Hybrid Deep Learning for Gastrointestinal Disease Classification and Lesion Segmentation Using Kvasir Dataset

Amna Noreen, Aqleema Mehmood

2022-BSE-005, 2022-BSE-006
Department of Software Engineering, [Fatima Jinnah Women University], [Rawalpindi, Pakistan]
Email: {amnanoreen353@gmail.com, aqleemamehmood@gmail.com

***Abstract*—Gastrointestinal diseases pose a substantial burden on the global disease burden and demand accurate and early diagnosis. Deep learning algorithms have made promising contributions to the diagnosis of such diseases based on endoscopic images, but there are some limitations such as low contrast, noise, or complicated structures. This paper presents a modified framework for the diagnosis of gastrointestinal diseases by combining modern image enhancement strategies like contrast enhancement, noise reduction, gamma correction, and edge fusion with the deep learning model. A classification model based on the convolutional neural network framework EfficientNet will be used, and a modified decoder based on the interpretation of the U-Net framework with the pseudo-mask generated by the U²-Net will also be implemented for secondary segmentation learning. The proposed framework has been tested on the Kvasir gastrointestinal image dataset, and the results have confirmed the substantial improvement made by combining image enhancement strategies and joint segmentation and classification learning for the diagnosis of gastrointestinal diseases. Various comparative analysis results also have validated the efficiency and effectiveness of the proposed approach for the diagnosis of gastrointestinal diseases.**

***Index Terms*—Gastrointestinal disease diagnosis, Endoscopic image analysis, Deep learning, Image enhancement, Multi-task learning, EfficientNet, U^2-Net**

I. Introduction

Detection and diagnosis of gastrointestinal diseases through endoscopic imaging remain among the most critical tasks in modern medical diagnostics. The broad range of gastrointestinal pathologies, from benign inflammatory processes to precancerous lesions and malignant transformations, stresses the need for early detection, valid diagnosis, and appropriate treatment for the improvement of patient outcomes. In traditional diagnosis, much reliance is placed on manual examination and interpretation by expert clinicians of observations made during endoscopy, which is time-consuming, subjective, and can be variable depending on the experience of the clinician and the conditions under which the procedure is conducted. In light of the recent rapid developments within machine learning and computer vision, there is an emerging potential to automate and standardize the interpretation of endoscopic images to minimize workload, enhance diagnostic uniformity, and probably catch some pathologies which human observers might have overlooked.

Most recently, deep learning has seen phenomenal success in various imaging domains like radiology, dermatology, or histopathology, which in turn has sparked growing interest in applying similar approaches to endoscopic GI imaging. The main limitation to deploying ML in this domain is the lack of publicly available and well-annotated datasets. Often, medical image data are limited due to issues such as patient privacy, variable imaging protocols, and expert annotation. The scarcity of such data severely limits the reproducibility of research, the possibility of making fair comparisons between different methods, and overall progress in general.

Within this context, it is the Kvasir dataset that has become a vital resource. From actual endoscopic examinations curated at its inception, Kvasir includes thousands of high-resolution images, each certified by experienced endoscopists, representing a wide range of anatomical landmarks, common GI pathologies, and endoscopic procedures. By providing a standard dataset for research, reproducible experimentation considering classification, segmentation, detection, and retrieval tasks is achieved through the Kvasir; these form the bases for developing CAD systems for GI diseases.

The barrier to entry for researchers and developers looking to create AI-based diagnostic tools just got lowered. It allows direct comparisons across studies, fosters benchmarking, and promotes best practices. Further, since Kvasir contains both normal and pathological cases across several disease types and procedural contexts, models trained on it are more likely to generalize better-a key characteristic of real-world clinical settings. In this paper, based on the Kvasir dataset, we seek to develop a deep-learning classifier, setting it in a wider context while aiming to contribute toward fully automated, accurate, and scalable GI disease diagnosis. We further conjecture that a system carefully combining strong pre-processing, data augmentation, and attention to model design detail may be capable of achieving performance rivaling - or even exceeding - that of human experts for some diagnostic/detection tasks, and therefore offers practical assistance for the clinician. The convergence of high-quality open datasets like Kvasir with modern ML/CV techniques promises a future in which automated diagnostics of GI will become more accessible, reducing workload for clinicians, enhancing diagnostic consistency, and bringing about early and reliable detection for improved patient outcomes.

II. Literature Review

A. Related Work

This work presents a critical review of existing research related to the analyses of gastrointestinal diseases, focusing mainly on polyp detection, segmentation, and classification. The reviewed works have been grouped into dataset development, segmentation approaches, detection frameworks, classification techniques, and surveys and clinical validation studies.

B. Public Gastrointestinal Datasets

Early concepts of vision-based localization and segmentation were presented by Sharma [1], inspiring subsequent medical segmentation tasks. In this regard, one of the significant contributions was the proposal of the Kvasir-SEG dataset by Jha et al. [2], presenting colonoscopy images with pixel-wise annotations for supervised segmentation of polyps. Borgli et al. [3] extended the availability of datasets with a new, extensive multi-class image and video dataset for gastrointestinal endoscopy entitled HyperKvasir.

Ali et al. [4] presented the PolypGen dataset, a multi-center dataset for cross-domain robustness evaluation to address the issue of limited generalization across clinical centers. These datasets collectively drive most of the recent advances in deep learning-based gastrointestinal image analysis.

C. Deep Learning-Based Polyp Segmentation

Polyp segmentation remains one of the essential issues since it plays a crucial role in clinical decision support. In this respect, Ramzan et al. [5] developed the Graft-U-Net model, improving the traditional U-Net architecture by enhancing skip connections and achieving high accuracy in segmentation. Later on, Tomar et al. [6] proposed DDANet, a dual-decoder attention network that performs better regarding boundary localization owing to parallel processing of the spatial and contextual information. Finally, Vezakis et al. [7] modified EffiSegNet from a pre-trained EfficientNet encoder by simplifying the decoder for computational efficiency.

Park et al. [8] proposed a semi-supervised consistency training framework with pseudo-label updates to reduce reliance on annotated datasets. Saad et al. [9] proposed PolySeg Plus, which integrates deep learning with active learning to reduce annotation costs. Manan et al. [10] proposed DPE-Net that uses a dual-parallel encoder to capture both fine and coarse features simultaneously.

Recently, several segmentation approaches have been based on transformers. Among them, Nachmani et al. [11] introduced a real-time segmentation approach based on a pyramid vision transformer with residual blocks. Rajasekar et al. [12], [13] combined wavelet transformation with AdaptUNet to enhance boundary feature extraction. Bernal et al. [14] presented one of the earliest structured discussions of polyp segmentation in colonoscopy images; thus, serving as a baseline reference. Song and Shin [15] contributed semantic polyp generation to enhance segmentation performance through synthetic data augmentation.

D. Polyp Detection Approaches

Object detection models allow for direct localization of the polyps within the colonoscopy frames. Sahoo et al. [16], [17] showed the performance of YOLOv11 in detecting polyps in real time. ELKarazle et al. [18] presented a comprehensive review of machine learning-based detection systems. Shen et al. [19] confirmed the clinical reliability of deep learning-based detection and classification systems by validating them at multi-site hospitals. Lalinia and Sahafi [20] further improved the detection performance based on an optimized YOLOv8 architecture.

Abbas et al. [21] proposed a high-performance semantic segmentation network capable of segmenting polyps and surgical instruments simultaneously, thus improving clinical usability. Detection systems continue to evolve with improved generalization and inference speed.

E. Gastrointestinal Disease and Polyp Classification

Beyond segmentation and detection, classification approaches also come into prominence. Al-Adhaileh et al. [22] applied deep learning models for multi-class gastrointestinal disease detection and classification using the Kvasir dataset. Demirbas et al. [23] proposed a spatial-attention ConvMixer architecture for classification and detection of gastrointestinal diseases using Kvasir.

Carvalho et al. [24] introduced a NICE-based polyp feature classification model aiming to support clinical screening improvements. Chen et al. [25] proposed a deep learning-based serrated polyp classification system using ordinary white-light endoscopy images, demonstrating promising diagnostic performance.

F. Surveys, Reviews, and Clinical Studies

Several recent survey and review studies have summarized advances on this topic. Qayoom et al. [26] presented a detailed review about the challenges, methodologies, and future directions of medical polyp segmentation. Mei et al. [27] provided a survey study on deep learning-based polyp segmentation with an emphasis on transformer-based architectures. Mameli et al. [23], [28] designed DeepPolyp, a framework for real-time clinical deployment and benchmarking of transformer-based segmentation models. Xu and He [29] explored the effect of clinical working hours on missed diagnosis rates of colorectal polyps to bring into light the demand for AI-assisted systems to minimize human errors.

TABLE I: Comparative Analysis of Existing Techniques on Kvasir-based Polyp Analysis

Ref	Year	Task	Method	Remarks
[2]	2019	Seg.	Kvasir-SEG	Benchmark segmentation dataset
[3]	2020	Multi-task	HyperKvasir	Large-scale GI dataset
[4]	2023	Det.+Seg.	PolypGen	Strong generalization
[5]	2022	Seg.	Graft-U-Net	Improved skip connections
[6]	2020	Seg.	DDANet	Dual-decoder attention
[7]	2024	Seg.	EffiSegNet	Lightweight EfficientNet
[8]	2022	Seg.	Semi-supervised CNN	Reduced annotation cost
[9]	2023	Seg.	PolySeg Plus	Active learning based
[10]	2024	Seg.	DPE-Net	Dual encoder structure
[11]	2023	Seg.	ResPVT	Transformer backbone
[12]	2024	Seg.	Wavelet AdaptUNet	Multi-scale features
[15]	2024	Seg.	Synthetic Polyp Gen.	Data augmentation
[16]	2025	Det.	YOLOv11	Real-time detection
[20]	2023	Det.	YOLOv8	High FPS, good accuracy
[21]	2024	Det.+Seg.	Multi-task CNN	Joint learning approach
[22]	2021	Class.	Deep CNN	GI disease classification
[24]	2025	Class.	NICE-based DL	Clinical feature-based
[25]	2024	Class.	Serrated Polyp Classifier	Specialized classification
[23]	2024	Det.+Class.	Attention ConvMixer	Lightweight hybrid model

G. *Summary and Research Gaps*

The reviewed literature has shown that deep learning-based models significantly outperform traditional image processing techniques for the detection, segmentation, and classification of polyps. Nevertheless, cross-dataset generalization, real-time inference on low-resource hardware, limited labeled data, and explainability still remain some open challenges. These limitations motivate the need for robust, efficient, and clinically deployable systems.

III. Related Dataset

Figure 1 For this work, we use the Kvasir dataset-a publicly available dataset consisting of a collection of gastrointestinal tract endoscopic images that are expert-annotated. The images in this original Kvasir dataset were captured in real procedures from a Norwegian hospital, and it includes all important anatomical landmarks, pathological findings, and therapeutic procedures. Concretely, there are 8 classes: three for anatomical landmarks, namely, Z-line, pylorus, and cecum; three relating to pathological findings: esophagitis, polyps, and ulcerative colitis; and two classes of procedures performed in the video, namely "dyed and lifted polyp" and "dyed resection margins." The dataset was originally composed of 4,000 images, according to original documentation-400 images per class, balanced across all classes. It supports different resolutions, ranging from 720 × 576 to 1920 × 1072 pixels; therefore, it is appropriate for a broad range of computer vision tasks. Kvasir provides a reliable benchmark for training and evaluating machine-learning or deep-learning models for disease detection, classification, and endoscopic image retrieval, since all images are labeled and verified by expert endoscopists. Besides, this availability of sample images for each class, as depicted above, ensures the transparency and reproducibility of the research outcomes. Beyond that original dataset, there is a greater family of datasets associated with Kvasir, including segmentation extensions, which further enhance different use cases independent of but including segmentation, polyp detection, instrument detection, and many other advanced tasks. Thus, Kvasir is a benchmark dataset for medical image analysis in gastrointestinal endoscopy, a basic resource that enables the development and benchmarking of computer-aided diagnostic systems.

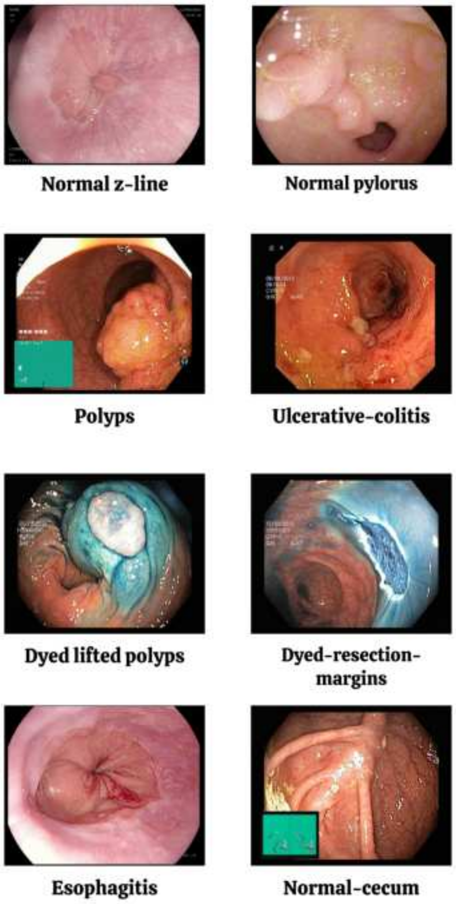

Fig. 1: 8 Classes in kvasir Dataset

IV. Methodology

A. *Research Design and Approach*

This study uses a quantitative experimental design based on the concept of deep learning as well as computer vision pertaining to the automatic diagnosis of gastrointestinal diseases via images from endoscopic searches. The study deals with the numerical evaluation of model performance measurements based on metrics such as accuracy, precision, recall, F1 score, and Dice coefficient.

A comparative modeling technique has been used, where four deep learning models have been designed, implemented, and tested under a controlled environment:

Model 1: Classification using spatial attention and explainability

Model 2: Strongly Supervised Hybrid Classification-Seg

Model 3: Multi-task learning with U²-Net-based strong supervision

Model 4: Fully optimized hybrid multitask architecture with improved regularization

This method is appropriate for the study aims because it allows the comparison of different architectures, management strategies, and explainability methods for the diagnosis of GI diseases.

B. Dataset and Sample Description

1) Dataset Description: For experimentation, the KVASIR dataset is used. It is a publicly available gastrointestinal endoscopy image database. It is comprised of labeled RGB images of various categories of gastrointestinal diseases. These images are normally obtained through endoscopic examinations.

Every image functions as an independent observation or a separate sample, making the data set suitable for supervised image classification and image segmentation using deep learning techniques.

2) Dataset Splitting Strategy: In order to avoid the leak of data and an unbiased evaluation, the class-wise stratified split was used:

Model 1:

- Training dataset: 80
- Validation dataset: 20

Models 2, 3, and 4:

- Training dataset: 70
- Validation set: 15
- Test set: 15

Splitting was carried out separately in each class to maintain the class distribution in each subset.

C. Data Collection Procedure

The process of data collection was as follows:

1) Extraction of Dataset: First, extract the KVASIR dataset from the downloadable compressed archive and organize it into class-specific directories.

2) Pre-processing of images: Each image was resized to 224 × 224 pixels and pre-processed by a multi-stage augmentation pipeline:

- Contrast stretching by CLAHE in LAB color space
- Noise reduction using Gaussian, median, and bilateral filtering
- Illumination normalization using gamma correction
- Original images with Canny edge detection superimposed highlighting lesion margins

3) Automatic Lesion Mask Generation (Models 2–4): Since pixel-level annotations are not available, the authors have used a pre-trained U²-Net model to generate lesion segmentation masks. The predicted probability maps were thresholded to binary masks that serve as pseudo ground truth for strong supervision.

4) Data Augmentation: Real-time augmentation was applied only to the training set, including:

- Random rotation
- Horizontal and vertical flipping
- Brightness and contrast adjustment
- Hue and saturation variation

Segmentations models were applied augmentations to both images and masks simultaneously to keep their spatial alignment.

D. Proposed Deep Learning Frameworks

1) Common Backbone Network: All models utilize EfficientNet-B0 pretrained on ImageNet as the feature extraction backbone owing to the high accuracy-to-parameter efficiency. Shallow layers were frozen to retain generic visual features, while deeper layers were fine-tuned for the recognition of GI disease.

2) Model-Specific Architectures:

a) Model 1: Classification with Spatial Attention:

- EfficientNet-B0 feature extraction backbone
- Spatial attention module applied to convolutional feature maps
- Global Average Pooling for reducing spatial dimensionality
- Fully connected layers with dropout regularization
- A softmax output for disease classification

b) Model 2: Strongly Supervised Hybrid Multi-Task Model:

- Shared EfficientNet encoder
- Segmentation decoder trained with masks generated by U^2-Net-
- Parallel classification branch
- The classification and segmentation are jointly optimized.

c) Model 3: Explicit Strong Supervision for Multi-Task Learning:

- Efficient encoder by using EfficientNet
- U-Net-like segmentation decoder
- Global pooling classification branch
- Segmentation-oriented end-to-end training

d) Model 4: Enhanced Hybrid Framework: Model 4 represents an optimized version of the hybrid multi-task architecture designed to enhance robustness and generalization. The model incorporates:

- Improved preprocessing, augmentation
- Strong regularization: dropout, L2, batch normalization
- Balanced multi-task loss weighting to harmonize classification and segmentation learning
- Improved Training Stability and Generalization

E. Training Strategy and Optimization

All models were trained with Adam optimizer with learning rates ranging from 1×10^{-5}, 1×10^{-4}, depending on model complexity.

Following are the callbacks that have been used:

- **Early stopping** to stop training when validation performance no longer improved
- **ReduceLROnPlateau** for adaptive learning rate adjustment
- **Model checkpointing** Saving the best-performing model

The models were optimized using task-specific loss functions:

- **Optimization Function:** Classification Loss- Categorical Cross-Entropy Loss
- **Segmentation loss:** Binary Cross-Entropy + Dice Loss

F. Data Analysis and Evaluation Metrics

1) Classification Evaluation: The performance of models is evaluated using different quantitative metrics, including:

- Overall classification accuracy
- Class-wise precision, recall, and F1-score
- Confusion matrix analysis to examine misclassifications patterns across disease categories

2) Segmentation Evaluation: Segmentation model performance was assessed using both quantitative and qualitative measures:

- Dice coefficient to evaluate spatial overlap between predicted and reference lesion masks
- Qualitative evaluation by visual inspection of the predicted masks.
- Lesion area and bounding box extraction

3) Explainability Analysis: Grad-CAM was applied to all the classification models in order to increase interpretability. Heatmaps were used to visualize discriminative regions that contribute to model decisions, allowing verifications about lesion-focused decision-making.

G. Methodology Framework Diagram

Figure 2 illustrates the detailed workflow of the proposed deep learning methodology.

V. Results and Outcomes

In this section, a detailed quantitative and qualitative assessment of the four deep learning models submitted for classifying gastrointestinal diseases using the KVASIR endoscopy database is presented. The models are tested for metrics such as classification accuracy, precision, recall, and F1 score. Furthermore, for segmentation-enabled models, metrics such as Dice score will also be calculated. A comparison analysis is carried out among the models to identify the best and most accurate method.

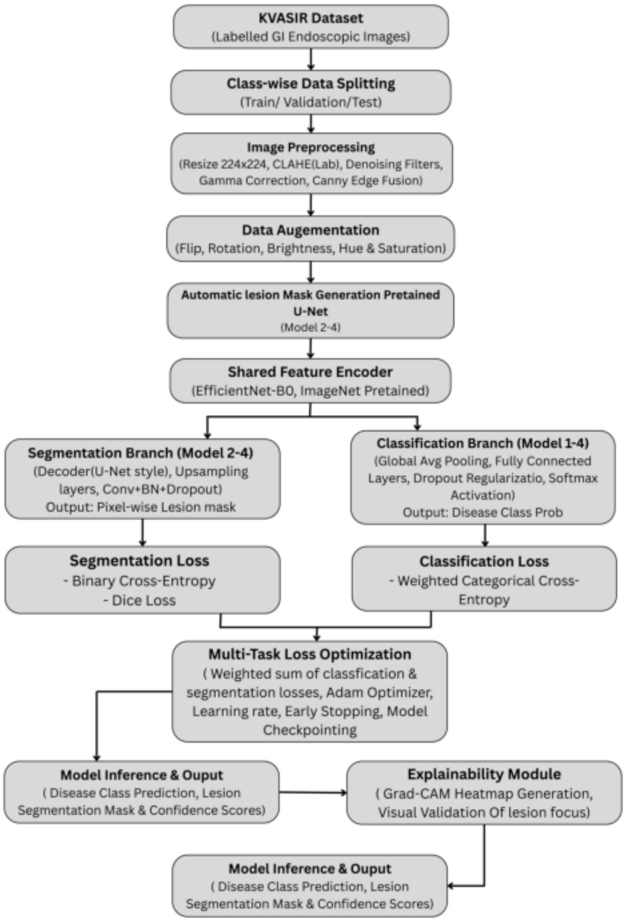
KVASIR Dataset
(Labelled GI Endoscopic Images)
Class-wise Data Splitting
(Train/ Validation/Test)
Image Preprocessing
(Resize 224x224, CLAHE(Lab), Denoising Filters, Gamma Correction, Canny Edge Fusion)
Data Augementation
(Flip, Rotation, Brightness, Hue & Saturation)
Automatic lesion Mask Generation Pretained U-Net
(Model 2-4)
Shared Feature Encoder
(EfficientNet-B0, ImageNet Pretained)
Segmentation Branch (Model 2-4)
(Decoder(U-Net style), Upsampling layers, Conv+BN+Dropout)
Output: Pixel-wise Lesion mask
Classification Branch (Model 1-4)
(Global Avg Pooling, Fully Connected Layers, Dropout Regularizatio, Softmax Activation)
Output: Disease Class Prob
Segmentation Loss
- Binary Cross-Entropy
- Dice Loss
Classification Loss
- Weighted Categorical Cross-Entropy
Multi-Task Loss Optimization
(Weighted sum of classfication & segmentation losses, Adam Optimizer, Learning rate, Early Stopping, Model Checkpointing
Model Inference & Ouput
(Disease Class Prediction, Lesion Segmentation Mask & Confidence Scores)
Explainability Module
(Grad-CAM Heatmap Generation, Visual Validation Of lesion focus)
Model Inference & Ouput
(Disease Class Prediction, Lesion Segmentation Mask & Confidence Scores)

Fig. 2: Detailed framework of the proposed hybrid deep learning methodology for gastrointestinal disease classification and lesion segmentation.

A. Training and Validation Performance

They demonstrate that all four models have absolutely stable convergence behavior: smooth increase in classification accuracy, a monotonic decrease of the training and validation loss.

Model 1 converged rapidly within 25 epochs with around 94% training accuracy and 90% validation accuracy. This close correspondence of the training and validation curves signifies a strong generalization and low overfitting.

Model 3 represented the best learning ability: training accuracy increased from 68.18% to 97.65%, and the validation accuracy tended to stabilize within the range from 88% to 90%, reaching its maximum value of 90.17%. This means effective feature learning is balancing with stable generalization.

By comparison, Model 2 and Model 4 realized relatively stable but lower validation accuracy, about 89%, and longer training schedules brought up less performance gain relative to their architectural complexity.

Focus Justification (Model 1 and Model 3): Model 1 and Model 3 have better convergence efficiency and validation ability compared to the other models and have good generalization. The models 2 and 4 have good convergence ability compared to the other models and poor generalization.

B. *Overall Classification Performance*

Table II summarizes the overall classification accuracy obtained by all models on the independent test set.

TABLE II: Overall Test Accuracy Comparison of the Evaluated Models

Model	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Test Accuracy (%)	90.00	89.17	**90.17**	88.50

Figure 3 Model 3 produced the best test accuracy, where it successfully classified 541 out of 600 test images, seconded by Model 1, which produced an accuracy of 90%. Both models showed excellent robustness in all eight categories of gastrointestinal disease.

Focus Justification (Model 1 and Model 3): Model 1 and Model 3 perform better than Models 2 and 4 regarding overall classification accuracy with simpler and more efficient training workflows.

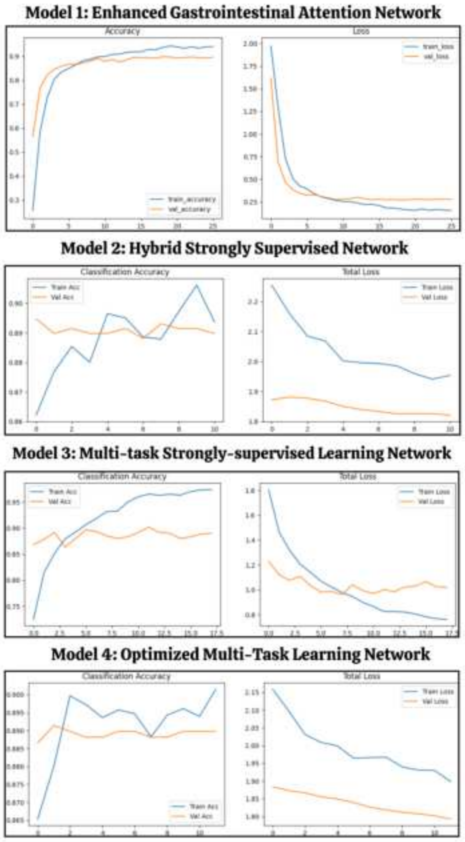

Fig. 3: Models Performance and Loss Graph

C. *Class-wise Performance Analysis*

Table III ClassName-wise analysis showed parallel trends for all the models.

Model 1 performed very well on the normal anatomical classes such as normal-cecum (100%) and normal-pylorus (96%), and also on the pathological classes such as ulcerative colitis (94%) and polyps (92%).

Model 3 showed very close accuracy results, with normal-pylorus (98.67%), normal-cecum (97.33%), and ulcerative-colitis (94.

In all models, there was a tendency for lower accuracy in visually similar categories like esophagitis and normal Z-line, which are inherently ambiguous in endoscopic images.

Focus Justification (Model 1 and Model 3): Model 1 and Model 3 perform comparatively better in both normal and pathological classes, while variability in Models 2 and 4 is higher in visually ambiguous classes.

TABLE III: Class-wise Prediction Accuracy of MSL-Net (Model 3) on the Test Set

Class	**Accuracy (%)**	**Count (Correct / Total)**
Dyed-lifted polyps	88.00	66 / 75
Dyed-resection margins	84.00	63 / 75
Esophagitis	81.33	61 / 75
Normal-cecum	97.33	73 / 75
Normal-pylorus	98.67	74 / 75
Normal Z-line	89.33	67 / 75
Polyps	88.00	66 / 75
Ulcerative colitis	94.67	71 / 75

D. *Confusion Matrix and Misclassification Analysis*

Figure 4 By confusion matrix analysis, high values on the diagonals can be observed in all models, which justify correct prediction rates.

In the case of Model 1 and Model 3, most mapping errors happened between the “dyed-lifted polyps” and “dyed-resection margins” classes, and between the “normal Z-line” class and the “esophagitis” class.

Lowest Rate of Misclassification: The model that recorded the lowest level of misclassification error is Model 3, followed by Model 2 and

Focus Justification (Model 1 and Model 3):

Model 1 and Model 3 show lower and more understandable misclassification, which accentuates the robustness and diagnosis integrity of these models.

E. *Segmentation and Localization Performance*

Models 2, 3, and 4 use the lesion segmentation branch based on the pseudo ground truth masks produced by

Of these models, Model 3 performed the best on segmentation, with a validation set Dice score above 0.92. Secondly, Model 2 and Model 4 followed with a slight edge, their validation set Dice score being above 0.91.

The greatly improved segmentation results of Model 3 led to better lesion boundary delineation and attention in the spatial domain, thereby promoting the improvement of classification accuracy.

Focus Justification (Model 1 and Model 3): Model 3 integrates segmentation and classification best, whereas Model 1 gives comparable results without the segmentation cost.

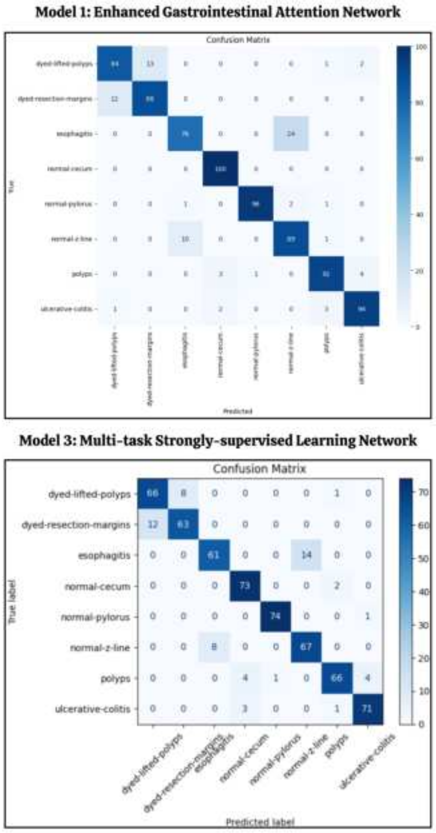

Fig. 4: Confusion matrix of Model 1 and 3 on the test set

Model 2 and Model 4 add extra complexity without improving corresponding performance.

F. Explainability and Visual Interpretation

Figure 5 Grad-CAM visualizations for Model 1 and Model 3 consistently highlight clinically relevant lesion regions, confirming model predictions are driven by pathological rather than by background artifacts.

With Model 3, the added interpretability by incorporating segmentation masks explicitly localizes lesion regions even in some misclassified cases, hence indicating correct spatial attention despite class ambiguity.

Focus Justification (Model 1 and Model 3): Model 1 uses attention-based classification, which would make the model explainable; Model 3 explicitly locates lesions to improve transparency for both models, which would be more appropriate for clinical interpretation.

G. Outcome Summary and Model Selection Rationale

Based on the experimental results, the most effective models for the classification of gastrointestinal diseases are found to be Model 1 and Model 3.

Model 1 has high classification accuracy (90%), excellent generalization capabilities, and a very simple model structure

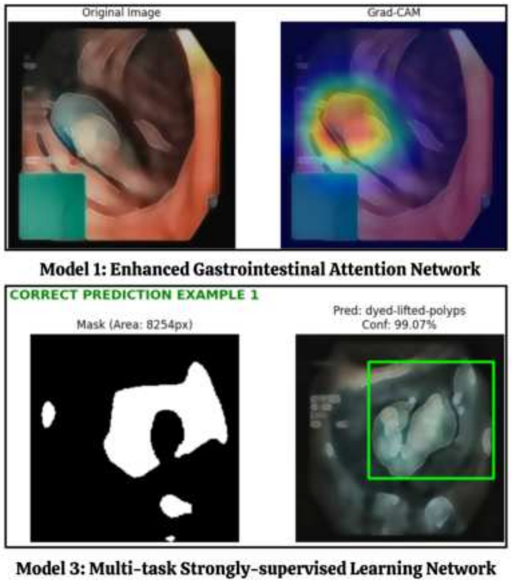
Original Image
Grad-CAM
Model 1: Enhanced Gastrointestinal Attention Network
CORRECT PREDICTION EXAMPLE 1
Mask (Area: 8254px)
Pred: dyed-lifted-polyps
Conf: 99.07%
Model 3: Multi-task Strongly-supervised Learning Network

Fig. 5: Grad-CAM visualization for Model 1 and 3

Model 3 performs best with the highest test accuracy of 90.17% along with the best lesion localization.

Even though **Model 2** and **Model 4** show stability in performance, the added complexity in architecture and slightly lower accuracy make these models less optimal for detailed analysis and implementation. On the basis of the trade-off between accuracy, robustness, interpretability, and computational efficiency, the most suitable candidates for clinical decision support systems, and extensions for future research, are revealed to be Models 1 and 3.

VI. Conclusion

In this study, we presented and discussed four different deep learning models for automated classifications of gastrointestinal diseases from KVASIR endoscopies. Various models employed a variety of different techniques, ranging from highly complex pre-processing, attention mechanisms, to multi-task learning and heavily-supervised segmentations.

From the experimental results, Model 1, EGA-Net was found to have higher effectiveness and accuracy than Models 2, HSS-Net; and Model 3, MSL-Net has been found to be more accurate than Model 4, OMTL-Net. Although Model 1 has been able to achieve a test accuracy of 90

In terms of analysis by class, it is found that the performance of Model 1 and Model 3 is good for the detection of normal anatomy and clinically significant lesions with only a slight drop of performance for the visually ambiguous classes such as esophagitis and normal Z-line. This is because the optimization of classification and localization tasks for Model 3 enabled the improvement of feature learning.

On the whole, this work has proved that advanced pre-processing, attention-based feature learning, and multi-task optimization combined lead to very positive changes regarding accuracy and interpretability of classification models of gastrointestinal diseases. EGA-Net and MSL-Net proposed models guarantee high accuracy and clinical validity that can be used for developing computer-aided diagnosis systems, which may help endoscopists decrease diagnostic mistakes and improve treatment outcomes accordingly.

Future studies will focus on working with larger datasets, allowing for a consideration of endoscopy video temporal data, as well as real-time inference, to optimize applicability to a clinical setting.

References

[1] A. Sharma, "One shot joint colocalization and cosegmentation," 2017. [Online]. Available: https://arxiv.org/abs/1705.06000

[2] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-seg: A segmented polyp dataset," 2019. [Online]. Available: https://arxiv.org/abs/1911.07069

[3] H. Borgli, V. Thambawita, P. Smedsrud, S. Hicks, D. Jha, S. Eskeland, K. Randel, K. Pogorelov, M. Lux, D. T. Dang Nguyen, D. Johansen, C. Griwodz, H. Stensland, E. Garcia Ceja, P. Schmidt, H. Hammer, M. Riegler, P. Halvorsen, and T. de Lange, "Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Scientific Data*, vol. 7, 08 2020.

[4] S. Ali, D. Jha, N. Ghatwary, S. Realdon, R. Cannizzaro, O. Salem, D. Lamarque, C. Daul, M. Riegler, K. V. Ånonsen, A. Petlund, P. Halvorsen, J. Rittscher, T. de Lange, and J. East, "A multi-centre polyp detection and segmentation dataset for generalisability assessment," *Scientific Data*, vol. 10, p. 19, 02 2023.

[5] M. Ramzan, M. Raza, M. Sharif, and S. Kadry, "Gastrointestinal tract polyp anomaly segmentation on colonoscopy images using graft-u-net," *Journal of Personalized Medicine*, vol. 12, p. 1459, 09 2022.

[6] N. K. Tomar, D. Jha, S. Ali, H. D. Johansen, D. Johansen, M. A. Riegler, and P. Halvorsen, "Ddanet: Dual decoder attention network for automatic polyp segmentation," 2020. [Online]. Available: https://arxiv.org/abs/2012.15245

[7] I. A. Vezakis, K. Georgas, D. Fotiadis, and G. K. Matsopoulos, "Effisegnet: Gastrointestinal polyp segmentation through a pre-trained efficientnet-based network with a simplified decoder," 2024. [Online]. Available: https://arxiv.org/abs/2407.16298

[8] H.-C. Park, S. Poudel, R. Ghimire, and S.-W. Lee, "Polyp segmentation with consistency training and continuous update of pseudo-label," *Scientific Reports*, vol. 12, 08 2022.

[9] A. Saad, F. Maghraby, and O. Badawy, "Polyseg plus: Polyp segmentation using deep learning with cost effective active learning," *International Journal of Computational Intelligence Systems*, vol. 16, 09 2023.

[10] M. A. Manan, F. Jinchao, S. Ahmed, and A. Raheem, "Dpe-net: Dual-parallel encoder based network for semantic segmentation of polyps," in *2024 9th International Conference on Signal and Image Processing (ICSIP)*. IEEE, Jul. 2024, p. 790–794. [Online]. Available: http://dx.doi.org/10.1109/ICSIP61881.2024.10671533

[11] R. Nachmani, I. Nidal, D. Robinson, M. Yassin, and D. Abookasis, "Segmentation of polyps based on pyramid vision transformers and residual block for real-time endoscopy imaging," *Journal of Pathology Informatics*, vol. 14, p. 100197, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2153353923000111

[12] D. Rajasekar, G. Theja, M. R. Prusty, and S. Chinara, "Efficient colorectal polyp segmentation using wavelet transformation and adaptunet: A hybrid u-net," *Heliyon*, vol. 10, no. 13, p. e33655, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405844024096865

[13] D. Rajasekar, G. Theja, M. Prusty, and S. Chinara, "Efficient colorectal polyp segmentation using wavelet transformation and adaptunet: A hybrid u-net," *Heliyon*, vol. 10, p. e33655, 06 2024.

[14] J. Bernal, G. Fernández, A. García-Rodríguez, and F. Sánchez, *Polyp Segmentation in Colonoscopy Images*, 07 2021, pp. 151–154.

[15] H. Song and Y. Shin, "Semantic polyp generation for improving polyp segmentation performance," *Journal of Medical and Biological Engineering*, vol. 44, pp. 1–13, 03 2024.

[16] A. R. Sahoo, S. S. Sahoo, and P. Chakraborty, "Polyp detection in colonoscopy images using yolov11," 2025. [Online]. Available: https://arxiv.org/abs/2501.09051

[17] ——, "Polyp detection in colonoscopy images using yolov11," 2025. [Online]. Available: https://arxiv.org/abs/2501.09051

[18] K. ELKarazle, V. Raman, P. H. H. Then, and C. Chua, "Detection of colorectal polyps from colonoscopy using machine learning: A survey on modern techniques," *Sensors (Basel, Switzerland)*, vol. 23, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:256217392

[19] M. Shen, C.-C. Huang, Y.-T. Chen, Y.-J. Tsai, F.-M. Liou, S.-C. Chang, and N. Phan, "Deep learning empowers endoscopic detection and polyps classification: A multiple-hospital study," *Diagnostics*, vol. 13, p. 1473, 04 2023.

[20] M. Lalinia and A. Sahafi, "Colorectal polyp detection in colonoscopy images using yolo-v8 network," *Signal, Image and Video Processing*, vol. 18, pp. 1–12, 12 2023.

[21] J. Abbas, Z. Abidin, R. Naqvi, and S.-W. Lee, "Unmasking colorectal cancer: A high-performance semantic network for polyp and surgical instrument segmentation," *Engineering Applications of Artificial Intelligence*, vol. 138, 09 2024.

[22] M. Al-Adhaileh, E. Senan, W. Alsaade, T. Aldhyani, N. Alsharif, A. Alqarni, M. I. Uddin, M. Alzahrani, E. Alzain, and M. Jadhav, "Deep learning algorithms for detection and classification of gastrointestinal diseases," *Complexity*, vol. 2021, 10 2021.

[23] A. Demirbas, H. Üzen, and H. Fırat, "Spatial-attention convmixer architecture for classification and detection of gastrointestinal diseases using the kvasir dataset," *Health Information Science and Systems*, vol. 12, 04 2024.

[24] T. Carvalho, R. Kader, P. Brandao, L. Lovat, P. Mountney, and D. Stoyanov, "Nice polyp feature classification for colonoscopy screening," *International Journal of Computer Assisted Radiology and Surgery*, vol. 20, 03 2025.

[25] C. Tsung-Hsing, Y.-T. Wang, C.-H. Wu, C.-F. Kuo, H.-T. Cheng, S.-W. Huang, and C. Lee, "A colonial serrated polyp classification model using white-light ordinary endoscopy images with an artificial intelligence model and tensorflow chart," *BMC Gastroenterology*, vol. 24, 03 2024.

[26] A. Qayoom, J. Xie, and H. Ali, "Polyp segmentation in medical imaging: challenges, approaches and future directions," *Artificial Intelligence Review*, vol. 58, 03 2025.

[27] J. Mei, T. Zhou, K. Huang, Y. Zhang, Y. Wu, Y. Zhou, and H. Fu, "A survey on deep learning for polyp segmentation: techniques, challenges and future trends," *Visual Intelligence*, vol. 3, pp. 1–20, 12 2025.

[28] M. Mameli, S. Shiralizadeh, M. Papi, and I. G. Coltea, "Deeppolyp: an artificial intelligence framework for polyp detection and segmentation," *Exploration of Digital Health Technologies*, vol. 3, 2025. [Online]. Available: https://www.explorationpub.com/Journals/edht/Article/101158

[29] Q. Xu and Z. He, "Effect of different working periods on missed diagnosis of colorectal polyps in colonoscopy," *BMC Gastroenterology*, vol. 24, 08 2024.