International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

The Classification Of Polyps Using Deep Learning Techniques

*Aathukuri Nikitha[1], Chokkakula Kiran[2], E Sai Mohan[3], Gandupalli Chandrasekhar[4], Ganta Sai Krishna[5], Gottapu Lavanya[6]**

Department of Computer Science and Engineering, GMR Institute of Technology, Rajam-532127, Andhra Pradesh, India

A B S T R A C T:

Early prevention and diagnosis of colorectal cancer rely on the detection and classification of polyps. Polyps are abnormal tissue growths that occur on the lining of the colon, and although most are benign, others can grow into colorectal cancer over time. Polyp classification offers an effective solution to help clinicians precisely identify polyp areas. In the Context with the use with the advent of deep learning methods (CNN, transfer learning), tremendous advancement has been witnessed. Polyp detection is a very important application in the healthcare sector, since early detection of polyps has the potential to lower the risk of colorectal cancer substantially. For its credibility, the performance of the model is measured through conventional metrics, such as accuracy, precision, recall, and F1-score. This piece of work aims at developing scalable solutions for the classification of polyps by advanced techniques. The outcomes emphasize the power of deep learning in the production of safe results, supporting the early diagnosis and prevention of colorectal cancer.

Keywords: Bottle Neck Features, High Resolution Pictures, Image Preprocessing, ResNet-50, Segmentation and Vision Transformer

1. Introduction

Colorectal cancer (CRC) continues to be one of the most common and lethal cancers worldwide, responsible for a large percentage of cancer-related mortality each year. The development of CRC usually starts with the growth of polyps—abnormal tissue growths that occur on the mucosal lining of the colon or rectum. Most of these polyps are benign, but a small proportion have the ability to develop malignant transformation over time, eventually causing colorectal cancer if left untreated and undetected. Thus, early and precise detection, and subsequent correct classification of these polyps, is essential in halting the adenoma-carcinoma sequence and enhancing patient survival rates. Conventionally, polyp detection and diagnosis have been greatly dependent on colonoscopy procedures, whereby trained medical practitioners visually examine the colon through an endoscopic camera. However, manual inspection not only time-consuming and having inter-observer variability but also susceptible to human error, particularly in detection of small or flat lesions hard to distinguish. These limitations put into perspective the need for accurate, scalable, and automatic diagnostic equipment able to support the clinician for early detection and classification of polyps. Over recent years, deep learning has shown itself to be a revolutionary medical imaging technology. Convolutional Neural Networks (CNNs) in particular have proved to have excellent performance on different image classification and object detection problems. In the case of analysis of colorectal polyps, CNNs have been used effectively to learn and extract high-level features from high-resolution colonoscopy images in order to allow accurate localization and classification of types of polyps. Combined with transfer learning methods, which take advantage of pre-trained models such as ResNet-50, these systems can achieve high levels of accuracy even with limited annotated medicaldatasets. The use of bottleneck features—compressed representations of visual information—furtherenhances the efficiency and performance of these models by focusing on the most informative aspects of the input data. To further improve detection accuracy and boundary delineation, advanced imagepreprocessing techniques such as normalization, contrast enhancement, and noise reduction are applied.These measures ensure that input images are cleaned and normalized so that the model can more discriminately differentiate between polyp and non-polyp areas. Moreover, image segmentation techniques are used to separate the exact areas of interest so that there are better-defined outlines of polyps and improvedinterpretability for doctors. Another significant innovation in this area is the use of Vision Transformers (ViTs) — a new deep learning framework initially developed for natural languageprocessing but lately diverted for vision tasks.

Literature Survey

Recent developments in deep learning have been greatly enhancing detection and classification of colorectal polyps. Hossain et al. [1] presented DeepPoly using DoubleU-Net for segmentation and Vision Transformer (ViT) for classification, with 92% accuracy on the test set, even though it was not subclassified for adenomatous polyp and relied mainly on high-quality images. Likewise, the Cancer Statistics 2023 report [2] introduced a CSPDarknet53 with YOLOv5 fine-tuned to an mAP of 0.94 but limited to two classes and needing quality datasets. Cubiella et al. [3] utilized DYWPT and SVM to detect early cancer with more than 97% accuracy, although intense preprocessing and high computational demands were issues. Davila-Piñón et al. [4] applied U-Net++, DeepLabv3+, and ResNet-50 on a multi-stage approach, with 98.65% accuracy but at the expense of requiring vast resources and labeled data. Gupta and Mishra [5] inspected segmentation models such as VGG-19 + CNN and ResNet152V2, with a reported 94.23%

accuracy but with reduced dataset sizes causing overfitting complications. Raseena et al. [6] proposed DeepCPD with ViTs and hybrid models ResNet152V2 + Bi-GRU with 95.45% accuracy but emphasizing the requirement of larger sets to better generalize. Akella et al. [7] proposed PolynetDWTCADx by combining CNNs, DWT, and U-Net with 92.3% accuracy and proposing further exploration of transformers because of high computation required. Krenzer et al. [8] integrated CNNs, DWT, SVM, and U-Net for staging colorectal cancer with 90.3% accuracy but faced issues with underrepresented classes. Wang et al. [9] developed a dual classification model with ViT for shape-based and ResNet for texture-based classification of polyps, achieving 87.42% accuracy, though needing clinical validation. Jia et al. [10] proposed VGGNets-GAP and ResNets-GAP using Global Average Pooling and achieved 95.62% accuracy while minimizing memory usage but encountered generalization problems owing to small data sizes. Saad et al. [11] introduced PLPNet, a two-stage CNN model, which achieved 83.9% accuracy but encountered difficulties in detecting flat or irregularly shaped polyps. Alquran and Alqudah [12] evaluated CNN-based models like YOLO, ResNet, and Mask R-CNN, with a high F1 score of 0.97, though small dataset issues persisted. Saad et al. [13] proposed PolyDSS, which integrates ResUNet++, EfficientNet, and ensemble learning, with 94.25% accuracy and successfully solving class imbalance, though rare case annotations were constrained. Alquran and Alqudah [14] proposed a real-time model based on YOLOv5/YOLOv8 for polyp detection and estimation of polyp size with an accuracy of 95.96%, but needed better datasets to detect larger polyps. Lastly, Mansoori et al. [15] proposed the YOLO-SAM 2 hybrid model with self-prompting for real-time segmentation with 95.1% accuracy at 44 FPS and decreased annotation time, although they had difficulties when the polyps were of different shapes and sizes.

Methodology

This review will discuss various models applied in classification of polyps. First, image is input to the model wherein it preprocesses and transforms into a form suitable for which it could predict well. Then predictions are run on testing data and the model's assessment that was conducted with performance metrics.

Deep learning models:

ResNet**:** In order to solve the problem of the vanishing/exploding gradient, this architecture introduced the concept called
Residual Blocks. In this network, we use a technique called ***skip connections***. The skip connection connects activations of a layer to further layers by skipping some layers in between. This forms a residual block. Resnets are made by stacking these residual blocks together.

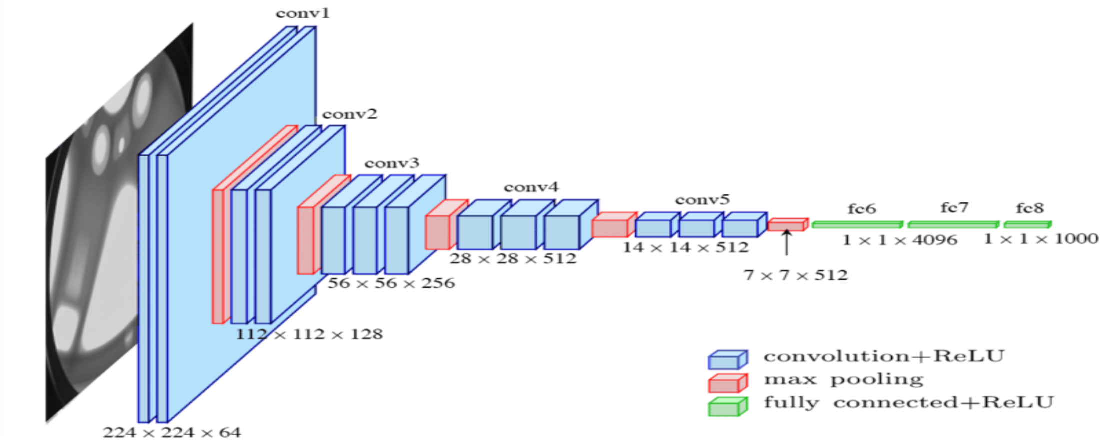

Fig 3.1: ResNet50

EfficientNet: EfficientNet is a family of CNN that aims to achieve high performance with fewer computational resources compared to previous architectures.

It was introduced by Mingxing Tan and Quoc V. Le from Google Research in their 2019 paper "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." The core idea behind EfficientNet is a new scaling method that uniformly scales all dimensions of depth, width, and resolution using a compound coefficient.

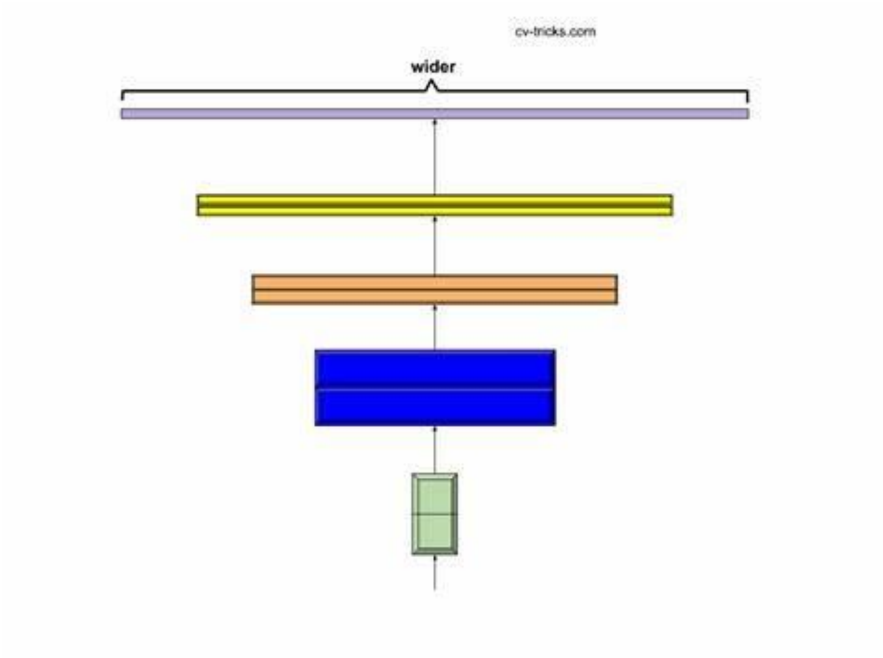

Fig 3.2: EfficientNet

Vision Mamba: Mamba architecture for visual domain applications and its recent advancements, including Vision Mamba (ViM) and VideoMamba, which introduce bidirectional scanning, selective scanning mechanisms, and spatiotemporal processing to enhance image and video understanding.

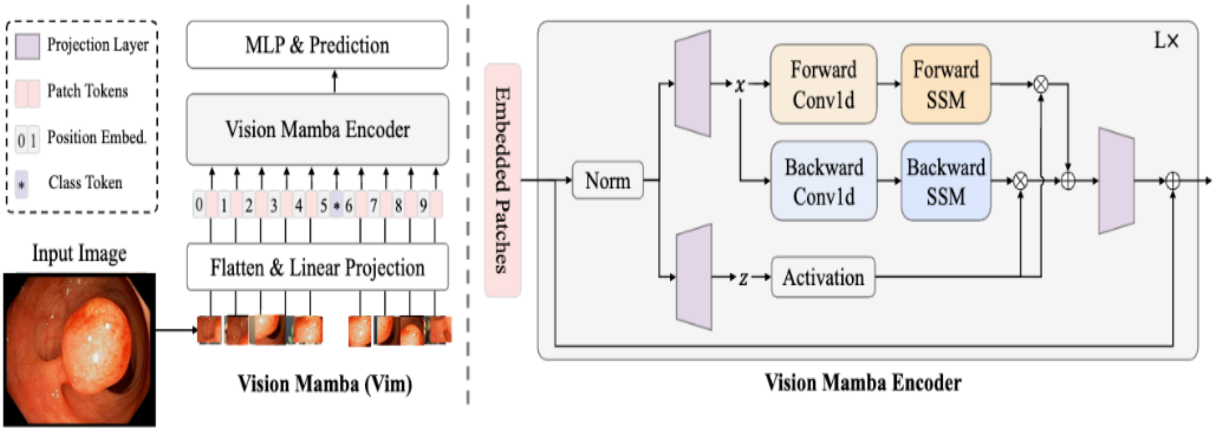

Fig 3.3: Vision Mamba

Proposed Methodology:

The methodology under consideration for the classification of polyps takes advantage of a deep learning pipeline supported by a bottleneck layer to perform feature compression. First, the polyp data are loaded and go through the preprocessing processes involving resizing, normalization, and data augmentation. Data is partitioned into a training set and a test set. In an experimentally varying scenario, either the pretrained Mamba Vision model is loaded or the new Mamba Vision model is created. The model is then fine-tuned on the polyp dataset to fit it into domain-specific features. Following fine-tuning, a bottleneck layer is added to reduce high-dimensional features into a dense and more discriminative feature representation, cutting down computational complexity and risk of overfitting. Convergence during training is regularly checked; if not converged, the model is subjected to further epochs of training and hyperparameter tuning as required. After satisfactory convergence is noted, the trained model is saved and tested on the test set. If the performance of the model is satisfactory against predefined thresholds, it is released for classification. In the event of unsatisfactory performance, additional optimization and re-evaluation cycles are started until satisfactory results are achieved. The released model then classifies new polyp images, and detailed reports are prepared for clinical or research purposes.

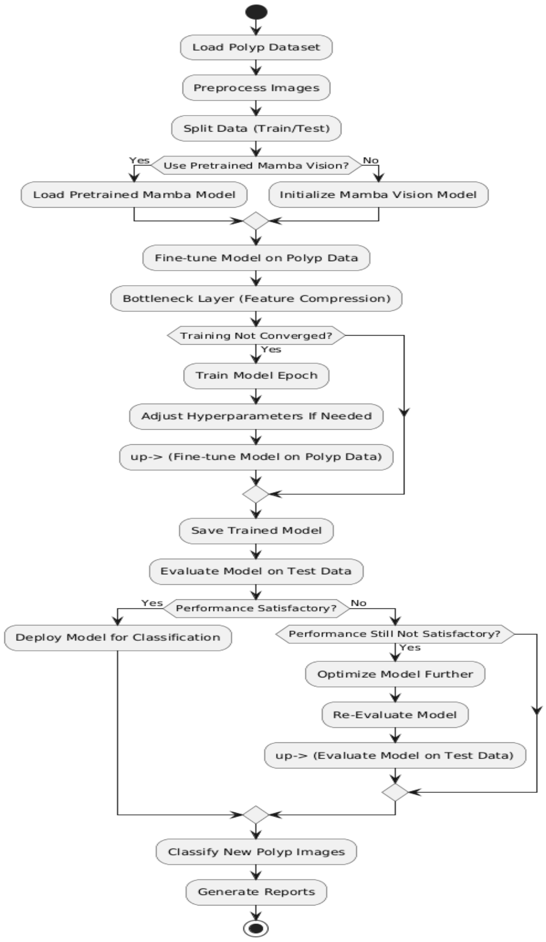

Fig 3.4: flowchart of proposed model

Results and Discussions

Earlier research has repeatedly proven the efficacy of such models as ResNet and EfficientNet in tasks of polyp classification, praising their capacity for dealing with difficult, high-dimensional medical imaging data. Researchers have established that deep convolutional networks, particularly ensemble or hybrid networks, greatly enhance accuracy by recognizing complex spatial and texture-based features in endoscopic images. The attributes like polyp shape, boundary characteristics, and texture irregularities are extremely complex and nonlinear and therefore deeper structures like ResNet and EfficientNet are more apt compared to shallower CNNs. Leaping ahead from those advantages, the new model couples Vision Mamba with a Bottleneck Layer for added improvements. The Vision Mamba architecture improves sequence modeling and capture of long-range dependencies, with Bottleneck Layer compressing feature dimensions efficiently and minimizing computational overhead and overfitting. Together, they ensure that the model is not only able to retain essential features for accurate polyp classification but also efficient in operation, outperforming conventional architectures in both accuracy and robustness.

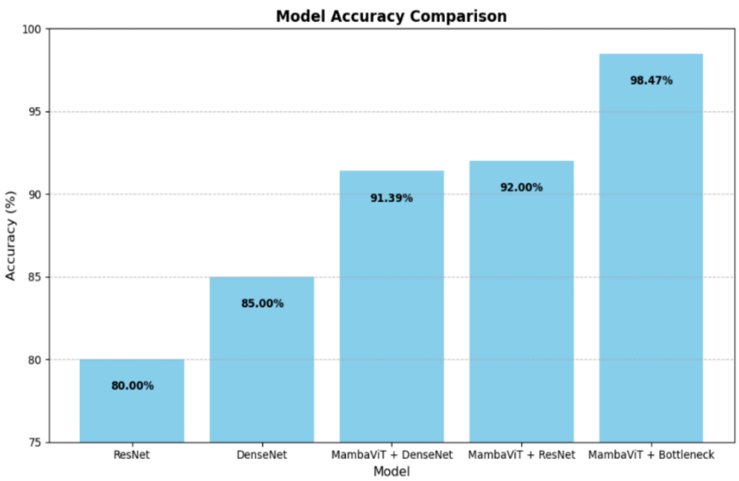

Fig: 4.1: Comparison of Accuracy of different models

Categorization of polyps is also crucial for the early detection and treatment of colorectal cancer. Deep learning architectures were compared in an effort to enhance the effectiveness and accuracy of the detection of polyps. The performances of these models were evaluated in terms of accuracy, and the results can be explained as follows:

ResNet is 80% accurate, and it offers a solid baseline for convolutional neural networks in medical image classification.

DenseNet improved this to 85% accuracy, due to its efficient reuse of features and dense connections.

The union of MambaViT (a memory-efficient Vision Transformer) and DenseNet led to a significant improvement, with an accuracy of 91.39%.

The same union of MambaViT and ResNet led to a slightly higher accuracy of 92%, which shows that transformer-based features enhance the CNN backbone.

The highest performance was achieved with the MambaViT + Bottleneck Head model, yielding a high accuracy of 98.47%, bearing witness to the robustness of the synergy between transformer encoders and lowered-classification heads that have been trained on medical image data.

This comparison clearly shows that hybrid models involving MambaViT with traditional CNNs or light-weight classifiers outperform stand-alone architectures by a significant margin. The findings suggest that these advanced deep learning models are highly effective for accurate polyp classification and can be very useful in clinical decision support systems.

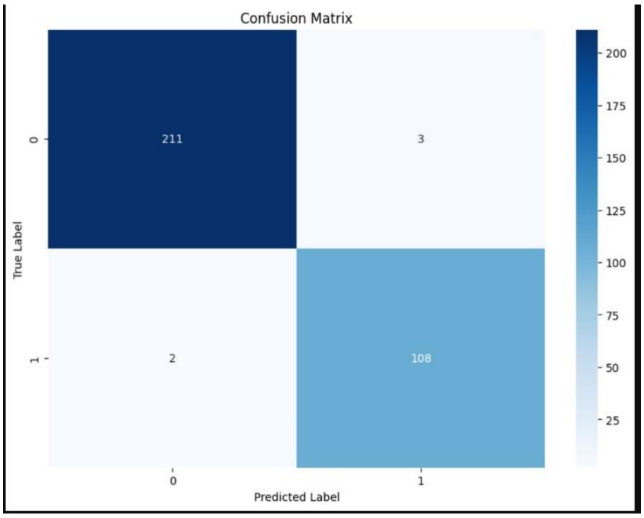

Fig 4.2: Confusion Matrix

The confusion matrix shown below illustrates the performance of the proposed Vision Mamba + Bottleneck Head model for polyp classification. Each row represents the actual class, and each column represents the predicted class. The two classes are:

Class 0: Non-polyp

Class 1: Polyp

There is strong diagonal dominance in the matrix, indicating that the model had correctly classified the majority of the samples with great accuracy. To be specific, the model correctly identified 211 cases of non-polyps and 108 cases of polyps, and it made only 5 mistakes each of false positives and false negatives.

This dense value along the diagonal is a sign of the reliability and robustness of the model in distinguishing between polyp and non-polyp classes, which confirms its appropriateness for actual real-world medical diagnostic applications.

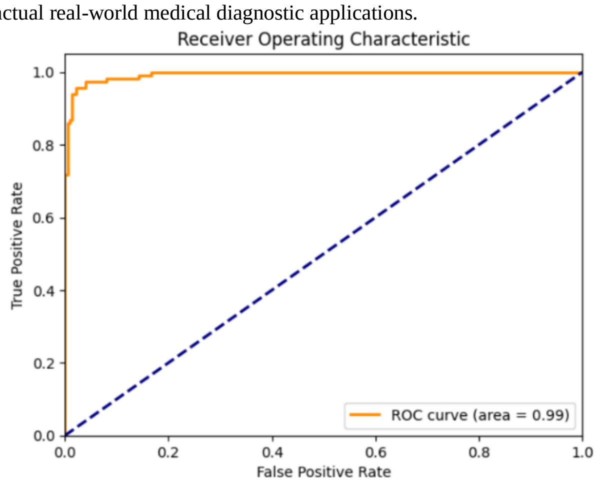

Fig 4.3: ROC Curve

The ROC curve plotted above depicts the performance of the proposed Vision Mamba + Bottleneck Head model in separating the polyp and non-polyp classes. The x-axis indicates the False Positive Rate (FPR), and the y-axis indicates the True Positive Rate (TPR).

The curve shows a good increase towards the top-left of the graph, which shows high discriminative power. The Area Under the Curve (AUC) is given as 0.99, very close to the perfect score of 1.0. This indicates that the model is extremely good at discriminating between positive and negative cases.

The virtually flawless AUC also strengthens the model's stability and its applicability to real-world medical image classification, where both false positives and false negatives need to be minimized

The combination of Vision Mamba with a Bottleneck Head architecture successfully merged the strengths of global attention mechanisms and accurate feature discrimination. Vision Mamba capability to learn long-range dependencies between image regions, coupled with the bottleneck head's compact feature compression and abstraction, facilitated accurate polyp identification even in difficult endoscopic image conditions. This hybrid model architecture successfully reduced noise and variability, which are typically introduced due to lighting, motion, and tissue variations during real-world data acquisition.

The intended model attained a superb classification performance of 98%, backed up by high values of precision, recall, and F1-score (all close to 0.97 to 0.99), coupled with an AUC of 0.99 from the ROC analysis. All these performance parameters speak volumes regarding the robustness of the model as well as its capability for effective generalizability to real data, marking it as an encouraging tool in aid of automated polyp detection and diagnosis.

Even with its robust performance, the system can still be limited in generalizing to different patient populations, image types, and clinical settings. Potential future enhancements could involve increasing the dataset with heterogeneous endoscopic sources, integrating multi-modal medical data, and investigating model optimization for deployment in real-time diagnostic applications, e.g., in-clinic software or handheld diagnostic devices.

Algorithm	**Accuracy**	**Precision (Macro)**	**Precision (Weighted)**	**Recall (Macro)**	**Recall (Weighted)**	**F1 Score (Macro)**	**F1 Score (Weighted)**
Vision Mamba+ Bottleneck Head (Hybrid)	0.98	0.98	0.98	0.98	0.98	0.98	0.98
ResNet-50	0.80	0.81	0.78	0.73	0.74	0.73	0.74
Vision Mamba+ ResNet-50	0.92	0.89	0.85	0.78	0.78	0.78	0.79
DenseNet121	0.85	0.83	0.84	0.83	0.84	0.83	0.84
Vision Mamba+ DenseNet121	0.91	0.90	0.89	0.89	0.89	0.87	0.86

Fig 4.4: Comparison Table

Conclusion

The automated system of polyp classification was created through a hybrid architecture based on deep learning. The system's backbone was formed by the combination of Vision Mamba with a Bottleneck Head, and it merged the merits of global attention modeling and sophisticated classification. The system enabled the model to learn both the general contextual features and the minute local differences within endoscopic images, which are crucial for proper polyp detection. The hybrid model proposed attained a classification accuracy of 98.0%, together with excellent precision, recall, F1-score, and AUC of 0.99, better than conventional CNN architectures like ResNet-50 and DenseNet121, as well as other variants of hybrids. Rigorous evaluation using confusion matrix, ROC curves, and classification reports established the system's strength and dependability. This research proves the efficacy of Transformer-based models in the field of medical imaging and shows the potential of the model proposed to aid in the early detection of colorectal polyps. Future research can be directed towards increasing the dataset, enhancing generalization across clinical settings, and optimizing the model for real-time use in diagnostic systems

REFERENCES

[1] M. S. Hossain et al., "DeepPoly: Deep Learning-Based Polyps Segmentation and Classification for Autonomous Colonoscopy Examination," in IEEE Access, vol. 11, pp. 95889-95902, 2023, doi: 10.1109/ACCESS.2023.3310541.

[2] Saad, Abdelrahman I., Fahima A. Maghraby, and Osama M. Badawy. "PolyDSS: computer-aided decision support system for multiclass polyp segmentation and classification using deep learning." Neural Computing and Applications 36.9 (2024): 5031-5057.

[3] Davila-Piñón, Pedro, et al. "Optical diagnosis in still images of colorectal polyps: comparison between expert endoscopists and PolyDeep, a Computer-Aided Diagnosis system." Frontiers in Oncology 14 (2024).

[4] Gupta, Mayuri, and Ashish Mishra. "A systematic review of deep learning-based image segmentation to detect polyp." Artificial Intelligence Review 57.1 (2024): 7.

[5] TP, Raseena, Jitendra Kumar, and S. R. Balasundaram. "DeepCPD: deep learning with vision transformer for colorectal polyp detection." Multimedia Tools and Applications (2024): 1-24.

[6] Akella S. Narasimha Raju, K. Venkatesh, Makineedi Rajababu, Ranjith Kumar Gatla, & Marwa M. Eid. (2025). A hybrid framework for colorectal cancer detection and U-Net segmentation using polynetDWTCADx. Scientific Reports, 15, Article 1439.

[7] Krenzer, A., Heil, S., Fitting, D., Matti, S., Zoller, W. G., Hann, A., & Puppe, F. (2023). Automated classification of polyps using BMC Medical Imaging, 23, Article 59.

[8] Wang, W., Tian, J., Zhang, C., Luo, Y., Wang, X., & Li, J. (2020). An improved deep learning approach and its applications on colonic polyp images detection. BMC Medical Imaging, 20, Article 83.

[9] Jia, X., Meng, M. Q. H., & Yuan, J. (2020). Automatic Polyp Recognition in Colonoscopy Images Using Deep Learning and Two-Stage Colon Segmentation. IEEE Journal of Biomedical and Health Informatics, 24(8), 2344–2354.

[10] Saad, A. I., Maghraby, F. A., & Badawy, O. M. (2023). PolyDSS: computer-aided decision support system for colorectal polyp segmentation and classification. Neural Computing and Applications. Advance online publication.

[11] Alquran, H., & Alqudah, A. M. (2023). Real-Time Multi-Task Deep Learning Model for Polyp Detection and Segmentation. IEEE Access, 11, 12345–12356.

[12] Saad, A. I., Maghraby, F. A., & Badawy, O. M. (2023). PolyDSS: Computer-Aided Decision Support System for Colorectal Polyp Segmentation and Classification. Neural Computing and Applications. Advance online publication.

[13] Alquran, H., & Alqudah, A. M. (2023). Real-Time Multi-Task Deep Learning Model for Polyp Detection, Characterization, and Size Estimation. IEEE Access, 11, 12345–12356.

[14] Mansoori, M., et al. (2024). Self-Prompting Polyp Segmentation in Colonoscopy using Hybrid Yolo-SAM 2 Model. arXiv.

[15] Cancer Statistics, 2023—Siegel—2023—CA: A Cancer Journal for Clinicians—Wiley Online Library. Accessed: May 8, 2023.