Forward convolutional network, **b** Recurrent-convolutional-network, **c** Residual-convolutional network, **d** Recurrent-Residual-convolutional-network (RRCN) (Alom et al. 2018)

4.2.2 R2U-Net

Md Zahangir et al. first introduced the RCNN and a recurrent-residual-convolutional-neural network (RRCNN) based on UNet models, known respectively as RUNet and R2UNet (Alom et al. 2018). In this, four various architectures are analyzed (Fig. 9). A novel model based on the recurrence residual UNet was proposed by Tran et al. (2022a). By reusing convolutional units, the proposed model reduces the network size and achieves better results. It is difficult to optimize feature maps due to the large number of filters included in the convolution node. To decrease the network size while maintaining the quality of the feature map, double blocks have been used. In comparison with the current model, the introduced model yields improved outcomes with a reduced size. The outcomes determine the effectiveness of the introduced model in polyp segmentation.

4.2.3 Attention-based Unet

The attention gate model described in Fig. 10 was first introduced by Oktay et al. (2018). Sang et al. introduce AG-CUResNeSt, a unique neural network architecture that improves Coupled UNets with enhanced techniques for integrating skip connections and attention gates (Sang et al. 2021). To suppress the redundant low-level information from the encoders, attention gates are implemented into skip connections within each UNet. To achieve high accuracy in polyp segmentation, the network is capable of efficiently merging multi-level characteristics and utilizing semantic information flow.

4.2.4 ResUnet

The ResUnet network is made up of stacked layers initially proposed by (Diakogiannis et al. 2020). Figure 11 shows the architecture of the ResUnet-a d6 network in which the encoder is situated on the left (downward) branch and the decoder is located on the right (upward) branch. Jha et al. (2019) present ResUNet++, an upgraded version of the ResUNet structure for the segmentation of colonoscopic images, for the development of an automated system for the segmentation of polyps at the pixel level. Jha et al. (2021b) show that

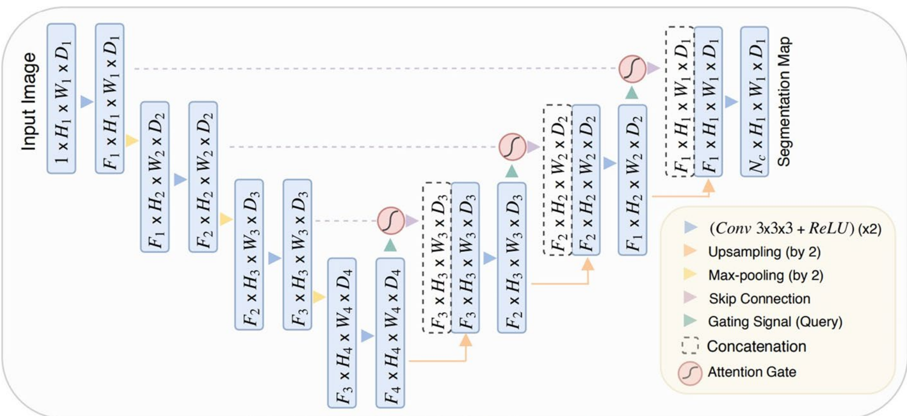


Fig. 10 An illustration of the Attention U-Net segmentation model (Oktay et al. 2018)

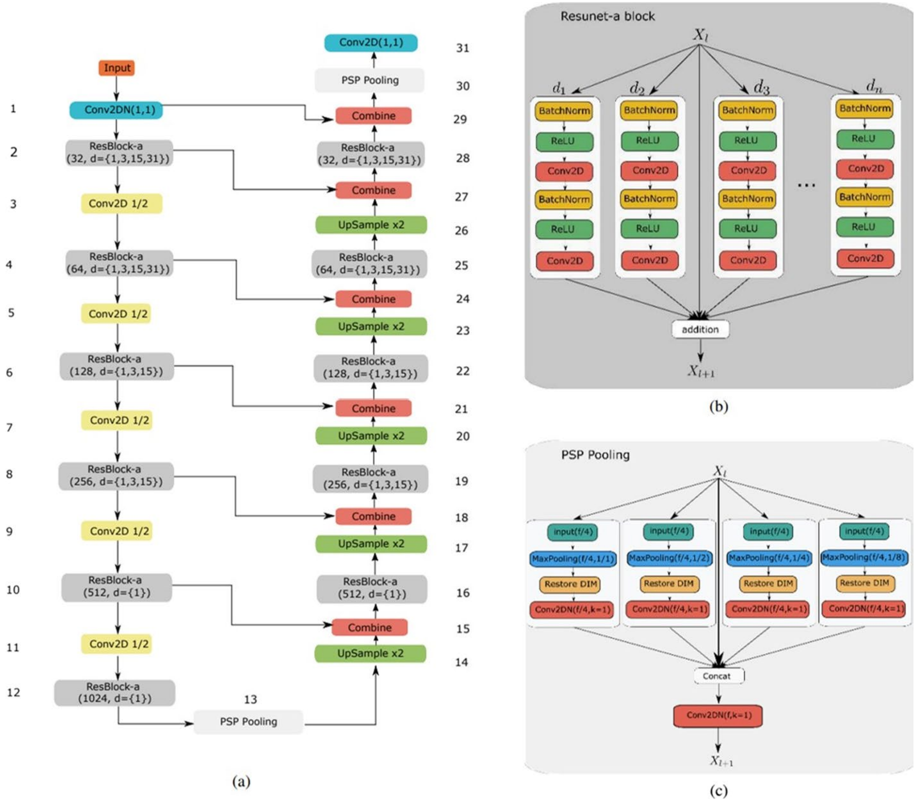


Fig. 11 A layout of the ResUNet-a d6 network. **a** The architecture's encoder is located on the left (downward) branch. The decoder is located on the right (upward) branch. There are as many channels in the final convolutional layer as there are distinct classes. **b** A ResUNet-a network building block. In the residual block, each unit has the same number of filters. Here, d1,…, dn represent a range of dilation rates. **c** Pooling layer for pyramid scene parsing. Pooling occurs in 1/1, 1/2, 1/4, and 1/8 sections of the original image (Diakogiannis et al. 2020)

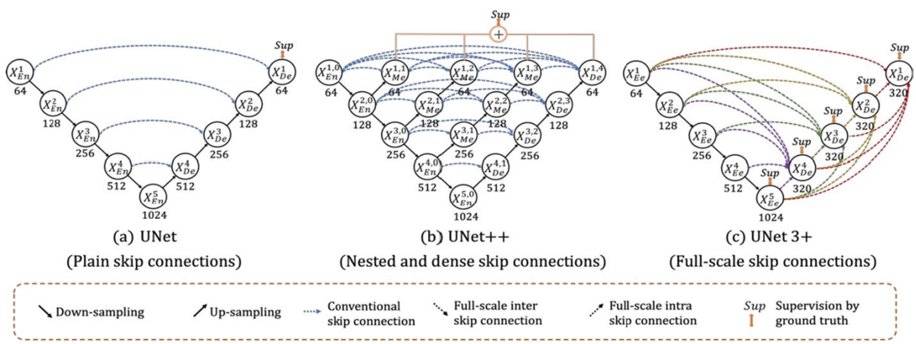


Fig. 12 Differentiation of **a** UNet, **b** UNet++, and **c** UNet3+ architectures (Huang et al. 2020)

Conditional Random Field (CRF) and Test-Time Augmentation (TTA) may significantly enhance the overall predicted performance of the Resunet++ architecture on colorectal polyp segmentation. As researchers designed the new ResUNet++ architecture, the residual block, ASPP, and attention block served as our inspirations.

4.2.5 UNet 3+

The UNet3+ model was first proposed by Huang et al. (2020), which can decrease network parameters to increase computing efficiency in addition to improving accuracy. Figure 12 represents the comparison of the framework of UNet, UNet++, and UNet3+ in which plain skip connections, nested and dense skip connections, and full-scale skip connections are used. M. Wang et al. provide an improved multi-scale network for efficient polyp segmentation in this research (Wang et al. 2021). It consists of four Local Context Attention (LCA) modules, three Receptive Field Block (RFB) modules, a multi-scale backbone (Res2Net), and a multi-scale linked baseline (U-Net3+).

4.2.6 TransUnet

Chen et al. (2021) introduced that TransUNet is a powerful alternative to Transformers and U-Net for segmenting medical images shown in Fig. 13. Tomar et al. present the TransResU-Net architecture (Tomar et al. 2022), which focusses on the benefits of the transformer's encoder block, residual block, and dilated convolution as its core component for segmenting polyps in real-time colonoscopies. The self-attention network built into the transformer and the dilated convolution block both improve the architecture's performance on polyp segmentation.

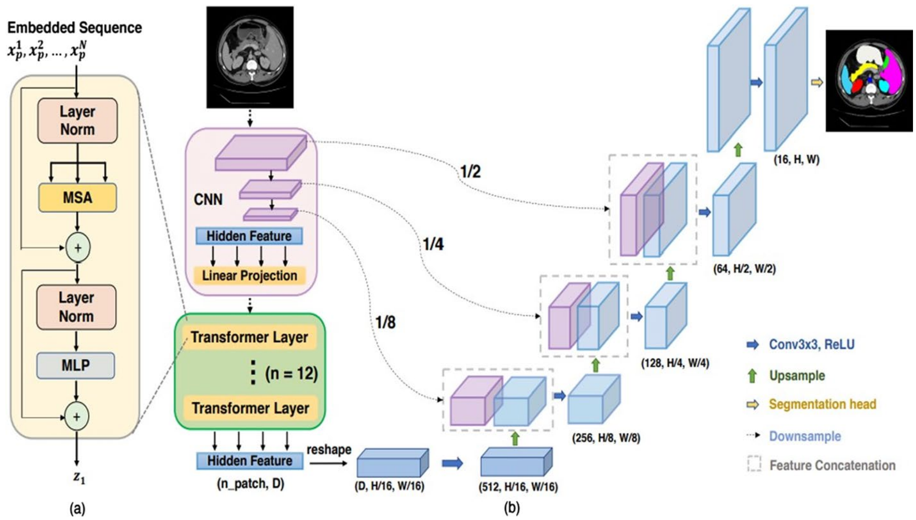


Fig. 13 A description of the framework. **a** Transformer layer layout; **b** Proposed TransUNet framework (Chen et al. 2021)

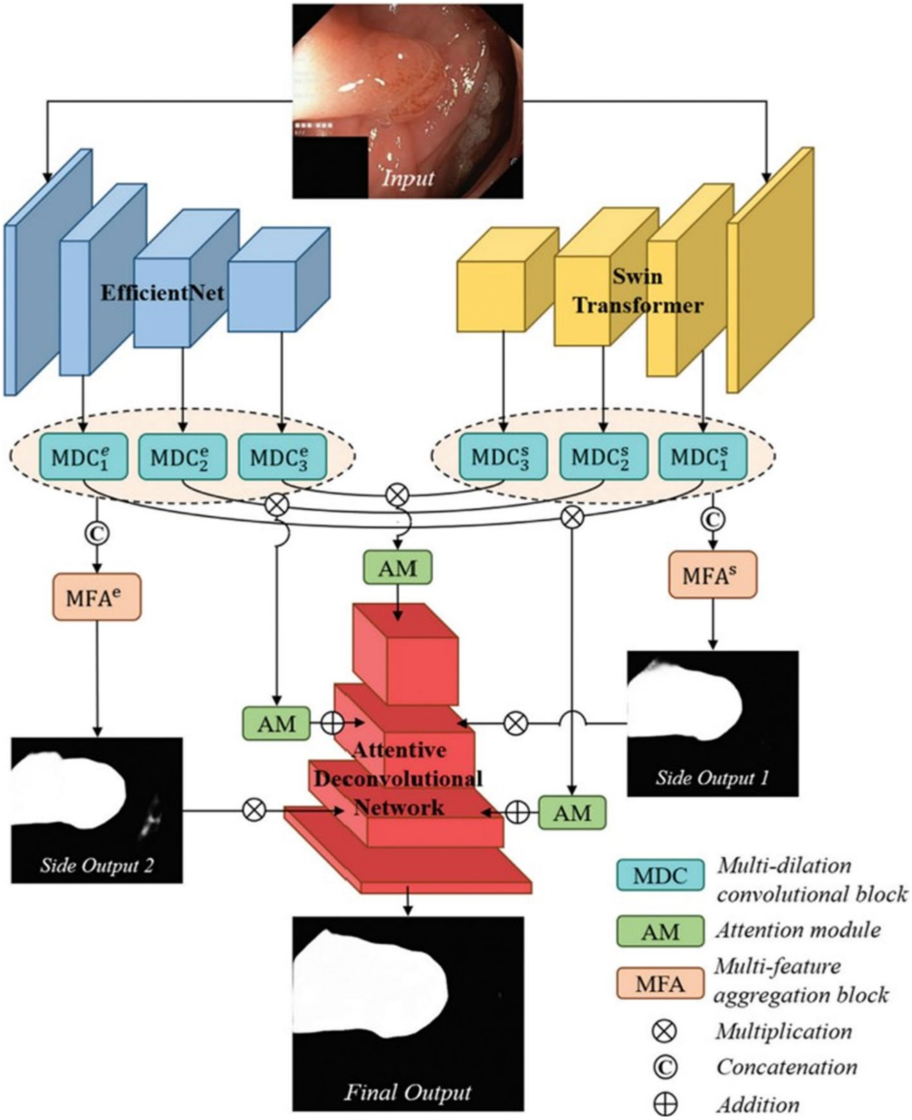


Fig. 14 The architecture of SwinUNet (Cao et al. 2021)

4.2.7 SwinUNet

SwinUnet, is a pure Transformer with Unet-like features (Cao et al. 2021) for medical image segmentation. Swin-Unet is the first entirely Transformer-based U-shaped design, and it comprises encoders, bottlenecks, decoders, and skip connections. The Swin Transformer block is the foundation on which the encoder, bottleneck, and decoder are built as represented in Fig. 14 (Liu et al. 2021b). Park et al. provide SwinE-Net, a new deep

Table 2 Comparison of various Encoder–Decoder-based models

S. no.	References	Models	Datasets	Performance metrics
1	Tran et al. (2022a) Tashk et al. (2019)	U-Net	Colonoscopy, dermoscopy, and	DSC: 90.84, F1-score: 91.25, mIOU: 83.38
			Magnetic Resonance Imaging (MRI)	
			CVC-ClinicDB, CVC-ColonDB and ETIS-Larib	Accuracy: 99.6, Precision:70.2, Recall: 90.9, F1 score: 79.23
2	Zhou et al. (2020) Hoorali et al. (2020)	U-Net++	Nodule segmentation is performed in low-dose CT scans of the chest, nuclei segmentation is performed in microscopy images, liver segmentation is performed in abdominal CT scans, and polyps are segmented in the video of colonoscopy procedures	Intersection over Union of colon polyp (IoU) i.e 30.08 (UNet), 30.14 (wide UNet), 33.45 (UNet++ w/o DS), 33.12 (UNet++ w/DS)
3	Alom et al. (2018)	R2-Unet	Image datasets for colonoscopies:	Dice score:
			CVC-ClinicDB	94.59
			ETIS-LaribPolypDB	92.73
			CVCColonDB	93.31
4	Oktay et al. (2018) Sang et al. (2021)	Attention U-Net	CVC-ClinincDB, ETIS-Larib, CVC-ColonDB, and Kvasir-SEG dataset	Dice: 0.815, mIoU: 0.730, Recall: 0.863 and Precision:0.825
5	Diakogiannis et al. (2020) Jha et al. (2019) Jha et al. (2021b)	ResUnet /ResUnet++	Databases including CVC-ClinicDB, CVC-ColonDB, Kvasir-SEG, CVC-VideoClinicDB, ETIS Larib Polyp Database, ASU-Mayo Clinic Colonoscopy Video Database	(1) In the Kvasir-SEG dataset, the dice coefficient is 81.33%, and the mean intersection over union (mIoU) is 79.27%. In the CVC-612 dataset, the dice coefficient is 79.55%, and the mIoU is 79.62% (2) For both image and video datasets, Kvasir-SEG provides an average maximum mIoU of 0.6817 and an average maximum DSC of 0.4779. All proposed models have an AUC greater than 0.93
6	Wang et al. (2021)	U-Net3+	EndoScene	Dice:0.900, mIoU: 0.834
			CVC-ClinicDB	Dice:0.923, mIoU: 0.874
			Kvasir-SEG	Dice:0.897, mIoU: 0.842

Table 2 (continued)

S. no.	References	Models	Datasets	Performance metrics
7	Tomar et al. (2022)	Trans-UNET	Kvasir-Seg	Kvasir-SEG dataset:
				0.8884 (DSC)
				0.8214(mIoU)
				0.9106 (Recall)
				0.9022(Precision)
				BKAI-IGH dataset:
				0.9154 (DSC)
				0.8568 (mIoU)
				0.9142 (Recall)
				0.9299 (Precision)
			BKAI-IGH	
8	Park and Lee (2022)	Swin-UNET	Kvasir	Kvasir dataset:
				0.920 (DSC)
				0.870 (mIoU)
			ClinicDB	0.913 (F-measure)
			ColonDB	0.926(Structure measure)
			ETIS	0.963 (Enhanced alignment measure)
			EndoScene	0.024 (MAE)

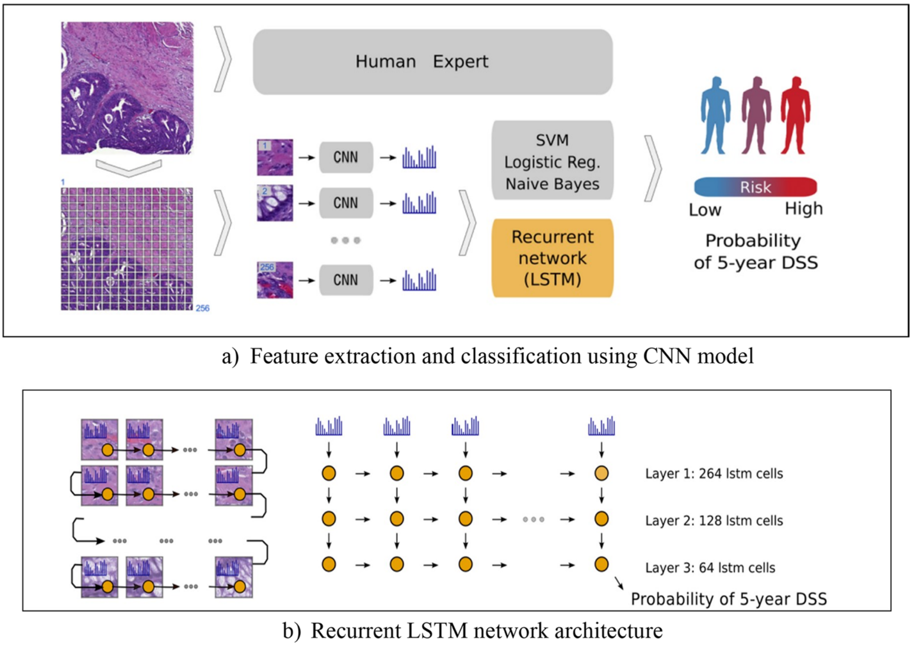


Fig. 15 A description of the pipeline used for image analysis and the LSTM prognosis model. An image of tissue microarray (TMA) areas is defined tile-by-tile using a pre-trained CNN (VGG-16). VGG-16 creates high-dimensional feature vectors for every tile from an input image. These parameters are then fed into classifiers that have been trained to predict five-year disease-specific survival (DSS) (**a**). LSTM Networks calculate the patient risk score based on the whole image of the tissue microarray area based on observed image tiles (**b**) (Bychkov et al. 2018)

learning model for polyp segmentation that successfully integrates a CNN-based Efficient Net and a Vision-Transformer (ViT)-based Swin Transformer (Park and Lee 2022).

Earlier in this section, all encoder–decoder-based U-Net variations with their architectures were discussed. This section describes how these models differ from U-Net and their functionalities.

The comparative study is presented in Table 2 and categorizes all the encoder decoder-based U-net variations relating to polyp datasets and performance metrics.

4.3 Recurrent neural network based models

Recurrent neural networks help to simulate the short and long-term connections among pixels to enhance the segmentation map estimate. In this paper, Bychkov et al. used a mix of convolutional and recurrent architectures (Fig. 15a) (Bychkov et al. 2018) to train the deep learning model to forecast the results of colorectal cancer using visuals of tumor tissue samples. To address two pertinent clinical imaging challenges, (Huang et al. 2018) trained two already existing deep neural networks, SegNet and DeepLab. The first involves locating colorectal polyps within colonoscopy images, while the second involves the delineation of axial lung structures within CT images. The inverted

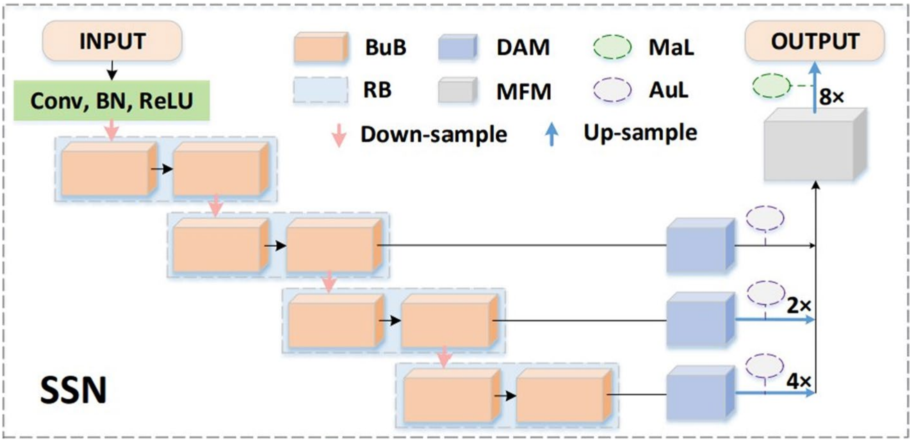


Fig. 16 An architecture of the SSN model. An input consists of polyp colonoscopy images, and an output consists of segmentation masks created. BuB is for Building Block in ResNet, RB stands for Residual Block, MaL stands for Main Loss, and AuL stands for Auxiliary Loss (Feng et al. 2020)

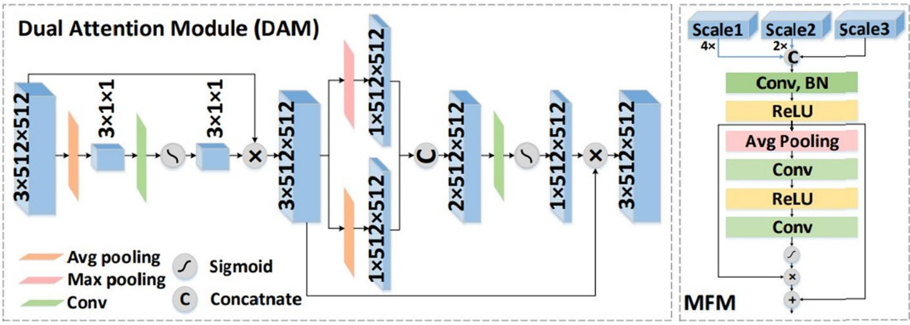


Fig. 17 The fundamental structures of DAM and MFM, with a 3×512×512 example (Feng et al. 2020)

type of the long short-term memory (LSTM) network is merged with the two networks through a parallel connection to improve their segmentation abilities. The adding of LSTM helps segment polyps but not so much for delineating the lung. In this paper (Öztürk and Özkaya 2021), an AI technique is used for accurately classifying datasets of GI Tract Images with a less amount of labeled data. The CNN architecture is used in the proposed AI technique, and each pooling layer's features are transported to an LSTM network represented in Fig. 15 (b). By integrating all LSTM layers, a classification is obtained. GoogLeNet, AlexNet, and ResNet are used in all investigations to properly assess the contribution of the introduced residual LSTM structure.

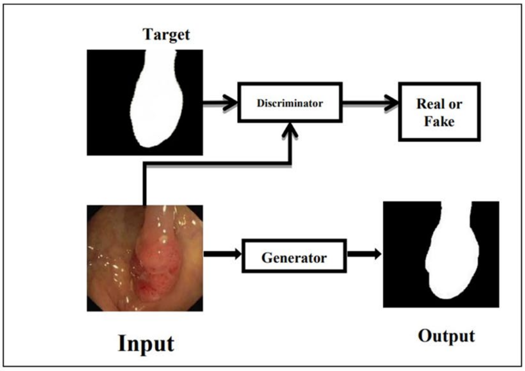
4.4 Attention-based models

Research on attention processes has been ongoing in image segmentation and has evolved through time, thus it is not unexpected to discover researchers that use them for semantic segmentation. Attention models improve on the average and maximum pooling baselines and allow one to visualize the relevance of features at different locations and sizes (Jha et al. 2021b).

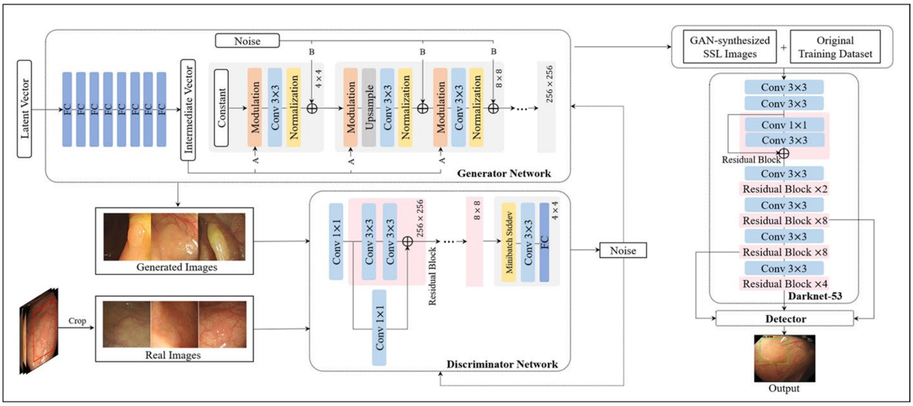
In this study, (Feng et al. 2020) create a new stair-shape network (SSN) for colonoscopy image polyp segmentation in real-time represented in Fig. 16. The novel model performs better for polyp segmentation than U-Net while being substantially faster. As part of the encoder stage, four blocks are used to extract spatial characteristics. The next skip connection is shown in Fig. 17, in which a Dual Attention Module is applied for every block and a Multi-scale Fusion Module is applied to fully fuse features of various scales. A polyp segmentation model based on the proposed method can learn more data due to the abundance of data augmentation and the effective supervision of auxiliary losses. Fan et al. proposed an effective polyp segmentation method which is a challenging process due to two main factors (Fan et al. 2020): (i) polyps of the same type exist in varying sizes, colors, and textures; and (ii) the boundary between a polyp and the mucosa around it is not well defined. So, a parallel reverse attention network (PraNet) was developed for precise polyp segmentation in colonoscopy images to address these issues. A deep convolutional neural network architecture, SR-AttNet, has been proposed in this paper (Alam and Fattah 2023) to facilitate the systematic segmentation of polyps with the help of a utilitarian attention system. Encoder and decoder pipelines use un-dilated and dilated filters to provide local context and depth perception.

4.5 Generative models and adversarial training

A generative adversarial network (GAN) has been proposed (Goodfellow et al. 2020) for training deep representations without requiring lengthy training data with annotations. The back-propagation signals are obtained by involving two networks in a competitive process. According to Z. Qian et al., an innovative polyp segmentation framework was proposed (Qian et al. 2022) based upon two principles: (i) extending the training datasets through the application of a Conditional-Generative-Adversarial Network (CGAN), with the Generator utilizing the Efficient-Spatial-Pyramid (ESP) and the Discriminator employing the Patch-GAN; (ii) improving YOLOv4 structure through dilated convolution and skip connections. J.M. Poorneshwaran et al. investigate the deep generative convolutional framework for polyp segmentation (Poomeshwaran et al. 2019). The pix2pix CGAN was used to investigate polyp segmentation. Due to their flat and inconspicuous shape, sessile-serrated-lesions (SSLs) are a colorectal cancer precursor with a substantially increasing miss rate. Endoscopists might benefit from colonoscopy CADe systems; nevertheless, present methods perform badly in identifying SSLs. To detect disorganized or simply missed polyps, a polyp segmentation technique is used that replicates the morphological properties of SSLs (Fig. 18). A generative- adversarial-network (GAN) was applied to simulate high-resolution complete endoscopic images, including SSL, to construct a well-trained system using unbalanced polyp data. According to the results of this study (Tang et al. 2023), the accuracy of polyp segmentation was greater after implementing the proposed GAN method to generate rare polyp images than after using a common augmentation technique. Results indicated that obtaining a better GAN outcome required a higher sample number as compared with non-GAN training. The sensitivity of colon polyp detection and classification was improved using GAN and DeblurGAN-v2 in hybrid with the YOLOv5 method. In this paper (Yoon et al. 2022), the issue of insufficient labeled data for computer-aided polyp segmentation tasks is taken into consideration. Fan He et al. developed a data augmentation



a)



b)

Fig. 18 **a** Conditional GANs polyp segmentation architecture (Poomeshwaran et al. 2019), **b** GAN is used to build an automated polyp segmentation system with SSL enhancement. "A" denotes a learned affine transform in the generator network, while "B" controls the noise broadcast (Yoon et al. 2022)

framework that uses GAN and the adversarial attack to directly generate false negative colonoscopy images (He et al. 2021).

5 Polyp datasets

Any research and implementation would be incomplete without a dataset. There are various types of polyp datasets available for the research community. Some of the datasets were publicly accessible, and some can be made available on request. In prior research, various polyp image and video datasets were found. In this study, all prominent datasets were classified into two forms: the image polyp dataset and the video polyp dataset. The historical development of the polyp datasets from the discovery to the present period has been comprehensively discussed in this section.

The various polyp datasets have been studied for this study. The prominent datasets were taken from various review studies on polyps and classified into two types: image polyp dataset and video polyp dataset shown in Fig. 19. First, image datasets consist of

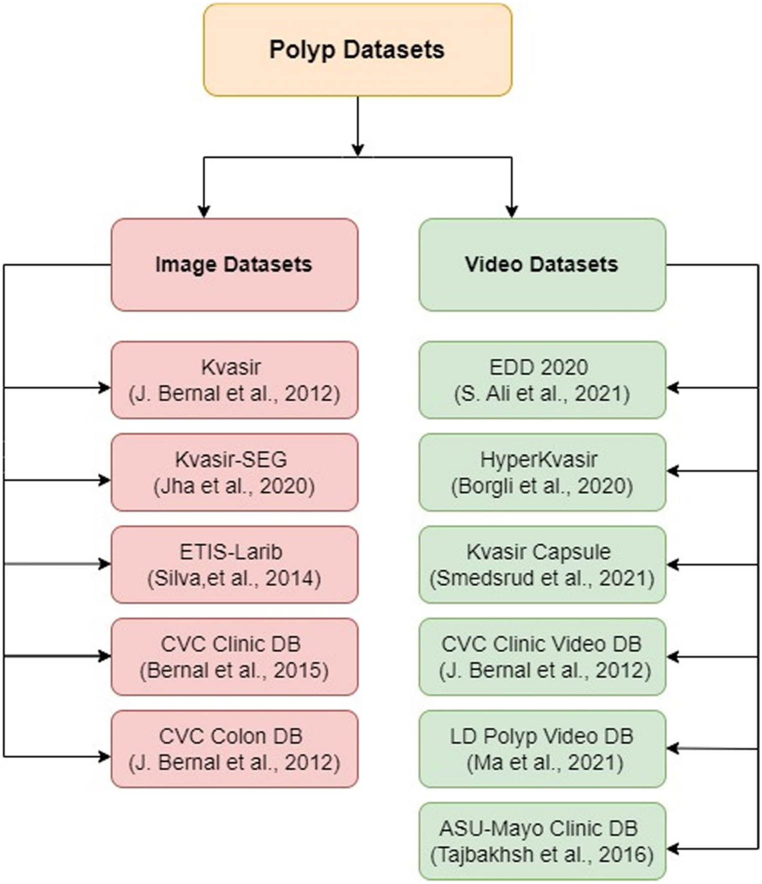


Fig. 19 Classification of all Polyp Datasets

five datasets including CVC-Colon DB (Bernal et al. 2012), ETIS-Larib (Silva et al. 2014), CVC Clinic DB (Jorge Bernal et al. 2015), Kvasir (Pogorelov et al. 2017), and Kvasir-SEG (Jha et al. 2020). Second, video datasets consist of mainly six datasets including CVC Clinic Video DB (Bernal et al. 2012), ASU-Mayo Clinic DB (Tajbakhsh et al. 2016), EDD 2020 (Ali et al. 2021), Hyperkvasir (Borgli et al. 2020), Kvasir-Capsule (Smedsrud et al. 2021), and LD Polyp Video DB (Ma et al. 2021). These datasets were discussed elaborately in further sections. A comprehensive history of polyp datasets has been described in Fig. 20 from the time of discovery to the present (2012–2022). Specifically, the first image polyp dataset was published in 2012, named CVC-Colon DB, and the latest dataset, Kvasir-SEG, was created in 2019. The initial video dataset was introduced in 2012 namely CVC Clinic Video DB (Bernal et al. 2012) and the recent was in the year 2021 consisting of the Kvasir capsule (Smedsrud et al. 2021) and LD Polyp Video DB (Ma et al. 2021). The LD Polyp Video DB was recently launched, so it contains limited research studies. Hyperkvasir and LD Polyp Video DB databases are less used in polyp segmentation.

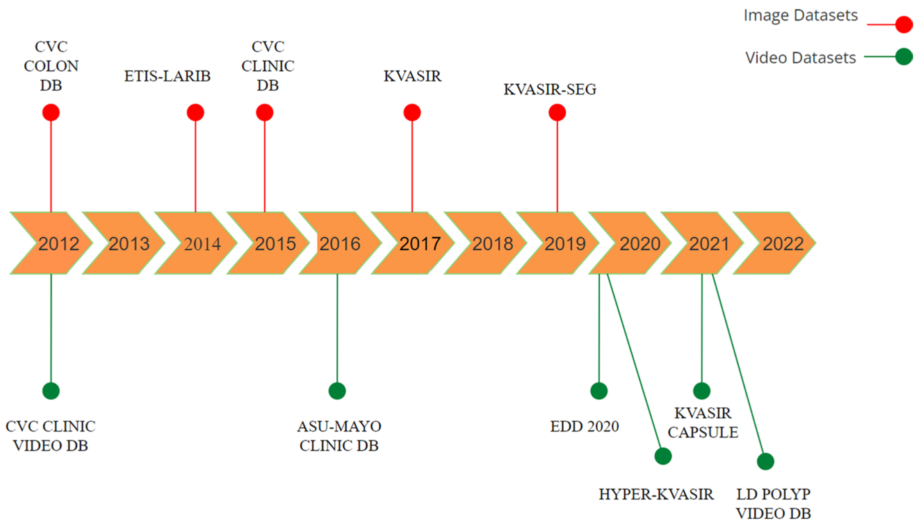


Fig. 20 Historical development of Image and video datasets on Polyp from the year 2012 to 2022

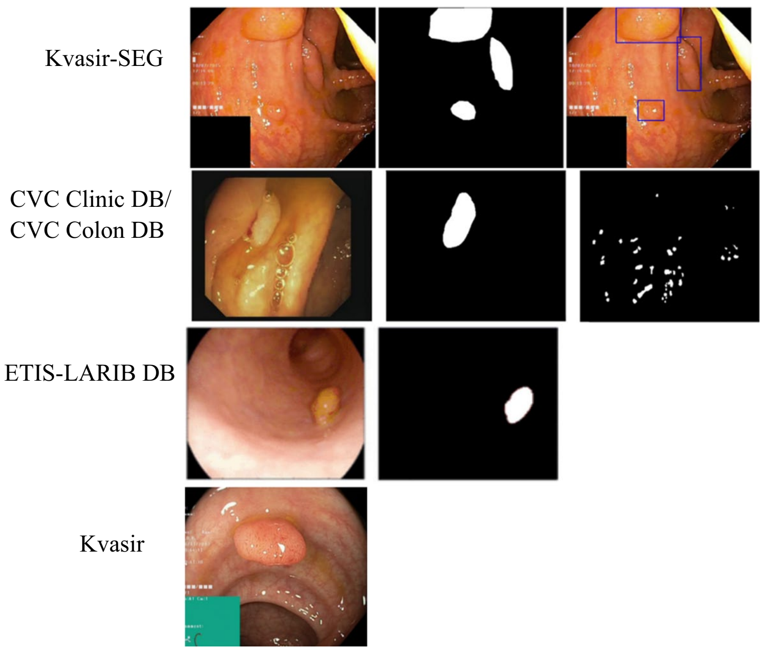


Fig. 21 Available image datasets on polyp

5.1 Image datasets on polyp

In this section, the datasets have been categorized and provide a comprehensive study of the image polyp datasets including CVC-Colon DB (Bernal et al. 2012), ETIS-Larib (Silva et al. 2014), CVC Clinic DB (Jorge Bernal et al. 2015), Kvasir (Pogorelov et al. 2017) and Kvasir-SEG (Jha et al. 2020). All image polyp datasets are openly accessible.

5.1.1 CVC- Colon DB

The CVC-Colon DB was the first dataset proposed for polyp segmentation. It was created at the Computer Vision Center and Computer Science (CS) Department of Universitat Autonoma de Barcelona situated in Barcelona, Spain by Bernal et al. (2012). The dataset contains fifteen random cases in which physicists annotate polyps in all sequences, as well as 20 random frames each with a frame size of 500 × 574 pixels represented in Fig. 21. A cropping procedure was used to remove the non-functioning black borders. By rejecting similar frames, the experts ensured that all 20 frames presented a significantly different angle within the image. Images from different databases were used to collect the data. Consequently, 300 different images must be included in the database to cover the wide variety of polyp presences.

5.1.2 ETIS-Larib

The ETIS-Larib Polyp DB database was developed at the ETIS lab. ETIS (Information Processing and System Teams) is a combined research laboratory of CY Cergy Paris University, ENSEA, and CNRS (UMR 8051) (Silva et al. 2014). There are a total of 1500 images in this database, 300 of which are polyps and 1200 of which are not polyps. The images are labeled by a specialist. According to existing studies, ETIS- Larib DB contains 196 polyp frames from colonoscopy videos with masks of 1225 × 966 and 44 different polyps from 34 sequences shown in Fig. 21. Registration is required to access this database. This dataset is based on the psycho-visual methodology employed by doctors during endoscopic examinations. A total of 196 ETIS-LARIB images have been included in the testing dataset, out of which only 156 images have been used for training with a resolution of 384x384 pixels.

5.1.3 CVC Clinic DB

The CVC-ClinicDB database was created in association with the Hospital Clinic of Barcelona, Spain introduced by Bernal et al. (Jorge Bernal et al. 2015). To create CVC-ClinicDB, 23 separate video studies were combined with white light. Across all research studies, all sequences containing a polyp were retrieved, resulting in 31 sequences, each containing a unique polyp (Fig. 21). There were no frames with exceptional patient preparation or poor visibility quality as a result of visual blurring that was discarded. Therefore, 31 frame sequences consisting of an average of 25 frames were created, with a special emphasis on obtaining as many polyp appearances as possible. The CVC-ClinicDB database contains 612 images of polyps with an overall size of 576 × 768. Ma et al. (2021) represent 612 images with a resolution of 384 × 288 with 29 videos on polyps. Park and Lee (2022) show that the training dataset contains 550 images and the testing dataset contains 62 images from the CVC ClinicDB with a resolution of 384 × 384 pixels.

5.1.4 Kvasir

The Kvasir dataset contains images of the gastrointestinal (GI) tract and was first introduced by Pogorelov et al. (2017). It is gathered using the apparatus called colonoscope at Vestre Viken Health Trust (VV) situated in Norway. The images are categorized according to three major anatomical landmarks and three clinically relevant observations. There are 4000 images in the dataset containing 8 classes displaying anatomical landmarks, pathological findings, or endoscopic processes within the gastrointestinal tract, with a total of 500 images per class. A range of pathological findings can be observed, including esophagitis, polyps, and ulcerative colitis, while anatomical markers include the Z-line, the pylorus, and the cecum. As shown in Fig. 21, the database contains images with resolutions ranging from 720 × 576 pixels to 1920 × 1072 pixels, which have been organized into various folders according to their content. Park and Lee (2022) show 900 training images of Kvasir and 100 testing images in the dataset. As pixel-by-pixel segmentations, the ground truth annotations for each image are included in the dataset. To facilitate training and testing, each image was reduced to 384 × 384.

5.1.5 Kvasir-SEG

Jha et al. proposed the Kvasir-SEG database derived from the prior Kvasir (Pogorelov et al. 2017) database. It is the latest launched image dataset in polyp segmentation given in Fig. 21. The Kvasir-SEG (Jha et al. 2020) is divided into two groups: images and masks. There are 1000 images in each folder interpreted by expert scientists from the Hospital of Oslo University placed in Norway (Jha et al. 2019). A JSON file contains the bounding boxes for the matching images. As a result, the Kvasir-SEG includes a folder of images, a folder of masks, and a JSON file. The image and its associated mask have an identical filename. JPEG compression is used to encode the image files, and web surfing is made easier. All images do have not the same resolution, they vary from 332 × 487 to 1920 × 1072 pixels. The dataset contains 196 polyps less than 10 mm in size that are categorized as Paris class 1 sessile or Paris class IIa (Jha et al. 2021b). Figure 21 represents some Kvasir-SEG dataset images and their corresponding mask from the actual image.

5.2 Video datasets on polyp

In this section, it provided a comprehensive description of the video datasets for polyp segmentation including CVC Clinic Video DB (Bernal et al. 2012), ASU-Mayo Clinic DB (Tajbakhsh et al. 2016), EDD 2020 (Ali et al. 2021), Hyperkvasir (Borgli et al. 2020), Kvasir-Capsule, (Smedsrud et al. 2021) and LD Polyp Video DB (Ma et al. 2021). Every dataset consists of video frames and sequences with their corresponding images. It includes segmented, labeled, and unlabelled images. In comparison to image datasets, video datasets are larger. In existing research, it was observed that the video datasets convert into image datasets to find the results.

5.2.1 CVC-Clinic Video DB

The CVC-ClinicVideoDB database is the first video-annotated database that is publicly available, introduced by Bernal et al. (2012). There are 15 separate standard-definition

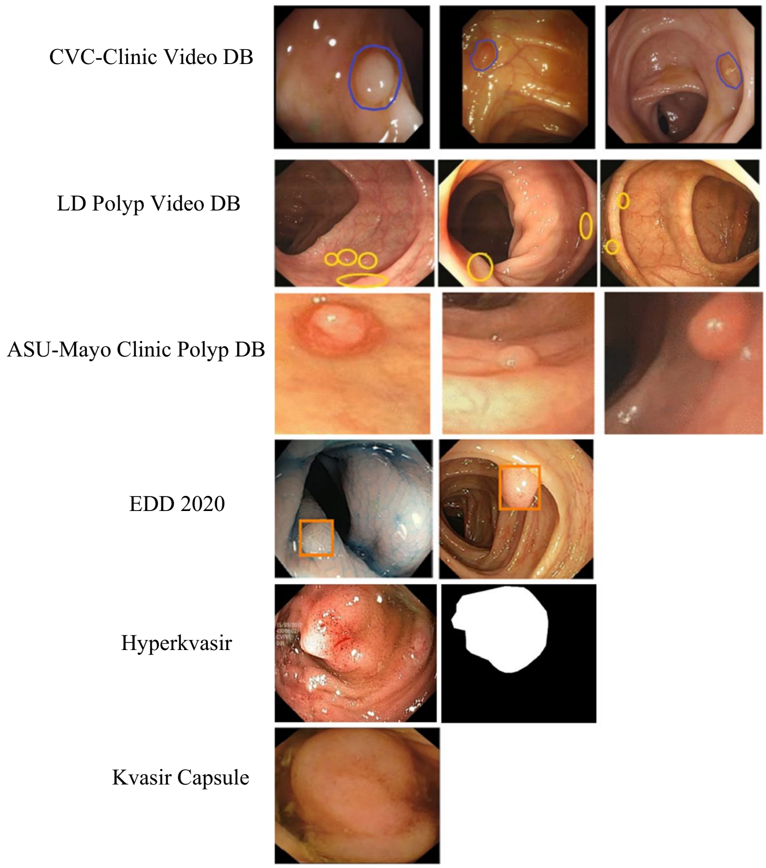


Fig. 22 Available video datasets samples on polyp

(SD) video sequences illustrating a polyp. The sequences were captured using OLYMPUS QF190 endoscopes and an Exera III video-grabber. There are 10924 frames in CVC-ClinicVideoDB, each measuring 768 × 576, of which 9221 have a polyp. For each frame, the ground truth corresponds to a binary image in which white pixels represent all pixels in the image. According to Jha et al. (2021b), only the information contained in the "CVC-VideoClinicDBtrainvalid" folder included ground truth masks. CVC-ClinicVideoDB uses an ellipse to estimate the polyp's border as its ground truth. Figure 22, exhibits various instances of original images and their associated ground truth. Ma et al. (2021) show that there are 18 videos, 18 polyps, and 11945 images in this dataset with a resolution of 560 × 480 pixels. A study published by Ma et al. (2021) found that the CVC-ClinicVideoDB, currently the largest database of colonoscopy videos in existence, was four times smaller than the LDPolypVideo dataset proposed in their study. According to the author, an intelligent annotation tool that utilizes object tracking has been developed to increase the effectiveness of polyp annotation.

5.2.2 ASU-Mayo clinic polyp database

The colonoscopy videos proposed by Tajbakhsh et al. (2016), are known as the ASU-Mayo database. The copyrighted dataset can only be accessed by contacting Arizona State University. It was not available publicly. There were ten videos without polyps and ten videos with polyps in the ASU-Mayo Clinic Colonoscopy Video Database (Borgli et al. 2020), entered as part of the Endovis 2015 sub-challenge. However, due to licensing problems, the test subset is not accessible. A total of 5200 instances of 10 different polyps are contained in 19,400 frames. There are 20 annotated short videos in the online database. The biggest polyp has an area of 29026 pixels, while the smallest has an area of 247 pixels. Polyp radius ranges between 96 and 9 pixels when assumed to be round. The polyp axes range from 8 and 119 pixels assuming an elliptical form with a 3/2 major to minor axes ratio. The median polyp area is 2440 pixels, while the mean polyp area is 3122 pixels. The authors use a scanning window of 120×120 pixels in size since the great majority of polyps have an area of fewer than 10,000 pixels (Figure 22). In the experiments, (El Khatib et al. 2015) display some statistics on the polyps detected in nine of the database's short videos (3477 frames with polyps). Jha et al. (2021b) used all 20 videos for experimentation while training, validation, and testing with an 80:10:10 split on the database ASU-Mayo. Jha et al. (2020) show that the dataset's images are extremely similar to one another, which raises the issue of overfitting.

5.2.3 EDD 2020

Ali et al. (2021) proposed the EDD2020 sub-challenge consists of endoscopic video frames gathered from seven universities across the world, illustrating three distinct modalities, as well as five different organ systems (Ali et al. 2021). For polyp segmentation, endoscopy video frames were annotated. An overall of 280 videos of patients were used in this dataset from various organs and institutes. For this task, 45,478 annotations were executed on frame-by-frame and sequence-by-sequence video recordings. The given training set (Fig. 22) included 385 video frames from 137 distinct patients used in the research studies, each having 817 unique annotations. Among the diagnoses were non-dysplastic Barrett's oesophagus (NDBE), suspicious, high-grade dysplasia (HGD), cancer, and polyps. There are three distinct endoscopic modalities (white light, narrow-band imaging, and chromoendoscopy) that were derived from four varieties of clinical facilities and involved four various gastrointestinal organs (Jha et al. 2021a).

Table 3 Description of the Hyperkvasir dataset

Dataset filenames	Number of files	Explanation	Dataset size
Labeled Images	10,662 images	23 classes of training	3.9 GB
Unlabelled images	99,417 images	Unlabelled records with different classes	29.9 GB
Segmented images	1000 images	Segmentation of original images with their mask	57 MB
Video dataset	374 videos	30 various classes	32.5 GB

5.2.4 Hyperkvasir

Hyperkvasir is the biggest image and video collection of the gastrointestinal system introduced by Borgli et al. (2020). The information is gathered during actual gastro- and colonoscopy tests at Norway's Baerum Hospital from 2008 to 2016 and has been partially labeled by experienced gastroenterologists. A total of 110,079 images and 374 videos are included in the dataset, which depicts both anatomical landmarks and pathological ones. The various organ datasets were included in anatomical landmarks i.e., cecum, ileum, and rectum images, and pathological findings, it includes hemorrhoids, polyps, and ulcerative-colitis organ datasets (Jha et al. 2021a). A total of one million images and video frames have been captured and shown in Fig. 22. Table 3 represents the full overview of the hyperkvasir dataset.

5.2.5 Kvasir-Capsule

The Kvasir Capsule database is accessible through the Open Science Framework and was introduced by Smedsrud et al. (2021) in their research study. The dataset contains 47,41,621 primary data records in total, including 47,238 images with bounding box masks and labels, 43 related labeled videos, and 74 unlabelled videos (Jha et al. 2021a). From all the videos combined, 46,94,266 unlabelled images may be recovered. The dataset is around 89 GB in size. During data uploading, unlabelled images are not included due to excessive data duplication but can be retrieved easily from video files. There are various types of datasets in samples of this dataset represented in Fig. 22. It includes labeled images that comprise archive files for each annotated class of images. Labeled videos database that contains all the videos from which results were retrieved, and an unlabelled videos database that contains all the videos from which results cannot be retrieved.

5.2.6 LD Polyp Video dataset

LDPolypVideo is a large-scale colonoscopy video library containing a wide range of polyps and immensely complex bowel conditions. In a research study conducted by Ma et al. (2021), clinical colonoscopy videos were gathered from normal clinical findings at a public hospital to create a large-scale database of colonoscopy videos. The database has 160 colonoscopy video clips and 40,266 polyp-annotated frames, making it four times in size as the major existing colonoscopy video repository i.e., CVC ClinicVideoDB. All patient-related metadata was eliminated. A professional clinician chooses video segments with polyps. Then resized the video frames to 560×480 after removing the black borders and sampled 160 videos with 40,266 frames each with 200 polyps, ensuring that each video had at least one polyp. There are a total of 33,884 frames with at least one polyp. The LDPolypVideo dataset (Fig. 22) sets a new standard for computer-aided polyp identification and diagnostic research.

5.3 Discussion about datasets

In this section, a detailed description of the datasets with the comparative study of all types of datasets is discussed in Sects. 5.1 and 5.2. The comparative study of datasets shows the

Table 4 Comparative analysis of polyp datasets in existing research studies

References	Dataset	Organ type	Categories of dataset	Dataset size with the pixel of images	Memory size	Method used	Availability of dataset	Dataset type
Pogorelov et al. (2017)	Kvasir	GI Tract	Polyps, esophagitis, ulcerative colitis, Z-line, pylorus, and cecum	8000 images in each class have 1000 images with pixels 720×576	1.2 GB	Classification	Open access dataset	Image dataset
Jha et al., (2020)	Kvasir-SEG	Colonoscopy	Polyp	2000 images (1000 ground truth and 1000 mask value) with 332×487 to 1920×1072 pixels	44.1 MB	Segmentation, detection, and localization	Open access dataset	
Jorge Bernal et al. (2015)	CVC Clinic DB	Colonoscopy	Polyp	672 images with 672 ground truth and masks with 576×768 pixels	258 MB	Segmentation	Open access dataset	
Silva et al. (2014)	ETIS Larib	Colonoscopy	Polyp	196 images of polyps with their associated masks with a resolution of 1225×966	–	Segmentation	Registration for dataset	
Ali et al. (2021)	EDD 2020	Complete GI	Polyp categories, non-dysplastic Barrett's oesophagus, suspicious, high-grade dysplasia, and cancer	386 images derived from 280 patient videos	57.36 MB	Detection, segmentation, and localization	Open access dataset	Video dataset
Borgli et al. (2020)	Hyper-kvasir	Complete GI	16 distinct classes from the upper GI and 24 from the lower GI tract	110,079 images and 374 videos with 224×224 pixels images	66.4 GB	Classification and segmentation	Open access dataset	

Table 4 (continued)

References	Dataset	Organ type	Categories of dataset	Dataset size with the pixel of images	Memory size	Method used	Availability of dataset	Dataset type
Smedsrud et al., (2021)	Kvasir Capsule	Complete GI	Polyp	4,741,621 primary data records in total, including a resolution of 336 × 336 pixels	89 GB	Classification	Open access dataset	
Bernal et al. (2012)	CVC Clinic Video DB	Colonoscopy	Polyp	15 different video sequences in which videos have 10,924 frames with 9221 having a polyp and 380 images of resolution 768 × 576	6.1 GB	Segmentation	Open access dataset	
Ma et al. (2021)	LDPolyp Video	Colonoscopy	Polyp	160 videos with 40,266 frames with a resolution of 560 × 480 each with 200 polyps	24.6 GB	Segmentation	Open access dataset	
Tajbakhsh et al. (2016)	ASU-Mayo Clinic Polyp database	Colonoscopy	Polyp	19,400 frames and a total of 5,200 instances of 10 different polyps with 18,781 images and 247 pixels of polyp in the area	–	Segmentation and classification	By request	

author's description, the type of organ through which the dataset belongs, categories, size of the dataset with the resolution, memory size, the method used, and availability of the dataset (Table 4). This module also describes the existing research study on datasets in detail.

From existing research, it was observed that CVC ColonDB shows good performance but is not better than CVC ClinicDB (Akbari et al. 2018; Alam and Fattah 2023; Feng et al. 2020; Qian et al. 2022; Tashk et al. 2019). (Pozdeev et al. 2019) trained their model using CVC ClinicDB and aimed to forecast segmentation masks for the dataset Kvasir but were unable to provide experimental results due to missing ground truth. Kvasir plays an essential role in training and testing for different image retrieval and objects localization technologies such as search-based systems, video analysis, neural-networks, deep learning, information retrieval, machine learning (Park and Lee 2022; Pogorelov et al. 2017), object identification, computer vision, big data processing and data fusion (Borgli et al. 2020; Jha et al. 2021a). The Kvasir-SEG dataset was designed recently so it was less used and progressed with new and enhanced methodologies for polyp segmentation, localization, and classification. ETIS-Larib is used only for segmentation which shows poor performance in most of the research studies as compared to other datasets (Jha et al. 2020, 2021a, b; Ma et al. 2021; Park and Lee 2022; Qian et al. 2022; Silva et al. 2014; Tashk et al. 2019; Tran et al. 2022a).

In this research literature survey, various video datasets for polyp segmentation were given. CVC Clinic Video DB, in which the ground truth is denoted by a circle or oval. It is clear, however, that pixel-wise annotations of this collection would require significant human labor from skilled endoscopists and engineers. It should be noted that CVC-VideoClinicDB, only used data from the "CVC Video ClinicDB trainvalid" folder since only this data had ground truth masks. CVC-VideoClinicDB is extremely unbalanced, which has resulted in poor performance (Sun et al. 2019; Jha et al. 2021b; Ma et al. 2021), (Bernal et al. 2012; Tajbakhsh et al. 2016). ASU-Mayo Clinic DB contains licensing problems due to which the test subset is not accessible (Borgli et al. 2020). Due to the similarity in images, the dataset shows overfitting (El Khatib et al. 2015; Jha et al. 2020; Jha et al. 2021b; Tajbakhsh et al. 2016). The EDD 2020 consists of four different types of GI organs. It does not target the polyp only, it includes four organs in which polyp is the one part of the dataset (Jha et al. 2021a; Ali et al. 2021). The dataset includes Hyperkvasir in which researchers can quickly start using the database for standard machine learning tasks such as classification using a series of text files and scripts. This dataset is available on GitHub (Borgli et al. 2020). The dataset is further split into three official splits for cross-validation experiments. To keep a fair contrast of results, it is important to maintain a consistent split between methods. Reproducibility and transparency are enhanced by including scripts for generating plots, folding data, and generating annotations. Current research in the field of GI image and video analysis encourages future contributions (Ali et al. 2020). A dataset can be used for comparative analysis and reproducibility of experiments, as well as for publication and sharing of novel data in the future (Jha et al. 2021a; Borgli et al. 2020). Kvasir-Capsule (Jha et al. 2021a; Smedsrud et al. 2021) is similar to Kvasir but it was detected by the instrument capsule and it contains video frames to detect polyps. Currently, less research is being conducted on LD Polyp Video DB. The dataset was four times wider than the CVC Video Clinic DB and should be explored further by the research community to determine its efficiency (Ma et al. 2021).

6 Performance metrics on polyp segmentation

For the polyp segmentation, researchers applied well-established image segmentation performance measures such as Dice Coefficient (DSC), Intersection over Union (IoU), Precision (P), Recall (R), Average Precision (AP), and pixel accuracy. To assess the clinical viability of the segmentation techniques in terms of inference time throughout the test, some authors additionally included Frame Per Second (FPS) for videos (Jha et al. 2021a). The following section summarizes the different evaluation metrics used for judging the performance of various methods found in the literature review for polyp segmentation.

6.1 Dice coefficient

The DSC (Jha et al. 2021a; Shamir et al. 2019) is a commonly used statistic for comparing the pixel-by-pixel outcomes of forecasted segmentation with ground truth. It is described as:

$$\text{DSC }(A,B)=\frac{2\times|A\cap B|}{|A|+|B|}=\frac{2\times TP}{(2\times TP)+FP+FN} \tag{3}$$

where A stands for the anticipated set of pixels and B represents the actual image's target object.

Let TP, FP, TN, and FN stand for true positives, false positives, true negatives, and false negatives, respectively, to define each metric.

6.2 Intersection over union (IoU)

There is a popular statistic in polyp segmentation known as intersection-over-union (IoU). The IoU metric measures the number of pixels common between the target and prediction masks divided by the total number of pixels present across both masks. It ranges from 0 to 1 where, a value of zero indicates that there is no overlap, while a value of one indicates flawless overlap. An average of the IoUs of each class is used to calculate the mean IoU of an image for binary segmentation (two classes) or multi-class segmentation. The overlap between two bounding boxes A and B is determined by calculating the ratio of their overlap areas (Rezatofighi et al. 2019).

$$IoU(A,B)=\frac{A\cap B}{A\cup B}=\frac{TP}{TP+FP+FN} \tag{4}$$

6.3 Pixel accuracy

A model's accuracy parameter measures the model's performance across a variety of classes. When all classes have equal significance, it is helpful. A prediction accuracy rate is calculated by dividing the number of accurate predictions by the number of predictions overall (Coleman et al. 2019).

One alternative method of evaluating image segmentation is to simply report the percentage of pixels in the image that were correctly classified. Each class's pixel accuracy is commonly reported separately, as well as on a global basis across all classes.

A binary mask is used to evaluate the per-class pixel accuracy. True positives are pixels that are accurately predicted to belong to a specific class (according to the target mask), and true negatives are pixels that are accurately identified as not belonging to that class (Ye et al. 2018).

$$Pixel\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

6.4 Precision

Precision is defined as the fraction of automatic segmentation boundary pixels that correspond to ground truth boundary pixels. Precision is calculated by dividing the number of Positive image samples that were correctly classified by the number of Positive image samples that were classified as Positive (Qadir et al. 2019; Wang et al. 2020; Xu et al. 2019).

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

6.5 Recall

The recall is defined as the percentage of ground truth boundary pixels that were correctly identified by automatic segmentation. A recall is a measure of the proportion of Positive image samples accurately classified as Positive compared to the total number of Positive image samples. It is a measure of how well the model can identify positive samples. The higher the recall, the more positive samples are detected (Aguiar et al. 2019; Xu et al. 2019).

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

6.6 Average precision

Average precision is defined as the weighted mean of precisions achieved at each threshold, with the increase in recall as the weight for each threshold:

$$\text{Average Precision} = \sum_n (R_n - R_{n-1})P_n \tag{8}$$

where P_n and R_n are the precision and recall at the nth threshold. This implementation is not interpolated and is different from computing the area under the precision-recall curve with the trapezoidal rule, which uses linear interpolation and can be too optimistic (Perez-Borrero et al. 2021; De Moura Lima et al. 2023).

This section explains the different metrics that are used to evaluate the performance of image segmentation methods. For polyp segmentation, the most relevant metrics are Dice Coefficient and Intersection over Union, found in existing research studies based on polyp segmentation. Other metrics, such as Precision, Average Precision, Recall, and Pixel Accuracy, are also used in polyp segmentation, but they are not as important.

7 Review outcomes and comparative analysis

This section aims to discuss the comparative analysis as well as the outcomes of the review study. As shown in Table 5, it was possible to compare the analysis of all the models and datasets with their performance metrics. This article also discusses a statistical analysis of polyp models based on popular datasets and their performance metrics.

This study provided the comparative analysis of polyp segmentation in which deep learning-based image segmentation techniques, methods, and models are represented in Table 5 with polyp datasets and evaluation metrics including dice coefficient, Intersection over the union, Precision, Recall, and F2 from prior review studies. The various research studies show that CVC-ColonDB, CVC-ClinicDB, Kvasir-SEG, and ETIS-LaribPolypDB are mostly used for training and testing for the segmentation of polyp. In the remaining section, this study presented the statistical analysis of prominently used performance metrics i.e., the dice coefficient of various deep learning-based models used in polyp datasets. From various research literature, it was found that the dice coefficient is the most important metric used in the polyp.

The statistical analysis in Fig. 23 shows the performance of different DL models used in polyp segmentation according to the dice coefficient and the CVC ColonDB dataset. In this study, according to results from existing studies, PraNet models show poor performance with a 70.9% dice coefficient, and SegNet-VGG models show the best performance with a 95.9% dice coefficient.

Figure 24 presents the results of the performance of various DL models used in polyp segmentation based on the dice coefficient and the CVC ClinicDB dataset. Based on previous studies, the hybrid model of ResUnet++ with conditional random field and test time augmentation has demonstrated poor performance with a 67.12% dice coefficient, and the FasterRCNN model has shown the best performance with a score of 99.6% dice coefficient.

As shown in Fig. 25, different deep-learning models are compared based on the dice coefficient and the Kvasir-SEG database. As reported in existing studies, the hybrid model of ResUnet++ with conditional random field (CRF) and test time augmentation (TTA) performed poorly with a 47.64% dice coefficient, and on the other hand, the SwinE-Net model has shown the best performance with a dice coefficient of 92%.

According to Fig. 26, different deep learning models are compared based on the dice coefficient as well as the ETIS-Larib dataset to determine the most effective dataset and model. Several studies have shown that the hybrid ResUnet++ model with CRF and TTA has performed poorly with a dice coefficient of 39.97%, whereas, SegNet-VGG has shown superior performance with a dice coefficient of 96.3%.

In this review study, Fig. 27 presents benchmark datasets including CVC-ClinicDB, CVC ColonDB, Kvasir-SEG, and ETIS-LaribPolyp used by most researchers for polyp segmentation, along with dice coefficients based on various deep learning polyp image segmentation models. Figure 27 aims to classify the minimum dice coefficient value, the maximum dice coefficient value, and the average value of the minimum and maximum dice coefficients of commonly used datasets. In comparison to other datasets, ETIS-Larib DB displays the lowest dice coefficient of 39.97%, while CVC Clinic DB displays the highest dice coefficient of 99.6%. However, CVC Clinic DB displays the best results with an average dice coefficient of 83.36% compared to the other datasets. According to all datasets, hybrid ResUnet++ models with conditional random field (CRF) and test time augmentation (TTA) performed poorly with fewer dice coefficients, whereas SegNet-VGG and Faster RCNN performed best with maximum dice coefficients.

8 Challenges and future trends

In current years, deep learning algorithms have assisted in the segmentation of polyps. The challenges and trends in this section are divided into three parts: the image segmentation models, the datasets, and the evaluation metrics. However, despite the achievement of deep learning models, the scientific community will face new challenges in the coming years. The purpose of this section is to determine the challenges, and future trends and organize discussions.

- *Based on DL-based image segmentation models:* A detailed description of image segmentation models used for polyp segmentation has been provided in this review study, including convolution neural network-based models, recurrent neural network-based models, attention-based models, encoder–decoder-based models, generative adversarial training models. In addition to supervised and unsupervised learning methods, the efficiency, and challenges of CNNs, encoder–decoders, RNNs, Attention, and GAN-based models for polyp segmentation have already been extensively investigated. In most research studies, CNNs and encoder–decoder-based models are used to segment polyps, including UNet and ResNet. According to various research studies, no research has been conducted on hybrid models that combine ML and DL algorithms in polyp segmentation. A novel segmentation model based on variations of U-Net and ResNet might be developed in the future using machine learning and deep learning algorithms to detect polyps. Researchers may be able to explore this area in the future. Nowadays, the transformer model, Segment Any model (SAM), and diffusion models are in trend for image segmentation in deep learning. So, it may be possible to use these models for polyp segmentation to enhance its evaluation metrics and efficiency. In addition, attention-based models are gaining popularity in polyp segmentation, where researchers may combine attention-based models with machine learning and DL algorithms to improve the performance of the model. The ability of recurrent networks to describe temporal connections may facilitate the model development for polyp videos that consider temporal information. Considering that GANs perform well in polyp segmentation, they can be used for future polyp segmentation research. GANs may also be combined with CNN-based models, deep learning, and machine learning algorithms, to produce effective results. It is anticipated that a hybrid GAN model using encoder- and decoder-based models will be developed in the future to provide better results and provide higher evaluation metrics.
- *Based on datasets used* : Multiple polyp segmentation datasets, including image and video datasets, were analyzed for this study. There are five types of image datasets commonly used in polyp segmentation prior research, namely Kvasir, Kvasir-SEG, CVCClinicDB, CVCColonDB, and ETIS-Larib, as well as six types of video datasets consisting of Hyperkvasir, CVC Clinic Video DB, EDD 2020, Kvasir-Capsule, LD Polyp Video DB, and ASU-Mayo Clinic DB. To make it easier for readers to distinguish between image datasets and video datasets, this review study compares all the image datasets and video datasets. The polyp datasets have several limitations, including the fact that they are small and have a texture or appearance that is like that of the skin. This makes it difficult to identify the exact location of the polyp in the region. According to existing research, most of the models proposed for polyp segmentation have been applied only to single image datasets, however, there may be potential for applying their proposed models to multiple image datasets. This may result in the tun-

Table 5 Comparative study of performance evaluation of various models for polyp segmentation

DL based segmentation	References	Models used	DSC/F1 (%)	IoU/JC (%)	Precision (%)	Recall (%)	F2 (%)	Accuracy (%)
CNN based models	Yunbo Guo et al. (2020)	FCN8s	63		68	65		
		ResFCN	71	–	75	74	–	–
		Dilated ResFCN	79		81	81		
	Akbari et al. (2018)	FCN8s with image + rotation + patch selection method	81	–	88.3	–	–	97.7
	Qadir et al. (2019)	1) Mask RCNN (Resnet50 +)	80.42	73.4	93.1	85.34		
		2) Mask RCNN (Resnet101 +)	77.48	70.1	95	84.87	–	–
		3) Mask RCNN (Inception Resnet +)	80.19	73.2	94.1	86.1		
	Jorge Bernal et al. (2017)	CUMED, OUS datasets	78.5	–	80.7	76.4	77.2	–
	Mo et al. (2018)	CVC15DB with faster RCNN	91.7	–	86.2	98.1	95.6	–
	Sun et al. (2019)	CNN with post-processing	82.48	–	96.7	95.51	–	–
		CNN without post-processing			94.6	96.13		
Encoder–decoder based models	Xiao et al. (2018)	DeepLab_v3 + and LSTM	–	93.21	–	–	–	–
	Mulliqi et al. (2020)	Dense-UNet + +	78.37	–	87.46	70.99	73.7	90.42
	Tashk et al. (2019)	FCN + UNet (CVC ClinicDB)	83		90	77		97
		FCN + UNet (ETIS-Larib)	79.4	–	73.6	86.3	–	99.1
		FCN + UNet (CVC ColonDB)	81		88.3	74.8		97.7
	Oktay et al. (2018)	AG-CUResNeSt fivefold cross-validation on CVC Clinic and Kvasir-SEG	94.6	90	94.4	84	–	–
			91.2	86	92.7	84		
	Jha et al. (2019)	ResUNet + + on Kvasir Seg and CVC ClinicDB	81.33	79.2	87.74	70.64	–	–
			79.55	79.6	87.85	70.22		
RNN based models	Tran et al. (2022a)	Modified Recurrent Residual Unet	94.59	89.80	94.63	94.95	–	–
Attention based models	Fan et al. (2020)	PraNet on Kvasir and CVC-Clinic	89.8	84	–	–	88	
			89.9	84.9			89	

Table 5 (continued)

DL based segmentation	References	Models used	DSC/F1 (%)	IoU/JC (%)	Precision (%)	Recall (%)	F2 (%)	Accuracy (%)
GAN based models	Qian et al. (2022)	Conditional Generative Adversarial Network (CGAN)	–	–	–	–	–	92.37
	Poomeshwaran et al. (2019)	Deep generative convolutional framework namaste	84.74	76.9				
			87.23	79.4	–	–	–	–
			88.48	81.2				
	Yoon et al. 2022)	Generative adversarial network applied for enhanced segmentation of sessile serrated lesions using convolutional neural network	92.31	–	–	–	–	91.22

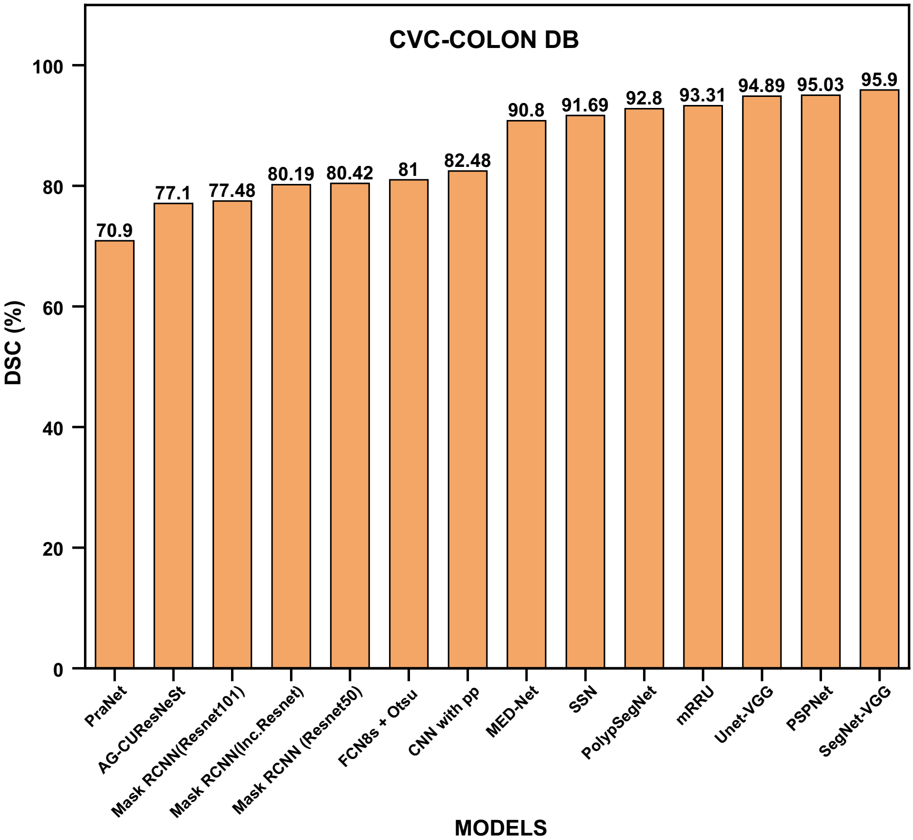


Fig. 23 Dice Coefficient of different DL models used in the CVC ColonDB dataset

ing of models for all types of datasets. The researchers may be able to generate a polyp dataset where classification or machine learning models have been applied since classification on polyp detection has not yet been explored. The available polyp datasets do not include information about a person's age or gender, if such information is included in the dataset, then it will be used for classification and segmentation. Therefore, if the gender of the individual in the polyp dataset is known, it may be possible to determine if the polyp is predominantly present in males or females. In addition, it may be possible to determine whether the polyp size is larger in males or females. In the same way, if the patient's age is available in the polyp database, it may be possible to determine the probability that the polyp will have an adverse effect on the patient based on its age. There will be further exploration of these datasets by the research community as well as the development of datasets that include other information on polyp segmentation. Based on the research work on the polyp video dataset, it has been found that fewer models have been applied to video frames for polyp segmentation. As a result, most of the researchers are facing complexities and related issues regarding video polyp datasets, so they have taken images from video datasets converted them into images, and used the images to polyp segmentation. Due to the limited use of video datasets in polyp segmentation, there will be more opportunities to develop some models that use

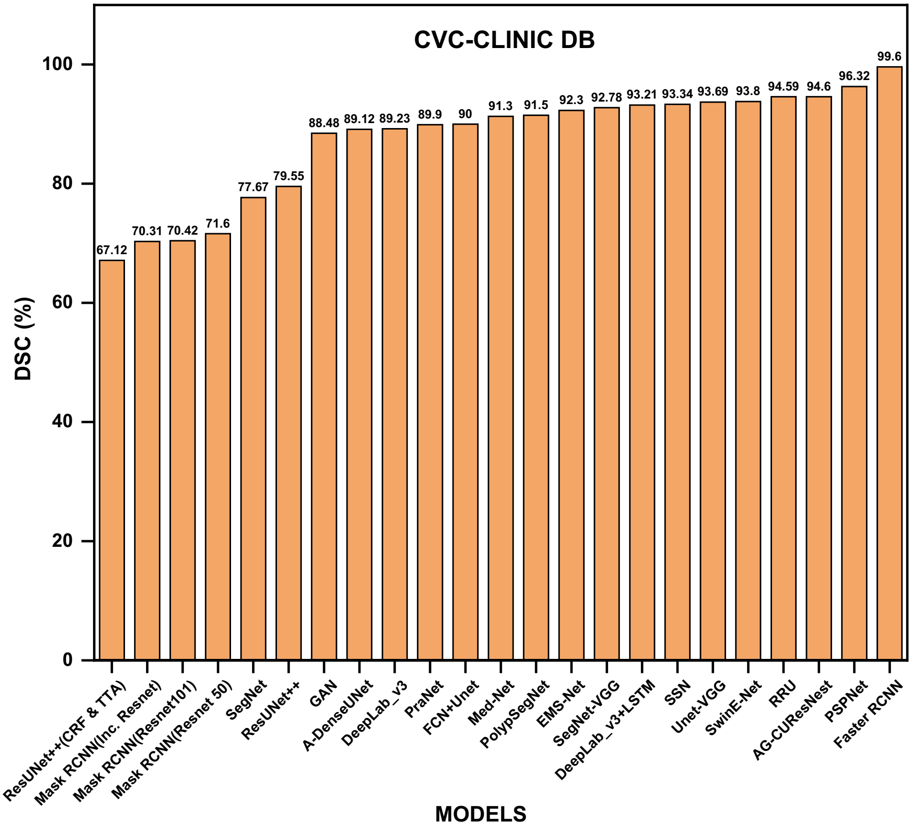


Fig. 24 Dice Coefficient of different DL models used in the CVC ClinicDB dataset

video data directly to detect polyps in a real-time manner. Many video datasets have been released recently, and only limited research has been conducted on those datasets. Therefore, they will be more useful in polyp segmentation research in the future.

- *Based on performance metrics:* This section aims to assess the performance metrics found in existing studies that are commonly used for polyp segmentation. As part of existing research on polyp segmentation, several metrics are available, including the dice coefficient, the intersection over union, the Jaccard index, precision, recall, F2, and accuracy. The DSC and IoU are the two most significant evaluation metrics used by various authors to assess the performance of polyp segmentation. The accuracy of the model used for polyp segmentation has also been used as a metric for evaluating polyp segmentation. As this is a computer vision problem, there may be a need to use more performance metrics to accurately locate or segment a polyp. The dice coefficient may be combined with intersection over union in the future to improve the performance of the evaluation metrics in polyp segmentation. Due to the limited research on video polyp datasets, frames per second were computed to identify the frames and to find the evaluation metrics. The evaluation metrics may show good and efficient results when the image dataset is larger in quantity. A new performance metric can be proposed that improves the accuracy, robustness, and effectiveness of comparing deep learning algo-

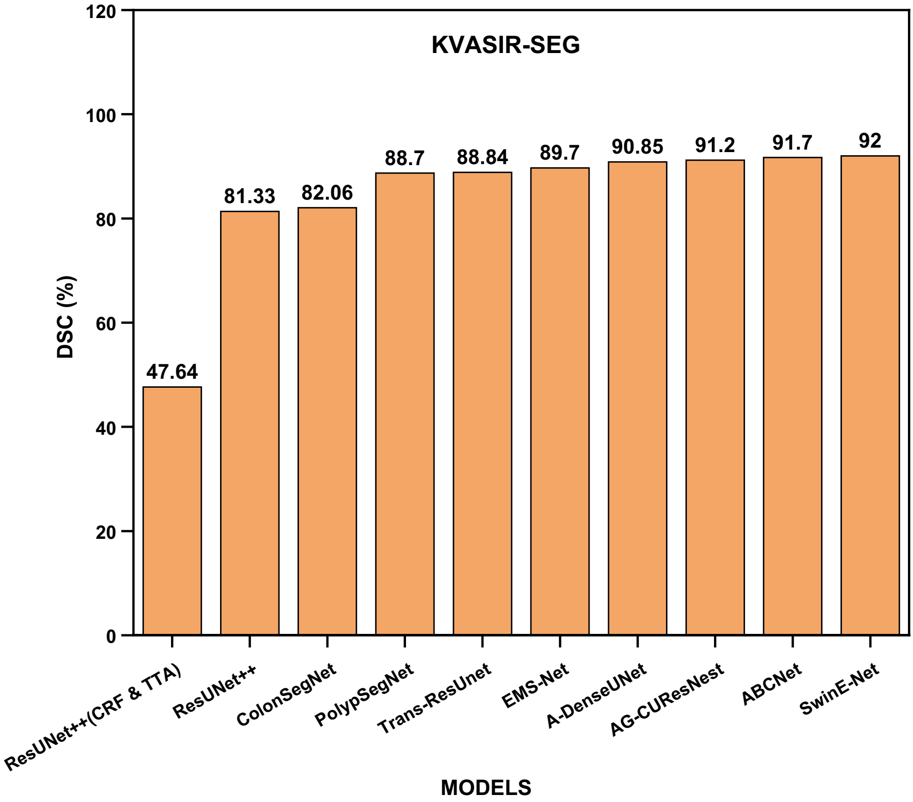


Fig. 25 Dice Coefficient of different DL models used in the Kvasir-SEG dataset

rithms for predicting returns over time and assets in polyp segmentation. It may be possible to classify polyps using machine learning classifiers and then show the performance of all ML algorithms based on their classification metrics. It may be possible for accuracy to produce effective results for both polyp segmentation and classification if it is combined with dice coefficients or IOUs. Therefore, if it is feasible, classification and segmentation can employ the same evaluation metrics to obtain reliable results for polyps.

9 Conclusions

Polyp cancer is the third most fatal and severe form of cancer in the world. It is crucial to diagnose cancer at an early stage to treat it effectively. In recent years, deep learning applications have gained popularity due to their advantages and accomplishments in the early diagnosis and screening of malignant tissues and organs. A review of the most recent research studies utilizing DL techniques and models for polyp segmentation has been conducted in this study.

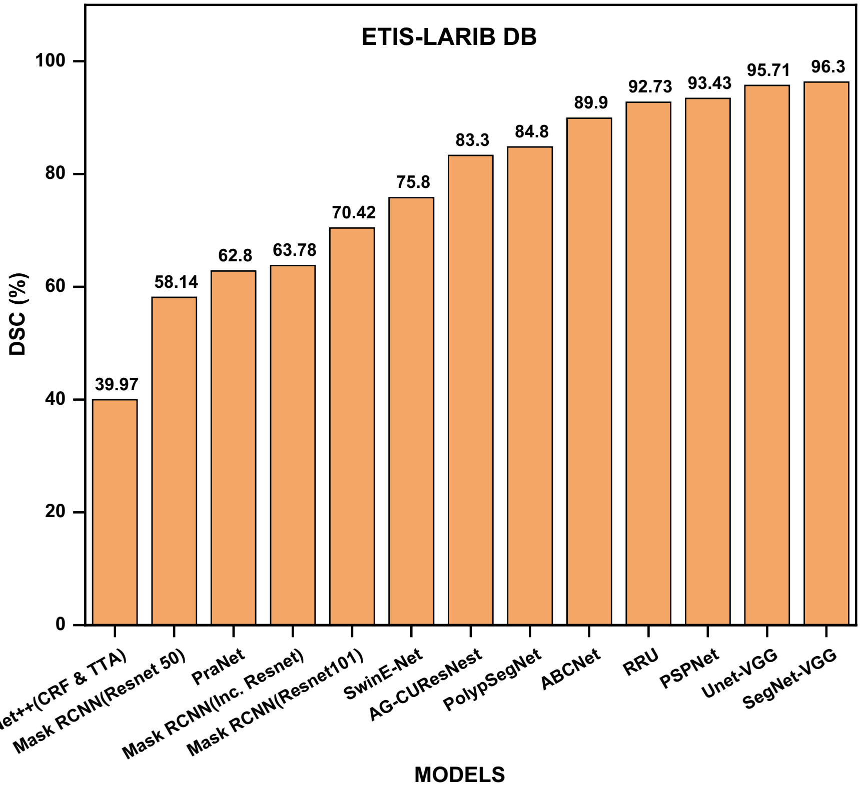


Fig. 26 Dice Coefficient of different DL models used in the ETIS-Larib DB dataset

The current study presents a comprehensive and systematic review of more than 100 research studies for the segmentation of polyps using DL approaches from 2018 to the current year. This review work was categorized into six major categories to make it easier to understand. First, a discussion of image segmentation methods was presented, including semantic segmentation, instance segmentation, and panoptic segmentation. Second, a comprehensive review of DL-based image segmentation models was conducted based on existing research studies that are most relevant to polyp segmentation. The most popular deep learning models found in prior research are categorized into convolutional neural networks (CNNs), encoder–decoder models, recurrent neural networks (RNNs), attention-based models, and generative models (GANs) in polyp segmentation. In this research study, encoder–decoder models play an important role in deep learning image segmentation as they are widely used in polyp segmentation research. U-Net is one of the most important encoder–decoder image segmentation models in polyp segmentation and various variations of U-Net have been found in existing research. As a result, a comparative and comprehensive analysis of U-Net and its variations was presented. The third is to provide a detailed classification of all the image and video datasets used in the polyp segmentation, as well as their associated issues and comparative analysis. The five image polyp datasets and six video polyp datasets are presented as public databases for polyp segmentation. Fourth, discusses evaluation metrics for assessing the effectiveness of different methods, models, and

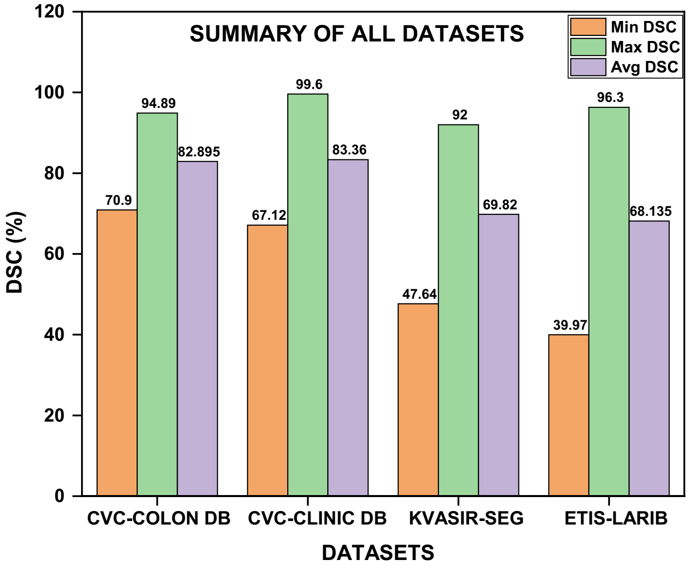


Fig. 27 Min, Max, and Avg dice coefficient of different DL models used in CVC-ColonDB, CVC-ClinicDB, Kvasir-SEG, and ETIS-LaribPolypDB datasets

techniques found in the literature review regarding polyp segmentation. Fifth discusses the comparative analysis as well as the outcomes of the review study and presents a statistical analysis of polyp models based on prominent datasets and their performance metrics. The sixth is to present a discussion of future research trends and limitations associated with deep learning models, polyp datasets, and performance metrics for polyp segmentation. While deep learning approaches have achieved great success, validation, testing, and application still need to be a priority. Further, the creation of larger, more diverse, public datasets, new algorithms requiring fewer training samples, and the creation of a common evaluation criterion will maintain the upward trend and will result in increased efficiency and polyp segmentation.

Acknowledgements Not applicable.

Author contributions MG: Literature search, Data analysis, Conceptualization, Methodology, Investigation Writing, Deep learning models review & editing, Supervision. AM: Literature search, Idea for the article, Resources, review & editing, Data curation, Writing—review & editing.

Funding This study received no specific grant from any funding agency i.e., public, commercial, or non-profit sector.

Data availability Data collected during the study are subject to a data-sharing mandate and available in a public repository that does not issue datasets with DOIs.

Declarations

Competing interests The authors declare no competing interests.

Ethical approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

Aguiar GJ, Mantovani RG, Mastelini SM, de Carvalho ACPFL, Campos GFC, Junior SB (2019) A meta-learning approach for selecting image segmentation algorithm. Pattern Recogn Lett 128:480–487. https://doi.org/10.1016/J.PATREC.2019.10.018

Akbari M, Mohrekesh M, Nasr-Esfahani E, Soroushmehr SMR, Karimi N, Samavi S, Najarian K (2018) Polyp segmentation in colonoscopy images using fully convolutional network. In: 2018 40th annual international conference of the ieee engineering in medicine and biology society (EMBC), 2018-July, 69–72. https://doi.org/10.1109/EMBC.2018.8512197

Al Jowair H, Alsulaiman M, Muhammad G (2023) Multi parallel U-net encoder network for effective polyp image segmentation. Image Vis Comput 137:104767. https://doi.org/10.1016/j.imavis.2023.104767

Alam MJ, Fattah SA (2023) SR-AttNet: an interpretable stretch-relax attention based deep neural network for polyp segmentation in colonoscopy images. Comput Biol Med 160:106945. https://doi.org/10.1016/j.compbiomed.2023.106945

Alfred Daniel J, Chandru Vignesh C, Muthu BA, Senthil Kumar R, Sivaparthipan C, Marin CEM (2023) Fully convolutional neural networks for LIDAR–camera fusion for pedestrian detection in autonomous vehicle. Multimed Tools Appl 82(16):25107–25130. https://doi.org/10.1007/S11042-023-14417-X

Ali H, Sharif M, Yasmin M, Rehmani MH, Riaz F (2020) A survey of feature extraction and fusion of deep learning for detection of abnormalities in video endoscopy of gastrointestinal-tract. Artif Intell Rev 53(4):2635–2707. https://doi.org/10.1007/S10462-019-09743-2

Ali S, Dmitrieva M, Ghatwary N, Bano S, Polat G, Temizel A, Krenzer A, Hekalo A, Guo YB, Matuszewski B, Gridach M, Voiculescu I, Yoganand V, Chavan A, Raj A, Nguyen NT, Tran DQ, Huynh LD, Boutry N, Rittscher J (2021) Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy. Med Image Anal 70:102002. https://doi.org/10.1016/J.MEDIA.2021.102002

Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK (2018) Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for medical image segmentation. arXiv: 1802.06955

Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 39(12):2481–2495. https://doi.org/10.1109/tpami.2016.2644615

Barua I, Vinsard DG, Jodal HC, Loberg M, Kalager M, Holme O, Misawa M, Bretthauer M, Mori Y (2021) Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. Endoscopy 53(3):277–284. https://doi.org/10.1055/a-1201-7165

Bernal J, Sánchez J, Vilariño F (2012) Towards automatic polyp detection with a polyp appearance model. Pattern Recogn 45(9):3166–3182. https://doi.org/10.1016/J.PATCOG.2012.03.002

Bernal J, Sánchez FJ, Fernández-Esparrach G, Gil D, Rodríguez C, Vilariño F (2015) WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. Comput Med Imaging Graph 43:99–111. https://doi.org/10.1016/J.COMPMEDIMAG.2015.02.007

Bernal Jorge, Tajkbaksh N, Sanchez FJ, Matuszewski BJ, Chen H, Yu L, Angermann Q, Romain O, Rustad B, Balasingham I, Pogorelov K, Choi S, Debard Q, Maier-Hein L, Speidel S, Stoyanov D, Brandao P, Cordova H, Sanchez-Montes C, Histace A (2017) Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge. IEEE Trans Med Imaging 36(6):1231–1249. https://doi.org/10.1109/TMI.2017.2664042

Borgli H, Thambawita V, Smedsrud PH, Hicks S, Jha D, Eskeland SL, Randel KR, Pogorelov K, Lux M, Nguyen DTD, Johansen D, Griwodz C, Stensland HK, Garcia-Ceja E, Schmidt PT, Hammer HL, Riegler MA, Halvorsen P, de Lange T (2020) HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. Sci Data 7(1):1–14. https://doi.org/10.1038/s41597-020-00622-y

Bychkov D, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, Walliander M, Lundin M, Haglund C, Lundin J (2018) Deep learning based tissue analysis predicts outcome in colorectal cancer. Sci Rep 8(1):1–11. https://doi.org/10.1038/s41598-018-21758-3

Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M (2021) Swin-Unet: Unet-like pure transformer for medical image segmentation. arXiv: 2105.05537

Castiglioni I, Rundo L, Codari M, Leo GD, Salvatore C, Interlenghi M, Gallivanone F, Cozzi A, Claudia D'amico N, Sardanelli F (2021) AI applications to medical images: from machine learning to deep learning. Physica Medica 83:1120–1797. https://doi.org/10.1016/j.ejmp.2021.02.006

Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y (2021) TransUNet: transformers make strong encoders for medical image segmentation. arXiv: 2102.04306

Chou YC, Chen CC (2022) Improving deep learning-based polyp detection using feature extraction and data augmentation. Multimed Tools Appl. https://doi.org/10.1007/S11042-022-13995-6

Choudhuri A, Chowdhary G, Schwing AG (2023) Context-aware relative object queries to unify video instance and panoptic segmentation. In: openaccess.thecvf.com.

Chu W, Williams H, Verghese K, Wever R, Glad W (2020) Tensions and opportunities: an activity theory perspective on date and storage label design through a literature review and co-creation sessions. Sustainability 12(3):1162. https://doi.org/10.3390/su12031162

Chuang Y, Zhang S, Zhao X (2023) Deep learning-based panoptic segmentation: recent advances and perspectives. In: IET image processing. Wiley.https://doi.org/10.1049/ipr2.12853

Coleman C, Kang D, Narayanan D, Nardi L, Zhao T, Zhang J, Bailis P, Olukotun K, Ré C, Zaharia M (2019) Analysis of DAWNBench, a time-to-accuracy machine learning performance benchmark. ACM SIGOPS Oper Sys Rev 53(1):14–25. https://doi.org/10.1145/3352020.3352024

Cong P, Li S, Zhou J, Lv K, Feng H (2023) Research on instance segmentation algorithm of greenhouse sweet pepper detection based on improved mask RCNN. Agronomy 13(1):196. https://doi.org/10.3390/agronomy13010196

Cufí Jou A (2022) Colorectal cancer screening among individuals aged 40–49 years: a randomised controlled field trial

De Moura Lima AC, De Paiva LF, Braz G, De Almeida JDS, Silva AC, Coimbra MT, De Paiva AC (2023) A two-stage method for polyp detection in colonoscopy images based on saliency object extraction and transformers. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3297097

Diakogiannis FI, Waldner F, Caccetta P, Wu C (2020) ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data. ISPRS J Photogramm Remote Sens 162:94–114. https://doi.org/10.1016/j.isprsjprs.2020.01.013

Dong X, Zhang C, Fang L, Yan Y (2022) A deep learning based framework for remote sensing image ground object segmentation. Appl Soft Comput 130:109695. https://doi.org/10.1016/J.ASOC.2022.109695

Elharrouss O, Al-Maadeed S, Subramanian N, Ottakath N, Almaadeed N, Himeur Y (2021) Panoptic segmentation: a review. arXiv: 2111.10250

ELKarazle K, Raman V, Then P, Chua C (2023) Detection of colorectal polyps from colonoscopy using machine learning: a survey on modern techniques. In: Sensors, vol 23, Issue 3. https://doi.org/10.3390/s23031225

El Khatib A, Werghi N, Al-Ahmad H (2015) Automatic polyp detection: a comparative study. In: Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, EMBS, 2015-November, pp 2669–2672. https://doi.org/10.1109/EMBC.2015.7318941

Eloranta S, Smedby KE, Dickman PW, Andersson TM (2021) Cancer survival statistics for patients and healthcare professionals—a tutorial of real-world data analysis. J Intern Med 289(1):12–28. https://doi.org/10.1111/JOIM.13139

Fan DP, Ji GP, Zhou T, Chen G, Fu H, Shen J, Shao L (2020) PraNet: parallel reverse attention network for polyp segmentation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12266 LNCS, pp 263–273. https://doi.org/10.1007/978-3-030-59725-2_26

Feng R, Lei B, Wang W, Chen T, Chen J, Chen DZ, Wu J (2020) SSN: a stair-shape network for real-time polyp segmentation in colonoscopy images. Proceedings—international symposium on biomedical imaging, 2020-April, pp 225–229. https://doi.org/10.1109/ISBI45749.2020.9098492

Fernando T, Gammulle H, Denman S, Sridharan S, Fookes C (2022) Deep Learning for medical anomaly detection—a survey. ACM Comput Surv 54(7):1–37. https://doi.org/10.1145/3464423

Gangcuangco LMA, Rivas T, Basnet A, Ryu DY, Qaiser M, Usman R, Costales VC (2022) Factors associated with colorectal cancer screening adherence and the impact of COVID-19 on screening patterns in Connecticut, USA. Intern Emerg Med. https://doi.org/10.1007/S11739-022-03053-2

Ghosh S, Das N, Das I, Maulik U (2020) Understanding deep learning techniques for image segmentation. ACM Comput Surv 52(4):1–35. https://doi.org/10.1145/3329784

Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. IEEE Conf Comput Vis Pattern Recognit 2014:580–587. https://doi.org/10.1109/CVPR.2014.81

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2020) Generative adversarial networks. Dl.Acm.Org, 63(11):139–144. https://doi.org/10.1145/3422622
Gopakumar G (2020) A review on polyp detection and segmentation in colonoscopy images using deep learning. 9(10):329–335
Gopinath B, Gokul K, Pumenitha ST, Vasanth SH (2023) Deep learning based automated parking lot space detection using aerial imagery. Ieeexplore.Ieee.Org
Guo Y, Matuszewski B (2019) GIANA polyp segmentation with fully convolutional dilation neural networks. In: Proceedings of the 14th international joint conference on computer vision, imaging and computer graphics theory and applications, vol 4, pp 632–641. https://doi.org/10.5220/0007698806320641
Guo Y, Liu Y, Georgiou T, Lew MS (2018) A review of semantic segmentation using deep neural networks. Int J Multimed Inf Retr 7(2):87–93. https://doi.org/10.1007/S13735-017-0141-Z
Guo Z, Li X, Huang H, Guo N, Li Q (2019) Deep learning-based image segmentation on multimodal medical imaging. IEEE Trans Radiat Plasma Med Sci 3(2):162–169. https://doi.org/10.1109/TRPMS.2018.2890359
Guo Y, Bernal J, Matuszewski B (2020) Polyp segmentation with fully convolutional deep neural networks—extended evaluation study. J Imaging 6(7):69. https://doi.org/10.3390/jimaging6070069
Gupta M, Sinha A (2022) Multi-class autoencoder-ensembled prediction model for detection of COVID-19 severity. Evolut Intell. https://doi.org/10.1007/s12065-022-00744-9
Gupta M, Singhal YK, Sinha A (2022) Assessing spatiotemporal transmission dynamics of COVID-19 outbreak using AI analytics, pp 829–838. https://doi.org/10.1007/978-981-16-3346-1_67
Gupta S, Sikka G, Malik A (2023) A review on deep learning-based polyp segmentation for efficient colorectal cancer screening. In: ICSCCC 2023—3rd international conference on secure cyber computing and communications, pp 501–506. https://doi.org/10.1109/ICSCCC58608.2023.10176721
Hafiz AM, Bhat GM (2020) A survey on instance segmentation: state of the art. Int J Multimed Inf Retr 9(3):171–189. https://doi.org/10.1007/S13735-020-00195-X
He F, Chen S, Li S, Zhou L, Zhang H, Peng H, Huang X (2021) Colonoscopic image synthesis for polyp detector enhancement via gan and adversarial training. In: Proceedings—international symposium on biomedical imaging, 2021-April, pp 1887–1891. https://doi.org/10.1109/ISBI48211.2021.9434050
Hoorali F, Khosravi H, Moradi B (2020) Automatic Bacillus anthracis bacteria detection and segmentation in microscopic images using UNet++. J Microbiol Methods 177:106056. https://doi.org/10.1016/J.MIMET.2020.106056
Hsu WY, Yang PY (2023) Pedestrian detection using multi-scale structure-enhanced super-resolution. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2023.3287574
Huang CH, Xiao WT, Chang LJ, Tsai WT, Liu WM (2018) Automatic tissue segmentation by deep learning: from colorectal polyps in colonoscopy to abdominal organs in CT exam. In: VCIP 2018—IEEE international conference on visual communications and image processing. https://doi.org/10.1109/VCIP.2018.8698645
Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, Han X, Chen YW, Wu J (2020) UNet 3+: a full-scale connected UNet for medical image segmentation. In: ICASSP, IEEE international conference on acoustics, speech and signal processing—proceedings, 2020-May(ii), pp 1055–1059. https://doi.org/10.1109/ICASSP40776.2020.9053405
Iqbal MJ, Javed Z, Sadia H, Qureshi IA, Irshad A, Ahmed R, Malik K, Raza S, Abbas A, Pezzani R, Sharifi-Rad J (2021) Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. Cancer Cell Int. https://doi.org/10.1186/S12935-021-01981-1
Jha D, Smedsrud PH, Riegler MA, Johansen D, De Lange T, Halvorsen P, Johansen HD (2019) ResUNet++: an advanced architecture for medical image segmentation. In: Proceedings—2019 IEEE international symposium on multimedia, ISM 2019, pp 225–230https://doi.org/10.1109/ISM46123.2019.00049
Jha D, Smedsrud PH, Riegler MA, Halvorsen P, de Lange T, Johansen D, Johansen HD (2020) Kvasir-SEG: a segmented polyp dataset. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11962 LNCS, pp 451–462. https://doi.org/10.1007/978-3-030-37734-2_37
Jha D, Ali S, Tomar NK, Johansen HD, Johansen D, Rittscher J, Riegler MA, Halvorsen P (2021a) Real-time polyp detection, localization and segmentation in colonoscopy using deep learning. IEEE Access 9:40496–40510. https://doi.org/10.1109/ACCESS.2021.3063716
Jha D, Smedsrud PH, Johansen D, De Lange T, Johansen HD, Halvorsen P, Riegler MA (2021b) A comprehensive study on colorectal polyp segmentation with ResUNet++, conditional random field

and test-time augmentation. IEEE J Biomed Health Inform 25(6):2029–2040. https://doi.org/10.1109/JBHI.2021.3049304
Lei H, He F, Jia B, Wu Q (2023) MFNet: panoptic segmentation network based on multiscale feature weighted fusion and frequency domain attention mechanism. IET Comput Vis 17(1):88–97. https://doi.org/10.1049/cvi2.12133
Lewandowska A, Rudzki M, Rudzki S, Lewandowski T, Laskowska B (2019) Environmental risk factors for cancer—review paper. Ann Agric Environ Med 26(1):1–7. https://doi.org/10.26444/aaem/94299
Li X, Chen D (2022) A survey on deep learning-based panoptic segmentation. In: Digital signal processing: a review journal, vol 120. https://doi.org/10.1016/j.dsp.2021.103283
Li Y, Hu M, Yang X (2023) Polyp-SAM: transfer SAM for polyp segmentation.
Liu X, Song L, Liu S, Zhang Y (2021a) A review of deep-learning-based medical image segmentation methods. Sustainability 13(3):1–29. https://doi.org/10.3390/su13031224
Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021b) Swin transformer: hierarchical vision transformer using shifted windows. In: 2021 IEEE/CVF international conference on computer vision (ICCV), pp 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986
Luca M, Ciobanu A, Drug V (2019) Deep learning and automatic polyp detection in colonoscopies: a review of recent contributions and future outlook. In: 2019 7th E-Health and Bioengineering Conference, EHB 2019, pp 43–46. https://doi.org/10.1109/EHB47216.2019.8970041
Ma Y, Chen X, Cheng K, Li Y, Sun B (2021) LDPolypVideo benchmark: a large-scale colonoscopy video dataset of diverse polyps. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12905 LNCS, pp 387–396. https://doi.org/10.1007/978-3-030-87240-3_37/FIGURES/6
Macháček R, Mozaffari L, Sepasdar Z, Parasa S, Halvorsen P, Riegler MA, Thambawita V (2023) Mask-conditioned latent diffusion for generating gastrointestinal polyp images
Manshadi MD, Mousavi M, Golzarian A, Soltani M, Mosavi A (2023) Colorectal Polyp localization: from image restoration to real-time detection with deep learning. Ieeexplore.Ieee.Org, pp 739–744. https://doi.org/10.1109/saci58269.2023.10158661
Mao L, Ren F, Yang D, Zhang R (2023) ChaInNet: deep chain instance segmentation network for panoptic segmentation. Neural Process Lett 55(1):615–630. https://doi.org/10.1007/s11063-022-10899-2
Mazumdar S, Sinha S, Jha S, Jagtap B (2023) Computer-aided automated diminutive colonic polyp detection in colonoscopy by using deep machine learning system; first indigenous algorithm developed in India. Indian J Gastroenterol. https://doi.org/10.1007/s12664-022-01331-7
MBiostat NB, Wong GY, Molloy C, Dieng M, Kelly PJ, Hugh TJ (2022) Lifetime direct healthcare costs of treating colorectal cancer: a systematic review. Eur J Health Econ. https://doi.org/10.1007/S10198-022-01497-Z
Messina P, Pino P, Parra D, Soto A, Besa C, Uribe S, Andía M, Tejos C, Prieto C, Capurro D (2022) A survey on deep learning and explainability for automatic report generation from medical images. ACM Comput Surv 54(10s):1–40. https://doi.org/10.1145/3522747
Mi J, Han X, Wang R, Ma R, Zhao D (2022) Diagnostic accuracy of wireless capsule endoscopy in polyp recognition using deep learning: a meta-analysis. Int J Clin Pract 2022:1–10. https://doi.org/10.1155/2022/9338139
Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D (2022) Image segmentation using deep learning: a survey. IEEE Trans Pattern Anal Mach Intell 44(7):3523–3542. https://doi.org/10.1109/TPAMI.2021.3059968
Mo X, Tao K, Wang Q, Wang G (2018) An Efficient approach for polyps detection in endoscopic videos based on faster r-CNN. In: 2018 24th international conference on pattern recognition (ICPR), 2018-August, pp 3929–3934. https://doi.org/10.1109/ICPR.2018.8545174
Mukherjee A, Chakraborty S, Saha SK (2019) Detection of loop closure in SLAM: a DeconvNet based approach. Appl Soft Comput J 80:650–656. https://doi.org/10.1016/j.asoc.2019.04.041
Mulliqi N, Yildirim S, Mohammed A, Ahmedi L, Wang H, Elezaj O, Hovde O (2020). The importance of skip connections in encoder-decoder architectures for colorectal polyp detection. In: 2020 IEEE international conference on image processing (ICIP), 2020-October, pp 380–384. https://doi.org/10.1109/ICIP40778.2020.9191310
Nie X, Liu L, He L, Zhao L, Lu H, Lou S, Xiong R, Wang Y (2023) Weakly-interactive-mixed learning: less labelling cost for better medical image segmentation. IEEE J Biomed Health Inform. https://doi.org/10.1109/JBHI.2023.3268157
Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D (2018) Attention U-Net: learning where to look for the pancreas. arXiv: 1804.03999

Öztürk Ş, Özkaya U (2021) Residual LSTM layered CNN for classification of gastrointestinal tract diseases. J Biomed Inform 113:103638. https://doi.org/10.1016/J.JBI.2020.103638

Pacal I, Karaboga D (2021) A robust real-time deep learning based automatic polyp detection system. Comput Biol Med 134:104519. https://doi.org/10.1016/j.compbiomed.2021.104519

Pacal I, Karaboga D, Basturk A, Akay B, Nalbantoglu U (2020) A comprehensive review of deep learning in colon cancer. Comput Biol Med 126:104003. https://doi.org/10.1016/j.compbiomed.2020.104003

Padilla R, Netto SL, Da Silva EAB (2020) A survey on performance metrics for object-detection algorithms. In: International conference on systems, signals, and image processing, 2020-July, pp 237–242. https://doi.org/10.1109/IWSSIP48289.2020.9145130

Palanisamy V, Thirunavukarasu R (2019) Implications of big data analytics in developing healthcare frameworks—a review. J King Saud Univ-Comput Inf Sci 31(4):415–425. https://doi.org/10.1016/J.JKSUCI.2017.12.007

Park KB, Lee JY (2022) SwinE-Net: hybrid deep learning approach to novel polyp segmentation using convolutional neural network and Swin Transformer. J Comput Des Eng 9(2):616–632. https://doi.org/10.1093/JCDE/QWAC018

Patel AR, Tyagi S (2022) The state of test automation in DevOps: a systematic literature review. In: ACM international conference proceeding series, pp 689–695. https://doi.org/10.1145/3549206.3549321

Perez-Borrero I, Marin-Santos D, Vasallo-Vazquez MJ, Gegundez-Arias ME (2021) A new deep-learning strawberry instance segmentation methodology based on a fully convolutional neural network. In: Neural computing and applications, vol 33, Issue 22. Springer Science and Business Media Deutschland GmbH, pp 15059–15071. https://doi.org/10.1007/s00521-021-06131-2

Pogorelov K, Ranheim Randel K, Griwodz C, Hospital B, Thomas de Lange N, Johansen D, Spampinato C, Dang-Nguyen D-T, Lux M, Thelin Schmidt Karolinska Institutet P, Karolinska Hospital S, Michael Riegler S, Halvorsen P, Losada Eskeland S, de Lange T, Thelin Schmidt P, Riegler M (2017) Kvasir: a multi-class image dataset for com-puter aided gastrointestinal disease detection. ACM Reference Format, 6. https://doi.org/10.1145/3083187.3083212

Poomeshwaran JM, Santhosh KS, Ram K, Joseph J, Sivaprakasam M (2019) Polyp segmentation using generative adversarial network. In: Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, EMBS, pp 7201–7204. https://doi.org/10.1109/EMBC.2019.8857958

Pozdeev AA, Obukhova NA, Motyko AA (2019) Automatic Analysis of endoscopic images for polyps detection and segmentation. In: 2019 IEEE conference of russian young researchers in electrical and electronic engineering (EIConRus), pp 1216–1220. https://doi.org/10.1109/EIConRus.2019.8657018

Prasath VBS (2017) Polyp detection and segmentation from video capsule endoscopy: a review. J Imaging. https://doi.org/10.3390/jimaging3010001

Qadir HA, Shin Y, Solhusvik J, Bergsland J, Aabakken L, Balasingham I (2019). Polyp detection and segmentation using mask R-CNN: does a deeper feature extractor CNN always perform better? In: 2019 13th international symposium on medical information and communication technology (ISMICT), 2019-May, pp 1–6. https://doi.org/10.1109/ISMICT.2019.8743694

Qian Z, Jing W, Lv Y, Zhang W (2022) Automatic polyp detection by combining conditional generative adversarial network and modified you-only-look-once. IEEE Sens J 22(11):10841–10849. https://doi.org/10.1109/JSEN.2022.3170034

Quan Q, Zhao S, Yao Q, Zhu H, Zhou SK (2023) Unsupervised augmentation optimization for few-shot medical image segmentation

Rahim T, Hassan SA, Shin SY (2021) A deep convolutional neural network for the detection of polyps in colonoscopy images. Biomed Signal Process Control 68:102654. https://doi.org/10.1016/j.bspc.2021.102654

Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I, Savarese S (2019) Generalized intersection over union: a metric and a loss for bounding box regression. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2019-June, pp 658–666. https://doi.org/10.1109/CVPR.2019.00075

Ruan J, Cui H, Huang Y, Li T, Wu C, Zhang K (2023) A review of occluded objects detection in real complex scenarios for autonomous driving. In: Green energy and intelligent transportation, vol 2, Issue 3. https://doi.org/10.1016/j.geits.2023.100092

Safarov S, Whangbo TK (2021) A-denseunet: adaptive densely connected unet for polyp segmentation in colonoscopy images with atrous convolution. Sensors 21(4):1–15. https://doi.org/10.3390/s21041441

Sánchez-Montes C, Bernal J, García-Rodríguez A, Córdova H, Fernández-Esparrach G (2020) Review of computational methods for the detection and classification of polyps in colonoscopy imaging. Gastroenterol Hepatol 43(4):222–232. https://doi.org/10.1016/j.gastre.2019.11.003

Sánchez-Peralta LF, Bote-Curiel L, Picón A, Sánchez-Margallo FM, Pagador JB (2020) Deep learning to find colorectal polyps in colonoscopy: a systematic literature review. Artif Intell Med 108:101923. https://doi.org/10.1016/j.artmed.2020.101923
Sang DV, Chung TQ, Lan PN, Hang DV, Van Long D, Thuy NT (2021) AG-CUResNeSt: a novel method for colon polyp segmentation. arXiv: 2105.00402
Segev L, Kalady MF, Plesec T, Mor E, Schtrechman G, Nissan A, Church JM (2020) The location of premalignant colorectal polyps under age 50: a further rationale for screening sigmoidoscopy. Int J Colorectal Dis 35(3):529–535. https://doi.org/10.1007/S00384-020-03504-2
Shamir RR, Duchin Y, Kim J, Sapiro G, Harel N (2019) Continuous dice coefficient: a method for evaluating probabilistic segmentations. https://doi.org/10.1101/306977
Shelhamer E, Long J, Darrell T (2017) Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 39(4):640–651. https://doi.org/10.1109/TPAMI.2016.2572683
Silva J, Histace A, Romain O, Dray X, Granado B (2014) Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. Int J Comput Assist Radiol Surg 9(2):283–293. https://doi.org/10.1007/S11548-013-0926-3/TABLES/5
Simeth J, Jiang J, Nosov A, Wibmer A, Zelefsky M, Tyagi N, Veeraraghavan H (2023) Deep learning-based dominant index lesion segmentation for MR-guided radiation therapy of prostate cancer. Med Phys. https://doi.org/10.1002/mp.16320
Singh S, Singh M, Gaur S (2022) Probiotics as multifaceted oral vaccines against colon cancer: a review. Front Immunol. https://doi.org/10.3389/fimmu.2022.1002674
Smedsrud PH, Thambawita V, Hicks SA, Gjestang H, Nedrejord OO, Næss E, Borgli H, Jha D, Berstad TJD, Eskeland SL, Lux M, Espeland H, Petlund A, Nguyen DTD, Garcia-Ceja E, Johansen D, Schmidt PT, Toth E, Hammer HL, Halvorsen P (2021) Kvasir-Capsule, a video capsule endoscopy dataset. Sci Data 8(1):142. https://doi.org/10.1038/s41597-021-00920-z
Spring M, Faulconbridge J, Sarwar A (2022) How information technology automates and augments processes: insights from Artificial-Intelligence-based systems in professional service operations. J Oper Manag 68(6–7):592–618. https://doi.org/10.1002/JOOM.1215
Sun X, Zhang P, Wang D, Cao Y, Liu B (2019) Colorectal polyp segmentation by U-Net with dilation convolution. In: 2019 18th IEEE international conference on machine learning and applications (ICMLA), pp 851–858. https://doi.org/10.1109/ICMLA.2019.00148
Tajbakhsh N, Gurudu SR, Liang J (2016) Automated polyp detection in colonoscopy videos using shape and context information. IEEE Trans Med Imaging 35(2):630–644. https://doi.org/10.1109/TMI.2015.2487997
Tang CP, Chang HY, Wang WC, Hu WX (2023) A novel computer-aided detection/diagnosis system for detection and classification of polyps in colonoscopy. Diagnostics 13(2):170. https://doi.org/10.3390/diagnostics13020170
Tashk A, Herp J, Nadimi E (2019) Fully automatic polyp detection based on a novel U-Net architecture and morphological post-process. In: Proceedings—2019 3rd international conference on control, artificial intelligence, robotics and optimization, ICCAIRO 2019, pp 37–41. https://doi.org/10.1109/ICCAIRO47923.2019.00015
Tomar NK, Shergill A, Rieders B, Bagci U, Jha D (2022) TransResU-Net: transformer based ResU-Net for real-time colonoscopy polyp segmentation. arXiv: 2206.08985
Tomar NK, Jha D, Bagci U (2023) DilatedSegNet: a deep dilated segmentation network for polyp segmentation. https://doi.org/10.1007/978-3-031-27077-2_26
Tran S-T, Nguyen M-H, Dang H-P, Nguyen T-T (2022a) Automatic polyp segmentation using modified recurrent residual Unet network. IEEE Access 10:65951–65961. https://doi.org/10.1109/ACCESS.2022.3184773
Tran S-T, Nguyen T-T, Le M-H, Cheng C-H, Liu D-G (2022b) TDC-Unet: triple unet with dilated convolution for medical image segmentation. Int J Pharma Med Biol Sci 11(1):1–7. https://doi.org/10.18178/ijpmbs.11.1.1-7
Uijlings JRR, Van De Sande KEA, Gevers T, Smeulders AWM (2013) Selective search for object recognition. Int J Comput Vis 104(2):154–171. https://doi.org/10.1007/S11263-013-0620-5
Wang Z, Wang E, Zhu Y (2020) Image segmentation evaluation: a survey of methods. Artif Intell Rev 53(8):5637–5674. https://doi.org/10.1007/S10462-020-09830-9/TABLES/6
Wang M, An X, Li Y, Li N, Hang W, Liu G (2021) EMS-Net: Enhanced multi-scale network for polyp segmentation. In: Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, EMBS, pp 2936–2939. https://doi.org/10.1109/EMBC46164.2021.9630787
Wen Y, Zhang L, Meng X, Ye X (2023) Rethinking the transfer learning for FCN based polyp segmentation in colonoscopy. IEEE Access 11:16183–16193. https://doi.org/10.1109/ACCESS.2023.3245519

Weng W, Zhu X (2015) U-Net: convolutional networks for biomedical image segmentation. IEEE Access 9:16591–16603

Wittenberg T, Raithel M (2020) Artificial intelligence-based polyp detection in colonoscopy: where have we been, where do we stand, and where are we headed? Visceral Med 36(6):428–438. https://doi.org/10.1159/000512438

Xiao WT, Chang LJ, Liu WM (2018) Semantic segmentation of colorectal polyps with DeepLab and LSTM networks. In: 2018 IEEE international conference on consumer electronics-Taiwan, ICCE-TW 2018. https://doi.org/10.1109/ICCE-CHINA.2018.8448568

Xu W, Li B, Liu S, Qiu W (2018) Real-time object detection and semantic segmentation for autonomous driving. Spiedigitallibrary.OrgB Li, S Liu, W Xu, W QiuMIPPR 2017: Automatic target recognition and navigation, 2018 spiedigitallibrary.Org, 44. https://doi.org/10.1117/12.2288713

Xu Y, Wang Y, Yuan J, Cheng Q, Wang X, Carson PL (2019) Medical breast ultrasound image segmentation by machine learning. Ultrasonics 91:1–9. https://doi.org/10.1016/J.ULTRAS.2018.07.006

Ye S, Pontius RG, Rakshit R (2018) A review of accuracy assessment for object-based image analysis: from per-pixel to per-polygon approaches. ISPRS J Photogramm Remote Sens 141:137–147. https://doi.org/10.1016/j.isprsjprs.2018.04.002

Yoon D, Kong HJ, Kim BS, Cho WS, Lee JC, Cho M, Lim MH, Yang SY, Lim SH, Lee J, Song JH, Chung GE, Choi JM, Kang HY, Bae JH, Kim S (2022) Colonoscopic image synthesis with generative adversarial network for enhanced detection of sessile serrated lesions using convolutional neural network. Sci Rep 12(1):1–12. https://doi.org/10.1038/s41598-021-04247-y

Yu C, Helwig EJ (2022) The role of AI technology in prediction, diagnosis and treatment of colorectal cancer. Artif Intell Rev 55(1):323–343. https://doi.org/10.1007/S10462-021-10034-Y

Zhou D, Tian F, Tian X, Sun L, Huang X, Zhao F, Zhou N, Chen Z, Zhang Q, Yang M, Yang Y, Guo X, Li Z, Liu J, Wang J, Wang J, Wang B, Zhang G, Sun B, Li X (2020) Diagnostic evaluation of a deep learning model for optical diagnosis of colorectal cancer. Nat Commun. https://doi.org/10.1038/s41467-020-16777-6