

Enhanced GI Tract Cancer Diagnosis Using CNNs and Machine Learning Models

Abdul Haseeb
0009-0002-6069-1467
Capital University of Science and
Technology, Islamabad
Expressway,
Kahuta Road, Zone-V Islamabad.
Email: abdul.haseeb@cust.edu.pk

Faheem Shehzad
0009-0003-7204-183X
Department of Electrical
Engineering and Information
Technology (DIETI),
University of Naples “Federico II”,
Naples, Italy.
Email: faheem.shehzad@unina.it

Sidra Naseem
Department of Computer Science,
COMSATS University Islamabad,
Vehari Campus, Mailsi Road,
Off Multan Road,
Vehari, Punjab, Pakistan
Email:
sidrafaheem228@gmail.com

Abstract—Gastrointestinal cancer exhibits the greatest mortality rate among all cancers, at 35.4%. Endoscopy is one of the few methods for obtaining visuals of gastrointestinal tract lesions. Manual cancer detection is arduous. Deep learning can autonomously diagnose gastrointestinal tract lesions. Automation produces erroneous detection results. This study used the challenging Hyper-Kvasir dataset for training and validation purposes. The dataset undergoes first preprocessing with Brightness Preserving Histogram Equalization. Furthermore, processed datasets comprise training and validation sets. For segmentation, pretrained backbone-based U-Net architecture is used. The U-Net backbones include EfficientNet-B0, EfficientNet-B7, and DenseNet201. The pre-trained models utilize ImageNet, so Hyper-Kvasir is employed for the fine-tuning of gastrointestinal tract segmentation. The optimal Intersection over Union (IoU) is 85.2% for the EfficientNet-B7 backbone inside the U-Net design. A custom convolutional neural network is employed to classify the hyper kvasir dataset. The suggested network derives profound features for classification using artificial neural networks. The proposed methodology surpassed state-of-the-art (SOTA) methods.

Index Terms—Gastrointestinal Cancer, U-Net, CNNs, Deep learning, Classification, Segmentation.

I. INTRODUCTION

GASTROINTESTINAL (GI) cancer occurs when malignant cells grow inside GI tract [1]. In the past fifty years, GI cancer has had the second highest mortality rates among different types of cancers [2]. According to global cancer index, GI cancer has the mortality rate of 35.4% whereas 26.3% of all the cancers are diagnosed as GI tract cancers. Over the past few years, the performance of artificial intelligence-driven computer-aided diagnosis (CAD) tools in various medical fields has been greatly improved by deep learning algorithms [3], particularly artificial neural networks (ANNs) [4]. Identifying gastrointestinal (GI) illnesses subjectively takes time and professional competence. By automating the detection and categorization of GI illnesses, computer-assisted diagnosis (CAD) technology may reduce these diagnostic obstacles. Such technologies might help doctors detect and cure serious medical diseases early

on. Medical practitioners benefit from CAD technology's precise diagnosis and appropriate action [5]. Deep learning (DL) [6] are statistical based methodologies that authorize computer systems to sovereignly identify patterns and properties from unprocessed data inputs, including structured data, images, text, and audio. The substantial progress in artificial intelligence (AI) based on DL has had a profound impact on numerous domains within clinical practice [7, 8].

In semantic segmentation, every pixel of an object is assigned a specific label corresponding to its class. This process involves categorizing each pixel in an image into predetermined classes. Semantic segmentation relies on the concept of a mask that incorporates edge detection, which helps identify the connected regions in an image that belong to the same class [9]. For the purposes of semantic segmentation, multiple architectures are being used by researchers. Recent studies shows that one of few most effective frameworks for image segmentation in medical domain is U-Net [10].

However, using state-of-the-art pretrained models like U-Net for semantic segmentation may provide considerable results. The design has two partitions: contraction (Encoder) and dilation (Decoder). To get image context, convolutional and pooling layers are used. While the latter half spreads the picture utilizing skip connections and anti-convolution (transpose convolution). The segmented image is the outcome. U-Net is a semantic segmentation benchmark. It improves outcomes with numerous fundamental architectural modifications. Several pretrained networks compose the U-Net architecture's backbone, improving performance. One of the best semantic segmentation designs is U-Net. This design is largely utilized in medical image segmentation.

This study introduces a deep learning methodology for segmenting gastrointestinal (GI) tract lesions in endoscopic images, employing a U-Net architecture with several pretrained models (EfficientNet-B0, EfficientNet-B7, and DenseNet201) utilized as fixed encoders. The decoding component employs the encoded information for precise segmentation. Brightness Preserving Histogram Equalization

(BPHE) is utilized as a preprocessing technique to improve image quality. A unique convolutional neural network is employed for feature extraction and classification utilizing the Hyper-Kvasir dataset. The efficacy of each model is assessed by several criteria and juxtaposed with one another and contemporary state-of-the-art techniques. The document is organized to encompass background, relevant research, methods, findings, and conclusions with prospective directions.

II. LITERATURE REVIEW

This section describes the recent advancements in GI tract segmentation. Researchers achieved significant results for GI cancer segmentation. In [11], authors used U-Net model with depth of five to perform semantic segmentation on hyper Kvasir dataset. Moreover, researchers used the image size of 96×96 at the start of the model. Additionally, in the encoder part the gradual decrease in size is observed till the size of the image becomes 6×6 . Furthermore, convolutional layers with filter size of 3×3 is used whereas filter size for pooling layers is 2×2 . Authors used loss score as evaluation parameter for results and obtained value of 0.69 for loss. A new pipeline for unsupervised domain adaptation (UDA) for the purpose of semantic segmentation is introduced in [12], that combines feature-level adaptation with image-level adaption. To address domain shifts at the image-level, the proposed approach includes a global photometric alignment and global texture alignment modules that are used to align images from the source domains and target domains based on their image-level properties. A global manifold alignment approach is used for feature-level domain change by mapping pixel features from the two domains onto the source domain's feature manifold. Additionally, desired domain consistency regularization is carried out on enhanced target domain images, and category centers in the source domain

are regularized using a class-oriented triplet loss. On hyper kvasir dataset authors achieved 81.5% mean IoU.

As described by Nguyen Thanh Duc et al. [13], polyps in the colorectal region can be detected rapidly and accurately using novel deep learning algorithms. To recognize lesions in colonoscopy images, the authors proposed Colon-Former, a deep learning architecture employing an encoder-decoder architecture. The encoder is a lightweight and efficient modeling framework for multi-scale global semantic connections. The decoder is a representation of visual data generated by a hierarchical network that has been enhanced. Five distinct reference datasets were used to construct the proposed system. This paradigm describes multiscale functionalities via transformers and convolutional neural networks (CNNs). It only supports one architecture and utilizes data from five unique collections. In addition, we observed that it produced the finest results compared to other methods, leading us to conclude that it is a cutting-edge method.

Experts in the field of colonoscopy image analysis emphasized the importance of image segmentation for the detection of lesions caused by colorectal cancer [14]. For image segmentation, regional dense-pixel classification and boundary-based polygon algorithms have previously been established. Using a graphical neural network (GNN) that is based on a deep neural network, the authors developed a novel polyp detection methodology. This technique identifies the polyp area using an attention-enhancement module (AEM). Using the AEM, border and area characteristics of polyps can be extracted. Each plot is data-driven, so the GNN, which functions as a weighted link between the nodes of various domains, preserves the global and local connections between the nodes. It focuses on the demographics and geographic boundaries of the region. The GNN outperforms competing methods and accurately detects malignant lesions in colonoscopy-acquired biomedical images. However, the complexity of the system makes it challenging to precisely

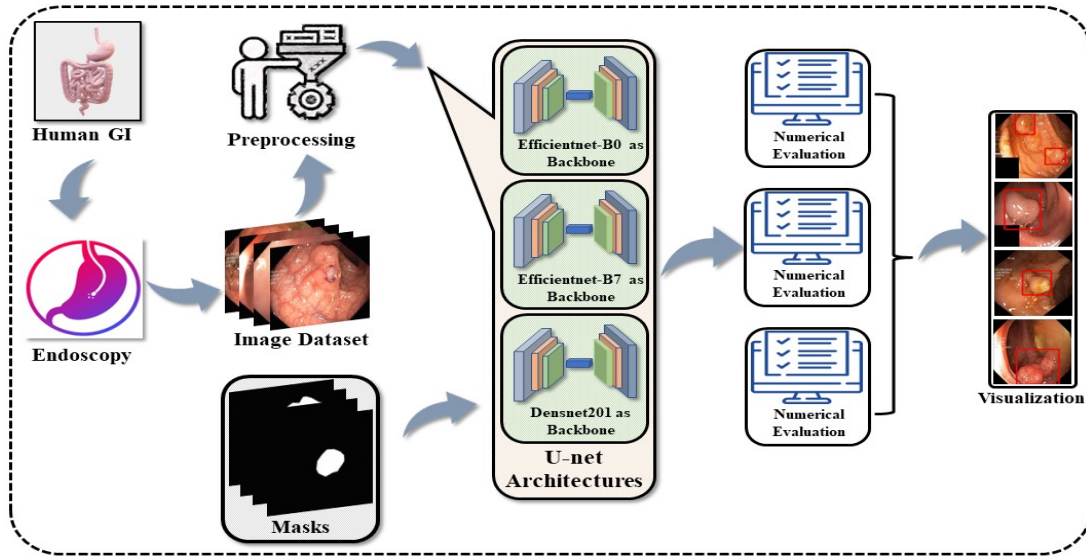


Figure 1: System Model for semantic segmentation of GI tract lesions.

identify the polyp region. Based on comparisons to other innovative methods, this GNN system is the superior model. The polyps could only be identified using specialized equipment.

III. METHODOLOGY

The proposed methodology comprised of multiple steps. In the first step the dataset is pre-processed to enhance the image quality. To achieve this purpose, histogram equalization technique is used to improve the spatial quality of image. The second step of the methodology is to feed these images with their corresponding masks to multiple U-Net based on different backbones. Evaluation of the system is completed using different evaluation parameters. In the last step the system is tested and obtained the visual representation of segmented region of GI tract cancer. Fig. 1 shows the methodology for the segmentation of GI tract lesion.

A. Dataset

Segmentation Dataset: This study explores with hyper kvasir segmentation dataset [19]. The collection comprises 1000 endoscopic images and masks from various GI tract locations from numerous people. Due of the dataset's inconsistent image sizes, images and masks are resized. The final size is 256x256 after resizing. Additionally, the dataset has two subgroups. Subset one has training images and subset two testing images. The training subset includes 80% of the data and the testing subset 20%. Randomization is considered while splitting data. Fig. 2 shows dataset examples of images and masks.

Classification Dataset: Additionally, hyper Kvasir classification dataset with 23 classes is utilized for classification. Due to class imbalance in dataset, dataset is enhanced with images to address this issue. These methods modify the spatial features of dataset images without changing their orientation. Dataset has training, validation, and testing subsets.

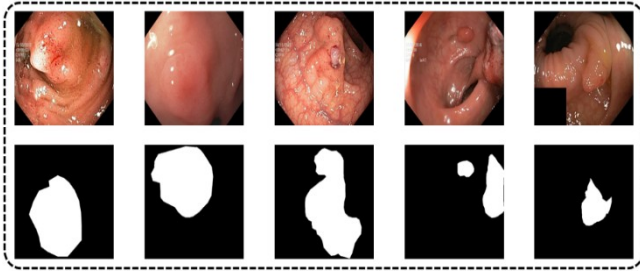


Figure 2: Hyper-Kvasir dataset sample images and masks

B. Data Pre-Processing

To achieve better results, one of the most common phenomena used is pre-processing. Brightness Preserving Histogram Equalization (BPHE) [15] is one of the efficient histogram equalizations to enhance the images. In the discipline of image processing, contrast is enhanced using brightness-preserving histogram equalization. Adjusting the image's histogram so that the intensity levels are distributed more

equitably is one method to improve the image's perceived quality. In contrast to conventional histogram equalization techniques, preserving bi-histogram equalization considers both the light and dark regions of an input image. It adjusts the histograms of each separately to increase contrast while preserving detail in the highlights and shadows. This method excels in situations where, maintaining the detail in both lighter and darker areas is essential, such as medical imaging. Preprocessing is applied only to classification dataset.

C. Convolutional Neural Networks

CNNs have emerged as a useful instrument for analyzing medical images in recent years [16]. If a neural network has at least one convolutional layer, we refer to it as a convolutional neural network (CNN). A convolution operation uses a sliding window technique to apply a fixed-size filter with multiple parameters to an input image. When a layer is complete, the resulting image is sent to the subsequent layer. Here is the mathematical expression for this process:

$$FM_{out}(H_{out} \times V_{out}) = (FM_{inp} * Filter_{op}) \quad (1)$$

The output matrix FM_{out} comprises the rows and columns designated H_{out} and V_{out} as shown in equation (1). Using the rectified linear unit function, the value of a negative feature is set to zero, as shown in the following equation.

$$Active_{ReLU} = \text{Maxof}(0, k), k \in FM_{out} \quad (2)$$

In addition, an aggregating technique is employed to reduce computational complexity and accelerate processing time. This procedure involves exchanging the input value at the center with the utmost or average value in a particular region. Using an entirely linked layer, the features are then transformed into a one-dimensional vector. In mathematical notation, it appears as follows:

$$(Vect_{flat})_0^{out} = FM_{out}(H_{out} \times V_{out}) \quad (3)$$

$$(Vect_{flat})_i^{in} = (Vect_{flat})_{i-1}^{out} * M_i + Vert_i \quad (4)$$

$$(Vect_{flat})_i^{out} = \Delta_i((Vect_{flat})_i^{in}) \quad (5)$$

In above equations, $(Vect_{flat})_0^{out}$ is flattened final one-dimensional vector, i is the layer number. Moreover, Δ shows the activation function used in the operation.

D. U-Net Architecture

U-Net architecture [10] is a popular CNN for image segmentation. U-Net was named because it resembles the letter U when diagrammed. The network's architecture may gain local and global context due to its encoding (contracting) and decoding (expanding) paths. The U-Net design excels in organ and tumor segmentation in biomedical imaging [17]. In addition to semantic and instance segmentation, this approach has been used for many additional segmentation tasks. U-Net may segment using low-level and high-level characteristics due to skip connections. Keep fine-grained

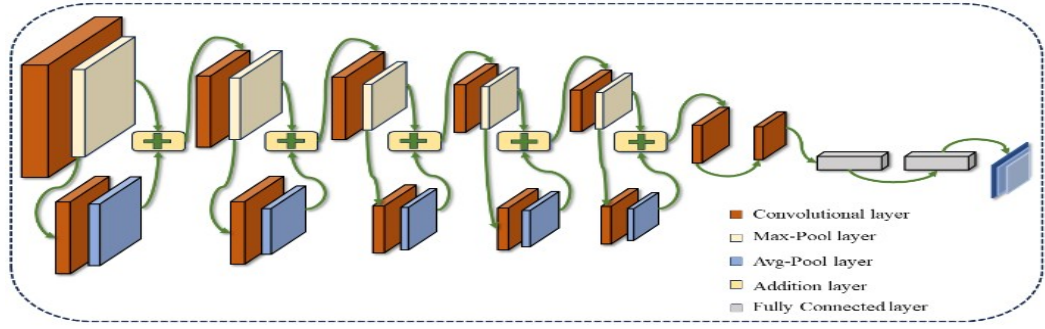


Figure 3: Custom 24 layered proposed architecture for hyper kvasir classification

information while collecting context for more accurate segmentation. This article uses three pretrained networks as U-Net backbones for segmentation. The study uses Densenet201 [18], Efficientnetb0 [19], and Efficientnetb7 [20].

E. Backbones Used for U-Net Architecture

U-Net architecture is a widely used framework for semantic segmentation. U-Net uses encoder and decoder structures to perform semantic segmentation. On the other hand, the Efficientnet models are some of the best models that is been used by researchers in different domains. Combining both networks can be robust and accurate in terms of results. In this paper, Efficientnet-B0, Efficientnet-B7 and Densenet201 are used as backbones. Pre-trained weights of ImageNet dataset are used as weights for encoder part. Skip connections are used from encoder part to decoder path. These skips connections combine the features from encoder part with the decoder part features. This increases the accuracy of segmentation by preserving the spatial information.

The most important concept that is used in U-Net architecture is deconvolution. The purpose of deconvolution is to find a specific solution for any convolution. This is achieved by using the following equation.

$$H_{sol} = (FM_{inp} * Filter_{op}) + \epsilon \quad (6)$$

In above equation, FM_{inp} is the input image while $Filter_{op}$ is the filter used for convolution. Moreover, ϵ is the noise that is added due to convolution operation and $*$ is used for convolution operation.

F. Proposed 24 Layered Architecture

To obtain the classification results pre-trained models are often used by implementing transfer learning techniques. Pre-trained models are trained on “ImageNet” dataset having millions of images categorized in 1000 classes. To achieve results, pre-trained models are fine-tuned on target dataset using transfer learning. The technique freezes the weights of all layers except the last few layers. Transfer learning is used to solve a wide range of problems in deep learning domain. However, there is major drawback in transfer learning technique which is referred to as domain mismatch problem. To resolve this, custom deep architectures are developed by researchers. However, the results with custom models are not

up to the mark. In this article a custom model with 24 layers is proposed through which features are extracted. Moreover, these features are used as input for artificial neural networks for classification.

Custom designed model consists of 24 layers combining 12 convolutional layers, 5 max_pooling layers, 5 average_pooling layers and 2 dense layers. 5 residual blocks are used in the architecture having one convolutional layer and one average pooling layer each. Similarly, linear blocks consist of one convolutional layer and one max pooling layer. Fig. 3 shows the model architecture diagram. Each convolutional layer is activated using relu activation function. However, the last convolutional layer is activated through soft plus function. Both linear and residual blocks are combined using addition operation. Network also have one input layer which takes the images having the dimension $224 \times 224 \times 3$ as input. Furthermore, a classification output layer is also included at the end of the network.

G. Training and validation

Hyper-kvasir classification dataset containing 24000 images is used for the training and validation of the proposed network. Dataset is divided into training and validation sets with 70% training and 30% validation data. Training is performed on the system having windows 10 and Nvidia RTX 3060 GPU. MATLAB R2022A is used as a programming platform. The hyper parameters used for training are “Learning Rate = 0.0001”, “Batch Size = 16”, “Optimizer = Adam”.

IV. RESULTS AND DISCUSSION

In this section the results achieved through the several experiments for segmentation and classification purposes are described. To achieve this Densenet201, Efficientnet-b0 and Efficientnet-b7 are used as backbone in application of U-Net architecture for semantic segmentation. Furthermore, boundary box for affected area is created to identify the lesion in endoscopy images of GI tract. Moreover, for classification custom network is designed and obtained the results.

A. System Setup

The experiments are performed using a system having core i7 processor with four cores and eight threads and 16 GB of RAM. Moreover, the Nvidia GTX 950M graphical

TABLE I.
EVALUATIONS FOR GI TRACT SEGMENTATION USING EFFICIENTNET-B7

Epoch	Validation Accuracy	Validation Dice Coef	Validation IoU	Validation Loss	Validation Precision	Validation Recall
1	0.885135	0.471594	0.313928	0.46715	0.610926	0.916307
10	0.958906	0.83735	0.721798	0.101103	0.911861	0.86235
20	0.962712	0.891788	0.806815	0.109419	0.924289	0.868229
30	0.964142	0.90777	0.834153	0.121391	0.924001	0.877824
40	0.963541	0.909911	0.837617	0.125882	0.918826	0.883313
50	0.96515	0.918313	0.852109	0.140427	0.924259	0.889357

TABLE II.
EVALUATIONS FOR GI TRACT SEGMENTATION USING DENSENET201

Epoch	Validation Accuracy	Validation Dice Coef	Validation IoU	Validation Loss	Validation Precision	Validation Recall
1	0.776002	0.403931	0.257738	0.647087	0.419642	0.943514
10	0.957872	0.82911	0.71018	0.122115	0.930014	0.828044
20	0.958587	0.871834	0.774337	0.137972	0.939886	0.820215
30	0.957271	0.884891	0.796025	0.160757	0.911674	0.843068
40	0.961018	0.894244	0.810935	0.118371	0.921321	0.853872
50	0.96024	0.901343	0.823115	0.140625	0.909625	0.870043

processing unit with 4 GB of VRAM is used for training purposes. All the experiments for segmentation are performed using Python 3.10 and TensorFlow. For training purposes, initial learning rate is set to the value of 0.0001, max epochs are set to 50 whereas batch size is set to the value of 8.

B. Results for Segmentation

Numerical Results are provided through multiple performance measures implemented. Table I shows the gradual depiction of results during the validation of the framework. To obtain the results Efficientnet-B7 is used as backbone for segmentation. IoU is the most important performance measure for semantic segmentation. Results shows that using Efficientnet-B7, the best validation IoU value of 0.85 is obtained while 96% validation accuracy is achieved. Additionally, the table shows gradual increase in validation accuracy, validation dice coefficient, validation precision and validation recall. Moreover, the loss is at lowest as the epochs increase. It is observed that from epoch 1 to 10 the increase in validation accuracy, dice coefficient, IoU, precision and recall values increase significantly whereas the validation loss decreases abruptly. However, the change in the evaluation parameters is less for over 10 epochs as compared to the values before 10 epochs. Another discrepancy is observed during the epochs 30 to 40 which is the sudden decrease in validation accuracy. Moreover, the improvement in evaluation parameters is significant between epoch number 40 to 50.

The best values achieved at the end of the validation are 0.96 for accuracy, 0.91 for dice coefficient, 0.85 for IoU, 0.92 for precision and 0.88 for recall. Additionally, the best loss value at the end of the validation is 0.14.

Table II demonstrates Densenet201-segmented GI tract lesions. Densenet201's maximum validation IoU is 0.82 and accuracy is 96%. Dice coefficient for validation data is 0.90. Increased epochs cause some disruption in steady evolution of outcomes. Compared to Efficientnet, Dense-net201 did not produce smooth results. Analyzing the evaluation parameters shows that the values vary exponentially in the first 10 epochs and less after 10 epochs. Densenet201-based framework training and validation showed considerable inconsistency between epochs 30 and 35. Best values after validation are 0.96 accuracy, 0.90 dice coefficient, 0.82 IoU, 0.90 precision, and 0.87 recall. However, validation loss is 0.

Additionally, Efficientnet-B0 is also used in the study for experiments to segment GI lesion detection. Table III describes that the highest value of validation IoU obtained through Efficientnet-B0 is 0.78 whereas the value of validation accuracy achieved by the framework is 95%. By analyzing the values, it is assessed that the highest validation accuracy achieved at the end of validation process is 0.95. Moreover, for dice coefficient the value is 0.87. Similarly, the value of validation IoU is 0.78. Also, the achieved value for validation precision is 0.93 and the value for validation recall is 0.81. In case of Efficientnet-B0 it is observed that

TABLE III.
EVALUATIONS FOR GI TRACT SEGMENTATION USING EFFICIENTNET-B0

Epoch	Validation Accuracy	Validation Dice Coef	Validation IoU	Validation Loss	Validation Precision	Validation Recall
1	0.776002	0.403931	0.257738	0.647087	0.419642	0.943514
10	0.957872	0.82911	0.71018	0.122115	0.930014	0.828044
20	0.958587	0.871834	0.774337	0.137972	0.939886	0.820215
30	0.957271	0.884891	0.796025	0.160757	0.911674	0.843068
40	0.961018	0.894244	0.810935	0.118371	0.921321	0.853872
50	0.96024	0.901343	0.823115	0.140625	0.909625	0.870043

there is slight gradual decrease in the value of accuracy, dice coefficient, IoU, precision and recall after the 30th epoch. The validation loss also increased after the 30th epoch and achieved the lowest value of 0.17 at the end of the validation process.

Visualization of segmented areas and corresponding bounding boxes are shown in Fig. 4. The first image shows the original image from hyper-kvasir dataset (Validation Data) whereas the second image shows the original masks given with the hyper-kvasir dataset. In the third image, the mask predicted by the Efficientnet-B7 is depicted. Finally, the image with bounding box (corresponds to the predicted mask) is achieved. The figure illustrates that the network predicted the GI tract lesions with phenomenal accuracy.

In table IV proposed system's results are compared with the state-of-the-art (SOTA) previous works. Researchers used different techniques including U-Net, DeeplabV3+, Transformers etc. to segment the GI tract lesion segmentation. It is analyzed that the proposed system outperforms the previous techniques.

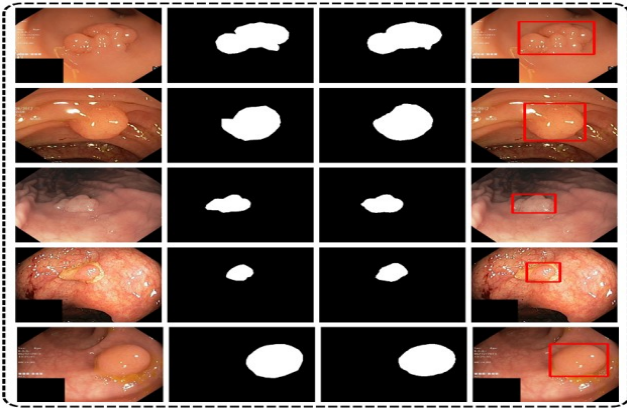


Figure 4: Visualizations of segmentation using U-Net with Efficientnet-B7 backbone: left to right: original image, original annotation, predicted annotation and bounding box around predicted lesions

TABLE IV.

COMPARISON OF SEGMENTATION RESULTS FOR PROPOSED METHODOLOGY WITH STATE-OF-THE-ART (SOTA)

Reference	Dataset	Technique	Results
[11]	Hyper-Kvasir	U-Net	Loss = 0.69
[12]	Hyper-Kvasir + Piccolo	DeeplabV3+	IoU = 0.84
[13]	Kvasir	ColonFormer-L (Transformers)	IoU = 0.87
[21]	Endocv2022 + CVC-Clinics	Improved-STCN Network	Dice = 0.76
[22]	Hyper-Kvasir	MSACL	IoU = 0.40
Proposed	Hyper-Kvasir	-	IoU = 0.85

C. Result for Classification

Custom network design for feature extraction purpose is used to extract the features from test dataset. The extracted feature vector is further fed to multiple Artificial Neural Network (ANNs) classifiers to achieve the best results for classification. Table V comprises accuracies, precisions, recalls and F1-Scores for all classifiers used in experiments. By analyzing the table, it is clear that the Narrow Neural Network classifier gives the best overall results with 92.70% accuracy. The values for precision, recall and F1-score are 92.87, 92.78 and 92.80 respectively. On the other hand, Tri-layered Neural Network has given the worst overall performance with 89.00% accuracy. Fig. 5 depicts the confusion matrix for classification of hyper-kvasir dataset using Narrow Neural Network classifier. Generally, the model performed exceptionally well. Yet some classes still have low classification accuracies which shows that class imbalance problem can alter the model's result drastically in clinical

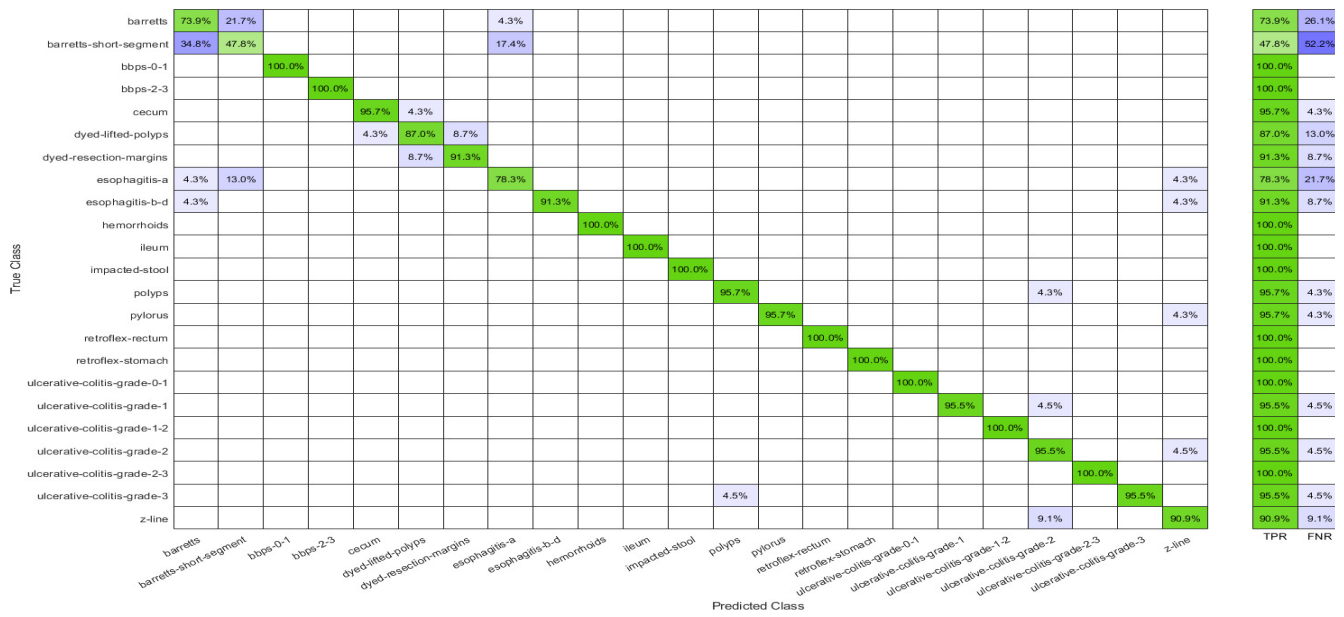


Figure 5: Confusion matrix of narrow neural network for hyper-kvasir dataset classification

environment. Moreover, noisy data can also be a hurdle in acquiring the accurate results in clinical experimentations.

TABLE V

CLASSIFICATION RESULTS FOR ARTIFICIAL NEURAL NETWORK CLASSIFIERS FOR HYPER-KVASIR CLASSIFICATION

Classifier	Accuracy	Precision	Recall	F1 Score	Time
Narrow Neural Network	92.70	92.87	92.78	92.80	39.50
Medium Neural Network	91.50	92.20	91.64	91.90	39.80
Wide Neural Network	92.30	92.56	92.36	92.45	87.27
Bilayered Neural Network	89.20	89.67	89.30	89.48	43.27
Trilayered Neural Network	89.00	89.53	89.13	89.32	51.70

Custom network design for feature extraction purpose is used to extract the features from test dataset. The extracted feature vector is further fed to multiple Artificial Neural Network (ANNs) classifiers to achieve the best results for classification. Table V comprises accuracies, precisions, recalls and F1-Scores for all classifiers used in experiments. By analyzing the table, it is clear that the Narrow Neural Network classifier gives the best overall results with 92.70% accuracy. The values for precision, recall and F1-score are 92.87, 92.78 and 92.80 respectively. On the other hand, Trilayered Neural Network has given the worst overall performance with 89.00% accuracy. Fig. 5 depicts the confusion matrix for classification of hyper-kvasir dataset using Narrow Neural Network classifier. Generally, the

model performed exceptionally well. Yet some classes still have low classification accuracies which shows that class imbalance problem can alter the model's result drastically in clinical environment. Moreover, noisy data can also be a hurdle in acquiring the accurate results in clinical experimentations.

TABLE VI

CLASSIFICATION RESULTS FOR ARTIFICIAL NEURAL NETWORK CLASSIFIERS FOR HYPER-KVASIR CLASSIFICATION

Classifier	Accuracy	Precision	Recall	F1 Score	Time
Narrow Neural Network	92.70	92.87	92.78	92.80	39.50
Medium Neural Network	91.50	92.20	91.64	91.90	39.80
Wide Neural Network	92.30	92.56	92.36	92.45	87.27
Bilayered Neural Network	89.20	89.67	89.30	89.48	43.27
Trilayered Neural Network	89.00	89.53	89.13	89.32	51.70

Table VI exhibits the comparison of the proposed technique with recent literature. By analyzing the results, it is clear that the proposed method outperforms the state-of-the-art (SOTA) with significant margins.

V. CONCLUSION

World-wide, GI tract cancer is frequent. This study introduces U-Net topologies with Efficientnet-B0, B7, and Densnet201 backbones to identify GI lesions. Model training

TABLE VII

COMPARISON OF RESULTS WITH STATE-OF-THE-ART (SOTA)

Reference	Dataset	Classes	Year	Accuracy (%)
[23]	Hyper-Kvasir	6	2023	87.45
[24]	Hyper-Kvasir	14	2020	73.66
[25]	Kvasir	5	2021	97.00
[26]	Hyper-Kvasir	23	2020	63.00 for macro
Proposed	Hyper-Kvasir	23	-	92.87

and validation employ Hyper-Kvasir segmentation dataset. Endoscopic photos of malignant regions are labeled by medical professionals in the collection. Brightness Preserving Histogram Equalization improves photos. To train models, the proposed U-Net with Efficientnet-B0, B7, and Densenet201 backbones receives enhanced pictures and masks. Custom deep models are used to classify and extract features. These characteristics also feed artificial neural networks for categorization. The paper also discusses validation data outcomes. The suggested strategy outperforms state-of-the-art methods. Visual differences in photos prevented certain models from performing as expected.

REFERENCES

- [1] M. Vajihinejad, "A Systematic Review of Clinic Pathology and Survival in Gastrointestinal Stromal Tumors," *International Journal of New Chemistry*, vol. 10, no. 1, pp. 33-61, 2023.
- [2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *Ca Cancer J Clin*, vol. 68, no. 6, pp. 394-424, 2018.
- [3] S. Aamir, A. Rahim, S. Bashir, and M. Naeem, "Prediction of Breast Cancer Using AI-Based Methods," in *Intelligent Environments 2021*: IOS Press, 2021, pp. 213-220.
- [4] H. Ko *et al.*, "COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image: model development and validation," *Journal of medical Internet research*, vol. 22, no. 6, p. e19569, 2020.
- [5] M. Owais, M. Arsalan, J. Choi, T. Mahmood, and K. R. Park, "Artificial intelligence-based classification of multiple gastrointestinal diseases using endoscopy videos for clinical diagnosis," *Journal of clinical medicine*, vol. 8, no. 7, p. 986, 2019.
- [6] H. Qayyum, S. T. H. Rizvi, M. Naeem, U. b. Khalid, M. Abbas, and A. Coronato, "Enhancing Diagnostic Accuracy for Skin Cancer and COVID-19 Detection: A Comparative Study Using a Stacked Ensemble Method," *Technologies*, vol. 12, no. 9, p. 142, 2024.
- [7] K. Sumiyama, T. Futakuchi, S. Kamba, H. Matsui, and N. Tamai, "Artificial intelligence in endoscopy: Present and future perspectives," *Digestive Endoscopy*, vol. 33, no. 2, pp. 218-230, 2021.
- [8] A. Ismail, M. Naeem, M. H. Syed, M. Abbas, and A. Coronato, "Advancing Patient Care with an Intelligent and Personalized Medication Engagement System," *Information*, vol. 15, no. 10, p. 609, 2024.
- [9] J. Cheng, H. Li, D. Li, S. Hua, and V. S. Sheng, "A Survey on Image Semantic Segmentation Using Deep Learning Techniques," *Computers, Materials and Continua*, vol. 74, no. 1, pp. 1941-1957, 2023.
- [10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 2015: Springer, pp. 234-241.
- [11] A. S. Narasimha Raju, K. Jayavel, and T. Rajalakshmi, "Dexterous Identification of Carcinoma through ColoRectalCADx with Dichotomous Fusion CNN and UNet Semantic Segmentation," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [12] H. Ma, X. Lin, and Y. Yu, "I2F: A Unified Image-to-Feature Approach for Domain Adaptive Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [13] N. T. Duc, N. T. Oanh, N. T. Thuy, T. M. Triet, and V. S. Dinh, "ColonFormer: an efficient transformer based method for colon polyp segmentation," *IEEE Access*, vol. 10, pp. 80575-80586, 2022.
- [14] Y. Meng *et al.*, "Graph-based region and boundary aggregation for biomedical image segmentation," *IEEE transactions on medical imaging*, vol. 41, no. 3, pp. 690-701, 2021.
- [15] Y.-T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE transactions on Consumer Electronics*, vol. 43, no. 1, pp. 1-8, 1997.
- [16] F. Shehzad *et al.*, "Two-stream deep learning architecture-based human action recognition," 2023.
- [17] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *Ieee Access*, vol. 9, pp. 82031-82057, 2021.
- [18] S. Jégou, M. Drozdal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 11-19.
- [19] W. Mulim, M. F. Revikasha, and N. Hanafiah, "Waste Classification Using EfficientNet-B0," in *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*, 2021, vol. 1: IEEE, pp. 253-257.
- [20] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019: PMLR, pp. 6105-6114.
- [21] Q. He, X. Hu, F. Sun, L. Zhou, J. Wang, and Q. Wan, "Improved-STCN Network with Enhanced Strategy for Sequence Polyp Segmentation," 2022.
- [22] Y. Tian *et al.*, "Self-supervised multi-class pre-training for unsupervised anomaly detection and segmentation in medical images," *arXiv preprint arXiv:2109.01303*, 2021.
- [23] T. Nguyen-DP, M. Luong, M. Kaaniche, J. Chaussard, and A. Beghdadi, "Self-supervised Learning for Gastrointestinal Pathologies Endoscopy Image Classification with Triplet Loss," *arXiv preprint arXiv:2303.01672*, 2023.
- [24] P. H. Smedsrud *et al.*, "Kvasir-Capsule, a video capsule endoscopy dataset," *Scientific Data*, vol. 8, no. 1, p. 142, 2021.
- [25] M. Hmoud Al-Adhaileh *et al.*, "Deep learning algorithms for detection and classification of gastrointestinal diseases," *Complexity*, vol. 2021, pp. 1-12, 2021.
- [26] H. Borgli *et al.*, "HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Scientific data*, vol. 7, no. 1, p. 283, 2020.