Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

Automated Detection and Classification of Gastrointestinal Diseases Using Transfer Learning

Nitun Kumar Podder
Dept. of Computer Science & Engineering
Pabna University of Science and Technology
Pabna, Bangladesh
E-mail: nituncse@gmail.com

Md. Sanzid Olioul Tanzid
Dept. of Computer Science & Engineering
Pabna University of Science and Technology
Pabna, Bangladesh
E-mail: sanzid94@gmail.com

Tania Sultana
Dept. of Computer Science & Engineering
Pabna University of Science and Technology
Pabna, Bangladesh
E-mail: taniasultanaanni@gmail.com

***Abstract*— Cancer is the most common cause of death nowadays. About 5 million new patients are affected, and 2 million patients die of gastrointestinal diseases worldwide. Till now, video endoscopy is an improved medical imaging technology, and this modality today has gained popularity for the diagnosis of gastrointestinal problems that include bleeding, stomach ulceration, and polyps. This technology uses a flexible, lighted tube to take high-definition visualization and real-time images of the digestive system for close examination and exact diagnosis of several gastrointestinal problems. Continuous monitoring of endoscopic images by an expert is the primary issue with video endoscopy, as the procedure lasts around 30 minutes, which can be costly and prone to human error. This paper presents a comprehensive analysis of deep learning models to detect and classify 5 types of gastrointestinal conditions automatically. The study evaluates the performance of various convolutional neural networks (CNNs), including VGG19, Resnet152V2, DenseNet201, InceptionResNetV2, and InceptionResNetV2, and a combination of DenseNet201 and ResNet152V2. Each model is trained using transfer learning with fine-tuning techniques optimized for accuracy and loss. The results demonstrate the effectiveness of deep learning in enhancing diagnostic accuracy, with the combination of DenseNet201 and ResNet152V2 performing better than other models with accuracy of 95.50%. This work underscores the potential of AI in clinical settings for early disease detection and improved patient outcomes.**

Keywords*— Deep learning, convolutional neural networks, gastrointestinal diseases, image classification, automated detection, transfer learning, fine-tuning.

I. INTRODUCTION

Cancer is a very common disease nowadays. It can be cured by early detection. Gastrointestinal cancer is the most commonly occurring type of cancer. The early detection of gastrointestinal diseases is important for effective patient management and treatment because the treatment of early-stage cancer increases the success rate.

Gastrointestinal cancer originates from gastrointestinal polyps, which are abnormal tissue growths on the mucosa of the stomach and colon. The growth of polyps is slow, and symptoms only manifest when they become large. However, early detection can prevent and cure polyps. Video endoscopy plays a crucial role in enhancing the early diagnosis of gastrointestinal polyps and reducing mortality rates. Endoscopy can assess the severity of ulcerative colitis by identifying mucosal patterns, including spatial variations in mucosal color and texture (the level of roughness on the mucosal surface in the gastrointestinal tract). A gastrointestinal video can yield hundreds of images, but only a few images show signs of disease, and it is impractical for medical professionals to dedicate the necessary time to review all the images.

Traditionally, diagnosing these conditions relies on manual examination of endoscopic images by healthcare professionals, which can be time-consuming and prone to human error. Recent advancements in artificial intelligence (AI) and machine learning, particularly Convolutional Neural Networks (CNNs), present an opportunity to automate this process and enhance diagnostic accuracy.

This study aims to explore the efficiency of various deep-learning models in the classification of gastrointestinal diseases. We focus on a comprehensive dataset that includes 5 categories of diseases (dyed lifted polyps, normal cecum, normal pylorus, polyps, and ulcerative colitis.). The primary contribution of this paper is the introduction of a computer-aided detection approach for lower gastrointestinal conditions. This method involves adjusting the criteria for extracting deep shape, color, and texture features and then applying a learning transfer technique for finely tuned and contoured transfers. Extensive experiments were carried out to identify pre-trained models for diagnosing lower gastrointestinal diseases. Additionally, new models were created to transfer features extracted from a non-medical deep learning dataset and adapt them to the new dataset.

II. METHODOLOGY

A. Dataset Description

The Vestre Viken Health Trust (VV) in Norway gathered the dataset, specifically from the gastroenterology department at Baerum Hospital, using endoscopic equipment. Experts from VV and the Cancer Registry of Norway (CRN) annotated all the images. The CRN, based at Oslo University Hospital, is responsible for screening and early cancer detection to prevent its spread.

The Kvasir dataset comprises interpreted images, featuring classes related to endoscopic procedures in the gastrointestinal tract and anatomical landmarks. It includes hundreds of images suitable for deep learning and transfer

learning. The images are in RGB color space and have resolutions ranging from 720×576 to 1920×1072 pixels with 8 categories of 8000 images with each containing 1000 images. In our study, the dataset contains 5,000 images equally distributed across five categories (dyed lifted polyps, normal cecum, normal pylorus, polyps, and ulcerative colitis).

Figure 1 displays samples from the Kvasir dataset. The dataset link: https://datasets.simula.no/kvasir/#download

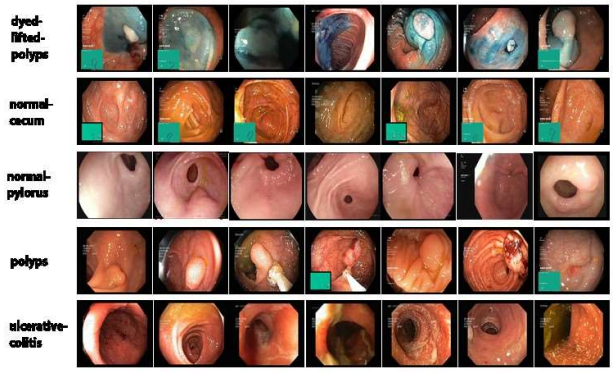

Fig 1: Dataset sample image of 5 categories of Gastrointestinal Diseases

B. Data Preprocessing

To train our deep learning model on the Kvasir dataset, we set up a thorough data pre-processing pipeline which involved normalization and data augmentation. We rescaled each image to a range of [0, 1] by dividing pixel values by 255 to enable quicker convergence. Various augmentation techniques, such as a shear range of (±20%), a zoom range of (±20%), a rotation range of (±30°), and random horizontal shift (±20%), were applied to improve the model's resilience to variations in image orientation and scale. Furthermore, 20% of the dataset was kept aside for validation to ensure accurate performance assessment of new data. These pre-processing steps collectively enhanced the quality and diversity of the dataset, leading to more effective detection of gastrointestinal diseases.

C. Model Architecture

We employed several pre-trained CNN architectures, each fine-tuned for optimal performance on our dataset. The models are as follows:

VGG19: VGG19 is a deep convolutional neural network model proposed by Simonyan and Zisserman in 2014 for large-scale image classification tasks. It consists of 19 layers, including 16 convolutional layers and 3 fully connected layers, and follows an extremely simple and uniform architecture of stacking 3×3 convolution filters. The VGG19 model has been trained on the ImageNet dataset, which has more than one million images for 1,000 object categories. The model achieved a top-5 accuracy of 92.7 on the ImageNet validation set, proving that it is very good at image classification. Because of its simplicity and depth, VGG19 has been used as a baseline in many transfer learning tasks. The uniformity of the model allows easy extension for different architectures or fine-tuning for new tasks.

ResNet152V2: ResNet152V2 is a deep residual network introduced by Kaiming He et al. in 2016, which would address the problem of vanishing gradients in very deep networks. The ResNet architecture first introduced residual connections so that the gradients can easily flow through the network. It is an improved version with optimized batch normalization and activation functions and includes 152 layers. The ResNet152V2 model has been trained on the ImageNet dataset and reaches one of the state-of-the-art results for image classification, 96.43% top-5 accuracy on the ImageNet validation. Due to residual connections, it is an effective deep network and does not face over-fitting problems when being trained. It has 152 layers and has become a standard for tasks that require very deep architectures due to strong performance and efficient training.

InceptionResNetV2: The inceptionResNetV2 model was proposed by Szegedy et al. in 2016, which combines the benefits of both the Inception architecture and the residual connections of ResNet. The Inception module extracts features at multiple scales using parallel convolutions with different kernel sizes, while the residual connections from ResNet enable better gradient flow through the network. InceptionResNetV2 is a network of 164 layers and was trained on the ImageNet dataset. It achieved 96.6% top-5 accuracy on the ImageNet validation set, which provides an excellent trade-off between model size, depth, and performance. The model is well-suited for large-scale image classification tasks and has been widely adopted for transfer learning and fine-tuning applications in various domains. Its efficient architecture enables faster training with high performance, especially in complex image recognition tasks.

DenseNet201: DenseNet201 is a deep convolutional network proposed by Huang et al. in 2017, which introduced dense connections between layers, meaning each layer receives input from all previous layers. This dense connectivity pattern helps to alleviate the vanishing gradient problem and improves feature reuse. DenseNet201 has 201 layers and was trained on the ImageNet dataset, achieving a top-5 accuracy of 96.4 on the ImageNet validation set. With dense connectivity, DenseNet201 enjoys fewer parameters than traditional architectures such as ResNet and VGG, which adds to its efficiency while preserving high performance. The signature unit of the model, a dense block, applies the concatenation of the input of each layer with the outputs of all previous ones, which leads to feature reusing and increases the representational power of the model. In addition, DenseNet201 has proved to be quite effective for image classification and transfer learning tasks.

Each model was initialized with pre-trained weights from the ImageNet dataset, followed by fine-tuning on our specific dataset.

D. Training Process

In this study, we utilized the pre-trained models VGG19, ResNet152V2, InceptionResNetV2, and DenseNet201, excluding their top fully connected layers, to adapt them for our specific dataset. The convolutional base of each model served as a feature extractor. We perform fine-tuning by unfreezing selected layers, allowing the deeper layers to adapt to the new task while preserving the low-level features learned from the ImageNet dataset. Specifically, the following layers were unfrozen for fine-tuning:

- *block5_conv1* for **VGG19**,
- *block8_4_conv* for **InceptionResNetV2**,
- *conv5_block1_preact_bn* **forInceptionResNetV2**,
- *conv5_block1_0_bn* for **DenseNet201**

To extend each model for multi-class classification, we appended the following layers:

- A flattened layer,
- A dense hidden layer with 256 units and ReLU activation,
- A final output layer with sigmoid activation.

The models were trained using the RMSprop optimizer with a low learning rate of 1×10^{5} to ensure minimal disruption to the pre-trained weights. A custom learning rate scheduler was employed to decrease learning rate by 10^{-1} at each epoch after the 5th epoch. To keep the learning rate constant for the first five epochs, followed by exponential decay. Early stopping was implemented to terminate training if the validation loss did not improve for four consecutive epochs.

We used a batch size of 32 and additionally, model check-pointing was utilized to save the best model based on validation accuracy. For the combination of the DenseNet201 and ResNet152V2 pre-trained model we follow as above-mentioned process and just combining two model features by using tensor-flow concatenate method.

These strategies, including selective fine-tuning, a carefully designed architecture, and efficient training techniques, ensured that the models generalized effectively to the new task while retaining valuable knowledge from the ImageNet pre-training. Training was conducted on a GPU to expedite the process.

III. Results Analysis

A. Model Evaluation

To evaluate the performance of our deep learning models, we employed the precision, recall, and F1 score metrics, which are standard measures in classification tasks. These metrics are derived from the confusion matrix of each model, specifically focusing on the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) [9]. Fig 2 shows our all-models confusion matrix. Precision, recall, and F1 scores are illustrated in Table 1.

In table1 we see that the combination of model DenseNet201 and ResNet152V2 contains the height metrics values. And DenseNet201 contains second height metrics values.

Each of the five models (VGG19, Resnet152V2, InceptionResNetV2, DenseNet201, and a combination of DenseNet201 and ResNet152V2) was evaluated for the automated detection and classification of gastrointestinal diseases. All models are trained many times by changing their parameters and fine-tuned by unfreezing the last few layers. Among the models' combination of DenseNet201 and ResNet152V2 performs better. The model results are summarized in Table 2

Table 1: Model average metrics values

Model	Precision	Recall	F1 Score
VGG19	91.91	91.93	91.90
Resnet152V2	92.62	92.65	92.61
InceptionResNetV2	92.49	92.53	92.49
DenseNet201	93.52	93.49	93.50
DenseNet201 and ResNet152V2	95.52	95.51	95.51

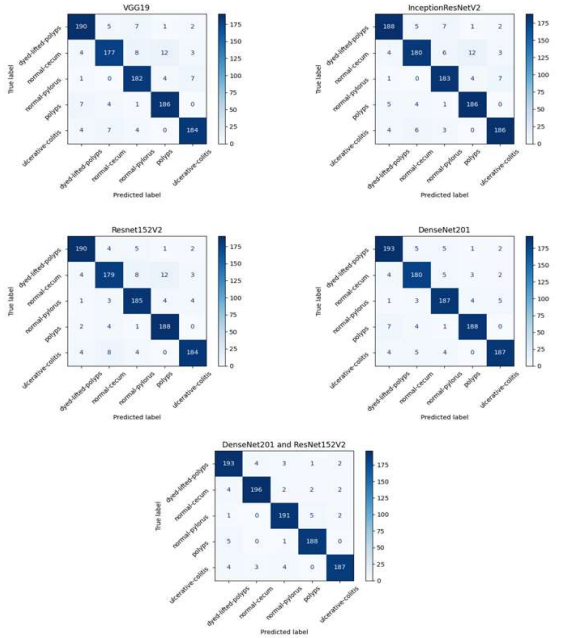

Fig 2: Confusion matrix of all model

Table 2: Model accuracy and loss function

Model	Training Accuracy %	Training Loss	Validation Accuracy %	Validation Loss
VGG19	92.10	0.2035	91.90	2509
Resnet-152V2	92.95	0.1956	92.60	0.2254
Inception-ResNetV2	93.62	1849	92.30	2349
DenseNet-201	94.66	0.1772	93.50	0.2494
DenseNet-201 and ResNet-152V2	97.60	0.0736	95.50	0.1789

B. Training History Visualization

The training and validation performance were visualized through plots depicting accuracy and loss over epochs. These visualizations illustrate the learning curves, aiding in the identification of potential over-fitting or under-fitting. Fig 3 presents the accuracy curves, while Fig 4 displays the loss curves. All models are trained in less than 10 epochs.

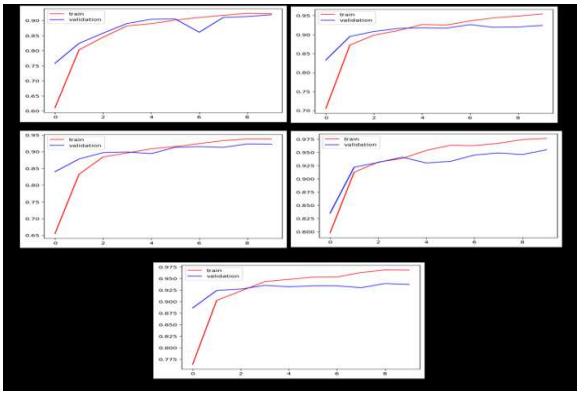

Fig 3: All models train and validation accuracy

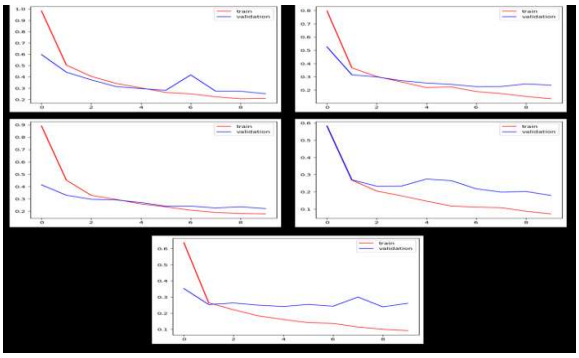

Fig 4: All models loss function

IV. DISCUSSION AND CONCLUSION

The results indicate that the combination of DenseNet201 and ResNet152V2 outperforms other models in terms of accuracy, making them suitable candidates for clinical applications in gastrointestinal disease detection. The combination of two different pertained models increases the efficiency of the model. The use of transfer learning significantly contributed to the models' high performance by leveraging pre-trained weights and learned features from a large dataset.

While the models showed promising results, challenges remain in ensuring the generalizability of these models to different populations and clinical settings. Future work should focus on testing these models on external datasets and exploring additional augmentation techniques to further improve robustness.

This study highlights the potential of deep learning models in the automated detection and classification of gastrointestinal diseases. By demonstrating the effectiveness of various CNN architectures, we emphasize the role of AI in enhancing diagnostic accuracy and patient care. By combining different pre-trained models the efficiency increased, further research will be necessary to implement these models in real-world clinical environments and to refine their performance through continuous learning and adaptation.

V. DISCUSSION AND CONCLUSION

[1] M. S. T. Rao, "Gastrointestinal diseases: A comprehensive review," *Journal of Clinical Gastroenterology*, vol. 25, no. 6, pp. 350-365, 2018.

[2] S. Kumar, R. Verma, and V. Sharma, "Gastritis and peptic ulcer disease: Pathophysiology and clinical management," *Indian Journal of Gastroenterology*, vol. 31, no. 4, pp. 148-157, 2019.

[3] S. K. Singal, "Hepatitis C virus infection and its impact on liver cirrhosis," *Journal of Hepatology*, vol. 47, no. 3, pp. 299-306, 2017.

[4] M. B. Lichtenstein, M. T. Farkas, and L. S. Furu, "Functional gastrointestinal disorders in children," *Pediatric Gastroenterology*, vol. 21, no. 2, pp. 104-112, 2019.

[5] N. R. Gill, "Gastroesophageal reflux disease: Current therapeutic strategies," *American Journal of Gastroenterology*, vol. 43, no. 1, pp. 56-64, 2018.

[6] P. H. Lee, "Irritable bowel syndrome: Pathophysiology and management strategies," *Gastroenterology Clinics of North America*, vol. 38, no. 4, pp. 523-537, 2020.

[7] A. H. George and J. E. Sutherland, "Chronic diarrhea: Etiology and management," *Digestive Diseases and Sciences*, vol. 44, no. 2, pp. 231-245, 2017.

[8] J. P. McLean, "Celiac disease: Diagnosis and management," *Gastroenterology Nursing*, vol. 36, no. 6, pp. 456-461, 2019.

[9] F. L. Williams and T. A. Olsen, "Crohn's disease: An overview and management options," *World Journal of Gastroenterology*, vol. 27, no. 1, pp. 115-123, 2021.

[10] D. S. Patel, "Gastric cancer and its association with Helicobacter pylori," *Journal of Cancer Research & Clinical Oncology*, vol. 144, no. 4, pp. 867-875, 2020