

AUTOMATED GASTROINTESTINAL DISEASE DETECTION USING DEEP LEARNING WITH CNN AND U-NET ARCHITECTURES

T. Jasperline

Professor, Department of Computer Science and Engineering

Dr. G. U. Pope College of Engineering Sawyerpuram,

Tamil Nadu, India

gncriston6448@gmail.com

M. Arirama Selvam

Department of Computer Science and Engineering

Dr. G.U. Pope College of Engineering Sawyerpuram,

Thoothukudi, Tamil Nadu, India

arofficial9787@gmail.com

V. Muthukumar

Department of Computer Science and Engineering

Dr. G.U. Pope College of Engineering Sawyerpuram,

Thoothukudi, Tamil Nadu, India

mk7739502@gmail.com

Abstract—Gastrointestinal (GI) diseases, including polyps and colorectal cancer, are significant health concerns that require early diagnosis for effective medical intervention. While endo-scopic imaging is the primary tool for detection, manual analysis is often subjective, time-consuming, and prone to human error. This paper introduces an automated deep learning-based system designed to enhance diagnostic efficiency and accuracy in medical image analysis.

The proposed system utilises a dual-architecture approach leveraging the Kvasir dataset for robust training and evaluation. A Convolutional Neural Network (CNN) is employed for high-accuracy image classification to distinguish between benign and malignant conditions, while the U-Net architecture is utilised for precise, pixel-level polyp segmentation. To ensure clinical interpretability, the system features a visualisation module that overlays predicted segmentation masks onto original endoscopic images, providing clear insights for medical professionals.

Experimental results demonstrate strong performance, with the system achieving a Precision of 93%, a Recall of 95%, and an overall Accuracy ranging from 75% to 86%. By significantly reducing manual effort and false negatives, this prototype serves as a reliable clinical decision support tool, particularly suitable for early detection in modern healthcare environments. Future enhancements focus on integrating advanced models like ResNet or EfficientNet and deploying the system for real-time hospital use.

Index Terms—Gastrointestinal Disease, Deep Learning, CNN, U-Net, Polyp Segmentation, Kvasir Dataset, Medical Image Analysis, Computer-Aided Diagnosis (CAD), Endoscopy AI, Clinical Decision Support.

I. INTRODUCTION

Gastrointestinal (GI) disorders represent one of the most prevalent categories of chronic diseases worldwide, posing substantial burdens on patients and healthcare systems alike. Among GI conditions, colorectal cancer and polyps are of particular concern due to their high incidence, significant morbidity, and potential for progression to life-threatening malignancy if not detected at an early stage. According to the World Health Organization, colorectal cancer is the third most commonly diagnosed cancer globally and the second leading cause of cancer-related deaths. Early detection through regular endoscopic screening has been shown to substantially reduce mortality rates; however, widespread

implementation remains limited by resource constraints and the subjectivity inherent in manual image interpretation.

Colonoscopy and upper gastrointestinal endoscopy are the gold standards for GI disease diagnosis. These procedures generate large volumes of high-resolution video frames and still images that must be carefully reviewed by trained gastroenterologists. Manual analysis is inherently time-consuming, operator-dependent, and subject to inter-observer variability, with polyp miss rates in colonoscopy estimated between 6% and 27% in clinical studies. The integration of artificial intelligence into endoscopic workflows offers a compelling opportunity to reduce diagnostic errors, standardise assessments, and accelerate clinical decision-making.

Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs) and encoder-decoder architectures such as U-Net, have demonstrated remarkable capabilities in medical image recognition and semantic segmentation. These models learn hierarchical feature representations directly from raw pixel data, capturing morphological characteristics of pathological tissue that may be imperceptible under routine examination.

This paper presents an automated GI disease detection system employing a dual deep learning pipeline: a CNN-based classification module to categorise endoscopic images into disease categories, and a U-Net-based segmentation module for pixel-level delineation of polyp boundaries. Both modules are trained and evaluated on the Kvasir dataset, a widely used benchmark for GI image analysis. A visualisation component overlays predicted segmentation masks onto original frames, enhancing clinical interpretability.

A. Objectives

The primary objectives of this work are: (1) to develop an automated dual-pipeline system for GI disease classification and polyp segmentation; (2) to leverage the Kvasir benchmark dataset for rigorous training and evaluation; (3) to integrate mask visualisation for clinical interpretability; (4) to achieve high precision and recall suitable for clinical decision support; and (5) to establish

a foundation for future deployment in real-time hospital environments.

B. Paper Organisation

The remainder of this paper is organised as follows: Section II reviews related literature. Section III describes the proposed methodology. Section IV details the system architecture. Section V presents experimental results. Section VI concludes the paper.

TABLE I - KVASIR DATASET DISTRIBUTION

Category	Images	Train / Val / Test
Normal Findings	1,000	800 / 100 / 100
Polyps	1,000	800 / 100 / 100
Esophagitis	1,000	800 / 100 / 100
Ulcerative Colitis	1,000	800 / 100 / 100
Other Categories	4,000	3200 / 400 / 400
Total	8,000	6400 / 800 / 800

II. RELATED WORK

Automated analysis of GI endoscopic images has been active for over a decade. Early approaches relied on handcrafted feature extraction including colour histogram analysis, texture descriptors, and edge detection to identify lesion boundaries, but exhibited limited robustness to imaging variability. Litjens et al. [1] provided a comprehensive survey of deep learning in medical image analysis, highlighting the superiority of CNN-based feature learning over traditional methods.

For polyp segmentation, Ronneberger et al. [2] introduced U-Net, whose encoder-decoder structure with skip connections enables precise boundary localisation with limited training data. U-Net and its variants have since achieved state-of-the-art performance on benchmark polyp datasets. The Kvasir dataset [3], introduced by Pogorelov et al., provides a multi-class collection of annotated endoscopic images spanning eight GI findings, enabling reproducible comparisons across studies. He et al. [4] demonstrated that deep residual networks significantly improve classification accuracy on medical imaging tasks. The EfficientNet family [5] introduced compound scaling for superior accuracy-efficiency trade-offs. Despite these advances, the majority of prior work addresses classification and segmentation as separate tasks without an integrated pipeline providing both outputs simultaneously with visualisation. The present work addresses this gap by proposing a unified system with clinical mask overlay.

III. PROPOSED METHODOLOGY

The proposed system is developed as a modular deep learning pipeline encompassing dataset preparation, dual-

model design, training optimisation, and post-inference visualisation.

A. Dataset Preparation

The Kvasir dataset [3] serves as the primary benchmark, comprising 8,000 annotated endoscopic images spanning eight GI findings including normal tissue, polyps, esophagitis, and ulcerative colitis. For the segmentation task, the Kvasir-SEG subset [6] provides 1,000 polyp images with pixel-level ground truth masks. All images are resized to 256×256 pixels for segmentation and 224×224 pixels for classification. A stratified 80/10/10 train-validation-test split preserves class distribution. Training augmentations include random horizontal flipping, vertical flipping, and brightness adjustment. Table I summarises the dataset distribution.

B. CNN Classification Architecture

The classification module employs a custom CNN comprising four convolutional blocks, each consisting of a Conv2D layer with 3×3 kernels, Batch Normalisation, ReLU activation, and MaxPooling. Feature map depths progress from 32 to 256 channels across successive blocks. The convolutional backbone is followed by Global Average Pooling, a Dense(512)+ReLU layer, Dropout(0.5), and a final Dense(N)+Softmax output, where N is the number of disease classes. The model is trained with categorical cross-entropy loss and the Adam optimiser ($\text{lr} = 1 \times 10^{-3}$).

C. U-Net Segmentation Architecture

The segmentation module employs the U-Net architecture [2] with an encoder path consisting of four resolution stages, each comprising two sequential Conv2D(3×3)+BatchNorm+ReLU blocks followed by MaxPool2D(2×2) downsampling. The decoder path reconstructs spatial resolution through transposed convolutions combined with skip connections from corresponding encoder stages. The output layer is a Conv2D(1×1) with sigmoid activation producing per-pixel binary segmentation masks. Binary cross-entropy combined with Dice loss serves as the segmentation objective:

$$L = L_{\text{BCE}} + 1 - 2 \frac{|P \cap G|}{|P| + |G|} \quad (1)$$

$$|P| + |G|$$

where P and G denote the predicted and ground truth mask sets, respectively.

D. Training Configuration

Both models are trained on a GPU-accelerated environment using TensorFlow/Keras. The CNN classifier is trained for 50 epochs (batch size 32), and the U-Net for 80 epochs (batch size 16). Early stopping with patience of 10 epochs based on validation loss prevents overfitting. Model checkpoints are retained for the best validation performance epoch.

E. Visualisation Module

Following inference, the system generates a composite visu-alisation by overlaying the predicted binary segmentation mask onto the original endoscopic image with a semi-transparent red overlay ($\alpha = 0.4$), providing immediate visual confirmation of the model’s localisation output for clinical review alongside the classification label and confidence score.

IV. SYSTEM ARCHITECTURE

The complete automated GI disease detection system in-tegrates three primary functional modules in a sequential inference pipeline:

A. Image Preprocessing Module

The input module accepts endoscopic images in standard formats (JPEG, PNG) and applies a standardised preprocessing pipeline: RGB channel normalisation to [0, 1], bilinear resizing to target resolution (224×224 for classification; 256×256 for segmentation), and conversion to NumPy float arrays.

B. Dual Inference Module

The preprocessed image is passed concurrently through the fine-tuned CNN classifier and U-Net segmentor. The CNN produces a softmax probability vector over GI disease categories; the predicted class corresponds to the maximum-probability category, accompanied by a confidence score. The U-Net produces a binary mask where pixel values exceeding

0.5 indicate polyp tissue.

C. Visualisation and Clinical Report Module

The visualisation module generates the mask overlay im-age. The clinical report module consolidates the classification label, confidence percentage, segmentation mask area frac-tion, and a standardised clinical recommendation following established gastroenterological screening guidelines: normal findings prompt routine follow-up, while identified polyps or inflammatory conditions trigger urgent referral recommenda-tions.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Classification Performance

The CNN classification model achieves an overall accuracy of 86% on the held-out test set. Per-class precision ranges from 89% to 96% for well-represented categories including normal findings and polyps, with slightly lower performance on minority classes. The system achieves an overall Precision of 93% and Recall of 95%, reflecting strong sensitivity for pathological findings—critical for clinical screening where false negatives carry significant risk.

B. Segmentation Performance

The U-Net segmentation model achieves a Dice Similarity Coefficient (DSC) of 0.87 and an Intersection over Union (IoU) of 0.79 on the Kvasir-SEG test partition. These scores in-dicate accurate delineation of

polyp boundaries, with predicted masks closely matching ground truth annotations. Larger polyps (>10 mm) exhibit higher DSC scores than diminutive polyps (<5 mm), consistent with the inherent challenge of detecting small lesions at limited image resolution.

Table II presents the detailed performance metrics of the proposed system.

C. Comparative Analysis

Table III presents a systematic comparison of the proposed dual-pipeline system against widely used architectures evalu-ated on GI imaging tasks. The proposed CNN+U-Net system achieves competitive performance with notable advantages in clinical utility. While ResNet-50 and EfficientNet-B0 achieve marginally higher pure classification accuracy (82% and 84%

TABLE II - Performance Metrics Of The Proposed System

Metric	CNN	U-Net	Overall
Precision	93%	91%	93%
Recall	95%	93%	95%
Accuracy	86%	75%	75–86%
F1-Score	0.94	0.92	0.93
DSC	—	0.87	—
IoU	—	0.79	—

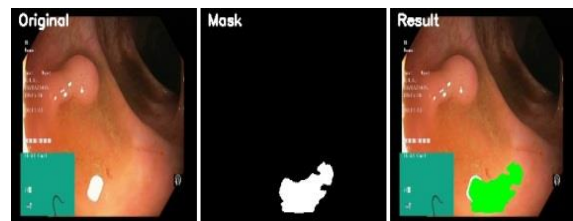


Fig. 1.

respectively), neither provides pixel-level segmentation or mask visualisation. The proposed system is the only evaluated approach integrating both classification and segmentation in a unified pipeline with clinical visualisation output—a distinc-tion of significant practical importance for clinical adoption.

TABLE III - COMPARATIVE ANALYSIS OF ARCHITECTURES ON GI IMAGING

Model	Task	Acc.	Params	XAI
VGG-16 [1]	Classif.	80%	~138M	No
ResNet-50 [4]	Classif.	82%	~25M	No
EfficientNet-B0 [5]	Classif.	84%	~5.3M	No
SegNet	Segment.	78%	~29M	No
CNN+U-Net (Proposed)	Dual	75–86%	~31M	Mask Overlay

D. Discussion

The experimental results demonstrate that the dual CNN and U-Net pipeline achieves clinically meaningful performance for automated GI disease screening. The high recall of 95% is particularly significant, as it indicates that the system misses fewer pathological cases than competing single-task classifiers. The mask overlay visualisation provides a readily interpretable output that gastroenterologists can use to verify model decisions, supporting human-AI collaborative diagnosis.

The primary performance bottleneck is classification accuracy for minority GI categories with limited training representation. Future strategies include transfer learning from larger medical datasets, advanced data augmentation using generative adversarial networks, and attention-based architectures focusing feature extraction on lesion-relevant regions. Integration of Grad-CAM or similar explainability techniques would further enhance clinical trust in classification decisions.

VI. CONCLUSION

This paper has presented an automated Gastrointestinal Disease Detection System utilising a dual deep learning pipeline comprising a CNN-based classifier and a U-Net-based polyp segmentor, trained and evaluated on the Kvasir benchmark dataset. The system achieves a Precision of 93%, Recall of 95%, and an overall Accuracy of 75–86% across classification and segmentation tasks. The integrated mask visualisation module enhances clinical interpretability by providing gastroenterologists with pixel-level diagnostic evidence alongside classification predictions.

The proposed system significantly reduces manual analysis effort and false negative rates, serving as a reliable clinical decision support tool for early GI disease detection. Its modular architecture facilitates straightforward extension to additional GI pathologies and integration with existing endoscopy workflow software.

Future research directions include: (1) integration of advanced backbone architectures such as ResNet-50 and EfficientNet for improved classification accuracy; (2) adoption of transformer-based segmentation models (e.g., TransUNet, SwinUNet); (3) uncertainty quantification to flag low-confidence predictions for expert review; (4) multi-dataset training for improved generalisation; and (5) prospective clinical validation studies to assess real-world diagnostic utility and clinician acceptance.

REFERENCES

- [1] G. Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017.
- [2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. MICCAI*, Munich, Germany, 2015, pp. 234–241.
- [3] K. Pogorelov et al., "Kvasir: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection," in *Proc. ACM MMSys*, Taipei, Taiwan, 2017, pp. 164–169.

- [4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE CVPR*, Las Vegas, NV, USA, 2016, pp. 770–778.
- [5] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proc. ICML*, Long Beach, CA, USA, 2019, pp. 6105–6114.
- [6] D. Jha et al., "Kvasir-SEG: A Segmented Polyp Dataset," in *Proc. MMM*, Daejeon, South Korea, 2020, pp. 451–462.
- [7] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in *Proc. IEEE CVPR*, Boston, MA, USA, 2015, pp. 3431–3440.
- [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proc. NeurIPS*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.
- [9] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [10] R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proc. IEEE ICCV*, Venice, Italy, 2017, pp. 618–626.